Measuring School Segregation[*]

David M. Frankel

Oscar Volij

Iowa State University

Ben Gurion University

August 4, 2010

Abstract

Using only ordinal axioms, we characterize several multigroup school segregation indices: the Atkinson Indices for the class of school districts with a given fixed number of ethnic groups and the Mutual Information Index for the class of all districts. Properties of other school segregation indices are also discussed. In an empirical application, we document a weakening of the effect of ethnicity on school assignment from 1987/8 to 2007/8. We also show that segregation between districts within cities currently accounts for 33% of total segregation. Segregation between states, driven mainly by the distinct residental patterns of Hispanics, contributes another 32%.

1 Introduction

Recent research supports the view that school segregation creates unequal opportunities: separate schools are not equal. By and large, students in schools with a higher proportion of minority students have lower educational attainment and subsequent wages (Boozer, Krueger, and Wolkin [6, pp. 303-6]; Hanushek, Kain, and Rivkin [24]; Hoxby [25]). Consistent with this, African Americans tend to have better educational outcomes in less segregated school districts (Card and Rothstein [8]; Guryan [23]).[1]

Given these concerns, it is important to answer a more fundamental question: what is segregation, and how should it be measured? The literature on segregation measurement has generated over 20 different indices (see Massey and Denton [33] and Flückiger and Silber [19]). While some papers have analyzed the properties of these indices, very few of them have provided a full axiomatization, and none of these have used purely ordinal axioms. Further, most of the existing axiomatizations treat only the two-group case. In this paper we provide two axiomatizations of multigroup school segregation indices using purely ordinal axioms.

Axiomatizations are important because they characterize an index in terms of basic properties and thus facilitate the comparison of different measures. Ordinal axioms are more appealing than cardinal ones because they refer to bilateral comparisons and not to their specific functional representations. Multigroup segregation orderings are important because they allow us to study units (cities, school districts, etc.) with more than two ethnic groups and to compare units with different numbers of groups.

Formally, we define a segregation ordering as a ranking of school districts from most to least segregated. We consider six substantive axioms. Scale Invariance states that the scale of a district does not matter: if the number of students in each ethnic group in each school is multiplied by the same positive factor, segregation is unaffected. A segregation ordering satisfies Symmetry if it is invariant to a renaming of the ethnic groups (e.g., if blacks are

---

[1]Evidence on the effects of residential segregation is more mixed; see, e.g., Cutler, Glaeser, and Vigdor [16].

renamed to whites and vice-versa). Independence states that if the students in a subset of schools are reallocated within that subset, then segregation in the whole district rises if and only if it rises within the subset. The School Division Property states that splitting a school into two schools (a) cannot lower segregation in the district, and (b) leaves segregation unchanged if the ethnic distributions of the two resulting schools are identical. Composition Invariance states that the segregation of a district does not change if the number of students in a given ethnic group is multiplied by the same positive constant throughout the district (e.g., if the number of blacks in every school rises by ten percent). Finally, the Group Division Property states that the segregation of a district does not change if an existing ethnic group is subdivided into two groups that have the same distribution across schools (e.g., if whites are divided by gender and white boys have the same distribution across schools as white girls).

We first show that a nontrivial segregation ordering satisfies Scale Invariance, Independence, the School Division Property, Composition Invariance, and a technical continuity property if and only it is represented by an Atkinson index (Atkinson [2]; James and Taeuber [30]). This index equals one minus the sum, over all the schools, of some weighted geometric average of the percentages of each group who attend the school. For instance, suppose 70% of Asians, 40% of blacks, and 10% of whites attend school A while the remainder attend school B. The index for this district is $1 - (.7)^a (.4)^b (.1)^w - (.3)^a (.6)^b (.9)^w$ where a, b, and w are arbitrary nonnegative weights, chosen by the researcher, that sum to one.

Most multigroup segregation indices weight a group according to its relative size within a district (Reardon and Firebaugh [42]). As this yields group weights that vary across districts, these indices violate Composition Invariance. An alternative is to use an Atkinson index with group weights that vary across groups but not across districts. For instance, a group's weight might equal the proportion of students in the universe of districts who belong to the group in some reference year. Since the weight on each group is constant across districts, such an index would satisfy Composition Invariance, and it would assign greater importance to larger groups. We illustrate this point empirically in section 6.

3

The Atkinson indices have some limitations. They allow comparisons only between districts with a fixed set of nonempty ethnic groups. In practice, an ethnic group may be present in one district but not in another. In order to rank such districts, one must either omit the ethnic group in question or combine it with some other group. The Atkinson indices are also sensitive to zeroes. If each school in a district lacks students from at least one ethnic group, then according to the Atkinson indices that district is as segregated as one in which there is no ethnic mixing at all. In this respect, the indices are analogous to the Cobb-Douglas utility functions, which take a constant value of zero if the quantity of any good is zero. Like Cobb-Douglas functions, the Atkinson indices should be used only with sufficiently aggregated data in which zeroes are rare.

We next turn to a characterization of the Mutual Information index, which does not have these limitations, although it is not Composition Invariant. This index represents the unique nontrivial ordering that satisfies Scale Invariance, Independence, the School Division Property, Symmetry, the Group Division Property, and a technical continuity property. It is related to the concept of entropy. Consider a discrete random variable x that takes K possible values. Let q_k be the probability of the kth value of x. For instance, if x is the ethnic group of a randomly selected student, then q_k is the proportion of district students who are in the kth group. The entropy of x is a measure of the uncertainty in x.[2]

The Mutual Information index is defined as follows. Suppose a student is drawn randomly from the district. Initially, we know nothing more about the student; our uncertainty about her race is measured by the entropy of her district's ethnic distribution. Now say that, in addition, we are told which school the student attends. Our uncertainty about her race is now measured by the entropy of her *school*'s ethnic distribution. If the schools in the district are at all segregated, then this entropy will tend to be lower: the student's school conveys some information about her race. The Mutual Information index, M, equals this

---

[2]The entropy of x is $\sum_{k=1}^{K} q_k \log_2\left(\frac{1}{q_k}\right)$. Among other things, it is an upper bound on the average number of bits needed to encode a series of i.i.d. realizations of x (Cover and Thomas [14]).

change in entropy, averaged over the students in the district:

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

Mutual information treats ethnic groups and schools symmetrically: either variable leads to the same reduction in uncertainty about the other (Cover and Thomas [14, pp. 18 ff.]). Hence, the Mutual Information index also equals the reduction in uncertainty about a student's *school* that comes from learning her *race*.

The rest of this paper is as follows. Section 2 defines notation and presents our axioms. The Atkinson and Mutual Information indices are defined and characterized in section 3. Section 4 discusses other school segregation indices, various decomposability properties, and the relation of our results to prior axiomatic treatments of inequality and segregation indices. Additional related literature is reviewed in section 5. Section 6 contains empirical applications using U.S. public school data and section 7 concludes.

2   Definitions and Axioms

We assume a continuum population. This is a reasonable approximation when ethnic groups are large. Formally, we define a (school) district as follows:

Definition 1 A *district* X is a triplet, $\left\langle \mathbf{N}, \mathbf{G}, \left((T_g^n)_{g \in \mathbf{G}} \right)_{n \in \mathbf{N}} \right\rangle$, where $\mathbf{N}$ is a finite, nonempty set of schools, $\mathbf{G}$ is a finite, nonempty set of ethnic groups, and, for each ethnic group $g \in \mathbf{G}$ and school $n \in \mathbf{N}$, T_g^n is a nonnegative real number that represents the number of members of ethnic group g that attend school n.

With some abuse of notation we will sometimes specify a district as a list of ethnic compositions of schools in the district. For instance, $\langle (10, 20), (30, 10) \rangle$ denotes a district with two schools and two ethnic groups - say, blacks and whites. The first school, $(10, 20)$,

contains ten blacks and twenty whites; the second, $(30, 10)$, contains thirty blacks and ten whites.

For any nonnegative scalar α, αX denotes the district in which the number of students in each group and school of X has been multiplied by α. If X and Y are two districts with the same set of ethnic groups, $X \uplus Y$ denotes the result of combining them into one district. For example, if $X = \langle (10, 20), (30, 10) \rangle$ and $Y = \langle (40, 50) \rangle$, then $2X = \langle (20, 40), (60, 20) \rangle$, 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

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

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

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

$t_g^n = \dfrac{T_g^n}{T_g}$: the proportion of students in ethnic group g who attend school n

$r_g^n = \dfrac{p_g^n}{P_g}$: the *disproportionality ratio* of group g in school n (Reardon and Firebaugh [42])

The *ethnic distribution of a district* is the vector $P = (P_g)_{g \in \mathbf{G}}$ 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 $p^n = (p_g^n)_{g \in \mathbf{G}}$ 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: if $p^n = P$. A district in which every school is representative is *completely integrated,* while a district with no ethnic mixing is *completely segregated.*

We will specify the district as an argument of any of the above quantities when this is needed for clarity. For instance, $T_g(X)$ denotes the number of group-g members in district

X, $\mathbf{G}(X)$ is the set of groups in district X, and so on.

We will state our axioms with respect to an unspecified class of districts, $\mathcal{C}$. Later we will apply them either to the class $\mathcal{C}_K$ of districts with exactly $K \geq 2$ nonempty groups and to the class of all districts with at least one nonempty group, $\mathcal{C}^A = \bigcup_{K=1}^{\infty} \mathcal{C}_K$.

Let $\mathcal{C}(\mathbf{G}, \mathbf{N})$ be the set of all districts with a fixed set $\mathbf{G}$ of groups and a fixed set $\mathbf{N}$ of schools. Let us fix some enumerations $g_1, ..., g_{|\mathbf{G}|}$ of the groups and $n_1, ..., n_{|\mathbf{N}|}$ of the schools. We can then identify each district X in $\mathcal{C}(\mathbf{G}, \mathbf{N})$ with the matrix whose entry in row r, column c equals the number of students in group g_r and school n_c, $T_{g_r}^{n_c}(X)$. Moreover, we can define the distance between two districts to be the Euclidean distance between their corresponding matrices in $\Re_+^{|\mathbf{G}| \times |\mathbf{N}|}$. Under this definition, $\mathcal{C}(\mathbf{G}, \mathbf{N})$ is a metric space.

A *segregation ordering* $\succcurlyeq$ on a class of districts is a complete and transitive binary relation on that set of districts. We interpret $X \succcurlyeq Y$ to mean "district X is at least as segregated as district Y." The relations $\sim$ and $\succ$ are derived from $\succcurlyeq$ in the usual way.[3] We restrict throughout to orderings that treat schools symmetrically.[4]

A related concept is the segregation *index*: a function $S : \mathcal{C} \to \Re$ that assigns to each district a number that is interpreted as the district's segregation level. The index S *represents* the segregation ordering $\succcurlyeq$ if, for any two districts $X, Y \in \mathcal{C}$, $X \succcurlyeq Y$ if and only if $S(X) \geq S(Y)$. While every index induces a segregation ordering, not every ordering is represented by an index.

We impose axioms not on the segregation index but on the underlying segregation ordering. These approaches are not equivalent. As in decision theory, an ordering may be represented by more than one index, and there are orderings that are not captured by any index. We will often say, for brevity, that an index satisfies some axiom. This means that

---

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

[4]More precisely, let X be a district and let σ be a permutation on the set of schools of X. Let X' be the district that results from X if all of the students in each school n are relocated to school $\sigma(n)$. That is, X' has the same sets of groups and schools as X, and the number of students in group g and school $\sigma(n)$ in X' equals the number of students in group g and school n in X. We restrict attention to orderings for which $X' \sim X$.

its underlying ordering satisfies the axiom.

A district's *segregation ranking* or simply its *segregation* is its place in the segregation ordering. We will sometimes say that if a transformation $\sigma : \mathcal{C} \to \mathcal{C}$ 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 $\sigma(X) \sim X$.

2.1 Axioms

Each of our characterizations uses a subset of the following axioms. The first two axioms are purely technical. Nontriviality is used to rule out the trivial segregation ordering, while Continuity ensures that an ordering can be represented by a continuous function.

Nontriviality (N) There exist districts $X, Y \in \mathcal{C}$ such that $X \succ Y$.

Continuity (CONT) For any district $Z \in \mathcal{C}$, the set of districts that have the same groups and schools as Z and that are at least as segregated as Z is closed, as is the set of districts that have the same groups and schools as Z and are no more segregated than Z.[5]

Scale Invariance (SI) 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, $X \sim \alpha X$.

Symmetry (SYM) The segregation in a district is invariant to any permutation of the groups in the district. More precisely, let X be a district and let σ be a permutation on the set of groups of X. Let X' be a district that has the same sets of groups and schools as X, such that the number of students in school n and group $\sigma(g)$ in district X' equals the number of students in school n and group g in district X. Then $X' \sim X$.

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

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 $X' \succsim X$. Furthermore, if either one of the new schools is empty ($T^{n_i} = 0$

---

[5]Formally, these sets are $\{X \in \mathcal{C}\left(\mathbf{G}\left(Z\right), \mathbf{N}\left(Z\right)\right) : Z \succsim X\}$ and $\{X \in \mathcal{C}\left(\mathbf{G}\left(Z\right), \mathbf{N}\left(Z\right)\right) : Z \precsim X\}$.

for some $i \in \{1, 2\}$) or the two schools have the same ethnic distribution ($p^{n_1} = p^{n_2}$), then $X' \sim X$.

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

Group Division Property (GDP) Let $X \in \mathcal{C}$ be a district in which the set of ethnic groups is $\mathbf{G}$. Let X' be the result of partitioning some group $g \in \mathbf{G}$ into two subgroups, g_1 and g_2, such that either one subgroup is empty ($T_{g_i} = 0$ for some $i \in \{1, 2\}$) or the two subgroups have the same distribution across schools ($t_{g_1}^n = t_{g_2}^n$ for all $n \in \mathbf{N}(X)$). Then $X' \sim X$.

Scale Invariance states that the scale of a district does not matter. This axiom is satisfied by all of the common school segregation indices. It is implied by two principles from the literature. The first, "Size Invariance", states that if a school district is duplicated and the resulting two districts are combined into a single large district, segregation should not change. The second, "Organizational Equivalence", states that if two schools with identical ethnic distributions are combined, segregation in the district should not change. Both principles were first proposed by James and Taeuber [30].

Symmetry states that a district's ranking should depend only on the number of each group who attend each school: labels such as "black", "white," etc., do not matter. Although it 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.

Independence is a standard separability axiom. It states that if the students in a subdistrict are reallocated among schools within that subdistrict, then segregation in the district rises if and only if segregation in the subdistrict rises.[6] In particular, what happens to

---

[6]Since X and Y have the same number of each ethnic group, $Y \uplus Z$ is the result of reallocating the students within the subdistrict X of the district $X \uplus Z$. As Independence does not require Y to have

districtwide segregation does not depend on the composition of the rest of the district. This axiom is analogous to the standard independence axiom of expected utility theory (von Neumann and Morgenstern [52]).[7]

The School Division Property states that a district cannot become *less* segregated if a school is split into two new schools. In addition, if the new schools have identical ethnic distributions, then segregation is unchanged. Finally, the presence of empty schools (those with no students) does not affect the segregation of a district. SDP is the only axiom that specifies that some districts are more segregated than others. If a segregation ordering satisfies any of the other axioms, then the reverse ordering does so as well; however, the same is not true for SDP.

The School Division Property is related to two properties: Organizational Equivalence, discussed above, and the Transfer Principle.[8] In the case of two ethnic groups - say, blacks and whites - the Transfer Principle states that if a black (white) person moves from one school to another school in which the proportion of blacks (whites) is higher, then segregation in the district rises. With two ethnic groups, Organizational Equivalence and the Transfer Principle jointly imply the School Division Property.[9] But while SDP and Organizational

---

the same number of schools as X, the reallocation might be accompanied by new school construction or conversion of some schools to other uses.

[7]The expected utility axiom states that if lottery L is weakly preferred to lottery M, then for any $p \in [0,1]$ and any lottery N, $pL + (1-p)N$ is weakly preferred to $pM + (1-p)N$. While the requirement that X and Y have the same size is analogous to the assumption of a constant weight p on L and M, the requirement that X and Y have the same ethnic distribution has no obvious analogue.

[8]The Transfer Principle is a translation by James and Taeuber [30] of an analogous property in the context of inequality, the Pigou-Dalton Transfer Principle (Dalton [17]).

[9]Let X' be the district that results from a district X if a school n in X is divided into two schools, n_1 and n_2. If $p^{n_1} = p^{n_2}$, then Organizational Equivalence (OE) implies $X' \sim X$. If not, let X'' be the district that results from X if school n is instead split into two schools n'_1 and n'_2 with identical ethnic distributions: $n'_1 = \left(T_1^n \frac{T_2^{n_1}}{T_2^{n_1} + T_2^{n_2}}, T_2^{n_1}\right)$ and $n'_2 = \left(T_1^n \frac{T_2^{n_2}}{T_2^{n_1} + T_2^{n_2}}, T_2^{n_2}\right)$. Assume w.l.o.g. that $p_1^{n_1} > p_1^{n_2}$. Then $p_1^{n'_1} \in (p_1^{n_1}, p_1^{n_2})$ and since $T_2^{n'_1} = T_2^{n_1}$, $T_1^{n'_1} < T_1^{n_1}$. Move members of group 1 from n'_2 to n'_1 until the two schools have the same number of members of group 1 as n_2 and n_1, respectively. The district that

Equivalence extend naturally to three or more groups, the Transfer Principle does not.[10]

Imposing both the School Division Property and the mild axiom of Scale Invariance amounts to requiring that a segregation ordering respects the informativeness criterion of Blackwell [4]: if a student's school is at least as informative of her race in X than in Y, then X is at least as segregated as Y. The reasoning is as follows. For an ordering that satisfies Scale Invariance, the segregation ranking of a school district X is determined solely by the proportion of students in each group in the district, $P = (P_g)_{g \in \mathbf{G}}$, and by the distribution of each group across schools, given by the $|\mathbf{G}| \times |\mathbf{N}|$ matrix $t = ((t_g^n)_{g \in \mathbf{G}})_{n \in \mathbf{N}}$. Suppose a student is selected at random and we are informed what school she attends. We can think of this school as a signal of the student's unknown race. The probability, given the student's race g, that the signal n is received is given by entry (g, n) of matrix t. Blackwell [4] calls such a matrix an "experiment": for each value of the unknown variable (here, the student's race), it gives a probability distribution of signals (in this case, schools).

Now let $\mathbf{G}$ be a set ethnic groups, and let $\mathcal{T}_{\mathbf{G}}$ be the set of experiments (matrices) whose rows correspond to the groups in $\mathbf{G}$. One can partially order the matrices in $\mathcal{T}_{\mathbf{G}}$ according to their informativeness.[11] Blackwell [4] shows that the $|\mathbf{G}| \times |\mathbf{N}_1|$ matrix $t \in \mathcal{T}_{\mathbf{G}}$ is at least as informative as the $|\mathbf{G}| \times |\mathbf{N}_2|$ matrix $t' \in \mathcal{T}_{\mathbf{G}}$ if and only if there is a $|\mathbf{N}_1| \times |\mathbf{N}_2|$ Markov probability matrix α such that $t' = t \cdot \alpha$: that is, if it is possible to obtain the signal structure t' by "garbling" the signal structure t. (This garbling consists of retransmitting the signal $n' \in \mathbf{N}_2$ with probability $\alpha_{n,n'}$ whenever we get signal $n \in \mathbf{N}_1$). Grant, Kajii, and Polak [22, Lemma A.1] show, further, that for any two signal structures $t, t' \in \mathcal{T}_{\mathbf{G}}$, there is a Markov matrix α such that $t' = t \cdot \alpha$ if and only if t' can be obtained from t by a finite sequence of

---

results is X'. Hence, by the Transfer Principle and OE, $X' \succ X'' \sim X$.

[10] For instance, consider a district with blacks, whites, and Asians. Suppose a black moves to a school that has higher proportions of both blacks and Asians. Since there are more blacks in the destination school, one might argue (using the Transfer Principle) that segregation has gone up. On the other hand, blacks are now more integrated with Asians, so perhaps segregation has fallen. One attempt to overcome this difficulty appears in Reardon and Firebaugh [42].

[11] See, e.g., Blackwell [4], or Bohnenblust, Shapley, and Sherman [5]

splitting columns apart and sticking equivalent columns together.[12]   But these are precisely the operations referred to in the School Division Property.   Hence, for any two districts X and X' with the same group distribution P, and with likelihood matrices $t \in \mathcal{T}_{\mathbf{G}}$ and $t' \in \mathcal{T}_{\mathbf{G}}$, respectively, t is at least as informative as t' if and only if X is at least as segregated as X' according to all segregation orderings that satisfy SDP and SI. That is, these axioms jointly require that the segregation ordering be consistent with the informativeness ordering of the likelihood matrices induced by districts with the same group distribution.

Composition Invariance requires that an ordering be insensitive to changes in the size of an ethnic group that leave that group's distribution across schools unchanged.   This means that the segregation of a district depends only on how its ethnic groups are distributed across schools in the district.   The first to propose this property were Jahn *et al* [28], who wrote: "a satisfactory measure of ecological segregation should ... not be distorted by the size of the total population, the proportion of Negroes, or the area of a city...." (Jahn *et al* [28]).

Composition Invariance has been controversial.   The dominant view, espoused by Taeuber and James [47], is that segregation refers to the effect of ethnic origins on destinations (schools, neighborhoods, etc.).   Accordingly, they favor Composition Invariance.   Others, such as Coleman, Hoffer, and Kilgore [11], view segregation as capturing different degrees of exposure of one ethnic group to another, and thus oppose the principle.

In order to incorporate these diverse points of view, we also study the effect of replacing Composition Invariance by the Group Division Property.   This axiom states that a segregation measure should not change if an ethnic group is divided into two groups that have identical distributions across schools.   In addition, the presence of empty groups (those with no members) do not affect the segregation of a district.   This axiom is related to the School Division Property, with the roles of groups and schools reversed.   It is our only axiom that tells us how to rank districts with different numbers of ethnic groups.   It does not appear to

---

[12]Splitting a column $(t_g^n)_{g \in \mathbf{G}}$ apart means replacing it by two columns $(t_g^{n_1})_{g \in \mathbf{G}}$, and $(t_g^{n_2})_{g \in \mathbf{G}}$ where $t_g^{n_1} + t_g^{n_2} = t_g^n$ for all $g \in \mathbf{G}$. Two columns $(t_g^{n_1})_{g \in \mathbf{G}}$, and $(t_g^{n_2})_{g \in \mathbf{G}}$ are equivalent if $t_g^{n_1}/t_g^{n_2}$ does not depend on g. Sticking two equivalent columns together means replacing them by their sum.

have been discussed in the prior literature, which has focused mainly on the two-group case.

3 Atkinson and Mutual Information Indices

3.1 Index Definitions

The Atkinson segregation indices were introduced by James and Taeuber [30] for the case of two ethnic groups. They are based on the Atkinson inequality indices (Atkinson [2]).[13] Let $\mathbf{w} = (w_1 \ldots w_K)$ be a vector of K fixed nonnegative weights that sum to one. The Atkinson index with weights $\mathbf{w}$, $A_{\mathbf{w}}$, is defined by

$$A_{\mathbf{w}}(X) = 1 - \sum_{n \in \mathbf{N}(X)} \prod_{g \in \mathbf{G}} \left(t_g^n \right)^{w_g} \tag{1}$$

When all weights are equal, we obtain the symmetric Atkinson index, denoted A. Massey and Denton [33] study properties of the Atkinson indices; Johnston, Poulsen, and Forrest [31] use them to study residential segregation.

The *entropy* of the discrete probability distribution $q = (q_1, \ldots, q_K)$ is defined by[14]

$$h(q) = \sum_{k=1}^{K} q_k \log_2 \left(\frac{1}{q_k} \right). \tag{2}$$

The Mutual Information index equals the entropy of a district's ethnic distribution minus

---

[13] In the case of two groups, the Atkinson index with weight $0 < \delta < 1$ on group one equals

$$1 - \left[\sum_{n \in \mathbf{N}(X)} \left(t_1^n \right)^{\delta} \left(t_2^n \right)^{1-\delta} \right]^{\frac{1}{1-\delta}}.$$

It is due to James and Taeuber [30, p. 9], who derive it from the inequality index of the same name. The Atkinson index is difficult to generalize to more than two groups since the outer exponent, $\frac{1}{1-\delta}$, is the reciprocal of the weight on a particular ethnic group. Instead, we generalize $1 - \sum_{n \in \mathbf{N}(X)} \left(t_1^n \right)^{\delta} \left(t_2^n \right)^{1-\delta}$. As this is an increasing transformation of the original index, it represents the same ordering.

[14] When $q_k = 0$, the term $q_k \log_2(1/q_k)$ is assigned the value zero.

13

the average entropy of the ethnic distributions of its schools:

$$M(X) = h(P) - \sum_{n \in \mathbf{N}(X)} \pi^n h(p^n) \tag{3}$$

This index was first proposed by Theil [49] and has been applied by Fuchs [21] and Mora and Ruiz-Castillo [35, 38].[15]

Properties of these indices are discussed in section 3.3.

3.2 Main Results

Our characterization results are as follows. Throughout, we assume the axioms of Scale Invariance, Independence, the School Division Property, and Nontriviality. When, in addition, we assume Composition Invariance and Continuity, we obtain the family of Atkinson orderings:

Theorem 1 *Let $K \geq 2$. An ordering $\succcurlyeq$ on $\mathcal{C}_K$ satisfies Scale Invariance, Independence, the School Division Property, Nontriviality, Composition Invariance, and Continuity if and only if there exist fixed weights $w_g \geq 0$ for $g = 1, ..., K$, adding up to one, such that $\succcurlyeq$ is represented by the Atkinson index $A_{\mathbf{w}}(X)$.*

If we replace Composition Invariance by the Group Division Property and add Symmetry, we obtain the Mutual Information ordering:

Theorem 2 *An ordering on $\mathcal{C}^A$ satisfies Scale Invariance, Independence, the School Division Property, Nontriviality, the Group Division Property, Symmetry,and Continuity if and only if it is represented by the Mutual Information index.*

---

[15]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 [36, 37].

3.3 Discussion

The Atkinson indices are Composition Invariant. This makes them a natural choice to study the effects of a student's ethnic origin on her school destination, as they are affected only by differences in how ethnic groups are distributed across schools. On the other hand, they are sensitive to zeroes in the case of three or more ethnic groups. For instance, they rank the districts $X = \langle(10, 10, 0), (0, 0, 10)\rangle$ and $Y = \langle(10, 0, 0), (0, 10, 0), (0, 0, 10)\rangle$ as equally segregated. One can see from this example that the Atkinson indices do not satisfy a stronger version of the School Division Property in which splitting a school into two schools with different ethnic distributions leads to *strictly* higher segregation. (The Atkinson indices are free of these limitations in the case of two ethnic groups.)

The Mutual Information index ranks Y as strictly more segregated than X and satisfies this stronger version of SDP. It also has several useful decompositions that the Atkinson and other indices lack (section 4.2). These make this index a good choice for studying the sources of segregation at different geographic and ethnic levels. However, the Mutual Information index violates Composition Invariance. Accordingly, the Mutual Information index is unsuitable for judging whether different ethnic groups are becoming more similarly distributed across schools. These differences are illustrated empirically in section 6.

Why does the Mutual Information index violate Composition Invariance? Multiplying the number of students in a given group by a common factor in every school alters the ethnic distributions of the schools and of the district as a whole. This changes both our initial uncertainty about a student's ethnicity, as well as our residual uncertainty after learning her school. These changes are not necessarily equal. Hence, their difference - the Mutual Information index - can change as well. For instance, consider the district $X = \langle(10, 0), (0, 1000)\rangle$ and let $Y = \langle(1000, 0), (0, 1000)\rangle$ be the result of scaling the first group up by a factor of 100. While this change greatly increases our initial uncertainty about a student's race, it has no effect on our residual uncertainty after learning her school as there is none. Hence, the Mutual Information index is higher in Y, while a Composition Invariant index would regard them as equally segregated.

The Atkinson indices are defined for a fixed set of groups, each of which has a fixed weight. Thus, they are not designed to compare districts with different numbers of nonempty groups. One can use the symmetric Atkinson index in an ad-hoc way to make such comparisons by lowering the weight on each group from $1/K$ to $1/(K+1)$ when a group is subdivided. However, this usage clearly violates the Group Division Property.

4   Other Indices

In this section we introduce other school segregation indices and several decomposability properties. All claims not proved here are proved in Appendix B.

4.1   Other Indices

We begin by stating some simple facts that will often let us verify an axiom by writing an index in a particular way. Denote by $\pi = (\pi^n)_{n \in \mathbf{N}}$ the vector of relative sizes of schools in a district. Let $r_g = (r_g^n)_{n \in \mathbf{N}}$ be the vector of disproportionality ratios of group g. Recall that $p^n = (p_g^n)_{g \in \mathbf{G}}$ is the ethnic distribution of school n and $t = ((t_g^n)_{g \in \mathbf{G}})_{n \in \mathbf{N}}$ is the matrix of distributions of groups across schools.

Claim 1   *1. If an index can be written in the form $F_0(P) + F_1(P) \sum_{n \in \mathbf{N}} \pi^n F_2(p^n)$ for some functions F_0, $F_1 > 0$, and F_2, then it satisfies Independence.*

   2. If an index can be written as $\sum_{n \in \mathbf{N}} \pi^n F_3(P, p^n)$ for some F_3 that is convex in p^n, then the index satisfies the School Division Property.

   3. If an index can be written as a function of t alone, then it satisfies Composition Invariance.

   4. If an index can be written in the form $\sum_{g \in \mathbf{G}} P_g F_4(\pi, r_g)$ for some function F_4, then it satisfies the Group Division Property.

By part 3 of Claim 1, the Atkinson indices satisfy Composition Invariance. Using $t_g^n = \pi^n \frac{p_g^n}{P_g}$, we can rewrite $A_{\mathbf{w}}(X)$ as $1 - \left(\prod_{g \in \mathbf{G}} (P_g)^{w_g}\right)^{-1} \sum_{n \in \mathbf{N}} \pi^n \prod_{g \in \mathbf{G}} (p_g^n)^{w_g}$. Hence, by parts 1 and 2 of Claim 1, the Atkinson indices also satisfy Independence and SDP.[16]

By part 1 of Claim 1, the Mutual Information index satisfies Independence. Letting $F_3(P, p^n) = h(P) - h(p^n)$, the index satisfies SDP by part 2. Finally, the index can be written as $\sum_{g \in \mathbf{G}} P_g \sum_{n \in \mathbf{N}(X)} \pi^n r_g^n \log_2 r_g^n$, so by part 4 it satisfies the Group Division Property.

The Entropy index $H(X)$ of Theil [50] and Theil and Finizza [51] equals $M(X)/h(P)$ where $M(X)$ is the Mutual Information index and $h(P)$ is the entropy of the group distribution of district X. While the Mutual Information index has no maximum value, the Entropy index has a maximum value of one. Like the Mutual Information index, the Entropy index satisfies Independence and SDP by Claim 1, parts 1 and 2,[17] but cannot be written as a function of t and thus violates Composition Invariance. As the Mutual Information index satisfies GDP, the presence of the factor $1/h(P)$, which is not invariant to group division, means that the Entropy index violates GDP.

For any group distribution P, let $I(P) = \sum_{g \in \mathbf{G}} P_g(1 - P_g)$ denote the Simpson Interaction index (Lieberson [32]). The Dissimilarity index $D(X) = \frac{1}{2I(P)} \sum_{g \in \mathbf{G}} P_g \sum_{n \in \mathbf{N}} \pi^n |r_g^n - 1|$ of Morgan [39] and Sakoda [44] is a generalization of the two-group Dissimilarity index of Jahn, Schmid, and Schrag [28]. It has an unnormalized version $D' = I(P)D$, which equals the minimum proportion of students who would have to change schools, keeping school sizes fixed, in order to completely integrate the district. $I(P)$ is what this proportion would be under complete segregation. Hence, the Dissimilarity index D is the result of normalizing D' to take a maximum value of one.

The Gini index $G(X) = \frac{1}{2I(P)} \sum_{g \in \mathbf{G}} P_g \sum_{m \in \mathbf{N}} \sum_{n \in \mathbf{N}} \pi^m \pi^n |r_g^m - r_g^n|$ of Reardon [41] is a generalization of the two-group Gini index of Jahn, Schmidt, and Schrag [28]. In the case of two ethnic groups - say, blacks and whites - the Gini index measures the area between the

---

[16]For SDP, $F_3(P, p^n) = 1 - \prod_{g \in \mathbf{G}} (p_g^n/P_g)^{w_g}$, which is convex in each p_g^n as $w_g \leq 1$.

[17]For part 2, let $F_3 = 1 - h(p^n)/h(P)$.

17

Lorenz curve and the 45 degree line, while the Dissimilarity index measures the maximum vertical distance between the Lorenz curve and the same line. (For a definition of Lorenz curves and a proof of this result, see James and Taeuber [30].) Since the Lorenz curve depends only on the distributions of the two groups across schools, the Gini and Dissimilarity indices satisfy Composition Invariance in this case. However, they violate it when there are three or more groups (Reardon and Firebaugh [42]).

Both the Dissimilarity and Gini indices satisfy SDP.[18] By Claim 1, part 4, the unnormalized versions of these indices, D' and $G' = I(P)G$, satisfy GDP. But since $I(P)$ is not invariant to group division, D and G violate GDP. Moreover, both indices violate Independence. Intuitively, the interaction of different ethnic distributions within the absolute value function creates a link between segregation within a subdistrict and the composition of the rest of the district.

The Normalized Exposure index was originally proposed by Bell [3] for the case of two groups. Let group 1 represent blacks and let $P^* = \sum_{n \in \mathbf{N}} t_1^n p_1^n$ be the percent black in the school attended by the average black student. Bell [3] calls P^* the index of Isolation. The Normalized Exposure index P for two groups is the result of normalizing P^* to lie between zero and one: $NE = \frac{P^* - P_1}{1 - P_1} = \sum_{n \in \mathbf{N}} t_1^n \frac{p_1^n - P_1}{1 - P_1} = \sum_{n \in \mathbf{N}} \pi^n r_1^n \frac{p_1^n - P_1}{1 - P_1}$. The index can be generalized to an arbitrary number of groups as $NE(X) = \sum_{n \in \mathbf{N}} \sum_{g \in \mathbf{G}} \pi^n P_g \frac{P_g}{1 - P_g} (r_g^n - 1)^2$ (James [29]). The factor $\frac{P_g}{1 - P_g}$ guarantees that this index takes a maximum value of one. If this factor is omitted, the resulting index satisfies GDP by Claim 1, part 4. Since the factor is not invariant to group division, NE violates GDP. It is well known that the index is not Composition Invariant (Taeuber and James [47]; Coleman, Hoffer, and Kilgore [11]). With two groups, NE equals $\frac{1}{P_1(1 - P_1)} \sum_{n \in \mathbf{N}} \pi^n (p_1^n)^2 - \frac{P_1}{1 - P_1}$, so it satisfies Independence in this case (Claim 1, part 1), but not with three or more groups. Finally, NE satisfies SDP by part 2 of Claim 1.[19]

---

[18]For D, let $F_3 = (2I(P))^{-1} \sum_{g \in \mathbf{G}} |p_g^n - P_g|$ in part 2 of Claim 1. For Gini, see Appendix B.

[19]Let $F_3 = \sum_{g \in \mathbf{G}} \frac{(p_g^n - P_g)^2}{1 - P_g}$, which is convex in p_g^n.

The last two indices are defined for a fixed set of groups, so the Group Division Property is not applicable. Clotfelter [10] has measured segregation as the percentage of blacks who attend schools in which at least some proportion $\kappa \in (0,1)$ of students are nonwhite. More formally, the index equals $C_\kappa(X) = \frac{1}{P_2} \sum_{n \in \mathbf{N}(X)} \pi^n p_2^n 1\, (p_2^n \geq \kappa)$, where group 2 denotes blacks or nonwhites. Clotfelter's Index satisfies Independence by Claim 1, part 1. However, it clearly violates Symmetry, Continuity, and Composition Invariance. It also violates the School Division Property: if a school whose proportion black is slightly above κ is split into two schools, one of which has a black proportion slightly below κ, the index falls rather than rises.

Card and Rothstein [8] measure segregation as the average fraction black or Hispanic in the schools attended by the typical black and white student, and define segregation as the difference between these figures. Thus, their index can be written $CR(X) = \sum_{n \in \mathbf{N}(X)} (t_2^n - t_1^n)(p_2^n + p_3^n)$ where groups 1, 2, and 3 denote whites, blacks, and Hispanics, respectively. The Card-Rothstein Index cannot be written in the forms given in Claim 1, parts 1, 2, and 3, and indeed violates IND, SDP, and CI. It also clearly violates Symmetry.

4.2 Decomposability

We now turn to cardinal decomposability properties. These are useful if one wants to study segregation at several levels simultaneously. For instance, one may be interested in how much of the segregation between schools in a metropolitan area is due to segregation between districts and how much is due to segregation within districts. The analogous property for ethnic groups might be used, e.g., to study the relative importance of race and religion in generating school segregation.

The first two properties are due to Hutchens [27] and are based on Shorrocks's [45, 46] analogous properties for inequality indices. The third property is analogous to the Theil Decomposability axiom used by Bourguignon [7] and by Foster [20] in their characterization of the Theil index of income inequality. It was previously discussed by Mora and Ruiz-Castillo [36]. When the roles of schools and groups in the third property are swapped, we

obtain the fourth property, a type of decomposability across groups.

For any district Z, let $T.(Z) = (T_g(Z))_{g \in \mathbf{G}}$ be the number of members of each ethnic group in Z and let $c(Z)$ be the district that results if the schools of Z are all merged into a single school. Recall that $T(Z)$ denotes the total population of Z.

Aggregation (AGG) An index S is Aggregative if there is a function F such that $S(X \uplus Y) = F(S(X), S(Y), T.(X), T.(Y))$, where F is continuous and strictly increasing in $S(X)$ and $S(Y)$.

Additive Decomposability (AD) An index S is Additively Decomposable if there are strictly positive functions $w_X(T.(X), T.(Y))$ and $w_Y(T.(X), T.(Y))$ such that, for all districts X and Y,

$$S(X \uplus Y) = S(c(X) \uplus c(Y)) + w_X(T.(X), T.(Y)) S(X) + w_Y(T.(X), T.(Y)) S(Y) \quad (4)$$

Strong School Decomposability (SSD) An index S satisfies Strong School Decomposability if, for all districts X and Y,

$$S(X \uplus Y) = S(c(X) \uplus c(Y)) + \frac{T(X)}{T(X) + T(Y)} S(X) + \frac{T(Y)}{T(X) + T(Y)} S(Y).$$

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

$$S(X) = S\left(\widehat{X}\right) + P_1 S(X_1) + P_2 S(X_2), \quad (5)$$

where P_k is the proportion of students who are in supergroup k, $\widehat{X}$ is the district that results from district X if each supergroup is treated as a group (i.e., ignoring within-supergroup ethnic differences), and X_k is the district that results from district X if all students not in supergroup k are removed.[20]

AGG states that for any partition of a district into subdistricts, overall segregation is some function of within-subdistrict segregation and the size and ethnic distributions of the

---

[20]For instance, let $X = \langle (1, 2, 3, 4), (5, 6, 7, 8) \rangle$ and let supergroup 1 consist of the first two groups and supergroup 2 comprise the last two. Then $\widehat{X} = \langle (1 + 2, 3 + 4), (5 + 6, 7 + 8) \rangle$, $X_1 = \langle (1, 2), (5, 6) \rangle$, and $X_2 = \langle (3, 4), (7, 8) \rangle$.

subdistricts. AD and SSD require, in addition, that this function equal the sum of between-subdistrict segregation and a weighted average of within-subdistrict segregation. While AD permits a subdistrict's weight to depend in a general way on the size and ethnic distributions of all subdistricts, SSD requires that this weight equal the proportion of students who attend schools in the subdistrict. Finally, SGD results from swapping the roles of schools and ethnic groups in SSD: for any partition of groups into supergroups, overall segregation is the sum of between-supergroup segregation and the population-weighted average of segregation within each supergroup.

SSD clearly implies AD, which in turn implies AGG. However, the opposite is not true. The square of any nontrivial index that satisfies AD must violate AD, but still satisfies AGG. Furthermore, the Atkinson indices satisfy AD[21] but violate SSD and SGD by the following claim.

Claim 2 *Let $\overline{X}^2$ be the district $\langle (1,0) , (0,1) \rangle$. Let the index S satisfy SI and SDP, and suppose that $S\left(\overline{X}^2\right) > 0$. Then either S is unbounded or it violates SSD and SGD.*

The Mutual Information index, the only unbounded index we consider, satisfies SSD, as previously shown by Mora and Ruiz-Castillo [36] in the case of two ethnic groups. By the duality of mutual information (Cover and Thomas [14, pp. 18 ff.]), it also satisfies SGD. We will illustrate some uses of these properties in section 6.

Analogous properties have been extensively discussed in the income inequality literature. Bourguignon [7] and Foster [20] show that Theil Decomposability, which is analogous to SSD, fully characterizes the Theil inequality index (Theil [48]) within the class of relative inequality indices.[22] However, our characterization of the Mutual Information index differs qualitatively from these prior results. While both Theil and Mutual Information measure reductions in uncertainty - about the owner of a dollar and the race of a random student, respectively - the baselines are different: the Theil index starts with a uniform prior distribution while the

---

[21]Hutchens [27] proves this in the two-group case.

[22]While Bourguignon [7] restricts to differentiable indices, Foster [20] assumes only continuity.

Mutual Information index begins with the districtwide ethnic distribution.[23]   In addition, unlike Foster [20] and Bourguignon [7], we do not assume SSD. As the following claim shows, our only separability axiom, Independence, is analogous not to SSD but rather to the weaker property of Aggregation:[24]

Claim 3 *Aggregation implies Independence. In the class of segregation orderings that are represented by continuous indices, Aggregation and Independence are equivalent.*

Finally, Foster [20] shows that the Theil index is the only continuous relative inequality index that satisfies Theil Decomposability. In this sense, Theil Decomposability is a defining property of the Theil index.   The same is not true of Independence: indeed, every axiom that we use to characterize the Mutual Information index is satisfied by at least one of the other segregation indices that we survey in section 4.1.[25]

Strong Group Decomposability also does not correspond to any of our axioms. While it implies GDP,[26] the reverse is not true: the unnormalized Dissimilarity and Gini indices D' and G' satisfy GDP but not SGD as they are bounded (by one) and satisfy the assumptions of Claim 2.

---

[23]Let T be the number of persons in a city, y_i be the wealth of person i, and $|y| = \sum_{i=1}^{T} y_i$ be the total wealth in the city.   Let $y = (y_i/|y|)_{i=1}^{T}$ be the distribution of wealth in the city.   The Theil index equals $h\left(\left(\frac{1}{T}, ..., \frac{1}{T}\right)\right) - h(y)$: the difference in entropies of the uniform and actual distributions of wealth.

[24]A segregation index is *continuous* if it is a continuous function of the numbers of members of each group in each school, $(T_g^n)_{n \in \mathbf{N}, g \in \mathbf{G}}$.

[25]This follows from Table 1 below, together with the fact that D' and G' satisfy GDP.

[26]More precisely, this holds for any index S that equals zero on districts that are completely integrated. In particular, suppose X' has K groups. Let X be the result of splitting some group g in X' into two subgroups, g_1 and g_2, such that either one subgroup is empty or the two subgroups have the same distribution across schools. Let us now partition X into K supergroups, such that each supergroup consists of students who were in a given group of X'. By SGD, $S(X)$ can be written as the sum of between- and within-supergroup terms as in (5). By construction, the between-supergroup term equals $S(X')$. Since each district X_k is completely integrated, each within-supergroup term is zero. Thus, $S(X) = S(X')$.

4.3 Discussion

A list of the indices we have discussed appears in Table 1, together with an indication of which properties they satisfy. Results not proved above are shown in Appendix B. The first six rows correspond to our axioms, while the last four pertain to different types of decomposability. Nontriviality and Scale Invariance are omitted as all of the indices satisfy them.

PROPERTY	Index $A_{\mathbf{w}}$	M	D	G	H	NE	C_{κ}	CR
Continuity	√	√	√	√	√	√	×	√
Symmetry	×	√	√	√	√	√	×	×
Independence	√	√	×	×	√	2	√	×
School Division Property	√	√	√	√	√	√	×	×
Composition Invariance	√	×	2	2	×	×	×	×
Group Division Property	N/A	√	×	×	×	×	N/A	N/A
Aggregation	√	√	×	×	√	2	√	×
Additive Decomposability	√	√	×	×	√	2	×	×
SSD	×	√	×	×	×	×	×	×
SGD	N/A	√	×	×	×	×	N/A	N/A

Table 1: Which Indices Satisfy Which Properties? The Atkinson index satisfies symmetry when the group weights are equal. A "√" means that the property holds; an "×" indicates that it does not. A "2" means that the property is satisfied only in the case of two ethnic groups. A "S" means that the property holds only for the symmetric version of the index.

We can draw two sorts of conclusions from this table. First, if one is interested in the effect of ethnic origin on school destination in the presence of three or more groups, one should use an Atkinson index since only they are Composition Invariant in this case. On the other hand, in order to study the geographic and ethnic sources of segregation, the Mutual Information index, with its decomposability properties, would be a good choice. These

conclusions are illustrated empirically in section 6.

The table also points to a tradeoff between intuitiveness and performance in choosing an index. While the Atkinson and Mutual Information are a bit complex, their formal properties make them well-behaved in a variety of situations. Some of the other indices, while highly intuitive, violate several axioms and so can behave in undesirable ways.

5   Related Literature

The first to study segregation axiomatically, Philipson [40], ordinally characterizes the family of segregation orderings that can be represented by a population-weighted average, across schools, of some fixed function u of the school's ethnic distribution: $S = \sum_{n \in \mathbf{N}} \pi^n u(p^n)$. However, the only way to write any of the common indices in this way is to let the function u depend also on the *districtwide* ethnic distribution. Since u is a fixed function, Philipson's analysis is relevant only to comparisons of districts that have the same ethnic distribution.

Hutchens [27] characterizes segregation indices that satisfy a set of cardinal axioms in the case of two ethnic groups. His Theorem 1 shows that a continuous index satisfies cardinal versions of Composition Invariance (Hutchens's axiom P1), symmetry across schools (P2), the Transfer Principle (P3), Organizational Equivalence (P4), and Aggregation (P5), if and only if it is an increasing transformation of the Atkinson index for the two-group case. As the Transfer Principle and Organizational Equivalence jointly imply the School Division Property (section 2.1) and Aggregation implies Independence (Claim 3), Hutchens's theorem is a kind of cardinal version of our Theorem 1 for the two-group case. However, while Hutchens begins with an index, we start with an ordering and prove that it is represented by an index. In addition, Hutchens's proof relies on the isomorphism between inequality measures and segregation indices in the two-group case. Hence, it cannot be generalized in a simple way to the case of three or more groups.[27]

---

[27]In an earlier paper, Hutchens [26] uses a nonstandard separability property in place of Aggregation. This yields a family of indices that includes the Atkinson indices as well as some indices that violate Independence.

More recently, Echenique and Fryer [18] characterize an index that measures the strength of an individual's isolation from members of other demographic groups. They also rely on cardinal axioms. As the inputs to their index are data on social networks, their analysis has little in common with ours.

6   Empirical School Segregation Patterns

In this section we study the empirical performance of the main school segregation indices. We use the Common Core of Data (CCD), which contains student ethnic counts for virtually all public schools in the U.S. from 1987/8 to 2007/8. Four ethnic groups are used: Asians, (non-Hispanic) whites, (non-Hispanic) blacks, and Hispanics.[28]

We first study changes in school segregation over the period. Attention is restricted to the 60,674 schools that reported positive attendance in every school year from 1987/8 to 2007/8.[29] We focus on total segregation among U.S. schools, essentially treating the U.S. as a single district and studying its evolution over time.

Three sets of indices are computed. The first consists of Composition Invariant indices. This set comprises the symmetric Atkinson index A, which gives equal weight to each ethnic group, and two asymmetric Atkinson indices $A_{\mathbf{w}}$, in which a group's weight equals the proportion of students in our sample who are in the given group in either 1987/8 or 2007/8.

Indices that are not Composition Invariant fall, heuristically, into two groups. Each such index begins with a quantity that captures some intuitive notion of segregation. Sometimes

---

[28]The CCD actually has five ethnic groups. The smallest, American Indian/Alaskan Native, is not represented in some school districts. Since most segregation indices (including the Atkinson indices) are not well defined on such districts, we excluded this ethnic group from our analysis.

[29]To aid in matching, for 1987/8 through 1998/99 we used the 13-year longitudinal version of this database (McLaughlin [34]). For subsequent years, we used the annual files. Schools that closed for one or more years and then reopened are excluded from our sample. Since parents and teachers may prefer not to move back after they have gotten used to new schools, the sense in which these are actually "the same schools" is open to debate.

this quantity itself is used as the segregation index: the index is unnormalized. This may be because the index already takes a maximum value of one, or because normalization destroys certain desirable properties. This set consists of the Mutual Information index M, the Clotfelter index C_κ, the Card-Rothstein index CR, and the unnormalized Dissimilarity index D'. In other cases, the intuitive quantity is normalized by dividing by the maximum value it can take, given the district's ethnic distribution. This set consists of the Gini index G, Dissimilarity D, Normalized Exposure index NE, and Entropy H.

Results appear in Table 2. Over the period we study, school segregation measures were affected by two important developments. First, ethnic groups were becoming more similarly distributed across schools: Panel 1 shows declines in the pairwise symmetric Atkinson indices for all six pairs of ethnic groups. As a result, we see declines in the Composition-Invariant indices in Panel 3. At the same time, ethnic diversity was growing significantly (Panel 2). This change dominated for the unnormalized Composition-Variant indices: they tend to show large *increases* over the period (Panel 4).[30] As for the normalized Composition-Variant indices, increased ethnic diversity led to offsetting increases in both the intuitive quantities on which these indices are based, as well as the maximum possible values of these quantities. The end result was little discernible change in the indices themselves (Panel 5).

We now turn to cross-sectional patterns of segregation in 2007/8. We restrict to school districts that contain at least two schools and that serve grades K-12. Schools not located in Core Based Statistical Areas (CBSA's) or that do not lie in the 50 U.S. states and the District of Columbia are excluded. We refer to the resulting set of schools as "urban schools".

Table 3 computes the Mutual Information index for all urban schools in the U.S. and decomposes it using the properties of Strong School and Group Decomposability. Since supergroup schemas must be nested in order to apply Strong Group Decomposability, we

---

[30]For instance, the increase in the Mutual Information index shows that a randomly selected student's school now conveys more information about her race. This is driven by the fact that there is now more information to convey: since ethnic diversity has increased, the initial uncertainty about a random student's race is now greater.

	School Year		
	1987/8	2007/8	Change
1. PAIRWISE ATKINSON INDICES			
White-Black	0.423	0.401	-0.022
White-Hispanic	0.498	0.409	-0.089
White-Asian	0.369	0.324	-0.046
Black-Hispanic	0.549	0.401	-0.148
Black-Asian	0.520	0.438	-0.083
Hispanic-Asian	0.410	0.344	-0.066
2. PUBLIC SCHOOL ETHNIC DISTRIBUTION			
Whites	72.1%	58.2%	-13.9%
Blacks	14.9%	16.5%	1.6%
Hispanics	9.9%	20.6%	10.7%
Asians	3.1%	4.7%	1.6%
TOTAL	100.0%	100.0%	0.0%
3. COMPOSITION INVARIANT INDICES			
Symmetric Atkinson	0.635	0.519	-0.116
Asymmetric Atkinson (2006/7 Weights)	0.619	0.469	-0.149
Asymmetric Atkinson (1987/8 Weights)	0.603	0.443	-0.160
4. UNNORMALIZED COMPOSITION-VARIANT INDICES			
Mutual Information	0.558	0.658	0.100
Clotfelter (50% Threshold)	0.556	0.680	0.123
Clotfelter (90% Threshold)	0.249	0.313	0.064
Card-Rothstein	0.447	0.471	0.024
Unnormalized Dissimilarity (D')	0.290	0.366	0.076
5. NORMALIZED COMPOSITION-VARIANT INDICES			
Gini	0.818	0.790	-0.028
Dissimilarity	0.649	0.621	-0.028
Normalized Exposure	0.435	0.442	0.007
Entropy Segregation Index	0.452	0.422	-0.030
6. MISCELLANEOUS			
Entropy of U.S. Public School Ethnic Distribution	1.236	1.561	0.325
Simpson Interaction Index	0.448	0.590	0.142
Number of Schools	60674	60674	0
Number of Students (millions)	32.252	32.709	0.457

Table 2: Summary Statistics for U.S. Public Schools, 1987/8 and 2007/8. Universe is set of U.S. public schools that report positive numbers of students in Common Core of Data for all school years from 1987/8 to 2007/8. Panel 1 shows the symmetric Atkinson index, computed separately for all six pairs of ethnic groups. Panel 2 shows the aggregate ethnic distribution. Three indices that satisfy Composition Invariance appear in Panel 3. Among indices that violate this axiom, unnormalized indices appear in Panel 4 and normalized indices appear in Panel 5. Panel 6 shows the entropy of the aggregate ethnic distribution $H(P)$, the Simpson Interaction Index I, the total number of schools, and the total number of students.

	Geographic Level	1 Total	2 Hispanic vs. Non Hispanic	3 Black vs. White & Asian	4 White vs. Asian
	Absolute Contribution to Total Segregation among U.S. Public Schools, 2007/8				
1	Total: Between Urban Schools in US*	**0.651**	0.293	0.273	0.084
2	Between States in US	0.208	**0.124**	**0.049**	**0.035**
3	Between CBSAs in States	0.104	**0.051**	**0.040**	**0.013**
4	Between Districts in CBSAs	0.210	**0.062**	**0.125**	**0.023**
5	Between Schools in Districts	0.129	**0.057**	**0.059**	**0.013**
	Percentage Contribution to Total Segregation among U.S. Public Schools, 2007/8				
6	Total: Between Urban Schools in US*	**100.0%**	45.1%	41.9%	13.0%
7	Between States in US	31.9%	**19.0%**	**7.6%**	**5.3%**
8	Between CBSAs in States	15.9%	**7.8%**	**6.1%**	**2.0%**
9	Between Districts in CBSAs	32.3%	**9.6%**	**19.2%**	**3.5%**
10	Between Schools in Districts	19.8%	**8.7%**	**9.0%**	**2.1%**

Table 3: Decomposition of Segregation Between Urban Schools in U.S., 2007/8 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. Source is the Common Core of Data. The Mutual Information Index is computed for all schools in universe defined above and decomposed into various components. Ethnic groups are mutually exclusive: Asians, non-Hispanic whites, non-Hispanic blacks, and Hispanics. The three terms in equation (6) appear in columns 2-4. Column 2 shows how much segregation at the given geographic level is due (in an accounting sense) to segregation between Hispanics and non-Hispanics. Column 2 shows the contribution of segregation between blacks, on the one hand, and whites and Asians, on the other. Column 3 shows the contribution of segregation between whites and Asians. The sum of these numbers appears in column 1 and (by Strong Group Decomposability) represents segregation between the four ethnic groups at the given geographic level. This analysis is performed at four geographic levels. Row 1 computes segregation among all U.S. public schools, treating the U.S. as a single "district". Row 2 computes segregation among U.S. states, treating each state as a single "school". For row 3, 51 state-level segregation indices are first computed, treating the CBSA's in a state as individual schools. The figures shown are the averages of these 51 state-level indices, weighted by the number of students in the state who come from the given ethnic groups. For row 4, we first compute segregation within each CBSA, treating each school district as an individual school. The figures shown are the averages of these CBSA-level indices, weighted by the number of students in the CBSA who belong to the given ethnic groups. Finally, for row 5 we first compute a segregation index for each school district. The figures shown are the averages of these indices, weighted by the number of district students who belong to the given ethnic groups. By Strong School Decomposability, the sum of rows 2-5 equals total segregation between schools in the U.S., which appears in row 1. In panel B, all indices are re-expressed as percentages of this total.

remove one ethnic group at a time. 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 (5) twice, we can decompose overall segregation into three terms:

$$
\begin{pmatrix} \text{Segregation} \\ \text{among} \\ H, B, W \ \& \ A \end{pmatrix} = \begin{pmatrix} \text{Segregation} \\ \text{between} \\ H \ \& \ \{B, W, A\} \end{pmatrix} + \begin{pmatrix} \text{Proportion} \\ \text{of students} \\ \text{in } B, W, \text{ and } A \end{pmatrix} \begin{pmatrix} \text{Segregation} \\ \text{between} \\ B \ \& \ \{W, A\} \end{pmatrix}
$$
$$
+ \begin{pmatrix} \text{Proportion} \\ \text{of students} \\ \text{in } W \text{ and } A \end{pmatrix} \begin{pmatrix} \text{Segregation} \\ \text{between} \\ W \ \& \ A \end{pmatrix} \quad (6)
$$

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

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

Total segregation among the four groups across schools in the U.S. is 0.651 (row 1, column 1). In panel B, all indices are re-expressed as percentages of this total. The most important source of school segregation is the ethnic differentiation of districts within CBSA's, which accounts for 32.3% of the total (row 9, column 1). A comparison of columns 2-4 of row 9 shows that this is mostly due to the separation of blacks from whites and Asians. Segregation

between the states is also important, accounting for 31.9% of total segregation (row 7, column 1). This is mainly due to the residential patterns of Hispanics (row 7, columns 2-4). Indeed, 51% of Hispanic public school students lived in Texas, California, or New Mexico in 2007/8, compared to only 14% of non-Hispanic students.

Using the Gini and Normalized Exposure indices, respectively, Rivkin [43] and Clotfelter [10] find that segregation between whites and blacks within cities is driven mainly by segregation between school districts. We reproduce this finding in rows 9 and 10 of column 3. However, by Strong School and Group Decomposability, the Mutual Information index can be decomposed across any number of geographic levels and ethnic groups simultaneously, using simple population weights. In contrast, Normalized Exposure is Additively Decomposable only in the two-group case; Gini is not even Aggregative (Table 1). As a result, a district's weight in Clotfelter [10] depends on its ethnic distribution, and Rivkin's [43] decomposition includes an enigmatic interaction term.[31]

RANK CORRELATIONS (KENDALL'S TAU-B), 2007/8									
INDEX	M	A	H	D	G	NE	C90	C50	CR
Mutual Information (M)	1	0.216	0.687	0.546	0.57	0.868	0.479	0.516	0.754
Symmetric Atkinson (A)	0.216	1	0.409	0.424	0.422	0.243	0.261	0.159	0.248
Entropy Index (H)	0.687	0.409	1	0.822	0.855	0.747	0.432	0.368	0.665
Dissimilarity (D)	0.546	0.424	0.822	1	0.913	0.617	0.348	0.262	0.563
Gini (G)	0.57	0.422	0.855	0.913	1	0.643	0.378	0.284	0.582
Normalized Exposure (NE)	0.868	0.243	0.747	0.617	0.643	1	0.467	0.496	0.784
Clotfelter (90% threshold) (C90)	0.479	0.261	0.432	0.348	0.378	0.467	1	0.617	0.457
Clotfelter (50% threshold) (C50)	0.516	0.159	0.368	0.262	0.284	0.496	0.617	1	0.495
Card-Rothstein (CR)	0.754	0.248	0.665	0.563	0.582	0.784	0.457	0.495	1
Mean (diagonal excluded)	0.515	0.265	0.554	0.499	0.516	0.541	0.382	0.355	0.505

Table 4: Kendall's Rank Correlation (τ_b) of CBSA Segregation Indices, 2007/8 School Year. $C50$ and $C90$ refer to Clotfelter index with threshold $\kappa = 0.5, 0.9$, 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, (non-Hispanic) whites, (non-Hispanic) blacks, and Hispanics.

Rank correlations of the major indices across CBSAs, using Kendall's τ_b, are shown in Table 4. The Mutual Information index is most correlated with Normalized Exposure index.

---

[31]See Reardon and Firebaugh [42, pp. 53-4].

The symmetric Atkinson index, the only index that is Composition Invariant, is only weakly correlated with the other indices.

Table 5 shows segregation indices for CBSA's with at least 200,000 public school students in K-12 districts. A city's rankings according to the Mutual Information and symmetric Atkinson indices appear in the first two columns. The entropy of the city's public school population appears in the second to last column and reflects the ethnic diversity of students in the city. Since the Mutual Information index cannot exceed this quantity, its 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 index. Cleveland and Detroit each has a high Mutual Information index despite its relatively low level of ethnic diversity.

7 Conclusion

In this paper we give an axiomatic foundation for multigroup segregation, based only on ordinal axioms. We first axiomatize the Atkinson segregation indices. These satisfy Composition Invariance: they depend not on the overall ethnic distribution of a district, but rather only on how each ethnic group is distributed across schools. These indices should be used only to compare districts with the same, fixed number of ethnic groups. In addition, they should not be used with highly disaggregated ethnic schemas, because of their sensitivity to zeroes. Empirically, segregation measured with the Atkinson indices shows a steep decline over the past twenty years, indicating a weakening of the effect of ethnicity on school assignment.

We also axiomatize the Mutual Information index. This index violates Composition Invariance but can be used to compare districts with different numbers of ethnic groups. As it is not sensitive to zeroes, it can be used with disaggregated data. The Mutual Information index has intuitive decompositions across locations and ethnic groups that make it suitable

Rank by M	Rank by A	Abbreviated CBSA Name	Segregation of Public Schools in CBSA, 2007/8								Entropy of CBSA	Public School Pop.
			M	A	H	D	G	NE	C50	CR		
1	2	Chicago	0.87	0.57	0.50	0.68	0.83	0.52	0.92	0.65	1.76	916875
4	1	Cleveland	0.62	0.57	0.51	0.73	0.87	0.58	0.86	0.67	1.21	286838
2	3	New York	0.72	0.49	0.38	0.59	0.75	0.40	0.86	0.61	1.87	2380186
3	4	Milwaukee	0.68	0.47	0.45	0.67	0.80	0.52	0.86	0.63	1.52	217910
6	7	Los Angeles	0.54	0.44	0.34	0.57	0.73	0.36	0.86	0.45	1.58	1598106
5	9	Philadelphia	0.57	0.43	0.37	0.59	0.75	0.43	0.74	0.54	1.53	698142
7	14	Atlanta	0.53	0.38	0.33	0.55	0.71	0.37	0.79	0.46	1.62	876725
13	10	Memphis	0.48	0.41	0.39	0.65	0.79	0.45	0.86	0.50	1.24	230947
11	13	Denver	0.50	0.41	0.44	0.65	0.81	0.53	0.72	0.60	1.12	387607
17	8	Detroit	0.45	0.43	0.39	0.58	0.75	0.43	0.63	0.54	1.15	586073
8	18	Washington	0.52	0.34	0.29	0.51	0.66	0.32	0.78	0.47	1.80	808730
22	5	Cincinnati	0.37	0.45	0.41	0.64	0.81	0.46	0.66	0.55	0.89	307831
16	12	Boston	0.47	0.41	0.35	0.59	0.74	0.39	0.64	0.47	1.34	539659
9	19	Miami	0.52	0.34	0.31	0.53	0.70	0.35	0.92	0.34	1.66	757016
14	15	Houston	0.47	0.38	0.27	0.48	0.64	0.28	0.82	0.38	1.75	1098990
19	11	Indianapolis	0.39	0.41	0.34	0.60	0.75	0.39	0.67	0.48	1.17	272227
10	20	Baltimore	0.50	0.33	0.36	0.58	0.75	0.44	0.75	0.52	1.42	382240
15	17	Dallas	0.47	0.35	0.27	0.49	0.65	0.30	0.75	0.38	1.73	1131521
12	21	San Francisco	0.49	0.33	0.25	0.46	0.61	0.25	0.73	0.35	1.95	398607
21	16	Columbus	0.38	0.37	0.34	0.56	0.74	0.39	0.58	0.48	1.10	265691
34	6	Pittsburgh	0.26	0.45	0.36	0.62	0.77	0.37	0.54	0.43	0.72	307888
20	22	Kansas City	0.39	0.33	0.31	0.53	0.68	0.37	0.64	0.47	1.26	321552
18	24	St. Louis	0.41	0.32	0.28	0.53	0.67	0.33	0.65	0.36	1.48	397309
25	23	Providence	0.35	0.32	0.30	0.57	0.72	0.38	0.50	0.39	1.16	213789
23	26	Nashville	0.36	0.30	0.28	0.54	0.69	0.33	0.64	0.40	1.28	219474
26	25	Austin	0.35	0.31	0.22	0.45	0.59	0.25	0.70	0.30	1.59	270394
24	28	San Diego	0.35	0.27	0.20	0.41	0.56	0.21	0.70	0.26	1.75	277529
27	29	Charlotte	0.34	0.26	0.22	0.46	0.60	0.27	0.64	0.35	1.53	265131
30	27	San Antonio	0.31	0.30	0.23	0.48	0.63	0.25	0.84	0.22	1.32	375797
28	33	Sacramento	0.31	0.22	0.17	0.39	0.52	0.18	0.45	0.22	1.85	259900
32	30	Minneapolis	0.29	0.26	0.23	0.47	0.62	0.28	0.35	0.29	1.26	500762
31	31	Tampa	0.29	0.23	0.19	0.41	0.55	0.21	0.60	0.28	1.53	369709
29	34	Orlando	0.31	0.20	0.18	0.39	0.53	0.20	0.69	0.25	1.70	319561
38	32	Portland, OR	0.19	0.22	0.15	0.37	0.51	0.15	0.25	0.20	1.25	315225
33	37	Virginia Beach	0.26	0.18	0.19	0.42	0.57	0.24	0.72	0.29	1.40	266285
36	35	Riverside	0.24	0.20	0.16	0.38	0.52	0.18	0.81	0.21	1.50	675464
37	36	Seattle	0.22	0.19	0.15	0.36	0.50	0.16	0.23	0.20	1.47	469854
35	39	Las Vegas	0.25	0.16	0.14	0.36	0.48	0.17	0.62	0.19	1.78	306603
39	38	Salt Lake City	0.19	0.16	0.17	0.43	0.56	0.21	0.21	0.16	1.15	210381
40	40	Phoenix	0.18	0.15	0.14	0.35	0.49	0.18	0.24	0.13	1.32	357194

Table 5: Segregation of Public Schools Within CBSA's, 2007/8 School Year. $C50$ refers to Clotfelter index with threshold $\kappa = 0.5$. 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, (non-Hispanic) whites, (non-Hispanic) blacks, and Hispanics.

for studying the sources of segregation in cross-section. Using these decompositions, we show that public school segregation in the U.S. is driven primarily by differences in the ethnic composition of districts within cities, and by segregation across U.S. states, where the latter is due mainly to the distinct residential patterns of Hispanics.

Appendix A. Proofs

Proof of Claim 1

For part 1, let $X, Y \in \mathcal{C}$ have equal populations and equal group distributions. Then $T(X) = T(Y)$ and for any $Z \in \mathcal{C}$, $P(X \uplus Z) = P(Y \uplus Z)$, so

$$S(X \uplus Z) \geq S(Y \uplus Z) \iff \sum_{n \in \mathbf{N}(X) \cup \mathbf{N}(Z)} T^n F_2(p^n) \geq \sum_{n \in \mathbf{N}(Y) \cup \mathbf{N}(Z)} T^n F_2(p^n) \iff S(X) \geq S(Y)$$

As for part 2, let X' be the result of partitioning some school $n \in \mathbf{G}(X)$ into two schools, n_1 and n_2. Then $S(X') - S(X)$ equals $\pi^{n_1} F_3(P, p^{n_1}) + \pi^{n_2} F_3(P, p^{n_2}) - \pi^n F_3(P, p^n)$, which is nonnegative as F_3 is convex. If (say) n_1 is empty, then $\pi^{n_1} = 0$, $\pi^{n_2} = \pi^n$, and $p^{n_2} = p^n$. If both schools have the same ethnic distribution, then $p^{n_1} = p^{n_2} = p^n$. In both cases, $S(X') - S(X) = 0$. Part 3 holds since rescaling a group has no effect on t. As for part 4, let $X \in \mathcal{C}$ be a district in which the set of ethnic groups is $\mathbf{G}$. Let X' be the result of partitioning some ethnic group $g \in \mathbf{G}$ into two ethnic groups, g_1 and g_2. Then

$$S(X') - S(X) = P_{g_1} F_4(\pi, r_{g_1}) + P_{g_2} F_4(\pi, r_{g_2}) - P_g F_4(\pi, r_g).$$

If (say) g_1 is empty, then $P_{g_1} = 0$ and $P_{g_2} = P_g$ and $r_{g_2} = r_g$. If both ethnic groups have the same distribution across schools, then $r_{g_1} = r_{g_2} = r_g$. In both cases, $S(X') - S(X) = 0$.

Notation and Auxiliary lemmas

Before we prove our main results, we need some notation. We say that a school is a *ghetto school* if all its students belong to the same group. Let $\underline{X}^K = \left\langle \left(\underbrace{1, 1, \ldots, 1}_{K \text{ groups}} \right) \right\rangle$ be

the district with K groups of unit size who all attend the same school. Let $\overline{X}^K = \left\langle \left(\underbrace{1,0,\ldots,0}_{K\text{ groups}}\right), \left(\underbrace{0,1,0,\ldots,0}_{K\text{ groups}}\right), \ldots, \left(\underbrace{0,\ldots,0,1}_{K\text{ groups}}\right)\right\rangle$ be the district with K ghetto schools, each of unit size. When the number of groups is clear from the context, we write $\overline{X}$ and $\underline{X}$, respectively. For any district X and any vector of nonnegative scalars $\boldsymbol{\alpha} = (\alpha_g)_{g\in\mathbf{G}}$, let $\boldsymbol{\alpha} * X$ denote the district in which the number of members of group g in school n is $\alpha_g T_g^n$. For example, if $X = \langle (1,2), (3,4)\rangle$, and $\boldsymbol{\alpha} = (2,3)$, then $\boldsymbol{\alpha} * X = \langle (2,6), (6,12)\rangle$. We will also apply this operation to individual schools; e.g., $\boldsymbol{\alpha} * (1,2) = (2,6)$. For any K-vector $\mathbf{t} = (t_1,\ldots,t_K) \in [0,1]^K$, let $X(\mathbf{t})$ denote the district $\mathbf{t} * \underline{X} \uplus (\mathbf{1}-\mathbf{t}) * \overline{X}$ where $\mathbf{1}$ denotes a K-vector of ones. City $X(\mathbf{t})$ consists of the mixed school $\mathbf{t}$, containing t_g students from each group g, and for each group g a ghetto school that contains $1 - t_g$ students of group g. For any scalar α, let $X(\alpha)$ denote the district $X(\alpha\mathbf{1})$ that contains one school with α students of each group and K ghetto schools, each with $1 - \alpha$ students.

We first prove some preliminary lemmas.

Lemma 1 *Let $\succcurlyeq$ be a segregation ordering that satisfies CONT. For any districts $X, Y, Z \in \mathcal{C}$, the sets $A = \{c \in [0,1] : cX \uplus (1-c)Y \succcurlyeq Z\}$ and $B = \{c \in [0,1] : Z \succcurlyeq cX \uplus (1-c)Y\}$ are closed.*

Proof. Let $\{c_k\}$ be a sequence of elements of A that converges to c. Then, $c_k X \uplus (1-c_k)Y$ is a sequence of districts in $\{X \in \mathcal{C}(\mathbf{G}, \mathbf{N}) : X \succcurlyeq Z\}$ that converges to $cX \uplus (1-c)Y$ (where $\mathbf{G}$ and N are the group and school sets of $cX \uplus (1-c)Y$). Since $\{X \in \mathcal{C}(\mathbf{G}, \mathbf{N}) : X \succcurlyeq Z\}$ is closed, $cX \uplus (1-c)Y \succcurlyeq Z$, so $c \in A$. The argument for B is analogous. Q.E.D.

Lemma 2 *Let $\succcurlyeq$ be either a segregation ordering on $\mathcal{C} = \mathcal{C}_K$ that satisfies SDP and CI or a segregation ordering on $\mathcal{C} = \mathcal{C}^A$ that satisfies SDP, SI and GDP.*

1. *All districts in which every school is representative are equally segregated under $\succcurlyeq$.*

2. *Any district in which every school is representative is weakly less segregated under $\succcurlyeq$ than any district in which some school is unrepresentative.*

34

Proof. 1. Consider any district Y in which every school is representative. Number the schools $1, ..., N$. For each $i = 1, ..., N$, let Y_i be the district that results from Y when the first i schools of Y are combined into a single school. By SDP, for each $i = 1, ..., N - 1$, $Y_i \sim Y_{i+1}$. Hence, by transitivity, $Y = Y_1 \sim Y_{\mathbf{N}}$. Y_N contains a single school. If $\mathcal{C} = \mathcal{C}_K$ and CI holds, then $Y_N \sim \underline{X}^K$ and hence $Y \sim \underline{X}^K$. If $\mathcal{C} = \mathcal{C}^A$, and SI and GDP hold, then $Y_N \sim \underline{X}^1$ and hence $Y \sim \underline{X}^1$.

2. Let Y be a district in which every school is representative and consider any district X in which at least one school is unrepresentative. The above reasoning yields $X \succcurlyeq X_N$, where X_N is the result of combining the students of X into a single school. By part 1, $X_N \sim Y$. Therefore, $X \succcurlyeq Y$. Q.E.D.

Lemma 3 *Let $\succcurlyeq$ be a segregation ordering on $\mathcal{C}$ that satisfies SI, IND, and SDP. 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 SI, $(\alpha - \beta)X \succ (\alpha - \beta)X'$. By IND,

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

The result then follows from SDP. Q.E.D.

Lemma 4 *Let $\succcurlyeq$ be a segregation ordering on $\mathcal{C}$ that satisfies SI, IND, SDP, and CONT. 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. By Lemma 1, $\{\alpha \in [0, 1] : \alpha Z \uplus (1 - \alpha)Y \succcurlyeq X\}$ and $\{\alpha \in [0, 1] : X \succcurlyeq \alpha Z \uplus (1 - \alpha)Y\}$ are closed sets. 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 sets are each nonempty as they contain 1 and 0, respectively, by SDP. Their union is the whole unit interval since $\succcurlyeq$ is complete. Since the interval $[0, 1]$ is connected, the intersection of the two sets must be nonempty. By Lemma 3, their intersection cannot contain more than one element. Q.E.D.

Proof of Theorem 1

The index $A_{\mathbf{w}}$ clearly satisfies N. It satisfies CONT since $A_{\mathbf{w}}$ is a continuous function. As shown in section 3.1, it also satisfies IND, SDP, and CI.

Now let $\succcurlyeq$ be a segregation ordering on $\mathcal{C}_K$ that satisfies IND, SDP, N, CI, and CONT. We will show that it must be represented by an index of the form $A_{\mathbf{w}}(X)$ by a series of lemmas.

Lemma 5 *All completely segregated districts have the same degree of segregation under $\succcurlyeq$, and they are weakly more segregated than any district in which any school is mixed.*

Proof. Consider a completely segregated district $X \in \mathcal{C}_K$. Let X' be the district that results from X when, for each group $g \in G$, all schools that contain only members of group g are combined into a single school. (X' thus consists of K schools, each of which contains all the members of a single group.) By iteratively applying SDP, $X \sim X'$. By CI, X' is as segregated as any other district that consists of K schools, each of which contains all the members of a single group. This implies that all completely segregated districts have the same degree of segregation. Now any district that has at least one mixed school can be converted into a completely segregated district by dividing each school n into K distinct schools, each of which includes all and only the members of a single group. By SDP, this procedure results in a weakly more segregated district. Q.E.D.

Lemma 6 *1. $\overline{X} \succ \underline{X}$. 2. For any $\alpha, \beta \in [0,1]$ such that $\alpha > \beta$, $X(\beta) \succ X(\alpha)$.*

Proof. By N, there exist districts X and Y such that $X \succ Y$. By lemmas 2 and 5, $\overline{X} \succcurlyeq X \succ Y \succcurlyeq \underline{X}$, so $\overline{X} \succ \underline{X}$. Part 2 then follows from Lemma 3. Q.E.D.

Lemma 7 *For any district X, there is a unique $\alpha_X \in [0,1]$ such that $X \sim X(\alpha_X)$.*

Proof. Follows from Lemma 4 and Lemma 6, part 1. Q.E.D.

Let the index $S : \mathcal{C} \to [0,1]$ be defined by $S(X) = 1 - \alpha_X$. By Lemma 6, $X \succcurlyeq Y$ if and only if $1 - \alpha_X > 1 - \alpha_Y$, so S represents $\succcurlyeq$. It remains to show that S equals $A_{\mathbf{w}}$ for some vector $\mathbf{w}$ of weights.

Lemma 8 *Let $\mathbf{t} = (\mathbf{t_1}, ..., \mathbf{t_K})$ and $\mathbf{v}$ be two vectors in $[0,1]^K$, such that $\mathbf{t} \leq \mathbf{v}$. Then, $X(\mathbf{t}) \succcurlyeq X(\mathbf{v})$. If $\mathbf{t} \in (0,1)^K$ then $\overline{X} \succ X(\mathbf{t}) \succ \underline{X}$.*

Proof. Applying SDP twice, we obtain

$$X(\mathbf{t}) = \mathbf{t} * \underline{X} \uplus (\mathbf{1} - \mathbf{t}) * \overline{X} \sim \mathbf{t} * \underline{X} \uplus (\mathbf{v} - \mathbf{t}) * \overline{X} \uplus (\mathbf{1} - \mathbf{v}) * \overline{X}$$

$$\succcurlyeq \mathbf{v} * \underline{X} \uplus (\mathbf{1} - \mathbf{v}) * \overline{X} = X(\mathbf{v}).$$

Assume now that $\mathbf{t} \in (0,1)^K$, and let $\overline{t} = \max\{t_1, ..., t_K\}$, $\underline{t} = \min\{t_1, ..., t_K\}$. Then $\overline{X} \succ X(\underline{t}) \succcurlyeq X(\mathbf{t}) \succcurlyeq X(\overline{t}) \succ \underline{X}$ by Lemma 3 since $0 < \underline{t} \leq \overline{t} < 1$. Q.E.D.

Lemma 9 *For any two vectors $\mathbf{t}, \mathbf{v} \in [0,1]^K$ and for any $\gamma \in (0,1]$,*

1. $\mathbf{v} * X(\mathbf{t}) \uplus (\mathbf{1} - \mathbf{v}) * \overline{X} \sim X(\mathbf{v} * \mathbf{t})$

2. If for some $\alpha \in [0,1]$, $X(\mathbf{t}) \sim X(\alpha)$, then $X(\mathbf{v} * \mathbf{t}) \sim X(\alpha \mathbf{v})$.

Proof. 1. By definition of $X(\mathbf{t})$, $\mathbf{v} * X(\mathbf{t}) \uplus (\mathbf{1} - \mathbf{v}) * \overline{X} = \mathbf{v} * (\mathbf{t} * \underline{X} \uplus (\mathbf{1} - \mathbf{t}) * \overline{X}) \uplus (\mathbf{1} - \mathbf{v}) * \overline{X}$, which by SDP is as segregated as $(\mathbf{v} * \mathbf{t}) * \underline{X} \uplus (\mathbf{1} - \mathbf{v} * \mathbf{t}) * \overline{X} = X(\mathbf{v} * \mathbf{t})$.

2. By CI and IND, $\mathbf{v} * X(\mathbf{t}) \uplus (\mathbf{1} - \mathbf{v}) * \overline{X} \sim \mathbf{v} * X(\alpha) \uplus (\mathbf{1} - \mathbf{v}) * \overline{X}$, which, by the previous steps, implies $X(\mathbf{v} * \mathbf{t}) \sim X(\alpha \mathbf{v})$. Q.E.D.

For any group g and scalar β, let $\mathbf{1}_g(\beta)$ denote the K-vector with β in the gth place and ones elsewhere.

Lemma 10 *For each group g there is a fixed constant $w_g \geq 0$ such that for any $\beta \in (0,1]$,*

$$X(\mathbf{1}_g(\beta)) \sim X(\beta^{w_g}).$$

Proof. By Lemma 7, for each $u \geq 0$ there is a unique scalar $f(u) \in (0,1]$ defined implicitly by $X(\mathbf{1}_g(e^{-u})) \sim X(f(u))$. Let $u, v \in \Re_+$. By Lemma 9, $X(\mathbf{1}_g(e^{-u}) * \mathbf{1}_g(e^{-v})) \sim X(f(v)\mathbf{1}_g(e^{-u})) \sim X(f(u)f(v))$. Therefore, f satisfies the functional equation

$$f(u + v) = f(u)f(v) \text{ for all } u, v \geq 0. \tag{7}$$

37

Further, by Lemma 8, if $u > v$, then $X(f(u)) \sim X(\mathbf{1}_g(e^{-u})) \succcurlyeq X(\mathbf{1}_g(e^{-v})) \sim X(f(v))$, which by Lemma 6 implies that $f(u) \leq f(v)$. Therefore, f is nonincreasing, so it is continuous at at least one point. Thus, by Theorem 1 in Aczél [1, pp. 38-39], either (a) f is identically zero, or (b) $f(0) = 1$ and, for all $u > 0$, $f(u) = 0$, or (c) there is a $w_g \geq 0$ such that $f(u) = e^{-w_g u}$. The function f cannot be identically zero because then $X(1) \sim X(\mathbf{1}_g(1)) \sim X(f(0)) = X(0)$, which contradicts Lemma 6. Further, $f(u)$ cannot equal zero for $u > 0$, because, by Lemma 8, $f(u) \geq e^{-u}$. Hence, since f is nonincreasing, there must be a $w_g \geq 0$ such that $f(u) = e^{-w_g u}$. But then by definition of f, $X(\mathbf{1}_g(e^{-u})) \sim X(f(u)) = X(e^{-w_g u})$. The claim follows by setting $\beta = e^{-u}$. Q.E.D.

Lemma 11 *There are fixed, non-negative weights $w_g \geq 0$ for $g = 1, ..., K$ such that for any $\mathbf{t} = (t_1, \dots, t_K) \in [0,1]^K$ the unique $\alpha \in [0,1]$ that satisfies $X(\mathbf{t}) \sim X(\alpha)$ is given by $\prod_{g=1}^{K} (t_g)^{w_g}$. Further, the weights add up to one.*

Proof. Assume first that $\mathbf{t} \in (0,1]^K$. By Lemma 10, $X(\mathbf{1}_g(t_g)) \sim X(t_g^{w_g})$ for all $g = 1, \dots K$. Note that $\mathbf{t} = \mathbf{1}_1(t_1) * \mathbf{1}_2(t_2) * \dots * \mathbf{1}_K(t_K)$. Repeated application of Lemma 9 yields $X(\mathbf{t}) = X\left(\prod_{g=1}^{K} t_g^{w_g}\right)$. In order to complete the proof we need to show that the weights w_g add up to one. Consider the district $X(\alpha)$ where $\alpha \in (0,1)$. By the previous conclusion $X(\alpha) \sim X\left(\prod_{g=1}^{K} \alpha^{w_g}\right)$. By Lemma 6, $\prod_{g=1}^{K} \alpha^{w_g} = \alpha$, so the weights w_g add up to one. Assume now that $\mathbf{t} \in [0,1]^K \backslash (0,1]^K$. By Lemma 7, there is an $\alpha \in [0,1]$ such that $X(\mathbf{t}) \sim X(\alpha)$. We need to show that $\alpha = 0$. Let $\mathbf{t}(\varepsilon) = (t_1(\varepsilon), ..., t_K(\varepsilon))$ be the school that results from $\mathbf{t}$ after replacing the 0 components by $\varepsilon > 0$. Since $\mathbf{t}(\varepsilon) \in (0,1]^K$, by the previous argument $X(\mathbf{t}(\varepsilon)) \sim X(\alpha(\varepsilon))$ where $\alpha(\varepsilon) = \prod_{g=1}^{K} t_g(\varepsilon)^{w_g}$. By Lemma 8, $X(\mathbf{t}) \succcurlyeq X(\mathbf{t}(\varepsilon))$, which implies $X(\alpha) \succcurlyeq X(\alpha(\varepsilon))$. Hence, by Lemma 6, $\alpha(\epsilon) \geq \alpha \geq 0$. Since $\alpha(\varepsilon) \to 0$ as $\varepsilon \to 0$, we obtain that $\alpha = 0$. Q.E.D.

Lemma 12 *For every district $X \in \mathcal{C}$ there is a unique $\alpha_X \in [0,1]$ such that $X \sim \alpha_X \underline{X} \uplus (1 - \alpha_X)\overline{X}$. Further, this unique α_X is $\sum_{n \in \mathbf{N}(X)} \prod_{g=1}^{K} (t_g^n)^{w_g}$, where the weights w_g are those found in Lemma 11.*

Proof. By CI it is enough to prove the statement for districts where all groups are of unit measure. Also, by SDP we can restrict attention to districts where for each group there is at most one ghetto. So let $X = \left(\biguplus_{n=1}^{r} (\mathbf{t}^n * \underline{X})\right) \uplus \left(\mathbf{1} - \sum_{n=1}^{r} \mathbf{t}^n\right) * \overline{X}$ be a district with r non-ghetto schools and where all groups are of unit measure. The proof is by induction on the number r of non-ghetto schools. By Lemma 11, the statement of the theorem holds for $r = 1$. Assume that the statement is true for all districts with $m - 1$ non-ghetto schools, and let $r = m$. There are two cases.

Case 1: there is at least one non-ghetto school n such that $t_g^n > 0$ for all $g = 1, \dots, K$. Assume without loss of generality that the totally mixed school is school r, and that $\prod_{g=1}^{K} \left(t_g^n\right)^{w_g} \leq \prod_{g=1}^{K} \left(t_g^r\right)^{w_g}$ for $n = 1, 2, \dots, r - 1$. Assume first that $t_g^n \leq 1/m$ for all $n = 1, 2, \dots, m$ and $g = 1, \dots, K$. Define $\widetilde{t_g^n} = t_g^n/(1 - t_g^m)$ for $g = 1, ..., K$ and $n = 1, 2, \dots, m$. Note that $\widetilde{t_g^n} \leq 1/(m - 1)$ and that $\prod_{g=1}^{K} \left(\widetilde{t_g^n}\right)^{w_g} \leq \prod_{g=1}^{K} \left(\widetilde{t_g^m}\right)^{w_g}$ for $n = 1, 2, \dots, m - 1$. Define $\tau_n = \prod_{g=1}^{K} \left(t_g^n/t_g^m\right)^{w_g} = \prod_{g=1}^{K} \left(\widetilde{t_g^n}/\widetilde{t_g^m}\right)^{w_g} \leq 1$ for $n = 1, 2, \dots, m - 1$. We can write $X = \biguplus_{n=1}^{m-1} (\mathbf{t}^n * \underline{X}) \uplus \left(\mathbf{1} - \sum_{n=1}^{m} \mathbf{t}^n\right) * \overline{X} \uplus (\mathbf{t}^m * \underline{X})$. Let $\widetilde{\mathbf{t}^n} = \left(\widetilde{t_g^n}\right)_{g=1}^{K}$. By CI,

$$X \sim Y \uplus \left(\widetilde{\mathbf{t}^m} * \underline{X}\right) \tag{8}$$

where $Y = \biguplus_{n=1}^{m-1} \left(\widetilde{\mathbf{t}^n} * \underline{X}\right) \uplus \left(\mathbf{1} - \sum_{n=1}^{m-1} \widetilde{\mathbf{t}^n}\right) * \overline{X}$.[32] District Y has $m - 1$ non-ghetto schools. Consequently, by the induction hypothesis,

$$Y \sim \alpha_Y \underline{X} \uplus (1 - \alpha_Y) \overline{X}. \tag{9}$$

where $\alpha_Y = \sum_{n=1}^{m-1} \prod_{g=1}^{K} \left(\widetilde{t_g^n}\right)^{w_g}$. Define

$$Y' = \biguplus_{n=1}^{m-1} \left(\tau_n \widetilde{\mathbf{t}^m} * \underline{X}\right) \uplus \left(\mathbf{1} - \sum_{n=1}^{m-1} \tau_n \widetilde{\mathbf{t}^m}\right) * \overline{X}. \tag{10}$$

All entries in Y' are nonnegative since $\tau_n \leq 1$ for all n and since $\widetilde{t_g^m} \leq 1/(m - 1)$ for all g. As $\sum_{n=1}^{m-1} \prod_{g=1}^{K} \left(\tau_n \widetilde{t_g^m}\right)^{w_g} = \sum_{n=1}^{m-1} \prod_{g=1}^{K} \left(\widetilde{t_g^n}\right)^{w_g} = \alpha_Y$, Lemma 11 implies that

$$Y' \sim \alpha_Y \underline{X} \uplus (1 - \alpha_Y) \overline{X}. \tag{11}$$

---

[32]Y has no negative entries since $0 \leq \widetilde{t_g^n} \leq 1/(m - 1)$ for all g and n.

It follows from (9) and (11) that $Y \sim Y'$. Consequently,

$$
\begin{aligned}
X \;\sim\; & Y \uplus \left(\widetilde{\mathbf{t}^m} * \underline{X} \right) && \text{by (8)} \\
\sim\; & Y' \uplus \left(\widetilde{\mathbf{t}^m} * \underline{X} \right) && \text{by IND} \\
\sim\; & \biguplus_{n=1}^{m-1} \left(\tau_n \widetilde{\mathbf{t}^m} * \underline{X} \right) \uplus \left(\mathbf{1} - \sum_{n=1}^{m-1} \tau_n \widetilde{\mathbf{t}^m} \right) * \overline{X} \uplus \left(\widetilde{\mathbf{t}^m} * \underline{X} \right) && \text{by (10)} \\
\sim\; & \left(\sum_{n=1}^{m-1} \tau_n + 1 \right) \widetilde{\mathbf{t}^m} * \underline{X} \uplus \left(\mathbf{1} - \sum_{n=1}^{m-1} \tau_n \widetilde{\mathbf{t}^m} \right) * \overline{X} && \text{by SDP} \\
\sim\; & \left(\sum_{n=1}^{m-1} \tau_n + 1 \right) \mathbf{t}^m * \underline{X} \uplus \left(\mathbf{1} - \left(1 + \sum_{n=1}^{m-1} \tau_n \right) \mathbf{t}^m \right) * \overline{X} && \text{by CI and definition of } \widetilde{\mathbf{t}^m}.
\end{aligned}
$$

Therefore, using Lemma 11, $X \sim \alpha_X \underline{X} \uplus (1 - \alpha_X) \overline{X}$, where

$$
\alpha_X = \left(\sum_{n=1}^{m-1} \tau_n + 1 \right) \prod_{g=1}^{K} \left(t_g^m \right)^{w_g} = \sum_{n=1}^{m-1} \prod_{g=1}^{K} \left(t_g^n \right)^{w_g} + \prod_{g=1}^{K} \left(t_g^m \right)^{w_g}.
$$

Consider now the case where $1/m < t_g^n \leq 1$ for some $n = 1, 2, \dots, m$ and $g = 1, \dots, K$. Define $\widehat{\mathbf{t}}^n = \frac{1}{m} \mathbf{t}^n$ for $n = 1, 2, \dots, m$. Let $\widehat{X} = \biguplus_{n=1}^{m} \left(\widehat{\mathbf{t}}^n * \underline{X} \right) \uplus \left(\mathbf{1} - \sum_{n=1}^{m} \widehat{\mathbf{t}}^n \right) * \overline{X}$. Each entry in each vector $\widehat{\mathbf{t}}^n$ is at most $1/m$. By the preceding argument, there is a unique $\widehat{\alpha}_X \in [0, 1]$ such that $\widehat{X} \sim \widehat{\alpha}_X \underline{X} \uplus (1 - \widehat{\alpha}_X) \overline{X}$ and this unique $\widehat{\alpha}_X$ is $\sum_{n=1}^{m} \prod_{g=1}^{K} \left(\widehat{t}_g^n \right)^{w_g}$. By SDP, $\widehat{X} \sim \frac{1}{m} X \uplus (1 - \frac{1}{m}) \overline{X}$. Therefore,

$$
\begin{aligned}
\frac{1}{m} X \uplus (1 - \frac{1}{m}) \overline{X} &\sim \widehat{\alpha}_X \underline{X} \uplus (1 - \widehat{\alpha}_X) \overline{X} \sim \frac{1}{m} \left(m \widehat{\alpha}_X \right) \underline{X} \uplus \left(1 - \frac{1}{m} \left(m \widehat{\alpha}_X \right) \right) \overline{X} \\
&\sim \frac{1}{m} \left(m \widehat{\alpha}_X \right) \underline{X} \uplus \frac{1}{m} (1 - (m \widehat{\alpha}_X)) \overline{X} \uplus (1 - \frac{1}{m}) \overline{X}
\end{aligned}
$$

by SDP. Consequently, by IND and CI, $X \sim (m \widehat{\alpha}_X) \underline{X} \uplus (1 - (m \widehat{\alpha}_X)) \overline{X}$, so the unique α_X that we are looking for is $\alpha_X = m \widehat{\alpha}_X = \sum_{n=1}^{m} \prod_{g=1}^{K} \left(t_g^n \right)^{w_g}$.

Case 2: for every non-ghetto school n there is a group g such that $t_g^n = 0$. By Lemma 7 there is an $\alpha \in [0, 1]$ such that $X \sim X(\alpha)$. We need to show that $\alpha = 0$. For any $\varepsilon \in (0, 1)$, let $X(m, \varepsilon)$ be the district that is obtained from X by transferring to school m, a representative proportion ε of the students of the other schools. Formally,

$$
X(m, \varepsilon) = [(\mathbf{t}^m + \varepsilon (\mathbf{1} - \mathbf{t}^m)) * \underline{X}] \uplus (1 - \varepsilon) \left[\biguplus_{n=1}^{m-1} (\mathbf{t}^n * \underline{X}) \uplus \left(\mathbf{1} - \sum_{n=1}^{r} \mathbf{t}^n \right) * \overline{X} \right].
$$

By SDP, applied twice, $X \succsim X(m, \varepsilon)$. By Lemma 7 there is an $\alpha(\varepsilon) \in [0, 1]$ such that $X(m, \varepsilon) \sim X(\alpha(\varepsilon))$. Consequently, $X(\alpha) \succsim X(\alpha(\varepsilon))$, which by Lemma 6 implies that

$\alpha(\varepsilon) \geq \alpha \geq 0$. But the district $X(m, \varepsilon)$ has school m with $t_g^m + \varepsilon(1 - t_g^m) > 0$ students of group g, while all other schools n have $t_g^n = 0$ for some g. So, by Case 1, $\alpha(\varepsilon) = \prod_{g=1}^{K} \left(t_g^m + \varepsilon(1 - t_g^m) \right)$. Since $\alpha(\varepsilon) \to \prod_{g=1}^{K} t_g^m = 0$ as $\varepsilon \to 0$, we obtain that $\alpha = 0$. Q.E.D.

This completes the proof of Theorem 1.

Proof of Theorem 2

The Mutual Information index M clearly satisfies N, SYM, and SI. It is a continuous function of the T_g^n's and thus satisfies CONT. As shown in section 3.1, it satisfies IND, SDP, and GDP.

Now let $\succsim$ be a segregation ordering that satisfies SI, IND, SDP, N, GDP, SYM, and CONT on $\mathcal{C}^A$. We will show that $\succsim$ is represented by the Mutual Information Index. For any district X, let the groups be numbered $g = 1, \dots, K$. For any ethnic distribution $P = (P_g)_{g=1}^{K}$, let $\overline{X}(P)$ denote a district consisting of K ghetto schools, where school $g = 1, ..., K$ contains P_g members of each group g. Let $\underline{X}(P)$ denote the one-school district with ethnic distribution P and unit population. That is, $\overline{X}(P) = \langle (P_1, 0, ..., 0), ...(0, ..., 0, P_K) \rangle$ and $\underline{X}(P) = \langle (P_1, ..., P_K) \rangle$. If X has group distribution P, then by SDP and SI, $\overline{X}(P) \succsim X$.

For any two distributions $P = (P_1, ..., P_K)$ and $P' = (P'_1, ..., P'_{K'})$, let $P \times P'$ denote the distribution $\left((P_g P'_{g'})_{g=1}^{K} \right)_{g'=1}^{K'}$. By SDP, $\overline{X}(P \times P')$ is at least as segregated as both $\overline{X}(P)$ and $\overline{X}(P')$.

Let X be a district with K' groups and ethnic distribution $P' = (P'_1, ..., P'_{K'})$. For any $K \geq 1$ and any distribution $P = (P_1, ..., P_K)$ let $\phi^P(X)$ be the district that results after splitting each ethnic group g in district X into K ethnic groups in proportions given by P, such that each of these K ethnic groups has the same distribution across schools. That is, the T_g^n members of each ethnic group g in each school n of X are split up into K ethnic groups of size $P_1 T_g^n, ..., P_K T_g^n$. District $\phi^P(X)$ has group distribution $P \times P'$.

By Nontriviality there exist districts $X_1 \succ X_0$. Let $P(X_1)$ be the group distribution of district X_1.[33] By GDP and SI, $\overline{X}(P(X_1)) \succsim X_1$. Let X be an arbitrary district and

---

[33] From now on, we fix X_1 and its distribution $P(X_1)$. We will show that the results are independent of

let $\widehat{P} = (\widehat{P}_1, ..., \widehat{P}_K)$ be a distribution such that $\overline{X}(\widehat{P}) \succcurlyeq X$ and $\overline{X}(\widehat{P}) \succcurlyeq \overline{X}(P(X_1))$. (For instance, let $\widehat{P} = P \times P(X_1)$, where P is the group distribution of X). By Lemma 4 there is a unique $\widehat{\alpha}$ such that

$$X \sim \widehat{\alpha}\overline{X}(\widehat{P}) \uplus (1 - \widehat{\alpha})\underline{X}(\widehat{P}). \tag{12}$$

and a unique $\widehat{\beta} \in [0, 1]$ such that $\overline{X}(P(X_1)) \sim \widehat{\beta}\overline{X}(\widehat{P}) \uplus (1 - \widehat{\beta})\underline{X}(\widehat{P})$. Moreover, $\overline{X}(P(X_1)) \succ X_0 \succcurlyeq \underline{X}(\widehat{P})$ by Lemma 2, so $\widehat{\beta} > 0$. Define the index $S : \mathcal{C} \to \Re$ by[34]

$$S(X) = h(P(X_1)) \frac{\widehat{\alpha}}{\widehat{\beta}} \tag{13}$$

We now verify that $\widehat{\alpha}/\widehat{\beta}$ does not depend on the particular choice of $\widehat{P}$, so S is well defined. Consider another distribution $\widetilde{P} = (\widetilde{P}_1, ..., \widetilde{P}_{K'})$ such that $\overline{X}(\widetilde{P}) \succcurlyeq X$ and $\overline{X}(\widetilde{P}) \succcurlyeq \overline{X}(P(X_1))$. Let $\widetilde{\alpha}$ uniquely satisfy $X \sim \widetilde{\alpha}\overline{X}(\widetilde{P}) \uplus (1 - \widetilde{\alpha})\underline{X}(\widetilde{P})$ and let $\widetilde{\beta} > 0$ uniquely satisfy $\overline{X}(P(X_1)) \sim \widetilde{\beta}\overline{X}(\widetilde{P}) \uplus (1 - \widetilde{\beta})\underline{X}(\widetilde{P})$. By GDP,

$$X \sim \widetilde{\alpha}\phi^{\widehat{P}}\left(\overline{X}(\widetilde{P})\right) \uplus (1 - \widetilde{\alpha})\phi^{\widehat{P}}\left(\underline{X}(\widetilde{P})\right) \tag{14}$$

Similarly, applying the transformation $\phi^{\widetilde{P}}$ to (12) and using GDP,

$$X \sim \widehat{\alpha}\phi^{\widetilde{P}}\left(\overline{X}(\widehat{P})\right) \uplus (1 - \widehat{\alpha})\phi^{\widetilde{P}}\left(\underline{X}(\widehat{P})\right). \tag{15}$$

The districts $\phi^{\widetilde{P}}\left(\underline{X}(\widehat{P})\right)$ and $\phi^{\widehat{P}}\left(\underline{X}(\widetilde{P})\right)$, as well as $\phi^{\widetilde{P}}\left(\overline{X}(\widehat{P})\right)$ and $\phi^{\widehat{P}}\left(\overline{X}(\widetilde{P})\right)$, have the same number of groups (KK') and (up to a permutation) the same ethnic distribution. Further, by Lemma 2, $\phi^{\widetilde{P}}\left(\underline{X}(\widehat{P})\right) \sim \phi^{\widehat{P}}\left(\underline{X}(\widetilde{P})\right)$. Assume w.l.o.g. that $\phi^{\widehat{P}}\left(\overline{X}(\widetilde{P})\right) \succcurlyeq \phi^{\widetilde{P}}\left(\overline{X}(\widehat{P})\right)$. By Lemma 4, there is a unique γ such that $\phi^{\widetilde{P}}\left(\overline{X}(\widehat{P})\right) \sim \gamma\phi^{\widehat{P}}\left(\overline{X}(\widetilde{P})\right) \uplus (1 - \gamma)\phi^{\widehat{P}}\left(\underline{X}(\widetilde{P})\right)$. Applying SI, IND (twice) and SDP, it follows from (15) that

$$X \sim \widehat{\alpha}\gamma\phi^{\widehat{P}}\left(\overline{X}(\widetilde{P})\right) \uplus (1 - \gamma\widehat{\alpha})\phi^{\widehat{P}}\left(\underline{X}(\widetilde{P})\right). \tag{16}$$

By (16), (14), and the uniqueness of $\widetilde{\alpha}$, $\widetilde{\alpha} = \widehat{\alpha}\gamma$. Exactly the same reasoning leads to $\widetilde{\beta} = \widehat{\beta}\gamma$. Consequently $\widehat{\alpha}/\widehat{\beta} = \widetilde{\alpha}/\widetilde{\beta}$, and S is well-defined.

---

these choices.

[34]In particular, $S\left(\overline{X}(P(X_1))\right) = h(P(X_1))$.

Lemma 13 *The index S represents $\succcurlyeq$.*

Proof. Let $X, Y \in \mathcal{C}$. Let the group distribution of X (respectively, Y) be P (P'). Let $\widetilde{P}$ denote the group distribution that results from dividing each group in the distribution P in proportions given by P', and then dividing each group in the resulting distribution by the proportions given by $P(X_1)$. By GDP and SI, $\overline{X}\left(\widetilde{P}\right) \succcurlyeq X$, $\overline{X}\left(\widetilde{P}\right) \succcurlyeq Y$, and $\overline{X}\left(\widetilde{P}\right) \succcurlyeq \overline{X}(P(X_1))$. Let α_X uniquely satisfy $X \sim \alpha_X \overline{X}\left(\widetilde{P}\right) \uplus (1 - \alpha_X)\underline{X}\left(\widetilde{P}\right)$, α_Y uniquely satisfy $Y \sim \alpha_Y \overline{X}\left(\widetilde{P}\right) \uplus (1 - \alpha_Y)\underline{X}\left(\widetilde{P}\right)$, and β uniquely satisfy $\overline{X}(P(X_1)) \sim \beta\overline{X}\left(\widetilde{P}\right) \uplus (1 - \beta)\underline{X}\left(\widetilde{P}\right)$. Then $X \succcurlyeq Y$ iff $\alpha_X \geq \alpha_Y$ by Lemma 3. This holds iff $S(X) \geq S(Y)$ since $\beta > 0$. Q.E.D.

The next lemma shows that S is linear among districts that contain K equal-size ethnic groups.

Lemma 14 *Assume $X, Y \in \mathcal{C}^A$ each contains exactly K ethnic groups, of equal sizes. Then*

$$S(X \uplus Y) = \frac{T(X)}{T(X) + T(Y)}S(X) + \frac{T(Y)}{T(X) + T(Y)}S(Y)$$

Proof. Let $Z \sim \alpha_Z \overline{X}(\widehat{P}) \uplus (1 - \alpha_Z)\underline{X}(\widehat{P})$ for $Z = X, Y$, where $\widehat{P}$ is a group distribution such that $\overline{X}(\widehat{P}) \succcurlyeq X$, $\overline{X}(\widehat{P}) \succcurlyeq Y$, and $\overline{X}(\widehat{P}) \succcurlyeq \overline{X}(P(X_1))$. Since $\underline{X}(\widehat{P})$ and $\overline{X}(\widehat{P})$ each has unit population, by IND, SDP, and SI,

$$
\begin{aligned}
X \uplus Y \quad \sim \quad & T(X)\left[\alpha_X \overline{X}(\widehat{P}) \uplus (1 - \alpha_X)\underline{X}(\widehat{P})\right] \uplus T(Y)\left[\alpha_Y \overline{X}(\widehat{P}) \uplus (1 - \alpha_Y)\underline{X}(\widehat{P})\right] \\
\sim \quad & \frac{\alpha_X T(X) + \alpha_Y T(Y)}{T(X) + T(Y)}\overline{X}(\widehat{P}) \uplus \left(1 - \frac{\alpha_X T(X) + \alpha_Y T(Y)}{T(X) + T(Y)}\right)\underline{X}(\widehat{P}).
\end{aligned}
$$

Q.E.D.

Let $\mathcal{C}^{\mathbf{Q}}$ consist of all districts X in $\mathcal{C}^A$ such that for each school n and group g in X, T_g^n is a rational number. We first show that $S(X)$ equals the Mutual Information index for all $X \in \mathcal{C}^{\mathbf{Q}}$. Let the "flattening" operator ψ be defined as follows. For any X in $\mathcal{C}^{\mathbf{Q}}$ and each group g in X, let $a_g \geq 0$ and $b_g > 0$ be the smallest non-negative integers such that $P_g(X) = \frac{a_g}{b_g}$. Let $\mathrm{lcm}(X)$ be the least common multiple of the denominators b_g. Let

$m = k * \mathrm{lcm}(X)$ for any positive integer k, and consider the district $\frac{m}{T(X)}X$. In this district, the number of students in each group g is $m\frac{a_g}{b_g}$, an integer. Let $\psi(X, m)$ be the result of splitting each group g in the district $\frac{m}{T(X)}X$ into $m\frac{a_g}{b_g}$ subgroups, each of size one and having the same distribution across schools By SI and GDP, $\psi(X, m) \sim X$. Note that $\psi(X, m)$ has m groups each of size one. Finally, define $\psi(X) = \psi(X, \mathrm{lcm}(X))$.

For any positive integer k, let Σ^k be the set of permutations $\sigma : \{1, \dots k\} \to \{1, \dots k\}$. By SYM, $\sigma(\psi(X)) \sim \psi(X)$ for all $\sigma \in \Sigma^{\mathrm{lcm}(X)}$, and moreover the two districts have the same group distribution. Let $\widehat{X} = \biguplus_{\sigma \in \Sigma^{\mathrm{lcm}(X)}} \sigma(\psi(X))$. By Lemma 14, $S(X) = S(\widehat{X})$.

Let $\psi(X)^n$ be the single-school district whose only school is the nth school of $\psi(X)$ (the flattened version of school n in X). That is, for each group $g \in \mathbf{G}$, the school in $\psi(X)^n$ contains $M(X)P_g(X)$ equal-sized ethnic groups, each consisting of $t_g^n(X)$ students. Note that in order to construct $\psi(X)^n$ it suffices to know the group distribution of X, which we write P, and the group distribution of school n, which we write p^n. This is so because both $M(X)$ and $t_g^n(X)$ can be written as functions of P and p^n only. Letting $\widehat{n} = \biguplus_{\sigma \in \Sigma^M} \sigma(\psi(X)^n)$ be the subdistrict that consists of all group permutations of $\psi(X)^n$, we can group the schools in $\widehat{X}$ into subdistricts $\widehat{X}(n)$ according to the school n in X from which they came: $\widehat{X} = \biguplus_{n \in \mathbf{N}(X)} \widehat{X}(n)$. By Lemma 14, $S(\widehat{X}) = \sum_{n \in \mathbf{N}(X)} \pi^n S(\widehat{X}(n))$ where $\pi^n = \frac{T^n(X)}{T(X)}$. In order to construct $\widehat{X}(n)$ it suffices to know the group distribution of X, and the group distribution of school n. Hence, we can write

$$S(X) = \sum_{n \in \mathbf{N}(X)} \pi^n f(p^n, P) \text{ where } p^n = (p_1^n, ..., p_K^n) \text{ and } P = (P_1, ..., P_K) \qquad (17)$$

and $f(p^n, P) = S\left(\biguplus_{\sigma \in \Sigma^{M(X)}} \sigma(\psi(X)^n)\right)$.

We now extend the domain of the function f to permit the sum of the P_g's to be less than one. For all K-tuples of nonnegative numbers $(p_1, ..., p_K)$ and $(P_1, ..., P_K)$ that satisfy $\sum_{g=1}^K p_g = 1$, $\sum_{g=1}^K P_g < 1$, and $p_g > 0 \Rightarrow P_g > 0$, define

$$f((p_1, ..., p_K), (P_1, ..., P_K)) = f\left((p_1, ..., p_K, 0), \left(P_1, ..., P_K, 1 - \sum_{g=1}^K P_g\right)\right). \qquad (18)$$

Lemma 15 *For any two K-tuples p and P of nonnegative rational numbers satisfying $\sum_{g=1}^{K} p_g = 1$, $\sum_{g=1}^{K} P_g \leq 1$, and such that $p_g > 0 \Rightarrow P_g > 0$, and for any rational $\alpha \in (0,1)$,*

$$f\left((p_1, ..., p_K), (P_1, ..., P_K)\right) = f\left((\alpha p_1, (1-\alpha)p_1, p_2, \dots, p_K), (\alpha P_1, (1-\alpha)P_1, P_2, \dots, P_K)\right).$$

Proof. Assume first that $\sum_{g=1}^{K} P_g = 1$. For $g = 1, ..., K$, let $a_g \geq 0$ and $b_g > 0$ be the smallest non-negative integers such that $P_g(X) = \frac{a_g}{b_g}$. Let $\mathrm{lcm}(X)$ be the least common multiple of $b_1, ..., b_K$. Consider a district $X \in \mathcal{C}^{\mathbf{Q}}$ with group distribution P, which contains a school n with group distribution p. Let $\alpha = \frac{c_0}{c_1} \in (0,1)$, where c_0 and c_1 are nonnegative integers. Let $X' \in \mathcal{C}^{\mathbf{Q}}$ be the district that results from splitting group 1 into two groups, 1_a and 1_b, in proportions α and $1-\alpha$, respectively. Note that $\mathrm{lcm}(X')$ is the least common multiple of $c_1 b_1$, b_2, ..., b_K. Importantly, $\mathrm{lcm}(X')$ is an integer multiple of $\mathrm{lcm}(X)$, and $\psi(X', \mathrm{lcm}(X'))$ and $\psi(X, \mathrm{lcm}(X'))$ are identical districts. Hence, by definition of f,

$$f\left((\alpha p_1, (1-\alpha)p_1, p_2, ..., p_K), (\alpha P_1, (1-\alpha)P_1, P_2, ..., P_K)\right)$$

$$= S\left(\biguplus_{\sigma \in \Sigma^{\mathrm{lcm}(X')}} \sigma\left(\psi(X', \mathrm{lcm}(X'))^n\right)\right) = S\left(\biguplus_{\sigma \in \Sigma^{\mathrm{lcm}(X')}} \sigma\left(\psi(X, \mathrm{lcm}(X'))^n\right)\right)$$

$$= S\left(\biguplus_{\sigma \in \Sigma^{\mathrm{lcm}(X)}} \sigma\left(\psi(X, \mathrm{lcm}(X))^n\right)\right) = f\left((p_1, p_2, ..., p_K), (P_1, P_2, ..., P_K)\right)$$

As for the penultimate equality, note that $\psi(X, \mathrm{lcm}(X'))$ is the result of splitting each group in $\psi(X, \mathrm{lcm}(X))$ into $\frac{\mathrm{lcm}(X')}{\mathrm{lcm}(X)}$ identically distributed subgroups, so $\psi(X, \mathrm{lcm}(X')) \sim \psi(X, \mathrm{lcm}(X))$ by GDP. The equality then follows from SYM and Lemma 14. This proves the claim for the case $\sum_{g=1}^{K} P_g = 1$. For general P,

$$f\left((p_1^n, ..., p_K^n), (P_1, ..., P_K)\right)$$
$$= f\left((p_1^n, ..., p_K^n, 0), \left(P_1, ..., P_K, 1 - \sum_{g=1}^{K} P_g\right)\right)$$
$$= f\left((\alpha p_1^n, (1-\alpha)p_1^n, p_2^n, ..., p_K^n, 0), \left(\alpha P_1, (1-\alpha)P_1, P_2, ..., P_K, 1 - \sum_{g=1}^{K} P_g\right)\right)$$
$$= f\left((\alpha p_1^n, (1-\alpha)p_1^n, p_2^n, ..., p_K^n), (\alpha P_1, (1-\alpha)P_1, P_2, ..., P_K)\right)$$

Q.E.D.

For any rational $c \in [0, 1]$, define $\phi(c) = f((1, 0), (c, 1 - c))$.

Lemma 16 *Fix an arbitrary probability distribution $p = (p_1, ..., p_K)$ with $p_1 > 0$, and two K-tuples of nonnegative numbers $P = (P_1, ..., P_K)$ and $\widetilde{P} = \left(\widetilde{P}_1, P_2, ..., P_K\right)$ such that $\sum_{g=1}^{K} P_g \leq 1$, $\sum_{g=1}^{K} \widetilde{P}_g \leq 1$, and $\widetilde{P}_1 > 0$. Assume that $P_g > 0$ whenever $p_g > 0$. Then*

$$\frac{\left(f(p, P) - f\left(p, \widetilde{P}\right)\right)}{p_1} = \phi(P_1) - \phi\left(\widetilde{P}_1\right). \tag{19}$$

Proof. Let e_1^K be the K-vector $(1, 0, ..., 0)$. We will show that

$$\frac{\left(f(p, P) - f\left(p, \widetilde{P}\right)\right)}{p_1} = f\left(e_1^K, P\right) - f\left(e_1^K, \widetilde{P}\right). \tag{20}$$

Equation (20) then implies (19) by Lemma 15. Assume without loss of generality that $\widetilde{P}_1 \leq P_1$. If $P_1 = 1$, then $p_1 = 1$, so (20) holds trivially. So assume that $P_1 < 1$ and $\sum_{g=1}^{K} P_g = 1$. By Lemma 15, we may assume that there are at least two nonempty groups other than group 1. Let $\lambda = \min\{P_g : P_g > 0\}$. For any $\pi^1 \in [0, \lambda]$ and $c \in [0, 1]$, we first construct two districts $X(\pi^1, c)$ and $Y(\pi^1, c)$ that have the same population of 1 and group distribution P. The first school in each district is constructed as follows. Let school 1 in $X(\pi^1, c)$ have group distribution p and a total of π^1 students, and let school 1 in district $Y(\pi^1, c)$ contain only group-1 students (so its group distribution is e_1^K), and let the two schools have the same number of members of group 1. Since π^1, the total number of students in school 1 of district $X(\pi^1, c)$, does not exceed λ, the number of students of each group g that attend that school, $\pi^1 p_g$, does not exceed the total number of students of group g, P_g, in that district. Also, since school 1 in $Y(\pi^1, c)$ has the same number of group-1 students as school 1 in $X(\pi^1, c)$ and contains no other students, the number of group-1 students that attend school 1 in $Y(\pi^1, c)$, $\pi^1 p_1$, will not exceed the total number of group-1 students, P_1, in district $Y(\pi^1, c)$.

In each district, let school 2 be a ghetto school that contains the remaining $P_1 - \pi^1 p_1$ students who belong to group 1. The remaining schools, which contain only members of

groups 2 through K, are constructed as follows. For any $c \in [0, 1]$, in each district let a proportion c of the students in each group $g > 1$ who are not in school 1 be allocated by themselves to a ghetto school, and let all the remaining students in each group $g > 1$ who are not in school 1 be allocated to in a single mixed school. Thus, each district contains $K + 2$ schools in total. The example of three ethnic groups appears below, where rows represent groups and columns represent schools:

$$X\left(\pi^1, c\right) = \begin{pmatrix} \pi^1 p_1 & P_1 - \pi^1 p_1 & 0 & 0 & 0 \\ \pi^1 p_2 & 0 & c\left(P_2 - \pi^1 p_2\right) & 0 & (1-c)\left(P_2 - \pi^1 p_2\right) \\ \pi^1 p_3 & 0 & 0 & c\left(P_3 - \pi^1 p_3\right) & (1-c)\left(P_3 - \pi^1 p_3\right) \end{pmatrix}$$

$$Y\left(\pi^1, c\right) = \begin{pmatrix} \pi^1 p_1 & P_1 - \pi^1 p_1 & 0 & 0 & 0 \\ 0 & 0 & cP_2 & 0 & (1-c) P_2 \\ 0 & 0 & 0 & cP_3 & (1-c) P_3 \end{pmatrix}$$

We will now show that there is a $\pi^1 \in (0, \lambda]$, and $c_0, c_1 \in [0, 1]$ such that $X\left(\pi^1, c_0\right) \sim Y\left(\pi^1, c_1\right)$. First, for any $\pi^1 \in [0, \lambda]$ and $1 \geq c' \geq c \geq 0$, SDP and IND imply that $X\left(\pi^1, c'\right) \succcurlyeq X\left(\pi^1, c\right)$ and $Y\left(\pi^1, c'\right) \succcurlyeq Y\left(\pi^1, c\right)$. In addition, for any $\pi^1 \in [0, \lambda]$ and $c \in [0, 1]$, $X(0, 1) \succcurlyeq X\left(\pi^1, c\right) \succcurlyeq X(0, 0)$ and $Y(0, 1) \succcurlyeq Y\left(\pi^1, c\right) \succcurlyeq Y(0, 0)$ by SDP and IND. Finally, for any $c \in [0, 1]$, $X(0, c) = Y(0, c)$.

If $X(0, 1) \sim X(0, 0)$, then $X\left(\pi^1, c_0\right) \sim Y\left(\pi^1, c_1\right)$ for all $\pi^1 \in (0, \lambda]$ and $c_0, c_1 \in [0, 1]$, so we are done. If $X(0, 1) \succ X(0, 0)$, then by Lemma 3 and IND, $X(0, 1) \succ X(0, 1/2) \succ X(0, 0)$. By Lemma 1, the sets $\{c \in [0, 1] : cY(\lambda, 1) \uplus (1-c) Y(0, 1) \preccurlyeq X(0, 1/2)\}$ and $\{c \in [0, 1] : cY(\lambda, 0) \uplus (1-c) Y(0, 0) \succcurlyeq X(0, 1/2)\}$ are both closed. Hence, their complements are the intersections of open sets with $[0, 1]$. Moreover, their complements are nonempty as they each contain $c = 0$. Thus, there is an $\varepsilon \in (0, 1]$ such that for all $c \in [0, \varepsilon)$, $cY(\lambda, 1) \uplus (1-c) Y(0, 1) \succ X(0, 1/2) \succ cY(\lambda, 0) \uplus (1-c) Y(0, 0)$. However, by SDP, for any $c, c' \in [0, 1]$, $cY(\lambda, c') \uplus (1-c) Y(0, c') \sim Y(c\lambda, c')$. We have shown that there is an $\varepsilon' \in (0, \lambda]$ such that for all $\pi^1 \in [0, \varepsilon')$, $Y\left(\pi^1, 1\right) \succ X(0, 1/2) \succ Y\left(\pi^1, 0\right)$. The same argument shows that there is an $\varepsilon'' \in (0, \lambda]$ such that for all $\pi^1 \in [0, \varepsilon)$, $X\left(\pi^1, 1\right) \succ X(0, 1/2) \succ X\left(\pi^1, 0\right)$. Let $\varepsilon = \min\{\varepsilon', \varepsilon''\}$. By Lemma 1, for all $\pi^1 \in [0, \varepsilon)$, the sets

$\{c \in [0,1] : cY\left(\pi^1, 1\right) \uplus (1-c) Y\left(\pi^1, 0\right) \succsim X\left(0, 1/2\right)\}$ and

$\{c \in [0,1] : cY\left(\pi^1, 1\right) \uplus (1-c) Y\left(\pi^1, 0\right) \precsim X\left(0, 1/2\right)\}$ are both closed. By the preceding, they are nonempty: the first set contains $c = 1$ while the second contains $c = 0$. Moreover, their union is the unit interval as $\succsim$ is complete. Since the unit interval is connected, the intersection of the two sets must be nonempty. Let c_1 be in this intersection. By SDP, we have shown that $Y\left(\pi^1, c_1\right) \sim X\left(0, 1/2\right)$. An analogous argument shows that there must be a $c_0 \in [0,1]$ such that $X\left(\pi^1, c_0\right) \sim X\left(0, 1/2\right)$. Let $X = X\left(\pi^1, c_0\right)$ and $Y = Y\left(\pi^1, c_1\right)$.

We have now constructed districts $X \sim Y$ with the same (unit) population and group distribution $(P_1, ..., P_K)$, such that the first school in X has group distribution p, the first school in Y has group distribution e_1^K, and all remaining group-1 students are located in a ghetto school ("school 2") in each district. Letting n index the schools in X and π^n denote the number of students in school n, by construction (and using (17)), we can write

$$S\left(X\right) = \pi^1 f\left(p, P\right) + \left(P_1 - \pi^1 p_1\right) f\left(e_1^K, P\right) + \sum_{n \in \mathbf{N}(X) \backslash \{1,2\}} \pi^n f\left(p^n, P\right).$$

Now let $T' > 0$ and $\beta \in (0,1)$ be such that $\frac{1+T'\beta}{1+T'} P_1 = \widetilde{P}_1$, and let Z consist of a single school with T' students and group distribution $p^Z = (\beta P_1, P_2, ..., P_K, (1-\beta) P_1)$. Note that district $X \uplus Z$ has group distribution $P' = \left(\frac{1+T'\beta}{1+T'} P_1, P_2, ..., P_K, \frac{T'(1-\beta)}{1+T'} P_1\right) = \left(\widetilde{P}_1, P_2, ..., P_K, P_1 - \widetilde{P}_1\right)$. Since by (18), $f\left((p_1, ..., p_K, 0), P'\right) = f\left(p, \left(\widetilde{P}_1, P_2, ..., P_K\right)\right) = f\left(p, \widetilde{P}\right)$, we can write

$$(1+T') S\left(X \uplus Z\right) = \pi^1 f\left(p, \widetilde{P}\right) + \left(P_1 - \pi^1 p_1\right) f\left(e_1^K, \widetilde{P}\right) + \sum_{n \in \mathbf{N}(X) \backslash \{1,2\}} \pi^n f\left(p^n, \widetilde{P}\right) + T' f\left(p^Z, P'\right)$$

But for all $n \in N(X) \backslash \{1,2\}$, $p_1^n = 0$, so by Lemma 15, $f\left(p^n, P\right) = f\left((p^n, 0), P'\right) = f\left(p^n, \widetilde{P}\right)$. Accordingly,

$$\begin{aligned}
& S\left(X\right) - (1+T') S\left(X \uplus Z\right) + T' f\left(p^Z, P'\right) \\
= \; & \pi^1 \left(f\left(p, P\right) - f(p, \widetilde{P})\right) + \left(P_1 - \pi^1 p_1\right) \left(f\left(e_1^K, P\right) - f(e_1^K, \widetilde{P})\right).
\end{aligned}$$

Similarly, letting m index the schools in Y and $\widehat{\pi}^m$ denote the number of students in school

m, we can obtain

$$S\left(Y\right) - \left(1 + T'\right)S\left(Y \uplus Z\right) + T'f\left(p^Z, P'\right)$$
$$= \widehat{\pi}^1\left(f\left(e_1^K, P\right) - f(e_1^K, \widetilde{P})\right) + \left(P_1 - \pi^1 p_1\right)\left(f\left(e_1^K, P\right) - f(e_1^K, \widetilde{P})\right).$$

But since $S(X) = S(Y)$ and since both X and Y have the same population and group distribution, by IND, $S(X \uplus Z) = S(Y \uplus Z)$. This implies that $S\left(X\right) - \left(1 + T'\right)S\left(X \uplus Z\right) = S\left(Y\right) - \left(1 + T'\right)S\left(Y \uplus Z\right)$, so $\pi^1\left(f\left(p, P\right) - f\left(p, \widetilde{P}\right)\right) = \widehat{\pi}^1\left(f\left(e_1^K, P\right) - f\left(e_1^K, \widetilde{P}\right)\right)$. Equation (20) then follows since $\widehat{\pi}^1 = \pi^1 p_1$.

For the case $\sum_{g=1}^K P_g < 1$, let $\widehat{P} = \left(1 - \sum_{g=2}^K P_g, P_2, ..., P_K\right)$. By the prior result (twice),

$$\frac{\left(f\left(p, P\right) - f\left(p, \widetilde{P}\right)\right)}{p_1} = \frac{\left(f\left(p, P\right) - f\left(p, \widehat{P}\right)\right)}{p_1} + \frac{\left(f\left(p, \widehat{P}\right) - f\left(p, \widetilde{P}\right)\right)}{p_1}$$
$$= \left(f\left(e_1^K, P\right) - f\left(e_1^K, \widehat{P}\right)\right) + \left(f\left(e_1^K, \widehat{P}\right) - f\left(e_1^K, \widetilde{P}\right)\right)$$
$$= f\left(e_1^K, P\right) - f\left(e_1^K, \widetilde{P}\right)$$

implying (20) in this case as well.   Q.E.D.

The preceding two lemmas imply that f can be disaggregated into a weighted sum of group-specific components.   For any probability distributions $p = (p_1, ..., p_K)$ and $P = (P_1, ..., P_K)$ such that the support of P contains the support of p, and for any $g = 1, ..., K$ let $Q^g = (P_1, ..., P_g, p_{g+1}, ..., p_K)$ and let $Q^0 = p$.   Since f satisfies SYM, we may assume the groups are arranged so that $p_g - P_g$ is nonincreasing in g; this implies that the sum of elements of each Q^g does not exceed one.   For any $g = 1, ..., K$, let e_g^K be a K-vector with 1 in the gth place and zeroes elsewhere.   Since $f\left(p, p\right) = 0$,

$$f\left(p, P\right) = \sum_{g=1}^K \left[f\left(p, Q^g\right) - f\left(p, Q^{g-1}\right)\right] = \sum_{\substack{g=1,...,K: \\ p_g > 0}} p_g \frac{f\left(p, Q^g\right) - f\left(p, Q^{g-1}\right)}{p_g}$$
$$= \sum_{\substack{g=1,...,K: \\ p_g > 0}} p_g \left[\phi\left(P_g\right) - \phi\left(p_g\right)\right] \text{ (by Lemma 16)} \tag{21}$$

(By Lemma 15, if $p_g = 0$, then $f\left(p, Q^g\right) = f\left(p, Q^{g-1}\right)$, which implies the second equality.) Given this disaggregation, it remains to show that ϕ is the logarithmic function.

By (21) and Lemma 15, for any rational $P_g, p_g \in (0, 1]$ and positive integer m,

$$p_g \left[\phi(P_g) - \phi(p_g) \right] = m \frac{p_g}{m} \left[\phi\left(\frac{P_g}{m} \right) - \phi\left(\frac{p_g}{m} \right) \right] = p_g \left[\phi\left(\frac{P_g}{m} \right) - \phi\left(\frac{p_g}{m} \right) \right].$$

Accordingly, for any rational $\alpha \in (0, 1]$ and positive integers $m_0 \leq m_1$,

$$\phi\left(\alpha \frac{m_0}{m_1} \right) - \phi\left(\frac{m_0}{m_1} \right) = \phi\left(\alpha \frac{1}{m_1} \right) - \phi\left(\frac{1}{m_1} \right) = \phi(\alpha) - \phi(1)$$

Since $\phi(1) = 0$, this implies that $\phi(\alpha\beta) = \phi(\alpha) + \phi(\beta)$ for all rational α, β in $(0, 1]$.

For any positive integer m, define $\widehat{\phi}(m) = \phi(1/m)$. Note that for any positive integers m_0 and m_1, $\widehat{\phi}(m_0 m_1) = \phi\left(\frac{1}{m_0 m_1} \right) = \phi\left(\frac{1}{m_0} \right) + \phi\left(\frac{1}{m_1} \right) = \widehat{\phi}(m_0) + \widehat{\phi}(m_1)$. Hence, $\widehat{\phi}$ is a completely additive number theoretic function (Aczel and Daroczy [1, def. 0.4.1, p. 16]). Moreover, for any positive integer m,

$$\phi\left(\frac{1}{m+1} \right) = \phi\left(\frac{1}{m} \frac{m}{m+1} \right) = \phi\left(\frac{1}{m} \right) + \phi\left(\frac{m}{m+1} \right)$$
$$\implies \widehat{\phi}(m+1) - \widehat{\phi}(m) = \phi\left(\frac{m}{m+1} \right) \geq 0$$

so $\widehat{\phi}$ is nondecreasing. Thus, by Corollary 0.4.17 in Aczel and Daroczy [1, p. 20], there is a nonnegative constant c such that $\widehat{\phi}(m) = c \log_2 m$ for positive integers m. But this implies that for any positive integers $m_0 \geq m_1$,

$$\phi\left(\frac{1}{m_1} \right) = \phi\left(\frac{1}{m_0} \frac{m_0}{m_1} \right) = \phi\left(\frac{1}{m_0} \right) + \phi\left(\frac{m_0}{m_1} \right)$$
$$\implies \phi\left(\frac{m_0}{m_1} \right) = \phi\left(\frac{1}{m_1} \right) - \phi\left(\frac{1}{m_0} \right) = \widehat{\phi}(m_1) - \widehat{\phi}(m_0) = -c \log_2 \left(\frac{m_0}{m_1} \right)$$

Recalling that $P(X_1) = (P_1(X_1), ..., P_K(X_1))$ is the group distribution of X_1, our above results imply that

$$S\left(\overline{X}(P(X_1)) \right) = \sum_{g=1}^{K} P_g(X_1) \left[\phi(P_g(X_1)) - \phi(1) \right] = -c \sum_{g=1}^{K} P_g(X_1) \log_2(P_g(X_1)) = cH(P(X_1))$$

However, $S\left(\overline{X}(P(X_1)) \right)$ also equals $h(P(X_1))$ (footnote 34). Accordingly, c must equal one. Thus, $\phi = \log_2$ and we conclude that S equals the Mutual Information index on $\mathcal{C}^{\mathbf{Q}}$.

It remains to extend this result to $\mathcal{C}^A$. Let $X \in \mathcal{C}^A$ be an arbitrary district with group set $\mathbf{G}$ and school set N. Let $\{X^k\}$ be a sequence of districts in $\mathcal{C}(\mathbf{G}, N)$ that converges to X, such that in each X^k the number of students T_g^n in each school n in each group g is rational. We want to show that $S(X^k) \to S(X)$. Since the Mutual Information index is a continuous function, this will show that S is the Mutual Information index at X as well. Assume for the moment (we will show this soon) that there is a group distribution $\widehat{P}$ such that $\overline{X}(\widehat{P}) \succcurlyeq X^k$ for all k and $\overline{X}(\widehat{P}) \succcurlyeq \overline{X}(P(X_1))$. Let α and α_k be the unique numbers such that $X \sim \alpha \overline{X}(\widehat{P}) \uplus (1 - \alpha)\underline{X}(\widehat{P})$ and $X^k \sim \alpha_k \overline{X}(\widehat{P}) \uplus (1 - \alpha_k)\underline{X}(\widehat{P})$. It is enough to show that $\alpha_k \to \alpha$. Assume not. Then, since $\alpha_k \in [0, 1]$, there must be a convergent subsequence $\alpha_{k_\ell} \to \alpha' \neq \alpha$. Suppose first that $\alpha' > \alpha$ and let $\alpha'' = (\alpha' + \alpha)/2$. This means that there is an $L < \infty$ such that for all $\ell > L$, $X_{k_\ell} \succcurlyeq \alpha'' \overline{X}(\widehat{P}) \uplus (1 - \alpha'')\underline{X}(\widehat{P})$. Since the ordering satisfies Continuity, $X \succcurlyeq \alpha'' \overline{X}(\widehat{P}) \uplus (1 - \alpha'')\underline{X}(\widehat{P})$ as well. But this implies, wrongly, that $\alpha \geq \alpha''$. The case in which $\alpha' < \alpha$ is analogous.

It remains to find a group distribution $\widehat{P}$ such that $\overline{X}(\widehat{P}) \succcurlyeq X^k$ for all k and $\overline{X}(\widehat{P}) \succcurlyeq \overline{X}(P(X_1))$. For each k, let $P(X^k)$ be the group distribution of X^k. By SDP, $\overline{X}(P(X^k)) \succcurlyeq X^k$. Since $\overline{X}(P(X^k))$ and $\overline{X}^{|\mathbf{G}|}$ are districts with rational entries, they are ordered by the Mutual Information index. Therefore, direct calculation shows that $\overline{X}^{|\mathbf{G}|} \succcurlyeq \overline{X}(P(X^k))$. Finally, let $\widehat{P}$ be the distribution that results from dividing each group in $P(X_1)$ into $|\mathbf{G}|$ equal-sized groups. By SDP, neither $\overline{X}(P(X_1))$ nor $\overline{X}^{|\mathbf{G}|}$ is more segregated than $\overline{X}(\widehat{P})$. Thus, $\widehat{P}$ is the group distribution that we are looking for. This concludes the proof of Theorem 2. Q.E.D.

Proof of Claim 2

If the index satisfies SI, SDP, and either SSD or SGD, then $S\left(\overline{X}^{K^2}\right) = 2S\left(\overline{X}^K\right)$ for any $K > 1$. The result follows by considering the sequence of values $S\left(\overline{X}^{2^i}\right)$ for $i = 1, 2, 4, 8, \dots$

Proof of Claim 3

Let S be Aggregative and let districts X and Y have equal populations and equal group distributions (i.e., $T.(X) = T.(Y)$). For any district Z,

$$\begin{aligned}
X \uplus Z \quad &\succcurlyeq \quad Y \uplus Z \Longleftrightarrow S(X \uplus Z) \geq S(Y \uplus Z) \\
&\Longleftrightarrow \quad F(S(X), S(Z), T.(X), T.(Z)) \geq F(S(Y), S(Z), T.(Y), T.(Z)) \\
&\Longleftrightarrow \quad S(X) \geq S(Y) \Longleftrightarrow X \succcurlyeq Y
\end{aligned}$$

so $\succcurlyeq$ satisfies IND. Conversely, if $\succcurlyeq$ is represented by the continuous index S and satisfies IND, then for any districts X and Y we can define $F(S(X), S(Z), T.(X), T.(Z))$ to be $S(X \uplus Z)$. Indeed, for any district Z such that $T.(Z) = T.(Y)$ and $S(Z) = S(Y)$, IND implies $S(X \uplus Z) = S(X \uplus Y)$. A similar argument applies to X, so $S(X \uplus Z)$ depends only on the arguments of F and hence is well defined. F is strictly increasing in its first two arguments by IND and inherits continuity from S. Thus, S is Aggregative.

Appendix B. Supplementary Material

This appendix proves claims made in Table 1 that are not shown in sections 3.3 and 4.1 or Appendix A. To avoid redundancy, these claims will be treated property by property.

Symmetry. It is obvious that Symmetry is satisfied by M, D, G, H, NE, but not by C_κ or CR. Also, $A_{\mathbf{w}}$ is clearly asymmetric unless $\mathbf{w} = (1/K, ..., 1/K)$.

Continuity. All of the indices except C_κ are continuous functions and thus satisfy CONT:

Lemma 17 *Any index S that is a continuous function of the T_g^n's (the numbers of each group g in each school n) satisfies CONT.*

Proof. Fix a district Z with group set $\mathbf{G}$ and school set $\mathbf{N}$. The sets $(-\infty, S(Z)]$, and $[S(Z), \infty)$ are closed in $\Re$. Consequently, the intersections of $S^{-1}((-\infty, S(Z)])$ and

$S^{-1}([S(Z),\infty))$ with $\mathcal{C}\left(\mathbf{G},\mathbf{N}\right)$ are closed in $\mathcal{C}\left(\mathbf{G},\mathbf{N}\right)$. (For continuous functions, their inverse image of closed sets are closed). But these are just the sets $\{X \in \mathcal{C}\left(\mathbf{G},\mathbf{N}\right) : X \succcurlyeq Z\}$ and $\{X \in \mathcal{C}\left(\mathbf{G},\mathbf{N}\right) : Z \succcurlyeq X\}$, respectively. Q.E.D.

C_κ violates CONT: let $\kappa = .5$ and let $X\left(\varepsilon\right) = \langle(1-\varepsilon,1),(0,1)\rangle$ and $Z = \langle(1,0),(1,2)\rangle$, where in each school the first entry is the number of blacks. The set $\{X \in \mathcal{C}\left(\mathbf{G},\mathbf{N}\right) : Z \succcurlyeq X\}$ is not closed since it contains $X\left(\varepsilon\right)$ for all $\varepsilon > 0$ but does not include $X\left(0\right)$.

Independence. Section 4.1 shows that IND is satisfied by $A_{\mathbf{w}}$, M, H, C_κ, and the two-group version of NE. However, NE violates IND in general: letting $X = \langle(0,2,3),(6,4,3)\rangle$, $Y = \langle(3,2,0),(3,4,6)\rangle$, and $Z = \langle(0,10,100)\rangle$, $NE(X) = NE(Y)$ since NE satisfies SYM, but one can verify that $NE(X \uplus Z) \neq NE(Y \uplus Z)$. Proofs that D and G violate AGG, and thus IND, appear in the prior literature (e.g., Hutchens [27, n. 12]). As for CR, let $X = \langle(2,4,6),(6,4,2)\rangle$, $Y = \langle(4,2,1),(4,6,7)\rangle$ and $Z = \langle(0,2,5)\rangle$. Although X and Y have the same population and ethnic distribution, $CR(X) = 1/12 < CR(Y) = 10/119$ while $CR(X \uplus Z) = 3/20 > CR(Y \uplus Z) = 88/595$. Hence, CR violates IND.

School Division Property. Section 4.1 shows that $A_{\mathbf{w}}$, M, H, D, and NE satisfy SDP, while C_κ violates it. CR violates SDP: the district $Y = \langle(1,0,4),(1,2,0),(4,1,1)\rangle$ is obtained by splitting the first school in the district $X = \langle(2,2,4),(4,1,1)\rangle$ in two, but $CR(X) > CR(Y)$. As for G, let X' be the district that results from some district X if school $n \in X$ is divided into two schools, n_1 and n_2, and let $\alpha = T^{n_1}/T^n$. Then

$$G(X') - G(X) = \frac{1}{I}\sum_{g=1}^{G}\frac{T^{n_1}T^{n_2}}{TT}\left|\frac{T_g^{n_1}}{T^{n_1}} - \frac{T_g^{n_2}}{T^{n_2}}\right|$$

$$+\frac{1}{I}\sum_{g=1}^{G}\sum_{\substack{m=1,\ldots,N \\ m\neq n}}\left(\frac{T^m}{TT}\left(\left|T^{n_1}\frac{T_g^m}{T^m} - T_g^{n_1}\right| + \left|T^{n_2}\frac{T_g^m}{T^m} - T_g^{n_2}\right| - \left|T^n\frac{T_g^m}{T^m} - T_g^n\right|\right)\right)$$

The first sum is nonnegative. The arguments of the first two absolute values in the second line sum to the argument of the third absolute value function. Since absolute value is a convex function, the summand is nonnegative for all g. Moreover, if the two schools have the same group distributions, then the arguments of the three absolute value functions are

proportional to each other and thus all of the same sign: the summand is zero. Hence, G satisfies SDP.

Composition Invariance. By Claim 1, $A_{\mathbf{w}}$ satisfies CI. The fact that D and G satisfy CI only in the two-group case is well known (Reardon and Firebaugh [42]) so we do not prove it. To see that M, H, NE, and C_κ violate CI, consider the districts $X = \langle (2,1), (1,2) \rangle$ and $Y = \langle (2,2), (1,4) \rangle$. It can be checked that $H(X) \neq H(Y)$, $M(X) \neq M(Y)$, $NE(X) \neq NE(Y)$, and $C_\kappa(X) \neq C_\kappa(Y)$ for thresholds $\kappa \in (1/2, 2/3)$.[35] CR violates CI: letting $X = \langle (9,5,1), (1,5,9) \rangle$ and $Y = \langle (9,5,10), (1,5,90) \rangle$, one can verify that $CR(X) = 16/75$ while $CR(Y) = 7/48$.

Group Division Property. The results for GDP have all been shown in sections 3.3 and 4.1.

Aggregation. Except for C_κ, AGG and IND are equivalent, as argued. As C_κ can be written

$$C_\kappa(X \uplus Y) = \frac{T_2(X)}{T_2(X) + T_2(Y)} C_\kappa(X) + \frac{T_2(Y)}{T_2(X) + T_2(Y)} C_\kappa(Y), \qquad (22)$$

it satisfies AGG.

Additive Decomposability. Since AD implies AGG, D, G, CR, and the 3+ group NE violate AD. On the other hand, Reardon and Firebaugh [42, pp. 53-4] show that with two groups, NE satisfies AD. We show below that M satisfies SSD, so it satisfies AD. Hence, H does as well. $A_{\mathbf{w}}$ satisfies (4) with weights $w_Z = \prod_{g \in G} \left(\frac{T_g(Z)}{T_g(X) + T_g(Y)} \right)^{w_g}$ for $Z = X, Y$, so it satisfies AD. As for the Clotfelter index, for any two districts X and Y, let X' any district that results from reallocating the students in X among the schools of X so that $C_\kappa(X') \neq C_\kappa(X)$. The change in the number of group-2 students who are in schools in which at least a proportion κ of students are in group 2 can be measured alternatively by $[T_2(X) + T_2(Y)][C_\kappa(X' \uplus Y) - C_\kappa(X \uplus Y)]$ or by $T_2(X)[C_\kappa(X') - C_\kappa(X)]$. As these must be equal, if C_κ satisfies AD then $w_X = \frac{T_2(X)}{T_2(X) + T_2(Y)}$ and, similarly, $w_Y = \frac{T_2(Y)}{T_2(X) + T_2(Y)}$. By (22) this implies, incorrectly, that $C_\kappa(c(X) \uplus c(Y))$ is identically zero: the Clotfelter

---

[35]In each school we list the numbers of blacks and whites, in that order.

index violates AD.

Strong School and Group Decomposability. All indices but M, CR, and C_κ take their maximum value of one on $\overline{X}^2$ and satisfy SI and SDP. Hence, by Claim 2, they cannot satisfy SSD or SGD. (As noted, SGD is undefined for $A_{\mathbf{w}}$.) CR and C_κ violate SSD as they violate AD; SGD is not applicable to them. To see that M satisfies SSD, let $X = X^1 \uplus X^2$ be a district composed of 2 subdistricts. By definition of M, $M(X) = H(P(X)) - \sum_{k=1}^{2} \sum_{n \in \mathbf{N}(X^k)} \pi^n H(p^n)$. Subtracting and adding $\sum_{k=1}^{2} \pi^k H(P(X^k))$ on the right hand side, we obtain

$$M(X) = H(P(X)) - \sum_{k=1}^{2} \pi^k H(P(X^k)) + \sum_{k=1}^{2} \pi^k H(P(X^k)) - \sum_{k=1}^{2} \sum_{n \in \mathbf{N}(X^k)} \pi^n H(p^n)$$

$$= H(P(X)) - \sum_{k=1}^{2} \pi^k H(P(X^k)) + \sum_{k=1}^{2} \pi^k \left(H(P(X^k)) - \sum_{n \in \mathbf{N}(X^k)} \pi^n H(p^n) \right)$$

$$= M(c(X^1) \uplus c(X^2)) + \sum_{k=1}^{2} \pi^k M(X^k),$$

so M satisfies SSD. That M satisfies SGD follows from symmetry of mutual information (Cover and Thomas [14, pp. 18 ff.]).

References

[1] Aczél, J. 1966. "Lectures on Functional Equations and their Applications." Academic Press, New York and London.

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

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

[4] Blackwell, D. 1953, "Equivalent Comparisons of Experiments." Ann. Math. Statist. 24, 265-272.

[5] Bohenblust, H. L. Shapley, and S. Sherman. 1949. "Reconnaissance in Game Theory." RAND Working Paper RM-208.

[6] 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, pp. 269-338.

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

[8] Card, David, and Jesse Rothstein. 2007. "Racial Segregation and the Black-White Test Score Gap." Journal of Public Economics 91:2158-84.

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

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

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

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

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

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

[15] Cutler, David, Edward Glaeser, and Jacob Vigdor. 1999. "The Rise and Decline of the American Ghetto" Journal of Political Economy 107: 455– 506.

[16] ——. 2008. "When are Ghettos Bad? Lessons from Immigration Segregation in the United States." Journal of Urban Economics 63:759-774.

[17] Dalton, Hugh. 1920. "The Measurement of the Inequality of Incomes." Economic Journal 30:348-61.

[18] Echenique, Federico, and Roland G. Fryer, Jr. 2007. "A Measure of Segregation Based on Social Interactions." Quarterly Journal of Economics 122: 441-485.

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

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

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

[22] Grant, S., A. Kajii, and B. Polak 1998. "Intrinsic Preference for Information," Journal of Economic Theory 83, 233-259.

[23] Guryan, Jonathan. 2005. "Desegregation and Black Dropout Rates." American Economic Review 94:919-943.

[24] 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." NBER Working Paper No. W8741.

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

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

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

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

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

[30] James, David R., and Karl E. Taeuber. 1985. "Measures of Segregation." Sociological Methodology 15:1–32.

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

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

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

[34] McLaughlin, Don. 2003. "CCD Data File: Thirteen-Year Longitudinal Common Core of Data Non-Fiscal Survey Database: School Years 1986-87 — 1998-99." Mimeo, NCES.

[35] 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.

[36] 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.

[37] Mora, Ricardo, and Javier Ruiz-Castillo. 2005. "An Evaluation of an Entropy Based index of Segregation." Mimeo, Universidad Carlos III de Madrid.

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

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

[40] Philipson, Tomas. 1993. "Social Welfare and Measurement of Segregation." Journal of Economic Theory 60:322-334.

[41] 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.

[42] Reardon, Sean F., and Glenn Firebaugh. 2002. "Measures of Multigroup Segregation." Sociological Methodology 32:33-67.

[43] Rivkin, Steven G. 1994. "Residential Segregation and School Integration." Sociology of Education 67:279-92.

[44] Sakoda. J. 1981. "A Generalized index of Dissimilarity." Demography 18:245-50.

[45] Shorrocks, A. F. 1980. "The Class of Additively Decomposable Inequality Measures." Econometrica 39:613-25.

[46] Shorrocks, A. F. 1984. "Inequality Decompositions by Population Subgroups." Econometrica 48:1369-86.

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

[48] Theil, Henri. 1967. Economics and Information Theory. Amsterdam: North-Holland.

[49] ——. 1971. Principles of Econometrics. New York: Wiley & Sons.

[50] ——. 1972. Statistical Decomposition Analysis. Amsterdam: North-Holland.

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

[52] von Neumann, John, and Oskar Morgenstern. 1944. Theory of Games and Economic Behavior. Princeton University Press.