



Working Paper 09-32
Economic Series (18)
November 2009

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

ENTROPY-BASED SEGREGATION INDICES

Ricardo Mora and Javier Ruiz-Castillo¹
Departamento de Economía, Universidad Carlos III de Madrid

Abstract

Recent research (Reardon and Firebaugh, 2002, Frankel and Volij, 2009, and Mora and Ruiz-Castillo, 2009a) has shown that two entropy-based segregation indices possess an appealing mixture of basic and subsidiary but useful properties. It would appear that the only fundamental difference between the mutual information, or M index, and the Entropy, Information or H index, is that the second is a normalized version of the first. This paper introduces another normalized index in that family, the H^* index that, contrary to what is often asserted in the literature, is the normalized entropy index that captures the notion of segregation as departures from evenness. More importantly, this paper shows that applied researchers do better using the M index than using either H or H^* in two circumstances: (i) if they are interested in the decomposability of segregation measures for any partition of organizational units into larger clusters and of demographic groups into supergroups, and (ii) if they are interested in the invariance properties of segregation measures to changes in the marginal distributions by demographic groups and by organizational units.

Keywords: Multigroup Segregation Measurement; Axiomatic Properties; Entropy Based Indicators; Econometric Models.

¹ The authors acknowledge financial support from the Spanish DGI, Grants SEJ2006-05710 and SEJ2007-67436.

I. INTRODUCTION

Segregation measures describe differences in the distribution of two or more demographic groups (genders, racial/ethnic groups) over a set of organizational units (occupations, neighborhoods, schools). As with the measurement of other complex, multifaceted phenomena in the social sciences – such as income inequality or economic poverty– it should come as no surprise that there exists a plethora of indicators capturing different aspects of the same phenomenon.² In some circumstances, this multiplicity of potential measures does not cause any practical problem. In most applications, however, different indices will lead to different conclusions, making it relevant to seek criteria to discriminate between the admissible alternatives.

One strategy followed in the income inequality literature starts by selecting a set of “basic” properties that any inequality index should satisfy. Afterwards, one may consider the contexts in which income inequality measures are often used, and to identify additional “subsidiary” properties that narrow down the options. This is important because, as one of the leading advocates of this approach indicates, “If this search is not undertaken, there is a tendency to continue using those measures that have been popular in the past. The index is then chosen by default, or historical accident, rather than by any assessment of its merits.” (Shorrocks, 1988, p. 433). Grusky and Charles (1998, p. 497) complain that this situation has indeed been prevalent in the history of research on occupational segregation by gender where “(...) scholars have effectively assumed that sex segregation is simply whatever (the index of dissimilarity [see Duncan and Duncan (1955)]) measures”.³

Two recent methodological contributions to the multigroup segregation literature have reached important results following the strategy of combining basic and interesting subsidiary properties.

² Surveys include James and Taeuber (1985), Massey and Denton (1988), and Flückiger and Silber (1999).

³ The quotation in full is enlightening: “For all its faddishness, the concept of path dependency proves useful in understanding the history of sex segregation research, and not merely because the index of dissimilarity (hereafter, *D*) has shaped and defined the methodology of segregation analysis over the last 25 years. It is perhaps more important that *D* has been so dominant during this period that it undermined all independent conceptual development. Indeed, segregation scholars have effectively assumed that sex segregation is simply whatever *D* measures.”

Reardon and Firebaugh (2002) –RF hereafter– extend to the multigroup case some basic properties of segregation measures discussed in the 1980s for the two-group case, and introduce some subsidiary decomposability properties for any partition of organizational units into clusters (such as schools into school districts, or three-digit occupations into two-digit ones), or of demographic groups into supergroups (such as precisely-defined ethnic categories, like Mexican or Puerto Rican, into a major category such as Hispanic). They find that, among six popular measures, only the entropy-based segregation index –known as the Information, the Entropy or the H index– satisfies the basic properties and a weak version of the two decomposability properties. Frankel and Volij (2009) –FV hereafter– characterize an entropy-based multigroup segregation ranking in terms of eight ordinal axioms that capture some basic, subsidiary, and technical properties. It is represented by a simple index called the Mutual Information or M index.⁴ It should be noted that the H index is a normalization of M . FV also establish that three of their subsidiary ordinal axioms are necessary for any segregation index to satisfy a strong cardinal version of the two decomposability properties. While the M index satisfies strong decomposability in FV’s sense, none of a list of nine alternative segregation measures, including the H index, does.

On the other hand, Mora and Ruiz-Castillo (2009a) –MRC hereafter– show that the M index can be characterized as a monotonic transformation of the likelihood ratio test for the independence between race status and school membership in a general statistical framework, and provide sufficient regularity conditions to ensure its consistency. Moreover, MRC extend the decomposition of M to isolate segregation conditional on any vector of (possibly continuous) socioeconomic characteristics, and propose consistent estimators for the terms in this decomposition. As a result, the M index now stands as the only index of segregation which has been fully characterized in terms of axiomatic

⁴ Very few segregation indices have been similarly characterized. In the two groups case, Chakravarty and Silber (1992) characterize an index of absolute segregation, while Chakravarty and Silber (2007) axiomatically derive a class of numerical indices of relative segregation that parallel the multidimensional Atkinson inequality indices. Two members of that class are monotonically related to the square root index, independently characterized by Hutchens (2004), and the M index. In the multigroup case, Frankel and Volij (2008) provide an ordinal characterization of two families of Atkinson indices.

properties, is well embedded into a general statistical framework, and can be used when samples are finite and a multivariate framework is required.

Given the newly found relevance of entropy-based indices in multigroup segregation studies, this paper makes two contributions to this literature. In the first place, it is shown that there are two ways to normalize the M index. Contrary to what is believed since Massey and Denton (1988), the H index captures the isolation or representative aspect of segregation. The second normalization, leading to what we call the H^* index, is the one that captures the evenness aspect of segregation in the classical sense of James and Taeuber (1985). Interestingly enough, the M index simultaneously captures the evenness and representative aspects of segregation.

In the second place, both the H and the M indexes violate the invariant properties according to which a segregation index should not vary when the only change between two situations under comparison is in the population marginal distributions by demographic groups or by organizational units. Taking this and the above results into account, it would appear that the problem of selecting an adequate segregation index in different practical circumstances is settled. Researchers are obviously free to choose whatever segregation indices they find suitable for their purposes independently of the good properties of entropy-based indices. But in a context where the decomposability properties are considered to be useful, they may profitably choose between the H and the M indices. On the other hand, as FV advise, if one's only goal is to study how the effect of origins (race) on destinations (school assignment) has changed over time, then one may prefer a measure that is unaffected by changes in the marginal distribution by demographic group.⁵ In our second contribution, this paper explains why the issue should not be settled in this way. It turns out that, except for Frankel and Volij (2008, 2009) and Mora and Ruiz-Castillo (2009a, b, c), the few authors that have used an entropy-

⁵ In particular, in a different paper Frankel and Volij (2008) have characterized an Atkinson segregation index in terms of a number of basic ordinal properties and a version of such an invariant property they call *Scale Invariance*, a property known as *Invariant 1* or **I1** in Mora and Ruiz-Castillo (2009b). No characterization is available in terms of a number of interesting properties and a second type of invariance to changes in the marginal distribution by organizational unit called *Invariant 2* or **I2** in Mora and Ruiz-Castillo (2009b).

based index to study multigroup segregation have preferred the H index.⁶ However, as argued in this paper, the conceptual and practical advantages of the M index are inescapable.

(i) Strongly decomposable indices are useful because they unambiguously address issues concerning the relationship between the overall segregation value and the characteristics of organizational units or demographic subgroups. For example, we may wish to assess the degree to which overall segregation is due to segregation within a large supergroup consisting of all minority races in the U.S. or how much is due to racial differences across clusters of different size. However, as pointed out in the income inequality literature, these deceptively simple questions raise a number of conceptual and methodological problems (Shorrocks, 1988, p. 435). In particular, it will be shown that the H index is not appropriately responsive to changes in the subgroups segregation levels in the partition of demographic groups into supergroups, while the H^* index is not responsive to those changes in the partition of organizational units into clusters. The M index has no such problem. Moreover, the empirical questions usually asked in decomposability analysis receive the more unambiguous answers that are possible in a segregation context under the additively decomposability properties that are only satisfied by the M index. In particular, when the empirical question refers to what extent can segregation be explained by the (possibly continuous) determinants of individual choice, the M index can be used to embed the econometric multivariate models used to study individual choice into the measurement of segregation.

(ii) The H and H^* measures mix up segregation changes with changes in the marginal distributions in segregation comparisons over time or across space, while the M index admits two decompositions that isolate terms that capture segregation changes net of the impact of pure demographic factors (Mora and Ruiz-Castillo, 2009b).

⁶ Reardon *et al.* (2000) distinguishes between the central city and the suburbs in a study of within-cities school segregation, while Miller and Quigley (1990) and Fisher (2003) on one hand, and Iceland (2002) on the other study within-cities and within-regions residential segregation. Fisher *et al.* (2004), which is the only contribution on residential segregation that develops a full multilevel approach using the H index, only reports pair-wise comparisons of racial/ethnic groups.

(iii) The H and H^* measures have not yet been axiomatically characterized as FV have done for the M index. On the other hand, only the M index and its between and within terms have been placed into a general statistical framework and their asymptotic properties fully characterized (MRC). This makes it possible to study the significance of the segregation measures under alternative hypothesis when samples are finite and a multivariate framework is required.

(iv) The only advantage the H and H^* measures can claim is normalization, a subsidiary property with serious undesirable implications (Clotfelter, 1979, and FV).

The rest of this paper is organized into four Sections. Section II introduces the notation, presents the M , H and H^* indices, and explores the normalization issue. Sections III and IV discuss decomposability and invariance to marginal distributions, while Section V concludes.

II. ENTROPY-BASED INDICES AND NORMALIZATION

II.1. Notation

It would be useful to refer to a specific segregation problem. To be faithful to the origins of entropy-based measures, the case discussed throughout the paper is the school segregation problem in Theil and Finizza (1971) –TF hereafter. Assume a city \mathbf{X} consisting of N schools, indexed by $n = 1, \dots, N$. Each student belongs to any of G racial groups, indexed by $g = 1, \dots, G$. In TF, $G = 2$. However, given the racial diversity existing in many countries, this paper studies the multigroup case where $G \geq 2$. The data available can be organized into the following $G \times N$ matrix:

$$\mathbf{X} = \{t_{gn}\} = \begin{bmatrix} t_{11} & \cdots & t_{1N} \\ \vdots & \ddots & \vdots \\ t_{G1} & \cdots & t_{GN} \end{bmatrix} \quad (1)$$

where t_{gn} is the number of individuals of racial group g attending school n , so that $t = \sum_{n=1}^N \sum_{g=1}^G t_{gn}$ is the total student population.

The information contained in the joint absolute frequencies of racial groups and schools, t_{gn} is usually summarized by means of numerical indices of segregation. Let $\mathfrak{E}(G, N)$ be the set of all cities with G groups and N schools. A segregation index S is a real valued function defined in $\mathfrak{E}(G, N)$, where $S(\mathbf{X})$ provides the extent of school segregation for any city $\mathbf{X} \in \mathfrak{E}(G, N)$. Let $p_{gn} = t_{gn}/t$, and denote by $P_{gn} = \{p_{gn}\}_{g=1, n=1}^{G, N}$ the joint distribution of racial groups and neighborhoods in a city $\mathbf{X} \in \mathfrak{E}(G, N)$. In the following, the discussion will be restricted to indices that capture a *relative* view of segregation in which all that matters is the joint distribution, i.e. indices which admit a representation as a function of P_{gn} .⁷

This paper considers two notions of segregation. Under the first one, referred to as “evenness”, segregation is viewed as the tendency of racial groups to have different distributions across schools.⁸ In contrast, the notion of “representativeness”, emphasized by FV, asks to what extent schools have different racial compositions from the population as a whole.⁹ As can be seen in expression (1), where the rows are racial groups and the columns are schools, evenness and representativeness are dual concepts: deviations from evenness (representativeness) correspond to differences in the row (column) percentages. The following observation indicates how close these two views are to each other.

Remark 1. If a segregation index S that captures the notion of evenness when applied to city $\mathbf{X} \in \mathfrak{E}(G, N)$ is applied to the city $\mathbf{X}' \in \mathfrak{E}(N, G)$, where the role of schools and racial groups are reversed so that $t_{gn} = t'_{ng}$ for all g and n , then what will be called the *reciprocal* index S^* applied to \mathbf{X}' captures equally well the notion of representativeness (and vice versa).

⁷ This property, satisfied by most segregation indices, is referred to as *Size Invariance* in James and Taeuber (1985) and as *Weak Scale Invariance* in FV and Frankel and Volij (2008). For a study that focuses on translation invariant segregation indices that represent an absolute view of segregation, see Chakravarty and Silber (1992).

⁸ It is generally agreed that residential and school segregation are multifaceted concepts whose measurement may require a battery of indices, one for each facet. In the context of residential segregation, Massey and Denton (1988) distinguish five notions or dimensions, of which evenness is the one that agrees with the classic definition of James and Taeuber (1985). Three of the other four dimensions –concentration, centralization, and clustering– require detailed geographic information about a city's neighborhoods.

⁹ FV view representativeness as the multigroup generalization of the notion of “isolation” proposed by Massey and Denton (1988) in the two-group case. Racially isolated schools are, by definition, not representative of the population. But unlike isolation, in the multigroup case representativeness is not based on the exposure of one specific group to another.

In general, $S(\mathbf{X})$ and $S^*(\mathbf{X}^*)$ will provide a different segregation value for the same data. When this is not the case, the segregation index under consideration is said to be *transpose-invariant*.

Before we present the entropy-based indices of segregation, the concept of entropy of a distribution must be introduced. Consider a discrete random variable x that takes Q probability values, indexed by $q = 1, \dots, Q$. Let p_q be the probability of the q th value with $p_q \geq 0$, $\sum_{q=1}^Q p_q = 1$, so that

$P = \{p_q\}_{q=1}^Q$ is the probability distribution for variable x . The *entropy* of the Q values of variable x is the real value function on the probability set $\mathbf{P}, E : \mathbf{P} \subset [0, 1]^Q \rightarrow [0, \log(Q)] \subset \mathbb{R}$, defined as

$$E(P) = -\sum_{q=1}^Q p_q \log(p_q)$$

with $0 \log(1/0) = 0$.¹⁰ Heuristically, the information brought about by observing the actual value of x is the opposite of the logarithm of its likelihood, $-\log(p_q)$: the observation of an unlikely value brings about a large amount of information once observed. Therefore, the entropy can be considered a measure of the expected information for the value of variable x brought about by an observation. It is straightforward to show that the entropy is bounded. It reaches its maximum value at the uniform distribution, i.e. $P = \{1/Q\}_{q=1}^Q$, whilst it takes its minimum value for each of the Q degenerate distributions defined by:¹¹

$$p_q = \begin{cases} 1 & \text{if } q = r \\ 0 & \text{if } q \neq r \end{cases}, \quad q, r = 1, \dots, Q.$$

Since the entropy captures the degree of uniformity in the probabilities of each possible event described by x , it can be interpreted as a measure of uncertainty or diversity of random variable x .

II. 2. Segregation as Departures from Representativeness

¹⁰ The base of the logarithm is irrelevant, providing essentially a unit of measure. In this paper the natural logarithm will be used.

¹¹ For proofs of these and other properties of the entropy function see, *inter alia*, Theil (1972).

Let $P_g = \{p_g\}_{g=1}^G$ be the marginal distribution by racial groups, $P_n = \{p_n\}_{n=1}^N$ the marginal distribution by school size, and $P_{g|n} = \{p_{g|n}\}_{g=1}^G$ the conditional distribution by race of students in school n . Consider the racial entropy at school n , $E(P_{g|n})$. This function can be interpreted as the degree of uniformity in the racial shares within school n . In contrast, the racial entropy at city level, $E(P_g)$, gives us a measure of the degree of uniformity in the racial mix for the city as a whole. The difference between these two entropies,

$$\Delta E(P_g, P_{g|n}) = E(P_g) - E(P_{g|n}) = \sum_{g=1}^G p_{g|n} \log(p_{g|n}) - \sum_{g=1}^G p_g \log(p_g),$$

can be interpreted as a local measure of discrepancy in racial-share uniformity between the city and school n . It reaches its maximum under the following two conditions: $E(P_{g|n}) = 0$ (only one race is present in school n), and $E(P_g) = \log(G)$ (all races are equally represented in the city). In contrast, $\Delta E(P_g, P_{g|n}) = 0$ if the racial mix is the same in the city and the school, in which case we may say that the school is representative of the city. Finally, note that a negative $\Delta E(P_g, P_{g|n})$ represents a situation whereby racial-share uniformity is larger at school than at city level. Thus, $\Delta E(P_g, P_{g|n})$ captures departures from representativeness in school n .

The weighted average of local measures $\Delta E(P_g, P_{g|n})$ over all schools with weights proportional to their student bodies gives an overall measure of discrepancy between the racial mix in the city and the racial mix within each school n .

$$M\left(P_g, \{P_{g|n}\}_{n=1}^N\right) = \sum_{n=1}^N p_n \Delta E(P_g, P_{g|n}) = E(P_g) - \sum_{n=1}^N p_n E(P_{g|n}). \quad (2)$$

This measure has been first proposed in the segregation literature by TF in the context of racial

segregation in schools for a given school district.¹² Expression (2) highlights that $M\left(P_g, \{P_{g|n}\}_{n=1}^N\right)$ is the difference between the racial entropy in the city and the weighted average of racial entropies for each school. In information theory, $M\left(P_g, \{P_{g|n}\}_{n=1}^N\right)$ represents the information lost when P_g is used to approximate $\{P_{g|n}\}_{n=1}^N$, or equivalently, the information gained about the race of a student when we are told about her school. As explained in FV, suppose that we do not know a student's race, g , but we are told which school n she attends; if the schools in the city are segregated, this will convey information about her race.

It can be shown that $M\left(P_g, \{P_{g|n}\}_{n=1}^N\right) \in [0, \log(G)]$. In particular, it reaches its minimum value whenever the racial composition in each school coincides with the racial mix at the city level, while it reaches its maximum value when the marginal racial distribution at city level is the uniform and there is no ethnic mix within schools. In other words, the notion of complete segregation for this measure demands two conditions: there must be no racial mix within organizational units, and races must be uniformly distributed at city level. For any given racial marginal distribution P_g , $M\left(P_g, \{P_{g|n}\}_{n=1}^N\right)$ attains its maximum at $E(P_g)$. This fact suggests normalizing the M measure by dividing it by the city's racial entropy:

¹² See equation (8), p. 191, in TF, as well as equations (7.21) and (7.22) in Theil, 1971, p. 653. In information theory, the expression $M^n(P_g, P_{g|n}) = \sum_{g=1}^G P_{g|n} \log(P_{g|n} / P_g)$ is the expected information of the message that transforms the marginal distribution of groups in the city, $P_g = \{P_g\}$, to the conditional distribution of racial groups in school n , $P_{g|n} = \{P_{g|n}\}$ (see equation (7) in TF, p 190; equation (7.1) in Theil, 1971, p. 645, and equation (2.5) in Theil, 1972, p. 65). As emphasized by TF, since $M^n(P_g, P_{g|n})$ measures the extent to which the racial composition in school n differs from the one for the city as a whole, it can be interpreted as a *local index of segregation* in school n when segregation focuses on representativeness. It can readily be seen that the demographically weighted average of these local measures of segregation coincides with the overall measure of discrepancy introduced in equation (2), that is, $M\left(P_g, \{P_{g|n}\}_{n=1}^N\right) = \sum_{n=1}^N P_n M^n(P_g, P_{g|n})$. Entropy-based and other local segregation indices are axiomatically characterized in Alonso-Villar and Del Río (2008).

$$H \equiv H\left(P_g, \{P_{g|n}\}_{n=1}^N\right) = \frac{1}{E(P_g)} M\left(P_g, \{P_{g|n}\}_{n=1}^N\right) = \sum_{n=1}^N p_n \frac{\Delta E(P_g, P_{g|n})}{E(P_g)}. \quad (3)$$

The H index, referred to as the Entropy or Information index, first appears in TF and Theil (1972) in the context of racial segregation in schools in a given school district. Intuitively, it captures the proportion of the racial mix uniformity in the city that is not due to racial mix uniformity at school level. Note that, in contrast to $M\left(P_g, \{P_{g|n}\}_{n=1}^N\right)$, it can only take values within the unit interval (regardless of the logarithmic base). More importantly, it reaches the unit whenever there is no racial mix within schools. On the other hand, equation (3) implies that, contrary to some previous claims in the literature, the entropy index $H = H\left(P_g, \{P_{g|n}\}_{n=1}^N\right)$ is a segregation index that measures departures from representativeness.¹³

One important point remains to be clarified.

Remark 2. Local measures of discrepancy in racial shares, $\Delta E(P_g, P_{g|n})$, are not independent from each other. First, an independent change in the racial mix in one school (through the addition or removal of one student) necessarily affects the racial composition in the city, and hence the local measure of discrepancy in racial shares in the remaining schools. Second, a change in the racial composition of a school maintaining the total number of students of each race in the city as a whole, necessarily affects the local measure of discrepancy in some other school.

Therefore, while equations (2) and (3) may seem to permit the decomposition of overall segregation at the city level in N components –each indicating the contribution to overall segregation made by each school– it is meaningless to talk of a single school’s contribution to overall city

¹³ We believe that the misunderstanding starts with Massey and Denton (1988, p. 304). For the data they analyzed, they found that the Entropy index shared a common latent factor, which they interpreted as an evenness factor, with some evenness measures of segregation.

segregation. Segregation as deviations from representativeness arises from the comparison of the racial composition in the N schools –not by the racial characteristics of a school in isolation.

II. 3. Segregation as Departures from Evenness

Let $P_{n|g} = \{p_{n|g}\}_{n=1}^N$ be the conditional distribution by schools of individuals in racial group g .

Following an argument similar to the one used in the previous section to construct a measure of representativeness, we can use the difference between the schools entropy at city level, $E(P_n)$, and the schools entropy of racial group g , $E(P_{n|g})$, to obtain a measure of departures from evenness for racial group g :

$$\Delta E(P_n, P_{n|g}) = E(P_n) - E(P_{n|g}) = \sum_{n=1}^N p_{n|g} \log(p_{n|g}) - \sum_{n=1}^N p_n \log(p_n).$$

The weighted average of these measures over all racial groups with weights equal to their city shares gives an overall measure of discrepancy between the marginal school distribution in the city and school distribution within each racial group g .

$$M\left(P_n, \{P_{n|g}\}_{g=1}^G\right) = \sum_{g=1}^G p_g \Delta E(P_n, P_{n|g}) = E(P_n) - \sum_{g=1}^G p_g E(P_{n|g}). \quad (4)$$

Expression (4) shows that $M\left(P_n, \{P_{n|g}\}_{g=1}^G\right)$ is the difference between the schools' entropy in the city and the weighted average of schools entropies for each racial group. In information theory, $M\left(P_n, \{P_{n|g}\}_{g=1}^G\right)$ represents the information lost when P_n is used to approximate $\{P_{n|g}\}_{g=1}^G$. As explained in FV, suppose that we do not know a student's school, n , but we are told which race g she belongs to; if races in the city are segregated, this will convey information about her school.

Note that $M\left(P_n, \{P_{n|g}\}_{g=1}^G\right) \in [0, \log(N)]$. It reaches its minimum value whenever the school distribution is the same for all racial groups, while it reaches its maximum value when the school

distribution at the city level is uniform but each racial group attends a disjoint set of schools (so that there is no ethnic mix within schools). Thus, the notion of complete segregation for $M\left(P_n, \{P_{n|g}\}_{g=1}^G\right)$ also demands two conditions: in addition to requiring no racial mix within organizational units, schools must be uniformly distributed at city level. For any given school distribution P_n , $M\left(P_n, \{P_{n|g}\}_{g=1}^G\right)$ attains its maximum at $E(P_n)$. This fact suggests normalizing $M\left(P_n, \{P_{n|g}\}_{g=1}^G\right)$ by dividing it by the schools entropy at the city level:

$$H^* \equiv H\left(P_n, \{P_{n|g}\}_{g=1}^G\right) = \frac{1}{E(P_n)} M\left(P_n, \{P_{n|g}\}_{g=1}^G\right) = \sum_{g=1}^G P_g \frac{\Delta E(P_n, P_{n|g})}{E(P_n)}. \quad (5)$$

The H^* index has not been defined previously. Intuitively, it captures the proportion of the school distribution uniformity in the city that is not due to school-share uniformity within racial groups. It can only take values within the unit interval, and it reaches the unity whenever there is no racial mix within schools. Equation (5) implies that H^* is a segregation index that measures departures from evenness. Of course, a remark similar to Remark 2 applies here as well: it is meaningless to talk of a single race's contribution to overall city segregation. In the words of RF, "segregation is defined by the *relationships* among the groups' distributions across organizational units –not by the distribution across units of each group in isolation".

II. 4. The Mutual Information Index

What is the relation between the entropy-based measures of deviations from representativeness and the entropy-based measures of deviations from evenness?

We start by noting that Mora and Ruiz-Castillo (2005) for the two-group case and FV for the multigroup case show that, for the same data set, the expressions $M\left(P_g, \{P_{g|n}\}_{n=1}^N\right)$ and $M\left(P_n, \{P_{n|g}\}_{g=1}^G\right)$ result in a unique value that we can denote by M , that is

$$M = M\left(P_g, \{P_{gn}\}_{n=1}^N\right) = M\left(P_n, \{P_{ng}\}_{g=1}^G\right). \quad (6)$$

From equations (3), (5), and (6) we obtain that:

$$H = \frac{E(P_g)}{E(P_n)} H^*.$$

Therefore, in terms of the definition introduced in Remark 1, H^* is the reciprocal index of H . Clearly, since the two indices are different whenever $E(P_n) \neq E(P_g)$, neither of them is transpose-invariant.

In contrast, equation (6) shows that the M index is transpose invariant and captures the criteria of evenness and representativeness in a symmetric fashion. It represents both the information gained about race when school is known, as well as the information gained about school when race is known. Hence, the M index is referred to as the *Mutual Information* index.

An alternative motivation of the M index arises from the observation (in line with note 12) that it can be expressed as the expected information of the message that transforms the unrestricted joint distribution of groups and schools in the city, $P_{gn} = \{p_{gn}\}$, to the joint distribution under the assumption of independence between racial group and school membership, $\tilde{P}_{gn} = \{p_g p_n\}$:

$$M = \sum_{n=1}^N \sum_{g=1}^G p_{gn} \log \left(\frac{p_{gn}}{p_g p_n} \right).$$

In Information Theory, this last expression is a particular case of the Kullback-Leibler Information measure for discrete distributions:

$$KL = \sum_{i=1}^I p_i \log \left(\frac{p_i}{\pi_i} \right).$$

KL captures the information lost when the restricted joint distribution $\Pi = \{\pi_i\}$ is used to approximate $P = \{p_i\}$. From a statistical perspective, Kullback (1959, page 158) shows that this measure multiplied by the number of observations can be interpreted as the likelihood ratio test for

$H_0 : P = \Pi$. Hence, the index M can be interpreted as the log-likelihood ratio test for the statistical independence of race status and school membership divided by the number of students.¹⁴ MRC establish the relation between the M index and the likelihood-ratio test of independence between race and school membership in a general statistical framework in the presence of conditioning discrete or continuous variables.

II. 5. The Normalization Issue

Normalization properties concern the bounds for the range of admissible values for an index of segregation. Most researchers would identify the absence of segregation with the situation where organizational units have the same racial composition or, equivalently, where demographic groups have the same distribution across organizational units. Similarly, most researchers would accept that demographic groups are completely segregated whenever they do not mix at all within organizational units. A segregation index is said to be *normalized in the unit interval*—or possess the **NOR** property— if it takes value 0 whenever there is no segregation and it takes value 1 whenever it reaches complete segregation as defined above.

It is important to understand that requiring the subsidiary property **NOR** has larger implications than simply rescaling the measure of segregation. In particular, the un-normalized and the normalized entropy-based indices do not generally give the same segregation ordering. Both H and H^* rank all cities with no racial mixing within schools as equally segregated, while M assigns a higher segregation level to cities in which there is more initial uncertainty about a student's racial group. Following an example in FV, consider city **A** with three schools and three racial groups and city **B** with two schools and two racial groups, such that

¹⁴ For a similar motivation based on the comparison of prior and posterior probabilities, see Flückiger and Silber (1999, pp. 69), and Zoloth (1974, pp 14-16 and 1976). See Reardon *et al.* (2000) for the relation between the M index and a contingency table test for the independence of two variables.

$$\mathbf{A} = \begin{bmatrix} 50 & 0 & 0 \\ 0 & 50 & 0 \\ 0 & 0 & 50 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 50 & 0 \\ 0 & 50 \end{bmatrix}.$$

Given each city's marginal distributions, segregation is at a maximum in both cities according to the three indexes. Both H and H^* assign to each city a segregation value of 1. However, learning a student's school (racial group) in \mathbf{A} conveys more information about a student's race (school) than in \mathbf{B} . Consequently, segregation in \mathbf{A} is larger than in \mathbf{B} according to the M index: $M(\mathbf{A}) = 1.10$ and $M(\mathbf{B}) = 0.69$. Consider now a third completely segregated city \mathbf{C} :

$$\mathbf{C} = \begin{bmatrix} 99 & 0 \\ 0 & 1 \end{bmatrix}$$

Both H and H^* assign again to \mathbf{C} a segregation value of 1. However, since there is much less uncertainty about a student's racial group (school) in \mathbf{C} than in either \mathbf{A} or \mathbf{B} , segregation in \mathbf{C} according to M is much smaller than before: $M(\mathbf{C}) = 0.06$.

As was pointed out in Clotfelter (1979), a critical problem with segregation indices that satisfy *NOR* is that they fail to capture well changes in inter-racial contact. Compare the effect of merging the two schools in city \mathbf{C} , yielding the one-school city represented by column vector $[99 \ 1]'$, with the effect of merging the two schools in \mathbf{B} , yielding the one-school city represented by $[50 \ 50]'$. The first merger has a very small effect on the inter-racial exposure of the average student, while the second one has a much larger effect: each student switches from a completely segregated school to one that is completely integrated. The M index reflects this difference, falling by 0.06 in \mathbf{C} versus 0.69 in \mathbf{B} . In contrast, H and H^* miss the difference because the segregation value they both assign decreases by 1 in the two cases.

It can be concluded that there are conceptual reasons for not requiring subsidiary property *NOR* from a segregation index. In the next section we will refer to a result by FV (about the

incompatibility of *NOR* and certain decomposability properties) to argue that in empirical studies we may want to avoid indexes that satisfy *NOR* for practical reasons.

On the other hand, it should be noted that all segregation indices that are bounded above can be weakly normalized, in the sense that they can be expressed as proportions of maximum segregation, by simply dividing them by its maximum value. In particular, in equation (6) it can be seen that the M index reaches its maximum at the smallest value between $\log(G)$ and $\log(N)$. Given that in most empirical applications $\log(G) < \log(N)$, normalizing M in this weak sense is simply equivalent to computing the logarithm in base G . The resulting measure can be interpreted as the proportion of maximum deviation from representativeness. However, this exercise is not very useful for two reasons. Firstly, the most robust feature of the index, namely the ranking it induces, is still the same and captures both departures from representativeness and evenness. Secondly, although the resulting index takes values in the unit interval it does not satisfy *NOR*.

III. DECOMPOSABILITY PROPERTIES

III.1. Strong School Decomposability

In many research situations it is useful to partition organizational units into subgroups or clusters of different size. Following TF, consider a partition of the N schools into $K < N$ school districts indexed by $k = 1, \dots, K$. Let \mathbf{X}^k be the set of schools which belong to district k , and N_k be the number of schools in \mathbf{X}^k with $\sum_{k=1}^K N_k = N$. The data available in \mathbf{X}^k can be organized into the following GN^k

matrix:

$$\mathbf{X}^k = \left\{ t_{gn^k} \right\} = \begin{bmatrix} t_{11^k} & \cdots & t_{1N^k} \\ \vdots & \ddots & \vdots \\ t_{G1^k} & \cdots & t_{GN^k} \end{bmatrix}$$

where t_{gn^k} denotes the number of individuals of racial group g attending school n^k in district k . School and race frequencies at city level simply result from horizontal grouping of the school and race frequencies from all K districts, $\mathbf{X} = [\mathbf{X}^1 \dots \mathbf{X}^k \dots \mathbf{X}^K]$. Assume now that all schools in district k have the same racial composition as the district as a whole, i. e. let $\tilde{\mathbf{X}}^k$ refer to the district such that $p_{g|n^k} = p_{g|k}$ for all n^k and all g , or the district in which the N^k original schools have been combined into a single school with conditional racial distribution $P_{g|k} = \{p_{g|k}\}$.¹⁵ Then $S(\tilde{\mathbf{X}}^k) = 0$ for every $k = 1, \dots, K$, according to any sensible segregation index S . Would this mean that city segregation should be equal to zero? As long as the racial composition of at least two districts differ from each other, it is to be expected that overall city segregation should be positive and equal to “between-districts” segregation, that is $S(\mathbf{X}) = S(\tilde{\mathbf{X}}^1, \dots, \tilde{\mathbf{X}}^K)$ where $S(\tilde{\mathbf{X}}^1, \dots, \tilde{\mathbf{X}}^K) \equiv S([\tilde{\mathbf{X}}^1 \dots \tilde{\mathbf{X}}^K])$.

These considerations motivate a decomposability property for a segregation index according to which, for any partition of the N schools into $K < N$ clusters, overall segregation can be expressed as the sum of two terms, one that captures between-groups segregation, and one that captures within-groups segregation and is equal to the weighted average of segregation levels within each of the clusters, with weights independent of the level of segregation within them.¹⁶ Generally, it would be convenient to have the weights adding up to unity. Moreover, it is natural to require that the weights coincide with the demographic importance of each cluster. Thus, we have

Definition 1. A school segregation index S is said to be *strongly school decomposable*, **D1**, if and only if for any partition of the set of N schools into $K < N$ clusters, so that

¹⁵ Alternatively, suppose that all racial groups in a district are equally distributed over the district schools.

¹⁶ The notion of cluster is introduced to refer to the elements of any partition of the set of schools. Examples of clusters are the set of public or private schools in a country, or the sets of schools in major regions, states, cities, school districts or neighborhoods.

$\mathbf{X} = [\mathbf{X}^1 \dots \mathbf{X}^k \dots \mathbf{X}^K] \in \mathfrak{E}(N, G)$, $\mathbf{X}^k \in \mathfrak{E}(G, N_k)$, $k = 1, \dots, K$, and $\sum_{k=1}^K N_k = N$, overall segregation,

$S(\mathbf{X})$, can be written as

$$S(\mathbf{X}) = S(\tilde{\mathbf{X}}^1, \dots, \tilde{\mathbf{X}}^K) + \sum_{k=1}^K p_k S(\mathbf{X}^k), \quad (7)$$

where $\tilde{\mathbf{X}}^k$ refers to the cluster in which $p_{g|n^k} = p_{g|k}$ for all n^k and g , and p_k is the proportion of

students in cluster k , $p_k = (1/t) \sum_{n^k \in \mathbf{X}^k} \sum_{g=1}^G t_{gn^k}$.

For any partition of schools into clusters, we have to make sure that the following three questions are well defined: (i) the contribution to overall segregation of any individual cluster; (ii) the part of overall segregation accounted for by segregation within all clusters, and (iii) how much segregation can be attributed to racial differences across clusters of different size.

In the first place, note that if we are merely interested in ranking clusters' segregation levels the decomposability requirement is quite inessential. However, if the analysis involves comparisons between subgroup and overall levels, then decomposability can be very useful indeed. As pointed out in the field of income inequality, a problem arises in the different interpretations that can be placed in statements like “ α per cent of overall segregation is attributed to cluster k ” (see, *inter alia*, Shorrocks 1980, 1984, 1988). Fortunately, definition **D1** implies a satisfactory way of assigning segregation contributions to the subgroups. For, when equation (7) holds for any partition of N schools into K clusters, it seems natural to define the contribution to overall segregation of cluster k by:

$$C_k = p_k S(\mathbf{X}^k). \quad (8)$$

It is easy to check that this definition of C_k is consistent with the other two obvious interpretations of the sentence “contribution to segregation of cluster k ”. First, consider the situation in which the original frequencies of students across races and schools in the city is replaced by one in which all

schools in cluster k are incorporated into a single school. Since in this case $S(\tilde{\mathbf{X}}^k) = 0$, then from equation (6) it is immediate to see that

$$C_k = S(\mathbf{X}) - S(\mathbf{X}^1, \dots, \mathbf{X}^{k-1}, \tilde{\mathbf{X}}^k, \mathbf{X}^{k+1}, \dots, \mathbf{X}^K),$$

i.e. the contribution C_k can also be interpreted as the amount by which overall segregation falls if the segregation within cluster k is eliminated. Second, consider the situation by which the original joint frequencies are replaced by one in which all clusters except k become single school clusters. Since in this situation $S(\tilde{\mathbf{X}}^j) = 0$, for all $j \neq k$, it follows that

$$C_k = S(\tilde{\mathbf{X}}^1, \dots, \tilde{\mathbf{X}}^{k-1}, \mathbf{X}^k, \tilde{\mathbf{X}}^{k+1}, \dots, \tilde{\mathbf{X}}^K) - S(\tilde{\mathbf{X}}^1, \dots, \tilde{\mathbf{X}}^K),$$

i.e. C_k can also be interpreted as the amount by which overall segregation increases if segregation within cluster k is introduced starting from the position of zero segregation within each cluster. Therefore, under **D1** it is possible to provide the same answer to different interpretations of what is meant by the contribution of each cluster to overall segregation. Consequently, the problem of unambiguously comparing individual clusters' contributions is solved. For example, the ratio $S(\mathbf{X}^k) / S(\mathbf{X})$ will be greater, equal or smaller than one whenever subgroup k 's contribution to the overall segregation level, $C_k / S(\mathbf{X})$, is greater than, equal to, or smaller than its demographic importance given by p_k .

In the second place, we must examine the contribution made to overall segregation by all clusters taken together, C . This question admits two sensible interpretations. First, a natural response is to compute the reduction in overall segregation that would arise if the segregation within all clusters were eliminated. In the partition into K clusters C will be:

$$C = S(\mathbf{X}) - S(\tilde{\mathbf{X}}^1, \dots, \tilde{\mathbf{X}}^K).$$

A second interpretation would consist of the sum of the individual contributions defined in expression (8), that is,

$$\sum_{k=1}^K C_k \equiv \sum_{k=1}^K p_k S(\mathbf{X}^k).$$

It is immediate to see that for any segregation measure S satisfying **D1**, $C = \sum_{k=1}^K C_k$ so that both interpretations provide the same answer.

Finally, consider the possibility of partitioning the set of schools in a country into clusters of different size, say regions, cities, or school districts. The empirical question to be addressed is “How much segregation can be attributed to racial differences across regions as opposed to other geographical levels.” The segregation attributed to racial differences across clusters of a given size may be interpreted as meaning: (i) how much less segregation would be observed if racial differences across clusters were the only source of school segregation, or (ii) by how much segregation would fall if racial differences at the cluster level were eliminated. Interpretation (i) suggests a comparison of overall segregation with the amount that would arise if segregation within each of K clusters were made equal to zero but racial differences across districts remained the same. As was seen before, for measures satisfying **D1** this would eliminate the total within-clusters term and leave only the between-clusters contribution so that $S(\mathbf{X}) = S(\tilde{\mathbf{X}}^1, \dots, \tilde{\mathbf{X}}^K)$. Interpretation (ii) suggests a comparison of overall segregation with the segregation level that would result if all clusters had the same racial composition, equal to the one for the nation as a whole, but the segregation within each cluster remained unchanged. Unfortunately, in contrast to the situation for relative measures of income inequality, this conceptual experiment is not possible for measures of segregation, a difficulty that deserves an explanation.

Let \mathbf{y} be a vector in Euclidean space E^N representing an income distribution for a population of N individuals, and let I be an index of *relative* income inequality, that is, a scale invariant inequality index for which $I(\lambda\mathbf{y}) = I(\mathbf{y})$ for all $\lambda > 0$. Consider any partition of $\mathbf{y} \in E^N$ into K subgroups, such that

$\mathbf{y} = (\mathbf{y}^1, \dots, \mathbf{y}^K)$ where $\mathbf{y}^k \in E^k$, $k \geq 2$, and $\sum_k E^k = E^N$. If I is an additively decomposable income inequality index in the sense of definition (7), then

$$I(\mathbf{y}) = B + \sum_r p_r I(\mathbf{y}^r), \quad (9)$$

where B is the inequality of the distribution where each individual is assigned the mean income of the subgroup to which she belongs, $\mu(\mathbf{y}^k)$, that is $B = I[\mu(\mathbf{y}^1), \dots, \mu(\mathbf{y}^K)]$. In that situation, let $\mathbf{y}^* = (\mathbf{y}^{*1}, \dots, \mathbf{y}^{*K})$ be the income distribution satisfying two conditions: (a) $\mu(\mathbf{y}^{*k}) = \mu(\mathbf{y})$, for all k , where $\mu(\mathbf{y})$ is the mean income for the entire population, and (b) $I(\mathbf{y}^k) = I(\mathbf{y}^{*k})$, so that income inequality within each subgroup is preserved.¹⁷ Then we must have

$$B = I(\mathbf{y}) - I(\mathbf{y}^*).$$

That is, according to interpretation (ii), between-groups income inequality is the amount by which overall income inequality is reduced when the differences between subgroup income means are eliminated by making them equal to the population income mean.¹⁸

The corresponding conceptual exercise in the segregation case is impossible. Starting from $\mathbf{X} = [\mathbf{X}^1 \dots \mathbf{X}^k \dots \mathbf{X}^K]$, assume that segregation falls if racial differences at the cluster level are eliminated. Let \mathbf{Y} be another country such that (a) the racial composition of every cluster k in \mathbf{Y} is the same as the one for the original population as a whole, that is, $p_{g|k} = p_g$ for all k and g , and (b) the level of segregation within each cluster remains as in the original country. Condition (a) implies that there is no between-group segregation in \mathbf{Y} . Since within-group segregation in \mathbf{Y} results from the comparison between the racial distribution at school level with the racial distribution in the original country, condition (b) implies that $S(\mathbf{Y}) = S(\mathbf{X})$ and, therefore, $S(\mathbf{X}) - S(\mathbf{Y}) = 0$, which contradicts the

¹⁷ Notice that this conceptual exercise is always admissible as long as income averages for all subgroups are strictly positive since then $\mathbf{y}^{*k} = [\mu(\mathbf{y})/\mu(\mathbf{y}^k)]\mathbf{y}^k$.

¹⁸ As a matter of fact, the answers to interpretations (i) and (ii) coincide and are equal to the between-groups term only when the weights in the within-groups term do not depend on the subgroup means, as in equation (9) above. This is only the case for one of the members of the entropy family of income inequality indicators: the mean logarithmic deviation (see Shorrocks, 1980).

assumption that segregation would fall if racial differences were eliminated at cluster level. In other words, in the segregation context it is impossible to eliminate the between-groups segregation maintaining the existing within-groups segregation as the former affects the latter. Nevertheless, this does not preclude the investigation of the original question about which geographical level accounts for a greater percentage of overall segregation. For any segregation measure satisfying **D1**, the size of the between-groups term at each geographical level provides a clear answer, if only in the sense of interpretation (i).

III.2. Strong Group Decomposability

In many research situations it is useful to partition demographic groups into supergroups. Consider a partition of the G groups in a city $\mathbf{X} \in \mathfrak{E}(G, N)$ into $L < G$ supergroups, indexed by $l = 1, \dots,$

L . Let \mathbf{X}_l be supergroup l and G_l its cardinal with $\sum_{l=1}^L G_l = G$. The data available in \mathbf{X}_l can be organized

into the following $G_l N$ matrix:

$$\mathbf{X}_l = \{t_{g_l n}\} = \begin{bmatrix} t_{1,1} & \cdots & t_{1,N} \\ \vdots & \ddots & \vdots \\ t_{G_l,1} & \cdots & t_{G_l,N} \end{bmatrix}.$$

where $t_{g_l n}$ denotes the number of individuals of racial group g_l in supergroup \mathbf{X}_l attending school n .

School and race frequencies at city level simply result from vertical grouping of the frequencies from

all L supergroups, $\mathbf{X} = [\tilde{\mathbf{X}}_1^T \quad \dots \quad \tilde{\mathbf{X}}_L^T]^T$, where superscript T stands for the transpose operator.

Suppose that the G_l groups in supergroup l have the same distribution over organizational units as the supergroup as a whole, i. e. let $\tilde{\mathbf{X}}_l$ be the supergroup in which $p_{n|g_l} = p_{n|l}$ for all g_l and n , or the supergroup in which the G_l original groups have been combined into a single group with conditional

school distribution $P_{nl} = \{p_{nl}\}$.¹⁹ Then $S(\tilde{\mathbf{X}}_l) = 0$ for every $l = 1, \dots, L$, according to any sensible segregation index S . Would this mean that city segregation should be equal to zero? As long as the spatial distribution of at least two supergroups would differ from each other, it is to be expected that overall city segregation should be positive and equal to “between-supergroups” segregation, or $S(\mathbf{X}) = S(\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_L)$ where $S(\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_L) \equiv S\left(\left[\tilde{\mathbf{X}}_1^T \ \dots \ \tilde{\mathbf{X}}_L^T\right]^T\right)$.

This motivates a decomposability property for a segregation index according to which, for any partition of the G racial groups into supergroups, overall city segregation can be expressed as the sum of two terms, one that captures between-supergroups segregation, and another that captures within-supergroups segregation and is equal to the weighted average of segregation within each of the supergroups. Again, it is natural to require that the weights coincide with the supergroups demographic importance. Thus, we have

Definition 2. A school segregation index S is said to be *strongly group decomposable (D2)* if and only if for any partition of the G groups into $L < G$ supergroups so that $\mathbf{X} = \left[\tilde{\mathbf{X}}_1^T \ \dots \ \tilde{\mathbf{X}}_L^T\right]^T \in \mathfrak{E}(G, N)$, $\mathbf{X}_l \in \mathfrak{E}(G_l, N)$, $l = 1, \dots, L$, and $\sum_{l=1}^L G_l = G$, overall segregation, $S(\mathbf{X})$, can be written as

$$S(\mathbf{X}) = S(\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_L) + \sum_{l=1}^L p_l S(\mathbf{X}_l), \quad (10)$$

where $\tilde{\mathbf{X}}_l$ refers to the supergroup in which $p_{n|g_l} = p_{nl}$ for all g_k and n , and p_l is the proportion of students in supergroup l , $p_l = (1/t) \sum_{g_l \in \tilde{\mathbf{X}}_l} \sum_{n=1}^N t_{g_l n}$.

As before, this definition implies a satisfactory way of assigning segregation contributions to the supergroups. For, when equation (10) holds, the definition $C_l = p_l S(\mathbf{X}_l)$, is consistent with all the

¹⁹ Alternatively, suppose that all organizational units have the same racial composition within each supergroup.

obvious interpretations of the concept “contribution to segregation by supergroup l ”: the amount by which overall segregation falls if the segregation within supergroup l is eliminated, or the amount by which overall segregation increases if segregation within supergroup l is introduced starting from the position of zero segregation within each supergroup. Also, as before, in the segregation context it is impossible to eliminate the between-supergroups segregation maintaining the existing within-supergroups segregation as the latter is affected by the former.

III.3. Entropy-Based Indexes and Decomposability Properties

Mora and Ruiz-Castillo (2003) show that the M index satisfies **D1** in the two-group case, while FV show that it satisfies both **D1** and **D2** in the multigroup case. The version of equation (7) satisfied by M admits two interpretations according to both notions of segregation, representativeness and evenness:

$$\begin{aligned} M &= M\left(P_g, \left\{P_{g|k}\right\}_{k=1}^K\right) + \sum_{k=1}^K p_k M\left(P_{g|k}, \left\{P_{g|n}\right\}_{n \in X^k}\right) \text{ (representativeness)} \\ &= M\left(P_k, \left\{P_{k|g}\right\}_{g=1}^G\right) + \sum_{k=1}^K p_k M\left(P_{n|k}, \left\{P_{n|g,k}\right\}_{g=1}^G\right) \text{ (evenness)} \end{aligned} \quad (11)$$

TF only refer to the within-districts weighted average at the city level of expression (11), i.e.

$$\sum_{k=1}^K p_k M\left(P_{g|k}, \left\{P_{g|n}\right\}_{n \in X^k}\right) \text{ (see the last paragraph in page 191, as well as equations (2.3) and (2.4) in$$

Theil, 1972, p. 123). This is as far as TF and Theil (1971, 1972) went: no citywide index was formally discussed. Afterwards, in an occupational context with two genders, Fuchs (1975) proposed a nationwide, entropy-based segregation index inspired in equation (2). Similarly, in the multigroup case the version of equation (10) satisfied by M the decomposition admits again a representativeness and an evenness interpretation:

$$\begin{aligned} M &= M\left(P_l, \left\{P_{l|n}\right\}_{n=1}^N\right) + \sum_{l=1}^L p_l M\left(P_{g|l}, \left\{P_{g|n,l}\right\}_{n=1}^N\right) \text{ (representativeness)} \\ &= M\left(P_n, \left\{P_{n|l}\right\}_{l=1}^L\right) + \sum_{l=1}^L p_l M\left(P_{n|l}, \left\{P_{n|g}\right\}_{g \in X_l}\right) \text{ (evenness)} \end{aligned} \quad (12)$$

Interestingly, FV fully characterize the ranking induced by the M index in terms of eight ordinal properties (Theorem 1, p.19). In particular, they indicate that although cardinal properties such as $D1$ and $D2$ might be very convenient in empirical work, their implications for how an index actually ranks pairs of cities are often unclear. Thus, they introduce three ordinal axioms that are meant to capture the rules that an ordering induced by a strongly decomposable segregation index in the sense of $D1$ and $D2$ should follow in ranking cities. For any partition of schools into clusters, the first axiom requires that the measure should be *responsive* to changes in the clusters, in the sense that if the students in a cluster are reallocated so as to raise segregation in that cluster, with no changes in the rest of the clusters, the segregation in the city should not fall. This is *Type I Independence* or $IND1$.²⁰ From a practical point of view, $IND1$ is needed to coordinate the efforts of a city's decentralization strategy towards, for example, a reduction of racial segregation. For such strategy may typically involve a set of policy measures targeted at, say, a reduction of segregation in each school district. If the segregation indicator is not responsive we may be faced with a situation in which each district achieves the objective of reducing its own segregation level, and yet the city's segregation level increases. Responsiveness may therefore be viewed as an essential counterpart to a coherent segregation-reducing policy program, since it merely ensures that the overall segregation value does not respond perversely to changes in the level of segregation within one cluster while the level in the other clusters stays constant. The second axiom requires that the index be *consistent* in the treatment of segregation within-subgroups and segregation between them. By this FV mean that the relative importance the measure assigns to these two aspects of segregation should not change if students are reallocated within a cluster. This is *Type II Independence* or $IND2$. $IND1$ and $IND2$ are shown to be necessary ordinal conditions for cardinal property $D1$ to hold. Finally, in any partition of the demographic

²⁰ It coincides with the *Subgroup Consistency* property in the income inequality or the poverty literature, as can be seen in Shorrocks (1988) and in Foster and Shorrocks (1991), respectively.

groups into supergroups FV suggest that a measure should be *unsuspicious* in the sense that it should not impute segregation where there is no direct evidence for it. For example, if the researcher only has data about whites and blacks, the measure should not impute a positive level of segregation between white subgroups, such as Hispanics versus white non-Hispanics. This axiom, called the *Group Division Property* or **GDP**, is shown to be necessary for **D2** to hold.²¹

On the other hand, FV obtain a very interesting result indicating that any segregation index which satisfies **NOR** and treats racial groups symmetrically must violate **D1** (Proposition 3, p. 22). Thus, any normalized index, including H and H^* , cannot be strongly decomposable in the **D1** sense. On the other hand, neither H nor H^* satisfy **D2**. However, they can be seen to satisfy some weaker decomposability properties.

Consider first the H index. Starting from the representativeness representation of decomposition (11) and noting that H can be computed by dividing the M index by the racial entropy, we have:

$$H = \frac{M\left(P_g, \{P_{g|k}\}_{k=1}^K\right)}{E(P_g)} + \sum_{k=1}^K p_k \frac{M\left(P_{g|k}, \{P_{g|n}\}_{n \in X^k}\right)}{E(P_g)}.$$

Multiplying and dividing by the racial entropy in every cluster k in the within-group term and using the relation between the un-normalized and the normalized indexes, then for every partition of the N schools into $K < N$ clusters we have:

$$H = H\left(P_g, \{P_{g|k}\}_{k=1}^K\right) + \sum_{k=1}^K p_k \frac{E(P_{g|k})}{E(P_g)} H\left(P_{g|k}, \{P_{g|n}\}_{n \in X^k}\right). \quad (13)$$

For any given joint distribution of racial groups and clusters, $P_{g|k}$, the weights in decomposition (13),

$p_k \frac{E(P_{g|k})}{E(P_g)}$, are constant so that if all within clusters segregation measures $H\left(P_{g|k}, \{P_{g|n}\}_{n \in X^k}\right)$ are

²¹ The remaining five axioms that FV employ to fully characterize the M index are three basic properties (*Weak Scale Independence*, the property already mentioned in note 7, *Symmetry*, and *School Division Property*), and two technical properties (*Continuity* and a property needed to rule out the trivial segregation ordering).

increased, then the entropy index H will unambiguously increase. Thus, H satisfies **IND1** (although the weights do not necessarily add up to unity). It remains to be seen whether the terms in decomposition (13) admit the same interpretations as those terms in any **DI** index. Define cluster k 's

contribution to overall segregation as $C_k^* = p_k \frac{E(P_{g|k})}{E(P_g)} H\left(P_{g|k}, \{P_{g|n}\}_{n \in X^k}\right)$. It is easy to show that C_k^*

can be interpreted both as the amount by which overall segregation falls if the segregation within cluster k is eliminated, and the amount by which overall segregation increases if segregation within cluster k is introduced starting from the position of zero segregation within each cluster. Likewise,

define the contribution of all clusters to segregation as $C^* = \sum_{k=1}^K C_k^*$. It turns out that C^* equals the

reduction in segregation that would arise if the segregation within all subgroups were eliminated.

Finally, the interpretation of the between-groups term in decomposition (12), $H\left(P_g, \{P_{g|k}\}_{k=1}^K\right)$, is also

subject to the conceptual limitation pointed out in the previous sub-section in relation to the decomposition of any **DI** index. Namely, the between-groups term can be interpreted as the level of

segregation if racial differences across clusters were the only source of school segregation (in which

case $H\left(P_{g|k}, \{P_{g|n}\}_{n \in X^k}\right) = 0$ for all $k = 1, \dots, K$). However, it cannot be interpreted as the decrease in

segregation if racial differences at the cluster level were eliminated. To summarize, the index H

satisfies **IND1** and its decomposition into clusters admits the same interpretation as that of any **DI**

index.

Let us now consider a partition of the racial groups into supergroups. It can be seen that, starting again from the representativeness representation of equation (12), for every partition of the G racial groups into $L < G$ supergroups we have:

$$H = \frac{E(P_l)}{E(P_g)} H\left(P_l, \{P_{l,n}\}_{n=1}^N\right) + \sum_{l=1}^L p_l \frac{E(P_{gl})}{E(P_g)} H\left(P_{gl}, \{P_{gl,n,l}\}_{n=1}^N\right).^{22} \quad (14)$$

Note that for a given joint distribution of supergroups and schools, P_{ln} , the weights in decomposition

(14), $p_l \frac{E(P_{gl})}{E(P_g)}$, generally change in response to exogenous changes in the joint distribution of

groups and schools within supergroups. Of course, supergroup demographic shares, p_b , remain

constant, but the overall racial entropy at group level, $E(P_g)$, or the racial entropy at group level in

supergroup l , $E(P_{gl})$, may change. Consequently, the H index may decrease even if all within

supergroups terms increase; in other words, H need not be responsive to all partitions of racial groups

into supergroups. Moreover, the contribution $C_l^* = p_l \frac{E(P_{gl})}{E(P_g)} H\left(P_{gl}, \{P_{gl,n,l}\}_{n=1}^N\right)$ cannot generally be

interpreted as the amount by which overall segregation falls if the segregation within supergroup l is

eliminated. The reason is that when the segregation within supergroup l is eliminated by changes in the

schools' racial composition, the overall racial entropy of the city $E(P_g)$ will usually change. This will

induce changes in the weights for all within terms and, hence, in contributions C_j^* , $j \neq l$. A similar

argument can be used to show that C_l^* cannot generally be interpreted as the amount by which overall

segregation increases if segregation within supergroup l is introduced starting from the position of

zero segregation within each racial supergroup. Finally, the between-supergroups term in

decomposition (13) cannot be interpreted either as the level of segregation if differences in the

supergroup distributions across schools were the only source of school segregation or as the decrease

in segregation if differences in the supergroups distributions into schools were eliminated.²³ Therefore,

²² Equations (13) and (14) figure prominently in Reardon *et al.* (2002) -see their expressions (5) and (4), respectively.

²³ For the sake of brevity, proofs of the statements in this paragraph will be only available on request.

equation (14) does not provide the H index with a decomposition that admits the same interpretation as that of any $D2$ index.

As far as the H^* index is concerned, it can be seen that, starting from the evenness representation of decomposition (12), for every partition of the G groups into $L < G$ supergroups we have:

$$H^* = H\left(P_n, \{P_{n|l}\}_{l=1}^L\right) + \sum_{l=1}^L p_l \frac{E(P_{n|l})}{E(P_n)} H\left(P_{n|l}, \{P_{n|g}\}_{g \in X_l}\right). \quad (15)$$

Similarly, starting from the evenness representation of equation (11), it can be seen that for every partition of the N schools into $K < N$ clusters we have:

$$H^* = \frac{E(P_k)}{E(P_n)} H\left(P_k, \{P_{k|g}\}_{g=1}^G\right) + \sum_{k=1}^K p_k \frac{E(P_{n|k})}{E(P_n)} H\left(P_{n|k}, \{P_{n|g,k}\}_{k=1}^K\right). \quad (16)$$

For reasons of brevity, the properties of decompositions (15) and (16) are not discussed in detail. Nevertheless, similar arguments to those provided for the decompositions of the H index can be used to see that the H^* index is responsive and the terms in decomposition (15) can be interpreted as those in the decomposition of any $D2$ index for any partition of racial groups into supergroups. However, the H^* index does not satisfy $IND1$ and the terms in decomposition (16) do not admit the interpretations of the terms in any $D1$ index for any partition of schools into clusters.

A final issue needs to be raised. Among others, the following two concerns have attracted the attention of applied researchers on segregation: the statistical significance of segregation indices, and the control for the statistical association between demographic groups and schools and other socioeconomic variables. The empirical decomposition of segregation indices is often carried out with small samples and the significance of the terms in the decomposition must be assessed. This can only be done under alternative hypothesis if the measure is explicitly embedded within a statistical framework. To our knowledge, only the M index and its decompositions have been placed into a general statistical framework,

and their asymptotic properties have been fully characterized. Furthermore, MRC generalize FV's decomposition to condition segregation on any vector of (possibly continuous) student and school characteristics x , therefore providing an intuitive unifying econometric framework for studies of segregation using segregation indices and econometric models.

IV. INVARIANCE PROPERTIES

The M , H , and H^* indexes may give a rather different picture of what is taking place in empirical studies of segregation trends. Consider, as an illustration, the evolution of the U.S. student population enrolled in public schools in Core-Based Statistical Areas (CBSAs) –urban clusters of 10,000 or more inhabitants, referred to in the sequel as *cities*– during the 1989-90 and 2005-06 academic years.²⁴ Table 1 provides a glimpse into the evolution of racial diversity of the student population: minorities (namely, Native Americans, blacks, Asians, and Hispanics) already represent 34.8% of the total population of 24.8 million in 1989; furthermore, since all of them grew more rapidly than whites during this period, they represent as much as 48.1% of the total population of 25.5 million in 2005.

Table 1 around here

Taking into account equations (3) and (5), the change in the mutual information index M (multiplied by 100) during this period can be written as

$$\begin{aligned} \Delta M &= (119.07 - 70.17) - (101.27 - 57.35) \\ &= (1035.72 - 986.82) - (1040.25 - 996.32) \\ &= 48.90 - 43.92 = 4.98. \end{aligned}$$

²⁴ Results pertain to those schools for which racial and ethnic information is available both in 1989 and in 2005. Given that a small proportion of schools did not report results in 1989, focusing on the schools which did probably gives a fairer comparison between the distributions observed in 1989 and in 2005 because it does not include those schools that did report in 2005 but failed to do so in 1989. However, interpretability of the results presented here is potentially compromised by the fact that some schools have been created whilst others have disappeared between 1989 and 2005. Nevertheless, results using all observations are qualitatively similar, suggesting that the selection mechanisms at work are not essential to our analysis. Results obtained using the full sample are available upon request.

Given equation (4) and the observed increase in the racial entropy, $119.07 - 101.27 = 17.80$, it comes as no surprise that according to the H index school segregation decreases during this period:

$$\begin{aligned}\Delta H &= (48.90/119.07) - (43.92/101.27) \\ &= 41.07 - 43.37 = -2.30\end{aligned}$$

Given equation (6) and the observed decrease in the spatial entropy, $1035.72 - 1040.25 = -4.53$, it comes as no surprise that according to the H^* index school segregation increases during this period:

$$\begin{aligned}\Delta H^* &= (48.90/1035.72) - (43.92/1040.25) \\ &= 4.72 - 4.22 = 0.50.\end{aligned}$$

As it is well known, neither of the entropy-based segregation indices are invariant to changes in the marginal distributions by racial group and/or schools. In terms of the subsidiary properties referred to in note 5, the M , H and H^* indices violate both **I1** and **I2**. Although in our opinion there are good reasons for recognizing the demographic importance of racial groups and schools in a measure of segregation at a given moment in time, it is quite clear that in segregation comparisons across time or space it becomes extremely useful to evaluate those changes in segregation which do not result from changes in the marginal distributions by race and schools.²⁵ To do so, two strategies can be followed. In the first place, segregation can be measured using invariant segregation indices. In practice, the literature has emphasized **I1** indices, such as the general Atkinson index (see Frankel and Volij, 2008, for a discussion and characterization results). The key shortcoming of this strategy is that **I1** indices only neutralize changes in the racial marginal distribution but in doing so they mix up the effect of changes in the conditional distribution that focuses on the school distribution by race, $\Delta P_{n|g}$ with the impact of changes in the marginal distribution by school, ΔP_n .

In the second place, one may use a segregation index that is neither **I1** nor **I2** but whose inter-

²⁵ For a full discussion of these issues, see Mora and Ruiz-Castillo (2009b), and for an empirical application to school segregation see Mora and Ruiz-Castillo (2009d).

temporal changes admit a decomposition that isolates a term that captures segregation changes net of the impact of pure demographic factors.²⁶ This is the case of the M index, which can be decomposed in two complementary ways: the first decomposition isolates an $I1$ term, maintaining constant the marginal distribution over racial groups and the entropy of the population school distribution, while the second decomposition isolates an $I2$ term, maintaining constant the marginal distribution over schools and the entropy of the population marginal distribution over races.

The results for both decompositions are the following. On the one hand, the change in the M index due to the change in the schools entropy is -4.53 while the change due to the change in the marginal distribution of racial groups is 10.63 . Therefore, the change in net segregation independent of these effects is

$$\Delta\text{Net}(I1) = 4.98 - (-4.53) - 10.63 = -1.11.$$

On the other hand, the change in the M index due to the change in the racial entropy is 17.80 while the change due to the change in the marginal distribution of schools is -0.59 . Therefore, the change in net segregation independent of these effects is

$$\Delta\text{Net}(I2) = 4.98 - (-0.59) - 17.80 = -12.23.$$

The first invariant term, $\Delta\text{Net}(I1)$, reflects changes in the groups' conditional distributions over schools and, therefore, can be interpreted as changes in evenness. The second invariant term, $\Delta\text{Net}(I2)$, reflects changes in the schools' racial mix and, therefore, can be interpreted as changes in representativeness. However, nothing of the sort is available for the H and H^* indices.

V. CONCLUSIONS

This paper adopts the methodological criterion that, as in the income inequality literature, one

²⁶ In the context of occupational segregation by gender, many authors have defended this strategy before (see, *inter alia*, Blau and Hendricks, 1979, Jonung, 1984, Beller, 1985, and Watts, 1992, 1998).

way to select an adequate segregation measure is to study which basic and subsidiary but useful properties different indices satisfy. We have discussed three types of subsidiary properties as they apply to three entropy-based segregation indices, M , H , and H^* .

First, it is often convenient to have segregation measures with the subsidiary property of additive decomposability. In a decomposition context, consider the notions of “contribution to overall segregation by a subgroup k , or by all subgroups together in a certain partition”, or consider the question of “how much segregation can be attributed to a given discrete variable”. As in the income inequality or the economic poverty literature, it is not always possible that all intuitive interpretations of these questions coincide under a certain decomposability property. For the first time in the literature, in this paper it has been shown that these questions receive the more unambiguous answers that are possible in a segregation context under the decomposability properties ***D1*** and ***D2*** that are only satisfied by the M index. Moreover, FV have unveiled the axioms that capture the rules that the ranking induced by a segregation index satisfying ***D1*** and ***D2*** follows when comparing any two cities. In particular, one of these axioms requires that the ordering should be responsive to segregation changes at the subgroup level. Since the M index satisfies ***D1***, it is always responsive in this sense –a situation that is not the case for the H and H^* indices.

Second, the invariance properties that require a segregation measure to be independent from changes in the relative importance of demographic groups or organizational units have also greatly concerned many authors in the segregation field. The M index is not invariant in this sense but changes in overall segregation according to the M index can be decomposed in two complementary ways to isolate terms that capture changes in net segregation independent of variations in the marginal distributions of schools and racial groups. No such decompositions are available to the H and the H^* indices. When such demographic changes are important, as it has been shown to be the case when assessing the change in school segregation in the U.S. during 1989-2005, this is a serious limitation.

Finally, many authors have insisted on the convenience of a third subsidiary property, namely, normalization. This can be easily achieved in our case by dividing the M index into an appropriate population entropy. If the racial entropy is chosen, then the H index is obtained. Similarly, if the entropy of the organizational units distribution is chosen, then the H^* index is obtained. However, the cost of either normalization is very high indeed.

1. At a conceptual or intuitive level, neither the H nor the H^* index captures well changes in inter-racial or inter-group exposure.

2. All normalized indices, including the H and the H^* indices, violate the strong decomposability properties **D1** and **D2**. Thus, it should not be surprising that the H index is not responsive to changes in the subgroups segregation levels in the partition of demographic groups into supergroups, while the H^* index is not responsive to those changes in the partition of organizational units into clusters. This could be important in practice, when the segregation in a supergroup (or in a school district) falls, while the segregation in the remaining supergroups (or school districts) remain constant, and overall segregation according to the H (or H^*) index increases.

In conclusion, applied researchers have available three appealing segregation indices based on the entropy notion first advocated by Theil and his co-author Finizza: the M index on one hand, and the H and H^* indices on the other hand. Only the M index has been formally characterized in terms of eight ordinal axioms –a result that allows us to know exactly which value judgments are invoked when using this rather than the remaining entropy-based indices. But beyond this conceptually convenient situation, when decomposability properties are desired in the empirical work nothing substantial is lost by focusing exclusively on the unnormalized M index. When, in addition, invariance properties are also thought to be useful, applied researchers would do better using the M index rather than using either H or H^* . Finally, the significance of the segregation differences and levels can only be studied under alternative hypothesis if the measure is explicitly embedded within a statistical framework.

Researchers with these considerations in mind can exploit the theoretical statistical established in MRC for the M index.

REFERENCES

- Alonso-Villar, O. and C. Del Río (2008), "Local versus Overall Segregation Measures," Documentos de Trabajo, no. 0802, Departamento de Economía Aplicada, Universidade de Vigo.
- Beller, A. (1985), "Changes in the Sex Composition of U.S. Occupations, 1960-1981," *Journal of Human Resources* **20**: 235-250.
- Blau, F. and W. Hendricks (1979), "Occupational Segregation by Sex: Trends and Prospects," *Journal of Human Resources* **12**: 197-210.
- Chakravarty, S.R. and J. Silber (1992), "Employment Segregation Indices: An Axiomatic Characterization" In Eichhorn, W. (ed), *Models and Measurement of Welfare and Inequality*, New York: Springer-Verlag.
- Chakravarty, S.R. and J. Silber (2007), "A Generalized Index of Employment Segregation," *Mathematical Social Sciences*, **53**: 185-195.
- Clotfelter, C. (1979), "Alternative Measures of School Desegregation: A Methodological Note," *Land Economics*, **54**: 373-380.
- Duncan, O. and Duncan, B. (1955), "A Methodological Analysis of Segregation Indices," *American Sociological Review* **20**: 210-217.
- Fisher, M. (2003), "The Relative Importance of Income and Race In Determining Residential Outcomes In U.S. Urban Areas, 1970-2000," *Urban Affairs Review*, **38**: 669-696.
- Fisher, C., G. Stockmayer, J. Stiles, and M. Hout (2004), "Distinguishing the Geographic Levels and Social Dimension of U.S. Metropolitan Segregation, 1960-2000", *Demography*, **41**: 37-59.
- Foster, J.E., and A.F. Shorrocks (1991), "Subgroup Consistent Poverty Indices," *Econometrica* **59**(3):687-709.
- Frankel, D and O. Volij (2008), "Scale Invariant Measures of Segregation," mimeo, August 2008.
- Frankel, D and O. Volij (2009), "Measuring School Segregation," mimeo, March 2009.
- Flückiger, Y. and Silber, J. (1999), *The Measurement of Segregation in the Labor Force*, Heidelberg, Physica-Verlag.
- Fuchs, V. (1975), "A Note on Sex Segregation in Professional Occupations," *Explorations in Economic Research* **2**: 105-111.

- Grusky, D.B. and Charles, M. (1998), "The Past, Present, and Future of Sex Segregation Methodology," *Demography* **35**: 497-504.
- Hutchens, R. M. (2004), "One Measure of Segregation," *International Economic Review*. **45**: 555-578.
- Iceland, J. (2002), "Beyond Black and White," paper presented at the American Sociological Association, Chicago, Illinois, 1-22.
- James, D.R. and Taeuber, K.E. (1985), "Measures of Segregation," in G. Schmid and R. Weitzel (eds.), *Sex Discrimination and Equal Opportunity: The Labor Market and Employment Policy*, London, Gower Publishing Company.
- Jonung, C. (1984), "Patterns of Occupational Segregation by Sex in the Labor Market," in N.B. Tuma (ed.), *Sociological Methodology*, San Francisco, Jossey-Bass.
- Kullback, S. (1959), *Information Theory and Statistics*, Gloucester, Mass.: Dover.
- Massey, D. and N. Denton (1988), "The Dimensions of Residential Segregation," *Social Forces* **67**: 281-315.
- Miller, V. and J. Quigley (1990), "Segregation by Racial and Demographic Group: Evidence From the San Francisco Bay Area," *Urban Studies* **27**: 3-21.
- Mora, R. and Ruiz-Castillo, J. (2003), "Additively Decomposable Segregation Indexes. The Case of Gender Segregation by Occupations in Spain," *The Journal of Economic Inequality*, **1**: 147-179, (2003).
- Mora, R. and Ruiz-Castillo, J. (2005), "Axiomatic Properties of an Entropy Based Index of Segregation," Working Paper 05-62, Economics Series 31, Universidad Carlos III.
- Mora, R. and Ruiz-Castillo, J. (2009a), "The Statistical Properties of the Mutual Information Index of Multigroup Segregation," Working Paper 07-74, Economics Series 43, Universidad Carlos III.
- Mora, R. and Ruiz-Castillo, J. (2009b), "The Invariance Properties of the Mutual Information Index of Multigroup Segregation," in Y. Flückiger, J. Silber and S. Reardon (eds.), *Research on Economic Inequality, Vol. 17, Occupational and Residential Segregation*, Emerald Books, 33-53.
- Mora, R. and Ruiz-Castillo, J. (2009c), "Multigroup and Multilevel School Segregation: the U.S. Case, 1989-2005," mimeo.
- Mora, R. and Ruiz-Castillo, J. (2009d), "Isolating Changes in Net Residential Segregation in the U.S. from the Effect of Demographic Factors, 1989-2005," Working Paper 08-62, Economics Series 29, Universidad Carlos III.
- Reardon, S., J. Yun and T. McNulty, (2000), "The Changing Structure of School Segregation: Measurement and Evidence of Multi-racial Metropolitan Area School Segregation, 1989-1999," *Demography* **37**: 351-364.

- Reardon, S. and G. Firebaugh, (2002), "Measures of Multigroup Segregation," *Sociological Methodology* **32**: 33-67.
- Shorrocks, A. (1980), "The Class of Additively Decomposable Inequality Measures," *Econometrica* **48**: 613-625.
- Shorrocks, A. (1984), "Inequality Decomposition by Population Subgroups," *Econometrica* **52**: 1369-1385.
- Shorrocks, A. (1988), "Aggregation Issues in Inequality Measurement" in W. Eichhorn (ed), *Measurement in Economics: Theory and Applications of Economic Indices*, Heidelberg: Physica.
- Theil, H. (1971), *Principles of Econometrics*, New York: Wiley and Sons.
- Theil, H. (1972), *Statistical Decomposition Analysis*, Amsterdam: North Holland.
- Theil, H. and Finizza, A.J. (1971), "A Note on the Measurement of Racial Integration of Schools by Means of Information Concepts," *Journal of Mathematical Sociology* **1**: 187-194.
- Watts, M. (1992), "How Should Occupational Segregation Be Measured?" *Work, Employment and Society* **6**: 475-487.
- Watts, M. (1998), "Occupational Gender Segregation: Index Measurement and Econometric Modelling," *Demography* **35**: 489-496.
- Zoloth, B.S. (1974), "An Investigation of Alternative Measures of School Segregation," IRP Discussion Paper No. 229-74, University of Wisconsin-Madison.
- Zoloth, B.S. (1976), "Alternative Measures of School Segregation," *Land Economics* **52**: 278-298.

Table 1. School Enrolment, Racial Mix, Entropies, and School Segregation in the U.S., 1989:2005

	No. of students (millions)			Racial Shares (%)			
	1989	2005	Change (%)	1989	2005	Change	
Minorities	8.61	12.24	42.10	34.78	48.05	13.27	
Native American	0.17	0.23	33.77	0.68	0.89	0.20	
Asian	1.03	1.40	36.11	4.15	5.49	1.34	
Black	3.99	4.53	13.70	16.10	17.80	1.70	
Hispanic	3.43	6.08	77.33	13.85	23.87	10.02	
White	16.14	13.23	-18.06	65.22	51.95	-13.27	
Total	24.76	25.47	2.87	100	100	0	
Entropies and Segregation Indexes							
	$E(P_g)$	$E(P_n)$	$\sum_{n=1}^N p_n E(P_{s n})$	$\sum_{g=1}^G p_g E(P_{n g})$	M	H	H^*
1989	101.27	1040.25	57.35	996.32	43.92	43.37	4.22
2005	119.07	1035.72	70.17	986.82	48.90	41.07	4.72
Change	17.80	-4.53	12.82	-9.50	4.98	-2.30	0.50

Notes: Ethnic shares are the percentages of students from every race/ethnic group. The terms Native American, Asian, Black, and White refer to non-Hispanic members of these racial groups. Asian includes Native Hawaiians and Pacific Islanders; Native American includes American Indians and Alaska Natives (Inuit or Aleut). The term Hispanic is an ethnic rather than a racial category since Hispanic persons may belong to any race. Minorities include all categories except White.