CEPS INSTEAD

WORKING PAPERS

Measuring Segregation: Basic Concepts and Extensions to Other Domains

Jacques SILBER[1, 2, 3]

Bar-Ilan University, Israel

Working Paper No 2012-20
April 2012

Measuring Segregation: Basic Concepts and Extensions to Other Domains[1]

Jacques Silber
Bar-Ilan University, Israel
Senior Research Fellow, CEPS/INSTEAD, Luxembourg.

April 2012

Abstract

This paper presents the main concepts used in measuring segregation. First it shows that the cardinal as well as the ordinal approach to the measurement of occupational segregation, when only two groups are considered (generally men and women), borrowed many ideas from the income inequality measurement literature. Second, it shows that more recent advances in segregation measurement, that were the consequence of an extension of segregation measures to the case of multi-group segregation and more recently to the analysis of ordinal segregation, could be the basis for additional approaches to the measurement of economic inequality, in particular inequality in life chances, health and happiness, and eventually also to the study of polarization. Finally because the measurement of spatial segregation is a field in itself, this paper only marginally mentions concepts that have been introduced in this no less fascinating domain.

Keywords: health inequality; inequality in happiness; inequality in life chances; multidimensional segregation; occupational segregation; ordinal segregation; polarization; residential segregation.

JEL classification codes: D31; D63;I14; J15; R23.

[1] This paper is based on my Presidential address given at the Fourth Meeting of the Society for the Study of Economic Inequality (ECINEQ) in Catania (Italy), July 18 2011. It was written while I was visiting the Department of Economic Sciences of the University of Geneva, Switzerland, which I thank for its very warm hospitality.

Introduction: On "Lateral Thinking"

In a recent paper entitled "On Lateral Thinking" Atkinson (2011) argued that Economics has benefited not only from borrowing ideas from other disciplines such as physics (e.g. Samuelson's Foundations of Economic Analysis) or psychology (e.g. the growing importance of behavioral economics) but also from applying ideas that appeared in one subfield of Economics to another domain of Economics. As examples of such a cross fertilization Atkinson cites duality theory where cost functions were applied to consumer theory or Harberger's (1962) model of tax incidence that was borrowed from international trade theory. Atkinson in fact cited a sentence from his famous 1970 (Atkinson, 1970) article: "My interest in the question of measuring inequality was originally stimulated by reading an early version of the paper by Rothschild and Stiglitz (1970; 1971)". The same parallelism between uncertainty and inequality had been drawn previously by Serge Kolm in his well-known presentation at the meeting of the International Economic Association in Biarritz, France, (see, Kolm, 1969) which was inspired by previous work of his on uncertainty (Kolm, 1966). Atkinson however stressed also the need for care in drawing parallels.

Though attempting to present the main concepts used in measuring segregation, this paper does not aim at being an exhaustive survey[2]. Its goal is first to show that the cardinal as well as the ordinal approach to the measurement of occupational segregation, when only two groups are considered (generally men and women) borrowed many ideas from the income inequality measurement literature. This paper aims however also at showing that more recent advances in segregation measurement, that were the consequence of an extension of segregation measures to the case of multi-group segregation and more recently to the analysis of ordinal segregation, could be the basis for additional approaches to the measurement of economic inequality, in particular inequality in life chances, health and happiness, and eventually also to the study of

[2] For a survey of the measurement of segregation in the labor force, published more than twelve years ago, see Flückiger, Y. and J. Silber, 1999.

polarization. Finally because the measurement of spatial segregation is a field in itself, this paper will only marginally mention concepts that have been introduced in this no less fascinating domain[3].

The present paper is organized as follows. The next section introduces the reader to the concept of "Segregation curve" and defines the Duncan and Duncan, Gini, "Generalized Gini" and entropy related indices of segregation. It also reviews the desirable properties of a measure of segregation. Section 2 looks then at the measurement of multidimensional segregation and shows first that there are at least four ways of apprehending this issue since an index of multidimensional segregation may be considered as measuring the degree of dependence between the population categories analyzed and, say, their occupations, the disproportionality in group proportions, the extent of diversity in the population or, when the emphasis is on income segregation, the relative importance of between groups income inequality. The last part of Section 2 is devoted to the comparison over time (or across geographic units) in the degree of segregation, the idea being to make a distinction between differences (changes) in the marginal distributions (of, say, the shares of the various occupations and different population subgroups examined) and variations in the "pure" (net of differences in the margins) joint distribution of, say, occupations and population subgroups. Sections 3 and 4 examine the case where an additional dimension is introduced in the analysis. Section 3 applies this idea to the measurement of spatial segregation and shows how it is possible to incorporate space in the measurement of segregation. Whereas traditional indices of segregation would usually compare the racial composition of the different neighborhoods with the average racial composition in the geographic area under study, spatial segregation indices have been proposed that take into account the spatial pattern of segregation, the emphasis being, for example, on the clustering of ethnic groups in separate geographical areas. Section 4 investigates another case where an additional dimension is introduced in the measurement of segregation, that where one assumes, for instance, that occupations may be ranked via, say, some occupational prestige scale. Section 5 finally shows how these recent advances in the measurement of segregation could be the basis for new ways of measuring inequality in life chances, health or

[3] For an excellent recent survey of the measurement of spatial segregation, see Reardon and O'Sullivan, 2004.

happiness, and eventually social polarization. The paper ends with some remarks on the respective advantages and shortcomings of specialization in research in the social sciences.

1. Basic Concepts: Measuring Segregation when there are only two groups

1.1. The Concept of Segregation Curve

This tool was introduced by Duncan and Duncan (1955). It is derived as follows from the traditional Lorenz Curve. Assume, for example, that the distribution of males among various occupations is given by a vector $\overrightarrow{\mathrm{M}} = [(\mathrm{M}_1/\mathrm{M}), \dots, (\mathrm{M}_\mathrm{k}/\mathrm{M}), \dots (\mathrm{M}_\mathrm{K}/\mathrm{M})]$ where M_k is the number of male workers in occupation k, M the total number of male workers in the labor force and K the total number of occupations. Similarly let the distribution of females among the various occupations be represented by the vector $\vec{\mathrm{F}} = [(\mathrm{F}_1/\mathrm{F}), \dots, (\mathrm{F}_\mathrm{k}/\mathrm{F}), \dots (\mathrm{F}_\mathrm{K}/\mathrm{F})]$ where F_k is the number of female workers in occupation k, and F the total number of female workers in the labor force. Let us now rank the occupations k by increasing ratios $(\mathrm{F}_\mathrm{k}/\mathrm{M}_\mathrm{k})$ and plot on the horizontal axis the cumulative values of the shares $(\mathrm{M}_\mathrm{k}/\mathrm{M})$ and on the vertical axis the cumulative values of the shares $(\mathrm{F}_\mathrm{k}/\mathrm{F})$. The curve obtained is what Duncan and Duncan (1955) called a segregation curve[4].

1.2. Indices of Segregation

Two indices of segregation are easily derived from the Segregation Curve: the Duncan and Duncan Index I_D and the Gini Segregation Index I_G .

1.2.1. The Duncan and Duncan Index $\mathbf{I}_\mathbf{D}$:

[4] We could also rank the occupations by increasing ratios (M_k/F_k) and plot the cumulative values of the shares (F_k/F) on the horizontal axis and the cumulative values of the shares (M_k/M) on the vertical axis. The curve obtained would evidently be the same segregation curve as that defined earlier.

It was introduced by Duncan and Duncan (1955) and remains the most popular measure of segregation. It is also called “dissimilarity index” and may be expressed, using the notations introduced previously, as

$$I_D = (1/2) \sum_{k=1}^{K} |(M_k/M) - (F_k/F)| \tag{1}$$

An intuitive interpretation can be given to the Duncan Index: it gives the percentage of the male (female) labor force that has to shift occupations so that the share of the male labor force employed in a given occupation will be equal to that of the females employed in this same occupation. It can be shown that the Duncan index corresponds to the greatest vertical distance between the Segregation Curve and the diagonal. Note that ID may be also written as

$$I_D = (1/2) \sum_{k=1}^{K} (M_k/M) \left| \frac{(F_k/M_k) - (F/M)}{(F/M)} \right| \tag{2}$$

or as

$$I_D = (1/2) \sum_{k=1}^{K} (F_k/F) \left| \frac{(M_k/F_k) - (M/F)}{(M/F)} \right| \tag{3}$$

The Duncan index ID is therefore a weighted relative mean deviation of the gender ratios (F_k/M_k) or (M_k/F_k).

1.2.2. The Gini segregation Index:

This index was originally proposed by Jahn, Schmid and Schrag (1947) and Duncan and Duncan (1955). It can be shown that the value of this index corresponds to twice the area lying between the Segregation Curve and the diagonal. This Gini segregation index can be expressed as

$$I_G = \sum_{h=1}^{K}\sum_{k=1}^{K}(M_h/M)(M_k/M)\left|\frac{(F_h/M_h)-(F_k/M_k)}{(F/M)}\right| \tag{4}$$

or as

$$I_G = (1/2)\sum_{h=1}^{K}\sum_{k=1}^{K}(F_h/F)(F_k/F)\left|\frac{(M_h/F_h)-(M_k/F_k)}{(M/F)}\right| \tag{5}$$

The Gini segregation index is hence also a measure of the inequality of the gender ratios in the various occupations. One can see here again the parallelism between the measurement of income inequality and that of occupational segregation, the gender ratio playing the role that the relative income of individual i, say, $(y_i/\overline{y})$, where y_i is the income of individual i and $\bar{y}$ the average income in the population, plays when computing the Gini index of income inequality.

The Gini Segregation index may be also expressed in a different way (Silber, 1989b) and written as

$$I_G = \vec{M}G\vec{F} \tag{6}$$

where $\vec{M}$ is a row vector of the shares of the male workers in the various occupations and $\vec{F}$ a column vector of the shares of the female workers in the various occupations. Note that the shares in $\vec{M}$ and $\vec{F}$ have to be ranked by decreasing values of the ratios (F_k/M_k). Finally G, called the G-matrix, is expressed as

0 -1 -1……..-1 -1
1 0 -1……. -1 -1

…………………………….

1 1 1………0 -1

1 1 1………1 0

In other words the typical element g_{ij} of the matrix G is equal to 0 if i = j, to -1 if $j > i$ and to 1 if $i > j$.

1.2.3. The Concept of Generalized Gini and the Measure of Segregation[5]:

Using Atkinson's (1970) concept of "equally distributed equivalent level of income" and following earlier work by Blackorby and Donaldson (1978), Donaldson and Weymark (1980) defined a generalized Gini index I_{GG} as

$$I_{GG} = \left[\sum_{i=1}^{n} \{(((i^{\delta}) - (i-1)^{\delta})/(n^{\delta}))y_i\}/\bar{y}\right] \tag{7}$$

where y_i is the income of individual i with $y_1 \gg \cdots \gg y_i \gg \cdots \gg y_n$, n being the number of individuals, and $\delta > 1$, while $\bar{y}$ is the arithmetic mean of the various incomes y_i . It can be shown (see, Donaldson and Weymark, 1980) that when $\delta = 2$, I_{GG} is equal to the Gini index. Note that, following Atkinson (1970), the numerator of the expression within square brackets in (7) represents the "equally distributed equivalent level of income" corresponding to the social welfare function defined by Donaldson and Weymark (1980).

It can also be proven that if there is more than one individual with income x_i expression (7) will be written as

$$I_{GG} = 1 - \left\{\left[\sum_{i=1}^{I}((\left(\sum_{j=1}^{i} n_j\right)^{\delta} - \left(\sum_{j=1}^{i-1} n_j\right)^{\delta})/(n^{\delta}))y_i\right]/\bar{y}\right\} \tag{8}$$

where n_i refers to the number of individuals with income y_i , I is the total number of income categories and $\sum_{j=1}^{I} n_j = n$.

If we now call q_i the relative frequency (n_i/n) and define a coefficient α_i as

$$\alpha_i = \frac{\left\{\left[\left(\sum_{j=1}^{i} n_j\right)^{\delta}\right] - \left[\left(\sum_{j=1}^{i-1} n_j\right)^{\delta}\right]\right\}}{n^{\delta}} = \left[(\sum_{j=1}^{i} q_j)^{\delta} - (\sum_{j=1}^{i-1} q_j)^{\delta}\right] \tag{9}$$

expression (9) may then be written as

[5] This section is based on Deutsch and Silber (2005).

$$I_{GG} = 1 - \{[\sum_{i=1}^{I} \alpha_i y_i]/\bar{y}\} \tag{10}$$

Call now $s_i = (n_i y_i)/(n\overline{y})$ the share of income y_i in total income. It is then easy to derive

$$I_{GG} = 1 - \left[\sum_{i=1}^{I} \alpha_i\,(s_i/q_i)\right] \tag{11}$$

so that (11) shows that computing a "generalized Gini" amounts to "transforming" population shares q_i into income shares s_i via the use of an operator α_i.

In such a "transformation" the population shares could be considered as "a priori shares" and the income shares as "a posteriori" shares.

Expression (11) may be easily extended to the measurement of occupational segregation by gender. The "a priori shares" q_k could, for example, be the shares $m_k = (M_k/M)$ of the males in the various occupations and the "a posteriori shares" s_k the shares $f_k = (F_k/F)$ of the females workers in these occupations. In such a case we would define a coefficient α_k as

$$\alpha_k = \left\{\left[(\sum_{j=1}^{k} m_j)^{\delta}\right] - \left[(\sum_{j=1}^{k-1} m_j)^{\delta}\right]\right\} \tag{12}$$

Note that in (12) the occupations j should be ranked by decreasing values of the ratios (f_j/m_j) in the same way that in Donaldson and Weymark's (1980) original paper, the (relative) incomes were ranked by decreasing values.

The "generalized Gini index of occupational segregation" I_{GGS} will therefore be expressed, combining (10), (11) and (12) as

$$I_{GGS} = 1 - \left[\sum_{k=1}^{K} \alpha_k (f_k/m_k)\right] \tag{13}$$

It can be shown that when δ=2, the index I_{GG} in (13) is identical to the index I_G defined in (4), (5) or (6). Note also that if δ=2 the index I_{GGS} would have been the same, had we assumed that the "a priori shares" are the shares f_k and the "a posteriori shares" the shares m_k.

However when $\delta \neq 2$, the index I_{GGS} will be different if it is assumed that the "a priori shares" are the shares m_k or the shares f_k. In the former case (when the "a priori shares" are the shares m_k), the higher δ, the greater the weight given to the occupations with low ratios (F_k/M_k) (these are hence "male-intensive" occupations). In the latter case however (when the "a priori shares" are the shares f_k), the higher the value of the parameter δ, the greater the weight given to the occupations with low ratios (M_k/F_k) (these are the "female-intensive" occupations).

The choice of "a priori" and "a posteriori" shares (as well as the selection of the parameter δ) introduces therefore normative elements in the computation of the degree of occupational segregation when using the index I_{GGS}.

Finally note that, since in expression (12) the sum $(\sum_{j=1}^{k} m_j)^{\delta}$ <1 whenever $< K$, while $(\sum_{j=1}^{K} m_j)^{\delta} = 1$, we may conclude that when $\delta \to \infty$, the sum $(\sum_{j=1}^{k} m_j)^{\delta} \to 0$ for any $k \neq K$ while it is equal to 1 when $k = K$. As a consequence when $\delta \to \infty$, expression (13) will be written as

$$I_{GSS} = 1 - (f_K/m_K) \tag{14}$$

This implies that when the "a priori" shares are those of the male workers and if $\delta \to \infty$, the generalized Gini-segregation index is equal to the complement to one of the ratio of the percentage of women employed in the occupation with the lowest gender ratio (F_K/M_K) over the percentage of men employed in this same occupation. In other words, when the "a priori shares" are the male shares, the higher the value of the parameter δ, the greater the weight given to the most "male-intensive" occupations.

It is easy to show that in the converse case, when the "a prior shares" are the female shares, the higher the value of the parameter δ, the greater the weight given to the most "female-intensive" occupations.

1.2.4. Measures of Segregation related to the concept of entropy

It is also possible to derive indices of segregation on the basis of concepts introduced in Economics by Theil (1967). More precisely, following Theil and Finizza (1971) and Mora and Ruiz-Castillo (2003), define the expected information of the message that transforms the proportions $[(F/(F+M)),(M/(F+M))]$ into the proportions $[(F_k/(F_k+M_k)),(M_k/(F_k+M_k))]$ as

$$I_k = (F_k/(F_k+M_k)) \log \frac{(F_k/(F_k+M_k))}{(F/M)} + (M_k/(F_k+M_k)) \log \frac{(M_k/(F_k+M_k))}{(M/(F+M))} \qquad (15)$$

It is easy to observe that the value of this expected information is zero whenever the two sets of proportions are identical. This expected information takes larger and larger positive values when the two sets are more different. Note that I_k may be interpreted as an index of local segregation in occupation *k*. A weighted average of these *K* indices of local segregation will then constitute an additive index of segregation. Such an index could, for example, be an index I_E , the weighted average of the information expectations, with weights proportional to the number of people in the occupations, that is, to $((F_k+M_k)/(F+M))$, so that

$$I_E = \sum_{k=1}^{K} \frac{(F_k+M_k)}{(F+M)} I_k \qquad (16)$$

1.3. The desirable properties of a segregation index

Desirable axioms for an index of occupational segregation have been proposed, for example, by James and Taeuber (1985), Siltanen et al. (1993), Kakwani (1994), Hutchens (1991, 2001) and Mora and Ruiz Castillo (2005).

Here is a list of some of the most common axioms that appeared in the literature.

Axiom 1: Size Invariance

As before, let $\vec{F}$ and $\vec{M}$ represent respectively the vectors $[(F_1/F), \dots (F_k/F), \dots, (F_K/F)]$ and $[(M_1/M), \dots, (M_k/M), \dots (M_K/M)]$ and let σ be a segregation index with $\sigma = \sigma(\vec{F}, \vec{M}, F.M)$

Then if, ∀k, $F_k^{'} = \gamma F_k$, $M_k^{'} = \gamma M_k$ so that $F^{'} = \gamma F$ and $M^{'} = \gamma M$, we will have σ '=σ.

Axiom 2: Complete Integration

If $(F_k/F) = (M_k/M)$ ∀k, then $\sigma = 0$.

Axiom 3: Complete Segregation

If $F_k(M_k) > 0$ implies $M_k(F_k) = 0$ ∀k, then $\sigma = 1$.

Axiom 4: Symmetry in Groups

Let $\overrightarrow{F'}$ and $\overrightarrow{M'}$ be two permutations of $\vec{F}$ and $\vec{M}$, respectively. Then $\sigma(\vec{F}, F, \vec{M}, M) = \sigma(\overrightarrow{F'}, F, \overrightarrow{M'}, M)$.

Axiom 5: Symmetry in Types

$\sigma(\vec{F}, F, \vec{M}, M) = \sigma(\vec{M}, M, \vec{F}, F)$

Axiom 6: Principle of Transfers

If there is a small shift of the female (male) labor force from a female- (male-) dominated occupation to a male- (female-) dominated occupation, the segregation index must decrease.

Axiom 7: Increasing Returns to a Movement Between Groups

This notion is analogous to the property of *decreasing returns of inequality in proximity* in Kolm (1999), or the *transfer sensitivity* property in Foster and Shorrocks (1987) in the income inequality literature.

Therefore if there is a small shift of the female (male) labor force from a female- (male-) dominated occupation to a male- (female-) dominated occupation, the segregation index will decrease more, the more male- (female-) dominated the "receiving" occupation is.

Axiom 8: Organizational Equivalence

This axiom was originally proposed by James and Taeuber (1985). It has been called *Insensitivity to Proportional Divisions by* Hutchens (2001).

The idea here is that an index of segregation should be unaffected by the division of an occupation into units with identical segregation patterns. Note that this axiom allows the comparison of economies with a different number of occupations by artificially equalizing those numbers with the help of a suitable division or combination of occupations.

Axiom 9: Additive Decomposability

Assume that the set of K occupations is partitioned into I groups, indexed by $i = 1,\ldots, I$ and denote by G_i the number of occupations in group i, so that $\sum_{i=1}^{I} G_I = K$.

This could, for example, be the case of a one- versus a two-digit classification of the occupations.

We can then make a distinction between an *overall* measure of segregation $\sigma(\vec{F}, F, \vec{M}, M)$, a *within-group* measure of segregation $\sigma_i(\vec{F}_i, F_i, \vec{M}_i, M_i)$ for each i and a *between-group* measure σ_{BET} of segregation computed as if every occupation j had the mean number of males and females of the group i to which it belongs.

The axiom of *Additive Decomposability*[6] then says that if there exists $\eta_i \geq 0$ for all i with $\sum_i \eta_i = 1$,

[6] The Gini index of segregation I_G cannot be decomposed into the sum of a between and within groups segregation. Such a breakdown includes generally a residual which can be interpreted as measuring the degree of overlap between the group-specific distributions of the gender ratios (F_k/M_k) or (M_k/F_k). For an illustration of the decomposition of the Gini segregation index, see, Deutsch et al. (1994).

then $\sigma(\vec{F}, F, \vec{M}, M) = \sum_i \eta_i\ \sigma_i(\vec{F}_i, F_i, \vec{M}_i, M_i) + \sigma_{BET}$

On the basis of at least some of these axioms, several papers have derived axiomatically indices of segregation (e.g., Chakravarty and Silber, 1994; Hutchens, 2001 and 2004, Chakravarty and Silber, 2007; Chakravarty, D'Ambrosio and Silber, 2009; Frankel and Volij, 2011). There are also papers taking an ordinal approach and deriving conditions for the dominance of a segregation curve over another (e.g. Hutchens, 1991).

2. The Multidimensional Analysis of Segregation[7]

2.1. Segregation as a measure of the degree of dependence

Assume we want to measure occupational (or residential) segregation by ethnic groups, when there are more than two ethnic groups.

To derive such a generalization let us first go back to the formulation of the Duncan Index but rather than comparing the shares of males (M_i/M) with the shares of the females (F_i/F), let us compare the shares (M_i/M) (or the shares (F_i/F)) with the shares of the various occupations *i* in the total labor force, that is, with the shares (T_i/T) where $T_i = M_i + F_i$ and $T = M + F$.

Moir and Selby Smith (1979) and Lewis (1982) suggested then using respectively the following segregation measures

$$I_{MSS} = (1/2)\sum_{k=1}^{K}|(F_k/F) - (T_k/T)| \tag{17}$$

$$I_L = (1/2)\sum_{k=1}^{K}|(M_k/M) - (T_k/T)| \tag{18}$$

[7] Subsections 2.2. and 2.3 were mainly inspired by the nice survey of Reardon and Firebaugh (2002).

Karmel and MacLachlan (1988) proposed a kind of mixture of these two formulas and Silber (1992) then proved that their proposition amounted to comparing the “actual” shares (M_k/T) (or (F_k/T)) of individuals of a given gender in a given occupation, in the total labor force, with the “expected” shares $(T_k/T)(M/T)$) or $(T_k/T)(F/T))$.

These are the expected share because if there was complete independence between the lines (occupations) and the columns (the gender) one would have expected the share of a given gender in a given occupation in the total labor force to be equal to the product of the share of this occupation times the share of this gender in the total labor force.

In other words the index proposed by Karmel and MacLachlan (1988) may be expressed as

$$I_{KM} = \sum_{k=1}^{K} |(M_k/T) - (M/T)(T_k/T)| \tag{19}$$

This interpretation shows clearly that measuring segregation amounts to measuring the degree of “dependence” between the occupations and the gender. But such an interpretation opens the way to a more generalized measure of segregation which does not have to be limited to two categories (genders). Assume you have *K* occupations and *J* categories (e.g. ethnic groups). Call T_{kj} the number of workers of category *j* working in occupation *k*. We can then generalize the Karmel and MacLachlan index as

$$I_{KMG} = \sum\nolimits_{k=1}^{K} \sum\nolimits_{j=1}^{J} \left|\left(T_{kj}/T\right) - ((T_{k.}/T)\left(T_{.j}/T\right))\right| \tag{20}$$

where $T_{k.} = \sum_{j=1}^{J} T_{kj}$ and $T_{.j} = \sum_{k=1}^{K} T_{kj}$. We could also call the index I_{KMG} a “Generalized Duncan Index”.

Once we interpret segregation as the comparison of “a priori” shares $(T_{k.}/T)\left(T_{.j}/T\right)$ with “a posteriori” shares $\left(T_{kj}/T\right)$ we are however not limited to using an extension of the Duncan index. We can also apply this idea to the Theil or Gini indices, for example.

One of Theil's two indices (see, Theil, 1967) could thus be expressed as

$$T_{SEG}^{1} = \sum_{k=1}^{K} \sum_{j=1}^{J} \left[(T_{k.}/T)(T_{.j}/T) \right] log \frac{\left[(T_{k.}/T)(T_{.j}/T) \right]}{\left[(T_{kj}/T) \right]} \tag{21}$$

while the second Theil index would be written as

$$T_{SEG}^{2} = \sum_{k=1}^{K} \sum_{j=1}^{J} \left[(T_{kj}/T) \right] log \frac{\left[(T_{kj}/T) \right]}{\left[(T_{k.}/T)(T_{.j}/T) \right]} \tag{22}$$

The multidimensional generalization $I_{G,MULTI}$ of the Gini segregation index (see, Boisso et al., 1994) would be expressed, using (6), as

$$I_{G,MULTI} = \{ \dots \left[(T_{k.}/T)\big(T_{.j}/T\big) \right] \dots \}^{'} G \, \{ \dots \left[(T_{kj}/T) \right] \dots \}$$

where $\{ \dots \left[(T_{k.}/T)\big(T_{.j}/T\big) \right] \dots \}^{'}$ is a row vector of the "a priori" shares $(T_{k.}/T)\big(T_{.j}/T\big)$, $\{ \dots \left[(T_{kj}/T) \right] \dots \}$ is a column vector of the "a posteriori" shares $\big(T_{kj}/T\big)$, G is a $(K \times J)$ by $(K \times J)$ G-matrix, and the $(K \times J)$ elements of the row and column vectors are classified by decreasing ratios $\left[T_{kj}/T \right] / \left[(T_{k.}/T)(T_{.j}/T) \right]$.

We can also define a "Generalized Segregation Curve" as follows. Put the cumulative values of the "a priori" shares $(T_{k.}/T)\big(T_{.j}/T\big)$ on the horizontal axis and the cumulative shares of the "a posteriori" shares $\big(T_{kj}/T\big)$ on the vertical axis, both sets of cumulative shares being ranked by increasing values of the ratios $(T_{kj}/T)/(T_{k.}/T)(T_{.j}/T)$. Note that this "Generalized Segregation curve" is what is often called a "relative concentration curve".

It is then easy to prove that the multidimensional generalization $I_{G,MULTI}$ of the Gini segregation index is equal to twice the area lying between the "Generalized Segregation curve" which has just been defined and the diagonal. Note also that the Karmel and MacLachlan generalized index I_{KMG} will be equal to the maximum distance between the diagonal and this "Generalized Segregation curve".

2.2. Segregation as Disproportionality in Group Proportions

To simplify the notations let us define t_{ij} as $t_{ij} = (T_{ij}/T)$ so that t_{ij} is is a typical element of a matrix whose lines *i* refer, for example, to the various occupations and whose columns *j* define, say, ethnic groups. So t_{ij} represents the share in the total labor force (all occupations and ethnic groups included) of individuals employed in occupation *i* and belonging to ethnic group *j*.

Let also $t_{i.}$ and $t_{.j}$ be respectively equal to $\sum_j t_{ij}$ (so that $t_{i.} = (T_{i.}/T)$) and $\sum_i t_{ij}$ (so that $t_{.j} = (T_{.i}/T)$). Note that we evidently assume that $\sum_i \sum_j t_{ij} = 1$.

Call now φ_{ij} the ratio $(t_{ij}/t_{i.})$ (the share in occupation i of those individuals belonging to ethnic group *j*) and call $\varphi_{.j}$ the share $(t_{.j}/1)$ of ethnic group *j* in the whole labor force. Define also the ratio ρ_{ij} as being equal to $(\varphi_{ij}/\varphi_{.j})$. Clearly if $\varphi_{ij} > 1$ ($\varphi_{ij} < 1$) ethnic group *j* is overrepresented (underrepresented) in occupation *i* since we can also express ρ_{ij} as $\rho_{ij} = (t_{ij}/t_{i.})/(t_{.j}/1)$. The ratio ρ_{ij} reflects therefore the extent to which ethnic group *j* is disproportionately represented in occupation *i*.

Occupational Segregation can then be considered as the mean disproportionality across groups and occupations**.**

To measure disproportionality Reardon and Firebaugh (2002) use[8] a function $f(\rho_{ij})$ such that $f(1) = 0$ and define the weighted average disproportionality D_W as the average value of $f(\rho_{ij})$ across all occupations and ethnic groups, the weights being the shares of the occupations and of the ethnic groups in the total labor force. More precisely we express D_W as

$$D_W = \sum_{i=1}^{I} t_{i.} \sum_{j=1}^{J} t_{.j} f(\rho_{ij}) \tag{23}$$

<u>Example 1</u>: Using an index derived from the relative mean deviation.

[8] Reardon and Firebaugh (2002) were more concerned by residential than occupational segregation.

Assume that $f(\rho_{ij}) = (1/2)|\rho_{ij} - 1|$. Then $D_W = (1/2)\sum_{i=1}^{I} t_{i.} \sum_{j=1}^{J} t_{.j} |\rho_{ij} - 1|$. Given the definition of ρ_{ij} we then derive that

$$D_W = (1/2)\sum_{i=1}^{I}\sum_{j=1}^{J}|t_{ij} - (t_{i.}t_{.j})| \tag{24}$$

and this clearly amounts to saying that D_W measures the degree of dependence between the lines *i* and the columns *j*.

<u>Example 2</u>: An index derived from the Gini index.

Assume we defined the function $f(\rho_{ij})$ as $f(\rho_{ij}) = |\rho_i - \rho_j|$. Then the disproportionality measure D_W will be expressed as

$$D_W = (1/2)\sum_{i=1}^{I} t_{i.}\sum_{h=1}^{J}\sum_{k=1}^{J}(t_{.h}t_{.k})|\rho_{ih} - \rho_{ik}| \tag{25}$$

so that

$$D_W = (1/2\sum_{i=1}^{I}\sum_{h=1}^{J}\sum_{k=1}^{J} t_{i.}\, t_{.h} t_{.k} \left|\frac{t_{ih}}{(t_{i.}\, t_{.h})} - \frac{t_{ik}}{(t_{i.}\, t_{.k})}\right| \tag{26}$$

It is then easy to observe that here again the measure of disproportionality D_W amounts to checking for independence between the lines and the columns.

<u>Example 3</u>: An index derived from the concept of entropy (Theil index).

Assume now that we define the function $f(\rho_{ij})$ as $(\rho_{ij}) = \rho_{ij}\, ln(\rho_{ij})$. The disproportionality measure D_W will then be expressed as

$$D_W = \sum_{I=1}^{I} t_{i.}\sum_{j=1}^{J} t_{.j}\, \rho_{ij}\, ln(\rho_{ij}) \tag{27}$$

and this clearly amounts again to checking for the independence between the lines *i* and the columns *j*.

Example 4: An index linked to the variance.

Assume finally that the function $f(\rho_{ij})$ is defined as $(\rho_{ij} - 1)^2$. The measure of disproportionality D_W will then be written as

$$D_W = \sum_{i=1}^{I} t_{i.} \sum_{j=1}^{J} t_{.j} \, (\rho_{ij} - 1)^2 \tag{28}$$

which again amounts to checking for the independence between the lines *i* and the columns *j*.

2.3. Segregation as a measure related to the concept of diversity

Using the notations defined previously we define the degree of diversity DIV in the whole labor force as

$$DIV = \sum_{j=1}^{J} t_{.j} (1 - t_{.j}) \tag{29}$$

The measure DIV is in fact equal to the probability that two individuals, taken randomly in the labor force, belong to two different ethnic groups.

We can similarly define the degree of diversity in occupation *i* as

$$DIV_{i.} = \sum_{j=1}^{J} \left(t_{ij}/t_{i.}\right)(1 - (t_{ij}/t_{i.})) \tag{30}$$

The average degree $\overline{DIV_{i.}}$ of diversity across all occupations may then be expressed as

$$\overline{DIV_{i.}} = \sum_{i=1}^{I} t_{i.} \sum_{j=1}^{J} \tfrac{t_{ij}}{t_{i.}} (1 - \frac{t_{ij}}{t_{i.}}) \tag{31}$$

so that

$$1-\left(\frac{\overline{DIV_{i.}}}{DIV}\right)=1-\frac{\sum_{i=1}^{I} t_{i.}\sum_{j=1}^{J}\frac{t_{ij}}{t_{i.}}(1-\frac{t_{ij}}{t_{i.}})}{\sum_{j=1}^{J} t_{.j}(1-t_{.j})} \tag{32}$$

This is known as the Goodman and Kruskal (1954) τ_B and is equal to 1 minus the probability that two individuals from the same occupation belong to different ethnic groups over the probability that any two individuals in the labor force belong to different ethnic groups. The previous measure may therefore be interpreted as the average difference between overall and within occupations diversity, divided by the overall diversity.

Such a residual diversity can be attributed only to between occupations differences in ethnic group proportions. It may therefore be interpreted as a measure of the proportion of total diversity attributable to between occupations differences. As expected this residual diversity will be equal to zero if each occupation has the same ethnic group proportions as the whole labor force and to 1 when each occupation has no diversity whatsoever.

2.4. Segregation as the ratio of inequality between groups of individuals to inequality among individuals[9]

This emphasis on the relative importance of between occupations differences appears also in a recent article by Jargowsky and Kim (2009). These authors start their analysis from Shannon's (1948) famous article on "A Mathematical Theory of Communication" stating that "the fundamental problem of communication is that of reproducing at one point…a message selected at another point". The reason for the existence of such a problem is evidently the possibility of a noisy transmission process over a medium that

[9] This section was inspired by ideas that appear in Jargowsky and Kim (2009).

has a limited capacity (e.g. a telephone cable). Shannon characterized the *information value* of a source as a function of the number of potential messages that the source could produce. If all individuals are identical, then choosing an individual at random will produce the same message every time. In Shannon's terms the information value of the message is zero because there is no uncertainty about the message. But if there is a lot of variation among individuals, many possible messages may have been sent so that the information value of the source of the message is high.

Jargowsky and Kim (2009) take as illustration the case of "income segregation". In other words they wanted to know to which extent there are poor and rich neighborhoods.

Call y_i the income of individual i, $\bar{y}$ the average income in the population and *n* the number of individuals.

The Gini index of income inequality (see, Kendall and Stuart, 1961) in the population will then be expressed as

$$G_{TOTAL} = (1/2)(1/\bar{y})(1/n^2)\sum_{h=1}^{n}\sum_{k=1}^{n}|y_h - y_k| \quad (33)$$

Call now $\overline{y_j}$ and $\overline{y_l}$ the mean incomes in areas *j* and *l*, n_j and n_l the number of individuals in areas *j* and *l* and *J* the total number of areas.

If every individual in a given area is assumed to receive the average income prevailing in the area, the Gini index for the whole population would then be expressed as

$$G_{BETWEEN} = (1/2)(1/\bar{y})(1/n^2)\sum_{j=1}^{J}\sum_{l=1}^{J}\left|\overline{y_j} - \overline{y_l}\right| \quad (34)$$

Jargowsky and Kim (2009) suggest then to measure income segregation via the ratio of $G_{BETWEEN}$ over G_{TOTAL}. In other words segregation is defined as the retention of information about inequality when comparing the group- and the individual-level information.

2.5. Comparing Degrees of Segregation

Let us now assume that we want to compare occupational segregation by ethnic groups at two different time periods. In other words we want to compare two matrices like those whose typical element was defined previously as t_{ij}. Karmel and MacLachlan (1988) and later on Watts (1998) have argued that it is not possible to directly compare these two matrices by computing a segregation index in each of the two cases. The reason is that the occupational structure by ethnic groups in a given country at two different points in times may have varied because

- the occupational structure changed

- the "ethnic structure" changed

- the "pure" degree of independence between occupations and ethnic groups (the essence of segregation) varied over time.

This is why Karmel and MacLachlan (1988) as well as Watts (1998) have argued that it is essential to make a difference between variations in the margins and what they called a change in the "internal structure" of the matrix. To solve this problem they suggested borrowing a technique originally proposed by Deming and Stephan (1940). A simple illustration of the technique proposed by Deming and Stephan (which is not the only technique available) is presented here. Let us assume we start with an "original" matrix t_{ij} and a "final" matrix v_{ij}. Both matrices are given below.

Original matrix t_{ij}.

0.05	0.35
0.20	0.40

Final matrix, say v_{ij}.

0.10	0.25
0.40	0.25

Stage 1: Multiply all the elements of the matrix t_{ij} by the ratios $(v_{i.}/t_{i.})$ and call x_{ij} the new matrix which is then

Matrix x_{ij} .

0.04375	0.30625
0.21666	0.40000

Stage 2: Multiply all the elements of the matrix x_{ij} by the ratios $(v_{.j}/x_{.j})$ and call y_{ij} the matrix just derived which is then

Matrix y_{ij} .

0.08509	0.21681
0.42139	0.28319

Stage 3: Multiply all the elements of the matrix y_{ij} by the ratios $(v_{i.}/y_{i.})$ and call z_{ij} the matrix just derived which is then

Matrix z_{ij} .

0.09865	0.25135
0.38875	0.26125

Stage 4: Multiply all the elements of the matrix z_{ij} by the ratios $(v_{.j}/z_{.j})$ and call w_{ij} the matrix just derived which is then

Matrix w_{ij} .

.10120	.24517
.39880	.25483

Note that already at this stage the horizontal margins of the matrix w_{ij} are respectively 0.34637 and 0.65363 whereas the horizontal margins of the matrix v_{ij} are 0.35 and 0.65.

The vertical margins of the matrix w_{ij} are, as expected at this stage, identical to those of the matrix v_{ij}, that is, 0.5 and 0.5. Assuming, for simplicity, that the matrix w_{ij} corresponds to the final stage of the iteration, we will say that this matrix w_{ij} has the "internal structure" of the original matrix t_{ij} but the margins of the matrix v_{ij} .

The Deming and Stephan (1940) technique allows us therefore making a distinction between variations over time in the "internal structure" and in the margins of the occupations by ethnic groups matrix. The technique that was just presented assumed that one started from an "original" matrix t_{ij} and ended with a "final" matrix v_{ij} . It would however have been also possible to start with an "original" matrix v_{ij} and end with a "final" matrix t_{ij} . This is a standard "index number problem" which can be solved using what is called a "Shapley decomposition" (see, Chantreuil and Trannoy, 2012, Shorrocks, 2012, or Sastre and Trannoy, 2002).

An Empirical Illustration: Changes in Occupational Segregation in Switzerland between 1970 and 2000

As an illustration of the application of the Deming and Stephan (1940) technique, we report here results that appeared in Deutsch et al. (2009). Using the Swiss Censuses for the years 1970 and 2000, the authors analyzed the changes over time in occupational segregation by gender, nationality and age. Table 1 gives the results of the decomposition they obtained.

This illustration shows clearly that there are cases where "gross segregation" seems to have increased while "net segregation" in fact decreased. There are also cases where the impacts of changes in the margins are in opposite directions (impact of changes in the occupational structure versus impact of changes in the relative shares of the genders). Deutsch et al. (2009) used here the Generalized Duncan Index but they could have used in a similar way the Generalized Gini Segregation Index or an entropy related index.

3. Measuring Spatial Segregation

Spatial Segregation is a good illustration of the case where an additional dimension has to be introduced to measure multidimensional segregation. Assume data are available on the distribution of ethnic groups across different geographical units (e.g. Census tracts) which are part of a bigger geographical area (e.g. a metropolitan area). The measures of multidimensional segregation introduced in Section 2 would not correctly measure the degree of spatial segregation because these indices would amount to comparing the ethnic composition of the different geographical units with the average ethnic composition in the bigger geographical area under study. A good measure of spatial segregation should however take into account the "geographical component" of the distribution of the ethnic groups across the geographical units. One might want to know, for example, whether the ethnic groups are evenly dispersed across the various geographical units or on the contrary clustered in a few specific areas. Another issue of interest may concern the location of the various ethnic groups with respect to the "center" of the bigger geographical area: does some ethnic group, for example, live in the suburbs of the metropolitan area and some other in the center of the city?

Massey and Denton (1988) considered thus five aspects of residential segregation: evenness, exposure, concentration, centralization and clustering. For them "evenness refers to the differential distribution of two social groups among areal units in a city" while "residential exposure refers to the degree of potential contact, or the possibility of interaction, between minority and majority group members within geographic areas of a city". "Concentration refers to the amount of physical space occupied by a minority group in the urban environment…Centralization is the degree to which a group is spatially located near the center of an urban area…". Finally the degree of spatial clustering exhibited by a minority group "is the extent to which areal units inhabited by minority groups adjoin one another, or cluster, in space" (Massey and Denton, 1988). These authors then surveyed twenty potential measures of segregation and checked which of these five aspects each segregation index was measuring. Massey and Denton (1988) in fact argued that evenness and exposure are "aspatial dimensions" while concentration, centralization and clustering are "spatial dimensions" of residential

Table 1: Decomposition of the change in Switzerland between 1970 and 2000 in the Generalized Duncan Index. (Occupational Segregation by gender, nationality and age)[10]

Criterion of Comparison of Populations	Value of the Index in 1970	Value of the Index in 2000	Change observed between 1970 and 2000	Component of the change due to a variation in the "internal structure" of the matrix	Component of the change due to a variation in the margins of the matrix	Component due to a variation in the occupational structure	Component sue to a variation in the shares of the subpopulations distinguished
Gender	0.4787	0.4875	0.0088	-0.0216	0.0304	-0.0237	0.0542
Nationality (Swiss versus Foreigners)	0.2449	0.1446	-0.1003	-0.0524	-0.0479	-0.0224	-0.0255
Age (below and above age 50)	0.1325	0.0651	-0.0673	-0.0691	0.0017	0.0036	-0.0019

[10] Deutsch et al. (2009) give for each number in the table confidence intervals based on the bootstrap approach.

segregation. Reardon and O'Sullivan (2004) criticized these distinctions and suggested an alternative classification. They recommended making a distinction between only two aspects of residential segregation: spatial exposure (as opposed to spatial isolation) and spatial evenness (the contrary of spatial clustering). "Spatial exposure refers to the extent that members of one group encounter members of another group…in their local spatial environments. Spatial evenness… refers to the extent to which groups are similarly distributed in the residential space" (Reardon and O'Sullivan, 2004). These authors reviewed then existing measures of spatial segregation and checked which (desirable) properties they had. They then proposed new indices of spatial segregation.

Since the present paper does not aim at offering also a survey of indices of spatial segregation, only one index of spatial segregation will be presented. It has been proposed by Dawkins (2004) and is a simple extension of the Gini index of Segregation I_G defined in (6).

Assume there are only two ethnic groups, *B* and *W*, and that data are available on the number B_k and W_k of individual belonging to groups *B* and *W* in each area *k* (e.g. Census tract). Using (6) the Gini Segregation index may then be expressed as

$$I_{G,SEG} = \overrightarrow{w}' G \vec{b} \tag{35}$$

where $\overrightarrow{w}'$ is a row vector of the shares (W_k/W) of ethnic group *W* in the various areas *k* and $\vec{b}$ is a column vector of the corresponding shares (B_k/B) for ethnic group *B*, both set of shares being ranked by decreasing ratios (B_k/W_k). The operator *G* in (35) is naturally the *G*-matrix that was previously defined. Assume, for example, a city with 9 neighborhoods and in each neighborhood there are either blacks or whites. The following graphs (Figures 3 and 4) draw two such cases with 5 black neighborhoods and four white neighborhoods.

Figure 1: Case 1

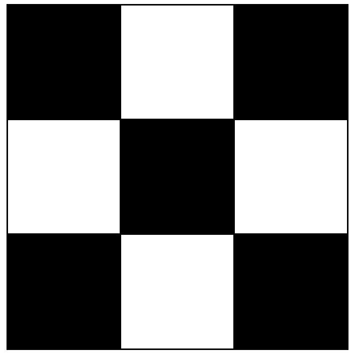
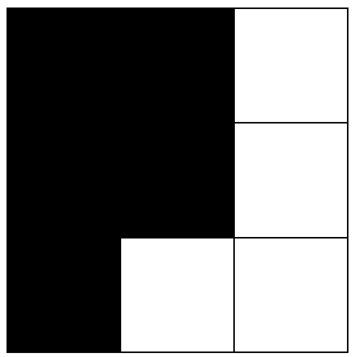


Figure 2: Case 2



Let us compute the Gini segregation index I_G for Case 1. Table 2 gives the "blacks/whites" ratios in each neighborhood in Case 1. One possible ordering[11] of the "black/white" ratios will then be given by Table 3.

Table 2: Black-White ratios in each neighborhood (Case 1).

∞	**0**	∞
0	∞	**0**
∞	**0**	∞

[11] Since there are five areas where the ratio is equal to ∞ we can order these five areas as we wish. Similarly there are four cases where the ratio is equal to 0 and here again we can order these four areas as we want.

Table 3: Ordering of the neighborhoods according to the black-white ratios (Case 1).

2	**7**	**1**
8	**5**	**6**

The traditional Gini segregation index will then be expressed as $\overrightarrow{w}' G\vec{b}$ where the row vectors $\overrightarrow{w}'$ and $\overrightarrow{b}'$ are $\overrightarrow{w}'$ = (0 0 0 0 0 .25 .25 .25 .25)' and $\overrightarrow{b}'$ = (.2 .2 .2 .2 0 0 0 0 0 0). It is easy to derive that in this case $\overrightarrow{w}' G\vec{b} =$1 so that residential segregation is maximal. Note that in Case 2 a similar computation would show that the Gini segregation index I_G is also equal to 1. This is not surprising since the Gini index I_G does not take into account the respective location of the two ethnic groups. It thus ignores the degree of clustering of these ethnic groups.

Assume now that we order the neighborhoods according to the order of the closest neighborhood. If, for example, the closest neighborhood to a black neighborhood is a white neighborhood, we will assume that the black neighborhood is not really segregated. A similar assumption will evidently be made for white neighborhoods.

Let us go back to Case 1 and assume, as a simple illustration, that the closest neighborhood is the one below and for the neighborhoods in the bottom line, we assume it is the one above. For Case 1 the relevant "black/white" ratios are then given in Table 4.

Table 4: New Black-White ratios in each neighborhood (Case 1).

0	∞	**0**
∞	**0**	∞
0	∞	**0**

so that the ordering of the neighborhoods is now the one given in Table 5.

Table 5: New Ordering of the neighborhoods according to the black-white ratios (Case 1).

6	**1**	**5**
2	**9**	**4**
7	**3**	**8**

In such a case we do not compute any more a Gini Segregation index but a "Pseudo-Gini" Segregation Index [12] which is expressed as $\breve{w}'G\breve{b}$ with $\breve{w}' = (.25\ .25\ .25\ .25\ 0\ 0\ 0\ 0\ 0)$ and $\breve{b}'$=(0 0 0 0 .2 .2 .2 .2 .2) so that the "Pseudo-Gini" Segregation index is equal to -1. Since "Pseudo-Ginis" vary between -1 and +1, we can conclude that the level of segregation obtained is now the lowest possible (no segregation). This should be clear because in Case 1, the neighborhood closest to one's neighborhood (the one below) is one with the opposite race, and hence integration is now assumed to be maximal.

The computation of the "Pseudo-Gini" for Case 2 will however give different results. Here also we will assume that the closest neighborhood is the one below and for

[12] For an exact definition of the concept of Pseudo-Gini, see, Silber, 1989a.

the neighborhoods in the bottom line, we assume it is the one above. The "black/white" ratios in Case 2 are given in Table 6 and the ordering of the neighborhoods appears in Table 7.

Table 6: New Black-White ratios in each neighborhood (Case 2).

∞	∞	**0**
∞	**0**	**0**
∞	∞	**0**

Table 7: New Ordering of the neighborhoods according to the black-white ratios (Case 2).

1	**2**	**7**
3	**6**	**8**
4	**5**	**9**

The "Pseudo-Gini" Segregation Index will as before be expressed as $\breve{w}'G\breve{b}$ but now $\breve{w}' = (0\ 0\ 0\ 0\ 0\ .25\ .25\ .25\ .25)$ while b'=(.2 .2 .2 .2 .2 0 0 0 0) so that now the Pseudo-Gini Segregation index is equal to 1. We thus find that, as expected, there is much less integration than in Case 1. In fact there is maximal segregation in Case 2. Naturally segregation can be computed in a similar way on the basis of other criteria. We could, for example, rank the neighborhoods on the basis of their distance from the center of the city.

4. Measures of Ordinal Segregation

Neither the binary nor the multiple group measures of segregation are appropriate when either the groups or the units have inherent orderings. This is another case where an additional dimension has to be introduced to correctly measure segregation. Assume for example, that the goal is to measure occupational segregation by gender, taking into account the fact that there is an ordering of the occupations (e.g. there is a prestige scale for the various occupations). Another illustration would be the case where one wants to analyze residential segregation by educational level when there is an ordering of the educational levels.

Let us take the case of occupational segregation by gender and imagine two possibilities (see, scenario A and scenario B in Table 8).

Table 8: Ordinal Segregation

Scenario A

	Occupation 1	**Occupation 2**	**Occupation 3**	**Occupation 4**	**Total**
Males	**10**	**20**	**30**	**40**	**100**
Females	**40**	**30**	**20**	**10**	**100**
Total	**50**	**50**	**50**	**50**	**200**

Scenario B

	Occupation 1	**Occupation 2**	**Occupation 3**	**Occupation 4**	**Total**
Males	**20**	**10**	**40**	**30**	**100**
Females	**30**	**40**	**10**	**20**	**100**
Total	**50**	**50**	**50**	**50**	**200**

In scenario A males are concentrated in prestigious occupations (3 and 4) and females in non prestigious occupations (1 and 2). In scenario B the data of occupations 3 and 4 were swapped as well as those of occupations 1 and 2. A traditional segregation index (e.g. the Duncan index) would give the same value in scenario A and scenario B, while intuition tells us that there should be more segregation in scenario A. Reardon (2009) proposed indices of segregation that would give different answers for these two scenarios:

Assume that the variable *j* denotes ordered categories (e.g. occupations), with *j*=1 to *J*, while *i* refers to the unordered categories (here gender, or ethnic groups) with *i* = 1 to *I*.

Let $n, n_{i.}, n_{.j}$ and n_{ij} refer respectively to the total population, that in group i, that in ordered category j and that in the cell (*i*, *j*).

Within each group (unordered category) *i* call π_{ij} the cumulative proportion of the population in *i* in ordered category *j*, that is,

$$\pi_{ij} = \sum_{h=1}^{j} (n_{ih}/n_{i.}) \tag{36}$$

In fact it is enough to characterize such a distribution by the [*J*-1]-tuple $\pi_i = (\pi_{i1}, \ldots, \pi_{ij}, \ldots, \pi_{i,J-1})$ since, *J* being the total number of categories, $\pi_{i,J}$ is always equal to 1.

Clearly segregation will be maximal if within each group (unordered category) *i* all individuals occupy a single ordered category[13], in which case π_{ij} is always either equal to 0 or to 1. Conversely segregation will be minimal if within each unordered category *i* the distribution of the individuals is equal to that of the population (in which case $\pi_{ij} = \pi_{.j}$ $\forall i$ and *j*).

Following the work of Reardon and Firebaugh (2002) on multi-group (unordered) segregation, Reardon proposes then to measure ordinal segregation as

$$\Psi = \sum_{i=1}^{I} (n_{i.}/n) \frac{(\xi - \xi_i)}{\xi} \tag{37}$$

where ξ_i and ξ refer respectively to a measure of variation (dispersion) in unit *i* and in the total population.

[13] Naturally the ordered category should be different for each unordered category.

How shall one measure variation? Variation will be assumed to be maximal when half of the population is located in the first column and half in the last column, that is, half the population belongs to the least prestigious occupation and half to the most prestigious one. Obviously variation will be minimal when all the observations are in the same column, that is, when all the individuals belong to the same occupation j with j =1, 2…, or J.

In other words there will be maximal variation when the [K-1]-tuple π is written as

$\pi_{MAX} = (0.5\ 0.5\ 0.5 \dots .0.5)$ and there will be minimal variation when π is written as

$\pi_{MIN} = (0\ 0\ 0.0 \dots ,1\ 1\ 1 \dots 1)$

The idea is then to measure variation as an inverse function of the distance from π to π_{MAX}. In other words we will write

$$\xi = (1 / (J-1)) \sum_{j=1}^{J-1} f(\pi_j) \tag{38}$$

where $f(\pi)$ is a continuous function on the interval [0,1] such that $f(\pi)$ is increasing for $\pi \in (0,0.5)$ and decreasing for $\pi \in (0.5,1)$. Note that $f(\pi)$ is maximal at π=(1/2), that is, f(1/2)=1 and minimal at π=0 and π=1, that is, f(0)=f(1)=0.

Reardon (2009) suggested then four possible functional forms for f:

$$f_1(\pi) = -[(\pi log_2 \pi) + ((1-\pi) log_2 (1-\pi)] \tag{39}$$

$$f_2(\pi) = 4\pi(1-\pi) \tag{40}$$

$$f_3(\pi) = 2\sqrt[2]{\pi(1-\pi)} \tag{41}$$

$$f_4(\pi) = 1 - |2\pi - 1| \tag{42}$$

Substituting these four functions in ξ yields four potential measures of ordinal variation:

$$\Omega_h = \sum_{i=1}^{I}(n_{i.}/n)(\frac{\xi_h - \xi_{hi}}{\xi_h} \tag{43}$$

with h =1 to 4, each h referring to one of the four functional forms which have just been described.

Note that Ω_1 is an ordinal generalization of the information theory index H and was called by Reardon the *ordinal information theory index.*

Similarly Ω_2 is an ordinal generalization of the diversity index *DIV* and was called by Reardon the *ordinal variation ratio index.*

5. Extensions to Other Domains: Segregation and Inequality in Life Chances

This section aims at showing that the concepts used in measuring multidimensional segregation may be applied to several other domains.

5.1. Inequality in Life Chances: The Case of Cardinal Variables

In a recent paper entitled "Inequality in Life Chances and the Measurement of Social Immobility" Silber and Spadaro (2011), in a book in honor of Serge Kolm, used as database a matrix whose lines correspond to the social category (e.g. occupation or educational level) of the parents and the columns to the income distribution of the children.

They then borrowed the concept of Generalized Segregation Curve which was presented previously, to define what they called Social Immobility Curves. More precisely they plotted on the horizontal axis the cumulative values of the "a priori" probability for an individual to belong to social origin i and income group j, this "a priori" probability being equal to the product of the probability to belong to social origin *i* and of the probability to belong to income group *j*. On the vertical axis they plotted cumulative values of the "a posteriori" probabilities, that is, of the actual probability to belong to social origin *i* and to income group *j*. On both axes the individuals were classified by increasing values of the ratio of the "a posteriori" over the "a priori" probabilities.

The empirical illustration given by Silber and Spadaro (2911) was in part based on a survey of 2000 individuals conducted in France by Thomas Piketty in the year 1998. To measure the social origin of the parents the authors used information on the profession of either the father or the mother. Eight professions were distinguished (farmer, businessman, store owner or "artisan", manager or independent professional, technician or middle rank manager, employee, blue collar worker, including salaried persons working in agriculture, not working outside the household and retired). The social status of the children's generation was measured via their monthly income classified in eight income categories (see, Silber and Spadaro, 2011, for more details).

The authors first computed what they called a "Gini index of Social Immobility" defined as being equal to twice the area lying between the Generalized Social Immobility Curve and the diagonal. They then applied the algorithm proposed by Deming and Stephan (1940) which allowed them to break down into three components the difference between, for example, the degree of social immobility from fathers to sons and that from mothers to daughters. In such an illustration the first component would reflect the impact of differences between the occupational distributions of the fathers and mothers. The second component would correspond to the impact of differences between the income distribution of sons and daughters. Finally the third element of the breakdown would actually measure differences between the two cases examined in the "pure degree" of social immobility, that is, differences that remain even after differences between the margins of the two cases under study are neutralized. Not surprisingly the authors often found that differences in "gross social immobility" (before "neutralizing" differences in the margins) were often of opposite sign to differences in "net social immobility" (in "pure" social immobility).

Note that in this illustration the fact that the income groups are ordered was not taken into account. The goal of the authors was simply to estimate the degree of independence between the social origin of the parents and the income group to which they belonged. It is however possible to extend such an analysis by taking into account the ranking of the income groups, as will now be shown.

5.2 Inequality in Life Chances: The Case of Ordinal Variables

In a recent paper Silber and Yalonetzky (2011) suggested that the four indices that Reardon (2009) had proposed to measure ordinal segregation could also be used to measure inequality in life chances when one deals with ordinal variables. Although the paper of Silber and Yalonetzky (2011) does not include any empirical illustration, one could, for example, apply their approach to the case of a matrix whose lines would refer to the unordered social origin of the parents (e.g. occupational category) and the columns to the (ordered) income group to which the individuals belong. Note that in addition to the four indices proposed by Reardon (2009), Silber and Yalonetzky (2011) also suggested two new indices to measure inequality in life chances. Needless to say, these indices could also be applied to the measurement of ordinal segregation.

5.3. Extensions to Other Domains: Segregation and Health Inequality

In a very important paper Allison and Foster (2004) showed that cardinal measures of inequality could not be used when attempting to measure the degree of inequality of the distribution of ordinal variable. They then defined a partial inequality ordering allowing one to decide whether a distribution was more “spread out” than another. Abul Naga and Yalcin (2008) then characterized the entire class of continuous inequality indices founded on the Allison and Foster ordering and proposed a parametric family of inequality indices that could be used with Self Rated Health Status data. More recently Lazar and Silber (forthcoming) showed that the indices of ordinal segregation recently proposed by Reardon (2009) could be also applied to the measurement of health inequality since they satisfied the four axioms specified by Abul Naga and Yalcin

(2008). They also suggested an extension of the family of indices proposed by Abul Naga and Yalcin (2008) and Reardon (2009) and gave an empirical illustration.

5.4. Segregation and Inequality in Happiness

It should be clear that concepts used to measure ordinal segregation may also be used to measure inequality in happiness. Dutta and Foster (2011) have thus applied the approaches of Allison and Foster (2004) and Abul Naga and Yalcin (2008) to the analysis of inequality in happiness in the United States during the 1972-2008 period.

5.5. Segregation and Polarization

Finally in an unpublished paper entitled "Ordinal Variables and the Measurement of Polarization" Fusco and Silber (2011) suggested applying Reardon's indices to the measurement of income polarization when only ordinal information was available on the income. Their basic idea is that the measurement of polarization, in particular bi-polarization, is based, as stressed already by Zhang and Kanbur (2001), on two basic principles:

- polarization increases with between groups inequality
- polarization decreases with within groups inequality

These two properties clearly show up in the cardinal indices of polarization proposed by Foster and Wolfson (2010). Since the denominator of the Reardon (2009) ordinal segregation indices refers to overall inequality ("variation") while the numerator is equal to the complement to one of the weighted within groups inequality, Fusco and Silber (2011) suggested that the Reardon indices could be also applied to the measurement of polarization when only ordinal information on income is available. Fusco and Silber (2011) based their empirical illustration on the 2008 cross-sectional data from the *European Union - Statistics on Income and Living Conditions* (EU-SILC). As ordinal variable they took the answers to the question where households' respondents are asked whether they are able to make ends meet. Six possible answers were proposed: (1) with great difficulty (2) with difficulty (3) with some difficulty (4) fairly easily (5) easily (6) very easily. The unordered categories refer to the citizenship of the household member who answered the household questionnaire (local versus foreigners).

Concluding Comments

In her book entitled *How to Be Human, Though an Economist* Deirdre McCloskey (2000) writes that "In the 1960s the sociologist of science Derek Price used the phrase Invisible College to describe the old-boy network of Big Science. Since then the rest of academic life has caught up to the social structure of physics, scattering old boys and old girls around the globe in each special field. The result has been damaging to the visible college and, in the end, damaging to science and scholarship. The experiment since the 1950s with turning intellectual life over to specialists has not worked, at least in the fields I know, especially economics….The odd thing about the way the advice has worked out in practice is that it has yielded a drearily uniform economics… The Kelly green golfing shoe of economics, on which all the best shoemakers agree, is microfoundations of overlapping generations in a game theoretic model with human capital and informational asymmetry… Good economics knows that specialization is not in itself good. The blessed Adam Smith (not to speak of Marx) was eloquent about the damage that specialization per se does to human spirit….What is good about specialization is that it allows more consumption, through trade….".

The lesson to be drawn is that although we have no choice but specialize, we should be aware of the usefulness of lateral thinking, as stressed by Atkinson (2011). Clearly those working on the measurement of segregation have greatly benefitted from borrowing ideas from the field of income inequality measurement. However, the extension of their analysis to multidimensional segregation is likely to be of great relevance to fields such as the inequality of opportunity, health and happiness or to that of social polarization. There is thus room for a greater level of interaction between those attempting to measure segregation and those focusing their interest on the measurement of income and more generally economic inequality.

References

Atkinson, A. B. (1970) “On the Measurement of Inequality,” *Journal of Economic Theory* 2: 244-263.

Atkinson, A. B. (2011) “On Lateral Thinking,” *Journal of Economic Inequality* 9(3): 319-328.

Blackorby, C. and D. Donaldson, 1978, Measures of relative equality and their meaning in terms of social welfare, *Journal of Economic Theory*, 18: 59-80.

Boisso, D., K. Hayes, J. Hirschberg and J. Silber (1994) “Occupational Segregation in the Multidimensional Case,” *Journal of Econom*etrics 61: 161-171.

Chakravarty, S. and J. Silber (1994) "Employment Segregation Indices: An Axiomatic Characterization," in *Models and Measurement of Welfare and Inequality*, W. Eichhorn, editor, Berlin: Springer-Verlag, pp. 912-920.

Chakravarty, S. and J. Silber (2007) “A Generalized Index of Employment Segregation,” *Mathematical Social Sciences*, 2007, 53(2): 185-195.

Chakravarty, S., C. D’Ambrosio and J. Silber (2009) "Generalized Gini Occupational Segregation Indices," in volume 17 of *Research on Economic Inequality* entitled *Occupational and Residential Segregation*, edited by Y. Flückiger, S. Reardon and J. Silber, Emerald, Bingley, UK, pp. 71-95.

Chantreuil, F. and Trannoy, A. (2012) “Inequality decomposition values: the trade-off between marginality and consistency,” *Journal of Economic Inequality*, forthcoming.

Dawkins, C. J. (2004) “Measuring the Spatial Pattern of Residential Segregation,” *Urban Studies* 41(4): 833-851.

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., Y. Flückiger and J. Silber (1994) “Measuring Occupational Segregation. Summary Statistics and the Impact of Classification Errors and Aggregation,” *Journal of Econometrics* 61: 133-146.

Deutsch, J. and J. Silber (2005) “Comparing Segregation by Gender in the Labor Force Across Ten European Countries in the 1990s. An analysis based on the use of normative segregation indices, with J. Deutsch, *International Journal of Manpower*, 2005, 26(3): 237-264.

Deutsch, J., Y. Flückiger and J. Silber (2009) "Analyzing Changes in Occupational Segregation: The Case of Switzerland (1970-2000)," in volume 17 of *Research on Economic Inequality* entitled *Occupational and Residential Segregation*, edited by Y. Flückiger, S. Reardon and J. Silber, Emerald, Bingley, UK, pp. 171-202.

Donaldson, D. and J. Weymark, 1980, A single parameter generalization of the Gini indices of inequality, *Journal of Economic Theory*, 22: 67-87.

Duncan, O.D. and Duncan B. (1955), “A methodological analysis of segregation indices”, *American Sociological Review*, Vol. 20, No. 2, pp. 210-217.

Dutta, I. and J. Foster (2010) " Inequality of Happiness in US: 1972-2008," mimeo, 2010.

Flückiger, Y. and J. Silber, 1999, *The Measurement of Segregation in the Labor Force*, Heidelberg: Physica Verlag.

Foster, J. E. and A.F. Shorrocks (1987) “Transfer Sensitive Inequality Measures,” *Review of Economic Studies* LIV: 485-497.

Foster, J. E. and M. C. Wolfson (2010) "Polarization and the Decline of the Middle Class: Canada and the U.S.," *Journal of Economic Inequality* 8: 247-273.

Frankel, D. and O. Volij (2011) “Measuring School Segregation,” *Journal of Economic Theory*, 146(1-38).

Fusco, A. and J. Silber (2011) “Quantifying Social Polarization When Variables Are Ordinal,” mimeo.

Goodman, L. A. and W. H. Kruskal (1954) “Measures of Association for Cross Classifications,” *Journal of the American Statistical Association* 49: 732-764.

Harberger, A. C. (1962) “The Incidence of the Corporate Income Tax,“ *Journal of Political Economy* 70(3): 215-240.

Hutchens, R. M. (1991) “Segregation curves, Lorenz curves and inequality in the distribution of people across occupations,” *Mathematical Social Sciences*, Vol. 21, No. 1, pp. 31-51.

Hutchens, R. M. (2001) “Numerical measures of segregation: Desirable properties and their implications,” *Mathematical Social Sciences*, Vol. 42, No. 1, pp. 13-29.

Hutchens, R.M. (2004) “One measure of segregation,” *International Economic Review*, Vol. 45, No. 2, pp. 555-578.

Jahn, J., C. F. Schmid and C. Schrag (1947) “The Measurement of Ecological Segregation,” *American Sociological Review* 12(3): 293-303.

James, D. R. and Taeuber, K. E. (1985) “Measures of segregation,” *Sociological Methodology*, 15, 1-32.

Jargowsky, P. A. and J. Kim (2009), “The information theory of segregation: uniting segregation and inequality in a common framework,” in Y. Flückiger, S. F. Reardon and J. Silber, editors, *Occupational and Residential Segregation*, vol.17 of *Research on Economic Inequality*, Emerald, Bingley, UK.

Kolm, S. C. *Monetary and Financial Choices* (in French), Dunod, Paris: 1966.

Kolm, S. C. **“**The Optimal Production of Social Justice**,”** in H. Guitton and J. Margolis eds. *Public Economics*, London: Macmillan, 1969.

Kakwani, N. C. (1994) “Segregation by sex: measurement and hypothesis testing,” in S. Neuman and J. Silber, (Eds), *Research on Economic Inequality*, JAI Press, Greenwich, pp. 1-26.

Karmel, T. and MacLachlan M. (1988) “Occupational sex segregation – increasing or decreasing,?” *Economic Record*, Vol. 64, No. 186, pp. 187-195.

Kendall, M. G. and A. Stuart (1961) *The Advanced Theory of Statistics*, London: Griffin and Company.

Kolm, S. C. (1999) “The Rational Foundations of Income Inequality Measures,” in J. Silber, *Handbook on Income Inequality Measurement*, Dordrecht, Kluwer Academic Publishers.

Lazar, A. and J. Silber (forthcoming) "On the cardinal measurement of health inequality when only ordinal information is available on individual health status," *Health Economics*.

Lewis, D. E. (1982) “The Measurement of the Occupational and Industrial Segregation of Women," *Journal of Industrial Relations* 24: 406-423.

Massey, D. S. and N. A. Denton (1988) "The Dimensions of Residential Segregation," *Social Forces* 67(2): 281-315.

McCloskey, D. *How to Be Human Though an Economist*. Ann Arbor: University of Michigan Press, 2000.

Moir, H. and J. Selby-Smith (1979) "Industrial Segregation in the Australian Labor Market," *Journal of Industrial Relations* 21: 281-291.

Mora, R. and J. Ruiz-Castillo (2003) "Additive Decomposable Segregation Indexes. The Case of Gender Segregation by Occupations and Human Capital Levels in Spain," *Journal of Economic Inequality* 1: 147-179.

Mora, R. and J. Ruiz-Castillo (2005) "An Evaluation of an Entropy Based Index of Segregation," mimeo.

Reardon, S. F. (2009) "Measures of Ordinal Segregation," in Y. Flückiger, S. F. Reardon and J. Silber, editors, *Occupational and Residential Segregation*, vol. 17 of *Research on Economic Inequality*, Emerald, Bingley, UK, pp. 129-55.

Reardon, S. F. and G. Firebaugh (2002) "Measures of multi-group segregation," *Sociological Methodology* 32: 33-67.

Reardon, S. F. and D. O'Sullivan (2004) "Measures of Spatial Segregation," *Sociological Methodology* 34(1): 121-162.

Rotschild, M. and J. E. Stiglitz (1970) "Increasing Risk: I. A Definition," *Journal of Economic Theory* 2(3): 225-243.

Rotschild, M. and J. E. Stiglitz (1971) "Increasing Risk: II. Its Economic Consequences," *Journal of Economic Theory* 3(1): 66-84.

Samuelson, P. *Foundations of Economic Analysis*, Harvard University Press, 1947.

Sastre, M. and A. Trannoy. (2002) "Shapley Inequality Decomposition by Factor Components: Some Methodological Issues," *Journal of Economics*, Supplement 9: 51-89.

Shannon, C. E. (1948) "A mathematical theory of communication," *Bell System Technical Journal*, Vol. 27, pp. 379-423, 623-656.

Shorrocks, A. F. (2012) "Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value," *Journal of Economic Inequality*, forthcoming.

Silber, J., (1989a) "Factors Components, Population Subgroups and the Computation of the Gini Index of Inequality," *The Review of Economics and Statistics*, 1989a, LXXI:107-115.

Silber, J. (1989b) "On the Measurement of Employment Segergation," Economics Letters 30(3): 237-243.

Silber, J. (1992) "Occupational Segregation Indices in the Multidimensional Case: A Note," *Economic Record* 68: 276-277.

Silber, J. and A. Spadaro (2011) "Inequality of Life Chances and the Measurement of Social Immobility," in Fleurbaey, M., M. Salles and J. Weymark editors, *Social Ethics and Normative Economics, Essays in Honour of Serge-Christophe Kolm*, Studies in Social Choice and Welfare, Springer Verlag, pp. 129-154.

Silber, J. and G. Yalonetzky (2011) "On Measuring Inequality in Life Chances when a Variable is Ordinal," Chapter 4 in Juan Gabriel Rodríguez (ed.) *Inequality of*

Opportunity: Theory and Measurement (Research on Economic Inequality),Volume 19, Emerald Group Publishing Limited, pp.77-98.

Siltanen, J., J. Jarman and R. M. Blackburn (1995) *Gender Inequality in the Labour Market. Occupational Concentration and Segregation. A Manual on Methodology*. International Labour Office, Geneva.

Theil, H. and Finizza, SA. J. (1971) "A Note on the Measurement of Racial Integration of Schools by Means of Information Concepts," *Journal of Mathematical Sociology* 1: 187-194.

Theil, H. (1967) *Economics and information theory.* Amsterdam: North-Holland.

Watts, M. (1998) "Occupational gender segregation: index measurement and econometric modelling," *Demography,* 35(4), 489-496.

Zhang, X. and R. Kanbur (2001) "What Differences Do Polarization Measures Make? An Application to China," Journal of Development Studies 37: 85-98.

CEPS
INSTEAD

3, avenue de la Fonte
L-4364 Esch-sur-Alzette
Tél.: +352 58.58.55-801
www.ceps.lu