STATISTICS IN TRANSITION-new series, December 2010
Vol. 11, No. 3, pp. 597—614

# ON THE LINK BETWEEN SILBER AND DAGUM DECOMPOSITIONS OF THE GINI INDEX

**Mauro Mussini**[1]

## ABSTRACT

In this paper, we combine two alternative procedures for decomposing the Gini index: the Silber matrix approach and the Dagum technique based on Gini mean difference. This combined decomposition can be used for overlapping as well as no overlapping population subgroups. The presented decomposition procedure derives the within-group, the between-group, and the overlapping inequalities by exploiting all of the information on income distribution contained in one matrix accounting for pair wise disparities between per capita income shares. Moreover, this matrix approach can serve the scope of counting the frequency of overlapping, providing new insights for a more complete analysis of overlapping. The suggested methodology is applied to survey data on Italian employee incomes in 2000 and 2008.

**Key words:** Inequality measure; Gini decomposition; overlapping.

## 1. Introduction

The literature offers various approaches for calculating the Gini index and decomposing it by subgroup (Bhattacharya and Mahalanobis, 1967; Giorgi, 1999; Shorrocks, 1984; Pyatt, 1976). Among those procedures using matrix algebra, there is Silber (1989) method. Otherwise, Dagum (1997) decomposed the Gini index starting from its expression in terms of relative mean difference. This paper combines Silber and Dagum approaches in a matrix form useful for the subgroup decomposition of the Gini index.

Introducing an algebraic linear operator, called *G*-matrix, Berrebi and Silber (1987) proposed a matrix compact form for computing the Gini index. Silber (1989) derived a decomposition of this matrix expression either by subgroup or by income source. Considering per capita income shares, Silber technique captures within and between group inequalities, the latter accounting for inequality between the average incomes of the population subgroups. Then,

[1] University of Milan Bicocca. Email: mauro.mussini1@unimib.it.

subtracting these two measures of inequality from the overall Gini index, one obtains the overlapping component of the overall inequality index.

Dagum decomposition yields a breakdown of the Gini index on the basis of Gini pair wise income difference criterion, reckoning the inequality within subgroups and the inequality across subgroups. While the within-group inequality is equivalent to the within-group component of total inequality measured by Silber technique, the across-group component measures the inequality due to the disparities between subgroup income distributions instead of the differences between subgroup average incomes (Mussard, 2004). By using the concept of relative economic affluence (Dagum, 1987), Dagum further decomposed the across-group inequality into the net across-group inequality and the inequality contribution attributable to transvariation. Monti (2008) pointed out that the net across-group and the transvariation components are equivalent to the well known between-group and overlapping inequalities, respectively. Costa (2008) thoroughly discussed Dagum methodology focusing on the role of overlapping component and introducing simplified expressions for Dagum decomposition when a population is partitioned in two subgroups.

This work links Silber and Dagum ways of decomposition, thereby providing a decomposable matrix expression for the Gini index relying on Gini pair wise difference criterion and *G*-matrix operator. This matrix decomposition has the appealing feature of deriving the overlapping (transvariation) component as an intuitive and straightforward contribution of the inequality across subgroup income distributions. A further attractiveness of this procedure is that it provides a matrix compact form for counting the absolute, relative frequency of overlapping phenomenon.

The rest of the paper is organized as follows. Section 2 outlines Silber decomposition of Gini index. Section 3 briefly recalls Dagum decomposition approach and introduces a new matrix-based decomposition technique combining Dagum and Silber procedures. Section 4 illustrates a matrix compact form for counting transvariation cases on the basis of the matrix approach introduced in Section 3, paying attention to its possible application when dealing with the aim of measuring stratification; in Subsection 4.1, we apply the proposed decomposition procedure to sample data providing information on Italian employee incomes in 2000 and 2008. Section 5 concludes.

## 2. Silber matrix form decomposition

In the literature, the use of matrix algebra for computing the Gini index was introduced by Pyatt (1976) who proposed a matrix decomposition approach on the basis of game theory. Silber (1989) suggested another matrix form for calculating and decomposing the Gini index. Given a population of size $n$ and mean income $\mu$, let $\mathbf{y}$ stand for the $n$ by 1 vector whose elements $y_1, y_2, \ldots, y_n$ are arranged in decreasing order. Being $s_i$ the proportion of total income earned by the income receiver whose income has the $i$-th rank in vector $\mathbf{y}$, Silber formula is as follows

$$G = \mathbf{e}'\mathbf{G}\mathbf{s}, \tag{1}$$

where $\mathbf{s}$ is the column vector of the per capita income shares $s_1, s_2, \ldots, s_n$ sorted in decreasing order, $\mathbf{e}$ is an $n$ by 1 column vector with elements equal to $1/n$, and $\mathbf{G}$ is an $n$ by $n$ matrix whose elements $g_{ij}$ are equal to -1 when $j > i$, to +1 when $j < i$, to 0 when $j = i$ (called $G$-matrix). Equation (1) can be decomposed by subgroup when income receivers are grouped according to some criterion (e.g. income class, socio-demographic characteristics). Suppose the $n$ income earners are partitioned into $h$ subgroups, with $n = \sum_{k=1}^{h} n_k$. Let us introduce the ordering criterion $A$ arranging incomes by subgroup mean income in decreasing order (first key of ordering) and by per capita income in decreasing order within each subgroup (second key of ordering). Let $\mathbf{s}_A$ denote the vector of income shares arranged by the ordering $A$. By partitioning the $n$ by $n$ $G$-matrix into $h^2$ sub matrices, one obtains

$$G = \begin{bmatrix} G(n_1, n_1) & \cdots & G(n_1, n_h) \\ \vdots & \ddots & \vdots \\ G(n_h, n_1) & \cdots & G(n_h, n_h) \end{bmatrix}, \tag{2}$$

where the main diagonal block matrix $G(n_k, n_k)$ is an $n_k$ by $n_k$ $G$-matrix, whereas the $n_k$ by $n_l$ off-diagonal block matrix $G(n_k, n_l)$ has all elements equal to -1 when $k < l$, to +1 when $k > l$. Now, the within-group component of the Gini index is

$$G^W = \sum_{k=1}^{h} \mathbf{e}'(n_k)\mathbf{G}(n_k, n_k)\mathbf{s}_A(n_k), \tag{3}$$

where $\mathbf{e}(n_k)$ is an $n_k$ by 1 vector with elements equal to $1/n$ and $\mathbf{s}_A(n_k)$ is an $n_k$ by 1 vector whose elements are the elements of $\mathbf{s}_A$ belonging to subgroup $k$ arranged in decreasing order. Indeed, the between-group Gini index can be written as

$$G^B = \sum_{k=1}^{h} \sum_{l \neq k}^{h} \mathbf{e}'(n_k)\mathbf{G}(n_k, n_l)\mathbf{s}_A(n_l). \tag{4}$$

Next, subtracting expressions in (3) and (4) from the Gini expression for ungrouped data in (1), one obtains

$$
\begin{aligned}
G^{T} &= \mathbf{e'Gs} - \sum_{k=1}^{h} \mathbf{e'}(n_k)\mathbf{G}(n_k, n_k)\mathbf{s}_A(n_k) - \sum_{k=1}^{h}\sum_{l \neq k}^{h} \mathbf{e'}(n_k)\mathbf{G}(n_k, n_l)\mathbf{s}_A(n_l) \\
&= \mathbf{e'Gs} - \mathbf{e'Gs}_A,
\end{aligned} \tag{5}
$$

which represents the transvariation (overlapping) component of the Gini index. Given two subgroups of income receivers, any pair formed by two income receivers belonging to different subgroups is said to transvary if the sign of the difference between their incomes is opposite to the sign of the difference between the corresponding subgroup mean incomes. If there is at least one transvarying pair, the two subgroups are said to transvary. When dealing with income, this definition of transvariation is equivalent to the definition of overlapping. In fact, two subgroups of income receivers are said to overlap if at least one income belonging to the subgroup with lower mean income is higher than at least one income of the subgroup with higher mean income. Thus, we use, throughout the paper, transvariation and overlapping as synonymous terms.

## 3. A matrix decomposition linking Dagum and Silber approaches

It is well known that the Gini index can be expressed as a function of Gini relative mean difference (Gini, 1912)

$$
G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} \left| y_i - y_j \right|}{2\mu n^2}. \tag{6}
$$

Starting from equation (6) based on the Gini pair wise difference criterion, Dagum (1997) decomposition yields a breakdown of the Gini index into a within-group contribution to the overall inequality

$$
G^{W} = \frac{\sum_{k=1}^{h}\left( \sum_{i=1}^{n_k}\sum_{j=1}^{n_k} \left| y_{ki} - y_{kj} \right| \right)}{2\mu n^2}, \tag{7}
$$

and an across-group contribution to the total inequality due to the disparities between subgroup income distributions

$$
G^{AG} = \frac{\sum_{k=1}^{h}\sum_{l \neq k}^{h}\left( \sum_{i=1}^{n_k}\sum_{j=1}^{n_l} \left| y_{k,i} - y_{l,j} \right| \right)}{2\mu n^2}. \tag{8}
$$

In general, equations (4) and (8) are not equivalent, the former accounts for inequality between subgroup mean incomes, the latter considers all possible pair wise disparities between incomes belonging to different subgroups. The between-group and the across-group components coincide only if subgroup income distributions do not overlap (Mookherjee and Shorrocks, 1982).

Now, we combine Silber matrix technique and Dagum approach by providing a new matrix compact form for the Gini index. Let us introduce the *n* by 1 vector $\mathbf{1}_n$ with elements equal to 1, then $\mathbf{D} = \mathbf{1}_n \mathbf{y}' - \mathbf{y}\mathbf{1}_n'$ represents the skew symmetric matrix containing the $n^2$ pair wise differences between incomes of vector $\mathbf{y}$ defined in the previous section. Using the Hadamard product operator (see Vernizzi (2009) and Vernizzi *et al*. (2010) for further details on the use of Hadamard product for Gini computation purpose), denoted by $\circ$, it can be verified that

$$G = \frac{1}{2\mu} \mathbf{e}' \left( \mathbf{G} \circ \mathbf{D} \right) \mathbf{e}, \tag{9}$$

where $\mathbf{G}$ is an *n* by *n* *G*-matrix and $\mathbf{e}$ has dimension equal to *n* by 1. In (9), the entry wise product between $\mathbf{G}$ and $\mathbf{D}$ ensures that all non-positive pair wise differences in the upper diagonal entries of $\mathbf{D}$ change their sign. Thus, the matrix $\mathbf{G} \circ \mathbf{D}$ contains the $n^2$ non-negative pair wise income differences considered in (6). Since $\frac{1}{\mu n} \mathbf{y} = \mathbf{s}$, expression (9) can be written in terms of per capita income shares,

$$G = \frac{n}{2} \mathbf{e}' \left( \mathbf{G} \circ \mathbf{S} \right) \mathbf{e}. \tag{10}$$

where $\mathbf{S} = \mathbf{1}_n \mathbf{s}' - \mathbf{s}\mathbf{1}_n'$ is the skew symmetric matrix containing the $n^2$ pair wise differences between per capita income shares. Expression (10) can be re-arranged by using the Hadamard product properties yielding the following matrix compact formula (see Appendix A)

$$G = \frac{1}{2n} tr\left( \mathbf{G}\mathbf{S}' \right). \tag{11}$$

Next, we introduce the *n* by 1 vector $\mathbf{w}_k$ with non-zero elements equal to 1 in the corresponding positions filled by income shares belonging to subgroup *k* in the vector $\mathbf{s}$. Then, the square matrix $\mathbf{W} = \sum_{k=1}^{h} \mathbf{w}_k \mathbf{w}_k'$ has 1 element in the entries corresponding to pair wise differences involving two income shares of a same subgroup in $\mathbf{S}$. By inserting $\mathbf{W}$ in equation (10), one obtains

$$\begin{aligned} G^W &= \frac{n}{2} \mathbf{e}' \left( \mathbf{G} \circ \mathbf{W} \circ \mathbf{S} \right) \mathbf{e} \\ &= \frac{1}{2n} tr\left( \mathbf{G}\mathbf{S}^{W\prime} \right), \end{aligned} \tag{12}$$

where $\mathbf{S}^{W} = \mathbf{W} \circ \mathbf{S}$. Expression (12) is equivalent to equations (3) and (7), yielding the within-group contribution to the overall inequality due to the disparities between income shares belonging to a same subgroup. By setting $\mathbf{J} = \mathbf{1}_n \mathbf{1}_n'$, the square matrix $\mathbf{J} - \mathbf{W} = \sum_{k=1}^{h} \sum_{l \neq k}^{h} \mathbf{w}_k \mathbf{w}_l'$ stands for the matrix selecting pair wise differences between income shares belonging to different subgroups from $\mathbf{S}$. Thus, the expression for the across-group component of the total inequality is

$$\begin{aligned} G^{AG} &= \frac{n}{2} \mathbf{e}' \left[ \mathbf{G} \circ \left( \mathbf{J} - \mathbf{W} \right) \circ \mathbf{S} \right] \mathbf{e} \\ &= \frac{1}{2n} tr\left( \mathbf{G} \mathbf{S}^{AG}{}' \right), \end{aligned} \tag{13}$$

where $\mathbf{S}^{AG} = \left( \mathbf{J} - \mathbf{W} \right) \circ \mathbf{S}$.

Now, let $\mathbf{A}$ stand for the *n* by *n* permutation matrix re-arranging the elements of $\mathbf{s}$ in accordance with the ordering criterion *A* defined in the previous section, $\mathbf{s}_A = \mathbf{A}\mathbf{s}$. In equation (13), by replacing $\mathbf{S}$ and $\mathbf{W}$ with $\mathbf{S}_A = \mathbf{1}_n \mathbf{s}_A' - \mathbf{s}_A \mathbf{1}_n' = \mathbf{A}\mathbf{S}\mathbf{A}'$ and $\mathbf{W}_A = \mathbf{A}\mathbf{W}\mathbf{A}'$ respectively, after some algebraic manipulations one obtains (see Appendix B)

$$\begin{aligned} G^{B} &= \frac{n}{2} \mathbf{e}' \left[ \mathbf{G} \circ \left( \mathbf{J} - \mathbf{W}_A \right) \circ \mathbf{S}_A \right] \mathbf{e} \\ &= \frac{1}{2n} tr\left( \mathbf{A}'\mathbf{G}\mathbf{A} \mathbf{S}^{AG}{}' \right), \end{aligned} \tag{14}$$

which is equivalent to Silber expression for the between-group Gini component defined in (4). Then, subtracting the expression for $G^B$ in (14) from the formula (13), we have the expression for the transvariation (overlapping) component

$$G^{T} = \frac{1}{2n} tr\left[ \left( \mathbf{G} - \mathbf{A}'\mathbf{G}\mathbf{A} \right) \mathbf{S}^{AG}{}' \right] \tag{15}$$

which is equivalent to (5). The matrix $\mathbf{G} - \mathbf{A}'\mathbf{G}\mathbf{A}$ detects the transvarying pairs from the matrix $\mathbf{S}$: it has non-zero elements equal to -2 (+2) in its upper (lower) diagonal entries corresponding to transvarying pairs in $\mathbf{S}$. The matrix $\mathbf{A}'\mathbf{G}\mathbf{A}$ identifies violations of the decreasing ordering of income shares in $\mathbf{s}_A$ by showing a value opposite to that of $\mathbf{G}$ in each entry corresponding to a difference in $\mathbf{S}$ between two incomes representing a transvarying pair, while $\mathbf{A}'\mathbf{G}\mathbf{A}$ and $\mathbf{G}$ have the same elements in the entries corresponding to non-transvarying pairs. It follows immediately that subtracting $\mathbf{A}'\mathbf{G}\mathbf{A}$ from $\mathbf{G}$, the non-zero elements (equal to -2 in the upper-diagonal entries, to +2 in the lower-diagonal ones) of $\mathbf{G} - \mathbf{A}'\mathbf{G}\mathbf{A}$ fill the entries corresponding to transvarying pairs in the matrix $\mathbf{S}$. A simple numerical example is shown to provide a practical explanation of the role of $\mathbf{A'GA}$. Consider four income receivers partitioned in two subgroups each containing two individuals and let the income shares of subgroup 1 be given by 0.4 and 0.2, and those of subgroup 2 by 0.3 and 0.1. The vector of income shares sorted by decreasing order is $\mathbf{s'} = \begin{pmatrix} 0.4 & 0.3 & 0.2 & 0.1 \end{pmatrix}$ whereas the vector of income shares sorted by the ordering $A$ is $\mathbf{s}_A' = \begin{pmatrix} 0.4 & 0.2 & 0.3 & 0.1 \end{pmatrix}$. Then, the pair of income shares ($s_2$=0.3, $s_3$=0.2) represents a transvarying pair. Thus the matrix $\mathbf{A'GA}$ has a +1 element in its (2,3)the entry and a -1 in the (3,2)the entry instead of, respectively, a -1 and a +1 element as shown in $\mathbf{G}$. The remaining elements of $\mathbf{A'GA}$ coincide with the elements of $\mathbf{G}$. Then, it can easy verified that $\mathbf{G} - \mathbf{A'GA}$ is a matrix with the (2,3)the element equal to -2 and the (3,2)the one equal to +2, while the other elements are equal to zero.

In the case of one pair of identical income shares belonging to different subgroups, this pair is denoted by non-zero elements in $\mathbf{G} - \mathbf{A'GA}$ only if the income share belonging to poorer (on average) subgroup is ranked ahead of the income share belonging to richer (on average) subgroup in vector $\mathbf{s}$. Regardless of the convention adopted for ranking identical income shares belonging to different subgroups in $\mathbf{s}$, one pair of identical shares provides a null contribution to the transvariation component because the corresponding elements of $\mathbf{S}$ are equal to zero. In the following, we shall come back to the choice of the ordering criterion for ranking identical income shares of different subgroups in order to obtain matrix expressions counting transvarying pairs.

If transvariation does not exist, the vector of income shares arranged by the ordering criterion $A$ coincides with the vector of income shares sorted in decreasing order, and then $\mathbf{A} = \mathbf{I}_n$ yielding $G^T = 0$.

Given a unique matrix containing all the information on the disparities concerning the income distribution, denoted by $\mathbf{S}$, the presented decomposition technique has the appealing feature of deriving the various components of the overall inequality by separating the pair wise inequality contributions contained in $\mathbf{S}$ by means of permutation and selection matrices defined according to a population partition into exclusive and exhaustive subgroups. Starting from the matrix expression for the across-group inequality, the transvariation inequality is achieved as a clear and straightforward contribution of disparities between subgroup distributions, while the between-group inequality is calculated without assuming that all income receivers of a subgroup have the same income. As noted in Ebert (2010), $G^{AG}$ and $G^B$ represent two extreme ways of measuring disparity between subgroup distributions, one accounting for inequality between all pairs of incomes belonging to different subgroups ($G^{AG}$), the other reducing disparity between subgroups to inequality between subgroup average incomes ($G^B$). Utilizing the full information available, the proposed decomposition technique reconciles these extreme possibilities of determining inequality between subgroup distributions by showing that both $G^{AG}$ and $G^{B}$ can be derived from the same standpoint, one matrix containing full information, and without employing average incomes.

## 4. A new matrix expression for counting transvariation and its application

In this section, we show that the proposed technique can be used to derive a convenient matrix form for counting the absolute, relative frequency of transvarying pairs.

By considering per capita income shares, the total number of transvarying pairs can be expressed as follows

$$T = \sum_{k=2}^{h} \sum_{l=1}^{k-1} \left( \sum_{i=1}^{n_k} \sum_{j=1}^{n_l} \eta\left(s_{ki}, s_{lj}\right) \right), \tag{16}$$

where

$$\eta\left(s_{ki}, s_{lj}\right) = \begin{cases} 1 & if \quad \left(s_{ki} - s_{lj}\right)\left(\overline{s}_k - \overline{s}_l\right) < 0 \\ 0 & if \quad \left(s_{ki} - s_{lj}\right)\left(\overline{s}_k - \overline{s}_l\right) > 0 \\ \dfrac{1}{2} & if \qquad s_{ki} = s_{lj} \end{cases},$$

with the adopted convention of counting one half each pair having identical members (Calò, 2006) and the mean income share of subgroup $k$ denoted by $\overline{s}_k$. In order to rewrite equation (16) in matrix compact form we introduce the following notation. Let $\mathbf{s}_U$ stand for the vector of income shares arranged in decreasing order, where the identical income shares are sorted by subgroup mean in decreasing order. Let $\mathbf{s}_L$ stand for the vector of income shares arranged in decreasing order, where the identical income shares are sorted by subgroup mean in increasing order. Then, $\mathbf{A}_U$ and $\mathbf{A}_L$ denote the permutation matrices satisfying the equalities $\mathbf{s}_A = \mathbf{A}\mathbf{s} = \mathbf{A}_U\mathbf{s}_U = \mathbf{A}_L\mathbf{s}_L$. It is worth emphasizing that by replacing $\mathbf{A}$ in (15) with $\mathbf{A}_U$ or $\mathbf{A}_L$, one has expressions equivalent to (15) as the ordering of elements in $\mathbf{s}$ does not pay attention to the reciprocal positions filled by identical income shares. However, the attempt of counting transvarying pairs in matrix form needs to specify the ordering criterion for identical income shares when sorting income shares in decreasing order.

By using $\mathbf{A}_U$, $\mathbf{A}_L$ and $\mathbf{G}$, a compact matrix form for counting the number of transvarying pairs can be written as (see appendix C)

$$T = \frac{1}{4} tr\left[\mathbf{GG'} - \mathbf{A}_U\mathbf{'GA}_U\mathbf{G'}\right] + \frac{1}{8} tr\left[\mathbf{A}_U\mathbf{'GA}_U\mathbf{G'} - \mathbf{A}_L\mathbf{'GA}_L\mathbf{G'}\right]$$
$$= T^A + T^C. \qquad (17)$$

In (17), $T^A$ represents the number of pairs whose members actually transvary, while $T^C$ accounts for pairs having identical members (conventional transvariation). Expression (17) can be re-arranged without the separation between actual and conventional transvariations, yielding

$$T = \frac{n(n-1)}{4} - \frac{1}{8} tr\left[\mathbf{A}_U\mathbf{'GA}_U\mathbf{G'} + \mathbf{A}_L\mathbf{'GA}_L\mathbf{G'}\right]. \qquad (18)$$

Given a population of size *n*, the total number of combinations by taking two members at a time out of population is equal to $n(n-1)/2$. Then, by dividing *T* by $n(n-1)/2$ one obtains the ratio between the number of transvarying pairs and the number of all possible pairs of income shares,

$$PT = \frac{1}{2} - \frac{1}{4n(n-1)} tr\left[\mathbf{A}_U\mathbf{'GA}_U\mathbf{G'} + \mathbf{A}_L\mathbf{'GA}_L\mathbf{G'}\right] . \qquad (19)$$

The presented matrix approach enables us to achieve the transvariation component as an intuitive and precise contribution to the overall inequality and, at the same time, to compute the absolute, relative frequency of the transvariation phenomenon. This can represent an advantage when dealing with the measurement of subgroup separation, such as the definition of stratification indices that requires an in-depth analysis of overlapping. As pointed out in the literature (Yitzhaki and Lerman, 1991; Yitzhaki, 1994; Frick *et al*., 2006), overlapping can be seen as the inverse of stratification. Since a raise in overlapping diminishes the separation of subgroup distributions, the greater the degree of overlapping between subgroup distributions the greater the distance from the perfect stratification between subgroup distributions. By interpreting overlapping as non-stratification between population subgroups, we notice that, while the overlapping concept is linked to between-group inequality (it is defined with respect to inequality between subgroup average incomes), the well known between-group inequality measure only accounts for differences in subgroup average incomes, neglecting overlapping degree between subgroup distributions. Recent studies (Liao, 2008; Monti and Santoro, 2009) argued that stratification measurement is closely related to between-group inequality measurement. Monti and Santoro (2009) provided a Gini decomposition in which the between-group inequality component captures changes in stratification degree between two population subgroups by counting the number of transvarying pairs. Starting from Dagum decomposition of the Gini index, Liao (2008) developed stratification indices based on the between-group inequality measure and the frequency of pair wise income comparisons involving incomes of different subgroups.

For instance, we consider the following stratification index proposed by Liao (2008) and defined in absence of transvariation,

$$I_3 = I_1 I_2 \tag{20}$$

where

$$I_1 = \frac{G - G^W}{G}$$

and

$$I_2 = \frac{\sum_{k=2}^{h} \sum_{l=1}^{k-1} \left( \sum_{i=1}^{n_k} \sum_{j=1}^{n_l} d \right)}{\left(n^2 - n\right)/2},$$

where *d* equals 1. The index $I_3$ is compounded by two other stratification indices, one accounting for the proportion of total inequality ascribed to between-group inequality, denoted by $I_1$, another reckoning the proportion of the number of pair wise comparisons between members of different subgroups, indicated by $I_2$. By means of the index $I_3$, the proportion of the Gini index due to between-group inequality is made relative to the proportion of the pair wise comparisons between members of different population subgroups. $I_3$ varies from 0, where no stratification exists, to 1, where there is full stratification. In the real world, no overlapping subgroup income ranges are seldom observed, and then we attempt to extend the above considered stratification index to the case of overlapping subgroups. In order to take into account the presence of overlapping (non-stratification), we modify the indices $I_1$ and $I_2$ as follows

$$I_1^* = I_1 - \frac{G^T}{G} \tag{21}$$

and

$$I_2^* = I_2 - PT\,; \tag{22}$$

thus, we have

$$I_3^* = I_1^* I_2^*\,. \tag{23}$$

The index $I_3^*$ ranges between 0 (no stratification) and 1 (full stratification), and it coincides with $I_3$ when dealing with no overlapping subgroup distributions. In (23) the proportion of the Gini ascribed to between-group inequality (measured by $I_1^*$) is made relative to the proportion of pair wise income differences between members belonging to different subgroups that actually contributes to subgroup separation, which is obtained by subtracting the proportion of transvarying pairs from the proportion of pair wise differences between members of different subgroups (index $I_2^*$).

### 4.1. Empirical application

We provide an empirical application of the methodology developed in this paper by using income data collected by the Survey of Household Income and Wealth (henceforth, SHIW) of the Bank of Italy in 2000 and 2008 (Banca d'Italia, 2010). By considering individual data concerning employee incomes, we decompose the Gini inequality index by gender. Furthermore, we apply the modified index of stratification ($I_3^*$) in order to measure the stratification between subgroups. As noticed in the literature (Sørensen, 1990; Monti and Santoro, 2009), gender discrimination on the labour market often implies that men earn more than women. This fact can be seen as an example of stratification, where income inequality between subgroups defined by gender represents an aspect linked to the existence of different strata. Since stratification is a phenomenon changing over time, it seems interesting to measure stratification variation over a given period (in this case the 2000–2008 period).

Both surveys were carried out by considering household as sample unit, so sample weights are referred to households. Even if household is the basic statistical unit, household weight can be used for each household member when considering individual incomes. In the performed analysis we applied weights provided by the central bank of Italy for each household. After deleting non-positive incomes, the samples are composed of 5911 individual incomes in 2008 and 6439 incomes in 2000. Incomes are expressed in Lire in the year 2000, whereas incomes referred to 2008 are expressed in Euro.

Tables 1 reports the results of the Gini index decomposition, the stratification indices, and some descriptive statistics for the 2000-2008 period. As expected, men average income is higher than women one in both 2000 and 2008. From Table 1, it emerges that overall inequality decreases between 2000 and 2008. This decrease is mainly attributable to the lower within-group inequality shown in 2008. By looking at the stratification index $I_3^*$ in Table 1, we observe that the year 2000 shows an index $I_3^*$ of 0.1111 while the year 2008

**Table 1**. Gini decomposition, stratification indices and descriptive statistics: Italy 2000–2008.

| Sample | 2008 | 2000 |
|---|---|---|
| $G$ | 0.2573898 | 0.2663702 |
| $G^W$ | 0.1286575 | 0.1368498 |
| $G^{AG}$ | 0.1287323 | 0.1295204 |
| $G^B$ | 0.07700322 | 0.09582038 |
| $G^T$ | 0.05172908 | 0.03370006 |
| $I_1^*$ | 0.2991697 | 0.3597264 |
| $I_2^*$ | 0.3172425 | 0.3089238 |
| $I_3^*$ | 0.0949034 | 0.1111280 |
| N. of observations | 5911 (women=2000, men=3911) | 6439 (women=2598, men=3841) |
| Women av. income | 14205.52 | 21594.68* |
| Men av. income | 18031.14 | 27750* |
| Av. income | 16373.34 | 25277.12* |

** Average income is expressed in thousands of Lire.*

*Source: elaborations on SHIW data.*

exhibits an index $I_3^*$ of 0.0949, suggesting that both the situations are far from full stratification. By comparing the index $I_3^*$ of 2000 with the index $I_3^*$ of 2008, we observe that the stratification decreased between 2000 and 2008. This decrease can be ascribed to the ascent of transvariation inequality in 2008. Even if the proportion of across-group pairwise differences contributing to subgroup separation is higher in 2008 (measured by $I_2^*$), its effect is counterbalanced by the fall of the between-group inequality due to the raise in $G^T$, as noted above. Thus, the intensity of transvariation, in terms of inequality, is more important than its relative frequency when we isolate the determinants of stratification variation moving from 2000 to 2008. Moreover, it is worth emphasizing that our empirical findings suggest that an increase in transvariation inequality component does not lead to a major role of within-group component on total inequality. In fact, the within-group component decreased between 2000 and 2008 whereas the overlapping component increased, underlining that a raise of the extent to which subgroup distributions overlap does not imply an increase of the inequality within subgroups.

## 5. Conclusions

The contribution of this article is two-fold. First, it has provided a matrix form decomposition of the Gini index which combines Silber matrix approach and Dagum decomposition based on the Gini pairwise difference criterion, thereby enabling to derive the within-group, the between-group and the transvariation components of total inequality by exploiting all of the information contained in one matrix accounting for disparities between per capita income shares of the overall population. Second, this matrix approach yields the quantification of transvariation in terms of inequality intensity and frequency of the event occurrence. This implies that the suggested matrix approach can be useful not only for decomposing inequality but also measuring phenomena closely linked with transvariation, such as stratification.

In conclusion, the proposed matrix approach offers a complete tool for the Gini decomposition which can be applied when dealing with no overlapping as well as overlapping subgroups.

### Appendix A

This appendix proves that

$$G = \frac{n}{2}\mathbf{e}'\left(\mathbf{G} \circ \mathbf{S}\right)\mathbf{e} = \frac{1}{2n}tr\left(\mathbf{GS}'\right). \tag{A1}$$

We recall some useful properties of the Hadamard product. Given the $n$ by $n$ matrices $\mathbf{O}$, $\mathbf{P}$, $\mathbf{Q}$, where $\mathbf{O}$ is a permutation matrix, the $n$ by 1 vector $\boldsymbol{\lambda}$ and the diagonal matrix $\boldsymbol{\Lambda}$ whose diagonal elements are the elements of $\boldsymbol{\lambda}$, the following properties hold:

*i)* $\mathbf{O}'\left(\mathbf{P} \circ \mathbf{Q}\right)\mathbf{O} = \left(\mathbf{O}'\mathbf{P}\mathbf{O}\right) \circ \left(\mathbf{O}'\mathbf{Q}\mathbf{O}\right)$ (see Faliva, 1996, pp. 157);

*ii)* $\boldsymbol{\lambda}'\left(\mathbf{P} \circ \mathbf{Q}\right)\boldsymbol{\lambda} = tr\left(\mathbf{P}\boldsymbol{\Lambda}\mathbf{Q}'\boldsymbol{\Lambda}\right)$ (see Abadir and Magnus, 2005, pp. 340).

By using the property *ii*), we have

$$G = \frac{n}{2}tr\left(\mathbf{GES}'\mathbf{E}\right), \tag{A2}$$

where $\mathbf{E}$ is a diagonal matrix with diagonal elements equal to *1/n*. Now, given that $\mathbf{E} = \frac{1}{n}\mathbf{I}_n$, it follows

$$\begin{aligned} G &= \frac{n}{2n^2}tr\left(\mathbf{GI}_n\mathbf{S}'\mathbf{I}_n\right) \\ &= \frac{1}{2n}tr\left(\mathbf{GS}'\right). \end{aligned} \tag{A3}$$

## Appendix B

This appendix proves that

$$G^B = \frac{n}{2}\mathbf{e}'\left[\mathbf{G}\circ\left(\mathbf{J}-\mathbf{W}_A\right)\circ\mathbf{S}_A\right]\mathbf{e} = \frac{1}{2n}tr\left(\mathbf{A}'\mathbf{GAS}^{AG}{}'\right). \tag{B1}$$

Let recall the ordering criterion $A$ defined in Section 2, for which incomes are sorted by decreasing subgroup mean income (first key) and, within each subgroup, in decreasing order (second key). Supposing that incomes are sorted by the ordering $A$, the well-known expression for the between-group inequality (Mookherjee and Shorrocks, 1982) can be written as

$$\begin{aligned} G^B &= \frac{1}{2n^2\mu}\sum_{k=1}^{h}\sum_{l\neq k}^{h} n_k n_l \left|\mu_k - \mu_l\right| \\ &= \frac{1}{2n^2\mu}\sum_{k=1}^{h}\sum_{l=k+1}^{h} 2n_k n_l \left(\mu_k - \mu_l\right) \\ &= \frac{1}{2n^2\mu}\sum_{k=1}^{h}\sum_{l=k+1}^{h}\sum_{i=1}^{n_k}\sum_{j=1}^{n_l} 2\left(y_{ki} - y_{lj}\right) \\ &= \frac{1}{2n}\sum_{k=1}^{h}\sum_{l=k+1}^{h}\sum_{i=1}^{n_k}\sum_{j=1}^{n_l} 2\left(s_{ki} - s_{lj}\right), \end{aligned} \tag{B2}$$

where $G^B$ is expressed as a function of the across-group pairwise share differences. Being $\mathbf{A}$ the permutation matrix re-arranging the element of $\mathbf{s}$ in accordance with the ordering $A$, it follows that $\mathbf{s}_A = \mathbf{As}$ is the vector of income shares sorted by the ordering $A$ and $\mathbf{w}_{Ak} = \mathbf{Aw}_k$ denotes the vector with non-zero elements equal to 1 in the positions filled by income shares of subgroup $k$ in $\mathbf{s}_A$. Thus, the matrices $\mathbf{S}_A = \mathbf{1}_n\mathbf{s}_A' - \mathbf{s}_A\mathbf{1}_n'$ and $\mathbf{W}_A = \sum_{k=1}^{h}\mathbf{w}_{Ak}\mathbf{w}_{Ak}'$ are, respectively, the matrix containing pairwise differences obtained from shares sorted by the ordering $A$ and the matrix selecting within-group pairwise differences from $\mathbf{S}_A$. Then, the matrix compact form accounting for across-group pairwise differences in $\mathbf{S}_A$ is equivalent to the between-group inequality component as expressed in (B2),

$$G^B = \frac{n}{2}\mathbf{e}'\left[\mathbf{G}\circ\left(\mathbf{J}-\mathbf{W}_A\right)\circ\mathbf{S}_A\right]\mathbf{e}. \tag{B3}$$

Then, since $\mathbf{J}-\mathbf{W}_A = \mathbf{A}\left(\mathbf{J}-\mathbf{W}\right)\mathbf{A}'$ and $\mathbf{S}_A = \mathbf{ASA}'$, using the property *i*) in Appendix A, expression (B3) can be re-written as

$$\begin{aligned} G^B &= \frac{n}{2}\mathbf{e}'\left\{\mathbf{G}\circ\mathbf{A}\left[\left(\mathbf{J}-\mathbf{W}\right)\circ\mathbf{S}\right]\mathbf{A}'\right\}\mathbf{e} \\ &= \frac{n}{2}\mathbf{e}'\left\{\mathbf{G}\circ\mathbf{A}\mathbf{S}^{AG}\mathbf{A}'\right\}\mathbf{e}. \end{aligned} \tag{B4}$$

We now use the result obtained in Appendix A, so that expression (B4) becomes

$$\begin{aligned} G^B &= \frac{1}{2n}tr\left(\mathbf{GAS}^{AG}{}'\mathbf{A}'\right) \\ &= \frac{1}{2n}tr\left(\mathbf{A}'\mathbf{GAS}^{AG}{}'\right) \end{aligned}.$$

## Appendix C

This appendix proves that

$$\begin{aligned} T &= \frac{1}{4}tr\left(\mathbf{GG}'-\mathbf{A}_U{}'\mathbf{GA}_U\mathbf{G}'\right)+\frac{1}{8}tr\left(\mathbf{A}_U{}'\mathbf{GA}_U\mathbf{G}'-\mathbf{A}_L{}'\mathbf{GA}_L\mathbf{G}'\right) \\ &= T^A+T^C. \end{aligned} \tag{C1}$$

The matrix $\mathbf{G}-\mathbf{A}'\mathbf{GA}$ entered in (15) detects transvarying pairs by identifying, with non-zero elements ($\pm 2$), violations of decreasing ordering of income shares when moving from $\mathbf{s}$ to $\mathbf{s}_A$. However, a further ordering criterion has to be added in order to detect pairs with identical members (income shares) belonging to different subgroups. By assuming identical shares are sorted by subgroup mean in decreasing order in $\mathbf{s}$, the permutation matrix $\mathbf{A}_U$ re-arranges them in $\mathbf{s}_A$ maintaining their reciprocal positions; then only actual transvarying pairs are detected by $\mathbf{G}-\mathbf{A}_U{}'\mathbf{GA}_U$, the number of which can be expressed as follows

$$\begin{aligned} T^A &= \frac{1}{4}\mathbf{1}_n{}'\left[\left(\mathbf{G}-\mathbf{A}_U{}'\mathbf{GA}_U\right)\circ\mathbf{G}\right]\mathbf{1}_n \\ &= \frac{1}{4}tr\left[\mathbf{GG}'-\mathbf{A}_U{}'\mathbf{GA}_U\mathbf{G}'\right]. \end{aligned} \tag{C2}$$

where $\mathbf{1}_n$ is an *n* by 1 vector with elements equal to 1. By sorting identical income shares by subgroup mean in increasing order in $\mathbf{s}$, their reciprocal positions in $\mathbf{s}_A$ are exchanged by means of the permutation matrix $\mathbf{A}_L$. Thus, the matrix $\mathbf{G}-\mathbf{A}_L{}'\mathbf{GA}_L$ detects pairs having identical members as transvarying pairs (conventional transvarying pairs), plus the actual transvarying pairs. In order to quantify the number of conventional transvarying pairs, one can subtract the number of actual transvarying pairs expressed in (C2) from the number of transvarying pairs obtained by replacing $\mathbf{G}-\mathbf{A}_U\mathbf{'GA}_U$ in (C2) with $\mathbf{G}-\mathbf{A}_L\mathbf{'GA}_L$ accounting for actual as well as conventional transvarying pairs. Therefore, transvarying pairs with identical members can be counted in matrix formula as follows

$$\begin{aligned} T^C &= \frac{1}{2}\left\{\frac{1}{4}\mathbf{1}_n\mathbf{'}\left[\left(\mathbf{G}-\mathbf{A}_L\mathbf{'GA}_L\right)\circ\mathbf{G}\right]\mathbf{1}_n - \frac{1}{4}\mathbf{1}_n\mathbf{'}\left[\left(\mathbf{G}-\mathbf{A}_U\mathbf{'GA}_U\right)\circ\mathbf{G}\right]\mathbf{1}_n\right\} \\ &= \frac{1}{8}\mathbf{1}_n\mathbf{'}\left\{\left[\left(\mathbf{G}-\mathbf{A}_L\mathbf{'GA}_L\right)-\left(\mathbf{G}-\mathbf{A}_U\mathbf{'GA}_U\right)\right]\circ\mathbf{G}\right\}\mathbf{1}_n \\ &= \frac{1}{8}\mathbf{1}_n\mathbf{'}\left[\left(\mathbf{A}_U\mathbf{'GA}_U-\mathbf{A}_L\mathbf{'GA}_L\right)\circ\mathbf{G}\right]\mathbf{1}_n, \end{aligned} \tag{C3}$$

where the denominator in (C3) is 8 instead of 4 since we adopt the convention of counting one half each transvarying pair with identical members, as said in the main text. Next, using the result obtained in Appendix A, we can re-arrange expression (C3) as follows

$$T^C = \frac{1}{8}tr\left(\mathbf{A}_U\mathbf{'GA}_U\mathbf{G'}-\mathbf{A}_L\mathbf{'GA}_L\mathbf{G'}\right).$$

## REFERENCES


ABADIR, K. M., MAGNUS, J. R., (2005), Matrix Algebra. Cambridge University Press, New York.

BANCA D'ITALIA, (2010), Survey on Households Income and Wealth. Supplements to the Statistical Bulletin – sample survey in 2008, sample survey in 2000. Available at http://www.bancaditalia.it/statistiche/indcamp/bilfait/boll_stat;internal&action=_setlanguage.action?LANGUAGE=en

BERREBI, Z. M., SILBER J., (1987), Regional Differences and the components of Growth and Inequality Change. *Economics Letters*, **25**, 295–298.

BHATTACHARYA, N., MAHALANOBIS, B., (1967), Regional disparity in household consumption in India. *American Statistical Association Journal*, March, 143–162.

CALÒ, D. G., (2006), On a Transvariation Based Measure of Group Separability. *Journal of Classification*, **23**, 143–167.

COSTA, M., (2008), Gini Index Decomposition for the Case of Two Subgroups. *Communications in Statistics – Simulation and Computation*, **37**(4), 631–644.

DAGUM, C., (1987), Measuring the Economic Affluence between Populations of Income Receivers. *Journal of Business & Economic Statistics*, **5**(1), 5–12.

DAGUM, C., (1997), A new approach to the decomposition of Gini income inequality ratio. *Empirical Economics*, **22**, 515–531.

EBERT, U., (2010), The decomposition of inequality reconsidered: Weakly decomposable measures. *Mathematical Social Sciences*, **60**, 94–103.

FALIVA, M., (1996), Hadamard matrix product, graph and system theories: motivations and role in Econometrics. In: Camiz, S., Stefani, S. (eds.) Matrices and Graphs, Theory and Applications to Economics, pp 152–175. World Scientific, London.

FRICK, J., GOEBEL, J., SCHECHTMAN, E., WAGNER, G., YITZHAKI, S., (2006), Using Analysis of Gini (ANOGI) for Detecting Whether Two Subsamples Represent the Same Universe: The German Socio-Economic Panel Study (SOEP) Experience. *Sociological Methods & Research*, **34**, 427–468.

GINI, C., (1912), Variabilità e Mutabilità. Bologna: Tipografia Paolo Cuppini.

GIORGI, G. M., (1999), Income Inequality Measurement: the statistical approach. In: Silber, J. (eds.) Handbook on Income Inequality Measurement, pp 245–267. Kluwer, Boston.

LIAO, T. F., (2008), The Gini unbound: analyzing class inequality with model-based clustering. In: Betti, G., Lemmi, A. (eds.) Advances on income inequality and concentration measures, pp 201-221. Routledge, New York.

MONTI, M. G., (2008), A note on the residual term R in the decomposition of the Gini Index. Argumenta Oeconomica, **20**, 107–138.

MONTI, M. G., SANTORO, A., (2009), A note on between-groups inequality with an application to genders. Working Paper Series Econpubblica CRPS, n. 135 October available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1338054

MOOKHERJEE, D., SHORROCKS, A.F., (1982), A decomposition analysis of the trend in UK income inequality. *Economic Journal*, **92**, 886–902.

MUSSARD, S., (2004), The Bidimensional Decomposition of the Gini Ratio. A case Study: Italy. *Applied Economics Letters*, **11**(8), 503–505.

PYATT, G., (1976), On the interpretation and disaggregation of Gini Coefficient. *The Economic Journal*, **86**, 243–255.

SØRENSEN, J., (1990), Perception of women's opportunity in five industrialized nations. *European Sociological Review*, **6**(2), 151–164.

SHORROCKS, A.F., (1984), Inequality decomposition by population subgroups. *Econometrica*, **52**, 1369–1385.

SILBER, J., (1989), Factor Components, Population Subgroups and the Computation of the Gini Index of Inequality. *Review of Economics and Statistics*, **71**, 107–115.

VERNIZZI, A., (2009), Applying the Hadamard product to decompose Gini, concentration, redistribution and re-ranking indexes. *Statistics in Transition*, **10**(3), 505–525.

VERNIZZI, A., MONTI, M.G., MUSSINI, M., (2010), A Gini and Concentration Index Decomposition with an Application to the APK Reranking Measure. DEAS Working Paper n. 2010-10, Available at http://www.economia.unimi.it/uploads/wp/DEAS-2010_10wp.pdf

YITZHAKI, S., LERMAN, R., (1991), Income stratification and income inequality. *Review of Income and Wealth*, **37**(3), 313–329.

YITZHAKI, S., (1994), Economic distance and overlapping distributions. *Journal of Econometrics*, **61**, 147–159.