Measuring Segregation*

David M. Frankel
Iowa State University

Oscar Volij
Iowa State University
and Ben Gurion University

May 5, 2008

*Email addresses: dfrankel@econ.iastate.edu; ovolij@bgu.ac.il. We thank seminar participants at University of the Basque Country, Brown University, University of Haifa, University of Iowa, Iowa State University, Rice University, Tel Aviv University, the 2007 Annual Meetings of the American Real Estate and Urban Economics Association, and the Murat Sertel Memorial Conference on Economic Theory. Volij thanks the Spanish Ministerio de Educación y Ciencia (project SEJ2006-05455) for research support.

Measuring Segregation

Abstract

We propose a set of axioms for the measurement of multigroup school segregation. They are motivated by two criteria: do ethnic groups have similar distributions across schools? And are schools ethnically representative of their district? Our axioms are satisfied by a unique ordering. It is represented by the Mutual Information index. This index, originally proposed by Henri Theil, has a more intuitive decomposition than other indices. As an application, we find that segregation between districts within cities accounts for 33% of school segregation. Segregation across states, driven mainly by the distinct residential patterns of Hispanics, contributes another 32%.

JEL Classification Numbers: C43, C81, D63.

Keywords: Segregation, indices, measurement, peer effects, schools, education.

1 Introduction

Segregation is a pervasive social issue. The segregation of men and women into different occupations helps explain the gender gap in earnings.[1] Racial segregation in schools is thought to contribute to low educational achievement among minorities.[2] Residential segregation has been blamed for black poverty, high black mortality, and increases in prejudice among whites.[3] In other contexts, segregation is viewed more positively. The formation

[1] See Cotter et al [13], Lewis [32], and Macpherson and Hirsh [34].

[2] Recent studies include Boozer, Krueger, and Wolkon [3] and Hanushek, Kain, and Rivkin [22].

[3] See Cutler and Glaeser [15], Collins and Williams [11], and Kinder and Mendelberg [31], respectively.

of homogeneous living areas has been discussed as a solution to highly polarized conflicts in the Middle East, Yugoslavia, and elsewhere.

In this paper we focus on contexts in which geography is unimportant. In some cases, such as residential neighborhoods, this might be a strong assumption. In others, it is more innocuous. For instance, the presence of other schools near a given student's school typically does not have a great effect on the student's educational outcomes. Hence, our presentation will focus on school segregation.

Segregation, like inequality, is a complex concept. The literature on segregation measurement has generated over 20 different indices.[4] Hence, it is useful to go back to basic principles: to look for simple and compelling axioms and see what their implications are. By characterizing segregation indices in terms of their basic properties, axiomatizations help researchers compare and choose between measures.

While some papers have analyzed the properties of various indices, very few of them have provided a full axiomatization. Those that have done so treat only the two-group case. Existing axiomatizations also rely in part on cardinal axioms—for instance, that the index be additively separable across schools. While such properties are convenient, their implications for how an index ranks pairs of districts are often unclear. Our axiomatization relies solely on ordinal axioms: rules that an index must follow in ranking districts.

Formally, we define a segregation ordering as a complete ordering on school districts: a ranking of districts from most segregated to least segregated. We propose a set of axioms that restrict this ranking in various ways. We then prove that there is a unique ordering that satisfies our axioms. It is represented by a simple index, which we call the "Mutual Information" index. While this index turns out to be additively separable—indeed, in a more intuitive way than other indices—this is a consequence of our axioms rather than an assumption.

The Mutual Information index is defined as follows. Consider a discrete random variable x that takes K possible values. Let p_k be the probability of the kth value of x. For

[4]Surveys include Massey and Denton [35], Flückiger and Silber [18], and Reardon and Firebaugh [43].

instance, if x is the ethnic group of a randomly selected student, then p_k is the proportion of district students who are in the kth group. The entropy of x is a measure of the uncertainty in x.[5] It is defined as $\sum_{k=1}^{K} p_k \log_2 \left(\frac{1}{p_k}\right)$. Now suppose that we do not know the student's race. We are told only which school, y, she attends. If the schools in the district are segregated, this will convey some information about her race. The *mutual information between x and y* is a measure of how much we learn. It is defined as the expected reduction in the entropy of the student's race that results from learning her school. In particular, once a student's school is known, the entropy of x will be based on the ethnic distribution *within* that school. The expected entropy of the student's race once her school is known is defined as the population-weighted average of these within-school entropies. The *mutual information between x and y* is thus defined as

$$M = \begin{pmatrix} \text{Entropy of district} \\ \text{ethnic distribution} \end{pmatrix} - \sum_{\substack{\text{schools } n \\ \text{in district}}} \begin{pmatrix} \text{Proportion of} \\ \text{district students} \\ \text{in school } n \end{pmatrix} \begin{pmatrix} \text{Entropy of} \\ \text{school } n\text{'s} \\ \text{ethnic distribution} \end{pmatrix}$$

We call this the *Mutual Information index* of segregation in the district: the reduction in uncertainty about a student's race that comes from learning which school she attends. Mutual information is a symmetric concept: either variable leads to the same reduction in uncertainty about the other (Cover and Thomas [14, pp. 18 ff.]). Hence, the Mutual Information index also equals the reduction in uncertainty about a student's *school* that comes from learning her *race*.

The Mutual Information index was first proposed by Theil [50] and was applied by Fuchs [21] and Mora and Ruiz-Castillo [36, 39] to study gender segregation in the labor force.[6] It is related to the more widely used Entropy index H (Theil [51]; Theil and Finizza [52]), which equals the ratio of the Mutual Information index to the entropy of the

[5] The entropy of x is, among other things, an upper bound on the average number of bits needed to encode a series of i.i.d. realizations of x. See Cover and Thomas [14] for this and other interpretations.

[6] See also Herranz, Mora, and Ruiz-Castillo [23]. Some of the properties of the Mutual Information index have been previously noted by Mora and Ruiz-Castillo in the case of two ethnic groups [37, 38].

districtwide ethnic distribution:

$$H = M \Bigg/ \left(\begin{array}{c} \text{Entropy of district} \\ \text{ethnic distribution} \end{array} \right)$$

While the Entropy index is normalized to reach a maximum value of one, the Mutual Information index has no maximum value. However, the Entropy index violates two of our axioms.

In order to judge our axioms, one must have an idea of what we are trying to measure. A starting point is James and Taeuber's [29] definition of segregation as the tendency of ethnic groups to have different distributions across locational units such as schools or neighborhoods. In a later paper, Massey and Denton [35] discern five different dimensions of segregation. The first, evenness, agrees with James and Taeuber's definition. The second dimension is isolation from the majority group. The three other dimensions rely on geographic information and thus are not relevant to our study.[7]

While evenness generalizes easily to the multigroup setting, isolation is more of a challenge, since there is more than one other ethnic group from which a student can be "isolated." Hence, we replace isolation with the concept of representativeness: to what extent do students attend schools that have different ethnic compositions than the district as a whole? The concepts are related, since racially isolated schools are, by definition, not representative of their districts. But unlike isolation, representativeness is not based on exposure to just one other group.

The concept of representativeness is connected to economic issues such as equality of opportunity. Boozer, Krueger, and Wolkin [3] and Hoxby [24] find that the ethnic composition of a school affects individual students' achievement. In the presence of such ethnic-based peer effects, a lack of representativeness can create unequal educational opportunities among students of different races. Evidence for this appears in Hanushek, Kain,

[7] These dimensions are concentration in a small area, centralization in the urban core, and clustering in a contiguous enclave.

and Rivkin [22], who find that the higher proportion of blacks in the school attended by the typical black student can explain a large portion of the black-white wage gap.

These two dimensions are also related to the two interpretations of the Mutual Information index. When schools are not representative of the district, a student's school conveys information about her race. When evenness is violated (i.e., when ethnic groups are not identically distributed across schools), a student's race conveys information about her school. Thus, a district's deviation from representativeness (evenness) determines how much information a random student's school (race) conveys about her race (school). In this sense, the Mutual Information index is affected by deviations from both representativeness and evenness.

Unlike other segregation indices, the Mutual Information index has no maximum value. This allows it to capture changes in interracial exposure better than normalized indices (section 4). It also affords the index an intuitive decomposition across geographic levels and ethnic groups that other indices lack (section 5). We illustrate by decomposing school segregation among urban schools in the United States simultaneously by geographic level (state, city, district, and school) and ethnic group (Asian, black, Hispanic, white). Rivkin [45] and Clotfelter [7] find that, within U.S. cities, segregation across districts exceeds segregation within districts. We confirm this finding. However, segregation among the districts of a city accounts for only 33% of total segregation. Another 32% is due to segregation across states. This is driven mainly by the distinct residential patterns of Hispanic students, 53% of whom attend schools in Texas, California, and New Mexico, compared to only 14% of non-Hispanic students. This and other empirical findings appear in section 6.

The first to study segregation axiomatically was Philipson [41], who provides an axiomatic characterization of a large family of segregation orderings that have an additively separable representation. The representation consists of a weighted average of a function that depends on the school's ethnic distribution only.

In two papers, Hutchens [25, 26] studies the measurement of segregation in the case of two ethnic groups. Hutchens [25] characterizes the family of indices that satisfy a set of mostly cardinal properties. Hutchens [26] strengthens one axiom and obtains a unique segregation index, which is based on the Atkinson inequality index [1]. Frankel and Volij [20] axiomatize multigroup versions of the Atkinson index using ordinal axioms. Finally, using cardinal axioms, Echenique and Fryer [16] characterize a segregation measure that relies on data on social networks to measure the isolation of an individual or group from other ethnic groups.

The rest of the paper is organized as follows. Section 2 presents our notation. In section 3 we explain our axioms. The main result appears in section 4. In section 5, we consider three other properties and show which properties are satisfied by existing school segregation indices. Section 6 applies the Mutual Information index to public schools in the U.S. Most results are proved in Appendix A. Properties of other indices (section 5), many of which are already known, are proved in a second, unpublished appendix.

2 Notation

We assume a continuum population. This is a reasonable approximation when ethnic groups are large. For brevity, we use small integers in our examples; each "person" should be interpreted as representing some large, fixed number of students.

Formally, we define a (school) district as follows:

Definition 1 A *district* X consists of

- A nonempty and finite set of ethnic groups $\mathbf{G}(X)$
- A nonempty and finite set of schools $\mathbf{N}(X)$
- For each ethnic group $g \in \mathbf{G}(X)$ and for each school $n \in \mathbf{N}(X)$, a nonnegative number T_g^n: the number of members of ethnic group g that attend school n.

We will sometimes specify a district in *list format*: $\left\langle (T_g^n)_{g\in\mathbf{G}} \right\rangle_{n\in\mathbf{N}}$. For instance, $\langle (10, 20), (30, 10)\rangle$ denotes a district with two ethnic groups (e.g., blacks and whites) and two schools. The first school, $(10, 20)$, contains ten blacks and twenty whites; the second, $(30, 10)$, contains thirty blacks and ten whites.

For any two districts X and Y, $X \uplus Y$ denotes the result of combining the schools in X and the schools in Y into a single district. If X is a district and α is a nonnegative scalar, then αX denotes the district in which the number of students in each group and school has been multiplied by α. Finally, $c(X)$ denotes the district that results from combining the schools in X into a single school. So, for instance, if $X = \langle (10, 20), (30, 10)\rangle$ and $Y = \langle (40, 50)\rangle$, then $X \uplus Y = \langle (10, 20), (30, 10), (40, 50)\rangle$, $2X = \langle (20, 40), (60, 20)\rangle$, and $c(X) = \langle (40, 30)\rangle$.

The following notation will be useful:

$T_g = \sum_{n\in\mathbf{N}} T_g^n$: the number of students in ethnic group g in the district

$T^n = \sum_{g\in\mathbf{G}} T_g^n$: the total number of students who attend school n

$T = \sum_{g\in\mathbf{G}} T_g$: the total number of students in the district

$P_g = \frac{T_g}{T}$: the proportion of students in the district who are in ethnic group g

$P^n = \frac{T^n}{T}$: the proportion of students in the district who are in school n

$p_g^n = \frac{T_g^n}{T^n}$ (for $T^n > 0$): the proportion of students in school n who are in ethnic group g

The *ethnic distribution of a district* X is the vector $(P_g)_{g\in\mathbf{G}}$ of proportions of the students in the district who are in each ethnic group. The *ethnic distribution of a nonempty school* n is the vector $\left(p_g^n\right)_{g\in\mathbf{G}}$ of proportions of students in school n who are in each ethnic group. A school is *representative* if it has the same ethnic distribution as the district that contains it.

3 Axioms

We now introduce our eight axioms.[8] Let $\mathcal{C}$ be the set of all districts. A *segregation ordering* $\succcurlyeq$ is a complete and transitive binary relation on $\mathcal{C}$. We interpret $X \succcurlyeq Y$ to mean "district X is at least as segregated as district Y." The relations $\sim$ and $\succ$ are derived from $\succcurlyeq$ in the usual way.[9] A related concept is the segregation *index*: a function $S : \mathcal{C} \rightarrow \Re$. The index S *represents* the segregation ordering $\succcurlyeq$ if, for any two districts $X, Y \in \mathcal{C}$,

$$X \succcurlyeq Y \Longleftrightarrow S(X) \geq S(Y) \tag{1}$$

Every index S induces a segregation ordering $\succcurlyeq$ that is defined by (1).

We impose axioms not on the segregation index but on the underlying segregation ordering. These approaches are not equivalent. As in utility theory, a segregation ordering may be represented by more than one index, and there are segregation orderings that are not captured by any index.

A district's *segregation ranking* or simply its *segregation* is its place in the segregation ordering. We will sometimes say that if a transformation $\sigma : \mathcal{C} \rightarrow \mathcal{C}$ is applied to a district X, then "the segregation of the district is unchanged" or "the district's segregation ranking is unaffected." By this we mean that $\sigma(X) \sim X$. If this holds for all districts X, then we will say that the segregation in a district is invariant to the transformation σ.

Evenness and representativeness are properties of the row and column percentages of the district matrix. Nothing in these concepts suggests that the rows or columns should be treated asymmetrically. Accordingly, our first axiom states that the order of the schools or groups and their labels such as "black", "Roosevelt School," etc., do not matter: all that matters is the number of each group who attend each school.

[8] Eight may seem a lot. However, the axiom count can always be reduced by combining related axioms, at some sacrifice of clarity. We ask that the reader judge our axioms based on their content and not on their quantity.

[9] That is $X \sim Y$ if both $X \succcurlyeq Y$ and $Y \succcurlyeq X$; $X \succ Y$ if $X \succcurlyeq Y$ but not $Y \succcurlyeq X$.

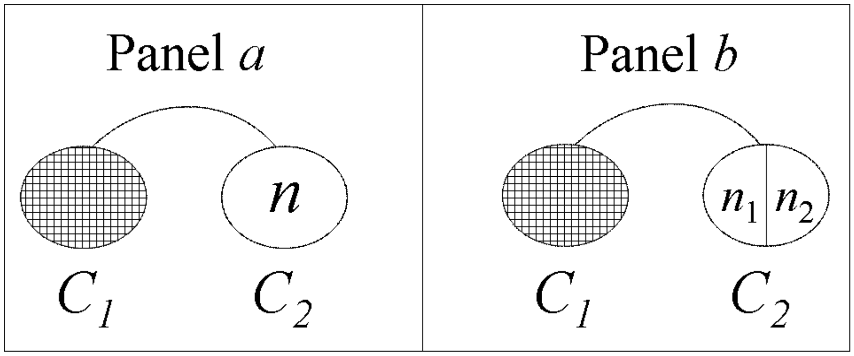

Figure 1: The School Division Property. In panel a, a district has been partitioned into two clusters, one containing a single school n. In panel b, school n has been divided into schools n_1 and n_2. The School Division Property states that segregation is no lower in panel b than in panel a and, moreover, that segregation is the same in the two panels if schools n_1 and n_2 have the same ethnic distribution.

Symmetry (SYM) The segregation in a district is invariant to any relabeling or reordering of the groups or the schools in the district.

One type of research for which this axiom may not be suitable is work that focuses on the problems that face a particular ethnic group. For instance, if one is interested in the social isolation of blacks from all other groups, then one may want to treat blacks differently.

We motivate the next three axioms with a brief discussion. Suppose we partition a district into K clusters, C_1 through C_K, each consisting of a subset of schools in the district. Define segregation within a given cluster as the segregation ranking of the cluster viewed in isolation, as a distinct school district. Define segregation between the K clusters as the segregation ranking of a district with K schools $k = 1, ..., K$, where school k consists of the students in cluster k in the original district. We would like the district's segregation ranking to be a function of segregation *within* each cluster, segregation *between* the clusters, and the relative sizes of the different clusters. Naturally, a district's segregation ranking should be a *nondecreasing* function of both segregation within each cluster and segregation between the K clusters.

The first axiom that uses this principle is illustrated in Figure 1. In panel a, we divide a district into two clusters. The first, cluster C_1, consists of all schools except a single school n. The second, cluster C_2, consists of school n alone. In panel b, school n has been torn down and replaced by two new schools, n_1 and n_2. Each student who formerly attended school n now attends either school n_1 or n_2; all other students attend the same schools as before. This change should not lower segregation in the district. Why? The only factor affected by the split is segregation within cluster C_2. There has been no change in segregation within cluster C_1, segregation between the clusters, or the relative sizes of the two clusters. Since initially cluster C_2 was not segregated at all, splitting school n cannot lower segregation in this cluster. Accordingly, splitting school n should not lower segregation in the district either. If schools n_1 and n_2 have the same ethnic distribution, then cluster C_2 is not segregated at all after the split, since each school is representative of the cluster. In this case, the segregation ranking of the district should not change. These conclusions are formalized in the following axiom, which is satisfied by all indices of school segregation of which we are aware (section 5).

School Division Property (SDP) Let $X \in \mathcal{C}$ be a district in which the set of schools is $\mathbf{N}$. Let X' be the result of splitting some school $n \in \mathbf{N}$ into two schools, n_1 and n_2. Then $X' \succcurlyeq X$. If both schools have the same ethnic distribution, then $X' \sim X$.

This axiom implies, for instance, that a district with 110 whites and ten blacks in a single school does not become more segregated if the ten blacks and an equal number of whites are relocated to a second school: the district $\langle (110, 10) \rangle$ is no more segregated than the district $\langle (100, 0), (10, 10) \rangle$. Of course, one can think of notions of "segregation" that would contradict this. A student in the school $(10, 10)$ might think that her new environment is more "integrated" since it has equal numbers of blacks and whites. No model can capture all possible notions of segregation. Our only hope is to specify a few well-defined criteria that are likely to be related to variables of economic interest. By the representativeness criterion, segregation in the district has indeed increased: the original school was representative of its district and the two new schools are not.[10] If peer effects are important, then the change is likely to widen the black-white achievement gap.

SDP is related to two properties that are discussed by James and Taeuber [29] and subsequent authors. The first is *organizational equivalence*: if a school is divided into two schools that have the same ethnic distribution, the district's level of segregation does not change. The second is the *transfer principle*. When there are two demographic groups, the transfer principle states that if a black (white) student moves from one school to another school in which the proportion of blacks (whites) is higher, then segregation in the district rises. In the case of two ethnic groups, SDP follows from organizational equivalence and the transfer principle.[11] But while SDP applies directly with any number of groups, it is unclear what form the transfer principle should take with more than two groups.[12]

Our next axiom is illustrated in Figure 2. In panel a, two districts, X and Y, are being compared. The districts are assumed to have the same number of students and ethnic distribution. In panel b, a cluster that contains a single school has been adjoined to each of these districts. The axiom states that the addition of this cluster should not affect which district is more segregated. That is, the district on the left hand side in panel b is more segregated than the district on the right hand side in panel b if and only if X is more segregated than Y. Intuitively, since X and Y have the same size and ethnic distribution, the *between-cluster* segregation is the same in each combined district in panel b. Moreover, segregation *within* cluster Z is the same in the two combined districts. Hence, which of the combined districts in panel b is more segregated reduces to whether segregation within

[10] With respect to evenness, ethnic groups are (trivially) distributed evenly across schools in the first district but not in the second.

[11] Proof available on request.

[12] For instance, suppose a black student moves to a school that has higher proportions of both blacks and Asians but fewer whites. Since there are more blacks, one might argue (using the transfer principle) that segregation has gone up. On the other hand, blacks are now more integrated with Asians. One attempt to overcome this difficulty appears in Reardon and Firebaugh [43].

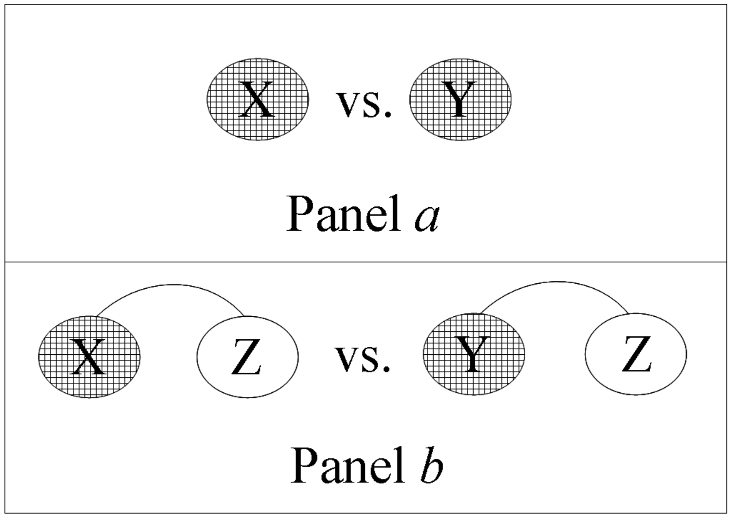

Figure 2: Type I Independence (IND1). Panel a shows two districts, X and Y, that have the same size and ethnic distribution. IND1 states that adjoining the same cluster containing a single school to the two districts (panel b) does not affect which district is more segregated.

cluster X is greater than segregation within cluster Y.

Type I Independence (IND1) Let $X, Y \in \mathcal{C}$ be two districts with equal populations and equal ethnic distributions. Then for any district Z that contains a single school, $X \succcurlyeq Y$ if and only if $X \uplus Z \succcurlyeq Y \uplus Z$.

Since X and Y have the same number of each ethnic group, one can think of Y as a reallocation of the students in X.[13] IND1 states that this reallocation raises segregation in the district if and only if it raises segregation within the cluster. The Dissimilarity Index violates this principle.[14] For instance, let $X = \langle (50, 100), (50, 0) \rangle$ and $Z = \langle (100, 0) \rangle$. Suppose that the students in cluster X are reallocated to yield $Y = \langle (100, 40), (0, 60) \rangle$.

[13] IND1 does not require Y to have the same number of schools as X. Hence, the reallocation might be accompanied by new school construction or conversion of some schools to other uses.

[14] With two ethnic groups, the Dissimilarity Index is defined as the minimum proportion of either group that would have to be reallocated in order for all schools to be representative of the district.

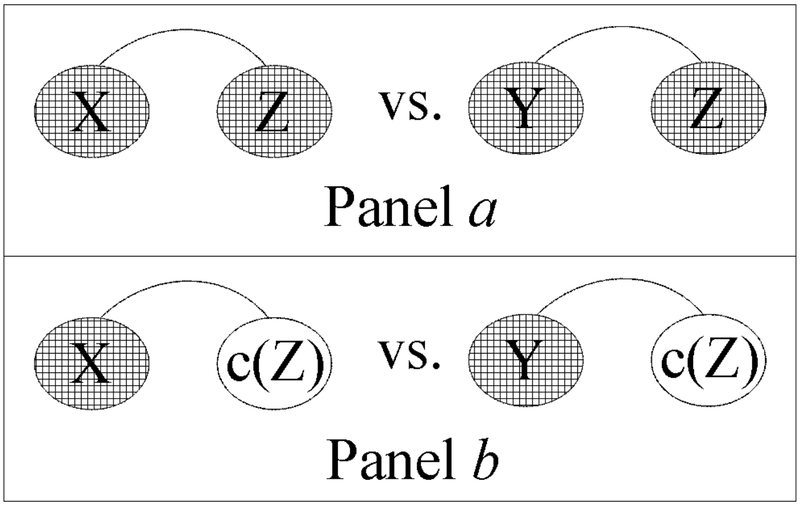

Figure 3: In panel a, a given district, Z, is combined with each of two districts, X and Y, which have the same total number of students but possibly different ethnic distributions. In panel b, all the schools in Z have been combined into a single school. Type II Independence states that this merger does not affect which combined district is more segregated.

The Dissimilarity Index within this cluster *rises* from 0.5 to 0.6, but the index for the full district *falls*, counterintuitively, from 0.75 to 0.6. IND1 rules out such behavior.

A second type of independence is depicted in Figure 3. In panel a, a given district, Z, is paired with each of two districts, X and Y. As in Figure 2, X and Y have the same total number of students; unlike that case, their ethnic distributions may differ. In panel b, all the schools in Z have been combined into a single school; the resulting cluster is denoted $c(Z)$. Type II Independence states that this merger of schools does not affect which combined district is more segregated.

Type II Independence (IND2) Let $X, Y, Z \in \mathcal{C}$ be three districts such that $T(X) = T(Y)$. Let $c(Z)$ be the cluster that results from combining the schools in Z into a single school. Then $X \uplus Z \succcurlyeq Y \uplus Z$ if and only if $X \uplus c(Z) \succcurlyeq Y \uplus c(Z)$.

A motivation is as follows. Suppose that, in panel a, the combination of X and Z is more segregated than the combination of Y and Z. What must be driving this? Segregation *within* cluster Z is the same in the two districts in panel a. So the combination of within-X segregation and between-X-and-Z segregation must exceed the combination of within-Y segregation and between-Y-and-Z segregation. Moreover, since X and Y are of the same size, the relative importance of within-cluster and between-cluster segregation is the same in the two cases. Now consider panel b. Merging the schools in Z does not affect segregation between this cluster and either X or Y. Consequently, if in panel b the district containing cluster X is more segregated than the district containing cluster Y, then the combination of within-X segregation and between-X-and-Z segregation must exceed the combination of within-Y segregation and between-Y-and-Z segregation, just as in panel a. Moreover, since the merger does not affect the size of any cluster, it does not change the relative importance of within-cluster and between-cluster segregation. Accordingly, merging the schools in Z should *not* affect which merged district is more segregated. In other words, the degree of segregation within a given cluster should not affect the relative importance of between-cluster segregation and segregation within the other clusters in the district. In section 5 we show that if an ordering violates Type II Independence, then an index that represents it cannot be decomposable across schools in a particular simple way (Proposition 1).

IND2 is violated by most existing school segregation indices. For instance, consider $X = \langle (50, 0), (0, 100) \rangle$, $Y = \langle (100, 0), (0, 50) \rangle$ and $Z = \langle (50, 0), (0, 100) \rangle$. Intuitively, X and Z have the same ethnic distribution, which differs from that of Y. Hence, between-cluster segregation is lower in $X \uplus Z$ than in $Y \uplus Z$. However, for most indices to reach their maximum value, it suffices for segregation *within* each cluster to be at a maximum: they regard $X \uplus Z$ and $Y \uplus Z$ as maximally, and thus equally, segregated. It is only as within-cluster segregation falls that these indices begin to reflect between-cluster segregation: they regard $X \uplus c(Z)$ as strictly less segregated than $Y \uplus c(Z)$, violating IND2.[15] IND2 implies

[15] This behavior characterizes all indices surveyed in section 5 except the Mutual Information index. For the Clotfelter index, this assumes that the first group is identified as blacks and the second as whites. Details are available on request.

that an index cannot ignore between-cluster segregation even if within-cluster segregation is at a maximum.

The next axiom is used to compare districts with different ethnic distributions. It states that segregation is invariant to the division of an existing ethnic group into two identically distributed subgroups. For instance, if white students are divided into those with blue eyes and those with brown, and these groups have the same distribution across schools, then the segregation of a district should not change.

Group Division Property (GDP) Let $X \in \mathcal{C}$ be a district in which the set of ethnic groups is G. Let X' be the result of partitioning some ethnic group $g \in G$ into two ethnic groups, g_1 and g_2, such that both ethnic groups have the same distribution across schools: $\frac{T^n_{g_1}}{T_{g_1}} = \frac{T^n_{g_2}}{T_{g_2}}$ for all $n \in \mathbf{N}$.[16] Then $X' \sim X$.

Intuitively, suppose we partition the ethnic groups of X into K sets or "supergroups." Define within-supergroup segregation to be the segregation of the district that would result if all students who are not members of the given supergroup were removed. Let between-supergroup segregation be the segregation of the district that would result from treating each supergroup as a single ethnic group. Then segregation in X should be a function of segregation *within* each supergroup, segregation *between* the supergroups, and the relative *sizes* of the supergroups.

This principle helps motivate GDP in the following way. Let us partition the ethnic groups of X into two supergroups, one consisting of group g alone and the other consisting of all other groups. Suppose group g is split into two groups, g_1 and g_2, which have the same distribution across schools. This change clearly does not affect segregation within either supergroup, nor does it affect segregation between the supergroups or the relative sizes of the two supergroups. Hence, the district's segregation ranking should not be affected by the split. In section 5 we show that an ordering that violates GDP cannot

[16]Note that X' has the same set N of schools as X and for each school $n \in N$, $T^n_g = T^n_{g_1} + T^n_{g_2}$.

be represented by an index that is decomposable over groups in a particular simple way (Proposition 1).

The criteria of evenness and representativeness pertain to the row and column percentages in the district matrix. Multiplying the whole matrix by a scalar does not affect these percentages, so it should not affect the segregation ranking of a district. Hence, we assume the following axiom, which is satisfied by all school segregation indices of which we are aware (section 5):

Weak Scale Invariance (WSI) The segregation ranking of a district is unchanged if the numbers of agents in all ethnic groups in all schools are multiplied by the same positive scalar: for any district $X \in \mathcal{C}$ and any positive scalar α, $X \sim \alpha X$.[17]

This axiom implies, e.g., that the districts $\langle (10^6, 0), (0, 10^6) \rangle$ and $\langle (100, 0), (0, 100) \rangle$ are equally segregated. One may argue that the first district is more segregated because it is less likely to be the outcome of random assignment of students to schools (see, e.g., Cortese, Falk, and Cohen [12]). However, our focus is not on the *ex ante* process by which students are assigned to schools, but rather on the *ex post* segregation of students among schools. Whether an *ex ante* or *ex post* measure is relevant depends on the context. If one cares about how segregated peer groups create unequal opportunities, an *ex post* measure is more suitable. This point has been made in the sociological literature by Taeuber and Taeuber [49, p. 886]. If one is trying to detect discriminatory policies, one may be interested instead in an *ex ante* segregation measure.

One weakness of *ex ante* segregation measures is that they can be manipulated: by simply increasing the number of schools, a district can make random assignment harder to

[17] One may argue that $\langle (1, 100), (100, 1) \rangle$ is more segregated than $\langle (2, 200), (200, 2) \rangle$ since in the first district, there are two students with no peers of their own race while in the second each has at least one. Recall, however, that we assume a continuum population: "1" person represents many identical students. This assumption is made for technical convenience. It is a good approximation in large districts for all but the smallest ethnic groups. And since the entropy term $p \log_2(1/p)$ is close to zero when p is small, our representation will not be sensitive to these small groups in any event.

rule out statistically. Under the hypothesis of color-blind assignment, the Mutual Information index (times a scale factor) is distributed χ^2 with degrees of freedom increasing in the number of schools. More precisely:

Claim 1 *Let X be a district with T students. Let H_0 be the hypothesis of color-blind assignment: that the probability that a random student is of race g and attends school n is the product of some constants α^n and β_g. Let H_1 be the alternative in which this probability is unrestricted. The log-likelihood ratio statistic for H_0 versus H_1 equals the Mutual Information index of the district, multiplied by $2T\ln(2)$. This test statistic is asymptotically distributed as χ^2 with $(N-1)(K-1)$ degrees of freedom, where N is the number of schools and K is the number of ethnic groups in X.*[18]

Suppose a district with a policy of racial separation subdivides one school into two new schools with the same ethnic distribution. The Mutual Information index is unchanged, by the School Division Property (Theorem 1, below). However, the increase in degrees of freedom makes it harder for a statistician to reject the hypothesis of color-blind assignment. The district has manipulated the test. In contrast, the only way the Mutual Information index can be reduced is by actually making schools more representative of their district.

The next axiom is a technical continuity property. We rely on this axiom to prove that the segregation ordering is represented by a segregation index.

Continuity (C) For any three districts $X, Y, Z \in \mathcal{C}$, $\{c \in [0,1] : cX \uplus (1-c)Y \succcurlyeq Z\}$ and $\{c \in [0,1] : Z \succcurlyeq cX \uplus (1-c)Y\}$ are closed sets.

Our final axiom states that there exist two districts with two nonempty ethnic groups that are not equally segregated. It is needed to rule out the trivial segregation ordering.

Nontriviality (N) There exist districts $X, Y \in \mathcal{C}$, each with exactly 2 nonempty ethnic groups, such that $X \succ Y$.

[18]This result is proven for the case of two ethnic groups by Mora and Ruiz-Castillo [38, pp. 32-33]. The general case is shown in our unpublished appendix.

4 Main Result

Our main result is that the segregation ordering represented by the Mutual Information index is the unique ordering that satisfies all of our axioms.

Theorem 1 *The Mutual Information ordering is the only segregation ordering that satisfies SYM, WSI, SDP, IND1, IND2, GDP, C, and N.*

As noted in the introduction, the Entropy index of Theil [51] and Theil and Finizza [52] is obtained by dividing the Mutual Information index by its maximum value, the entropy of the district ethnic distribution. Thus, the Entropy index takes a maximum value of one, while the Mutual Information index has no maximum value. As a result, these indices do not give the same segregation ordering. The Entropy index ranks all districts with no ethnic mixing as equally segregated, while the Mutual Information index assigns a higher segregation level to districts in which there is more initial uncertainty about a student's ethnicity. Consider, e.g., the two districts $\langle (50,0,0), (0,50,0), (0,0,50) \rangle$ and $\langle (50,0), (0,50) \rangle$. In each, segregation is at a maximum given the district ethnic distribution. Accordingly, the Entropy index assigns each a value of one. In contrast, the Mutual Information index equals 1.6 for the first district but 1.0 for the second. This difference arises since in the first district, learning a student's school conveys more information about a student's ethnicity.

Now consider the two districts $X = \langle (990,0), (0,10) \rangle$ and $Y = \langle (500,0), (0,500) \rangle$. Once again, the Entropy index assigns each a value of one. However, there is much less uncertainty about a random student's ethnicity in X, so learning her school conveys less information. Accordingly, the Mutual Information index is lower for X than for Y ($M = 0.08$ versus $M = 1.0$, respectively).

In the context of school segregation, normalized indices have two important disadvantages. First, they are not decomposable in a certain intuitive sense (section 5, below). Second, they do not do a good job of capturing changes in interracial contact.[19] To illustrate the second point, compare the effect of merging two schools in X, yielding the

[19] This argument is due to Clotfelter [6].

one-school district $\langle(990, 10)\rangle$, with the effect of merging the two schools in Y, yielding the one-school district $\langle(500, 500)\rangle$. The first merger has a tiny effect on the interracial exposure of the average student: 99% of students see only a 1% change in the percentage of minorities. The second merger has a much larger effect: each student switches from a completely segregated school to one that is half black, half white. The Mutual Information index reflects this difference, falling by 0.08 in district X versus 1.0 in Y. In contrast, the Entropy index misses the difference entirely, decreasing by 1.0 in both cases.

5 Other Indices

In an unpublished appendix, we discuss the other indices that have been used to study school segregation and show which of our axioms are violated by them. We also consider an additional property, Scale Invariance, and two decomposability properties. Scale Invariance states that the segregation of a district is invariant to proportional changes in ethnic group size:[20]

Scale Invariance (SI) For any district X, ethnic group $g \in \mathbf{G}(X)$, and constant $\alpha > 0$, let X' be the result of multiplying the number of students in group g in each school n in district X by α. Then $X' \sim X$.

This property has both supporters and opponents in the field of school segregation (Taeuber and James [48, p. 134]; Coleman, Hoffer, and Kilgore [9, p. 178]). One can easily verify that the Mutual Information index violates it.

The next property states that, for any partition of a district's schools into clusters, total segregation in the district is the sum of between-cluster and within-cluster segregation:

Strong School Decomposability (SSD) An index S satisfies Strong School Decomposability if, for any partition $X = X^1 \uplus \cdots \uplus X^K$ of the schools of a district into K

[20]This property is also known as Compositional Invariance (e.g., James and Taeuber [29, pp. 15-16]).

clusters,

$$S(X) = S(c(X^1) \uplus \cdots \uplus c(X^K)) + \sum_{k=1}^{K} P^k S(X^k) \tag{2}$$

where $S(c(X^1) \uplus \cdots \uplus c(X^K))$ is segregation between the K clusters, $S(X^k)$ is segregation within cluster k, and P^k is the proportion of students in cluster k.

Mora and Ruiz-Castillo [37] show that the Mutual Information index satisfies Strong School Decomposability in the case of two groups. This and weaker forms of separability have also been extensively discussed in the literature of the measurement of income inequality. Bourguignon [4], for instance, shows that a property analogous to Strong School Decomposability fully characterizes the Theil inequality index (a close relative of the Mutual Information index) within the class of differentiable relative inequality indices. Foster [19] obtains a further characterization of the Theil inequality index by replacing the differentiability requirement by a more appealing transfer principle. Hutchens [26] uses a weaker version of separability to help characterize the Atkinson segregation index in the two-group case.

The second, analogous property states that, for any partition of a district's groups into sets or "supergroups," total segregation is the sum of between-supergroup and within-supergroup segregation:

Strong Group Decomposability (SGD) An index S satisfies satisfies Strong Group Decomposability if, for any partition of the ethnic groups of a district X into K supergroups,

$$S = S_K + \sum_{k=1}^{K} P_k S_k \tag{3}$$

where P_k is the proportion of students who are in supergroup k; S_K is the segregation of the district that would result from treating each supergroup as a single group; and S_k is the segregation of the district that would result if all students not in supergroup k were removed.

These decomposability properties are related to the two types of Independence and the Group Division Property in the following way.

Proposition 1 *If S is a segregation index that satisfies Strong School Decomposability, then the segregation ordering represented by S satisfies Type I and Type II Independence. If S satisfies Strong Group Decomposability, then the induced segregation ordering satisfies the Group Division Property.*

Hence, if a segregation ordering violates the Group Division Property (respectively, either Type I or Type II Independence), then it cannot be represented by an index that satisfies the Strong Group (respectively, School) Decomposability. The Mutual Information index is decomposable in both ways:

Proposition 2 *M satisfies Strong School and Group Decomposability.*

The properties of the Mutual Information index and other existing school segregation indices are summarized in Table 1. Proofs appear in an unpublished appendix. Of all indices considered, only M satisfies Strong School Decomposability. M is also the only index that has no maximum value. These two properties are related, by the following result:

Proposition 3 *Let $S : \mathcal{C} \rightarrow \Re_+$ be a function with the following properties:*

1. *it attains a maximum value;*
2. *it treats ethnic groups symmetrically;*
3. *it equals zero only on districts in which all schools have the same ethnic distribution.*

Then S violates Strong School Decomposability.

Most existing indices of school segregation satisfy properties 1-3 and so cannot satisfy Strong School Decomposability. In contrast, the Mutual Information index takes no maximum value and satisfies SSD. The next section contains an empirical illustration of the uses of SSD and SGD.

6 U.S. School Segregation

In this section we show some uses of the Mutual Information index for the study of school segregation in the U.S. We restrict to school districts that contain at least two schools and that serve grades K-12. Schools not located in Core Based Statistical Areas (CBSA's) or that do not lie in the 50 U.S. states and the District of Columbia are excluded. We refer to the resulting set of schools as "urban schools". Data are for the 2005-6 school year and come from the Common Core of Data (CCD) [46].

Table 2 computes the Mutual Information index for all urban schools in the U.S. and decomposes it into various components. We use four, mutually exclusive ethnic groups: Asians, (non-Hispanic) whites, (non-Hispanic) blacks, and Hispanics.[21] Since supergroup schemas must be nested in order to apply Strong Group Decomposability, we remove one ethnic group at a time.[22] Let each ethnic group be denoted by its initials: A(sians), W(hites), B(lacks), and H(ispanics). Let curly braces denote a supergroup; e.g., $\{W, B, H\}$ denotes the set of non-Asians. Applying equation (3) twice, we can decompose overall segregation into three terms:

$$\begin{pmatrix} \text{Segregation} \\ \text{among} \\ A, W, B, \text{ \& } H \end{pmatrix} = \begin{pmatrix} \text{Segregation} \\ \text{between} \\ A \text{ \& } \{W, B, H\} \end{pmatrix} + \begin{pmatrix} \text{Proportion} \\ \text{of students} \\ \text{in } W, B, \text{ or } H \end{pmatrix} \begin{pmatrix} \text{Segregation} \\ \text{between} \\ \text{W \& } \{B, H\} \end{pmatrix} + \begin{pmatrix} \text{Proportion} \\ \text{of students} \\ \text{in } \{B, H\} \end{pmatrix} \begin{pmatrix} \text{Segregation} \\ \text{between} \\ B \text{ \& } H \end{pmatrix} \tag{4}$$

[21] The CCD actually has five ethnic groups; the smallest, American Indian/Alaskan Native, is not represented in some school districts. This does not create problems for the Mutual Information index, which ignores empty groups. However, some of the other indices require all groups to be nonempty. Hence, we include this group with the second smallest group, Asians.

[22] We successively remove the most advantaged of the remaining ethnic groups, based on child poverty rates in 2006. These were 12.2%, 14.1%, 26.9%, and 33.4% for Asians, whites, Hispanics, and blacks, respectively (U.S. Census Bureau [53]).

These three terms appear, in this order, in columns 1, 2, and 3. They represent, respectively, the contribution to total segregation of segregation between (1) Asians and non-Asians; (2) whites and non-Asian minorities; and (3) blacks and Hispanics. Their sum appears in column 4 and represents segregation among all four ethnic groups at the given geographic level.

We simultaneously compute segregation at four geographic levels: states, CBSA's, districts, and schools. The first row of Table 2 computes segregation between states, treating each state as a single "school". For row 2, the Mutual Information index across CBSA's is first computed for each state. We then compute the average of these 51 indices, weighted by state population. This average, which appears in row 2, is the within-state, between-CBSA segregation. Row 3 show segregation at within CBSA's, between districts. Row 4 shows segregation within districts, between schools. By SSD, the sum of rows 1-4 equals total segregation between schools in the U.S., which appears in row 5.

Total segregation among the four groups across schools in the U.S. is 0.665 (row 5, column 4). In panel B, all indices are re-expressed as percentages of this total. The most important source of school segregation is the racial differentation of districts within CBSA's, which accounts for 32.9% of the total. A comparison of columns 1-3 of row 8 shows that this is mostly due to the separation of whites from blacks and Hispanics. Segregation between the states is also important, accounting for 31.7% of total segregation. This is mainly due to the residential patterns of Hispanics: if we change the decomposition order, removing Hispanics first instead of Asians, we find that 59% of segregation across states is due to the segregation of Hispanics from non-Hispanics (results not shown). Indeed, 53% of Hispanic students live in Texas, California, and New Mexico, while only 14% of non-Hispanic students live in these states.

Rivkin [45] and Clotfelter [7] find that segregation between whites and nonwhites is mainly between districts within cities, rather than between schools within districts. This is reflected in the difference between rows 3 and 4 of column 2.[23] However, the properties

[23] Rivkin [45] and Clotfelter [7] also present detailed analyses of how segregation at these two levels varies

of SSD and SGD allow us to compare more than two ethnic groups and more than two geographic levels at once. In addition, the mutual information index affords a more intuitive decomposition than is available with other indices: the within-district term in a CBSA is simply the average segregation level of the districts in the CBSA, weighted by student populations. In Rivkin's decomposition of the Gini index, the "within-district" term also includes an enigmatic interaction term, making interpretation difficult. Clotfelter [7] uses the Normalized Exposure index, which can be decomposed only in the case of two ethnic groups. In addition, as with the Gini index, the within-district term is not simply a population-weighted average of the district Gini indices. Rather, the weight on a district depends on the district's ethnic distribution. The same problem afflicts the Entropy index, H.[24]

Table 3 analyzes segregation between pairs of ethnic groups. The least segregated pair is Asians and whites ($M = 0.144$). The most segregated pair is blacks and Hispanics ($M = 0.475$). The most important geographic level depends on the pair being considered. Blacks and whites are primarily segregated across districts within CBSA's: M equals 0.183, the highest district-level segregation of any ethnic group pair. This is also true to a lesser extent for Asians and Hispanics ($M = 0.091$), though segregation across states is almost as important ($M = 0.087$). For every other pair, the state is the most important level, with blacks and Hispanics the most segregated pair at this level ($M = 0.243$).

Rank correlations among the indices in Table 1, using Kendall's τ_b, are shown in Table 4. Each segregation index is computed across the full set of schools in each CBSA. The Mutual Information index is most highly correlated with the Normalized Exposure index, followed by the Card-Rothstein index and the Entropy index. The mean correlation between M and the other indices, 0.561, is the third highest in the table.

Table 5 ranks the large CBSA's (those with at least 200,000 students in K-12 districts) according to the Mutual Information index. The other indices in Table 1 are also shown.

across cities. We do not pursue such an analysis here.

[24] These observations are due to Reardon and Firebaugh [43, pp. 53-4].

The entropy of the CBSA's public school ethnic distribution, h, appears in the final column and reflects the ethnic diversity of students in the CBSA. Since the Mutual Information index cannot exceed h, the high ranking of Chicago and New York are made possible by their diverse ethnic compositions. However, this relation is not monotonic. San Francisco, Sacramento, and Las Vegas all have diverse (high-entropy) ethnic distributions but low rankings by the Mutual Information indices. Cleveland and Detroit have high Mutual Information indices despite their lower levels of ethnic diversity.

7 Conclusions

In this paper we give an axiomatic foundation for multigroup segregation, based on the criteria of evenness (how differently are ethnic groups distributed across schools?) and representativeness (how different are the ethnic distributions of individual schools from that of the district?).

We assume only ordinal axioms. We show that a unique ordering satisfies these axioms. This ordering has many representations. We focus on one of them, the Mutual Information index.[25] It equals the *mutual information* of a student's race and her school when these are viewed as random variables (Cover and Thomas [14]). It can be interpreted both as the information that a student's school conveys about her race and, vice-versa, as the information that her race conveys about her school. These dual intuitions should facilitate the index's application to empirical work on the causes and effects of school segregation.

The Mutual Information index is unusual in that it is not normalized to take a maximum value. This allows it to capture interracial exposure better than normalized indices. It also affords the index intuitive decompositions across ethnic groups and geographic levels. These decompositions are not possessed by other common indices.

We illustrated the use of these decompositions by studying the sources of segregation

[25] An index represents the ordering if and only if it is a strictly increasing transformation of the Mutual Information index.

across urban public schools in the U.S. in 2005-6. As Rivkin [45] and Clotfelter [7] find, segregation between districts within cities is indeed important, accounting for about a third of total segregation. But segregation across states is nearly as important. This is driven mainly by the distinct residential patterns of Hispanics, who are disproportionately concentrated in the southwestern states of Texas, California, and New Mexico.

The Mutual Information index is not scale invariant: it is sensitive to the overall ethnic distribution of the district. For axiomatizations of multigroup segregation that assume scale invariance, see Hutchens [25, 26] and Frankel and Volij [20].

A Proofs

Proof of Claim 1. Let γ_g^n be the probability that a random student is in ethnic group g and attends school n. Under H_0, $\gamma_g^n = \alpha^n \beta_g$. The log-likelihoods for the two hypothesis are thus:

$$H_0 \;:\; \ln\left(\prod_{n\in N}\prod_{g\in G}\left(\alpha^n\beta_g\right)^{T_g^n}\right) = \sum_{n\in N} T^n \ln\alpha^n + \sum_{g\in G} T_g \ln\beta_g$$
$$H_1 \;:\; \ln\left(\prod_{n\in N}\prod_{g\in G}\left(\gamma_g^n\right)^{T_g^n}\right) = \sum_{n\in N}\sum_{g\in G} T_g^n \ln\gamma_g^n$$

with constraints $\sum_{n\in N}\alpha^n = \sum_{g\in G}\beta_g = 1$ and $\sum_{n\in N}\sum_{g\in G}\gamma_g^n = 1$. The likelihood-maximizing parameters are $\alpha^n = T^n/T$, $\beta_g = T_g/T$, and $\gamma_g^n = T_g^n/T$. The log-likelihood ratio for a test of H_1 versus H_0 thus equals

$$\begin{aligned} LR &= -2\left(\sum_{n\in N} T^n \ln\frac{T^n}{T} + \sum_{g\in G} T_g \ln\frac{T_g}{T} - \sum_{n\in N}\sum_{g\in G} T_g^n \ln\frac{T_g^n}{T}\right) \\ &= 2\left(\sum_{g\in G} T_g \ln\frac{T}{T_g} - \sum_{n\in N}\sum_{g\in G} T_g^n \ln\frac{T^n}{T_g^n}\right) = 2T\ln(2)M(X) \end{aligned}$$

The unrestricted (respectively, restricted) model has $NK-1$ (respectively, $N+K-2$) degrees of freedom. Hence, $2T\ln(2)M(X)$ is asymptotically distributed as χ^2 with $(N-1)(K-1)$ degrees of freedom. Q.E.D.

Proof of Theorem 1. We first show that the ordering represented by the Mutual Information index satisfies the axioms. Axioms N, SYM, and WSI are trivial, and C follows from the fact that the index M is a continuous function of the T_g^n's (the number of students of each group in each school). Axioms IND1, IND2,and GDP follow from Propositions 1 and 2. So it remains to show that SDP is satisfied. Let $X \in \mathcal{C}$ be a district and let n be a school of X. Let X' be the district that results from dividing n into two schools, n_1 and n_2. Since X and X' have the same ethnic distribution,

$$\begin{aligned}M(X') - M(X) &= P^n h((p_g^n)_{g\in\mathbf{G}(X)}) - P^{n_1} h((p_g^{n_1})_{g\in\mathbf{G}(X)}) - P^{n_2} h((p_g^{n_2})_{g\in\mathbf{G}(X)}) \\ &= P^n \left(h((p_g^n)_{g\in\mathbf{G}(X)}) - \frac{P^{n_1}}{P^n} h((p_g^{n_1})_{g\in\mathbf{G}(X)}) - \frac{P^{n_2}}{P^n} h((p_g^{n_2})_{g\in\mathbf{G}(X)}) \right)\end{aligned}$$

But for all g, $p_g^n = \frac{P^{n_1}}{P^n} p_g^{n_1} + \frac{P^{n_2}}{P^n} p_g^{n_2}$ so, recalling that $h((q_g)_{g\in\mathbf{G}}) = \sum_{g\in\mathbf{G}} q_g \log_2(\frac{1}{q_g})$ is a concave function, $M(X') - M(X) \geq 0$, with strict inequality only if schools n_1 and n_2 have different ethnic distributions. This verifies SDP.

We now show that the Mutual Information ordering is the only segregation ordering that satisfies all the axioms. Let $\succcurlyeq$ be a segregation ordering that satisfies them. For any district X, let the schools be numbered $n = 1, \ldots, N$ and the groups $g = 1, \ldots, G$.

For any ethnic distribution $P = (P_g)_{g=1}^G$, let $\overline{X}(P)$ denote the district, with population 1, with group distribution P, and with G uniracial schools, and let $\underline{X}(P)$ denote the one-school district with ethnic distribution P and population 1:

$$\overline{X}(P) = \langle (P_1, 0, ..., 0), ...(0, ..., 0, P_G) \rangle \quad \text{and} \quad \underline{X}(P) = \langle (P_1, ..., P_G) \rangle \,.$$

For any integer $G \geq 1$, let $\overline{X}^G = \langle (1/G, 0, ..., 0), ..., (0, ..., 0, 1/G) \rangle$ denote the completely segregated district of population 1 with G equal sized ethnic groups. Let $\underline{X}^G = \langle (1/G, ..., 1/G) \rangle$ denote the one-school district with the same ethnic distribution and population.

We first state and prove some preliminary lemmas. By applying IND1 repeatedly, one can show the following apparently stronger (but actually equivalent) property, which will be used interchangeably with IND1.

Lemma 1 *Suppose the segregation ordering $\succcurlyeq$ satisfies IND1. Let $X, Y \in C$ be two districts with equal populations and equal ethnic distributions. Then for all districts $Z \in C$ containing any number of schools, $X \succcurlyeq Y$ if and only if $X \uplus Z \succcurlyeq Y \uplus Z$.*

Proof. Let the schools of Z be enumerated: $n_1, ..., n_N$. By IND1, $X \succcurlyeq Y$ if and only if $X \uplus \langle n_1 \rangle \succcurlyeq Y \uplus \langle n_1 \rangle$, where $\langle n_1 \rangle$ denotes a district that consists of school n_1 alone. The districts $X' = X \uplus \langle n_1 \rangle$ and $Y' = Y \uplus \langle n_1 \rangle$ have the same size and ethnic distribution since X and Y do. Hence, by IND1, $X' \succcurlyeq Y'$ if and only if $X' \uplus \langle n_2 \rangle \succcurlyeq Y' \uplus \langle n_2 \rangle$. The result follows by repeating the same argument for schools $n_3, ..., n_N$. Q.E.D.

Lemma 2

1. *All districts in which every school is representative have the same degree of segregation under $\succcurlyeq$.*
2. *Any district in which every school is representative is weakly less segregated under $\succcurlyeq$ than any district in which some school is unrepresentative.*

Proof.

1. Consider any district Y that consists of N representative schools. By WSI we can assume w.l.o.g. that $T(Y) = 1$. For each $i = 1, ..., N$, let Y_i be the school district consisting of schools $i+1$ through N of Y as well as a single school that contains the students in schools 1 through i of Y. By SDP, for each $i = 1, ..., N-1$, $Y_i \sim Y_{i+1}$. Hence, by transitivity, $Y = Y_1 \sim Y_N$. Y_N contains a single school. By GDP, $Y_N \sim \underline{X}^1$, and hence $Y \sim \underline{X}^1$.
2. Consider any district Y in which at least one school is unrepresentative. The above procedure yields $Y = Y_1 \succcurlyeq Y_2 \succcurlyeq \cdots \succcurlyeq Y_N \sim \underline{X}^1$. By transitivity, $Y \succcurlyeq \underline{X}^1$.

Q.E.D.

Lemma 3 *For any district Z with G ethnic groups, let $\sigma(Z) \in \mathcal{C}$ be such that the number of persons of ethnic group g in school n in Z equals the number of persons of ethnic group*

$(g+1) \bmod G$ *in school* n *in* $\sigma(Z)$. *Define* $\sigma^1(Z) = \sigma(Z)$ *and, for integers* $j > 1$, *let* $\sigma^j(Z) = \sigma(\sigma^{j-1}(Z))$.[26] *Then* $\frac{1}{G} \biguplus_{j=1}^{G} \sigma^j(Z) \succcurlyeq Z$.

Proof. Consider the following statement:

$$\left(\biguplus_{j=1}^{n} Z\right) \uplus \left(\biguplus_{j=n+1}^{G} c(Z)\right) \preccurlyeq \left(\biguplus_{j=1}^{n} \sigma^j(Z)\right) \uplus \left(\biguplus_{j=n+1}^{G} \sigma^j(c(Z))\right) \tag{5}$$

For $n = 0$, (5) simply states $\biguplus_{j=1}^{G} c(Z) \preccurlyeq \left(\biguplus_{j=1}^{G} \sigma^j(c(Z))\right)$, which holds by Lemma 2. Assume that (5) holds for some $n = k$, with $0 \leq k < G-1$. Then, taking into account that σ^G is the identity permutation,

$$\begin{aligned}
\left(\biguplus_{j=1}^{n} Z\right) \uplus \left(\biguplus_{j=n+2}^{G} c(Z)\right) \uplus c(Z) &\preccurlyeq \left(\biguplus_{j=1}^{n} \sigma^j(Z)\right) \uplus \left(\biguplus_{j=n+1}^{G-1} \sigma^j(c(Z))\right) \uplus c(Z) \\
\Longrightarrow \left(\biguplus_{j=1}^{n} Z\right) \uplus \left(\biguplus_{j=n+2}^{G} c(Z)\right) \uplus Z &\preccurlyeq \left(\biguplus_{j=1}^{n} \sigma^j(Z)\right) \uplus \left(\biguplus_{j=n+1}^{G-1} \sigma^j(c(Z))\right) \uplus Z \qquad \text{by IND2} \\
&\sim \sigma\left(\left(\biguplus_{j=1}^{n} \sigma^j(Z)\right) \uplus \left(\biguplus_{j=n+1}^{G-1} \sigma^j(c(Z))\right) \uplus Z\right) \qquad \text{by SYM} \\
&\sim \left(\biguplus_{j=2}^{n+1} \sigma^j(Z)\right) \uplus \left(\biguplus_{j=n+2}^{G} \sigma^j(c(Z))\right) \uplus \sigma(Z) \qquad \text{by def. of } \sigma \\
\left(\biguplus_{j=1}^{n+1} Z\right) \uplus \left(\biguplus_{j=n+2}^{G} c(Z)\right) &\preccurlyeq \left(\biguplus_{j=1}^{n+1} \sigma^j(Z)\right) \uplus \left(\biguplus_{j=n+2}^{G} \sigma^j(c(Z))\right)
\end{aligned}$$

That is, (5) also holds for $n = k+1$. By induction it also holds for $n = G-1$. That is, $\biguplus_{j=1}^{G} Z \preccurlyeq \biguplus_{j=1}^{G} \sigma^j(Z)$ which, by SDP and WSI implies $Z \preccurlyeq \frac{1}{G} \biguplus_{j=1}^{G} \sigma^j(Z)$. Q.E.D.

Lemma 4 *For any district* X *with* G *groups and group distribution* P, $\overline{X}^G \succcurlyeq \overline{X}(P) \succcurlyeq X$.

Proof. By WSI, w.l.o.g. we can assume that $T(X) = 1$. X can be converted into a completely segregated district by dividing each school n into G distinct schools, each of which includes all and only the members of a single ethnic group. By SDP, this procedure results in a weakly more segregated district. By then combining all schools containing a given ethnic group, this can be converted to $\overline{X}(P)$ without changing the segregation level (by SDP). To see that $\overline{X}^G \succcurlyeq X$, note that by Lemma 3, $\frac{1}{G} \biguplus_{j=1}^{G} \sigma^j(\overline{X}(P)) \succcurlyeq \overline{X}(P)$. But by SDP, the left hand side district is as segregated as $\overline{X}^G$. Q.E.D.

[26] Note that $\sigma^G(Z) = Z$.

Lemma 5 *For any integer $G \geq 1$, $\overline{X}^G \preccurlyeq \overline{X}^{G+1}$.*

Proof. Let X be the $(G+1)$-group district that results after splitting one ethnic group in $\overline{X}^G$ up into two equally distributed subgroups. By Lemma 4 and GDP, $\overline{X}^{G+1} \succcurlyeq X \sim \overline{X}^G$. Q.E.D.

Lemma 6 *Let X and X' be two districts with the same size and ethnic distribution such that $X \succ X'$. Let $1 \geq \alpha > \beta \geq 0$. Then $\alpha X \uplus (1-\alpha)X' \succ \beta X \uplus (1-\beta)X'$*

Proof. By WSI, $(\alpha - \beta)X \succ (\alpha - \beta)X'$. Since X and X' have the same size and ethnic distribution, so do $(\alpha - \beta)X$ and $(\alpha - \beta)X'$. So by IND1,

$$\beta X \uplus (\alpha - \beta)X \uplus (1-\alpha)X' \succ \beta X \uplus (\alpha - \beta)X' \uplus (1-\alpha)X'.$$

By SDP, $\alpha X \uplus (1-\alpha)X' \succ \beta X \uplus (1-\beta)X'$. Q.E.D.

Lemma 7 *For any districts $Z \succcurlyeq X \succcurlyeq Y$ such that $Z \succ Y$ and Y and Z have the same size and ethnic distribution, there is a unique $\alpha \in [0,1]$ such that $X \sim \alpha Z \uplus (1-\alpha)Y$.*

Proof. The sets $\{\alpha \in [0,1] : \alpha Z \uplus (1-\alpha)Y \succcurlyeq X\}$ and $\{\alpha \in [0,1] : X \succcurlyeq \alpha Z \uplus (1-\alpha)Y\}$ are closed by C. Any α satisfies $X \sim \alpha Z \uplus (1-\alpha)Y$ if and only if it is in the intersection of these two sets. Given that $Z \succcurlyeq X \succcurlyeq Y$, these sets are each nonempty. Their union is the whole unit interval since $\succcurlyeq$ is complete. Since the interval $[0,1]$ is connected, the intersection of the two sets must be nonempty. By Lemma 6, their intersection cannot contain more than one element. Thus, their intersection contains a single element α. Q.E.D.

Let X be a district with G groups and ethnic distribution $\widehat{P} = (\widehat{P}_1, ..., \widehat{P}_G)$. For any $G' \geq 1$ and any distribution $P = (P_1, ..., P_{G'})$ let $\phi^P(X)$ be the district that results after splitting each ethnic group g in district X into G' ethnic groups in proportions given by P. That is, the T_g^n members of each ethnic group g in each school n of X are split up into G' ethnic groups of size $P_1 T_g^n, ..., P_{G'} T_g^n$. The resulting district $\phi^P(X)$ has GG' groups with distribution $\left((\widehat{P}_g P_{g'})_{g=1}^G\right)_{g'=1}^{G'}$.

Let X be a district and let $\widehat{P} = (\widehat{P}_1, ..., \widehat{P}_G)$ be an arbitrary distribution such that $\overline{X}(\widehat{P}) \succcurlyeq X$ and $\overline{X}(\widehat{P}) \succcurlyeq \overline{X}^2$. By lemmas 4 and 5 such a distribution exists. By Non-triviality, Lemma 4, and Lemma 2, $\overline{X}^2 \succ \underline{X}^2 \sim \underline{X}(\widehat{P})$. Therefore, by Lemma 7 there is a unique $\widehat{\alpha}$ such that

$$X \sim \widehat{\alpha}\overline{X}(\widehat{P}) \uplus (1-\widehat{\alpha})\underline{X}(\widehat{P}). \tag{6}$$

Similarly, by Lemma 7 there is a unique $\widehat{\beta}$ such that $\overline{X}^2 \sim \widehat{\beta}\overline{X}(\widehat{P}) \uplus (1-\widehat{\beta})\underline{X}(\widehat{P})$. By Lemma 6, $\widehat{\beta} > 0$, as $\overline{X}(\widehat{P}) \succcurlyeq \overline{X}^2$.

Define the index $S : \mathcal{C} \to \Re$ by

$$S(X) = \widehat{\alpha}/\widehat{\beta} \tag{7}$$

For S to be well defined, $\widehat{\alpha}/\widehat{\beta}$ cannot depend on the particular choice of $\widehat{P}$. We now verify this. Consider another distribution $\widetilde{P} = (\widetilde{P}_1, ..., \widetilde{P}_{G'})$ such that $\overline{X}(\widetilde{P}) \succcurlyeq X$ and $\overline{X}(\widetilde{P}) \succcurlyeq \overline{X}^2$ and let $\widetilde{\alpha}$ and $\widetilde{\beta}$ the unique numbers such that $X \sim \widetilde{\alpha}\overline{X}(\widetilde{P}) \uplus (1-\widetilde{\alpha})\underline{X}(\widetilde{P})$ and $\overline{X}^2 \sim \widetilde{\beta}\overline{X}(\widetilde{P}) \uplus (1-\widetilde{\beta})\underline{X}(\widetilde{P})$. By GDP

$$X \sim \phi^{\widehat{P}}\left(\widetilde{\alpha}\overline{X}(\widetilde{P}) \uplus (1-\widetilde{\alpha})\underline{X}(\widetilde{P})\right) \sim \widetilde{\alpha}\phi^{\widehat{P}}\left(\overline{X}(\widetilde{P})\right) \uplus (1-\widetilde{\alpha})\phi^{\widehat{P}}\left(\underline{X}(\widetilde{P})\right) \tag{8}$$

Similarly, applying the transformation $\phi^{\widetilde{P}}$ to (6) and using GDP,

$$X \sim \widehat{\alpha}\phi^{\widetilde{P}}\left(\overline{X}(\widehat{P})\right) \uplus (1-\widehat{\alpha})\phi^{\widetilde{P}}\left(\underline{X}(\widehat{P})\right) \tag{9}$$

Both $\phi^{\widetilde{P}}\left(\underline{X}(\widehat{P})\right)$ and $\phi^{\widehat{P}}\left(\underline{X}(\widetilde{P})\right)$ are districts with the same number of groups $(G * G')$ and (up to a permutation) the same ethnic distribution. Further by SYM, $\phi^{\widetilde{P}}\left(\underline{X}(\widehat{P})\right) \sim \phi^{\widehat{P}}\left(\underline{X}(\widetilde{P})\right)$. Similarly, both $\phi^{\widetilde{P}}\left(\overline{X}(\widehat{P})\right)$ and $\phi^{\widehat{P}}\left(\overline{X}(\widetilde{P})\right)$ are districts with the same number of groups and (up to a permutation) the same ethnic distribution. Assume w.l.o.g. that $\phi^{\widehat{P}}\left(\overline{X}(\widetilde{P})\right) \succcurlyeq \phi^{\widetilde{P}}\left(\overline{X}(\widehat{P})\right)$ and let γ be the unique number such that

$$\phi^{\widetilde{P}}\left(\overline{X}(\widehat{P})\right) \sim \gamma\phi^{\widehat{P}}\left(\overline{X}(\widetilde{P})\right) \uplus (1-\gamma)\phi^{\widehat{P}}\left(\underline{X}(\widetilde{P})\right)$$

Then, applying WSI, IND1 (twice) and SDP, it follows from (9) that

$$\begin{aligned} X &\sim \widehat{\alpha}\left[\gamma\phi^{\widehat{P}}\left(\overline{X}(\widetilde{P})\right) \uplus (1-\gamma)\phi^{\widehat{P}}\left(\underline{X}(\widetilde{P})\right)\right] \uplus (1-\widehat{\alpha})\phi^{\widehat{P}}\left(\underline{X}(\widetilde{P})\right) \\ &\sim \widehat{\alpha}\gamma\phi^{\widehat{P}}\left(\overline{X}(\widetilde{P})\right) \uplus (1-\gamma\widehat{\alpha})\phi^{\widehat{P}}\left(\underline{X}(\widetilde{P})\right) \end{aligned} \tag{10}$$

Comparing (10) and (8) we obtain that $\widetilde{\alpha} = \widehat{\alpha}\gamma$. Exactly the same reasoning leads to $\widetilde{\beta} = \widehat{\beta}\gamma$. Consequently $\widehat{\alpha}/\widehat{\beta} = \widetilde{\alpha}/\widetilde{\beta}$. This establishes that S is well-defined.

Lemma 8 *The index S defined in (7) represents $\succcurlyeq$.*

Proof. Let $X, Y \in \mathcal{C}$ and let G be at least as large as the number of groups in X or Y. Then, by lemmas 4 and 5, $\overline{X}^G \succcurlyeq \overline{X}^2$, $\overline{X}^G \succcurlyeq X$ and $\overline{X}^G \succcurlyeq Y$. Define α_X, α_Y and β by

$$
\begin{aligned}
X &\sim \alpha_X \overline{X}^G \uplus (1-\alpha_X)\underline{X}^G \\
Y &\sim \alpha_Y \overline{X}^G \uplus (1-\alpha_Y)\underline{X}^G \\
\overline{X}^2 &\sim \beta \overline{X}^G \uplus (1-\beta)\underline{X}^G.
\end{aligned}
$$

Then,

$$
\begin{aligned}
X \succcurlyeq Y &\iff \alpha_X \overline{X}^G \uplus (1-\alpha_X)\underline{X}^G \succcurlyeq \alpha_Y \overline{X}^G \uplus (1-\alpha_Y)\underline{X}^G && \text{by definition of } \alpha_X \text{ and } \alpha_Y \\
&\iff \alpha_X \geq \alpha_Y && \text{by Lemma 6} \\
&\iff \alpha_X/\beta \geq \alpha_Y/\beta && \text{since } \beta > 0 \\
&\iff S(X) \geq S(Y) && \text{by definition of } S
\end{aligned}
$$

Q.E.D.

The following results will be used to show that S is the Mutual Information index.

Lemma 9 *For any ethnic distribution $P = (P_1, ..., P_G)$ (in which some entries may be zero), let $\widehat{P} = \left(\frac{P_1}{G}, ..., \frac{P_1}{G}, ..., \frac{P_G}{G}, ..., \frac{P_G}{G}\right)$ be the ethnic distribution that results from dividing each ethnic group in P into G equal sized groups. Then $\overline{X}(\widehat{P}) \succcurlyeq \overline{X}^G$ and $\overline{X}(\widehat{P}) \succcurlyeq \overline{X}(P)$.*

Proof. For the first claim, first subdivide each ethnic group in $\overline{X}^G$ into G groups in proportions given by P. For instance, the first group is divided into G groups of sizes $P_1\frac{1}{G}, ..., P_G\frac{1}{G}$. Now put each resulting group in a separate school. The group distribution of the resulting district, $\left(P_1\frac{1}{G}, ..., P_G\frac{1}{G}, ..., P_1\frac{1}{G}, ..., P_G\frac{1}{G}\right)$, is just a permutation of $\widehat{P}$. Hence, by GDP and SDP, $\overline{X}(\widehat{P}) \succcurlyeq \overline{X}^G$. The second claim follows from the first one after noting that by Lemma 4, $\overline{X}^G \succcurlyeq \overline{X}(P)$. Q.E.D.

Lemma 10 *Let districts Z_1, Z_2, Z_3, and Z_4 all have the same population and ethnic distribution and let $Z_1 \sim Z_2$ and $Z_3 \sim Z_4$. Let Z_5, Z_6 be two districts with equal populations. Then $Z_1 \uplus Z_5 \sim Z_2 \uplus Z_6$ if and only if $Z_3 \uplus Z_5 \sim Z_4 \uplus Z_6$.*

Proof. By IND2 applied twice, $Z_1 \uplus Z_5 \sim Z_1 \uplus Z_6$ if and only if $Z_3 \uplus Z_5 \sim Z_3 \uplus Z_6$. But by IND1, $Z_1 \uplus Z_6 \sim Z_2 \uplus Z_6$ and $Z_3 \uplus Z_6 \sim Z_4 \uplus Z_6$. Q.E.D.

Lemma 11 *For any districts X and Y, $S(X \uplus Y) = S(c(X) \uplus Y) + \frac{T(X)}{T(X)+T(Y)} S(X)$.*

Proof. Let X and Y be any two districts. Let $X \uplus Y$ have G ethnic groups. By adding an empty group if needed, we can assume WLOG that $G \geq 2$. For any district Z, let $\phi^G(Z)$ be the result of splitting each group g in Z into G equal-size groups, each of which has the same school distribution as g. Let $\widehat{P}$ be the group distribution of $\phi^G(X)$. By Lemma 9, $\overline{X}(\widehat{P}) \succcurlyeq \overline{X}^G$. Define $\widehat{\alpha}_X$ by $X \sim \widehat{\alpha}_X \overline{X}(\widehat{P}) \uplus (1 - \widehat{\alpha}_X) \underline{X}(\widehat{P})$ and γ by $c(X) \uplus Y \sim \gamma \overline{X}(\widehat{P}) \uplus (1 - \gamma) \underline{X}(\widehat{P})$. Define

$$\begin{aligned}
Z_1 &= \phi^G(X) \\
Z_2 &= T(X) \left(\widehat{\alpha}_X \overline{X}(\widehat{P}) \uplus (1 - \widehat{\alpha}_X) \underline{X}(\widehat{P}) \right) \\
Z_3 &= c(\phi^G(X)) = \phi^G(c(X)) \\
Z_4 &= T(X) \underline{X}(\widehat{P}) \\
Z_5 &= \phi^G(Y) \\
Z_6 &= T(X \uplus Y) \left(\gamma \overline{X}(\widehat{P}) \uplus \left(1 - \frac{T(X)}{T(X \uplus Y)} - \gamma \right) \underline{X}(\widehat{P}) \right)
\end{aligned}$$

To show that Z_6 is well defined, we must show that $\gamma \leq 1 - \frac{T(X)}{T(X \uplus Y)} = \frac{T(Y)}{T(X \uplus Y)}$. For this, by Lemma 6, it is enough to show that

$$\gamma \overline{X}(\widehat{P}) \uplus (1 - \gamma) \underline{X}(\widehat{P}) \preccurlyeq \frac{T(Y)}{T(X \uplus Y)} \overline{X}(\widehat{P}) \uplus \frac{T(X)}{T(X \uplus Y)} \underline{X}(\widehat{P}). \tag{11}$$

The district $c(X) \uplus Y$ has G groups since $X \uplus Y$ does. By Lemma 3,

$$c(X) \uplus Y \preccurlyeq \frac{1}{G} \biguplus_{j=1}^{G} \sigma^j (c(X) \uplus Y) = \frac{1}{G} \biguplus_{j=1}^{G} \sigma^j (c(X)) \uplus \frac{1}{G} \biguplus_{j=1}^{G} \sigma^j (Y)$$

Let $\widetilde{c(X)} = \frac{1}{G} \biguplus_{j=1}^{G} \sigma^j(c(X))$ and $\tilde{Y} = \frac{1}{G} \biguplus_{j=1}^{G} \sigma^j(Y)$. Each of $\widetilde{c(X)}$ and $\tilde{Y}$ has G groups of equal size. By SDP, $\widetilde{c(X)} \sim T(X)\underline{X}^G$ and both of these districts have the same population, $T(X)$, and the same ethnic distribution. Since $\tilde{Y}$ has G equal size groups, it is not more segregated than $T(Y)\overline{X}^G$ and both of these districts have the same population, $T(Y)$, and the same group distribution. Therefore,

$$\begin{array}{rcll} c(X) \uplus Y & \preccurlyeq & T(X)\underline{X}^G \uplus T(Y)\overline{X}^G & \text{by IND1 (twice)} \\ & \sim & \phi^G \left(T(X)\underline{X}^G \uplus T(Y)\overline{X}^G\right) & \text{by GDP} \\ & \sim & T(X)\phi^G \left(\underline{X}^G\right) \uplus T(Y)\phi^G \left(\overline{X}^G\right) & \text{by definition of } \phi^G \\ & \preccurlyeq & T(X)\underline{X}(\widehat{P}) \uplus T(Y)\overline{X}(\widehat{P}) & \text{by SDP} \\ & \sim & \frac{T(X)}{T(X \uplus Y)}\underline{X}(\widehat{P}) \uplus \frac{T(Y)}{T(X \uplus Y)}\overline{X}(\widehat{P}) & \text{by WSI} \end{array}$$

But $c(X) \uplus Y \sim \gamma\overline{X}(\widehat{P}) \uplus (1-\gamma)\underline{X}(\widehat{P})$ so (11) holds and $\gamma \leq \frac{T(Y)}{T(X \uplus Y)}$, as claimed.

By construction, Z_1, Z_2, Z_3, and Z_4 all have the same population and ethnic distribution. By GDP, $Z_1 \sim Z_2$. Clearly, $Z_3 \sim Z_4$ since these are actually the same district. Also, the population of Z_6 is $T(Y)$, which equals the population of Z_5. Moreover,

$$\begin{array}{rcll} Z_4 \uplus Z_6 & = & T(X)\underline{X}(\widehat{P}) \uplus T(X \uplus Y)\left(\gamma\overline{X}(\widehat{P}) \uplus \left(1 - \frac{T(X)}{T(X \uplus Y)} - \gamma\right)\underline{X}(\widehat{P})\right) & \\ & = & T(X \uplus Y)\left(\gamma\overline{X}(\widehat{P}) \uplus (1-\gamma)\underline{X}(\widehat{P})\right) & \text{by SDP} \\ & \sim & c(X) \uplus Y & \text{by WSI} \\ & \sim & \phi^G(c(X)) \uplus \phi^G(Y) & \text{by GDP} \\ & = & Z_3 \uplus Z_5 & \end{array}$$

So by Lemma 10,

$$Z_1 \uplus Z_5 \sim Z_2 \uplus Z_6 \tag{12}$$

Now,

$$
\begin{aligned}
X \uplus Y &\sim \phi^G(X \uplus Y) \text{ by GDP} \\
&= Z_1 \uplus Z_5 \\
&\sim Z_2 \uplus Z_6 \text{ by (12)} \\
&= T(X)\left(\widehat{\alpha}_X \overline{X}(\widehat{P}) \uplus (1-\widehat{\alpha}_X)\underline{X}(\widehat{P})\right) \\
&\uplus T(X \uplus Y)\left(\gamma \overline{X}(\widehat{P}) \uplus \left(1 - \frac{T(X)}{T(X \uplus Y)} - \gamma\right)\underline{X}(\widehat{P})\right) \\
&\sim (T(X \uplus Y)\gamma + T(X)\widehat{\alpha}_X)\,\overline{X}(\widehat{P}) \uplus T(X \uplus Y)\left(1 - \gamma - \frac{T(X)}{T(X \uplus Y)}\widehat{\alpha}_X\right)\underline{X}(\widehat{P}) \text{ by SDP} \\
&\sim \left(\gamma + \frac{T(X)}{T(X \uplus Y)}\widehat{\alpha}_X\right)\overline{X}(\widehat{P}) \uplus \left(1 - \gamma - \frac{T(X)}{T(X \uplus Y)}\widehat{\alpha}_X\right)\underline{X}(\widehat{P}) \text{ by WSI.}
\end{aligned}
$$

We have shown that $X \uplus Y \sim \left(\gamma + \frac{T(X)}{T(X \uplus Y)}\widehat{\alpha}_X\right)\overline{X}(\widehat{P}) \uplus \left(1 - \gamma - \frac{T(X)}{T(X \uplus Y)}\widehat{\alpha}_X\right)\underline{X}(\widehat{P})$. By definition of γ and $\widehat{\alpha}_X$, $c(X) \uplus Y \sim \gamma\overline{X}(\widehat{P}) \uplus (1-\gamma)\underline{X}(\widehat{P})$ and $X \sim \widehat{\alpha}_X\overline{X}(\widehat{P}) \uplus (1-\widehat{\alpha}_X)\underline{X}(\widehat{P})$. By Lemma 7, there is a unique β such that $\overline{X}^2 \sim \beta\overline{X}(\widehat{P}) \uplus (1-\beta)\underline{X}(\widehat{P})$. By definition of S, $S(X \uplus Y) = \frac{1}{\beta}\left(\gamma + \frac{T}{T+T(Y)}\widehat{\alpha}_X\right) = S(c(X) \uplus Y) + \frac{T}{T+T(Y)}S(X)$, as claimed. Q.E.D.

For any discrete probability distribution $P = (P_1, ..., P_G)$, define the function $s(P)$ to equal $S(\overline{X}(P))$.

Claim 2 *The function s is the entropy function. Namely, $s(P) = h(P) = \sum_{i=1}^n P_i \log_2 \frac{1}{P_i}$.*

Proof. It is known that the entropy function is the only function that satisfies the following three properties.[27]

1. $h(1/2, 1/2) = 1$.

2. $h(p, 1-p)$ is continuous in p.

[27] The statement of this result appears as an exercise in Cover and Thomas [14]. For the original proof, see Faddeev [17].

3. $h(p_1, ..., p_n) = h(p_1 + p_2, p_3, ..., p_n) + (p_1 + p_2)\, h\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$.

So it is enough to show that s satisfies them. Property 1 follows from the definition of S and the fact that $S(\overline{X}(1/2, 1/2)) = S(\overline{X}^2)$. Property 3 follows from Lemma 11. It remains to show property 2. Let us write $\overline{X}(p, 1-p)$ as Z^p for brevity. By Lemma 7, there is a unique α_p such that $Z^p \sim \alpha_p \overline{X}^2 \uplus (1-\alpha_p)\underline{X}^2$. By definition of S, $\alpha_p = S(\overline{X}(p, 1-p))$. For all p, the sets $\{q : Z^q \succcurlyeq Z^p\}$ and $\{q : Z^q \preccurlyeq Z^p\}$ are closed by Continuity. Note that $Z^q \succcurlyeq Z^p$ if and only if $\alpha_q \geq \alpha_p$ by Lemma 6. So the sets $\{q : \alpha_q \geq \alpha_p\}$ and $\{q : \alpha_q \leq \alpha_p\}$ are closed. If α_p is not a continuous function of p, then let the sequence $(p_k)_{k=1}^{\infty}$ converge to some p. By restricting to an appropriate subsequence, we may assume that $\lim_{k\to\infty} \alpha_{p_k}$ exists. Let this limit be c and assume by contradiction that $c \neq \alpha_p$. Assume that $c > \alpha_p$ (the other case is analogous). Since $\lim_{k\to\infty} \alpha_{p_k} = c > \frac{c+\alpha_p}{2}$, there is a k^* such that $\alpha_{p_k} > \frac{c+\alpha_p}{2}$ for all $k > k^*$. So the sequence $\{p_k : k > k^*\}$ lies in $\left\{q : \alpha_q \geq \frac{c+\alpha_p}{2}\right\}$. But $\lim_{k\to\infty} p_k = p$ does not lie in this set, which contradicts the fact that this set is closed. Q.E.D.

We now show that S is the Mutual Information index. Consider any district X with N schools, G ethnic groups, and ethnic distribution P. Let $X_0 = X$. Let X_n be the result of separating the students in each school $m \leq n$ into G uniracial schools. For instance, if $X = \langle (1,2), (3,4) \rangle$, then $X_1 = \langle (1,0), (0,2), (3,4) \rangle$ and $X_2 = \langle (1,0), (0,2), (3,0), (0,4) \rangle$. Note that X_N is completely segregated and has ethnic distribution P, so $X_N \sim \overline{X}(P)$. By Lemma 11,

$$S(X_n) = S(X_{n-1}) + P^n S(\overline{X}(p^n)) \qquad \text{for } n = 1, ..., N.$$

Thus,

$$S(X_N) = S(X) + \sum_{n=1}^{N} P^n S(\overline{X}(p^n))$$
$$\implies S(X) = S(X_N) - \sum_{n=1}^{N} P^n S(\overline{X}(p^n))$$
$$= S(\overline{X}(P)) - \sum_{n=1}^{N} P^n S(\overline{X}(p^n))$$
$$= \sum_{g=1}^{G} P_g \log_2 \frac{1}{P_g} - \sum_{n=1}^{N} P^n \sum_{g=1}^{G} p_g^n \log_2 \frac{1}{p_g^n}.$$

where the last line follows from Claim 2. This completes the proof of Theorem 1. Q.E.D.

Proof of Proposition 1: IND1: Let X and Y have the same size and ethnic distribution, and let Z be another district. Then $c(X) = c(Y)$ and $T(X)/T(X \uplus Z) = T(Y)/T(Y \uplus Z) = p$. Then, applying SSD, $M(X \uplus Z) \geq M(Y \uplus Z)$ if and only if

$$M(c(X) \uplus c(Z)) + pM(X) + (1-p)M(Z) \geq M(c(Y) \uplus c(Z)) + pM(Y) + (1-p)M(Z)$$
$$\Leftrightarrow M(X) \geq M(Y)$$

IND2: Let $W, X, Y \in \mathcal{C}$ be three districts such that $T(W) = T(X)$. Then, $T(W)/T(W \uplus Y) = T(X)/T(X \uplus Y) = p$. Now, applying SSD,

$$M(W \uplus c(Y)) \geq M(X \uplus c(Y)) \Leftrightarrow M(c(W) \uplus c(Y)) + pM(W) \geq M(c(X) \uplus c(Y)) + pM(X)$$
$$\Leftrightarrow M(c(W) \uplus c(Y)) + pM(W) + (1-p)M(Y) \geq M(c(X) \uplus c(Y)) + pM(X) + (1-p)M(Y)$$
$$\Leftrightarrow M(W \uplus Y) \geq M(X \uplus Y)$$

The proof of GDP is similar and is left to the reader. Q.E.D.

Proof of Proposition 2: Let $X = X^1 \uplus \cdots \uplus X^K$ be district composed of K clusters. By definition of M,

$$M(X) = h(P(X)) - \sum_{k=1}^{K} \sum_{n \in \mathbf{N}(X^k)} P^n h(p^n)$$

Subtracting and adding $\sum_{k=1}^{K} P^k h(P(X^k))$ on the right hand side, we obtain

$$\begin{aligned} M(X) &= h(P(X)) - \sum_{k=1}^{K} P^k h(P(X^k)) + \sum_{k=1}^{K} P^k h(P(X^k)) - \sum_{k=1}^{K} \sum_{n \in \mathbf{N}(X^k)} P^n h(p^n) \\ &= h(P(X)) - \sum_{k=1}^{K} P^k h(P(X^k)) + \sum_{k=1}^{K} P^k \left(h(P(X^k)) - \sum_{n \in \mathbf{N}(X^k)} P^n h(p^n) \right) \\ &= M(c(X^1) \uplus \cdots \uplus c(X^K)) + \sum_{k=1}^{K} P^k M(X^k). \end{aligned}$$

This shows that M satisfies SSD. That M satisfies SGD as well now follows from the symmetry of mutual information (Cover and Thomas [14, pp. 18 ff.]). Q.E.D.

Proof of Proposition 3: Suppose S satisfies SSD and properties 1-3. Let the maximum value of S be attained by the district $\overline{X}$. Define another district, $\overline{X}'$, that is a copy of $\overline{X}$ in which each ethnic group has been replaced by a new ethnic group not in $\overline{X}$. (For instance, if $\overline{X}$ has K groups that go to K separate schools, then let $\overline{X}'$ consist of a *different* K groups that go to K separate schools.) Then by SSD, $S(\overline{X} \uplus \overline{X}') = S(c(\overline{X}) \uplus c(\overline{X}')) + \frac{1}{2}S(\overline{X}) + \frac{1}{2}S(\overline{X}')$. The sum of the second and third terms equals $S(\overline{X})$ by symmetry. But the first term is strictly positive by property 3. This contradicts the hypothesis that S attains its maximum at $\overline{X}$. Q.E.D.

References

[1] Atkinson, A.B. 1970. "On the Measurement of Inequality." *Journal of Economic Theory* 2: 244-63.

[2] Bell, Wendell. 1954. "A Probability Model for the Measurement of Ecological Segregation." *Social Forces* 32:357-364.

[3] Boozer, Michael A., Alan B. Krueger, and Shari Wolkin. 1992. "Race and School Quality Since Brown v. Board of Education." *Brookings Papers on Economic Activity: Microeconomics 1992*, pp. 269-338.

[4] Bourguignon, F. 1979. "Decomposable Income Inequality Measures." *Econometrica* 47, 901-920.

[5] Card, David, and Jesse Rothstein. 2006. "Racial Segregation and the Black-White Test Score Gap." NBER Working Paper No. 12078.

[6] Clotfelter, Charles T. 1979. "Alternative Measures of School Desegregation: A Methodological Note." *Land Economics* 54: 373-380.

[7] Clotfelter, Charles T. 1999. "Public School Segregation in Metropolitan Areas." *Land Economics* 75:487-504.

[8] Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2002. "Segregation and Resegregation in North Carolina's Public School Classrooms." Terry Sanford Institute of Public Policy working paper no. SAN02-03, Duke University.

[9] Coleman, James S., Thomas Hoffer, and Sally Kilgore. 1982. "Achievement and Segregation in Secondary Schools: A Further Look at Public and Private School Differences." *Sociology of Education* 55:162-182.

[10] Coleman, James S., Sara D. Kelly, and John Moore. 1975. *Trends in School Segregation, 1968-1973.* Washington, D.C.: Urban Institute.

[11] Collins, Chiquita, and David R. Williams. 1999. "Segregation and Mortality: The Deadly Effects of Racism." *Sociological Forum* 74: 495–523.

[12] Cortese, Charles F., R. Frank Falk, and Jack K. Cohen. 1976. "Further Considerations on the Methodological Analysis of Segregation Indices." *American Sociological Review* 41:630-637.

[13] Cotter, David A., JoAnn DeFiore, Joan M. Hermsen, Brenda M. Kowalewski, and Reeve Vanneman. 1997. "All Women Benefit: The Macro-Level Effect of Occupational Integration on Gender Earnings Equality." *American Sociological Review* 62:714–734.

[14] Cover, Thomas M. and Joy A. Thomas. (1991) *Elements of Information Theory.* United States: Wiley.

[15] Cutler, David, and Edward Glaeser. 1997. "Are Ghettos Good or Bad?" *Quarterly Journal of Economics* 112: 827– 872.

[16] Echenique, Federico, and Roland G. Fryer, Jr. 2007. "A Measure of Segregation Based on Social Interactions." *Quarterly Journal of Economics* (forthcoming).

[17] Faddeev, D.K. 1956. "On the Notion of Entropy of a Finite Probability Schema [in Russian]." *Usp. Mat. Nauk* 11:227-231.

[18] Flückiger, Yves and Jacques Silber 1999. "The Measurement of Segregation in the Labor Force." Physica-Verlag. Heidelberg New York.

[19] Foster, James E. 1983. "An Axiomatic Characterization of the Theil Measure of Income Inequality." *Journal of Economic Theory* 65, 105-121.

[20] Frankel, David M., and Oscar Volij. 2008. "An Axiomatization of the Multigroup Atkinson Segregation Indices." Mimeo.

[21] Fuchs, Victor. 1975. "A Note on Sex Segregation in Professional Occupations." *Explorations in Economic Research* 2:105-111.

[22] Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2004. "New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement." Mimeo.

[23] Herranz, Neus, Ricardo Mora, and Javier Ruiz-Castillo. 2005. "An Algorithm to Reduce the Occupational Space in Gender Segregation Studies." *Journal of Applied Econometrics* 20:25-37.

[24] Hoxby, Caroline M. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." NBER Working Paper 7867.

[25] Hutchens, Robert. 2001. "Numerical Measures of Segregation: Desirable Properties and Their Implications." *Mathematical Social Sciences* 42:13–29.

[26] Hutchens, Robert. 2003. "One Measure of Segregation." *International Economic Review* 45:555–578.

[27] Jahn, Julius, and Schmid, Calvin F., and Schrag, Clarence. 1947. "The Measurement of Ecological Segregation." *American Sociological Review* 12:293–303.

[28] James, Franklin J. 1986. "A New Generalized 'Exposure-based' Segregation Index." *Sociological Methods and Research* 14:301-316.

[29] James, David R., and Taueber, Karl E. 1985. "Measures of Segregation." In *Sociological Methodology* Jossey-Bass. Nancy Tuma Ed.

[30] Johnson, Ron, Michael Poulsen, and James Forrest. 2007. Ethnic and Racial Segregation in U.S. Metropolitan Areas, 1980-2000: The Dimensions of Segregation Revisited. Urban Affairs Review 42:479-504.

[31] Kinder, Donald R., and Tali Mendelberg. 1995. "Cracks in American Apartheid: The Political Impact of Prejudice among Desegregated Whites." *Journal of Politics* 57:402–424.

[32] Lewis, Gregory B. 1996. "Gender Integration of Occupations in the Federal Civil Service: Extent and Effects on Male-Female Earnings." *Industrial and Labor Relations Review* 49:472–483.

[33] Lieberson, Stanley. 1969. "Measuring Population Diversity." *American Sociological Review* 34:850-862.

[34] Macpherson, David A., and Barry T. Hirsh, 1995. "Wages and Gender Composition: Why do Women's jobs Pay Less? " *Journal of Labor Economics* 13:426–471.

[35] Massey, Douglas S. and Nancy A. Denton. 1988. “The Dimensions of Racial Segregation.” *Social Forces* 67:281–315.

[36] Mora, Ricardo, and Javier Ruiz-Castillo. 2003. “Additively Decomposable Segregation Indices: The Case of Gender Segregation by Occupations and Human Capital Levels in Spain.” *Journal of Economic Inequality* 1: 147–179.

[37] Mora, Ricardo, and Javier Ruiz-Castillo. 2005. “The Axiomatic Properties of an Entropy Based Index of Segregation.” Universidad Carlos III de Madrid working paper no. 05-62.

[38] Mora, Ricardo, and Javier Ruiz-Castillo. 2005. “An Evaluation of an Entropy Based Index of Segregation.” Mimeo, Universidad Carlos III de Madrid.

[39] Mora, Ricardo, and Javier Ruiz-Castillo. 2004. "Gender Segregation by Occupations in the Public and the Private Sector: The Case of Spain." *Investigaciones Economićas* 28:399-428.

[40] Morgan, Barrie S. 1975. “The Segregation of Socioeconomic Groups in Urban Areas: A Comparative Analysis.” *Urban Studies* 12:47-60.

[41] Philipson, Tomas. 1993. “Social Welfare and Measurement of Segregation.” *Journal of Economic Theory* 60:322-334.

[42] Reardon, Sean F. 1998. “Measures of Racial Diversity and Segregation in Multigroup and Hierarchically Structured Populations.” Presented at the Annual Meeting of the Eastern Sociological Society, Philadelphia, PA.

[43] Reardon, Sean F., and Glenn Firebaugh. 2002. “Measures of Multigroup Segregation.” *Sociological Methodology* 32:33-67.

[44] Reardon, Sean F., and John T. Yun. 2001. “Suburban Racial Change and Suburban School Segregation, 1987-95.” *Sociology of Education* 74:79-101.

[45] Rivkin, Steven G. 1994. “Residential Segregation and School Integration.” *Sociology of Education* 67:279-92.

[46] Sable, J., Gaviola, N., and Garofano, A. 2007. *Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 200506 (NCES 2007-365).* U.S. Department of Education, Washington, D.C.: National Center for Education Statistics.

[47] Sakoda. J. 1981. “A Generalized Index of Dissimilarity.” *Demography* 18:245-50.

[48] Taeuber, Karl E., and David R. James. 1982. “Racial Segregation Among Public and Private Schools.” *Sociology of Education* 55:133-143.

[49] Taeuber, Karl E., and Alma F. Taeuber. 1976. “A Practitioner’s Perspective on the Index of Dissimilarity.” *American Sociological Review* 41:884-889

[50] Theil, Henri. 1971. “Principles of Econometrics.” Wiley & Sons. New York.

[51] Theil, Henri. 1972. “Statistical Decomposition Analysis.” North Holland.

[52] Theil, Henri., and A. J. Finizza 1971. “A Note on the Measurement of Racial Integration in Schools.” *Journal of Mathematical Sociology* 1:187–1193.

[53] U.S. Census Bureau. 2007. *Current Population Survey 2007, Annual Social and Economic Supplement, Detailed Poverty Tables*, Table POV01.

[54] Zoloth, Barbara S. 1976. “Alternative Measures of School Segregation.” *Land Economics* 52:278-298.

	Mutual Information	Entropy Index	Weighted Dissimilarity	Gini
Symbol	M	H	D^W	G
Defined by	[50]	[51, 52]	[27, 40, 47]	[27, 42]
Used by	[21, 36, 39]	[44, 48, 54]	[7, 44, 48, 54]	[7, 44, 48]
Formula	$h(P) - \sum_n P^n h(p^n)$	$1 - \frac{\sum_n P^n h(p^n)}{h(P)}$	$\frac{1}{2I} \sum_g \sum_n P^n \left\|p_g^n - P_g\right\|$	$\frac{1}{2I} \sum_g \sum_m \sum_n P^m P^n \left\|p_g^m - p_g^n\right\|$
SYM	✓	✓	✓	✓
WSI	✓	✓	✓	✓
SDP	✓	✓	✓	✓
IND1	✓	✓	×	×
IND2	✓	×	×	×
GDP	✓	×	×	×
C	✓	✓	✓	✓
N	✓	✓	✓	✓
SI	×	×	2	2
SSD	✓	×	×	×
SGD	✓	×	×	×
Name	Normalized Exposure	Clotfelter	Card-Rothstein	Symmetric Atkinson
Symbol	P	C	CR	A
Defined by	[2, 28]	[7]	[5]	[20, 29]
Used by	[7, 10, 9, 8, 54]	[3, 7, 8]	[5]	[30]
Formula	$\sum_g \sum_n P^n \frac{(p_g^n - P_g)^2}{1-P_g}$	$\frac{1}{T_2} \sum_{n: p_2^n \geq \kappa} T_2^n$	$\sum_n \left(\frac{T_2^n}{T_2} - \frac{T_1^n}{T_1}\right) \frac{T_2^n + T_3^n}{T^n}$	$1 - \sum_n \left(\prod_{g \in G} t_g^n\right)^{\frac{1}{\|G\|}}$
SYM	✓	×	×	✓
WSI	✓	✓	✓	✓
SDP	✓	×	×	✓
IND1	2	✓	×	✓
IND2	×	×	×	×
GDP	×	N/A	N/A	×
C	✓	✓	✓	✓
N	✓	✓	✓	✓
SI	×	×	×	✓
SSD	×	×	×	×
SGD	×	N/A	N/A	×

Table 1: Properties of School Segregation Indices. A check mark indicates that the property is satisfied by the index. An “×” indicates that it is not. “2”indicates that it is satisfied only in the 2-group case. The notation I denotes the Simpson Iteraction Index, $I = \sum_{g \in \mathbf{G}} P_g(1 - P_g)$ (Lieberson [33]). The properties are Symmetry (SYM), Weak Scale Invariance (WSI), the School Division Property (SDP), Type I Independence (IND1), Type II Independence (IND2), the Group Division Property (GDP), Continuity (C), Nontriviality (N), Scale Invariance (SI), Strong School Decomposability (SSD), and Strong Group Decomposability (SGD).

DECOMPOSITION OF URBAN SCHOOL SEGREGATION IN U.S.				
	1	2	3	4
Geographic Level	Asian vs. White, Black, and Hispanic	White vs. Black and Hispanic	Black vs. Hispanic	Total
A. ABSOLUTE MUTUAL INFORMATION INDICES				
1 Between States in US	0.030	0.089	0.093	0.211
2 Between CBSAs in States	0.013	0.067	0.028	0.108
3 Between Districts in CBSAs	0.022	0.173	0.023	0.219
4 Between Schools in Districts	0.016	0.073	0.038	0.127
5 Total: Between Urban Schools in U.S.	0.081	0.402	0.181	0.665
B. PERCENTAGES OF TOTAL URBAN SEGREGATION				
6 Between States in US	4.4%	13.3%	13.9%	31.7%
7 Between CBSAs in States	2.0%	10.1%	4.2%	16.2%
8 Between Districts in CBSAs	3.4%	26.1%	3.4%	32.9%
9 Between Schools in Districts	2.4%	11.0%	5.7%	19.2%
10 Total: Between Urban Schools in U.S.	12.2%	60.6%	27.3%	100.0%

Table 2: Decomposition of Segregation Between Urban Schools in U.S., 2005-6 School Year. Analysis is restricted to K-12 districts that contain at least two schools. Schools not located in CBSA's or that do not lie in the 50 U.S. states and the District of Columbia are excluded. Data are from the Common Core of Data (CCD). Mutual Information index is computed for all schools in universe defined above and decomposed into various components. Ethnic groups are mutually exclusive: Asians (including American Indians/Alaskan natives), (non-Hispanic) whites, (non-Hispanic) blacks, and Hispanics. The three terms in equation (4) appear in columns 1, 2, and 3. Column 1 shows how much segregation at the given geographic level is due (in an accounting sense) to segregation between Asians and non-Asians. Column 2 shows the contribution of segregation between whites, on the one hand, and blacks and Hispanics on the other. Column 3 shows the contribution of segregation between blacks and Hispanics. For precise definitions, see text. The sum of these numbers, appears in column 4 and (by Strong Group Decomposability) represents segregation between the four ethnic groups at the given geographic level. Four geographic levels are used. The first row computes segregation between states, treating each state as a single "school". For row 2, the Mutual Information index across CBSA's is first computed for each state. We then compute the average of these 51 indices, weighted by state population. This average, which appears in row 2, is the within-state, between-CBSA segregation. Row 3, computed analogously, shows segregation within CBSA's, between districts. Row 4 shows segregation within districts, between schools. By Strong School Decomposability, the sum of rows 1-4 equals total segregation between schools in the U.S., which appears in row 5. Total segregation among the four groups across schools in the U.S. is 0.665 (row 5, column 4). In panel B, all indices are re-expressed as percentages of this total.

URBAN SCHOOL SEGREGATION BETWEEN PAIRS OF ETHNIC GROUPS						
	1	2	3	4	5	6
	Asian vs.			White vs.		Black vs.
Geographic Level	White	Black	Hispanic	Black	Hispanic	Hispanic
A. MUTUAL INFORMATION INDICES						
Between States in US	0.059	0.173	0.087	0.062	0.168	0.243
Between CBSAs in States	0.024	0.036	0.056	0.060	0.069	0.073
Between Districts in CBSAs	0.040	0.106	0.091	0.183	0.115	0.060
Between Schools in Districts	0.022	0.068	0.067	0.077	0.060	0.100
Total: Between Urban Schools in U.S.	0.144	0.383	0.301	0.382	0.411	0.475

Table 3: Urban School Segregation between Pairs of Ethnic Groups, 2005-6 School Year. Analysis is restricted to K-12 districts that contain at least two schools. Schools not located in CBSA's or that do not lie in the 50 U.S. states and the District of Columbia are excluded. Data are from the Common Core of Data (CCD). Ethnic groups are mutually exclusive: Asians (including American Indians/Alaskan natives), (non-Hispanic) whites, (non-Hispanic) blacks, and Hispanics. The first row computes segregation between states, treating each state as a single "school". For row 2, the Mutual Information index across CBSA's is first computed for each state. We then compute the average of these 51 indices, weighted by state population. This average, which appears in row 2, is the within-state, between-CBSA segregation. Row 3, computed analogously, shows segregation within CBSA's, between districts. Row 4 shows segregation within districts, between schools. By Strong School Decomposability, the sum of rows 1-4 equals total segregation between schools in the U.S., which appears in row 5.

RANK CORRELATIONS (KENDALL'S TAU-B)									
INDEX	M	H	D	G	P	C90	C50	CR	A
Mutual Information (M)	1	0.668	0.521	0.539	0.859	0.467	0.5	0.706	0.23
Entropy Index (H)	0.668	1	0.817	0.843	0.733	0.418	0.352	0.603	0.453
Weighted Dissimilarity (D)	0.521	0.817	1	0.918	0.602	0.341	0.253	0.51	0.472
Gini (G)	0.539	0.843	0.918	1	0.622	0.367	0.271	0.524	0.473
Normalized Exposure (P)	0.859	0.733	0.602	0.622	1	0.462	0.485	0.727	0.258
Clotfelter (90% threshold) (Cl90)	0.467	0.418	0.341	0.367	0.462	1	0.629	0.46	0.243
Clotfelter (50% threshold) (Cl50)	0.5	0.352	0.253	0.271	0.485	0.629	1	0.502	0.129
Card-Rothstein (CR)	0.706	0.603	0.51	0.524	0.727	0.46	0.502	1	0.214
Symmetric Atkinson (A)	0.23	0.453	0.472	0.473	0.258	0.243	0.129	0.214	1
Mean (diagonal excluded)	0.561	0.611	0.554	0.570	0.594	0.423	0.390	0.531	0.309

Table 4: Kendall's Rank Correlation (τ_b) Between Pairs of Multigroup Segregation Indices, 2005-6 School Year. $C50$ and $C90$ refer to Clotfelter index with thresholds $\kappa = 0.5, 0.9$, respectively. Universe is set of Core Based Statistical Areas that lie in 50 U.S. states and District of Columbia. Schools that do not lie in a K-12 district that contains at least two schools are excluded. Data are from the Common Core of Data (CCD). Ethnic groups are mutually exclusive: Asians (including American Indians/Alaskan natives), (non-Hispanic) whites, (non-Hispanic) blacks, and Hispanics.

Rank by M	Abbreviated CBSA Name	Segregation of Public Schools Within Large CBSA's (Various Indices)									h
		M	H	D	G	P	CI90	CI50	CR	A	
1	Chicago	0.905	0.516	0.696	0.847	0.549	0.777	0.923	0.668	0.847	1.753
2	New York	0.728	0.390	0.596	0.755	0.406	0.586	0.859	0.612	0.755	1.866
3	Milwaukee	0.685	0.448	0.674	0.804	0.524	0.439	0.880	0.635	0.804	1.529
4	Cleveland	0.633	0.527	0.738	0.873	0.596	0.517	0.871	0.682	0.873	1.202
5	Detroit	0.632	0.506	0.695	0.832	0.583	0.685	0.802	0.695	0.832	1.248
6	Philadelphia	0.586	0.388	0.602	0.764	0.448	0.450	0.744	0.558	0.764	1.510
7	L.A.	0.552	0.349	0.570	0.736	0.357	0.497	0.892	0.469	0.736	1.583
8	Miami-Ft. Lauderdale	0.530	0.315	0.534	0.695	0.348	0.555	0.912	0.347	0.695	1.682
9	Atlanta	0.524	0.332	0.549	0.716	0.381	0.450	0.773	0.469	0.716	1.579
10	Washington	0.519	0.292	0.511	0.667	0.324	0.458	0.767	0.470	0.667	1.776
11	St. Louis	0.516	0.461	0.670	0.827	0.548	0.528	0.751	0.609	0.827	1.119
12	Baltimore	0.511	0.371	0.593	0.761	0.457	0.519	0.744	0.542	0.761	1.376
13	San Francisco	0.506	0.259	0.470	0.623	0.257	0.172	0.721	0.360	0.623	1.956
14	Houston	0.491	0.279	0.493	0.649	0.289	0.426	0.815	0.402	0.649	1.759
15	Memphis	0.486	0.398	0.663	0.793	0.463	0.673	0.838	0.514	0.793	1.221
16	Dallas-Ft. Worth	0.471	0.273	0.489	0.647	0.300	0.317	0.755	0.385	0.647	1.726
17	Boston	0.459	0.350	0.592	0.749	0.397	0.228	0.645	0.479	0.749	1.312
18	Kansas City	0.423	0.336	0.566	0.712	0.405	0.313	0.698	0.515	0.712	1.258
19	Denver	0.407	0.272	0.521	0.657	0.329	0.170	0.673	0.370	0.657	1.496
20	Columbus	0.375	0.349	0.571	0.752	0.400	0.259	0.608	0.483	0.752	1.075
21	Indianapolis	0.374	0.333	0.603	0.749	0.381	0.149	0.617	0.463	0.749	1.123
22	Cincinnati	0.368	0.428	0.655	0.814	0.475	0.365	0.660	0.560	0.814	0.861
23	Providence	0.352	0.304	0.572	0.722	0.383	0.016	0.489	0.380	0.722	1.159
24	San Diego	0.350	0.200	0.419	0.561	0.215	0.086	0.689	0.248	0.561	1.747
25	Nashville	0.337	0.275	0.541	0.691	0.330	0.117	0.611	0.389	0.691	1.228
26	Austin	0.332	0.208	0.442	0.580	0.240	0.223	0.650	0.293	0.580	1.592
27	San Antonio	0.325	0.242	0.489	0.639	0.269	0.236	0.834	0.233	0.639	1.343
28	Orlando	0.312	0.186	0.400	0.533	0.208	0.243	0.655	0.259	0.533	1.684
29	Carlotte	0.311	0.208	0.441	0.586	0.261	0.133	0.624	0.321	0.586	1.495
30	Jacksonville, FL	0.306	0.227	0.458	0.618	0.290	0.201	0.589	0.347	0.618	1.349
31	Sacramento	0.301	0.164	0.380	0.510	0.178	0.011	0.461	0.209	0.510	1.839
32	Minneapolis-St. Paul	0.288	0.234	0.476	0.630	0.289	0.019	0.342	0.289	0.630	1.232
33	Tampa-St. Petersburg	0.281	0.189	0.410	0.553	0.213	0.117	0.528	0.276	0.553	1.487
34	Pittsburgh	0.281	0.383	0.643	0.785	0.405	0.191	0.568	0.456	0.785	0.733
35	Virginia Beach	0.254	0.184	0.414	0.563	0.241	0.129	0.723	0.281	0.563	1.380
36	Riverside, CA	0.252	0.162	0.387	0.523	0.185	0.141	0.785	0.212	0.523	1.553
37	Las Vegas-Paradise	0.242	0.136	0.348	0.468	0.165	0.090	0.610	0.188	0.468	1.781
38	Phoenix	0.199	0.146	0.360	0.493	0.189	0.015	0.265	0.138	0.493	1.362
39	Seattle	0.199	0.137	0.347	0.476	0.155	0.009	0.180	0.194	0.476	1.453
40	Portland, OR	0.172	0.137	0.345	0.476	0.133	0.005	0.243	0.186	0.476	1.252

Table 5: Segregation of Public Schools Within CBSA's (Various Indices), 2005-6 School Year. $C50$ and $C90$ refer to Clotfelter index with thresholds $\kappa = 0.5, 0.9$, respectively. Universe is set of Core Based Statistical Areas that lie in 50 U.S. states and District of Columbia with at least 200,000 students. Schools that do not lie in a K-12 district that contains at least two schools are excluded. Data are from the Common Core of Data (CCD). Ethnic groups are mutually exclusive: Asians (including American Indians/Alaskan natives), (non-Hispanic) whites, (non-Hispanic) blacks, and Hispanics. h is the entropy of the ethnic distribution of public school students in the CBSA.