

Measuring School Segregation*

David M. Frankel

Oscar Volij

Iowa State University

Ben Gurion University

March 5, 2009

Abstract

We propose a set of axioms for the measurement of school-based segregation with any number of ethnic groups. These axioms are motivated by two criteria. The first is evenness: how much do ethnic groups' distributions across schools differ? The second is representativeness: how different are schools' ethnic distributions from one another? We prove that a unique ordering satisfies our axioms. It is represented by the Mutual Information index, which was originally proposed by Henri Theil. This index is decomposable in a more intuitive way than other segregation indices. As an application, we find that segregation between districts within cities accounts for 33% of urban school segregation in the U.S. 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, equal opportunity.

*Email addresses: dfrankel@econ.iastate.edu; ovolij@bgu.ac.il. We thank participants at the 2009 North American Winter Meeting of the Econometric Society and other conferences and workshops. Volij thanks the Spanish Ministerio de Educación y Ciencia (project SEJ2006-05455) for research support.

1 Introduction

Recent research has found that minority students have poorer academic outcomes in segregated schools (Boozer, Krueger, and Wolkin [4]; Hoxby [20]). School segregation can also explain a large portion of the wage gap between blacks and whites (Hanushek, Kain, and Rivkin [18]).

Despite the consensus that school segregation is important, there is little agreement about how to measure it: in a recent survey, twenty competing indices are considered (Massey and Denton [26]). In order to choose, one needs more information about how these indices behave. In the best case, one would like an axiomatization: a set of simple, intuitive properties that uniquely define a given index.

In this paper, we provide an axiomatization of school segregation for any number of ethnic groups. Formally, we define a segregation ordering as 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.

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 k th value of x . For 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 k th group. The entropy of x is a measure of the uncertainty in x .¹ 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*

¹The 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 [11] for this and other interpretations.

that school. The *mutual information between x and y* is thus defined as

$$M = \sum_{\substack{\text{Schools } n \\ \text{in district}}} \left(\frac{\text{Students in school } n}{\text{Students in district}} \right) \left(\begin{array}{c} \text{Entropy of district's ethnic distribution} \\ \text{minus} \\ \text{Entropy of school } n\text{'s ethnic distribution} \end{array} \right).$$

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.

As we assume only ordinal axioms, we obtain a unique segregation ranking rather than a unique index. There are many indices (each an increasing transformation of the other) that represent this ranking. We focus on the Mutual Information index because of its intuitive information-theoretic interpretation, discussed above. In addition, it has a very useful decomposition across geographic levels and ethnic groups that we illustrate using data from U.S. public schools.

The Mutual Information index was first proposed by Theil [39] and was applied by Fuchs [17] and Mora and Ruiz-Castillo [27, 28] to study gender segregation in the labor force.² It is related to the more widely used Entropy index H (Theil [40]; Theil and Finizza [41]), which equals the ratio of the Mutual Information index to the entropy of the districtwide ethnic distribution. 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 [24] definition of segregation as the tendency of ethnic groups to have different distributions across locational units such as schools or neighborhoods. Massey and Denton [26] call this the property of *evenness*. For instance, suppose 99% of a district's blacks are in school A while 1% are in school B, and whites are divided equally between the schools. By the evenness criterion, the district is segregated because the distribution of blacks across schools, (0.99, 0.01), differs from the distribution

²See also Herranz, Mora, and Ruiz-Castillo [19]. 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 [30, 29].

of whites, $(0.5, 0.5)$. As this judgment does not depend on how many blacks and whites there are overall in the district, James and Taeuber [24] argue that an index should be Scale Invariant: if the number of students of a given race is multiplied by the same positive constant in all schools, the index should not change.

Massey and Denton [26] also identify a second dimension of segregation: the extent to which minorities are *isolated* from, or lack *exposure* to, the majority group.³ In contrast to evenness, this notion does depend on the districtwide ethnic distribution. If the above district contains 100 blacks and 1,000 whites, then blacks are not very isolated in school A, which contains 99 blacks but 500 whites. If instead the district contains 1,000 blacks and 100 whites, school A contains 990 blacks but only 50 whites: blacks are more isolated. Accordingly, proponents of isolation typically do not regard Scale Invariance as a desirable property (e.g., Coleman, Hoffer, and Kilgore [9]).

It is not clear how to think of isolation in the case of three or more groups, since there is more than one group from which a student may be isolated. Hence, we replace isolation with the related concept of representativeness: to what extent do the ethnic distributions of individual schools differ from that of the district? The concepts are related, since racially isolated schools are, by definition, not representative of their districts. But unlike isolation, representativeness is not based on the exposure of one specific group to another.

We aim to craft a measure that reflects both evenness and representativeness. In addition, we need to impose a few other criteria in order to obtain a unique measure. The first is that the measure should treat ethnic groups *symmetrically*. Although symmetry is a standard property which is satisfied by most indices, it may not be suitable for 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 prefer an index that treats blacks differently.

³Massey and Denton [26] also discuss three other dimensions that rely on geographic information, and thus are not relevant to our study. They are concentration in a small area, centralization in the urban core, and clustering in a contiguous enclave.

We also require that the measure be *responsive* to changes within subdistricts: if the students in a subdistrict are reallocated so as to raise segregation in that subdistrict, with no changes to the rest of the district, then segregation in the district should not fall. This is our axiom of Type I Independence. Another requirement is that the index be *consistent* in its treatment of segregation within subdistricts and segregation between them. By this we mean that the relative importance the measure assigns to these two aspects of segregation should not change if students are reallocated within a subdistrict. This is our axiom of Type II Independence.

Finally, a measure should be *unsuspicious*: it should not impute segregation where there is no direct evidence for it. For instance, if the researcher only has data about whites and blacks, the measure should not impute a positive level of segregation between white subgroups (such as Hispanics vs. Anglos). We call this axiom the Group Division Property. While unsuspectingness will sometimes turn out to be naive, it would seem to be no worse than ascribing an arbitrary positive level of segregation between unobserved subgroups. In addition, it makes the segregation measure more predictable by ensuring that replacing a coarse ethnic schema with a fine one cannot lead to lower segregation.

Together with evenness and representativeness, unsuspectingness also helps motivate our Weak School Division Property (WSDP). This axiom states that if one starts with a one-school district and then splits the one school into two schools, segregation cannot fall. Intuitively, since all data are at the school level, an unsuspecting measure will not detect any segregation in a one-school district.⁴ Hence, splitting the school cannot lower the level of segregation. WSDP also states that if the two new schools have the same ethnic distribution, then the split has no effect on segregation in the district. This is motivated by representativeness and evenness: since the two schools are representative of their district, and since all ethnic groups have the same distribution across the schools, the new district is

⁴This discussion assumes that there are no data on tracking and other forms of within-school segregation. If such data were available, one could use our approach to study these phenomena by redefining the basic locational unit to be the classroom or ability group.

completely unsegregated by both criteria.

WSDP implies that a one-school district with 110 whites and ten blacks does not become less segregated if the ten blacks and an equal number of whites are relocated to a new school. Of course, a student in the new school might feel that her environment is now less segregated since it has equal numbers of blacks and whites. No model can capture all possible notions of segregation. By the representativeness criterion, segregation has increased: the original school was representative of its district but the two resulting schools are not.⁵

WSDP is related to two properties that are discussed by James and Taeuber [24] 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, WSDP follows from organizational equivalence and the transfer principle.⁶ But while WSDP applies directly with any number of groups, it is unclear what form the transfer principle should take with more than two groups.⁷

We also require Weak Scale Invariance: if the number of students in each ethnic group in every school is multiplied by the *same* positive constant, then segregation in the district does not change. Intuitively, the new district has the same distributions of races across

⁵With respect to evenness, ethnic groups are (trivially) distributed evenly across schools before the shift but not after.

⁶A rough intuition runs as follows. The first part of WSDP is just organizational equivalence itself. As for the second part, any division of a school into two new schools with differing ethnic distributions can be broken into two steps. First, create two schools with the desired sizes but the *same* ethnic distribution. By organizational equivalence, segregation is unchanged. In the second step, swap black students with white students until the desired ethnic distributions are attained. Since each swap moves students to schools in which their groups are overrepresented, segregation must rise by the transfer principle.

⁷For 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 [34].

schools and distributions of races within schools. By the criteria of evenness and representativeness, the two districts are equally segregated.

Unlike James and Taeuber [24], we do not focus solely on the dimension of evenness. We also seek a measure that reflects a student's exposure to other ethnic groups, as captured by the representativeness criterion. Accordingly, we do not impose the stronger axiom of Scale Invariance. Indeed, fixing the distribution of each ethnic group *across* schools, the Mutual Information index tends to be higher when there is more ethnic diversity in the district, since this leads to greater ethnic differences across schools (section 4). In this sense, it resembles the Normalized Exposure index of Bell [2] and James [25], with which it has a rank correlation of 0.859 (section 6). However, Normalized Exposure lacks the intuitive decomposition of the Mutual Information index and violates our principles of consistency, responsiveness, and unsuspectingness.⁸

Frankel and Volij [16] study a related axiomatization that does include Scale Invariance. They show that the Atkinson segregation indices uniquely satisfy the remaining axioms of this paper, with the exception of the Group Division Property and Type II Independence. Hence, these indices violate our principles of unsuspectingness and consistency. On the other hand, if one's only goal is to study whether the effect of origins (race) on destinations (school assignment) has changed over time, then one may prefer a Scale Invariant segregation measure: one that is unaffected by changes in the proportions of different races in the district.

Unlike other segregation indices, the Mutual Information index can be decomposed across geographic levels and ethnic groups in an intuitive way. We illustrate this property by decomposing 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 [35] and Clotfelter [8] find that within U.S. cities, segregation across districts exceeds segregation within districts. Our analysis confirms this finding. However,

⁸More precisely, it violates Type I Independence in the case of three or more groups, as well as Type II Independence and the Group Division Property (section 5.1).

we also show that segregation across districts accounts for only 33% of total segregation in the U.S. 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.

Several prior papers have analyzed the properties of segregation indices. However, only a few 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.

The first paper in this literature was Philipson [32], 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 [21, 22] studies the measurement of segregation in the case of two ethnic groups. Hutchens [21] characterizes the family of indices that satisfy a set of mostly cardinal properties, and Hutchens [22] strengthens one axiom to obtain a unique segregation index, which is based on the Atkinson inequality index [1]. Using cardinal axioms, Echenique and Fryer [13] 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. We present our notation in section 2, our axioms in section 3, and our main result in section 4. In section 5, we consider two convenient decomposability properties and indicate which of our axioms are satisfied by other multigroup segregation indices. We also discuss the Mutual Information index’s relation to a test of color-blind school assignment. Section 6 applies the Mutual Information index to public schools in the U.S. and section 7 concludes. Most results are proved in Appendix A.

2 Notation

We assume a continuum population. This is a reasonable approximation when ethnic groups are large. 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*: $\langle (T_g^n)_{g \in \mathbf{G}} \rangle_{n \in \mathbf{N}}$. 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 district X , $c(X)$ denotes the district that results from combining the schools in X into a single school. For any nonnegative scalar α , αX denotes the district in which the number of students in each group and school has been multiplied by α . If Y is another district, $X \uplus Y$ denotes the result of combining districts X and Y . For example, if $X = \langle (10, 20), (30, 10) \rangle$ and $Y = \langle (40, 50) \rangle$, then $c(X) = \langle (40, 30) \rangle$, $2X = \langle (20, 40), (60, 20) \rangle$, and $X \uplus Y = \langle (10, 20), (30, 10), (40, 50) \rangle$.

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 .

When referring to a particular district X , we will write $T(X)$ as the student population of the district. 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 $(p_g^n)_{g \in \mathbf{G}}$ 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 axioms. Let \mathcal{C} be the set of all districts. A *segregation ordering* \succsim is a complete and transitive binary relation on \mathcal{C} . We interpret $X \succsim Y$ to mean “district X is at least as segregated as district Y .” The relations \sim and \succ are derived from \succsim in the usual way.⁹ A related concept is the *segregation index*: a function $S : \mathcal{C} \rightarrow \mathfrak{R}$. The index

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

S represents the segregation ordering \succsim if, for any two districts $X, Y \in \mathcal{C}$,

$$X \succsim Y \iff S(X) \geq S(Y). \quad (1)$$

Every index S induces a segregation ordering \succsim 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 .

Our first axiom states that a district's ranking should depend only on the number of each group who attend each school: labels such as “black”, “Roosevelt School,” etc., do not matter.

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

Our second axiom states that the scale of a district does not matter.

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 , $X \sim \alpha X$.

This axiom is satisfied by all of the common school segregation indices (section 5.1). It implies, e.g., that the districts $\langle (1000, 0), (0, 1000) \rangle$ and $\langle (100, 0), (0, 100) \rangle$ are equally segregated. An alternative view is that the first district is more segregated because it is

less likely to be the outcome of random assignment of students to schools. This notion of segregation is discussed in section 5.2.

Our next axiom states that a one-school district cannot become *less* segregated if the school is split into two new schools. In addition, if the new schools have identical ethnic distributions, then segregation is unchanged.

Weak School Division Property (WSDP) Let $X \in \mathcal{C}$ be a district consisting of a single school. Let X' be the district that results from subdividing this school into two schools, n_1 and n_2 . Then, $X' \succcurlyeq X$. Further, if n_1 and n_2 have the same group distributions (i.e., $p_g^{n_1} = p_g^{n_2}$ for all $g \in \mathbf{G}$), then $X' \sim X$.

The following axiom states that reallocating the students in a subdistrict so as to raise segregation in that subdistrict, *ceteris paribus*, should not lower segregation in the district as a whole.

Type I Independence (IND1) Let $X, Y \in \mathcal{C}$ have equal populations and equal group distributions. Then for any $Z \in \mathcal{C}$, $X \uplus Z \succcurlyeq Y \uplus Z$ if and only if $X \succcurlyeq Y$.

Since X and Y have the same number of each ethnic group, Y can be seen as a reallocation of the students of X .¹⁰ Perhaps the most widely used segregation index, the Dissimilarity Index, violates this axiom.¹¹ For instance, suppose a district is composed of two areas: $X = \langle (50, 100), (50, 0) \rangle$ and $Z = \langle (100, 0) \rangle$. Suppose that the students in the first area are reallocated to yield $Y = \langle (100, 40), (0, 60) \rangle$. The Dissimilarity Index within this area *rises* from 0.5 to 0.6, but the index for the full district *falls* from 0.75 to 0.6.

Our next axiom states that a measure should treat segregation within and between subdistricts in a consistent way. It should not change the relative importance it assigns to segregation between subdistricts simply because segregation falls within a subdistrict.

¹⁰Since Y is not required to have the same number of schools as X , the reallocation might be accompanied by new school construction or conversion of some schools to other uses.

¹¹With two ethnic groups, the Dissimilarity Index is defined as the percentage of either group that must change schools in order for all schools to have the same ethnic distribution.

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

For instance, consider the districts $X \uplus Y$ and $X \uplus Z$ where $X = Y = \langle (50, 0), (0, 100) \rangle$ and $Z = \langle (100, 0), (0, 50) \rangle$. Since X and Y are identical, there is no segregation between them. However, there is some between X and Z since their ethnic distributions differ. Most indices ignore this distinction: since there is no ethnic mixing *within* X , Y , or Z , they regard $X \uplus Y$ and $X \uplus Z$ as maximally, and thus equally, segregated. When within-subdistrict segregation is reduced, these indices begin to reflect between-subdistrict segregation: they regard $c(X) \uplus Y$ as strictly less segregated than $c(X) \uplus Z$.¹² Type II Independence rules out this behavior: a segregation measure's sensitivity to between-subdistrict segregation should not depend on the level of segregation within subdistricts.

The next axiom is based on the principle of unsuspectingness: a segregation measure should not assume some positive level of within-subgroup segregation in the absence of any data on how subgroups are distributed. Hence, if data on subgroups becomes available, and the data show that the subgroups are identically distributed across schools, then the segregation measure 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: $\frac{T_{g_1}^n}{T_{g_1}} = \frac{T_{g_2}^n}{T_{g_2}}$ for all $n \in \mathbf{N}$.¹³ Then $X' \sim X$.

In section 5.1, we show that if a segregation index can be decomposed in a particular intuitive way across ethnic groups, then it must satisfy GDP (Proposition 1). In addition, if

¹²This behavior characterizes all indices surveyed in section 5.1, Table 1, 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.

¹³Note that X' has the same set N of schools as X and for each school $n \in N$, $T_g^n = T_{g_1}^n + T_{g_2}^n$.

a segregation index has a particular intuitive decomposition across geographic levels, then it must satisfy IND1 and IND2 (Proposition 1).

The next axiom is a technical continuity property. We use it 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 $X \succ Y$.

4 Main Result

We are now ready to state our main result.

Theorem 1 *The ordering that is represented by the Mutual Information index is the only ordering that satisfies Symmetry, Weak Scale Invariance, the Weak School Division Property, Type I and II Independence, the Group Division Property, Continuity, and Nontriviality.*

The Entropy index of Theil [40] and Theil and Finizza [41] is obtained by dividing the Mutual Information index by 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. This implies that 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.

For instance, consider 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 lack the intuitive decompositions of the Mutual Information index (section 5.1). Second, they do not capture changes in interracial contact well.¹⁴ To illustrate the second point, compare the effect of merging two schools in X , yielding the 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. Reflecting this difference, the Mutual Information index falls by 0.08 in district X versus 1.0 in Y . In contrast, the Entropy index falls by the same amount (1.0) in both cases.

5 Other Properties of the Index

In this section we discuss two useful ways in which the Mutual Information index is decomposable. We also discuss its relation to a statistical test of color-blind assignment of students to schools.

¹⁴This argument is due to Clotfelter [7].

5.1 Decomposability

Our main result characterizes the Mutual Information ordering exclusively by means of ordinal axioms. For empirical applications, however, a segregation ordering is not enough; one needs to choose a cardinal representation out of the many that represent the ordering. One approach is to require some convenient cardinal properties. In this section we present two properties that, if satisfied, allow us to decompose total segregation into several meaningful components. We further show that the Mutual Information index satisfies these properties. In the next section we use these decompositions to analyze school segregation in the US.

The first decomposability property states that, for any partition of a district's schools into subdistricts, total segregation in the district is the sum of between-subdistrict and within-subdistrict segregation:

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

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

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

Mora and Ruiz-Castillo [30] 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 [5], 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 [15] obtains a further characterization of the Theil inequality index by replacing the dif-

ferentiability requirement by a more appealing transfer principle. Hutchens [22] 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 Strong Group Decomposability if, for any partition of the ethnic groups of a district X into K supergroups,

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

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.*

Note that decomposability is not one of our axioms. Rather, our axioms, taken together, imply it. More precisely, they imply that the segregation ordering must be represented by an index that is decomposable in the above sense.

| | Mutual Information | Entropy Index | Weighted Dissimilarity | Gini |
|------------|---|---|---|---|
| Symbol | M | H | D^W | G |
| Defined by | [39] | [40, 41] | [23, 31, 37] | [23, 33] |
| 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 p_g^n - P_g $ | $\frac{1}{2I} \sum_g \sum_m \sum_n P^m P^n p_g^m - p_g^n $ |
| SYM | ✓ | ✓ | ✓ | ✓ |
| WSI | ✓ | ✓ | ✓ | ✓ |
| WSDP | ✓ | ✓ | ✓ | ✓ |
| IND1 | ✓ | ✓ | × | × |
| IND2 | ✓ | × | × | × |
| GDP | ✓ | × | × | × |
| C | ✓ | ✓ | ✓ | ✓ |
| N | ✓ | ✓ | ✓ | ✓ |
| SI | × | × | 2 | 2 |
| SSD | ✓ | × | × | × |
| SGD | ✓ | × | × | × |
| | Normalized Exposure | Clotfelter | Card-Rothstein | Symmetric Atkinson |
| Symbol | P | C | CR | A |
| Defined by | [2, 25] | [8] | [6] | [24] |
| 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 | ✓ | ✓ | ✓ | ✓ |
| WSDP | ✓ | × | × | ✓ |
| 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 Interaction Index, $I = \sum_{g \in G} P_g (1 - P_g)$. The properties are Symmetry (SYM), Weak Scale Invariance (WSI), the Weak School Division Property (WSDP), 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).

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. This property is necessary for Strong School Decomposability, in the following sense:

Proposition 3 *Let $S : \mathcal{C} \rightarrow \mathfrak{R}_+$ 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. Section 6 contains an empirical illustration of the uses of SSD and SGD.

5.2 Color Blind Assignment

We focus on two questions in this paper: are ethnic groups distributed differently across schools? And do schools have different ethnic distributions from one another? These questions relate to the realized, *ex post* allocation of students to schools. This allocation is important since the composition of a student's peer group can affect her outcomes in school and, later, in the labor market.

However, one can also ask a different question: does a district assign students to schools in a random, color-blind manner? If this is the question of interest, then some deviation from evenness and representativeness may be tolerated if it is likely the result of randomness in the assignment process.¹⁵ The Mutual Information Index also plays a role in such a test: under the hypothesis of color-blind assignment, the Mutual Information index, appropriately scaled, has a χ^2 distribution. 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*

¹⁵See, e.g., Berlowitz and Sapp [3] and Cortese, Falk, and Cohen [10].

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 .¹⁶

The main argument against using such a test to measure segregation is that random fluctuations, while not the result of discrimination, do matter for students' outcomes. As Taeuber and Taeuber [38] state,

Induction into the army, an unequal allocation of resources or a segregated geographic distribution may be the outcome of a random process, but the future consequences often depend on the resulting situation rather than on the randomness of the generating process.

In addition, a test of color-blind assignment can easily be manipulated: if a district subdivides a school into two new schools with identical ethnic distributions, the increase in degrees of freedom makes it less likely that the test will reject the null hypothesis. In contrast, such a split has no effect on most *ex post* segregation measures, including the Mutual Information index.

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. Data are for the 2005-6 school year and come from the Common Core of Data (CCD) [36]. 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)

¹⁶The proof of this claim is a straightforward generalization of the proof for the case of two ethnic groups by Mora and Ruiz-Castillo [29, pp. 32-33]. It is available from the authors on request.

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”.

| 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.

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.¹⁷ Since supergroup

¹⁷The CCD actually has five ethnic groups; the smallest, American Indian/Alaskan Native, is not repre-

schemas must be nested in order to apply Strong Group Decomposability, we remove one ethnic group at a time.¹⁸ 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:

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

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

sented 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.

¹⁸We 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 [42]).

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 differentiation 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 [35] and Clotfelter [8] 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 2.¹⁹ However, the properties 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 [8] 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,

¹⁹Rivkin [35] and Clotfelter [8] also present detailed analyses of how segregation at these two levels varies across cities. We do not pursue such an analysis here.

H .²⁰

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 ($M = 0.243$).

| URBAN SCHOOL SEGREGATION BETWEEN PAIRS OF ETHNIC GROUPS | | | | | | |
|---|-----------|-------|----------|-----------|----------|-----------|
| Geographic Level | 1 | 2 | 3 | 4 | 5 | 6 |
| | Asian vs. | | | White vs. | | Black vs. |
| | 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 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)

²⁰These observations are due to Reardon and Firebaugh [34, pp. 53-4].

| 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) (C190) | 0.467 | 0.418 | 0.341 | 0.367 | 0.462 | 1 | 0.629 | 0.46 | 0.243 |
| Clotfelter (50% threshold) (C150) | 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.

according to the Mutual Information index. The other indices in Table 1 are also shown. 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.²¹ It equals the *mutual information* of a student's race and her school when these are viewed as random variables (Cover and Thomas [11]). 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 facilitate the index's application to empirical work on the causes and effects of school segregation.

The dual interpretation is also directly related to the criteria of evenness and representativeness. When schools are not representative of their 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, the concepts are directly related to the two types of information that our index measures.

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 illustrate the use of these decompositions by studying the sources of segregation across urban public schools in the U.S. in 2005-6. As Rivkin [35] and Clotfelter [8] 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. A different axiomatization that assumes this property is Frankel and Volij [16]. That paper drops the Group Division Property, as the two axioms together have undesirable implications.²² As a result, the orderings characterized in that paper do

²¹An index represents the ordering if and only if it is a strictly increasing transformation of the Mutual Information index.

²²In particular, by repeatedly scaling and then subdividing, one can make an arbitrary number of copies of

not permit comparison between districts with different numbers of ethnic groups. They are also not decomposable in the sense of section 5.1.

A Proofs

Proof of Theorem 1. We first define a slight strengthening of WSDP:

School Division Property (SDP) Let $X \in \mathcal{C}$ be any district and let n be a school in X .

Let X' be the district that results from X if school n is subdivided into two schools, n_1 and n_2 . Then, $X' \succcurlyeq X$. Further, if n_1 and n_2 have the same group distributions (i.e., $p_g^{n_1} = p_g^{n_2}$ for all $g \in \mathbf{G}$), then $X' \sim X$.

SDP follows from WSDP and IND1:

Lemma 1 *Suppose the segregation ordering \succcurlyeq satisfies Type I Independence and the Weak School Division Property. Then \succcurlyeq also satisfies the School Division Property.*

Proof. Let Y denote the district X less the school n : $X = Y \uplus \langle n \rangle$ and $X' = Y \uplus \langle n_1, n_2 \rangle$. By WSDP, $\langle n_1, n_2 \rangle \succcurlyeq \langle n \rangle$. By IND1, $Y \uplus \langle n_1, n_2 \rangle \succcurlyeq Y \uplus \langle n \rangle$. If n_1 and n_2 have the same population distribution then the symbol \succcurlyeq can be replaced by \sim . Q.E.D.

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. As for WSDP, let $X \in \mathcal{C}$ contain the single school n and 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,

$$M(X') - M(X) = h((p_g^n)_{g \in \mathbf{G}(X)}) - \frac{P^{n_1}}{P^n} h((p_g^{n_1})_{g \in \mathbf{G}(X)}) - \frac{P^{n_2}}{P^n} h((p_g^{n_2})_{g \in \mathbf{G}(X)}).$$

any ethnic group without changing a district's segregation ranking

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 WSDP.

We now show that the Mutual Information ordering is the only segregation ordering that satisfies all the axioms. Let \succsim be a segregation ordering that satisfies them; by Lemma 1, it also satisfies SDP. For any district X , let the schools be numbered $n = 1, \dots, N$ and the groups $g = 1, \dots, G$. 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:

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

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

- Lemma 2** 1. *All districts in which every school is representative have the same degree of segregation under \succsim .*
2. *Any district in which every school is representative is weakly less segregated under \succsim 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 $i = 1, \dots, N$, 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 $i = 1, \dots, N-1$, $Y_i \sim Y_{i+1}$. 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 $Y = Y_1 \succcurlyeq Y_2 \succcurlyeq \dots \succcurlyeq Y_N \sim \underline{X}^1$. 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 $\sigma^j(Z) = \sigma(\sigma^{j-1}(Z))$.²³ Then $\frac{1}{G} \uplus_{j=1}^G \sigma^j(Z) \succcurlyeq Z$.

Proof. Consider the following statement:

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

For $n = 0$, (5) simply states $\uplus_{j=1}^G c(Z) \preccurlyeq \left(\uplus_{j=1}^G \sigma^j(c(Z)) \right)$, 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,

$$\begin{aligned} \left(\uplus_{j=1}^k Z \right) \uplus \left(\uplus_{j=k+2}^G c(Z) \right) \uplus c(Z) &\preccurlyeq \left(\uplus_{j=1}^k \sigma^j(Z) \right) \uplus \left(\uplus_{j=k+1}^{G-1} \sigma^j(c(Z)) \right) \uplus c(Z) \\ \implies \left(\uplus_{j=1}^{k+1} Z \right) \uplus \left(\uplus_{j=k+2}^G c(Z) \right) &\preccurlyeq \left(\uplus_{j=1}^k \sigma^j(Z) \right) \uplus \left(\uplus_{j=k+1}^{G-1} \sigma^j(c(Z)) \right) \uplus Z \\ &\sim \sigma \left(\left(\uplus_{j=1}^k \sigma^j(Z) \right) \uplus \left(\uplus_{j=k+1}^{G-1} \sigma^j(c(Z)) \right) \uplus Z \right) \\ &\sim \left(\uplus_{j=1}^{k+1} \sigma^j(Z) \right) \uplus \left(\uplus_{j=k+2}^G \sigma^j(c(Z)) \right) \end{aligned}$$

where the first line follows by hypothesis; the second by IND2; the third by SYM; and the fourth by definition of σ . This proves (5) for $n = k + 1$ and thus for all n . Hence, $\uplus_{j=1}^G Z \preccurlyeq \uplus_{j=1}^G \sigma^j(Z)$, so $Z \preccurlyeq \frac{1}{G} \uplus_{j=1}^G \sigma^j(Z)$ by SDP and WSI. Q.E.D.

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

²³Note that $\sigma^G(Z) = Z$.

Proof. By WSI, w.l.o.g. we can assume that $T(X) = 1$. X can be converted into $\overline{X}(P)$ by dividing each school into G racially isolated schools and then combining the schools appropriately. By SDP, this results in a weakly more segregated district. Finally, by SDP and Lemma 3, $\overline{X}^G \sim \frac{1}{G} \uplus_{j=1}^G \sigma^j(\overline{X}(P)) \succcurlyeq \overline{X}(P)$. Q.E.D.

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 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, $(\alpha - \beta)X \succ (\alpha - \beta)X'$. 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 nonempty by assumption and 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. The union of the sets is the whole unit interval as \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 $\widehat{P} = (\widehat{P}_1, \dots, \widehat{P}_G)$. For any $G' \geq 1$ and any distribution $P = (P_1, \dots, P_{G'})$ 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 $P_1 T_g^n, \dots, P_{G'} T_g^n$. The resulting district $\phi^P(X)$ has GG' groups with distribution $\left((\hat{P}_g P_{g'})_{g=1}^{G'} \right)_{g'=1}^{G'}$.

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

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

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

Define the index $S : \mathcal{C} \rightarrow \mathfrak{R}$ by

$$S(X) = \hat{\alpha} / \hat{\beta}. \quad (7)$$

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

$$X \sim \phi^{\hat{P}} \left(\tilde{\alpha} \overline{X}(\tilde{P}) \uplus (1 - \tilde{\alpha}) \underline{X}(\tilde{P}) \right) \sim \tilde{\alpha} \phi^{\hat{P}} \left(\overline{X}(\tilde{P}) \right) \uplus (1 - \tilde{\alpha}) \phi^{\hat{P}} \left(\underline{X}(\tilde{P}) \right) \quad (8)$$

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

$$X \sim \hat{\alpha} \phi^{\tilde{P}} \left(\overline{X}(\hat{P}) \right) \uplus (1 - \hat{\alpha}) \phi^{\tilde{P}} \left(\underline{X}(\hat{P}) \right). \quad (9)$$

The districts $\phi^{\tilde{P}} \left(\underline{X}(\hat{P}) \right)$ and $\phi^{\tilde{P}} \left(\underline{X}(\tilde{P}) \right)$, as well as $\phi^{\tilde{P}} \left(\overline{X}(\hat{P}) \right)$ and $\phi^{\tilde{P}} \left(\overline{X}(\tilde{P}) \right)$, have the same number of groups ($G * G'$) and (up to a permutation) the same ethnic distribution.

Further, by SYM, $\phi^{\hat{P}}(\underline{X}(\hat{P})) \sim \phi^{\hat{P}}(\underline{X}(\tilde{P}))$. Assume w.l.o.g. that $\phi^{\hat{P}}(\overline{X}(\tilde{P})) \succcurlyeq \phi^{\hat{P}}(\overline{X}(\hat{P}))$ and let γ uniquely satisfy $\phi^{\hat{P}}(\overline{X}(\hat{P})) \sim \gamma \phi^{\hat{P}}(\overline{X}(\tilde{P})) \uplus (1-\gamma) \phi^{\hat{P}}(\underline{X}(\tilde{P}))$. Applying WSI, IND1 (twice) and SDP, it follows from (9) that

$$X \sim \hat{\alpha} \gamma \phi^{\hat{P}}(\overline{X}(\tilde{P})) \uplus (1 - \gamma \hat{\alpha}) \phi^{\hat{P}}(\underline{X}(\tilde{P})). \quad (10)$$

By (10) and (8), $\tilde{\alpha} = \hat{\alpha} \gamma$. Exactly the same reasoning leads to $\tilde{\beta} = \hat{\beta} \gamma$. Consequently $\hat{\alpha}/\hat{\beta} = \tilde{\alpha}/\tilde{\beta}$: 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 $X \sim \alpha_X \overline{X}^G \uplus (1 - \alpha_X) \underline{X}^G$, $Y \sim \alpha_Y \overline{X}^G \uplus (1 - \alpha_Y) \underline{X}^G$, and $\overline{X}^2 \sim \beta \overline{X}^G \uplus (1 - \beta) \underline{X}^G$. Then $X \succcurlyeq Y$ iff $\alpha_X \geq \alpha_Y$ by Lemma 6, which holds iff $S(X) \geq S(Y)$ since $\beta > 0$ and by definition of S . Q.E.D.

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

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

Proof. For the first claim, first subdivide each ethnic group in \overline{X}^G into G groups in proportions given by P . Now put each resulting group in a separate school. The group distribution of the resulting district, $(P_1 \frac{1}{G}, \dots, P_G \frac{1}{G}, \dots, P_1 \frac{1}{G}, \dots, P_G \frac{1}{G})$, is just a permutation of \hat{P} . Hence, by GDP and SDP, $\overline{X}(\hat{P}) \succcurlyeq \overline{X}^G$. 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 $X \uplus Y$ 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 groups of equal size and identically distributed across schools. Let \hat{P} be the group distribution of $\phi^G(X)$. By Lemma 9, $\bar{X}(\hat{P}) \succcurlyeq \bar{X}^G$. Define $\hat{\alpha}_X$ by $X \sim \hat{\alpha}_X \bar{X}(\hat{P}) \uplus (1 - \hat{\alpha}_X) \underline{X}(\hat{P})$ and γ by $c(X) \uplus Y \sim \gamma \bar{X}(\hat{P}) \uplus (1 - \gamma) \underline{X}(\hat{P})$. Define $Z_1 = \phi^G(X)$, $Z_3 = c(\phi^G(X)) = \phi^G(c(X))$, $Z_4 = T(X) \underline{X}(\hat{P})$, $Z_5 = \phi^G(Y)$,

$$Z_2 = T(X) \left(\hat{\alpha}_X \bar{X}(\hat{P}) \uplus (1 - \hat{\alpha}_X) \underline{X}(\hat{P}) \right), \text{ and}$$

$$Z_6 = T(X \uplus Y) \left(\gamma \bar{X}(\hat{P}) \uplus \left(1 - \frac{T(X)}{T(X \uplus Y)} - \gamma \right) \underline{X}(\hat{P}) \right).$$

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

$$\gamma \bar{X}(\hat{P}) \uplus (1 - \gamma) \underline{X}(\hat{P}) \preccurlyeq \frac{T(Y)}{T(X \uplus Y)} \bar{X}(\hat{P}) \uplus \frac{T(X)}{T(X \uplus Y)} \underline{X}(\hat{P}). \quad (11)$$

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

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

Let $\widetilde{c(X)} = \frac{1}{G} \uplus_{j=1}^G \sigma^j(c(X))$ and $\tilde{Y} = \frac{1}{G} \uplus_{j=1}^G \sigma^j(Y)$. 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 $T(Y) \bar{X}^G$ and both of these districts have the same population,

$T(Y)$, and the same group distribution. Therefore,

$$\begin{aligned}
c(X) \uplus Y &\preceq T(X)\underline{X}^G \uplus T(Y)\overline{X}^G && \text{by IND1 (twice)} \\
&\sim T(X)\phi^G(\underline{X}^G) \uplus T(Y)\phi^G(\overline{X}^G) && \text{by GDP and definition of } \phi^G \\
&\preceq \frac{T(X)}{T(X \uplus Y)}\underline{X}(\hat{P}) \uplus \frac{T(Y)}{T(X \uplus Y)}\overline{X}(\hat{P}) && \text{by SDP and WSI.}
\end{aligned}$$

But $c(X) \uplus Y \sim \gamma\overline{X}(\hat{P}) \uplus (1 - \gamma)\underline{X}(\hat{P})$ 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, $Z_1 \sim Z_2$. Since they are the same district, $Z_3 \sim Z_4$. The populations of Z_5 and Z_6 are both $T(Y)$. Moreover, $Z_4 \uplus Z_6 \sim T(X \uplus Y) \left(\gamma\overline{X}(\hat{P}) \uplus (1 - \gamma)\underline{X}(\hat{P}) \right)$ by SDP, which is as segregated as $\phi^G(c(X)) \uplus \phi^G(Y) = Z_3 \uplus Z_5$ by WSI and GDP. So by Lemma 10,

$$Z_1 \uplus Z_5 \sim Z_2 \uplus Z_6. \quad (12)$$

Now, $X \uplus Y \sim \phi^G(X \uplus Y) = Z_1 \uplus Z_5$ by GDP, but by (12), $Z_1 \uplus Z_5 \sim Z_2 \uplus Z_6$, which is as segregated as $\left(\gamma + \frac{T(X)}{T(X \uplus Y)}\hat{\alpha}_X \right) \overline{X}(\hat{P}) \uplus \left(1 - \gamma - \frac{T(X)}{T(X \uplus Y)}\hat{\alpha}_X \right) \underline{X}(\hat{P})$ by SDP and WSI. By definition of γ and $\hat{\alpha}_X$, $c(X) \uplus Y \sim \gamma\overline{X}(\hat{P}) \uplus (1 - \gamma)\underline{X}(\hat{P})$ and $X \sim \hat{\alpha}_X\overline{X}(\hat{P}) \uplus (1 - \hat{\alpha}_X)\underline{X}(\hat{P})$. By Lemma 7, there is a unique β such that $\overline{X}^2 \sim \beta\overline{X}(\hat{P}) \uplus (1 - \beta)\underline{X}(\hat{P})$. By definition of S , $S(X \uplus Y) = \frac{1}{\beta} \left(\gamma + \frac{T}{T+T(Y)}\hat{\alpha}_X \right) = S(c(X) \uplus Y) + \frac{T}{T+T(Y)}S(X)$, as claimed. Q.E.D.

For any discrete probability distribution $P = (P_1, \dots, P_G)$, 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 satisfies²⁴

²⁴The statement of this result appears as an exercise in Cover and Thomas [11]. For the original proof, see Faddeev [14].

1. $h(1/2, 1/2) = 1$,
2. $h(p, 1 - p)$ is continuous in p , and
3. $h(p_1, \dots, p_n) = h(p_1 + p_2, p_3, \dots, p_n) + (p_1 + p_2) h\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$.

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. As for property 2, let $Z^p = \overline{X}(p, 1 - p)$. 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 , $\{q : Z^q \succcurlyeq Z^p\}$ and $\{q : Z^q \preccurlyeq Z^p\}$ are closed by Continuity. By Lemma 6, $\{q : \alpha_q \geq \alpha_p\}$ 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 $\lim_{k \rightarrow \infty} \alpha_{p_k}$ 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 $\lim_{k \rightarrow \infty} \alpha_{p_k} = c > \frac{c + \alpha_p}{2}$, 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 $\{q : \alpha_q \geq \frac{c + \alpha_p}{2}\}$. But $\lim_{k \rightarrow \infty} p_k = p$ does not lie in this set, which contradicts the fact that this set is closed. Q.E.D.

To see 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. Clearly, $X_N \sim \overline{X}(P)$. By Lemma 11, $S(X_n) = S(X_{n-1}) + P^n S(\overline{X}(p^n))$ for $n = 1, \dots, N$. Thus,

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

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

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 $T(X)/T(X \uplus Z) = T(Y)/T(Y \uplus Z)$.

$Z) = p$. Then, applying SSD, $S(X \uplus Z) \geq S(Y \uplus Z)$ if and only if

$$\begin{aligned} S(c(X) \uplus c(Z)) + pS(X) + (1-p)S(Z) &\geq S(c(Y) \uplus c(Z)) + pS(Y) + (1-p)S(Z) \\ &\Leftrightarrow S(X) \geq S(Y) \end{aligned}$$

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

$$\begin{aligned} S(W \uplus c(Y)) &\geq S(X \uplus c(Y)) \\ \Leftrightarrow S(c(W) \uplus c(Y)) + pS(W) + (1-p)S(c(Y)) &\geq S(c(X) \uplus c(Y)) + pS(X) + (1-p)S(c(Y)) \\ \Leftrightarrow S(c(W) \uplus c(Y)) + pS(W) + (1-p)S(Y) &\geq S(c(X) \uplus c(Y)) + pS(X) + (1-p)S(Y) \\ \Leftrightarrow S(W \uplus Y) &\geq S(X \uplus Y) \end{aligned}$$

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

Proof of Proposition 2: Let $X = X^1 \uplus \dots \uplus X^K$ be a district composed of K subdistricts. 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

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

This shows that M satisfies SSD. That M satisfies SGD as well now follows from the symmetry of mutual information (Cover and Thomas [11, 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 \bar{X} . Define another district, \bar{X}' , that is a copy of \bar{X} in which each ethnic group has been replaced by a new ethnic group not in \bar{X} . (For instance, if \bar{X} has K groups that go to K separate schools, then let \bar{X}' consist of a *different* K groups that go to K separate schools.) Then by SSD, $S(\bar{X} \uplus \bar{X}') = S(c(\bar{X}) \uplus c(\bar{X}')) + \frac{1}{2}S(\bar{X}) + \frac{1}{2}S(\bar{X}')$. The sum of the second and third terms equals $S(\bar{X})$ by symmetry. But the first term is strictly positive by property 3. This contradicts the hypothesis that S attains its maximum at \bar{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] Berlowitz, Marvin J., and Martin L. Sapp. 1987. "A Critique of the Taeuber Index as a Measure of School Desegregation: A Case Study." *Journal of Negro Education* 56:475-484.
- [4] 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.
- [5] Bourguignon, F. 1979. "Decomposable Income Inequality Measures." *Econometrica* 47, 901-920.
- [6] Card, David, and Jesse Rothstein. 2006. "Racial Segregation and the Black-White Test Score Gap." NBER Working Paper No. 12078.

- [7] Clotfelter, Charles T. 1979. "Alternative Measures of School Desegregation: A Methodological Note." *Land Economics* 54: 373-380.
- [8] Clotfelter, Charles T. 1999. "Public School Segregation in Metropolitan Areas." *Land Economics* 75:487-504.
- [9] Coleman, James, 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] 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.
- [11] Cover, Thomas M. and Joy A. Thomas. (1991) *Elements of Information Theory*. United States: Wiley.
- [12] Cutler, David, and Edward Glaeser. 1997. "Are Ghettos Good or Bad?" *Quarterly Journal of Economics* 112: 827– 872.
- [13] Echenique, Federico, and Roland G. Fryer, Jr. 2007. "A Measure of Segregation Based on Social Interactions." *Quarterly Journal of Economics* 122: 441-485.
- [14] Faddeev, D.K. 1956. "On the Notion of Entropy of a Finite Probability Schema [in Russian]." *Usp. Mat. Nauk* 11:227-231.
- [15] Foster, James E. 1983. "An Axiomatic Characterization of the Theil Measure of Income Inequality." *Journal of Economic Theory* 65, 105-121.
- [16] Frankel, David M., and Oscar Volij. 2008. "An Axiomatization of the Multigroup Atkinson Segregation Indices." Mimeo.
- [17] Fuchs, Victor. 1975. "A Note on Sex Segregation in Professional Occupations." *Explorations in Economic Research* 2:105-111.

- [18] 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.
- [19] 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.
- [20] Hoxby, Caroline M. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." NBER Working Paper 7867.
- [21] Hutchens, Robert. 2001. "Numerical Measures of Segregation: Desirable Properties and Their Implications." *Mathematical Social Sciences* 42:13–29.
- [22] Hutchens, Robert. 2003. "One Measure of Segregation." *International Economic Review* 45:555–578.
- [23] Jahn, Julius, and Schmid, Calvin F., and Schrag, Clarence. 1947. "The Measurement of Ecological Segregation." *American Sociological Review* 12:293–303.
- [24] James, David R., and Taueber, Karl E. 1985. "Measures of Segregation." In *Sociological Methodology* Jossey-Bass. Nancy Tuma Ed.
- [25] James, Franklin J. 1986. "A New Generalized 'Exposure-based' Segregation Index." *Sociological Methods and Research* 14:301-316.
- [26] Massey, Douglas S. and Nancy A. Denton. 1988. "The Dimensions of Racial Segregation." *Social Forces* 67:281–315.
- [27] 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.

- [28] Mora, Ricardo, and Javier Ruiz-Castillo. 2004. "Gender Segregation by Occupations in the Public and the Private Sector: The Case of Spain." *Investigaciones Económicas* 28:399-428.
- [29] Mora, Ricardo, and Javier Ruiz-Castillo. 2005. "An Evaluation of an Entropy Based Index of Segregation." Mimeo, Universidad Carlos III de Madrid.
- [30] 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.
- [31] Morgan, Barrie S. 1975. "The Segregation of Socioeconomic Groups in Urban Areas: A Comparative Analysis." *Urban Studies* 12:47-60.
- [32] Philipson, Tomas. 1993. "Social Welfare and Measurement of Segregation." *Journal of Economic Theory* 60:322-334.
- [33] 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.
- [34] Reardon, Sean F., and Glenn Firebaugh. 2002. "Measures of Multigroup Segregation." *Sociological Methodology* 32:33-67.
- [35] Rivkin, Steven G. 1994. "Residential Segregation and School Integration." *Sociology of Education* 67:279-92.
- [36] 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.
- [37] Sakoda, J. 1981. "A Generalized Index of Dissimilarity." *Demography* 18:245-50.

- [38] Taeuber, Karl E., and Alma F. Taeuber. 1976. "A Practitioner's Perspective on the Index of Dissimilarity." *American Sociological Review* 41:884-889
- [39] Theil, Henri. 1971. "Principles of Econometrics." Wiley & Sons. New York.
- [40] Theil, Henri. 1972. "Statistical Decomposition Analysis." North Holland.
- [41] Theil, Henri., and A. J. Finizza 1971. "A Note on the Measurement of Racial Integration in Schools." *Journal of Mathematical Sociology* 1:187-1193.
- [42] U.S. Census Bureau. 2007. *Current Population Survey 2007, Annual Social and Economic Supplement, Detailed Poverty Tables*, Table POV01.

| 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 | Charlotte | 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. C_{50} and C_{90} 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.