

THE MEASUREMENT OF INCOME SEGREGATION

Casilda Lasso de la Vega and
Oscar Volij

Discussion Paper No. 17-04

October 2017

Monaster Center for
Economic Research
Ben-Gurion University of the Negev
P.O. Box 653
Beer Sheva, Israel

Fax: 972-8-6472941
Tel: 972-8-6472286

The Measurement of Income Segregation*

Casilda Lasso de la Vega

Oscar Volij

University of the Basque Country

Ben Gurion University

October 18, 2018

Abstract

We examine the problem of measuring the extent to which students with different income levels attend separate schools. Unless rich and poor attend the same schools in the same proportions, some segregation will exist. Since income is a continuous cardinal variable, however, the rich-poor dichotomy is necessarily arbitrary and renders any application of a binary segregation measure artificial. This paper provides an axiomatic characterization of a measure of income segregation that takes into account the cardinal nature of income.

This measure satisfies an empirically useful decomposition by sub-districts.

*We are indebted to David M. Frankel. This paper benefited greatly from his generous collaboration. We also thank the Spanish Ministerio de Economía y Competitividad (project ECO2015-67519-P) and the Gobierno Vasco (project IT568-13) for research support. This research has used as a source of information databases of the Education Quality Agency of Chile (Agencia de Calidad de la Educación). The authors thanks the Education Quality Agency for access to information. All the results of the study are the responsibility of the authors and do not compromise the Institution. Email addresses: casilda.lassodelavega@ehu.es; ovolij@bgu.ac.il.

1 Introduction

Segregation is an attribute of school districts.¹ It refers to the extent to which pupils belonging to different demographic groups attend separate schools. When demographic groups are classified according to ethnicity, we are dealing with ethnic segregation. When they are classified according to gender, segregation is labeled as gender segregation. In this paper we are interested in income segregation, which can be observed when groups are classified according to income levels.

The criterion according to which we choose to classify individuals is not an innocuous one. When dealing with ethnicity or gender, for instance, there is no natural order of groups and indeed most of the ethnic segregation indices in the literature treat ethnic groups symmetrically. In other contexts, however, groups can be ordered according to some natural criterion. For example, pupils could be classified according to the educational level of their parents into having completed a primary, secondary or higher education. In these cases, it may not be appropriate to treat groups symmetrically, and in fact, indices have been developed that take into account the ordering of the groups. A richer context yet is the one of income segregation. Not only does income induce an order of the groups but it also induces a natural metric on them. Here too, segregation indices have been proposed that take into account the ordering of income levels and also their magnitude.²

School segregation, and its counterpart school diversity, are twin topics that regularly arise in political forums and in the media. Diversity and segregation are not restricted to race. In the US, for instance, programs exist that aim at increasing socioeconomic diversity in schools and creating more integrated public schools. Recently, in its concern that elite institutions enroll students who are diverse in every

¹More generally, segregation is an attribute of a collection of organizational units. For expositional purposes we focus on school districts, whose organizational units are, unsurprisingly, schools.

²For a necessarily incomplete list of ethnic segregation indices see Massey and Denton [16] and Reardon and Firebaugh [20]. For segregation among ordered categories, see for example, Reardon [18, 19]. For an index that exploits the cardinal nature of income, see Jargowsky [12].

aspect except economically, the New York Times has developed the College Access Index which attempts to measure economic diversity at top colleges, and which is published every year.

Recent empirical studies suggest that income segregation may affect educational outcomes. Students who have higher-quality peer groups tend to have better educational outcomes (Coleman *et al.* [1]), an effect for which evidence has been found to be causal (Hoxby [7]; Hanushek, Kain, Markman, and Rivkin [6]; Imberman, Kugler, and Sacerdote [10]; Lavy, Paserman, and Schlosser [14]). As pupils with higher family income tend to have higher ability, income segregation may be a significant source of differential peer effects across schools. Indeed, the findings of Mayer [17] suggest that an increase in income segregation between census tracts or school districts tends to lower the achievement of low ability pupils and raise that of high ability pupils.

Despite the potential importance of income segregation, there is wide disagreement about how to measure it. Several income segregation indices have been proposed in the literature and some of their properties have been pointed out. Some researchers have used ethnic segregation indices, such as Dissimilarity Index of Jahn, Schmid, and Schrag [11]. Other indices, notably the rank-order information theory index of Reardon [19], take account of the ordinal nature of income categories. Finally, some indices treat income as a cardinal variable, the main example being Jargowsky's [12] neighborhood sorting index.

Most of the attempts to measure income segregation, however, consist of transforming a district in which individuals are classified by income into one in which they are classified by dichotomous categories so that a standard ethnic segregation index can be applied. For instance, some authors propose to establish a poverty line to partition the population into rich and poor and to apply an existing two-group segregation index. A more sophisticated approach measures income segregation by averaging the two-group segregation indices associated with all possible poverty lines. The problem with these attempts is that the resulting indices may and do fail to

satisfy certain basic properties that any income segregation index should satisfy.

In this paper, instead of directly adapting an ethnic segregation index to the context of income segregation, we adapt the *properties* of standard ethnic segregation measures and apply them to the new context. Specifically, we adapt a number of properties satisfied by several indices of ethnic segregation and use them to characterize an index of income segregation, which we call the School Separation index. This index measures segregation as the difference between the district’s variability and the average variability of its schools, variability being measured by the mean logarithmic deviation. To the best of our knowledge this is the first axiomatic derivation of an income segregation measure.³

Before we move to the formal model, we discuss the concept of income segregation we have in mind and its relation to income inequality. Though different concepts, they are closely related to each other. Changes in the latter, however measured, will typically affect the former. Yet, some authors propose to disentangle the two concepts as much as possible. Reardon [19], for instance, proposes that income segregation be maximal if and only if within each school, all pupils have the same income, no matter what the income distribution of the district may be. To illustrate this requirement, which Reardon [19] calls *scale interpretability*, consider the following districts.

X	\$10	\$20
School 1	100	0
School 2	0	100

Y	\$10	$\$20 \times 10^6$
School 1	100	0
School 2	0	100

Both districts have two schools, one attended by the rich and the other by the poor. However, whereas the poor in both districts have an income of \$10, the rich have an income of \$20 in district X and an income of \$20 million in district Y . By virtue of scale interpretability, they are equally and maximally segregated. This is so despite the fact that the difference between rich and poor in X is negligible compared to the

³Measures of segregation among unordered categories such as ethnic groups have been axiomatized by Echenique and Fryer [2], Frankel and Volij [5], and Hutchens [8, 9].

corresponding difference in Y . The idea of income segregation that we have in mind is inconsistent with the above requirement. In fact our axioms will imply that district X exhibits less income segregation than district Y since, although in both districts poor and rich attend separate schools, district Y exhibits a much higher income inequality than X . In other words, according to our concept of income segregation, the extent to which students with different incomes attend different schools is magnified by the inequality of students' incomes.

To further illustrate the difference between a concept of income segregation that fulfills scale interpretability and the concept that we propose, consider the following two districts.

X	\$200	\$300
School 1	20	0
School 2	0	20

Y	\$100	\$200	\$300	\$400
School 1	10	10	0	0
School 2	0	0	10	10

District X consists of two schools, one attended by the rich and one attended by the poor. Since all the poor have an income of \$200 and all the rich have an income of \$300, according to the above requirement, X has maximum segregation. If we now make half the poor even poorer and half the rich even richer, by transferring \$100 from the former to the latter, we obtain district Y . According to scale interpretability segregation is reduced. The reason for this reduction is that although nobody moved from one school to the other, and although the poor and the rich still go to separate schools, the school attended by the poor became “more diverse” as a result of the pauperization of half of the already poor, and similarly the school attended by the rich also became “more diverse” as a result of the enrichment of half of the already rich. In contrast, according to our notion of segregation, the increase in income inequality observed in the transition from X to Y magnifies the income segregation already existing in X ; the poor and the rich still attend separate schools, and the difference between rich and poor became more striking.

The paper is organized as follows. After introducing the basic notation in Sec-

tion 2, Section 3 gives a brief review of the approaches followed so far to measure income segregation. Section 4 introduces, among others, the income segregation indices that are the focus of the paper. After proposing a list of axioms in Section 5, Section 6 presents our characterization result.

2 Notation

An *income group* is characterized by a pair (n, y) where $n \in \mathbb{R}_+$ is the group's population size and $y > 0$ is the income of each pupil in the group. A *school* $\langle (n_g, y_g) \rangle_{g \in G}$ is a finite collection of income groups where G is the set of groups. If two income groups with the same income in a school are combined, the school does not change; e.g., the schools $\langle (n, y), (n', y) \rangle$ and $\langle (n+n', y) \rangle$ are regarded as the same school. Also, if we permute the income groups the school does not change; e.g., for any $\pi : G \rightarrow G$, $\langle (n_g, y_g) \rangle_{g \in G} = \langle (n_{\pi(g)}, y_{\pi(g)}) \rangle_{g \in G}$.

For any school $c = \langle (n_g, y_g) \rangle_{g \in G}$, let $|c| = \sum_{g \in G} n_g y_g$ denote the total income of school c , and $n_c = \sum_{g \in G} n_g$ its total enrollment. If $n_c = 0$, c is an empty school. Empty schools will play no role in the paper, but are needed for notational convenience. If c is not empty, we denote by $\mu_c = |c|/n_c$ its mean income, and by $\bar{c} = \langle (n_c, \mu_c) \rangle$ denote the *smoothed* school that is obtained from c by redistributing c 's total income equally among its pupils. For any school $c = \langle (n_g, y_g) \rangle_{g \in G}$ and scalar $\lambda > 0$, let $\lambda c = \langle (\lambda n_g, y_g) \rangle_{g \in G}$ denote the school that is obtained from c by multiplying the number of people in each income group by λ and let $c * \lambda = \langle (n_g, \lambda y_g) \rangle_{g \in G}$ denote the school that is obtained from c by multiplying each pupil's income by λ . For any two schools $c = \langle (n_g, y_g) \rangle_{g \in G}$ and $c' = \langle (n_{g'}, y_{g'}) \rangle_{g' \in G'}$, let $c \# c' = \langle (n_g, y_g) \rangle_{g \in G \cup G'}$ denote the result of combining the two schools into a single school. We say that a sequence of schools $c^m = \langle (n_g^m, y_g^m) \rangle_{g \in G}$ converges to school $c = \langle (n_g, y_g) \rangle_{g \in G}$, denoted $c^m \rightarrow c$, if for all $g \in G$, the income groups (n_g^m, y_g^m) converges to (n_g, y_g) . We denote by \mathcal{C}_+ the class of nonempty schools where pupils have positive incomes.

A *district* $\{c_k\}_{k \in K}$ is a finite collection of schools at least one of which is not empty. We identify any district with the district that is obtained from it by deleting all its empty schools. If we permute the schools, the district does not change; e.g., for any $\pi : K \rightarrow K$, $\{c_k\}_{k \in K} = \{c_{\pi(k)}\}_{k \in K}$. With some abuse of notation we will denote a typical district by $X = \{c_1, \dots, c_K\}$. For any district X , let $n_X = \sum_{c \in X} n_c$ denote the total attendance of X , let $|X| = \sum_{c \in X} |c|$ denote its total income, and let $\mu_X = |X|/n_X$ denote its mean income. For any district X and scalar $\lambda > 0$, let $\lambda X = \{\lambda c\}_{c \in X}$ denote the district that is obtained from X by multiplying the number of people in each school by λ and let $X * \lambda = \{c * \lambda\}_{c \in X}$ denote the district that is obtained from X by multiplying each pupil's income by λ . For any two districts $X = \{c_1, \dots, c_K\}$ and $Y = \{c'_1, \dots, c'_{K'}\}$, let $X \uplus Y = \{c_1, \dots, c_K, c'_1, \dots, c'_{K'}\}$ denote the district that results from combining the schools of X and Y into a single district. We denote by \mathcal{D}_+ the set of all districts where all students have positive incomes.

A district is *simple* if it is of the form $\{\langle(n_1, y_1)\rangle, \dots, \langle(n_K, y_K)\rangle\}$ for some K ; that is if each school contains a single income group. A district is *completely integrated* if it consists of a single school, and thus can be written as $\{c\}$ for some $c \in \mathcal{C}_+$. For any district $X = \{c_1, \dots, c_K\}$, let $C(X) = \{c_1 \# \dots \# c_K\}$ denote the completely integrated district that results from combining the schools of X into a single school. For any school $c = \langle(n_g, y_g)\rangle_{g \in G}$, let $d(c) = \{\langle(n_g, y_g)\rangle\}_{g \in G}$ denote the simple district that results from placing each income group in c into its own school. Lastly, for a district $X = \{c_1, \dots, c_K\}$, let $d(X)$ denote $\uplus_{c \in X} d(c)$: the district that results from applying the operation d to each school in X . We will refer to $d(c)$ and $d(X)$ as the *simple versions* of c and X . For instance, if $c_1 = \langle(2, 1), (3, 2)\rangle$ and $c_2 = \langle(5, 4)\rangle$ are two schools, and $X = \{c_1, c_2\}$, then $C(X) = \{\langle(2, 1), (3, 2), (5, 4)\rangle\}$ and $d(X) = d(c_1) \uplus d(c_2) = \{\langle(2, 1)\rangle, \langle(3, 2)\rangle, \langle(5, 4)\rangle\}$.

3 From ethnic segregation to income segregation

Income segregation of a district refers to the extent to which its schools differ in the way pupils are distributed across income groups. As opposed to ethnic categories, income allows us not only to order the different income groups from poorest to richest but also to measure the distance between them. Despite this flexibility, the initial approach to measuring income segregation has been to classify pupils into dichotomous categories and apply an existing two-group segregation measure. In this section we describe this approach and subsequent methods that adapt two-group segregation measures to the measurement of income segregation. Readers just interested in the main results may move, without loss of continuity, to Section 4.

Segregation is an attribute of a collection of organizational units whose population is classified into several groups. When the number of groups is two, e.g., rich and poor, a typical such collection takes the form $\{\langle P_1, R_1 \rangle, \dots, \langle P_K, R_K \rangle\}$, where for each organizational unit $k = 1, \dots, K$, P_k and R_k are the number of poor and rich, respectively in it. Given a collection $\{\langle P_1, R_1 \rangle, \dots, \langle P_K, R_K \rangle\}$, let p_k stand for the proportion of poor in k , namely $p_k = P_k / (P_k + R_k)$, and let p stand for the overall proportion of poor $p = \sum_k P_k / \sum_k (P_k + R_k)$. Also, let $\pi_k = (P_k + R_k) / \sum_j (P_j + R_j)$ be the proportion of the population in organizational unit k .

Measures of segregation assign to each such collection a number that aims to capture its level of segregation between poor and rich. One example is the celebrated index of dissimilarity. Another example is the Mutual Information index, which is given by

$$MI(\{\langle P_1, R_1 \rangle, \dots, \langle P_K, R_K \rangle\}) = E(p) - \sum \pi_k E(p_k) \quad (1)$$

where $E(p) = -p \log_2(p) - (1-p) \log_2(1-p)$.

Note that the Mutual Information index assigns a value of $E(p)$ to a completely segregated district whose poor constitute a proportion p of the population. That is, the higher the variability of the district's distribution as measured by the entropy,

the higher its potential segregation. Theil and Finizza [24] propose to normalize the Mutual Information index so that the potential segregation of any district is always one. The resulting Entropy index is given by

$$H(\{\langle P_1, R_1 \rangle, \dots, \langle P_K, R_K \rangle\}) = \frac{E(p) - \sum \pi_k E(p_k)}{E(p)}. \quad (2)$$

As mentioned earlier, the first approach to measuring income segregation has been to divide the population into two groups and compute segregation by using a standard two-group segregation measure. Specifically, given a poverty line, we can transform any school into a binary organizational unit as follows. First classify pupils into poor and rich, the poor being those whose incomes are lower or equal the poverty line, and then record the number of poor and rich students. Formally, given a poverty line ℓ , we transform a school $c = \langle (n_g, y_g) \rangle_{g \in G}$ into the pair $c(\ell) = \langle P(\ell), R(\ell) \rangle$ where

$$P(\ell) = \sum_{g: y_g \leq \ell} n_g \quad \text{and} \quad R(\ell) = \sum_{g: y_g > \ell} n_g$$

are the number of poor and rich, respectively, in the school. By applying this transformation to all the schools in a district $X = \{c_1, \dots, c_K\}$, we obtain a collection of binary organizational units $X(\ell) = \{\langle P_1(\ell), R_1(\ell) \rangle, \dots, \langle P_K(\ell), R_K(\ell) \rangle\}$. For instance, consider school district $X = \{c_1, c_2\}$ where $c_1 = \{(20, 2), (10, 4)\}$ and $c_2 = \{(30, 3), (20, 6)\}$. If the poverty line is defined to be $\ell = 3$, then school c_1 has 20 poor and 10 rich, and school c_2 has 30 poor and 20 rich. We can then translate district X into the collection $X(3) = \{\langle 20, 10 \rangle, \langle 30, 20 \rangle\}$. Similarly, if the poverty line is $\ell = 2$, we translate X into $X(2) = \{\langle 20, 10 \rangle, \langle 0, 50 \rangle\}$.

Once we have performed this translation, we can somewhat crudely define the income segregation of X as the segregation between poor and rich as measured by some given binary segregation measure. For instance, if we use the entropy index

defined in (2) we obtain the following income segregation index:

$$\mathcal{H}_\ell(X) = H[X(\ell)].$$

This approach has been extensively applied in the sociology, geography and economics literature. See, for instance, Jenkins et al. [13], and Massey [15].

Needless to say, this approach is problematic for many reasons. Among them, to mention just one, the arbitrariness of the poverty line. An alternative approach, an application of which can be found in Fong and Shibuya [3] among others, consists of dividing the population into several income groups and applying a multigroup segregation index. The main problem with this approach, however, is that since multigroup segregation indices are generally symmetric in groups, it ignores not only the magnitude of the income levels but also their natural ordering. For instance, if the rich become the poor, the poor become the middle class and the middle class become the rich, according to this approach income segregation would remain the same, no matter how many people originally belonged to each class and whatever their level of income is.

Reardon [19] proposes to fix the problems caused by the arbitrary choice of the poverty line, by averaging the indices corresponding to all poverty lines. This approach results in what is known as rank-order segregation indices, which have the virtue of taking into account the fact that income categories are ordered. One example of such indices is the *rank-order information theory segregation index* which is given by the following average of the \mathcal{H}_ℓ indices

$$\mathcal{H}^R(X) = \int_0^\infty \frac{E(p(\ell))}{\int_0^\infty E(p(l))dp(l)} H[X(\ell)] dp(\ell)$$

where p is the cumulative distribution of income in X , or alternatively, $p(\ell)$ is the proportion of poor in $X(\ell)$. Given that X consists of a finite number of schools each with a finite number of income groups, the number of different income levels in X

is finite. Thus, letting $Y = \{y_g : g \in G(c), c \in X\}$ denote the set of all the income levels present in the population, the above expression can be written as

$$\mathcal{H}^R(X) = \sum_{y \in Y} \frac{n_y E(p(y))}{\sum_{l \in Y} n_l E(p(l))} H[X(y)]. \quad (3)$$

Although the rank-order segregation indices, and in particular \mathcal{H}^R , take into account the ordering of the income categories, they neglect their relative magnitude, thus discarding relevant information.⁴ One way to take into account both the ordering of income categories and their relative magnitude would be to use the income variation between schools to measure income segregation. Jargowsky [12] pioneered this approach and proposed an index which uses the variance as a measure of income variation. We introduce this and other income segregation indices that follow this approach in the following section.

4 Segregation and Inequality

4.1 Inequality Indices

Income segregation is related to income inequality in two ways. On the one hand, the higher the income inequality of a district is, the higher the potential for income segregation in it. On the other hand, *ceteris paribus*, for any given level of income inequality of a district, the more economically diverse are its schools, the lower its income segregation. Given this relation, before we define measures of income segregation, we need to introduce indices of income inequality.

An *inequality index* I assigns to each school c a real number, $I(c)$ which is meant to capture its level of income inequality. The following are examples of prominent income inequality indices. The first one consists of the class of generalized entropy

⁴Another serious drawback is that that they are not continuous in the income distribution. See Section 6 for details.

indices. The second is the variance.

Example 1 For $\alpha \in \mathbb{R}$ the *Generalized Entropy index*, $I^\alpha : \mathcal{C}_+ \rightarrow [0, \infty)$, is defined as follows: For all $c = \langle (n_1, y_1), \dots, (n_G, y_G) \rangle \in \mathcal{C}_+$,

$$I^\alpha(c) = \begin{cases} \frac{1}{\alpha(\alpha-1)} \sum_{g=1}^G \frac{n_g}{n_c} \left[\left(\frac{y_g}{\mu_c} \right)^\alpha - 1 \right] & \text{if } \alpha \notin \{0, 1\} \\ \sum_{g=1}^G \frac{n_g}{n_c} \ln \left(\frac{\mu_c}{y_g} \right) & \text{if } \alpha = 0 \\ \sum_{g=1}^G \frac{n_g y_g}{|c|} \ln \left(\frac{y_g}{\mu_c} \right) & \text{if } \alpha = 1 \end{cases}$$

When $\alpha = 0$, the associated generalized entropy index I^0 is known as *Theil's second measure of income inequality*.

Example 2 The *variance* assigns to each school the variance of its income distribution. Formally, var , is defined as follows: For all $c = \langle (n_1, y_1), \dots, (n_G, y_G) \rangle$,

$$var(c) = \frac{1}{n_c} \sum_{g=1}^G n_g (y_g - \mu_c)^2.$$

We shall sometimes speak of the income inequality in the whole district, and in order to measure it we will apply an inequality index to the combination of all its schools into a single school. In particular, with a slight abuse of notation for any district $X = \{c_1, \dots, c_K\}$, we will write $I(X)$ for $I(c_1 \# \dots \# c_K)$, the inequality of the district's income distribution.

4.2 Segregation orders and indices

A segregation order defined on \mathcal{D}_+ is a complete and transitive relation \succeq on \mathcal{D}_+ . An *income segregation index*, or segregation index for short, \mathcal{S} assigns to each district, X a real number, $\mathcal{S}(X)$ which is meant to capture its level of segregation. We shall

maintain the convention of using calligraphic capital letters to denote segregation indices. A segregation index represents a segregation order \succeq if for any two districts X, Y , $X \succeq Y \Leftrightarrow \mathcal{S}(X) \geq \mathcal{S}(Y)$. The following are examples of income segregation indices. For any district $X = \{c_1, \dots, c_K\}$,

- the *school separation index* is defined by

$$\mathcal{S}\mathcal{S}\mathcal{I}(X) = \sum_{c \in X} \frac{n_c}{n_X} \ln \left(\frac{\mu_X}{\mu_c} \right);$$

- the *Variance segregation index* is defined by

$$\mathcal{V}(X) = \frac{1}{n_X} \sum_{c \in X} n_c (\mu_c - \mu_X)^2;$$

- *Jargowsky's neighborhood sorting index* is defined (on the class of districts whose income distribution has positive variance) by

$$\mathcal{N}\mathcal{S}\mathcal{I}(X) = \sqrt{\frac{\mathcal{V}(X)}{\text{var}(X)}}.$$

And in general, given any income inequality index I ,

- the *segregation index induced by I* is defined by

$$\mathcal{I}(X) = I(X) - \sum_{c \in X} \frac{n_c}{n_X} I(c).$$

To understand the idea behind the last class of indices, note that the sum $\sum_{c \in X} \frac{n_c}{n_X} I(c)$ is an average of the level of income inequality, as measured by I , within the schools of X , and can be seen as a measure of the economic diversity of such schools. Clearly, this diversity cannot contribute to the segregation of X . Thus, the segregation of X as measured by \mathcal{I} is what remains from the district's income inequality after we deduct the economic diversity exhibited by the schools.

An interesting feature of the segregation index induced by I is that when each school has zero income variation, the district's segregation coincides with its income inequality as measured by I , and as a result the higher the income inequality is, the higher the district's segregation is. This implies that the segregation index induced by I is not “pure” in the sense that it does not satisfy the requirement of mentioned in the introduction that segregation be maximal when schools exhibit no diversity. Nevertheless, one could use the segregation index induced by I to define a measure of “pure” segregation by the ratio $\mathcal{I}(X)/I(X)$. With this definition we see that the segregation index induced by an inequality index I is in fact the product of an index of pure segregation and the associated income inequality of the district.

Interestingly, the segregation index induced by the variance, var , is \mathcal{V} .⁵ Also the segregation index induced by the generalized entropy index I^0 is \mathcal{SSI} . To see this, note that

$$\begin{aligned}
I^0(X) &= I^0(c_1 \# \dots \# c_K) \\
&= \sum_{c \in X} \sum_{g \in G(c)} \frac{n_g}{n_X} \ln \left(\frac{\mu_X}{y_g} \right) \\
&= \sum_{c \in X} \frac{n_c}{n_X} \sum_{g \in G(c)} \frac{n_g}{n_c} \ln \left(\frac{\mu_X}{\mu_c} \frac{\mu_c}{y_g} \right) \\
&= \sum_{c \in X} \frac{n_c}{n_X} \ln \left(\frac{\mu_X}{\mu_c} \right) + \sum_{c \in X} \frac{n_c}{n_X} \sum_{g \in G(c)} \frac{n_g}{n_c} \ln \left(\frac{\mu_c}{y_g} \right) \\
&= \mathcal{SSI}(X) + \sum_{c \in X} \frac{n_c}{n_X} I^0(c),
\end{aligned}$$

which, after a rearrangement of terms, yields the desired result. The proof that \mathcal{V} is the segregation index induced by the variance is similar. Lastly, although the entropy function is not an income inequality index, one can regard the Mutual Information index of ethnic segregation as the segregation index induced by the entropy (see

⁵This implies that Jargowsky's \mathcal{NSI} is ordinally equivalent to the pure segregation index induced by var .

equation 1).

5 Axioms

We now list several desirable properties of an income segregation order. We start with two fundamental axioms that convey the basic idea of what it means for a district to become less segregated. Recall that for any district X , $C(X)$ is its completely integrated version since it is obtained by combining all the schools in X into a single one. On the other hand, $d(X)$ can be interpreted as the completely segregated version of X since it is obtained by dividing its schools into single-income-group schools. The first axiom imposes two requirements. One is that if all the schools of a given district are combined into a single school segregation does not increase. The second is that the completely segregated and the completely integrated versions of a district are equally segregated if and only if all pupils have the same income.

Single-School Property (SSP) For any district X , $X \succeq C(X)$. Furthermore, if $d(X) = \{\langle(n_1, y_1)\rangle, \dots, \langle(n_K, y_K)\rangle\}$, then $d(X) \sim C(X)$ if and only if $y_1 = \dots = y_K$.

SSP is a very weak axiom. It places a restriction on \succeq only when comparing a district or its most segregated version to their completely integrated version. It says nothing about districts with different distributions of pupils across income groups. Also, it seems a very natural requirement for any segregation measure to satisfy. It is difficult to imagine segregation increasing when all the district's pupils are sent to the same school. And it is equally difficult to imagine segregation staying the same if that school is divided into many schools, each consisting of a single income group.

The next axiom complements the single-school property by requiring that all single-school districts be equally segregated. As a result, minimal segregation is attained by the single-school districts.

Equivalence of Single-School Districts (ESSD) If X and Y are single-school districts, then $X \sim Y$.

Note that when restricted to the class of simple districts, neither of the above two axioms have any bite. This is so because, except for trivial cases, single-school districts are not simple districts.

Next, we list two axioms that require invariance to certain changes in units of measurement. The first one states that changes in population that leave the relative attendances of the schools unchanged do not affect segregation.

Population Homogeneity (PH) For any district X and scalar $\lambda > 0$, $X \sim \lambda X$.

The next axiom states that changes in household incomes that keep the students' relative incomes unchanged do not affect segregation.

Income Homogeneity (IH) For any district X and scalar $\lambda > 0$, $X \sim X * \lambda$.

It can be easily checked that \mathcal{SSI} satisfies this and the previous axioms.

The next two axioms impose segregation comparisons be independent of irrelevant sub-districts. To motivate the first one, consider a school district partitioned into two sub-districts. Suppose that a reorganization within each sub-district reduces segregation in every one of them. It stands to reason that such reorganization does not result in a higher districtwide segregation. Otherwise we would be witnessing the outcome of a rather perverse policy. If we believe that no such policies exist, the index of segregation must satisfy the following.

Independence (IND) For any three districts X, Y, Z such that $|X| = |Y|$ and $n_X = n_Y$, $X \succeq Y \Leftrightarrow X \uplus Z \succeq Y \uplus Z$.

Independence is an eminently reasonable requirement. It guarantees that any policy that reduces segregation in one sub-district does not result in a higher districtwide segregation. Versions of this axiom appear in several contexts. For instance,

Shorrocks's [23] and Foster and Shorrocks's [4] subgroup consistency axioms are essentially the independence axiom in the context of income inequality and poverty measurement, respectively. Hutchens [8] and Frankel and Volij [5] use variations of this axiom in their characterizations of ethnic segregation measures.

Both \mathcal{SSI} and \mathcal{V} satisfy IND. To see that this is so for \mathcal{SSI} , let X, Y , and Z be three districts as described in the axiom. Denoting $n_X = n_Y = n$ and $\mu_X = \mu_Y = \mu$, and taking into account that $n_{X \uplus Z} = n_{Y \uplus Z}$ and $\mu_{X \uplus Z} = \mu_{Y \uplus Z}$, we have

$$\begin{aligned}
\mathcal{SSI}(X \uplus Z) \geq \mathcal{SSI}(Y \uplus Z) &\Leftrightarrow \sum_{c \in X \uplus Z} \frac{n_c}{n_{X \uplus Z}} \ln \left(\frac{\mu_{X \uplus Z}}{\mu_c} \right) \geq \sum_{c' \in Y \uplus Z} \frac{n_{c'}}{n_{Y \uplus Z}} \ln \left(\frac{\mu_{Y \uplus Z}}{\mu_{c'}} \right) \\
&\Leftrightarrow \sum_{c \in X} \frac{n_c}{n_{X \uplus Z}} \ln \left(\frac{\mu_{X \uplus Z}}{\mu_c} \right) \geq \sum_{c' \in Y} \frac{n_{c'}}{n_{Y \uplus Z}} \ln \left(\frac{\mu_{Y \uplus Z}}{\mu_{c'}} \right) \\
&\Leftrightarrow \sum_{c \in X} \frac{n_c}{n} \ln \left(\frac{\mu_{X \uplus Z}}{\mu} \frac{\mu}{\mu_c} \right) \geq \sum_{c' \in Y} \frac{n_{c'}}{n} \ln \left(\frac{\mu_{Y \uplus Z}}{\mu} \frac{\mu}{\mu_{c'}} \right) \\
&\Leftrightarrow \ln \left(\frac{\mu_{X \uplus Z}}{\mu} \right) + \sum_{c \in X} \frac{n_c}{n} \ln \left(\frac{\mu}{\mu_c} \right) \geq \ln \left(\frac{\mu_{Y \uplus Z}}{\mu} \right) + \sum_{c' \in Y} \frac{n_{c'}}{n} \ln \left(\frac{\mu}{\mu_{c'}} \right) \\
&\Leftrightarrow \sum_{c \in X} \frac{n_c}{n} \ln \left(\frac{\mu}{\mu_c} \right) \geq \sum_{c' \in Y} \frac{n_{c'}}{n} \ln \left(\frac{\mu}{\mu_{c'}} \right) \\
&\Leftrightarrow \mathcal{SSI}(X) \geq \mathcal{SSI}(Y).
\end{aligned}$$

The reader can verify that \mathcal{V} also satisfies IND. Jargowsky's \mathcal{NSI} , on the other hand, does not satisfy independence. To see this, consider the following districts: $X = \{\langle(10, 3), (20, 6)\rangle, \langle(20, 2), (30, 4)\rangle\}$, and $Y = \{\langle(10, 3), (20, 5)\rangle, \langle(20, 3), (30, 4)\rangle\}$. Both districts have a population of 80 and an income of 310. It can be checked that $\mathcal{NSI}(X) = 3/5 > 0.454794 = \mathcal{NSI}(Y)$. Now, if we append district $Z = \{\langle(20, 2), (10, 3)\rangle, \langle(25, 4), (5, 6)\rangle\}$ both to X and Y we obtain that $\mathcal{NSI}(X \uplus Z) = 0.7064 < 0.750879 = \mathcal{NSI}(Y \uplus Z)$.

Though similar, the next axiom is different from independence. Consider a district composed of two sub-districts $X \uplus Y$ and assume that a policy is applied to Y that leaves its attendance unchanged. The axiom states that whether or not this policy

increases districtwide segregation does not depend on the segregation within sub-district X . As it was the case with SSP and ESSD, this axiom is toothless when restricted to simple districts since it deals with comparisons involving districts that are not simple.

Separability (SEP) For any three districts X, Y, Z such that $n_Y = n_Z$, $X \uplus Y \succeq X \uplus Z \Leftrightarrow C(X) \uplus Y \succeq C(X) \uplus Z$.

Any segregation index induced by an inequality index, in particular \mathcal{SSI} and \mathcal{V} , satisfies separability. To see this consider three districts as described in the axiom, and let's denote $n = n_X + n_Y = n_X + n_Z$. Then,

$$\begin{aligned}
\mathcal{I}(X \uplus Y) \geq \mathcal{I}(X \uplus Z) &\Leftrightarrow I(X \uplus Y) - \sum_{c \in X \uplus Y} \frac{n_c}{n} I(c) \geq I(X \uplus Z) - \sum_{c \in X \uplus Z} \frac{n_c}{n} I(c) \\
&\Leftrightarrow I(X \uplus Y) - \sum_{c \in Y} \frac{n_c}{n} I(c) \geq I(X \uplus Z) - \sum_{c \in Z} \frac{n_c}{n} I(c) \\
&\Leftrightarrow I(C(X) \uplus Y) - \sum_{c \in C(X) \uplus Y} \frac{n_c}{n} I(c) \geq I(C(X) \uplus Z) - \sum_{c \in C(X) \uplus Z} \frac{n_c}{n} I(c) \\
&\Leftrightarrow \mathcal{I}(C(X) \uplus Y) \geq \mathcal{I}(C(X) \uplus Z)
\end{aligned}$$

The last axiom requires that similar districts have similar levels of segregation. It will allow us to find a continuous representation of an order that satisfies the previous axioms.

Continuity (CONT) Let $X = \{c_1, \dots, c_K\}$ be a district and let $X^n = \{c_1^n, \dots, c_K^n\}$, for $n = 1, 2, \dots$ be a sequence of districts such that $c_k^n \rightarrow c_k$ for $k = 1, \dots, K$. For any district Y , if $X^n \succeq Y$ for all n , then $X \succeq Y$, and if $Y \succeq X^n$ for all n , then $Y \succeq X$.

It is worth noticing that whereas most of the indices mentioned above satisfy this axiom, the rank-order information theory segregation index, \mathcal{H}^r does not, the reason being that small changes in a household income may induce a change in this

household's income rank. Consider, for instance, the following family of districts: $X(y) = \{\langle(20, 2), (30, 4)\rangle, \langle(10, y), (20, 6)\rangle\}$. When $y \in (2, 4)$, the income group $(10, y)$ represents the lower middle class, and when $y \in (4, 6)$, the income group $(10, y)$ represents the upper middle class. Now, let X and X' be the following districts:

$$X = \{\langle(10, 1), (10, 3)\rangle, \langle(20, 2)\rangle\}, \quad X' = \{\langle(5, 1), (5, 3)\rangle, \langle(20, 2)\rangle\}.$$

It can be checked that for any $y \in (2, 4)$ and $y' \in (4, 6)$

$$\mathcal{H}^r(X(y)) < \mathcal{H}^r(X) < \mathcal{H}^r(X(4)) < \mathcal{H}^r(X') < \mathcal{H}^r(X(y'))$$

which shows that the order represented by \mathcal{H}^r does not satisfy continuity.

6 An ordinal characterization of \mathcal{SSI}

We now state our main result.

Theorem 1 Let \succeq be a segregation order on \mathcal{D}_+ . It satisfies the single-school property, equivalence of single-school districts, independence, separability, population homogeneity, income homogeneity, and continuity if and only if it is represented by the school separation index. Namely, for all districts X, Y , $X \succeq Y \Leftrightarrow \mathcal{SSI}(X) \geq \mathcal{SSI}(Y)$.

Proof : As was shown earlier, the order represented by \mathcal{SSI} satisfies all the axioms listed in Theorem 1. We now show that the only order that satisfies this list is \mathcal{SSI} . Let \succeq be an order that satisfies all the axioms listed in Theorem 1. The proof consists of four steps. First we build an index \mathcal{S} that represents \succeq . Second we prove that \mathcal{S} satisfies a very strong separability property. Third, we show that when restricted to the family of simple districts, \mathcal{S} satisfies several properties and as a result it has a

particular form. Lastly, we show that the only extension of this restricted index to the class of all districts, if it is to satisfy all the axioms, is \mathcal{SSI} .

We first show that as long as a segregation order satisfies independence, equivalence of single-school districts and the single school property, merging schools with the same distribution of income does not affect segregation.

Claim 1 Let \succeq be a segregation order that satisfies the single-school property, equivalence of single-school districts and independence. Then, for any district X and for any $\alpha, \beta > 0$, we have $\alpha X \uplus \beta X \sim (\alpha + \beta)X$.

Proof : Let $X = \{c_1, \dots, c_K\}$ be a district and let $\alpha, \beta > 0$.

$$\begin{aligned} \alpha X \uplus \beta X &= \{\alpha c_1, \dots, \alpha c_K\} \uplus \{\beta c_1, \dots, \beta c_K\} \quad \text{by definition} \\ &\sim \{\alpha \bar{c}_1, \dots, \alpha \bar{c}_K\} \uplus \{\beta \bar{c}_1, \dots, \beta \bar{c}_K\} \quad \text{by ESSD and IND} \\ &\sim \uplus_{i=1}^K \{\alpha \bar{c}_i, \beta \bar{c}_i\}. \end{aligned}$$

By SSP we have that $\{\alpha \bar{c}_i, \beta \bar{c}_i\} \sim \{(\alpha + \beta) \bar{c}_i\}$ for $i = 1, \dots, K$. Therefore, by IND and ESSD

$$\begin{aligned} \uplus_{i=1}^K \{\alpha \bar{c}_i, \beta \bar{c}_i\} &\sim \uplus_{i=1}^K \{(\alpha + \beta) \bar{c}_i\} \\ &\sim \uplus_{i=1}^K \{(\alpha + \beta) c_i\} = (\alpha + \beta)X. \end{aligned}$$

□

We now start building the index. Let \mathcal{D}_1 denote the class of districts X with $n_X = |X| = 1$. Also let $X_0 = \{\langle(1, 1)\rangle\}$ be the district with a single school which has a single student with income 1. Note that by SSP and ESSD, $X \succeq X_0$ for all districts X .

Lemma 1 Let $X' \in \mathcal{D}_1$ be a district such that $X' \succ X_0$. If $0 \leq \alpha < \beta < 1$, then $\beta X' \uplus (1 - \beta)X_0 \succ \alpha X' \uplus (1 - \alpha)X_0$.

Proof : By PH, $(\beta - \alpha)X' \succ (\beta - \alpha)X_0$. By IND,

$$\alpha X' \uplus (\beta - \alpha)X' \uplus (1 - \beta)X_0 \succ \alpha X' \uplus (\beta - \alpha)X_0 \uplus (1 - \beta)X_0.$$

By Claim 1 and IND, $\beta X' \uplus (1 - \beta)X_0 \succ \alpha X' \uplus (1 - \alpha)X_0$. \square

Lemma 2 Let X' be a district in \mathcal{D}_1 such that $X' \succ X_0$. For any district X such that $X' \succeq X$, there is a unique $\alpha' \in [0, 1]$ such that $X \sim \alpha' X' \uplus (1 - \alpha')X_0$.

Proof : The sets $\{\alpha \in [0, 1] : \alpha X' \uplus (1 - \alpha)X_0 \succeq X\}$ and $\{\alpha \in [0, 1] : X \succeq \alpha X' \uplus (1 - \alpha)X_0\}$ are closed by CONT. Since $X' \succeq X \succeq X_0$, they are not empty. Since \succeq is complete, their union is $[0, 1]$. Therefore, since the unit interval is connected, the intersection of the two sets is not empty. By Lemma 1, this intersection must contain a single element. This single element is the α' we are looking for. \square

Lemma 3 Let X' and X'' be two districts in \mathcal{D}_1 such that $X'' \succeq X' \succ X_0$. Let X be a district such that $X' \succeq X$, and let α' and α'' be the unique numbers identified in Lemma 2 defined, respectively, by

$$X \sim \alpha' X' \uplus (1 - \alpha')X_0 \quad \text{and} \quad X \sim \alpha'' X'' \uplus (1 - \alpha'')X_0.$$

Let β be the unique number identified in Lemma 2 such that $X' \sim \beta X'' \uplus (1 - \beta)X_0$. Then, $\alpha'' = \alpha'\beta$.

Proof : By definition of α' and IND, $X \sim \alpha'(\beta X'' \uplus (1 - \beta)X_0) \uplus (1 - \alpha')X_0$. Therefore, by Claim 1, $X \sim \alpha'\beta X'' \uplus (1 - \alpha'\beta)X_0$. \square

We can now proceed to the definition of a segregation index. Fix the following district: $X_{1/2} = \{\langle(1/2, 1/2)\rangle, \langle(1/2, 3/2)\rangle\}$. Let X be a district, and let $X' \in \mathcal{D}_1$ be

a district that satisfies $X' \succeq X$ and $X' \succeq X_{1/2}$. Let α' and β' be the unique numbers identified in Lemma 2 that satisfy

$$X \sim \alpha' X' \uplus (1 - \alpha') X_0 \quad \text{and} \quad X_{1/2} \sim \beta' X' \uplus (1 - \beta') X_0.$$

Note that by SSP, $X_{1/2} \succ X_0$ and as a result $\beta' > 0$. We can thus assign to every district X the number α'/β' . It turns out that this number does not depend on the choice of X' . Indeed, let $X'' \in \mathcal{D}_1$ be another district such that $X'' \succeq X$ and $X'' \succeq X_{1/2}$ and let α'' and β'' be defined by

$$X \sim \alpha'' X'' \uplus (1 - \alpha'') X_0 \quad \text{and} \quad X_{1/2} \sim \beta'' X'' \uplus (1 - \beta'') X_0.$$

Assume without loss of generality that $X'' \succeq X'$. Let δ be defined by

$$X' \sim \delta X'' \uplus (1 - \delta) X_0.$$

By Lemma 3, $\alpha'' = \alpha' \delta$ and $\beta'' = \beta' \delta$. Therefore, $\alpha'/\beta' = \alpha''/\beta''$.

The above discussion allows us to define the segregation index \mathcal{S} by $\mathcal{S}(X) = \alpha'/\beta'$, where α'/β' is the ratio built above.

Lemma 4 The index \mathcal{S} represents the segregation order \succeq .

Proof : Let X and X' be two districts and assume that $X' \succ X$. Let $X'' \in \mathcal{D}_1$ be a district such that $X'' \succeq X'$ and $X'' \succeq X_{1/2}$. Let α and α' be defined by

$$\begin{aligned} X &\sim \alpha X'' \uplus (1 - \alpha) X_0 \\ X' &\sim \alpha' X'' \uplus (1 - \alpha') X_0. \end{aligned}$$

By Lemma 1, $\alpha' > \alpha$ which implies that $\mathcal{S}(X') > \mathcal{S}(X)$. □

The following property follows from the way the index was constructed and by ESSD. The proof is left to the reader.

Claim 2 For any district X , $\mathcal{S}(X) \geq 0$. Furthermore, if X is a single-school district then $\mathcal{S}(X) = 0$.

We now start the second part of the proof. The next proposition shows that the index \mathcal{S} satisfies a very strong separability property.

Proposition 1 Let X and X' be two districts. Then

$$\mathcal{S}(X \uplus X') = \frac{n_X}{n_{X \uplus X'}} \mathcal{S}(X) + \mathcal{S}(C(X) \uplus X').$$

Proof : Let X and X' be two districts with populations $n_X = n$ and $n_{X'} = m$, respectively. By PH, we can assume without loss of generality that $n + m = 1$. Since \succeq satisfies IH, we can also assume without loss of generality that $|X \uplus X'| = 1$. Let X'' be a district in \mathcal{D}_1 such that $X'' \succeq X$, $X'' \succeq X \uplus X'$ and $X'' \succeq X_{1/2}$, and let α, γ and δ be such that

$$X \sim \alpha X'' \uplus (1 - \alpha) X_0 \tag{4}$$

$$C(X) \uplus X' \sim \gamma X'' \uplus (1 - \gamma) X_0 \tag{5}$$

$$X \uplus X' \sim \delta X'' \uplus (1 - \delta) X_0. \tag{6}$$

Then, $\mathcal{S}(X) = \alpha/\beta$, $\mathcal{S}(C(X) \uplus X') = \gamma/\beta$ and $\mathcal{S}(X \uplus X') = \delta/\beta$ for some $\beta > 0$. To prove the result it is enough to show that $\delta = n\alpha + \gamma$.

Denote $X_0^* = \left\{ \left\langle \left(n, \frac{|X|}{n} \right) \right\rangle \right\}$. This district has the same population and income as X and it is obtained from X_0 by multiplying its population by n and the income of each pupil by $|X|/n$. Recall that \succeq satisfies IH and denote

$$X^* = nX'' * (|X|/n).$$

This district has the same population and income as X . It is obtained from X'' by multiplying its population by n and by multiplying the income of each pupil by $|X|/n$. It follows from 4, using PH and IH, that

$$X \sim \alpha X^* \uplus (1 - \alpha)X_0^* \quad (7)$$

Choose $k \in \mathbb{N}$ such that $k > n + \gamma$. By concatenating $(k - 1)X_0$ to both sides of equation 5, we obtain

$$\begin{aligned} C(X) \uplus \overbrace{X' \uplus (k - 1)X_0}^Y &\sim \gamma X'' \uplus (k - \gamma)X_0 && \text{by IND and Claim 1} \\ &\sim \frac{\gamma}{n}X^* \uplus \frac{k - \gamma}{n}X_0^* && \text{by IH} \\ &\sim \frac{\gamma}{n}X^* \uplus \frac{k - \gamma}{n}C(X) && \text{by ESSD and IND} \\ &\sim C(X) \uplus \overbrace{\frac{\gamma}{n}X^* \uplus \left(\frac{k - \gamma}{n} - 1\right)C(X)}^Z && \text{by Claim 1.} \\ &&& \underbrace{\hspace{1.5cm}}_{>0} \end{aligned}$$

Note that since $k > n + \gamma$ sub-district Z is well-defined. Since $n_Y = n_Z = m + (k - 1)$, by SEP,

$$X \uplus \overbrace{X' \uplus (k - 1)X_0}^Y \sim X \uplus \overbrace{\frac{\gamma}{n}X^* \uplus \left(\frac{k - \gamma}{n} - 1\right)C(X)}^Z.$$

By equations 6, 7 and IND,

$$\begin{aligned} \delta X'' \uplus (1 - \delta)X_0 \uplus (k - 1)X_0 &\sim \alpha X^* \uplus (1 - \alpha)X_0^* \uplus \frac{\gamma}{n}X^* \uplus \left(\frac{k - \gamma}{n} - 1\right)C(X) \\ &\sim \alpha X^* \uplus (1 - \alpha)X_0^* \uplus \frac{\gamma}{n}X^* \uplus \left(\frac{k - \gamma}{n} - 1\right)X_0^* \\ &\sim n\alpha X'' \uplus n(1 - \alpha)X_0 \uplus \gamma X'' \uplus (k - \gamma - n)X_0 \end{aligned}$$

where the second line follows from IND and the last one from IH. Applying Claim 1

to both sides, we obtain

$$\delta X'' \uplus (k - \delta)X_0 \sim (n\alpha + \gamma)X'' \uplus (k - \gamma - n\alpha)X_0.$$

By PH and Lemma 2, we conclude that $\delta = n\alpha + \gamma$. \square

Corollary 1 Let X_1, \dots, X_J be J districts. Then

$$\mathcal{S}(\uplus_{j=1}^J X_j) = \sum_{j=1}^J \frac{n_{X_j}}{n_X} \mathcal{S}(X_j) + \mathcal{S}(\uplus_{j=1}^J C(X_j)) \quad (8)$$

Proof : See Appendix. \square

Corollary 2 For any district $X = \{c_1, \dots, c_K\}$,

$$\mathcal{S}(X) = \mathcal{S}(d(c_1 \# \dots \# c_K)) - \sum_{k=1}^K \frac{n_{c_k}}{n_X} \mathcal{S}(d(c_k)).$$

Proof : By Corollary 1, $\mathcal{S}(\uplus_{k=1}^K d(c_k)) = \mathcal{S}(\uplus_{c=c_k}^K C(d(c_k))) + \sum_{k=1}^K \frac{n_{d(c_k)}}{n_X} \mathcal{S}(d(c_k))$. For any school c_k , the districts $\{c_k\}$ and $C(d(c_k))$ are identical, hence the equality $X = \uplus_{c=1}^K C(d(c_k))$. Noting that $n_{c_k} = n_{d(c_k)}$, and $d(c_1 \# \dots \# c_K) = \uplus_{k=1}^K d(c_k)$, rearranging yields the desired result. \square

We now start the third part of the proof. We will show that, restricted to the class of simple districts, \mathcal{S} has a very particular form. Let I be the inequality index defined by $I(c) = \mathcal{S}(d(c))$.

We now show that I is a monotone transformation of a member of the generalized entropy family defined in Example 1. The proof is based on Theorem 5 in Shorrocks

[22, p. 1381]. In order to apply it, we will show that the inequality index $I(c)$ satisfies the following properties.

Anonymity For all permutations $\pi : G \rightarrow G$, $I((n_g, y_g)_{g \in G}) = I((n_{\pi(g)}, y_{\pi(g)})_{g \in G})$.

The reason is that $\langle (n_g, y_g)_{g \in G} \rangle = \langle (n_{\pi(g)}, y_{\pi(g)})_{g \in G} \rangle$.

Normalization For any school $c = \langle (n_g, y_g) \rangle_{g \in G}$, we have that $I(\bar{c}) = 0$. Indeed,

$I(\bar{c}) = \mathcal{S}(d(\bar{c})) = 0$ where the last equality follows from Claim 2.

Replication invariance For any c , we have that $I(c) = I(c \# c)$. To see this, note

that by PH and Claim 1, $\mathcal{S}(d(c)) = \mathcal{S}(2d(c)) = \mathcal{S}(d(c) \uplus d(c)) = \mathcal{S}(d(c \# c))$.

Homogeneity For any $\alpha > 0$ and any c , $I(c * \alpha) = I(c)$. Indeed, by IH, $\mathcal{S}(d(c * \alpha)) =$

$\mathcal{S}(d(c) * \alpha) = \mathcal{S}(d(c))$.

Aggregativity There is a continuous aggregator $A : R \rightarrow \mathbb{R}$ for some subset $R \subset \mathbb{R}_+^6$

such that for all schools c, c'

$$I(c \# c') = A(I(c), n_c, \mu_c, I(c'), n_{c'}, \mu_{c'}). \quad (9)$$

Furthermore this aggregator is increasing in its first and fourth arguments. Indeed, consider the function $A : R \rightarrow \mathbb{R}$ defined by $A(x, n, \mu, y, m, \nu) = \mathcal{S}(X \uplus Y)$ for some districts X, Y such that $\mathcal{S}(X) = x$, $n_X = n$, $\mu_X = \mu$, and $\mathcal{S}(Y) = y$, $n_Y = m$, $\mu_Y = \nu$. This function is well defined. Indeed, if we let Z and W be two districts such that $(\mathcal{S}(W), n_W, \mu_W) = (\mathcal{S}(X), n_X, \mu_X)$ and $(\mathcal{S}(Z), n_Z, \mu_Z) = (\mathcal{S}(Y), n_Y, \mu_Y)$, by IND applied twice, $\mathcal{S}(X \uplus Y) = \mathcal{S}(X \uplus Z) = \mathcal{S}(W \uplus Z)$. To see that the aggregator A is increasing in its first argument note that by IND, $\mathcal{S}(W \uplus Y) > \mathcal{S}(X \uplus Y)$ whenever $\mathcal{S}(W) > \mathcal{S}(X)$ and $(n_W, \mu_W) = (n_X, \mu_X)$. A similar argument shows that A is increasing in its fourth argument.

To see that equation 9 holds, note that $I(c \# c') = \mathcal{S}(d(c \# c')) = \mathcal{S}(d(c) \uplus d(c'))$

and that by definition of the aggregator A

$$\mathcal{S}(d(c) \uplus d(c')) = A(\mathcal{S}(d(c)), n_c, \mu_c, \mathcal{S}(d(c')), n_{c'}, \mu_{c'}) = A(I(c), n_c, \mu_c, I(c'), n_{c'}, \mu_{c'}).$$

The next proposition show that I satisfies the Pigou-Dalton principle of transfers. Namely, if school c is obtained from school c' by means of a progressive transfer, then $I(c) < I(c')$. Formally,

Proposition 2 For any two schools $c = \langle (n_1, y_1), (n_2, y_2) \rangle$ and $c' = \langle (n_1, y_1 - \Delta/n_1), (n_2, y_2 + \Delta/n_2) \rangle$ such that $0 < y_1 \leq y_2$ and $\Delta \in (0, n_1 y_1)$, we have that $I(c') > I(c)$.

Proof : Let $b_1 = \frac{n_2 \Delta}{(n_1 + n_2) \Delta + n_1 n_2 (y_2 - y_1)}$ and $b_2 = \frac{n_1 \Delta}{(n_1 + n_2) \Delta + n_1 n_2 (y_2 - y_1)}$, and consider the following subdivision of school $c_1 = \langle (n_1, y_1) \rangle$ into

$$c_{11} = \langle ((1 - b_1)n_1, y_1 - \Delta/n_1) \rangle \text{ and } c_{12} = \langle (b_1 n_1, y_2 + \Delta/n_2) \rangle$$

Since $n_1 = n_{c_{11}} + n_{c_{12}}$ and $|c_1| = |c_{11}| + |c_{12}|$, this subdivision is feasible. By SSP we have that $\mathcal{S}(\{c_{11}, c_{12}\}) > \mathcal{S}(C(\{c_{11}, c_{12}\}))$. Since by Claim 2, $\mathcal{S}(C(\{c_{11}, c_{12}\})) = 0 = \mathcal{S}(\{c_1\})$ we obtain that $\mathcal{S}(\{c_{11}, c_{12}\}) > \mathcal{S}(\{c_1\})$. Similarly, if we subdivide school $c_2 = \langle (n_2, y_2) \rangle$ into the following two schools

$$c_{21} = \langle (b_2 n_2, y_1 - \Delta/n_1) \rangle, \quad c_{22} = \langle ((1 - b_2)n_2, y_2 + \Delta/n_2) \rangle$$

we obtain that $\mathcal{S}(\{c_{21}, c_{22}\}) > \mathcal{S}(\{c_2\})$. Therefore, by IND

$$\begin{aligned} \mathcal{S}(d(c)) = \mathcal{S}(\{c_1\} \uplus \{c_2\}) &< \mathcal{S}(\{c_{11}, c_{12}\} \uplus \{c_{21}, c_{22}\}) \\ &= \mathcal{S}(\{c_{11}, c_{21}\} \uplus \{c_{12}, c_{22}\}). \end{aligned} \quad (10)$$

Since $n_{c_{11}} + n_{c_{21}} = n_1$ and $n_{c_{12}} + n_{c_{22}} = n_2$, we have that $C(\{c_{11}, c_{21}\}) =$

$\{\langle(n_1, y_1 - \Delta/n_1)\rangle\}$ and $C(\{c_{12}, c_{22}\}) = \{\langle(n_2, y_2 + \Delta/n_2)\rangle\}$.

By SSP, $\mathcal{S}(\{c_{11}, c_{21}\}) = \mathcal{S}(C(\{c_{11}, c_{21}\})) = \mathcal{S}(\{\langle(n_1, y_1 - \Delta/n_1)\rangle\})$ and $\mathcal{S}(\{c_{12}, c_{22}\}) = \mathcal{S}(C(\{c_{12}, c_{22}\})) = \mathcal{S}(\{\langle(n_2, y_2 + \Delta/n_2)\rangle\})$. By IND

$$\begin{aligned} \mathcal{S}(\{c_{11}, c_{21}\} \uplus \{c_{12}, c_{22}\}) &= \mathcal{S}(\{\langle(n_1, y_1 - \Delta/n_1)\rangle\} \uplus \{\langle(n_2, y_2 + \Delta/n_2)\rangle\}) \\ &= \mathcal{S}(d(c')). \end{aligned} \tag{11}$$

From inequalities 10 and 11 we obtain that $I(c) = \mathcal{S}(d(c)) < \mathcal{S}(d(c')) = I(c')$ which is what we wanted to show. \square

Finally, the next proposition uses the fact that I satisfies the Pigou-Dalton principle to show that it is continuous.

Proposition 3 For all $c = \langle(n_g, y_g)\rangle_{g \in G}$, the value $I(c)$ depends continuously on its arguments (n_g, y_g) .

Proof: Let $c = \langle(n_1, y_1), \dots, (n_G, y_G)\rangle$ be a school and let $c^k = \langle(n_1^k, y_1^k), \dots, (n_G^k, y_G^k)\rangle$, for $k = 1, 2, \dots$ be a sequence of schools that converges to c . We need to show that $I(c^k) \rightarrow I(c)$. We can assume without loss of generality that $|c| = |c^k| = 1$ and $n_c = n_{c^k} = 1$, for $k = 1, 2, \dots$. Indeed, since \succeq satisfies IH we can define \hat{c} , and \hat{c}^k to be the schools that are obtained from c and c^k , respectively by normalizing both their attendance and income to be one as follows: $\hat{c} = (1/n_c) c * (n_c/|c|)$ and $\hat{c}^k = (1/n_{c^k}) c^k * (n_{c^k}/|c^k|)$. Since $c^k \rightarrow c$ we have that $\{\hat{c}^k\} \rightarrow \hat{c}$. By PH and by IH, $I(\hat{c}^k) = I(c^k)$ and $I(\hat{c}) = I(c)$ for all k .

Let $n = \min\{n_1, \dots, n_G, 1/2\}$ and $y = \min\{y_1, \dots, y_G, 1/2\}$. Let $\varepsilon \in (0, \min\{n, y\})$ and let k_0 be such that for all $k > k_0$, $\|c^k - c\| < \varepsilon$. Consider school

$$c^* = \langle(n_p, y_p), (n_r, y_r)\rangle$$

where $n_p = 1 - (n - \varepsilon)$, $y_p = y - \varepsilon$, and (n_r, y_r) is chosen so that $n_{c^*} = |c^*| = 1$. School

c^* is a school with two income groups, the poor being poorer than every student both in c and in c^k , and the rich being richer than the rich both in c and c^k , for every $k > k_0$. Also, the number of rich in c^* is smaller than the number of members in every income group both in c and c^k .

By construction, the income distribution of c^* is Lorenz-dominated by that of $c_{1/2} = \langle (1/2, 1/2), (1/2, 3/2) \rangle$. That is, the Lorenz curve associated with c^* is nowhere above the one associated with $c_{1/2}$. Therefore, there is a sequence of schools c_0, c_1, \dots, c_N with $c_0 = c^*$ and $c_N = c_{1/2}$ such that c_{t+1} is obtained from c_t by means of a progressive transfer. Therefore, by Proposition 2, $I(c^*) > I(c_{1/2})$. Noting that $X_{1/2} = d(c_{1/2})$ we conclude that $d(c^*) \succ X_{1/2}$.

Similarly, by construction the income distribution of c^* is Lorenz-dominated by that of c^k for all $k > k_0$. As a result, $d(c^*) \succ d(c^k)$ for all $k > k_0$. Furthermore, since c^k converges to c , by CONT $d(c^*) \succeq d(c)$.

Taking into account these relations, by Lemma 2 there are unique $\alpha, \alpha^k, \beta \in [0, 1]$ such that

$$d(c) \sim \alpha d(c^*) \uplus (1 - \alpha)X_0 \quad (12)$$

$$d(c^k) \sim \alpha^k d(c^*) \uplus (1 - \alpha^k)X_0 \quad k > k_0 \quad (13)$$

$$X_{1/2} \sim \beta d(c^*) \uplus (1 - \beta)X_0. \quad (14)$$

We end the proof by showing that α^k converges to α . Since $I(c^k) = \alpha^k/\beta$ and $I(c) = \alpha/\beta$ this will imply that $I(c^k)$ converges to $I(c)$ which is what we want to show. The argument is standard. Since $\alpha^k \in [0, 1]$ for $k > k_0$, the sequence $\{\alpha^k\}_{k > k_0}$ has a convergent sub-sequence. We now argue that all its convergent sub-sequences converge to α . Assume by contradiction that there is a sub-sequence $\alpha^{k(\ell)}$ that converges to $\lambda > \alpha$ (the case where $\lambda < \alpha$ is similar and left to the reader). Let

$\widehat{\lambda} = \frac{\lambda + \alpha}{2}$. Since $\widehat{\lambda} > \alpha$ by Lemma 1 and equation 12

$$\widehat{\lambda}d(c^*) + (1 - \widehat{\lambda})X_0 \succ \alpha d(c^*) + (1 - \alpha)X_0 \sim d(c). \quad (15)$$

Since $\alpha^{k(\ell)}$ converges to $\lambda > \widehat{\lambda}$, there is an ℓ_0 such that for all $\ell > \ell_0$, $\alpha^{k(\ell)} > \widehat{\lambda}$. Therefore, by equation 13 and Lemma 1, for all $\ell > \ell_0$,

$$d(c^{k(\ell)}) \sim \alpha^{k(\ell)}d(c^*) \uplus (1 - \alpha^{k(\ell)})X_0 \succ \widehat{\lambda}d(c^*) \uplus (1 - \widehat{\lambda})X_0.$$

Since $c^{k(\ell)}$ converges to c , by CONT we have that $d(c) \succeq \widehat{\lambda}d(c^*) \uplus (1 - \widehat{\lambda})X_0$, which contradicts (15). \square

Since I satisfies the above properties on \mathcal{C}_+ , they are also satisfied on the subclass of schools $\mathcal{C}_{+\mathbb{Z}}$, where the population n_g of each of its groups is an integer. It now follows from Theorem 5 in Shorrocks [22, p. 1381] that there exists a parameter α in \mathbb{R} and an increasing, continuous function $F : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ satisfying $F(0) = 0$ such that for any school c in $\mathcal{C}_{+\mathbb{Z}}$,

$$I(c) = F[I^\alpha(c)] \quad (16)$$

where I^α is the generalized entropy inequality index with parameter α .

We now show that equation 16 also holds for all schools c where the number of pupils in each school is a rational number. To see this, note that when the number of pupils in each group of school c is rational, $kc \in \mathcal{C}_{+\mathbb{Z}}$ for some positive integer k . By replication invariance, $I(c) = I(\underbrace{c \# \dots \# c}_k)$, which, by combining all the groups with the same income into one group, can be written as $I(kc)$. Then, using equation 16 we have that $I(c) = I(kc) = F[I^\alpha(kc)] = F[I^\alpha(c)]$ where the last equality follows from the fact that I^α also satisfies replication invariance. Finally, equation 16 also holds for all schools $c \in \mathcal{C}_+$ since $F \circ I^\alpha$ is continuous and \mathbb{Q} is dense in \mathbb{R} .

We now start the last step of the proof. We show that \mathcal{S} is a positive multiple of

the \mathcal{SSI} . Given that equation 16 holds for all schools $c \in \mathcal{C}_+$, applying Corollary 2, we obtain that the segregation index is of the form

$$\mathcal{S}(X) = F[I^\alpha(X)] - \sum_{c \in X} \frac{n_c}{n_X} F(I^\alpha(c)), \quad (17)$$

that is, \mathcal{S} is the segregation index induced by $F(I^\alpha)$. Since the \mathcal{SSI} is the segregation index induced by I^0 , it is enough to show that F is linear and that $\alpha = 0$.

In order to show that F is linear we will make use of the following well-known decomposability property of the generalized entropy indices I^α . See, for instance equations 32 and 4 in Shorrocks [21].

Observation 1 For any two schools c_1 and c_2 , let $c = c_1 \# c_2$. Then

$$I^\alpha(c) = \frac{n_{c_1}}{n_c} \left(\frac{\mu_{c_1}}{\mu_c} \right)^\alpha I^\alpha(c_1) + \frac{n_{c_2}}{n_c} \left(\frac{\mu_{c_2}}{\mu_c} \right)^\alpha I^\alpha(c_2) + I^\alpha(\bar{c}_1 \# \bar{c}_2).$$

The proof of this observation follows from a routine manipulation of the formula of I^α and is left to the reader.

We now show that F must be both concave and convex. Let z, z' be in the range of I^α (which is known to be an interval), and $\gamma \in (0, 1)$. Assume without loss of generality that $z > z'$. Pick two simple districts, $X = \{c_1, \dots, c_K\}$ and $Y = \{c'_1, \dots, c'_{K'}\}$, each with unit population and unit income, such that $I^\alpha(X) = z$ and $I^\alpha(Y) = z'$. Since X and Y are simple districts, we have that $I^\alpha(c) = 0$ for all $c \in X$ and for all $c' \in Y$. Therefore, since F is increasing, $\mathcal{S}(X) = F(I^\alpha(X)) = F(z) > F(z') = F(I^\alpha(Y)) = \mathcal{S}(Y)$. Since $\mathcal{S}(X) > 0$, X has at least two schools. Pick one school, say c_1 , and transfer a proportion p of pupils from each of the other schools to school c_1 to obtain district $X(p) = \{c_1 \# p(c_2 \# \dots \# c_n), (1-p)c_2, \dots, (1-p)c_n\}$.

Denoting $c_1(p) = c_1 \# p(c_2 \# \cdots \# c_n)$ we have by equation 17

$$\begin{aligned}\mathcal{S}(X(p)) &= F(I^\alpha(X(p))) - (n_1 + p(1 - n_1))F(I^\alpha(c_1(p))) \\ &= F(I^\alpha(X)) - (n_1 + p(1 - n_1))F(I^\alpha(c_1(p))).\end{aligned}$$

Note that when $p = 0$, $\mathcal{S}(X(0)) = \mathcal{S}(X) = F(z) > F(z') = \mathcal{S}(Y)$, and when $p = 1$, $\mathcal{S}(X(1)) = F(I^\alpha(X)) - F(I^\alpha(X)) = 0$. Consequently, by the intermediate value theorem, there is a $p^* \in (0, 1)$ such that $\mathcal{S}(X(p^*)) = \mathcal{S}(Y)$. Let $Z = X(p^*)$ and note that $n_Z = n_X = 1 = n_Y$, $|Z| = |X| = |Y| = 1$ and $I^\alpha(Z) = I^\alpha(X) = z$. Then, given that Y is a simple district,

$$\begin{aligned}\mathcal{S}(\gamma Z \uplus (1 - \gamma)Y) &= F(I^\alpha(\gamma Z \uplus (1 - \gamma)Y)) - \sum_{c \in Z} \gamma n_c F(I^\alpha(\gamma c)) \\ &= F\left(\gamma I^\alpha(Z) + (1 - \gamma)I^\alpha(Y)\right) - \sum_{c \in Z} \gamma n_c F(I^\alpha(\gamma c)) \quad (18)\end{aligned}$$

where the second equality made use of Observation 1 and the fact that $\mu_{\gamma Z} = \mu_{(1-\gamma)Y}$. On the other hand, since by PH, $\mathcal{S}(\gamma Z) = \mathcal{S}(\gamma Y)$, by IND,

$$\begin{aligned}\mathcal{S}(\gamma Z \uplus (1 - \gamma)Y) &= \mathcal{S}(\gamma Y \uplus (1 - \gamma)Y) \\ &= \mathcal{S}(Y) \\ &= \gamma \mathcal{S}(Z) + (1 - \gamma)\mathcal{S}(Y)\end{aligned}$$

where the second equality follows from SSP and IND. Using equation 17, and taking into account that Y is a simple district,

$$\begin{aligned}\mathcal{S}(\gamma Z \uplus (1 - \gamma)Y) &= \gamma \left[F(I^\alpha(Z)) - \sum_{c \in Z} n_c F(I^\alpha(c)) \right] + (1 - \gamma)F(I^\alpha(Y)) \\ &= \gamma F(I^\alpha(Z)) + (1 - \gamma)F(I^\alpha(Y)) - \sum_{c \in Z} \gamma n_c F(I^\alpha(c)). \quad (19)\end{aligned}$$

Comparing equations 18 and 19, and taking into account that $I^\alpha(\gamma c) = I^\alpha(c)$ we

conclude that

$$\gamma F(I^\alpha(Z)) + (1 - \gamma)F(I^\alpha(Y)) = F\left(\gamma I^\alpha(Z) + (1 - \gamma)I^\alpha(Y)\right).$$

Recalling that $I^\alpha(Z) = z$ and $I^\alpha(Y) = z'$ we conclude that F is both concave and convex. Furthermore, since $F(0) = 0$, we have that $F(z) = az$ for some $a > 0$.

It remains to show that $\alpha = 0$. We will show that unless this is the case, there exist two schools, c_1 and c_2 such that $\mathcal{S}(\{c_1, c_2\}) < 0$, which contradicts Claim 2.

Let $\alpha \neq 0$. Let $n_1 = n_2 = 1$, let $\mu_1 > 0$ be such that $\mu_1^\alpha \in (0, 1)$, and let μ_2 be implicitly defined by $n_1\mu_1 + n_2\mu_2 = 1$. Also let $c_1 = \left\langle (p, \varepsilon\mu_1), \left((1 - p), \frac{\mu_1(1 - \varepsilon p)}{(1 - p)}\right) \right\rangle$ and $c_2 = \langle (1, \mu_2) \rangle$ be two schools where $0 < p < 1$ and $0 < \varepsilon < 1$. School c_1 has two income groups. The proportion of pupils in the lower income group is p . The total population is 1 and the mean income is μ_1 . It can be checked that the closer p is to 1 and ε to 0, the higher is the income inequality as measured by I^α , both because the proportion of low income pupils becomes large and their incomes become low. For the moment assume that p is chosen to be close enough to 1 and ε is chosen to be close enough to 0 so that

$$I^\alpha(c_1) > \frac{I^\alpha(\bar{c}_1 \# \bar{c}_2)}{1/2(1 - \mu_1^\alpha)}. \quad (20)$$

We will later show that this can be done. Now let $X = \{c_1, c_2\}$. Then, using equation 17 and the fact that $F(z) = az$, we have that

$$\mathcal{S}(\{c_1, c_2\}) = a\left(I^\alpha(X) - \sum_{s=1}^2 \frac{I^\alpha(c_s)}{2}\right).$$

By Observation 1 and since $I^\alpha(c_2) = 0$,

$$\begin{aligned}
\mathcal{S}(\{c_1, c_2\}) &= a(I^\alpha(\bar{c}_1 \# \bar{c}_2) + \frac{1}{2}\mu_1^\alpha I^\alpha(c_1) + \frac{1}{2}\mu_2^\alpha I^\alpha(c_2) - \frac{1}{2}I^\alpha(c_1) - \frac{1}{2}I^\alpha(c_2)) \\
&= a(I^\alpha(\bar{c}_1 \# \bar{c}_2) + \frac{1}{2}\mu_1^\alpha I^\alpha(c_1) - \frac{1}{2}I^\alpha(c_1)) \\
&= a(I^\alpha(\bar{c}_1 \# \bar{c}_2) - \frac{1}{2}(1 - \mu_1^\alpha)I^\alpha(c_1)) \\
&< 0
\end{aligned}$$

where the last inequality follows from inequality 20. As mentioned before, this inequality contradicts Claim 2.

It remains to show that $p < 1$ and $\varepsilon > 0$ can be chosen so that inequality 20 holds. To see this, note first that since $n_{\bar{c}_1} = 1$ and $\mu_{\bar{c}_1} = \mu_1$, we have that $I^\alpha(\bar{c}_1 \# \bar{c}_2)$ is independent of p and of ε . Also, by direct computation, we have that

$$I^\alpha(c_1) = \begin{cases} \frac{p\varepsilon^\alpha + (1-p)^{1-\alpha}(1-p\varepsilon)^\alpha - 1}{(\alpha-1)\alpha} & \text{if } \alpha \neq 1 \\ p\varepsilon \log(\varepsilon) + (1-p\varepsilon) \log\left(\frac{1-p\varepsilon}{1-p}\right) & \text{if } \alpha = 1 \end{cases}.$$

Case 1: $\alpha \geq 1$. In this case we have that $\lim_{p \rightarrow 1} I^\alpha(c_1) = \infty$ and therefore, inequality 20 can be satisfied.

Case 2: $\alpha < 0$. In this case we have that $\lim_{p \rightarrow 1} I^\alpha(c_1) = \frac{\varepsilon^\alpha - 1}{(\alpha-1)\alpha}$ and therefore for ε close enough to 0, inequality 20 holds.

Case 3: $\alpha \in (0, 1)$. In this case we have that $\lim_{\substack{p \rightarrow 1 \\ \varepsilon \rightarrow 0}} I^\alpha(c_1) = \frac{1}{(1-\alpha)\alpha}$, and noting that

$$\frac{(1 - \mu_1^\alpha) + (1 - \mu_2^\alpha)}{(1 - \mu_1^\alpha)} < 1$$

we have that

$$\frac{1}{(1-\alpha)\alpha} > \frac{(1 - \mu_1^\alpha) + (1 - \mu_2^\alpha)}{(1 - \mu_1^\alpha)} \frac{1}{(1-\alpha)\alpha} = \frac{I^\alpha(\bar{c}_1 \# \bar{c}_2)}{1/2(1 - \mu_1^\alpha)}.$$

Therefore, for p close enough to 1 and ε close enough to 0, inequality 20 holds. This

completes the proof of the theorem. □

6.1 Discussion

a) *Independence of the axioms.* Table 1 lists a number of segregation indices that satisfy all the axioms but one. It also summarizes the axioms satisfied by Jargowsky's $\mathcal{N}\mathcal{S}\mathcal{I}$, and Reardon's \mathcal{H}^r . Apart from the segregation indices introduced in Section 4, the table includes the following indices:

- $n\mathcal{S}\mathcal{S}\mathcal{I}(X) = n_X\mathcal{S}\mathcal{S}\mathcal{I}(X)$
- $\mathcal{N}(X)$ = number of schools of X
- $\mathcal{F}(X) = \sqrt{I^0(X)} - \sum_{c \in X} \frac{n_c}{n_X} \sqrt{I^0(c)}$
- $\mathcal{C}_\mathcal{V}(X) = \frac{\sqrt{\mathcal{V}(X)}}{\mu_X}$
- $\mathcal{W}(X) = - \sum_{c \in X} \frac{n_c}{n_X} I^0(c)$.

	PH	IH	SSP	ESSD	IND	SEP	CONT
$\mathcal{S}\mathcal{S}\mathcal{I}$	✓	✓	✓	✓	✓	✓	✓
$n\mathcal{S}\mathcal{S}\mathcal{I}$		✓	✓	✓	✓	✓	✓
\mathcal{V}	✓		✓	✓	✓	✓	✓
\mathcal{N}	✓	✓		✓	✓	✓	✓
\mathcal{W}	✓	✓	✓		✓	✓	✓
\mathcal{F}	✓	✓	✓	✓		✓	✓
$\mathcal{C}_\mathcal{V}$	✓	✓	✓	✓	✓		✓
$\mathcal{N}\mathcal{S}\mathcal{I}$	✓	✓	✓	✓			✓
\mathcal{H}^r	✓	✓	✓	✓			

Table 1: Independence of the axioms

It can be seen that $\mathcal{S}\mathcal{S}\mathcal{I}$ satisfies all the axioms introduced in Section 5 and that if we do not require either population homogeneity, income homogeneity, the single-school property, equivalence of single-school districts, independence, or separability,

then a segregation index can be found that satisfies all the remaining axioms. We have not been able to show that continuity is not implied by the other axioms. Our main results states that \mathcal{SSI} index is essentially the only segregation index that satisfies all of them.

b) *Strength of the axioms.* Our axioms impose restrictions on segregation indices, not on inequality indices. Nevertheless, restricted to the class of simple districts (those with no income variation within schools), any segregation index \mathcal{S} induces a income inequality index I as follows: $I(c) = \mathcal{S}(d(c))$. Theorem 1 implies that our axioms characterize a segregation index whose induced index of income inequality is Theil's second measure. One may wonder whether our axioms restricted to the class of simple districts are strong enough to characterize directly Theil's second measure or any other income inequality index. The answer is negative. In fact, restricted to the class of simple districts our axioms are so weak that do not even imply the Pigou-Dalton principle. Indeed, as mentioned in Section 5, SSP, EESD, and SEP are axioms that are toothless when applied to indices defined on the class of simple districts since they deal with comparisons between districts that are not in that class. The only axioms that have any bite on the subclass of simple districts are PH, IH, IND, and CONT which are not sufficient to even imply the Pigou-Dalton principle.

c) *Additive separability of the index.* Given a partition of a district into two sub-districts, the within-district segregation is the population-weighted average of the segregation of the districts. The between-district segregation, on the other hand, is the segregation that would result if the segregation within each of the sub-districts were to be removed by combining all its schools into a single school. Corollary 1 shows that our axioms imply a very useful additive separability property. The index is the sum of the between-district and within-district segregation. The next section illustrates this separability property using data from Chile.

7 An empirical illustration

In this section we illustrate the decomposability property of the \mathcal{SSI} . We use data from SIMCE (Sistema de medición de la calidad de la educación) which contains student data from virtually all schools in Chile. Chile has fifty four provinces, grouped into fifteen regions. For our analysis we restrict attention to all provinces of the regions of Santiago, Valparaíso and Biobío (except for the province of Isla de Pascua, which has only three schools). These three regions represent a 60% of the Chilean population. Data include for each student, the school he attends and the income bracket his parents belong to. Income levels are partitioned into fifteen income brackets. For each province we estimate the mean income in each bracket by assuming that income is distributed according to a log-normal distribution, as follows. For an initial guess $(\bar{y}_1, \dots, \bar{y}_{15})$ of the mean incomes, we fit a log-normal distribution assuming that all households in income bracket i have an income of \bar{y}_i , for $i = 1, \dots, 15$. Then, we calculate the mean incomes of each bracket induced by the estimated distribution, and repeat the process using the estimated mean incomes as a new guess until the process converges. Chilean schools are classified according to their degree of dependence on public funding into three categories: public, semi-public and private.⁶

Table 2 shows for the regions of Santiago, Biobío and Valparaíso, their income segregation as measured by the \mathcal{SSI} (column 1), and its decomposition into segregation between provinces (column 2) and segregation within them (column 3). As can be seen, the Metropolitan region of Santiago exhibits more segregation than the other two. Also, for all the three regions more than 90% of the segregation can be attributed to the segregation within provinces, reflecting the fact that for each region the mean incomes of its provinces are roughly the same. Recall that any segregation index that is induced by an inequality index can be factored into a pure segregation and an inequality indices. Columns 4 and 5 report the result of this factorization for

⁶Public schools are funded by the city, and the semi-public category consist of private schools that are subsidized by public funds.

each of the regions. As can be seen, the tiny difference in the segregation exhibited by the regions of Biobío and Valparaíso is mainly due to a difference in their income inequality rather than a difference in their pure segregation.

Table 2: Segregation in selected Chilean regions for 2013. Columns 2 and 3 show its decomposition into segregation between provinces and within provinces for the \mathcal{SST} . Columns 4 and 5 show the decomposition into income inequality and pure segregation induced by \mathcal{SST} .

Region	\mathcal{SST}				
	Total	Segregation Breakdown			
	1	Between	Within	Inequality	Pure
Biobío	0.259	0.023	0.236	0.441	0.587
Valparaíso	0.236	0.016	0.220	0.402	0.586
Santiago	0.390	0.024	0.365	0.560	0.696

The fact that for most of the regions segregation is attributed to the segregation within their provinces suggests an analysis of this component. Table 3 reports for each of the provinces of the above three regions, their income segregation in 2013 as measured by the \mathcal{SST} , and its decomposition into between- and within-school categories.⁷

We can see that in most provinces, a large proportion of income segregation is due to the segregation between categories. This indicates that the mean incomes of the public, semi-public and private schools are substantially different from each other. The mean incomes of the schools within each category, on the other hand, are similar to each other as evidenced by the small segregation within categories exhibited by most provinces.

⁷As mentioned above, schools are classified into public, semi-public and private.

Table 3: Segregation in selected Chilean provinces for 2013. Columns 2 and 3 show its decomposition into segregation between- and within-school categories for the \mathcal{SSL} . Columns 4 and 5 show the decomposition into income inequality and pure segregation induced by \mathcal{SSL} .

Province	\mathcal{SSL}				
	Total	Segregation Breakdown			
		Between	Within	Inequality	Pure
	1	2	3	4	5
Arauco	0.123	0.054	0.070	0.310	0.398
Biobío	0.223	0.143	0.080	0.411	0.542
Concepcion	0.269	0.189	0.080	0.443	0.607
Ñuble	0.233	0.123	0.110	0.424	0.551
Chacabuco	0.617	0.555	0.062	0.805	0.766
Cordillera	0.147	0.066	0.081	0.298	0.493
Maipo	0.275	0.190	0.085	0.441	0.623
Melipilla	0.207	0.171	0.036	0.380	0.544
Santiago	0.402	0.300	0.101	0.573	0.701
Talagante	0.248	0.160	0.088	0.419	0.593
Los Andes	0.213	0.160	0.052	0.393	0.542
Marga Marga	0.170	0.104	0.065	0.351	0.484
Petorca	0.074	0.038	0.037	0.256	0.290
Quillota	0.189	0.139	0.050	0.361	0.523
S. Antonio	0.113	0.048	0.065	0.285	0.396
San Felipe	0.180	0.137	0.042	0.346	0.520
Valparaiso	0.311	0.252	0.059	0.462	0.673

References

- [1] J. S. Coleman, E. Q. Campbell, C. J. Hobson, J. J. McPartland, A. M. Mood, F. D. Weinfeld, and R. L. York. *Equality of educational opportunity*, volume 2. JSTOR, 1966.
- [2] F. Echenique and R. G. Fryer. A measure of segregation based on social interactions. *The Quarterly Journal of Economics*, 122(2):441–485, 2007.
- [3] E. Fong and K. Shibuya. The spatial separation of the poor in canadian cities. *Demography*, 37(4):449–459, 2000.
- [4] J. E. Foster and A. F. Shorrocks. Subgroup consistent poverty indices. *Econometrica*, 59(3):687–709, 1991.
- [5] D. M. Frankel and O. Volij. Measuring school segregation. *Journal of Economic Theory*, 146(1):1–38, 2011.
- [6] E. A. Hanushek, J. F. Kain, J. M. Markman, and S. G. Rivkin. Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18(5):527–544, 2003.
- [7] C. Hoxby. Peer effects in the classroom: Learning from gender and race variation. Technical report, National Bureau of Economic Research, 2000.
- [8] R. Hutchens. Numerical measures of segregation: desirable properties and their implications. *Mathematical Social Sciences*, 42(1):13–29, 2001.
- [9] R. Hutchens. One measure of segregation. *International Economic Review*, 45(2):555–578, 2004.
- [10] S. A. Imberman, A. D. Kugler, and B. I. Sacerdote. Katrina’s children: Evidence on the structure of peer effects from hurricane evacuees. *The American Economic Review*, 102(5):2048–2082, 2012.

- [11] J. Jahn, C. F. Schmid, and C. Schrag. The measurement of ecological segregation. *American Sociological Review*, 12(3):293–303, 1947.
- [12] P. A. Jargowsky. Take the money and run: Economic segregation in US metropolitan areas. *American Sociological Review*, 61(6):984–998, 1996.
- [13] S. P. Jenkins, J. Micklewright, and S. V. Schnepf. Social segregation in secondary schools: how does england compare with other countries? *Oxford Review of Education*, 34(1):21–37, 2008.
- [14] V. Lavy, M. D. Paserman, and A. Schlosser. Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal*, 122(559):208–237, 2012.
- [15] D. S. Massey. The age of extremes: Concentrated affluence and poverty in the twenty-first century. *Demography*, 33(4):395–412, 1996.
- [16] D. S. Massey and N. A. Denton. The dimensions of residential segregation. *Social Forces*, 67(2):281–315, 1988.
- [17] S. E. Mayer. How economic segregation affects children’s educational attainment. *Social Forces*, 81(1):153–176, 2002.
- [18] S. F. Reardon. Measures of ordinal segregation. In *Occupational and residential segregation*, pages 129–155. Emerald Group Publishing Limited, 2009.
- [19] S. F. Reardon. Measures of income segregation. *Unpublished Working Paper. Stanford Center for Education Policy Analysis*, 2011.
- [20] S. F. Reardon and G. Firebaugh. Measures of multigroup segregation. *Sociological Methodology*, 32(1):33–67, 2002.
- [21] A. F. Shorrocks. The class of additively decomposable inequality measures. *Econometrica*, 48:613–625, 1980.

- [22] A. F. Shorrocks. Inequality decomposition by population subgroups. *Econometrica*, 52(6):1369–1385, 1984.
- [23] A. F. Shorrocks. Aggregation issues in inequality measurement. In W. Eichhorn, editor, *Measurement in Economics*, pages 429–451. Springer, Heidelberg, 1988.
- [24] H. Theil and A. J. Finizza. A note on the measurement of racial integration of schools by means of informational concepts. *Journal of Mathematical Sociology*, 1:187–1193, 1971.

A Appendix

Proof of Corollary 1

Applying Proposition 1 to the district $X_1 \uplus C(X_2)$, we obtain the statement for $J = 2$.

Assume that the statement is true for $J = m - 1$. Then, denoting $n = n_X$,

$$\begin{aligned}
\mathcal{S}\left(\uplus_{j=1}^m X_j\right) &= \mathcal{S}\left(C\left(\uplus_{j=1}^{m-1} X_j\right) \uplus C(X_m)\right) + \frac{\sum_{j=1}^{m-1} n_{X_j}}{n} \mathcal{S}\left(\uplus_{j=1}^{m-1} X_j\right) + \frac{n_{X_m}}{n} \mathcal{S}(X_m) \\
&= \mathcal{S}\left(C\left(\uplus_{j=1}^{m-1} X_j\right) \uplus C(X_m)\right) + \frac{\sum_{j=1}^{m-1} n_{X_j}}{n} \left[\mathcal{S}\left(\uplus_{j=1}^{m-1} C(X_j)\right) + \sum_{j=1}^{m-1} \frac{n_{X_j}}{\sum_{j=1}^{m-1} n_{X_j}} \mathcal{S}(X_j) \right] + \\
&\quad + \frac{n_{X_m}}{n} \mathcal{S}(X_m) \\
&= \mathcal{S}\left(C\left(\uplus_{j=1}^{m-1} X_j\right) \uplus C(X_m)\right) + \frac{\sum_{j=1}^{m-1} n_{X_j}}{n} \mathcal{S}\left(\uplus_{j=1}^{m-1} C(X_j)\right) + \sum_{j=1}^m \frac{n_{X_j}}{n} \mathcal{S}(X_j). \quad (21)
\end{aligned}$$

Applying this expression to $\uplus_{j=1}^m C(X_j)$, and noting that $C(C(X_j)) = C(X_j)$ we

obtain that

$$\mathcal{S}\left(\biguplus_{j=1}^m C(X_j)\right) = \mathcal{S}\left(C\left(\biguplus_{j=1}^{m-1} C(X_j)\right) \uplus C(X_m)\right) + \frac{\sum_{j=1}^{m-1} n_{X_j}}{n} \mathcal{S}\left(\biguplus_{j=1}^{m-1} C(X_j)\right) + \sum_{j=1}^m \frac{n_{X_j}}{n} \mathcal{S}(C(X_j)).$$

Since $C\left(\biguplus_{j=1}^{m-1} C(X_j)\right) = C\left(\biguplus_{j=1}^{m-1} X_j\right)$ and since by Claim 2 $\mathcal{S}(C(X_j)) = 0$, rearranging we obtain

$$\mathcal{S}\left(C\left(\biguplus_{j=1}^{m-1} X_j\right) \uplus C(X_m)\right) = \mathcal{S}\left(\biguplus_{j=1}^m C(X_j)\right) - \frac{\sum_{j=1}^{m-1} n_{X_j}}{n} \mathcal{S}\left(\biguplus_{j=1}^{m-1} C(X_j)\right).$$

Replacing this expression in equation 21 we get

$$\mathcal{S}\left(\biguplus_{j=1}^m X_j\right) = \mathcal{S}\left(\biguplus_{j=1}^m C(X_j)\right) + \sum_{j=1}^m \frac{n_{X_j}}{n} \mathcal{S}(X_j).$$