

**THE INVARIANCE PROPERTIES OF THE MUTUAL INFORMATION INDEX OF
MULTIGROUP SEGREGATION**



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

Abstract

In the context of educational segregation by ethnic group, it has been argued that rigorous pair wise segregation comparisons over time or across space should be invariant in two situations: when the ethnic composition of the population changes while the distribution of each ethnic group over the schools remains constant (invariance 1), or when the size distribution of schools changes while the ethnic composition of each school remains constant (invariance 2). This paper makes two contributions to this literature. First, it argues by means of the Mutual Information or M index, which is neither invariant 1 or 2, that both properties have strong implications and it provides reasons to defend that the overall segregation index need not satisfy either one. Second, nevertheless, it is shown that in pair wise comparisons this index admits two decompositions into three terms. In the first decomposition, a term is invariant 1 and also satisfies a weak version of invariance 2. In the second decomposition, a term is invariant 2 and also satisfies a weak version of invariance 1. It is shown that these decompositions can be used to reach the analogous ones obtained in Deutsch *et al.* (2006).

Keywords: Gender Segregation Measurement; Axiomatic Properties; Econometric Models.

JEL Classification: J16, J24.

INTRODUCTION

The phenomenon of segregation has preoccupied sociologists and economists during the last 50 years in two different contexts: the occupational segregation by gender in the labor market, and the residential or school segregation by ethnic group.¹ For concreteness, and given the increasing ethnic diversity in many countries, this paper focuses on school segregation in the multigroup case, that is, when there are more than two ethnic groups.

This paper studies two invariance properties of a segregation index, originally discussed in the context of pair wise segregation comparisons over time or across space. Consider for a moment the special but important case of occupational segregation by gender, and assume that segregation in 1950 and 2000 are being compared in a given country. The two questions often asked are the following (see, *inter alia*, Watts, 1998).

First, should the measurement of occupational segregation be independent of the fact that female labor participation has greatly increased over time? Many people would agree that, as long as the male and female distributions over occupations remain constant, the degree of segregation should be the same in the two situations (this is known as composition invariance, or invariance 1). In the multigroup case, the question becomes: should segregation be invariant to changes in the ethnic composition of the population as long as the distribution of each group over the schools remains constant?

Second, should occupational segregation be independent from the fact that agricultural and industrial occupations are much more important in 1950 than in 2000, while services occupations carry much more weight in 2000 than in 1950? Many people would agree that, as long as the gender composition of each occupation remains constant, the degree of segregation should be the same in the two situations (this is known as occupational invariance, or invariance 2). In the multigroup case, the question becomes: should segregation be invariant to changes in the size distribution of schools as long

¹ For a recent treatise in occupational segregation by gender, see Fluckiger and Silber (1999), and for references to residential and educational segregation, see Reardon and Firebaugh (2002).

as the ethnic composition of each school remains constant?

The paper makes two contributions to the literature in the homogeneous case in which the two situations under comparison share a common structure, that is, the same number of demographic groups and organizational units.

In the first place, it is argued that both invariance properties have strong implications, and hypothetical examples are given to defend that the overall segregation index need not be invariant 1 or invariant 2.² This position is substantiated by means of the Mutual Information or M index, based on the entropy concept used in information theory that was first introduced in the segregation literature by Theil and Finizza (1971). Although the M index violates both invariance properties, it is very appealing for the following reasons. First, Frankel and Volij (2008a) have characterized the underlying segregation ordering in terms of eight axioms. Second, the M index has been shown to satisfy a number of other desirable properties discussed in the literature, including two additive separability properties into between- and within-group terms for any partition of either the schools or the ethnic groups (see Mora and Ruiz-Castillo, 2008a, 2008c). Third, Mora and Ruiz-Castillo (2008b) show that a test on segregation in an unrestricted statistical model leads to the M index, and that differences in segregation over time or across space according to the M index can be tested in an appropriate statistical framework.

It may be argued that intertemporal or international comparisons of segregation levels using the M index are “contaminated” by changes in the marginal distributions. However, in its second contribution, this paper presents two decompositions of the M index into three terms in pair wise comparisons over time or across space which isolate the effects of the changes in the marginal distributions.³ In the first decomposition, the first term captures the effects on the index M of changes in the ethnic groups’ marginal distribution. The second term captures the effects of the changes in the

² We are not alone in this position. See, *inter alia*, Coleman *et al.* (1982), Flückiger and Silber (1999, pp 84-85), Watts (1992, 1998), and Frankel and Volij (2005).

³ Using other segregation indices, many authors have defended this strategy before. See, *inter alia*, Blau and Hendricks (1979), Jonung (1984), Beller (1985), and Watts (1992, 1998).

entropy, or concentration, of the schools. The last term captures changes in the M index when there are no changes in the groups' marginal distributions and no changes in the concentration of schools. Thus, the third term is invariant 1 and also satisfies a weak version of invariance 2. In the second decomposition, the first term captures the effects on the index M of changes in the organizational units' marginal distribution. The second term captures the effects of the changes in the concentration of the groups, while the last term captures changes in the M index otherwise. Thus, the last term is invariant 2 and also satisfies a weak version of invariance 1. It is shown that these decompositions can be used to reach the analogous ones obtained in Deutsch *et al.* (2006).

The remaining of this paper is organized in four sections. In the first section, the notation, the notions of segregation most often used in the literature, and a variety of invariance axioms are presented. In the second section the M index is introduced and a number of examples given that illustrate how demanding invariance 1 and invariance 2 really are, and why it may be justified that a segregation index violates them. The following section is devoted to the presentation of the two decompositions of the M index which isolate either an invariance 1 or an invariance 2 term for the purpose of making pair wise segregation comparisons over time or across space. The last section concludes.

NOTATION AND INVARIANCE AXIOMS

Notation

Let X be a *city* with G ethnic groups distributed over N schools, indexed by $g = 1, \dots, G$ and $n = 1, \dots, N$, respectively. Denote by T_g^n the number of students of ethnic group g in school n . The data available in empirical situations can be organized as follows:

$$\begin{array}{c}
\mathbf{X} = \begin{array}{ccc}
\begin{bmatrix} T_1^1 & \cdots & T_1^N \\ \vdots & \ddots & \vdots \\ T_G^1 & \cdots & T_G^N \\ 1 & \cdots & N \end{bmatrix} & \begin{array}{c} \text{Ethnic groups} \\ \begin{bmatrix} 1 \\ \vdots \\ G \end{bmatrix} \end{array} \\
\text{Schools}
\end{array}
\end{array} \quad (1)$$

$$T_g = \sum_{n=1}^N T_g^n : \text{number of students of ethnic group } g$$

$$T^n = \sum_{g=1}^G T_g^n : \text{number of students attending school } n$$

$$T = \sum_{g=1}^G T_g = \sum_{n=1}^N T^n : \text{total number of students in the city.}$$

We are often interested in spatial or intertemporal segregation comparisons between heterogeneous situations, that is, between cities with a different set of demographic groups and/or schools. However, in what follows it will be assumed that we have data about two comparable cities that share the same number of ethnic groups and schools. A discussion of the possible empirical strategies to deal with the heterogeneous case is beyond the scope of this paper.⁴

The information contained in the joint distribution of ethnic groups and schools is usually summarized by means of numerical indices of segregation. Let \mathbf{X} be the set of all cities with G ethnic groups and N schools. A *segregation index* S is a real valued function defined in \mathbf{X} , $S(X)$, providing the extent of school segregation for any city X .

The concept of segregation used in this paper embraces two views. First, the notion advocated by James and Tauber (1985), according to which segregation is seen as the tendency of ethnic groups to have different distributions across schools. Second, the idea of “representativeness” emphasized by Frankel and Volij (2008a), which asks to what extent schools have different ethnic compositions than

the population as a whole.⁵ As can be seen in expression (1), where the rows are ethnic 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 of each other.

Remark 1. If a segregation index S that captures the notion of evenness when applied to the $G \times N$ array X in \mathbf{X} is applied to the $N \times G$ array X' where the role of demographic groups and schools are reversed, so that $T_g^{n'} = T_n^g$ for all n and g , then what will be called the reverse index, S^* , captures equally well the notion of representativeness (and vice versa).

In general, $S(X)$ will be different from $S^*(X')$. Otherwise, that is, when $S(X) = S^*(X')$, we say that the segregation index S is *transpose invariant*.

Relative and Absolute Segregation Views

It is customary to ask: should segregation depend on the population size T ? Most answers in the literature have been in the negative, that is, most segregation indices satisfy the following invariance property.

Size Invariance (James and Tauber, 1985; *Weak Scale Invariance* in Frankel and Volij, 2008a). If X and X' are two cities in \mathbf{X} such that $T_g^{n'} = \lambda T_g^n$ for all n and g with $\lambda > 0$, then $S(X) = S(X')$.

This property resembles the Population Principle in the income inequality literature, according to which replications of the income distribution do not alter income inequality. In our context, size invariance captures a *relative* view of segregation in which all that matters are the joint, conditional and

⁴ Herranz *et al.* (2005) may provide a starting point for the treatment of heterogeneity. Using the M index in the two-group case, these authors propose an algorithm to reduce the number of occupations in two different time periods without significantly reducing the segregation values.

⁵ These two notions are closely connected with the “evenness” and “isolation” dimensions distinguished in Massey and Denton (1988) in the two-group case.

marginal frequency distributions:⁶

$$p_{gn} = \frac{T_g^n}{T} : \text{proportion of students of group } g \text{ and school } n \text{ in the city}$$

$$P_{gn} = \{p_{gn}\}_{g=1, n=1}^{G, N} : \text{joint distribution of ethnic groups and schools}$$

$$p_{g\bullet} = \sum_{n=1}^N p_{gn} : \text{proportion of students of group } g \text{ in the population}$$

$$P_{g\bullet} = \{p_{g\bullet}\}_{g=1}^G : \text{marginal distribution by ethnic groups}$$

$$p_{\bullet n} = \sum_{g=1}^G p_{gn} : \text{proportion of students attending school } n \text{ in the population}$$

$$P_{\bullet n} = \{p_{\bullet n}\}_{n=1}^N : \text{marginal distribution by school size}$$

$$p_{g|n} = \frac{p_{gn}}{p_{\bullet n}} : \text{proportion of students in school } n \text{ who belong to group } g$$

$$P_{g|n} = \{p_{g|n}\}_{g=1}^G : \text{conditional distribution by groups of students at school } n$$

$$p_{n|g} = \frac{p_{gn}}{p_{g\bullet}} : \text{proportion of students in group } g \text{ who attend school } n$$

$$P_{n|g} = \{p_{n|g}\}_{n=1}^N : \text{conditional distribution by schools of students in group } g.$$

The Main Invariance Properties

As pointed out in the Introduction, in the empirical literature on gender segregation it has been noticed that both the overall gender composition of employment, as well as the distribution of the employed population across occupations, typically change over time and/or space. Similar phenomena are present in other segregation contexts. Due to differential fertility or to the vagaries of domestic or

⁶ For a study that focuses on translation invariant segregation indices that represent an absolute view of segregation, see Chakravarty and Silber (1992).

international migration flows, the relative size of certain ethnic groups may drastically change over time and across space. Similarly, the size distribution of schools in two cities or in a given city in two moments in time may be very different indeed. Consequently, it has been forcefully argued that rigorous segregation comparisons in spatial and time-series studies should be made independent of changes in the marginal distributions, be it the overall gender shares of the employed population and its distribution across occupations in studies of occupational segregation by gender, or the population ethnic composition, $P_{g\bullet}$, and the population distribution across schools, $P_{\bullet n}$, in studies of school segregation by ethnic group.⁷

The following two axioms have been proposed to capture these ideas. To motivate the first one, consider a situation in which only the size of one or more ethnic groups vary, so that the marginal distribution $P_{g\bullet}$ changes, but the allocation of ethnic groups across schools, $P_{n|g}$, remains constant. Under these circumstances, it is frequently argued that the segregation level should be unchanged.

Invariance 1 (II): (*Composition Invariance* in James and Taeuber, 1985, and Watts, 1998; *Homogeneity* in Hutchens, 1991; *Scale Independence* in Frankel and Volij, 2008b). If X and X' are two cities in \mathbf{X} such that $T_g^{n'} = \lambda_g T_g^n$ for all n and g with $\lambda_g > 0$ for each g , then $S(X) = S(X')$.

Under **II**, differences in segregation cannot be due solely to differences in citywide ethnic composition. In other words, a group's weight in the segregation index cannot depend only on its size. As a matter of fact, the only relevant magnitudes in the domain of an **II** index are the conditional school distributions by ethnic group, $P_{n|g}$; that is, if S_1 satisfies **II**, then $S_1(X) \neq S_1(X')$ only if the conditional distributions $P_{n|g}$ and $P_{n|g}'$ differ for some group g . This implies that users of segregation indices that satisfy **II** restrict themselves to an evenness notion of segregation.

In the important $G = 2$ case, there is a very close relation between **II** indices and segregation

⁷ See, *inter alia*, James and Taeuber (1985), Charles (1992, 1998), Charles and Grusky (1995, 2004), Grusky and Charles

curves, first suggested by Duncan and Duncan (1955). In the context of occupational segregation by gender, a *segregation curve* represents the cumulative fraction of females (on the ordinate) and the cumulative fraction of males (on the abscissa) with occupations sorted in ascending order according to the ratios $\left(\frac{P_{n|female}}{P_{n|male}}\right)$. A segregation curve is said to dominate another if it lies at no point below and at some point above the other. Just as with Lorenz curves, segregation curves provide an incomplete ranking of distributions of employed people across occupations. Hutchens (1991, 2001) show that a segregation index is consistent with the ranking obtained from segregation curves only if it satisfies **I**. Thus, the failure to satisfy **I** implies that a segregation index is not consistent with the ordering provided by segregation curves. Notice, however, that since segregation curves are only well defined when $G = 2$, this defense of **I** does not carry over to the multigroup case with $G > 2$.⁸

For the next invariance axiom, consider situations in which the school size distribution, $P_{\bullet n}$, changes, while the ethnic mix within each school, $P_{g|n}$, remains constant. It has also been argued that under these conditions the segregation index should not change.

Invariance 2 (I2): (*Occupational Invariance* in Watts 1998, and Blackburn *et al.* 1993, 1995). If X and X' are two cities in \mathbf{X} such that $T_g^{n'} = \lambda^n T_g^n$ for all n and g with $\lambda^n > 0$ for each n , then $S(X) = S(X')$.

Under **I2**, differences in segregation should not be due solely to differences in school sizes; consequently, a school's weight in a segregation index cannot depend only on its size. The only relevant magnitudes in the domain of **I2** indices are the schools' ethnic composition, $P_{g|n}$; that is, if S_2 satisfies **I2**, then $S_2(X) \neq S_2(X')$ only if the conditional distributions $P_{g|n}$ and $P'_{g|n}$ differ for some school n . This implies that users of segregation indices that satisfy **I2** restrict themselves to a representativeness

(1998), and Hutchens (1991, 2001, 2004).

notion of segregation.

Remark 2. In view of Remark 1, the reciprocal segregation index S^* of any **II** index S becomes an **I2** index (and vice versa).

THE MUTUAL INFORMATION INDEX OF SEGREGATION

In information theory, the expression

$$M^n = \sum_{g=1}^G p_{g|n} \log \left(\frac{p_{g|n}}{p_{g\bullet}} \right) \quad (2)$$

is known as the expected information of the message that transforms the set of proportions $P_{g\bullet}$ to the set of proportions $P_{g|n}$. The value of this expected information is zero whenever the two sets of proportions are identical, and it takes larger and larger positive values when the two sets are more different. In other words, the term M^n measures the extent to which the ethnic distribution of individuals in school n , $P_{g|n}$, differs from the overall ethnic distribution, $P_{g\bullet}$. Therefore, M^n can be interpreted as an index of (local) segregation in school n when segregation is interpreted as deviation from representativeness.⁹

Because M^n indicates the existence of local segregation whenever $P_{g|n}$ differs from $P_{g\bullet}$, M^n is obviously not independent from changes in the ethnic distribution of the population, $P_{g\bullet}$. Thus, the M^n index is not **II**. To justify this choice, consider the following two examples with 2 demographic groups (black and white), who attend 3 schools. In example **A**, a student population of 4 people, consisting of 2 black and 2 white students, are distributed in the 3 schools as follows:

⁸ For another notion of segregation curves in the multigroup case, see Alonso-Villar and Del Rio (2008).

⁹ In principle, the logarithm could be computed in any base. In the following examples, natural logarithms will be used.

$$\begin{array}{ccc}
& & \text{Ethnic groups} \\
X^A = & \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ s1 & s2 & s3 \end{bmatrix} & \begin{bmatrix} \text{black} \\ \text{white} \end{bmatrix} \\
& & \text{Schools}
\end{array}$$

Rows represent the distributions of blacks and whites across the three schools. Columns inform about the ethnic mix for each of the three schools. School 1 is entirely black, $p_{black|s1} = 1$, while school 3 is entirely white, $p_{black|s3} = 0$. Since the proportion of black students in the population is $p_{black\bullet} = 0.5$, the local indices $M^{s1} = M^{s3} = \log(2) = 0.69$ indicate large segregation levels in both schools. Since $p_{black|s2} = 0.5$, $M^{s2} = 0$, indicating the absence of segregation in school 2. Consider now example **B** where, relative to example **A**, the size of the white population increases by a factor of 100:

$$\begin{array}{ccc}
& & \text{Ethnic groups} \\
X^B = & \begin{bmatrix} 1 & 1 & 0 \\ 0 & 100 & 100 \\ s1 & s2 & s3 \end{bmatrix} & \begin{bmatrix} \text{black} \\ \text{white} \end{bmatrix} \\
& & \text{Schools}
\end{array}$$

Since now $p_{black\bullet} = 1/101$, the event $p_{black|s1} = 1$, becomes much more anomalous than before when $p_{black\bullet} = 0.5$. Correspondingly, it is reasonable to have $M^{s1} = \log(101) = 4.62$ in school 1, a larger value than before. Similarly, the event $p_{black|s3} = 0$ fits more into the overall pattern than in example **A**, which causes $M^{s3} = 0.99 \times 10^{-2}$ in school 3 to be much lower than before. Finally, since $p_{black|s2} = p_{black\bullet} = 1/101$, indicating that school 2 is characterized as before by the absence of segregation, we have that $M^{s2} = 0$. Thus, in spite of the fact that the distributions of both ethnic groups across schools remain constant, the change in the overall black share from $p_{black\bullet} = 0.5$ in example **A** to $p_{black\bullet} = 1/101$ in example **B** drastically alters the way segregation in certain schools is

measured by the local index M^n . Of course, all indices of (local) segregation for demographic groups satisfying **I1** will indicate that the extent of segregation among blacks and whites remain constant between situations A and B .

A weighted average of the M^n indices of local segregation defined in (2) will constitute an index of segregation for the city as a whole. The selection of the weights is an important issue. One possible option is to give the same weight to each school, thus ensuring that the index satisfies **I2**. However, we agree with the argument provided by England (1981) in the context of occupational segregation by gender. Suppose that schools that are more segregated grow faster over time. An index that reveals this increase in segregation seems to us preferable over one that adjusts the change out because it resulted from an increase in the relative size of highly segregated schools. Thus, the M index of overall segregation is defined by

$$M = \sum_{n=1}^N p_{\bullet n} M^n. \quad (3)$$

That is to say, M is the weighted average of information expectations, with weights equal to the number of people in each school.¹⁰

Consider the following illustration of the role of such weighting scheme. Relative to example A , in example C the size of school 2 is increased by a factor of 100:

$$X^C = \begin{array}{ccc} & \text{Ethnic groups} & \\ & \begin{bmatrix} \textit{black} \\ \textit{white} \end{bmatrix} & \\ \begin{bmatrix} 1 & 100 & 0 \\ 0 & 100 & 1 \\ s1 & s2 & s3 \end{bmatrix} & & \\ \text{Schools} & & \end{array}$$

The black share in the three schools and in the population as a whole coincides with those of example

¹⁰ Notice that, with this aggregation scheme, the citywide segregation in examples A and B above become $M(X^A) = 0.35$ and $M(X^B) = 0.03$, respectively, indicating larger overall segregation in example A than in example B . In contrast, an index which satisfies **I1** would give the same level of segregation in the two examples.

A so that the local indices of segregation M^n remain constant in each school. However, the relative sizes of schools 1 and 3 have now drastically decreased with respect to those in example A. Correspondingly, the overall segregation index is now $M(X^C) = 0.68 \times 10^{-3} < M(X^A)$. Of course, an index which satisfies **I2** would give the same level of segregation in the two examples.

Looking at the three examples together, advocates of segregation indexes that satisfy **I1** but not **I2** would recommend that segregation remains constant in examples A and B, but it should typically decrease in example C. Advocates of segregation indexes that satisfy **I2** but not **I1** would recommend that segregation remains constant in examples A and C, but it should typically decrease in example B. Interestingly enough, advocates of segregation indexes that satisfy both invariance properties, would insist on segregation remaining constant through the three examples.¹¹ As we have seen, the M index favoured in this paper would indicate instead that segregation continuously decreases in the sequence from example A to example C.

To further illustrate the differences between a segregation index that satisfies one or both invariance properties and the M index, consider the following example D (used for other purposes in Frankel and Volij, 2008b). There are two cities Y^1 and Y^2 with the following data matrixes:

$$\begin{array}{ccc}
 & \text{Ethnic groups} & \\
 Y^1 = & \begin{bmatrix} 99 & 0 \\ 0 & 1 \\ s1 & s2 \end{bmatrix} & \begin{bmatrix} \textit{black} \\ \textit{white} \end{bmatrix} \\
 & \text{Schools} & \\
 & & \\
 & & \text{Ethnic groups} \\
 Y^2 = & \begin{bmatrix} 50 & 0 \\ 0 & 50 \\ s1 & s2 \end{bmatrix} & \begin{bmatrix} \textit{black} \\ \textit{white} \end{bmatrix} \\
 & \text{Schools} &
 \end{array}
 ,$$

All conditional distributions in each city are equal to

$$P_{n|g} = P_{g|n} = \begin{Bmatrix} 1 & 0 \\ 0 & 1 \end{Bmatrix}$$

¹¹ For a segregation index that satisfies both **I1** and **I2** in the $G = 2$ case, see Charles (1992, 1998), Charles and Grusky (1995, 2004), and Grusky and Charles (1998).

Therefore, a student's ethnicity is determined by her school in both cities. Thus, according to an index satisfying both invariance properties, segregation is the same in both cities.¹² However, as Frankel and Volij (2008b) indicate, because the population ethnic distribution $P_{g\bullet}$ is $(0.99, 0.01)$ in Y^1 and $(0.50, 0.50)$ in Y^2 , there is initially less uncertainty about a student's ethnicity in Y^1 than in Y^2 . Correspondingly, the M index reveals that the second city is more segregated, $M(Y^2) = 0.69$, than the first one, $M(Y^1) = 0.56 \times 10^{-1}$. As long as *any* aspect of the joint distribution of ethnic group and school changes –including changes solely in the marginal distributions with both conditional distributions remaining constant– the M index reflects a change in the pattern of segregation.

Finally, one may wonder how the notions of evenness and representativeness are handled by the M index. For that purpose, notice that the expression

$$M_g = \sum_{n=1}^N p_{\bullet n} \log \left(\frac{p_{ng}}{p_{\bullet n}} \right) \quad (4)$$

is the expected message that transforms the set of proportions $P_{\bullet n}$ to the set of proportions P_{ng} . The term M_g measures the extent to which the distribution of students in group g across schools differs from the school size distribution for the population as a whole. Therefore, M_g can be interpreted as a (local) index of segregation in ethnic group g when segregation is taken to mean deviation from evenness. It turns out that the weighted average of information expectations defined in (4), with weights equal to the number of people in each group, coincides with the overall index of segregation for the population as a whole defined in (3), that is:

$$M = \sum_{g=1}^G p_{g\bullet} M_g. \quad (5)$$

Equations (3) and (5) show that the M index is transpose invariant (for the two groups case, this is

¹² Of course, indices which are either \mathbf{II} or $\mathbf{I2}$ will also give the same measurement of segregation in the two cities.

shown in Mora and Ruiz-Castillo, 2005). Consequently, we may say that it treats evenness and representativeness in a symmetric fashion. According to the alternative motivation in Frankel and Volij (2008a), if the ethnic group and school of a randomly selected student are thought of as random variables, then the M index equals the mutual information of these variables: the reduction in uncertainty about one variable that occurs when one learns the value of the other (Cover and Thomas, 1991). Since mutual information is a symmetric concept, these authors establish that the M index is transpose invariant. Therefore, it can be interpreted both as the average amount of information that a student's school conveys about her ethnicity, as well as what a student's ethnicity reveals about her school.

A final example E , inspired in another one from Frankel and Volij (2008b), will help to understand the difference between the M index and any index which satisfies \mathbf{II} , according to which a change in the segregation scheme in a relatively small ethnic group may lead to a large change in school segregation. Consider two cities Z^1 and Z^2 with the following data matrixes:

$$\begin{array}{ccc}
 & \text{Ethnic groups} & \\
 Z^1 = & \begin{bmatrix} 89 & 10 \\ 10 & 89 \\ 1 & 1 \\ s1 & s2 \end{bmatrix} & \begin{bmatrix} \textit{black} \\ \textit{white} \\ \textit{mixed} \end{bmatrix} \\
 & \text{Schools} & \\
 & & \\
 & & \text{Ethnic groups} \\
 Z^2 = & \begin{bmatrix} 89 & 10 \\ 10 & 89 \\ 10 & 1 \\ s1 & s2 \end{bmatrix} & \begin{bmatrix} \textit{black} \\ \textit{white} \\ \textit{mixed} \end{bmatrix} \\
 & \text{Schools} &
 \end{array}$$

Since the black and white distributions are very similar in the two situations and the mixed group is very small, the weighted M index does not change much: $M(Z^1) = 0.36$ and $M(Z^2) = 0.37$.¹³ This is not the case for an \mathbf{II} segregation index for which a group's weight cannot depend on its size. For

¹³ Since the ethnic mix in both schools and in the population as a whole do not change much in the two cities Z^1 and Z^2 , it can be conjectured that the extent of segregation according to any $\mathbf{I2}$ index will also be of a similar order of magnitude.

example, take the unweighted Atkinson index A defined by:¹⁴

$$A(X) = 1 - \sum_{n=1}^N \left(\prod_{g=1}^G (p_{n|g})^{1/G} \right).$$

In this case, $A(Z^1) = 0.29$, while $A(Z^2) = 0.37$. Since a group's size does not matter, the small mixed-race group receives the same weight, $1/3$, in the geometric mean as each of the two larger groups. Because this group's segregation pattern considerably changes from Z^1 to Z^2 , the overall segregation changes by as much as 27.6%.

It is reasonable to search for segregation indices that satisfy either **I1** or **I2**. However, we believe that it is equally reasonable to work with segregation indices that violate both properties. In particular, as the above examples illustrate, there are good *a priori* reasons to work with the M index where the influence of any index of (local) segregation at the group or school level in overall segregation depends on the group or the school size.

DECOMPOSITIONS OF PAIRWISE COMPARISONS OF THE MUTUAL INFORMATION INDEX

Differences in the index of segregation between any two situations may result from differences in the marginal distributions, $P_{g\bullet}$ and $P_{\bullet n}$, as well as from differences in the conditional distributions, $P_{n|g}$ and $P_{g|n}$. As indicated in the Introduction, there are reasons to argue that pair wise comparisons of segregation should net out the effect of differences in the marginal distributions. Such comparisons can be accomplished in at least two ways. First, if the index is invariant, then the comparison of the index will also be invariant. Second, as advocated, *inter alia*, by Blau and Hendricks (1979), Jonung (1984), Beller (1985), and Watts (1992, 1998), if the index is not invariant observed differences may be decomposed so that one of the terms in the decomposition reflects changes in segregation which are

¹⁴ Frankel and Volij (2008b) characterize the underlying Atkinson ordering in terms of an ordinal version of the **I1** property and five other ordinal axioms.

due only to changes in one of the conditional distributions. This is the strategy applied for the M index in the rest of this Section.

For the sake of concreteness, assume that there is data on a city X in \mathbf{X} during two periods, $X(t)$, $t = t^1, t^2$. Applying the M index to this data for any t we have:

$$\begin{aligned} M(X(t)) &= \sum_{g=1}^G p_{g\bullet}(t) \sum_{n=1}^N p_{n|g}(t) \log\left(\frac{p_{n|g}(t)}{p_{\bullet n}(t)}\right) \\ &= \sum_{n=1}^N p_{\bullet n}(t) \sum_{g=1}^G p_{g|n}(t) \log\left(\frac{p_{g|n}(t)}{p_{g\bullet}(t)}\right) \end{aligned}$$

As we know, a change in either $p_{\bullet n}(t)$ and/or $p_{g\bullet}(t)$ keeping $p_{g|n}(t)$ and/or $p_{n|g}(t)$ invariant will result in a change in the index, i.e. M does not satisfy **I1** or **I2**. However, as will be seen presently, there are ways to decompose pair wise comparisons using the M index in order to isolate either an invariant 1 or an invariant 2 term.

In the first place, to identify an invariant 1 term in a decomposition of a pair wise comparison, note that the M index can be written as:

$$M(X(t)) = \sum_{g=1}^G \left\{ p_{g\bullet}(t) \sum_{n=1}^N p_{n|g}(t) \log(p_{n|g}(t)) \right\} + T^c(t) \quad (6)$$

where

$$T^c(t) = \sum_{n=1}^N p_{\bullet n}(t) \log\left(\frac{1}{p_{\bullet n}(t)}\right)$$

is Theil's index of concentration for the marginal distribution of students by schools. For the moment,

let $\Pi_g \equiv \{\pi_g\}_{g=1}^G$ be any G array of real numbers such that $0 \leq \pi_g \leq 1$, and $\sum_{g=1}^G \pi_g = 1$.¹⁵ Given equation

(6), the intertemporal change in overall segregation can be decomposed into the following three terms:

¹⁵ Alternative choices for Π_g will be discussed below.

$$M(X(t^2)) - M(X(t^1)) = \Delta N(\Pi_g) + \Delta SC_g + \Delta EG(\Pi_g) \quad (7)$$

where:

$$\begin{aligned} \Delta N(\Pi_g) &= \sum_{g=1}^G \pi_g \sum_{n=1}^N \{p_{n|g}(t^2) \log(p_{n|g}(t^2)) - p_{n|g}(t^1) \log(p_{n|g}(t^1))\}, \\ \Delta SC_g &= T^c(t^2) - T^c(t^1), \\ \Delta EG(\Pi_g) &= \sum_{t=t^1, t^2} (-1)^{1(t=t^1)} \left\{ \sum_{g=1}^G (p_{g\bullet}(t) - \pi_g) \sum_{n=1}^N p_{n|g}(t) \log(p_{n|g}(t)) \right\} \end{aligned}$$

and $1(t=t^1)$ is the indicator function. The three terms in equation (7) can be interpreted as follows.

First, ΔSC_g isolates segregation changes due to intertemporal changes in the concentration of students across schools. Second, $\Delta EG(\Pi_g)$ isolates segregation changes due to differences between the marginal distributions of students by ethnic groups, $P_{g\bullet}(t)$, and Π_g . Third, $\Delta N(\Pi_g)$ gives the difference in segregation that arises from changes in $P_{n|g}(t)$ when both the Theil's index of concentration $T^c(t)$ and the marginal distribution $P_{g\bullet}(t)$ remain constant, and the latter equals Π_g .

Notice that $\Delta N(\Pi_g)$ satisfies **I1** in the sense that it equals zero if $T_g^n(t^2) = \lambda_g T_g^n(t^1)$ for all n and g with $\lambda_g > 0$. In addition, $\Delta N(\Pi_g)$ satisfies a weaker version of **I2** in the following sense. Suppose that the conditional distribution $P_{n|g}(t)$ remains constant but the Theil's index of concentration $T^c(t)$ changes. Then $\Delta N(\Pi_g) = 0$.

In the second place, to identify an invariant 2 term in a decomposition of a pair wise comparison, note that the M index can also be written as:

$$M(X(t)) = \sum_{n=1}^N \left\{ p_{\bullet n}(t) \sum_{g=1}^G p_{g|n}(t) \log(p_{g|n}(t)) \right\} + T_c(t) \quad (8)$$

where

$$T_c(t) = \sum_{g=1}^G p_{g\bullet}(t) \log \left(\frac{1}{p_{g\bullet}(t)} \right)$$

is Theil's index of concentration for the marginal distribution of students by ethnic groups. For the moment, denote by $\Pi^n \equiv \{\pi^n\}_{n=1}^N$ any N array of real numbers such that $0 \leq \pi^n \leq 1$, and $\sum_{g=1}^G \pi^n = 1$.

Given equation (8), the intertemporal change in overall segregation can be decomposed into the following three terms:

$$M(X(t^2)) - M(X(t^1)) = \Delta N(\Pi^n) + \Delta EG^n + \Delta SC(\Pi^n) \quad (9)$$

where:

$$\begin{aligned} \Delta N(\Pi^n) &= \sum_{n=1}^N \pi^n \sum_{g=1}^G \left\{ p_{g|n}(t^2) \log(p_{g|n}(t^2)) - p_{g|n}(t^1) \log(p_{g|n}(t^1)) \right\}, \\ \Delta EG^n &= T_c(t^2) - T_c(t^1), \\ \Delta SC(\Pi^n) &= \sum_{t=t^1, t^2} (-1)^{I(t=t^1)} \left\{ \sum_{n=1}^N (P_{\bullet n}(t) - \pi^n) \sum_{g=1}^G p_{g|n}(t) \log(p_{g|n}(t)) \right\} \end{aligned}$$

The three terms in equation (9) can be interpreted as follows. First, ΔEG^n isolates segregation changes due to changes in the concentration of students across ethnic groups. Second, $\Delta SC(\Pi^n)$ isolates segregation changes due to differences between the marginal distributions of individuals by schools, $P_{\bullet n}(t)$, and Π^n . Third, $\Delta N(\Pi^n)$ gives the difference in segregation that arises from changes in $P_{g|n}(t)$ when the Theil's index of concentration $T_c(t)$ and the marginal distribution of students by ethnic groups $P_{\bullet n}(t)$ remain constant, and the latter equals Π^n . Note that $\Delta N(\Pi^n)$ is **I2** in the sense that it equals zero if $T_g^n(t^2) = \lambda_n T_g^n(t^1)$ for all n and g with $\lambda^n > 0$. In addition, $\Delta N(\Pi^n)$ satisfies a weaker version of **II** in the following sense. If the conditional distribution $P_{g|n}(t)$ remains constant but the Theil's index of concentration $T_c(t)$ changes, then $\Delta N(\Pi^n) = 0$.

Decompositions (7) and (9) can only be computed after specific values for Π_g and Π^n are chosen. Consider the decomposition in equation (7). A simple choice for Π_g would be $P_{g\bullet}(t^1)$, in which case the terms $\Delta N(\Pi_g)$ and $\Delta EG(\Pi_g)$ in the right hand-side of equation (7) become:

$$\begin{aligned}\Delta N(P_g(t^1)) &= \sum_{g=1}^G p_{g\bullet}(t^1) \sum_{n=1}^N \left\{ p_{n|g}(t^2) \log(p_{n|g}(t^2)) - p_{n|g}(t^1) \log(p_{n|g}(t^1)) \right\}, \\ \Delta EG(P_{g\bullet}(t^2)) &= \sum_{g=1}^G (p_{g\bullet}(t^2) - p_{g\bullet}(t^1)) \sum_{n=1}^N p_{n|g}(t^2) \log(p_{n|g}(t^2)).\end{aligned}$$

The interpretation of $\Delta N(P_g(t^1))$ as a change in segregation when marginals are kept constant requires additional conditions. First, $P_{g\bullet}(t^2) = \{p_{g\bullet}(t^1)\}$, i.e. the marginal distribution of students across ethnic groups in the second period is assumed to be the same as the marginal distribution in the first period. Second, the Theil's index of concentration for the marginal distribution of students by schools in both periods must be equal, $T_c(t^2) = T_c(t^1)$. Finally, the conditional distribution by schools of students from group g in the second period must equal the actual one, $P_{n|g}(t^2) = \{p_{n|g}(t^2)\}$. As Karmel and MacLachlan (1988) argue, this approach will not be generally appropriate because marginal and conditional distributions are not independent from each other:

$$\begin{aligned}p_{g\bullet}(t) &= \sum_{n=1}^N p_{g|n}(t) p_{\bullet n}(t), \\ p_{\bullet n}(t) &= \sum_{g=1}^G p_{n|g}(t) p_{g\bullet}(t),\end{aligned}$$

and the additional conditions will almost never be met. Instead, they propose the use of an iterative numerical procedure originally suggested by Deming and Stephan (1940) to identify a joint distribution of schools and ethnic groups for, say, period t^1 , that maintains the original association structure between schools and ethnic groups but *closely* approximates the marginal distributions of, say, period t^2 .

Using this transformation, these authors establish that changes in segregation can be decomposed into four terms that capture variation in segregation due to a variation in the relative weights of the schools; a change in the overall proportions of ethnic groups; a net change in segregation, plus an interaction term.

More recently, Deutsch *et al.* (2006) combine the Karmel and McLachlan approach with the concept of Shapley value to provide an alternative decomposition without any interaction term. These authors accomplish this valuable task applying this approach to the Dissimilarity segregation index first proposed in Duncan and Duncan (1955). Interestingly enough, when this approach is applied to pairwise comparisons of the M index to isolate $I1$ and $I2$ terms, it turns out that the decomposition thus achieved can be obtained from expressions (7) and (9) for certain specifications of the sets Π_g and Π_n , respectively. The remaining of this section is devoted to establish the close connection between the two approaches.

First, consider the computation of an $I1$ term. Following the iterative algorithm described in Deming and Stephan (1940), assume that the matrix $X(t^1)$ converges into a new matrix $X(t^1, t^2)$ such that the conditional distribution $P_{n|g}$ closely resembles the original conditional distribution in $X(t^1)$, while the marginal distributions $P_{g\bullet}$ and $P_{\bullet n}$ are equal to those of period t^2 . Similarly, matrix $X(t^2)$ can also be transformed into a new matrix $X(t^2, t^1)$ such that the conditional distribution $P_{n|g}$ remains approximately constant but the marginal distributions equal those of period t^1 . Define

$$\Delta N_g = (1/2) \left[M \left(X(t^2, t^1) \right) - M \left(X(t^1) \right) \right] + (1/2) \left[M \left(X(t^2) \right) - M \left(X(t^1, t^2) \right) \right]$$

and

$$\Delta M_g = (1/2) \left[M \left(X(t^2) \right) - M \left(X(t^2, t^1) \right) \right] + (1/2) \left[M \left(X(t^1, t^2) \right) - M \left(X(t^1) \right) \right].$$

ΔN_g measures changes in segregation due to changes in the conditional school distribution by ethnic group, $P_{n|g}$, while ΔM_g captures the effect on segregation of changes across periods in both

marginals, $P_{g\bullet}$ and $P_{\bullet n}$. It is easy to observe that

$$M(X(t^2)) - M(X(t^1)) = \Delta N_g + \Delta M_g \quad (10)$$

To establish the relation between ΔN_g and $\Delta N(\Pi_g)$ in equation (7), assume that the algorithm used to obtain $X(t^2, t^1)$ actually delivers a new matrix that exactly replicates the conditional distribution in period t^2 , i.e. $p_{n|g}(t^2, t^1) = p_{n|g}(t^2)$. In this case, we have:

$$M(X(t^2, t^1)) = \sum_{g=1}^G \left\{ p_{g\bullet}(t^1) \sum_{n=1}^N p_{n|g}(t^2) \log(p_{n|g}(t^2)) \right\} + T^c(t^1).$$

If we now choose $\Pi_g = P_{g\bullet}(t^1)$, then it can be seen that

$$\left[M(X(t^2, t^1)) - M(X(t^1)) \right] = \Delta N(P_{g\bullet}(t^1)).$$

Similarly, assuming that the algorithm used to obtain $X(t^1, t^2)$ actually delivers a new matrix that exactly replicates the conditional distribution in period t^1 , i.e. $p_{n|g}(t^2, t^1) = p_{n|g}(t^1)$, and choosing

$\Pi_g = P_{g\bullet}(t^2)$, then it can be seen that

$$\left[M(X(t^2)) - M(X(t^1, t^2)) \right] = \Delta N(P_{g\bullet}(t^2)).$$

Therefore,

$$\Delta N_g = (1/2)\Delta N(P_{g\bullet}(t^1)) + (1/2)\Delta N(P_{g\bullet}(t^2)) \quad (11)$$

On the other hand, it can be seen that

$$\Delta M_g = \Delta SC_g + \Delta EG_g \quad (12)$$

where ΔSC_g is defined in equation (7) and

$$\Delta EG_g = (1/2)\Delta EG(P_{g\bullet}(t^1)) + (1/2)\Delta EG(P_{g\bullet}(t^2)).$$

Equations (11) and (12) imply that a decomposition that isolates an \mathbf{II} term following the approach in

Deutsch *et al.* (2006), can be achieved by applying equation (7) for $\Pi_g = P_{g\bullet}(t^1)$ and $\Pi_g = P_{g\bullet}(t^2)$, and then computing the average for each of the terms in the decomposition.

Second, consider the computation of an **I2** term following the Deutsch *et al.* approach. Using a similar argument as in the previous case, it can be shown that:

$$M(X(t^2)) - M(X(t^1)) = \Delta N^n + \Delta EG^n + \Delta SC^n$$

where ΔEG^n is defined in equation (9) and

$$\begin{aligned} \Delta N^n &= (1/2)\Delta N(P_{\bullet n}(t^1)) + (1/2)\Delta N(P_{\bullet n}(t^2)) \\ \Delta SC^n &= (1/2)\Delta SC(P_{\bullet n}(t^1)) + (1/2)\Delta SC(P_{\bullet n}(t^2)). \end{aligned} \tag{13}$$

Equation (13) shows that a decomposition that isolates an **I2** term can be achieved by evaluating decomposition (9) for $\Pi^n = P_{\bullet n}(t^1)$ and $\Pi^n = P_{\bullet n}(t^2)$, and then computing the average for each term in the decomposition.

CONCLUSIONS

In the segregation literature it has been advocated that rigorous pair wise segregation comparisons over time or across space in the multigroup case should be invariant in two situations: when the ethnic composition of the population changes while the distribution of each ethnic group over the schools remains constant (invariance 1), or when the size distribution of schools changes while the ethnic composition of each school remains constant (invariance 2).

This paper has made two contributions to this literature. First, it has argued that both properties have strong implications, and that there are reasons to defend that the overall segregation index need not satisfy either one. The contrast between invariant segregation indices and the Mutual Information segregation index that violates both properties has been illustrated with a number of examples. Second, it has nevertheless been shown that in pair wise comparisons this index admits two decompositions

into three terms. In the first decomposition, a term is invariant 1 and also satisfies a weak version of invariance 2. In the second decomposition, a term is invariant 2 and also satisfies a weak version of invariance 1. It is shown that these decompositions can be used to reach the analogous ones obtained in Deutsch *et al.* (2006).

ACKNOWLEDGEMENTS

The authors acknowledge financial support from the Spanish DGI, Grants SEJ2006-05710 and SEJ2004-01959. We wish to thank an anonymous reviewer whose insightful comments helped us improving the paper.

REFERENCES

- Alonso-Villar, O. and Del Rio, C. (2008), "Local versus overall segregation measures", Working Paper No. 08/02. 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.
- Blackburn, R.M., Jarman, J., Siltanen, J. (1993), "The Analysis of Occupational Gender Segregation over Time and Place: Considerations of Measurement and Some New Evidence", *Work, Employment and Society* **7**: 335-36.
- Blackburn, R.M., Siltanen, J. and Jarman, J. (1995), "The Measurement of Occupational Gender Segregation: Current Problems and a New Approach", *Journal of the Royal Statistical Society A*, Part 2 **158**: 319-331.
- Blau, F. and W. Hendricks (1979), "Occupational Segregation by Sex: Trends and Prospects", *Journal of Human Resources* **12**: 197-210.
- Chakravarty, S.R. and Silber, J. (1992), "Employment Segregation Indices: An Axiomatic Characterization" In Eichhorn, W. (ed), *Models and Measurement of Welfare and Inequality*, New York: Springer-Verlag.
- Charles, M. (1992), "Cross-National Variation in Occupational Sex Segregation", *American Sociological Review* **57**: 483-502.
- Charles, M. (1998), "Structure, Culture, and Sex Segregation in Europe", *Research in Social Stratification and Mobility* **16**: 89-116.
- Charles, M. and Grusky, D. (1995), "Models for Describing the Underlying Structure of Sex Segregation", *American Journal of Sociology* **100**: 931-971.
- Charles, M. and Grusky, D. (2004), *Occupational Ghettos*, Stanford University Press.
- Coleman, J., T. Hoffer, and S. Kilgore (1982), "Achievement and Segregation in Secondary Schools: A Further Look at Public and Private School Differences", *Sociology of Education* **55**: 162-182.
- Cover, T. and Thomas, J. (1991), *Elements of Information Theory*, Wiley.
- Deming, W. E. and Stephan, F. F. (1940), "On a Least Squares Adjustment of a Sampled Frequency Table when the Expected Marginals are Known", *Annals of Mathematical Statistics* **11**: 427-444.
- Deutsch, J, Flückiger, Y. and Silber, J. (2006), "The Concept of Shapley Decomposition and the Study of Occupational Segregation: Methodological Considerations with an Application to Swiss Data", mimeo.
- Duncan, O. and Duncan, B. (1955), "A Methodological Analysis of Segregation Indices", *American Sociological Review* **20**: 210-217.

- England, P. (1981), "Assessing Trends in Occupational Sex Segregation, 1900-1976", in I. Berg (ed.), *Sociological Perspectives on Labor Markets*, New York, Academic Press.
- Frankel, D and O. Volij (2008a), "Measuring School Segregation", mimeo, September 2008.
- Frankel, D and O. Volij (2008b), "Scale Invariant Measures of Segregation", mimeo, June 2008.
- Flückiger, Y. and Silber, J. (1999), *The Measurement of Segregation in the Labor Force*, Heidelberg, Physica-Verlag.
- Grusky, D.B. and Charles, M. (1998), "The Past, Present, and Future of Sex Segregation Methodology", *Demography* **35**: 497-504.
- Herranz, N., Mora, R., and Ruiz-Castillo, J. (2005), "An Algorithm to Reduce the Occupational Space in Gender Segregation Studies", *Journal of Applied Econometrics*. **20**: 25-37
- Hutchens, R. M. (1991), "Segregation Curves, Lorenz Curves and Inequality in the Distribution of People Across Occupations", *Mathematical Social Sciences* **21**: 31-51.
- Hutchens, R. M. (2001), "Numerical Measures of Segregation: Desirable Properties and Their Implications", *Mathematical Social Sciences* **42**: 13-29.
- Hutchens, R. M. (2004), "One Measure of Segregation", *International Economic Review*. **45**: 555-578.
- 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.
- Karmel, T. and MacLachlan (1988), M., "Occupational Sex Segregation: Increasing or Decreasing?", *Economic Record* **64**:187-195.
- Massey, D. and N. Denton (1988), "The Dimensions of Residential Segregation", *Social Forces* **67**: 281-315.
- 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. (2008a), "A Defense of an Entropy Based Index of Multigroup Segregation", Working Paper 07-76, Economics Series 45, Universidad Carlos III.
- Mora, R. and Ruiz-Castillo, J. (2008b), "The Statistical Properties of an Entropy Based Index of Multigroup Segregation", Working Paper 07-74, Economics Series 43, Universidad Carlos III.

Mora, R. and Ruiz-Castillo, J. (2008c), "Multigroup and Multilevel Residential Segregation: the U.S. Case, 1989-2005", Working Paper 08-61, Economics Series 28, Universidad Carlos III.

Mora, R. and Ruiz-Castillo, J. (2008d), "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. and G. Firebaugh, (2006), "Measures of Multigroup Segregation", *Sociological Methodology* **32**: 33-67.

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.