The Concept of Shapley Decomposition and the Study of Occupational Segregation

Joseph Deutsch*

Yves Flückiger**

Jacques Silber*

* Department of Economics, Bar-Ilan University, 52900 Ramat-Gan, Israel.
Emails: jdeutsch@mail.biu.ac.il and silberj@mail.biu.ac.il

** Département d'éçonomie politique, Université de Genève, Unimail, 40 bd du Pont d'Arve, 1211 Genève 4, Switzerland.
Email: yves.flueckiger@ecopo.unige.ch

September 2005

Abstract

This paper generalizes a decomposition procedure originally proposed by Karmel and McLachlan. The idea is to combine their approach with what is now known as the Shapley decomposition. Such a generalization offers a clear breakdown of the variation over time in occupational segregation (change in gross segregation) into a component measuring changes in net segregation and another one corresponding to changes in the margin, the latter itself including variations in the occupational structure and in the shares of the subpopulations (e.g. the genders) in the labor force. This new decomposition may easily be extended to the cases where more than two categories are distinguished or when there are more than three dimensions. The results of the empirical illustration based on Swiss data for 1970 and 2000 proved the usefulness of the approach. They stressed in particular that in several instances variations in gross and net segregation worked in opposite directions.

I) Introduction

The study of occupational segregation by gender has been mainly the work of sociologists. In recent years some economists have however shown some interest in the analysis of gender differences in the occupational structure (see, for example, Anker, 1998). Moreover several economists (see, Butler, 1987, and Silber, 1989a) have stressed that the study of occupational segregation could greatly benefit from the "import" of techniques used when measuring income inequality (see, Flückiger and Silber, 1999, for a survey on the measurement of segregation in the labor force). Silber (1992) also showed how a generalization of the famous Duncan dissimilarity index allowed one to study multidimensional segregation which refers either to the case when more than two categories are distinguished (e.g. more than two ethnic groups rather than only two genders) or to that where segregation is analyzed along more than two dimensions (e.g. segregation by occupation, industry and gender). More recently Deutsch et al. (2005) took an additional step in adapting the approach used in income inequality measurement to the study of occupational segregation, by proposing a "normative approach" to the analysis of segregation. Such a normative view of segregation allows one to determine whether occupational segregation by gender is mainly the consequence of the presence of "male-intensive" occupations or whether it is rather due to the existence of "female-intensive" occupations. Moreover it gives policy makers the possibility to select the weight they want to give to very "male- or female-intensive" occupations.

The purpose of this paper is to show how an additional tool recently introduced in income inequality analysis, the game theory concept of Shapley value, could be combined with a technique originally suggested by Deming and Stephan (1940) and applied to the field of segregation by gender by Karmel and McLachlan (1988) and Watts (1998a), to make a very clear distinction between changes in "gross segregation" and changes in "net segregation", the term "net" referring here to "net of changes in occupational weights and in the labor force participation rates of both genders". The paper is organized as follows. Section II recalls quickly how occupational segregation by gender has been usually analyzed in the literature.

Section III explains then how the traditional Duncan and Gini index of occupational segregation have been generalized to take into account more than two dimensions. Section IV recalls how Karmel and McLachlan (1988) suggested to decompose variations over time in the generalized Duncan index and shows how this decomposition procedure may be made systematic once the concept of Shapley decomposition is applied. Section V offers then an empirical illustration based on Swiss Census data for the years 1970 and 2000 and studies variations in occupational segregation by gender, nationality and age as well as changes in occupational segregation by gender, separately for Swiss and foreign workers. Concluding comments are given in Section VI.

II) The Traditional Analysis of Occupational Segregation by Gender

The most popular measure of occupational segregation by gender is the so-called Duncan Index (Duncan and Duncan, 1955) which is defined as

$$I_D = (1/2) \sum_{i=1 \text{ to } n} \left| (M_i / M) - (F_i / F) \right| \quad (1)$$

where M_i and F_i represent respectively the number of men and women in occupation i while M and F refer to the total number of men and women in the labor force, n being the number of occupations. One can prove (see, Flückiger and Silber, 1999) that

$$I_D = (1/2) \sum_{i=1 \text{ to } n} (M_i / M) \left[\left| (F_i / M_i) - (F / M) \right| / (F/M) \right] \quad (2)$$

Expression (2) shows clearly that the Duncan index belongs to the family of relative mean deviations (with respect to the mean) since for each occupation i the ratio (F_i/M_i) is compared with the overall ratio (F/M) in the labor force, the occupations being weighted by the relative share (M_i/M) of men in this occupation and standardized by the overall ratio (F/M).

Despite its extreme popularity the Duncan index has quite a few shortcomings (see, Flückiger and Silber, 1999) and this is why Silber (1989a) recommended using another index, called the G-segregation index, which is directly derived from Gini's concentration ratio. This G-segregation index is defined (see, Deutsch et al., 1994) as

$$I_G = \sum_{i=1 \text{ to } n} \sum_{j=1 \text{ to } n} (1/2)[(M_i/M)(M_j/M) \left| (F_i / M_i) - (F_j / M_j) \right| / (F/M)] \quad (3)$$

One may observe the similarity between the definitions of the Duncan and the G-segregation index since instead of comparing, for each occupation i, the ratio (F_i/M_i) with the average ratio (F/M), one makes all binary comparisons between the ratios (M_i/M) and (M_j/M), these comparisons being weighted by the product of the shares (M_i /M) and (M_j /M) and standardized by the overall ratio (F/M). Silber (1989a) proved that

$$I_G = [(M_1/M)\ldots(M_k/M)\ldots(M_K/M)]' \; G \; [(W_1/W)\ldots(W_k/W)\ldots(W_K/W)] \qquad (4)$$

where $[(M_1/M)\ldots(M_k/M)\ldots(M_K/M)]'$ is a row vector of the K shares corresponding to the percentage of males working in the various occupations. Similarly, in expression (4), $[(W_1/W)\ldots(W_k/W)\ldots(W_K/W)]$ is a column vector giving the percentage of females in the various occupations. Note that in these two vectors the shares are classified by decreasing values of the ratios (W_k/M_k). The operator G in (4) is called the G-matrix and its typical element g_{hk} is equal to 0 if h=k, to –1 if k>h and to +1 if h>k (see, Silber, 1989b). Expression (4) indicates that occupational segregation by gender is estimated by comparing a set of “prior shares” $\{(M_1/M),\ldots,(M_k/M),\ldots,(M_K/M)\}$ with a set of “posterior shares” $\{(W_1/W),\ldots,(W_k/W),\ldots,(W_K/W)\}$, the comparison being made via the mathematical operator G. Such an interpretation of the Gini-segregation index is also at the basis of what is known as the Segregation Curve (see, Duncan and Duncan, 1955).

The Segregation curve is a graphical tool that was originally introduced by Duncan and Duncan (1955). It is an extension to the analysis of segregation of the famous Lorenz curve used to analyze income inequality. The segregation curve is drawn as follows.

Let us first classify the occupations by increasing ratios (F_i/M_i). On the horizontal axis we will put the cumulated values of the shares (M_i /M) while on the vertical axis we put the cumulative values of the shares (F_i /F), the occupations being classified by increasing values of the ratios (F_i/M_i). The segregation curve is then the plot of these two sets of cumulative values. The segregation curve clearly starts at the point (0,0) and ends at the point (1,1) and its slope will never be decreasing since we classified the occupations by increasing ratios (F_i/M_i). It can be observed that if for each occupation the share of males (in the total number of males) is equal to the share of females (in the total number of females) the segregation curve will be identical to the diagonal line joining the points (0,0) to the point (1,1). Moreover the further away this Segregation Curve is from this diagonal, the more segregation there is. This is why the area between a Segregation Curve and the diagonal has been proposed as a measure of segregation. In fact one may show that the Gini segregation index is equal to twice the area lying between a segregation curve and the diagonal. Moreover it can also be proven that the highest vertical distance between the diagonal and the Segregation Curve is equal to the previously defined Duncan index (see, Flückiger and Silber, 1999).

III) On Generalizations of the Duncan and Gini Segregation Indices:

One can imagine cases where one is interested in comparing the occupational distribution of more than two categories (e.g. men, single women and married women or four ethnic groups). To solve this problem Silber (1992), following a suggestion of Karmel et McLachlan (1988), proposed a generalization of the Duncan Index. The basic idea of this generalization is as follows. Assume we know the occupational distribution of the two genders. We have hence a data matrix where the lines i correspond to the various occupations and the columns j to the two genders. The typical element T_{ij} of this matrix will then tell us how many individuals of gender j are employed in occupation i. Let $T = \sum_{i=1 \text{ to } n} \sum_{j=1 \text{ to } 2} T_{ij}$ be the total number of individuals in the labor force. The ratio (T_{ij}/T) will then give us the proportion of individuals employed in occupation i and of gender j. This ratio may be also interpreted as the "posterior probability" that an individual of gender j is employed in occupation i. It should however be clear that if there is independence between the genders and the occupations, the probability that an individual of gender j will be employed in occupation i will be equal to the product $(T_{i.}/T)$ $(T_{.j}/T)$ where $T_{i.}$ refers to the total number of individuals employed in occupation i, whatever their gender, and $T_{.j}$ the total number of individuals of gender j. If there is at least one element (i,j) of the matrix for which there is no identity between the posterior probability (T_{ij}/T) and the prior probability $(T_{i.}/T)$ $(T_{.j}/T)$ one will conclude that there is occupational segregation by gender. Such an approach may clearly be generalized to the case where there are more than two categories. Such a generalization of the Duncan index will be expressed as

$$I_{Dg} = \sum_{i=1 \text{ to } n} \sum_{j=1 \text{ to } n} [\,|(T_{ij} / T) - ((T_{i.} /T)(T_{.j} /T))|\,] \quad (10)$$

IV) The Decomposition of Changes over Time in Occupational Segregation:

A) Traditional Decomposition Approaches

Assume we know the value of one of the segregation indices at times t and t' and that the index decreased over time. It may then be interesting to find out why it decreased. One may think of two sets of causes: either the relative weight of the various occupations varied over time (e.g. the share of occupations where the ratio (F_i /M_i) was very high or very low decreased) or the occupations that were essentially "female occupations" became less "female" ((F_i /M_i) decreased) and/or those that were essentially "male" became less "male" ((F_i /M_i) increased). Such a decomposition stressing the respective role of the occupational weights and the occupation specific "gender ratios (F_i /M_i)" has been proposed by Blau et al. (1979) who used the Duncan index and by Boisso et al. (1994) who used a similar generalization of the G-segregation index while Flückiger and Silber (1999) used both generalizations.

B) The Shapley Value and Variations over Time in the Index I_{Dg}

Karmel and McLachlan (1988) proposed however a different approach based on the idea of a "marginal free" decomposition of the variation in segregation. A segregation index is said to be "marginal-free" if it is not affected by changes in the overall gender and occupational composition of the labor force (see, Charles, 1992, Blackburn et al., 1993, Watts, 1998a, Flückiger and Silber, 1999, for more details on this concept). In other words Karmel and MacLachlan (1988) made a distinction between a variation in occupational segregation by gender that is due to a variation in the relative weights of the occupations or to a change in the overall proportions of men and women in the labor force and a "net change in segregation", a change that has nothing to do with changes in the relative weights of the occupations or of the genders. In addition to these three changes there usually is also an interaction term. Flückiger and Silber (1999) but also Watts (1998a) adopted also Karmel and McLachlan's approach. We will now show that it is possible to combine the Karmel and McLachlan approach with an income inequality decomposition technique based on the concept of Shapley value so popular in cooperative game theory (see Chantreuil and Trannoy, 1999, Shorrocks, 1999 and Sastre et Trannoy, 2002, for a presentation of this application of the concept of Shapley value). Such a decomposition has no interaction term and makes a distinction between three sources of variation. A first impact is the consequence of variations over time in the relative weights of the different occupations. The second effect is the consequence of variations over time in the relative weights of the genders in the total labor force. Finally a third element measures the "net change" in segregation (net of the two first impacts) and refers to changes in the "internal structure" of the matrix because it is assumed in this case that no change in the margins of the matrix took place.

The sum of these three components will then be called "gross variation" in occupational segregation by gender. The approach taken by Karmel and Maclachlan (1988) is based on a technique originally introduced by Deming and Stephan (1940). It is however possible to derive a decomposition that is more general than the one introduced by Karmel and Maclachlan (1988). To simplify let p_{ij}, $p_{i.}$ and $p_{.j}$ refer respectively to the ratios (T_{ij}/T), $(T_{i.}/T)$ and $(T_{.j}/T)$ defined previously. Since the product $(p_{i.}\, p_{.j})$ is in fact equal to the product of the margins i and j of the matrix $\{p_{ij}\}$ whose typical element is p_{ij}, we will call q_{ij} the product $(p_{i.}\, p_{.j})$. The generalized Duncan index may then be expressed as

$$I_{Dg} = h\,(p_{ij},q_{ij}) = h\,(p_{ij},q_{ij}) = \sum_{i=1 \text{ to } m} \sum_{j=1 \text{ to } 2} \left| p_{ij} - q_{ij} \right| \qquad (11)$$

To compare occupation segregation by gender at two periods 0 and 1 Karmel and Maclachlan (1988), following Deming and Stephan (1940), proceeded as follows. The idea, when comparing two matrices of proportions $\{p_{ij}\}$ and $\{v_{ij}\}$, is to build a third matrix $\{s_{ij}\}$ that will have the « internal sructure » of the matrix $\{p_{ij}\}$but the margins of the matrix $\{v_{ij}\}$. To derive $\{s_{ij}\}$ one has to multiply first all the elements (p_{ij}) of the matrix $\{p_{ij}\}$ by the ratios $(v_{i.}/p_{i.})$

where $v_{i.}$ and $p_{i.}$ refer to the horizontal margins of the matrices $\{p_{ij}\}$and $\{v_{ij}\}$. Call $\{x_{ij}\}$the matrix you get after such a multiplication. Multiply then all the elements (x_{ij}) of this matrix $\{x_{ij}\}$by the ratios ($v_{.j}/x_{.j}$) where $v_{.j}$ and $x_{.j}$ refer to the vertical margins of the matrices $\{v_{ij}\}$and $\{x_{ij}\}$. Call $\{y_{ij}\}$the matrix you get after this second multiplication. If we continue this procedure multiplying now the elements (y_{ij}) of the matrix $\{y_{ij}\}$ by the ratios ($v_{i.}/y_{i.}$), where ($v_{i.}$) and ($y_{i.}$) are the horizontal margins of the matrices $\{v_{ij}\}$and $\{y_{ij}\}$, and so on, the matrices one successively derives will converge, as proven by Deming and Stephan (1940), towards a matrix $\{s_{ij}\}$ that will have the margins of the matrix $\{v_{ij}\}$ but the internal structure of the matrix $\{p_{ij}\}$.

We could evidently have started with the matrix $\{v_{ij}\}$ and end up with a matrix $\{w_{ij}\}$ that would have the margins of the matrix $\{p_{ij}\}$ but the internal structure of the matrix $\{v_{ij}\}$. As was previously explained this transition from a matrix $\{p_{ij}\}$ to a matrix $\{v_{ij}\}$ includes in fact two stages: a first one where only the margins of the matrix $\{p_{ij}\}$ vary and a second one where the internal structure of this matrix is modified.

Call $\Delta I = I(v) - I(p)$ the overall change between two periods in the degree of occupational segregation by gender, I being the generalized Duncan index I_{DG}. ΔI may also be expressed as being equal to $\Delta I = f(\Delta m, \Delta is)$ where Δm and Δis measure respectively the change in the margins and the variation in the internal structure of the original matrix.

We can now borrow a decomposition technique based on the concept of Shapley value that was originally suggested by Chantreuil and Trannoy (1999), extended by Shorrocks (1999) and applied by Sastre and Trannoy (2002).

A short summary of the concept of Shapley decomposition

Let F(a,b) be a function depending on two variables a and b. Such a function need not be linear. Although Chantreuil and Trannoy (1999) and Sastre et Trannoy (2002) limited their application of the Shapley value to the decomposition of income inequality, Shorrocks (1999) has shown that such a decomposition could be applied to any function.

The idea of the Shapley value is to consider all the possible sequences allowing us to eliminate the variables a and b. Let us start with the elimination of the variable a. This variable may be the first one or the second one to be eliminated. If it is eliminated first, the function F(a,b) will become equal to F(b) since the variable a has been eliminated so that in this case the contribution of a to the function F(a,b) is equal to F(a,b) – F(b). If the variable a is the second one to be eliminated the function F will then be equal to F(a). Since both elimination sequences are possible and assuming the probability of these two sequences is the same, we may conclude that the contribution C(a) of the variable a to the function F(a,b) is equal to

$$C(a) = (1/2)[F(a,b) - F(b)] + (1/2)F(a) \quad (12)$$

Similarly one can prove that the contribution C(b) of the variable b to the function F(a.b) is

$$C(b) = (1/2)[F(a,b) - F(a)] + (1/2) F(b) \quad (13)$$

Combining (12) and (13) we observe that

$$C(a) + C(b) = F(a,b) \quad (14)$$

Applying Shapley's decomposition to the analysis of variations over time in the value of the generalized Duncan index

Using expressions (12) to (14) we may express the contribution $C_{\Delta m}$ of the variations of the margins to the overall change ΔI in occupational segregation by gender as

$$C_{\Delta m} = (1/2) f(\Delta m) + (1/2) [f(\Delta m, \Delta is) - f(\Delta is)] \quad (15)$$

where Δm and Δis refer respectively to the change in the margins and to that in the internal structure of the original matrix.

Similarly the contribution $C_{\Delta is}$ of the variation in the internal structure of the matrix to the overall change ΔI in occupational segregation by gender will be

$C_{\Delta is} = (1/2)\, f(\Delta is) + (1/2)\, [f(\Delta m, \Delta is) - f(\Delta m)]$ (16)

It is easy to observe that $C_{\Delta m} + C_{\Delta is} = \Delta I$

Using the various matrices that were defined previously, one may prove that the contributions $C_{\Delta m}$ and $C_{\Delta is}$ may be also expressed as

$C_{\Delta m} = (1/2)\, \{[I(s) - I(p)] + [I(v) - I(w)]\}$ (17)

$C_{\Delta is} = (1/2)\, \{[I(w) - I(p)] + [I(v) - I(s)]\}$ (18)

so that, as expected,

$C_{\Delta m} + C_{\Delta is} = I(v) - I(p)$ (19)

If we now apply the concept of "Nested Shapley Decomposition", as suggested by Sastre and Trannoy (2002), we can also decompose the contribution $C_{\Delta m}$ into two components corresponding respectively to the contributions of the horizontal and vertical margins.

The idea is to derive first a matrix l that would have the internal structure of the matrix p, the vertical margins of this same matrix p but the horizontal margins of the matrix v. We therefore need to build a matrix that will have the vertical margins of the matrix p and the horizontal margins of the matrix v. There are many such matrices, among which a matrix n where each element n_{ij} is equal to the product of the margins $v_{.j}$ and $p_{i.}$. If we now apply the technique proposed by Deming et Stephan (1940) to the case where the original matrix is p and the final matrix is n, the matrix p will converge towards a matrix l that will have the internal structure of the matrix p, the vertical margins of this same matrix p but the horizontal margins of the matrix v since the matrix n has the horizontal margins of the matrix v and the vertical margins of the matrix p.

We can use the same procedure to define

- a matrix k that will have the internal structure of the matrix p, the vertical margins of the matrix v and the horizontal margins of the matrix p

- a matrix c that will have the internal structure of the matrix v, the vertical margins of the matrix v and the horizontal margins of the matrix p
- a matrix f that will have the internal structure of the matrix v, the vertical margins of the matrix p and the horizontal margins of the matrix v

Let D_{m1} be the change defined as

$$D_{m1} = [I(s) – I(p)] \tag{20}$$

Since the matrices s and p that were previously defined have the same internal structure (is) we may write that

$$D_{m1} = g[(\Delta h\neq 0),(\Delta t\neq 0), (\Delta is=0)] \tag{21}$$

where Δh and Δt correspond to horizontal and vertical variations of the margins.

The contributions $C_{\Delta h1}$ and $C_{\Delta t1}$ to the difference D_{m1} may therefore be expressed as

$$C_{\Delta h1}= (1/2)\ g[(\Delta h\neq 0),(\Delta t=0),(\Delta is=0)] \\ +(1/2)\{g[(\Delta h\neq 0),(\Delta t\neq 0),(\Delta is=0)]–g[(\Delta h=0),(\Delta t\neq 0), (\Delta is=0)]\} \tag{22}$$

and

$$C_{\Delta t1}= (1/2)\ g[(\Delta h=0),(\Delta t\neq 0),(\Delta is=0)] \\ +(1/2)\{g[(\Delta h\neq 0),(\Delta t\neq 0),(\Delta is=0)]–g[(\Delta h\neq 0),(\Delta t=0), (\Delta is=0)]\} \tag{23}$$

We must now derive the expressions corresponding to $g[(\Delta h\neq 0),(\Delta t=0),(\Delta is=0)]$ and $g[(\Delta h=0),(\Delta t\neq 0), (\Delta is=0)]\}$.

Using the definitions of the matrices l and p that were given previously we conclude that

$$g[(\Delta h\neq 0),(\Delta t=0),(\Delta is=0)]\} = I(l) – I(p) \tag{24}$$

Similarly we derive that

$$g[(\Delta h=0),(\Delta t\neq 0),(\Delta is=0)]\}=I(k)– I(p) \tag{25}$$

Combining the expressions (22) to(25) we derive

$$C_{\Delta h1}= (1/2)\ g[(\Delta h\neq 0),(\Delta t=0),(\Delta is=0)]$$

$$+(1/2)\{g[(\Delta h\neq 0),(\Delta t\neq 0),(\Delta is=0)]-g[(\Delta h=0),(\Delta t\neq 0),(\Delta is=0)]\}$$

$$\Leftrightarrow C_{\Delta h1}=(1/2)[I(l)-I(p)]+(1/2)\{[I(s)-I(p)]-[I(k)-I(p)]\}$$

$$\Leftrightarrow C_{\Delta h1}=(1/2)[I(l)-I(p)]+(1/2)\{[I(s)-I(k)] \tag{26}$$

and

$$C_{\Delta t1}=(1/2)\,g[(\Delta h=0),(\Delta t\neq 0),(\Delta is=0)]$$

$$+(1/2)\{g[(\Delta h\neq 0),(\Delta t\neq 0),(\Delta is=0)]-g[(\Delta h\neq 0),(\Delta t=0),(\Delta is=0)]\}$$

$$\Leftrightarrow C_{\Delta t1}=(1/2)[I(k)-I(p)]+(1/2)\{[I(s)\text{-}I(p)]-[I(l)\text{-}I(p)]\}$$

$$\Leftrightarrow C_{\Delta t1}=(1/2)[I(k)-I(p)]+(1/2)\{[I(s)\text{-}I(l)] \tag{27}$$

It is easy to observe that

$$C_{\Delta h1}+C_{\Delta t1}=I(s)-I(p) \tag{28}$$

Let us now similarly decompose the difference D_{m2} defined previously as being equal to

$$D_{m2}=[I(v)-I(w)] \tag{29}$$

Given the definitions of the matrices c and f that were given previously, we define the contributions $C_{\Delta h2}$ and $C_{\Delta t2}$ to the difference D_{m2} as being equal to

$$C_{\Delta h2}=(1/2)\,g[(\Delta h\neq 0),(\Delta t=0),(\Delta is=0)]$$

$$+(1/2)\{g[(\Delta h\neq 0),(\Delta t\neq 0),(\Delta is=0)]-g[(\Delta h=0),(\Delta t\neq 0),(\Delta is=0)]\} \tag{30}$$

$$\Leftrightarrow C_{\Delta h2}=(1/2)[I(v)-I(c)]+(1/2)\{[I(v)-I(w)]-[I(v)-I(f)]\}$$

$$\Leftrightarrow C_{\Delta h2}=(1/2)[I(v)-I(c)]+(1/2)\{[I(f)-I(w)] \tag{31}$$

and

$$C_{\Delta t2}=(1/2)g[(\Delta h=0),(\Delta t\neq 0),(\Delta is=0)]$$

$$+(1/2)\{g[(\Delta h\neq 0),(\Delta t\neq 0),(\Delta is=0)]-g[(\Delta h\neq 0),(\Delta t=0),(\Delta is=0)]\}$$

$$\Leftrightarrow C_{\Delta t2}=(1/2)[I(v)-I(f)]+\{[I(v)\text{-}I(w)]-[I(v)-I(c)]\}$$

$$\Leftrightarrow C_{\Delta t2}=(1/2)[I(v)-I(f)]+\{[I(c)\text{-}I(w)] \tag{32}$$

It easy to observe that

$$C_{\Delta h2} + C_{\Delta t2} = I(v) - I(w) \quad (33)$$

Combining now equations (16) to (33) we conclude that the contribution $C_{\Delta m}$ may be defined as being equal to

$$C_{\Delta m} = C_h + C_t \quad (34)$$

where

$$C_h = (1/2)\ [C_{\Delta h1} + C_{\Delta h2}]$$

$$\Leftrightarrow C_h = (1/2)\ (1/2)\ [\{[I(l)–I(p)] + [I(s)–I(k)]\} + \{[I(v) - I(c)] + [I(f) - I(w)]\}] \quad (31)$$

and

$$C_t = (1/2)\ [C_{\Delta t1} + C_{\Delta t2}]$$

$$\Leftrightarrow C_t = (1/2)\ (1/2)\ [\{[I(k) - I(p)] + [I(s)-I(l)]\} + \{[I(v) - I(f)] + [I(c)-I(w)]\}] \quad (36)$$

Combining (34), (35) and (36) we finally derive, as expected, that

$$C_{\Delta m} = (1/2)\ \{[I(s) - I(p)] + [I(v) - I(w)]\} \quad (37)$$

The decomposition that was proposed here has the advantage of being systematic and general since it takes into account the possibility that the variation in segregation will be measured by going from period 1 to period 0. Moreover it is easy to generalize this analysis to the case where the matrix has more than two dimensions (e.g. segregation by occupation, industry and gender).

V) An Empirical Illustration:

The data sources we used were the Swiss censuses for the years 1970 and 2000. We first computed the value of the Duncan generalized index of occupational segregation by gender for both years, Then we similarly computed the value of this index for occupational segregation by nationality (Swiss versus foreigners, making no distinction between the genders of the workers). Finally we repeated the exercise to compute the value of the Duncan generalized index to measure occupational segregation by age, making no distinction between the genders or the nationalities of the workers). Then in each of these three cases, we computed the change observed, over this thirty years period, in the value of the Generalized Duncan Index and, using the methodology previously described, we decomposed this change into two components, one measuring the changes in the margins, the other the change in the "internal structure". Finally the change in the margins was itself broken down into a change into a component measuring the change in the occupational structure and another one measuring the variations in the shares of the subpopulations analyzed. The results of this decomposition are given in Table 1.

Let us first take a look at occupational segregation by gender. Remembering that the overall variation in the value of the Generalized Duncan Index will be called "gross variation in segregation" while the change in the internal structure is labeled "net variation in occupational segregation" we observe opposite trends between these gross and net variations. Table 1 indicates a slight increase in gross segregation by gender but a decrease in net segregation. The overall increase (change in gross segregation) is due to the fact that the changes in the margins more than compensated that in the internal structure. But even the change in the margins is the consequence of opposite forces. Variations in the occupational structure would have per se led to a decrease in gross segregation. This impact of the occupational structure is probably due to the fact that the share of professions belonging to the "third sector" increased and in this sector the share of women is relatively high. The change in the relative shares of the genders in the labor force, which clearly is a consequence of an increase in the labor force participation of women, worked in the opposite direction, was stronger in absolute value and would per se have led to an increase in gross segregation. This is probably a consequence of the fact that the women who entered the labor force between 1970 and 2000 worked in occupations traditionally considered as feminine. This first illustration shows thus clearly how important it is to make a distinction between changes in gross and net segregation.

The second illustration in Table 2 refers to changes in occupational segregation by nationality. Here we observe first, not surprisingly, that this type of segregation is much smaller than occupational segregation by gender, second that the variations in gross and net segregation are in the same direction (of an important decrease in segregation). The respective impacts of variations in the “internal structure” and in the margins are quite similar both in magnitude and in sign. Even the respective impact of variations in the occupational structure and in the shares of the Swiss and foreigners in the labor force are similar in magnitude and sign.

The third illustration concerns occupational segregation by age, a distinction being made between workers, whatever their gender or nationality, younger than 50 and those who are 50 years old or more. Table 2 indicates that there was more than a halving of this type of occupational segregation (which in any case is much smaller than the two other ones) and that the whole change was a consequence of a change in the “internal structure”. In other words in this third illustration the picture given by changes in gross segregation is identical to that given by changes in net segregation.

In table 2 we look only at occupational segregation by gender but we make a distinction between Swiss and foreign workers. It then appears that among Swiss workers there was an increase between 1970 and 2000 in the gross variation in occupational segregation by gender but a decrease in the net variation in this occupational segregation. So here is another illustration of the usefulness of making a distinction between gross and net variation. The data thus show that “pure” occupational segregation by gender decreased, among Swiss workers, between 1970 and 2000. This decrease however was more than compensated by variations in the "margins", that is, by changes in the occupational structure or in the shares of the genders in the overall Swiss labor force. Note however that changes in the occupational structure per se would have led to a decrease in the overall segregation by gender but this variation was more than compensated by changes in the shares of men and women in the labor force (of workers having the Swiss nationality) so that the total impact of changes in the "margins" is an increase in segregation.

Among foreign workers the results are simpler in the sense that both variations in the "internal structure" and in the "margins" were of almost equal magnitude and of equal sign and led both to a decrease in segregation. Not however that here also variations in the occupational structure would have per se led to a decrease in segregation while changes in the shares of the genders would have per se led to an increase in segregation, the magnitude (in absolute value) of this latter change being higher than than that of the former.

The illustrations given in Tables 1 and 2 indicate therefore quite clearly how useful the methodology, originally proposed by Karmel and McLachlan and extended in the present study, is and how important it is to make a distinction between gross and net variations in occupational segregation.

VI) Concluding Comments:

This paper tried to generalize the decomposition procedure originally proposed by Karmel and McLachlan (1988) when they decomposed the so-called generalized Duncan index. The idea is to combine their approach with what is now known as the Shapley decomposition. Such a generalization offers a clear breakdown of the variation over time in occupational segregation (change in gross segregation) into a component measuring changes in net segregation and another one corresponding to changes in the margin, the latter itself including variations in the occupational structure and in the shares of the subpopulations (e.g. the genders) in the labor force. Such a decomposition includes no interaction term whereas that suggested by Karmel and McLachlan (1988) includes such a term. This decomposition may easily be extended to the cases where more than two categories are distinguished (e.g. several ethnic groups) or when there are more than three dimensions (e.g. segregation by gender, occupation and economic branch). The results of the empirical illustration, which looked at Swiss data for the years 1970 and 2000, proved the usefulness of such an approach. They stressed in particular that in several instances variations in gross and net segregation worked in opposite directions, the same being sometimes true of changes in the two margins, that referring to variations in the occupational structure and that measuring variations in the relative shares of the subpopulations distinguished.

Bibliography

Anker, R., 1998, *Gender and Jobs: Sex Segregation of Occupations in the World*. Geneva, International Labor Organization.

Blackburn, R. M., J. Jarman and J. Siltanen, 1993, The Analysis of Occupational Gender Segregation over Time and Place: Considerations of Measurement and Some New Evidence, *Work, Employment and Society* 7.

Blau, F., D. Hendricks and E. Wallace, 1979, Occupational Segregation by Sex: Trends and Prospects, *Journal of Human Resources*, 14(2): 197-210.

Boisso, D., K. Hayes, J. Hirschberg and J. Silber, 1994, Occupational Segregation in the Multidimensional Case: Decomposition and Tests of Statistical Significance, *Journal of Econometrics*, 1994, 61: 161-171.

Butler, R. J.,, 1987, New Indices of Segregation, *Economics Letters* 24: 359-362.

Chantreuil, F. and A. Trannoy, 1999, Inequality Decomposition Values: The Trade-Off Between Marginality and Consistency. THEMA Discussion Paper, Université de Cergy-Pontoise.

Charles, M., 1992, Cross-National Variation in Occupational Sex Segregation, *American Sociological Review* 57: 482-503.

Deming, W. E. and F. F. Stephan, 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 Indices and the Impact of Classification Errors and of Aggregation, *Journal of Econometrics*, 1994, 61: 133-146.

Deutsch, J. and J. Silber, 2005, Comparing segregation by gender in the labour force across ten European countries in the late 1990s: An analysis based on the use of normative segregation indices, *International Journal of Manpower*, 26(3): 237-264.

Duncan, O. D. and B. Duncan, 1955, A Methodological Analysis of Segregation Indices, *American Sociological Review*, 20: 210-217.

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

Karmel, T. and M. Maclachlan, 1988, Occupational Sex Segregation – Increasing or Decreasing, *Economic Record* 64: 187-195.

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

Shorrocks, A. F. (1999), Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value, mimeo, University of Essex.

Silber, J., 1989a, On the Measurement of Employment Segregation, *Economic Letters*, 30:237-243.

Silber, J., 1989b, Factor components, population subgroups and the computation of the Gini index of inequality, *Review of Economics and Statistics* 71: 107-115.

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

Watts, M., 1992, How Should Occupational Segregation be Measured, *Work, Employment and Society* 6: 475-487.

Watts, M., 1998a, Occupational Gender Segregation: Index Measurement and Econometric Modelling, *Demography* 35 (4): 489-496.

Watts, M., 1998b, The Analysis of Sex Segregation: When is Index Measurement Not Index Measurement?, *Demography* 35(4): 505-508.

Table 1: Decomposition of the Change in Switzerland between 1970 and 2000 in the Generalized Duncan Index (Occupational Segregation by Gender, Nationality or Age)

Criterion of Comparison of Subpopulations	Value of the Index in 1970	Value of the Index in 2000	Change observed between 1970 and 2000	Component of the change due to variations in the "internal structure"	Component of the change due to variations in the "margins"	Component due to variation in the occupational structure	Component due to variations in the shares of the subpopulations
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 (up to 50 and above 50)	0.1325	0.0651	-0.0673	-0.0691	0.0017	0.0036	-0.0019

Table 2 : Decomposition of the variation in Switzerland between 1970 and 2000 of the degree occupational segregation by gender, separately for Swiss and foreign workers (based on the use of the Generalized Duncan Index)

Criterion of Comparison of Subpopulations	Value of the Index in 1970	Value of the Index in 2000	Change observed between 1970 and 2000	Component of the change due to variations in the "internal structure"	Component of the change due to variations in the "margins"	Component due to variation in the occupational structure	Component due to variations in the shares of the subpopulations
Swiss	0.4683	0.4905	0.0223	-0.0224	0.0447	-0.0114	0.0561
Foreigners	0.5210	0.4705	-0.0505	-0.0269	-0.0235	-0.0617	0.0382