ARGUMENTA OECONOMICA
No 1 (20) 2008
PL ISSN 1233-5835

*Maria Monti**

A NOTE ON THE RESIDUAL TERM *R* IN THE DECOMPOSITION OF THE GINI INDEX

In this paper we reconsider the Dagum decomposition of the Gini index and we compare this decomposition with the decompositions proposed by Mookherjee and Shorrock and by Lambert and Aronson. In so doing, a deeper insight into the meaning of the overlapping term is given and an alternative expression for this term is obtained. The results are applied to a sub-sample of the C.S.O. survey: the families taken into consideration are selected and grouped according to the number of their components. The aim is to analyze how the present Polish personal income tax modifies the overall income distribution and to evaluate the changes in the within and between groups income inequalities.

Keywords: Gini coefficient, inequality decomposition, Gini residual
JEL: D63

INTRODUCTION

Since the first years of the last century in the analysis of income inequality, two particular instruments have played a very important role: the Lorenz curve and the Gini coefficient. They are particularly useful when income inequality has to be analyzed considering population's homogeneity with respect to other individual features. However, in the analysis of income inequality, it may be relevant to analyse the quantitative significance of income variations associated with other socio-economic characteristics of individuals such as age, sex, occupation, composition of their household, ethnic groups and so on. Overall inequality has to be attributed to population groups and to their properties. In this case, a decomposable inequality measure has to be used allowing separating the within-group inequality from the between-group inequality. Such decomposition may be used either to better understand economic inequality or to guide the design of economic policy. If the adopted inequality measure is additive decomposable, overall inequality is equal to the sum of within and between groups inequality.

* Department of Economics, University of Milan

Theil was a pioneer in proposing a decomposable inequality measure. In 1967, applying the entropy law, he decomposed the total inequality into the sum of a within inequality and a between subpopulations inequality. This decomposition may be simply expressed saying that to measure the inequality between groups, inequality within groups has to be neglected. Therefore, one smooths the income distributions of each group so that each member has the same income and one applies the inequality measure to the smoothed income distribution. In the same year (1967), Bhattacharya and Mahalanobis provided a decomposition of Gini index.

The two works stimulated further researches. For the sake of simplicity, we can divide these studies into two groups, the former deals with the more general argument of the class of decomposable indexes, and the latter specifically concerns the Gini coefficient decomposition. In the literature on the additive decomposable indexes, following Theil's first suggestion, the between group inequality is generally based on the (fictious) assumption that each individual receives the mean income of his own group. For this reason, the indexes belonging to this class are often called μ-decomposable. Among the contributions on this argument, one can quote Bourguignon (1979), Cowell (1980), Shorrock (1980, 1984), Russel (1985). In 1999, Ebert proposed, following Blockorby, Donaldson, and Auersperg (1981), a new single parameter family of decomposable inequality measures. These measures are based on the so-called normative approach to inequality measure and in Ebert's (1999) decomposition, a representative income, related to the welfare level attained by each group, is used instead of the group mean income.

Following Bahattacharya and Mahalanobis, the Gini decomposition was explored by Rao (1969), Pyatt (1976), Mookherjee and Shorrocks (1982), Silber (1989), Yitzhaki and Lernan (1991), Lambert and Aronson (1993), Ytzhaki (1994).

Pyatt (1976) and Silber's (1989) analysis rest upon matrix algebra, those of Bhattacharya and Mahalanobis (1967) and Mookherjee and Shorrocks (1982) are combinatoric, whereas Lambert and Aronson (1993) follow a geometrical approach. Yitzhaki and Lernan (1991) and Ytzhaki (1994) developed a pseudo-Gini coefficient, which mimics the Gini proper coefficient. In Giorgi (1990, 1993) can be found detailed background material and an interesting history of Gini decomposition. The Gini decompositions proposed in the above cited articles are μ-decompositions, but, as is well known, when the Gini coefficient is decomposed into the within and the between-groups inequality indexes a residual term arises if the group ranges overlap.

Two questions come up when dealing with the Gini decomposition. The first question holds in all μ-decomposition and it concerns with measuring income inequality between subpopulations by their means only. This is an oversimplification because in so doing the different variances and asymmetries of income distributions are ignored. The second question is the overlapping term. In this note it has been shown how the Dagum decomposition of the Gini index gives an important contribution to the overlapping term understanding. Moreover, and this is the most important result here obtained, we derive from the Dagum decomposition an alternative way to calculate the overlapping term which allows to decompose this term as a weighted sum of the overlapping terms calculated between each pair of groups. Using the proposed expression (19), inequality variations may be analyzed on considering either the whole population or referring to each groups.

The paper is organized as follows. In the second section, we present two important decompositions of the Gini index. These decompositions were proposed respectively by Mookherjee and Shorrock (1982) and by Lambert and Aronson (1993). Dagum found his decomposition on the relative economic affluence (REA) concept. In the same section, both the Dagum decomposition and REA are defined. In the third section, the three decompositions are compared showing that the overlapping term may be calculated using the expression proposed by Dagum to measure the contribution to the total inequality given by transvariation. On comparing the three decompositions, we suggest two important remarks at least. Firstly, one can observe that Dagum does not obtain the between groups inequality index starting from the hypothesis of equidistributed income groups. Then, it has to be noted that the Dagum decomposition shows clearly how the overlapping term is connected both with between groups and within group inequality. In the same section three, the expression used by Dagum to calculate the overlapping term is simplified, and the alternative expression (19) is proposed to calculate this term.

In section four, the suggested Gini index decomposition is used to analyse a particular source of inequality changes in income distributions, that is the taxation. We consider a Polish population subset composed by families with a different number of members and we try to evaluate the effects that the present Polish tax system has on the income inequality with reference to the whole population and to each group of families. The tax system is applied on the data collected in 2001 by the Polish Central Statistical Office, Household Budget Survey.

1. GINI INDEX DECOMPOSITIONS AND DAGUM DECOMPOSITION

Let us consider a population of n individuals. Let y_i be the income of individual i and $\mu = \sum_{i=1}^{n} y_i / n$ the population mean income. The overall Gini inequality index is defined as

$$G = \frac{1}{2n^2\mu} \sum_{j=1}^{n} \sum_{h=1}^{n} |y_j - y_h| = \frac{\Delta}{2\mu}$$

where Δ is the Gini mean difference, i.e. the mean of the absolute value of the income difference between $n \times n$ binary combinations of economic "units" belonging to the overall population.

Considering another population characteristic, different from income, we now partition the n individuals into k groups of sizes n_j (j=1, 2,..., k), with $\sum_{j=1}^{k} n_j = n$, defining $\mu_j = \sum_{i=1}^{n_j} y_{ij} / n_j$ i.e. the average income for the j^{th} group. The Gini index within each group is defined as

$$G_j = \frac{1}{2n_j^2\mu_j} \sum_{i=1}^{n_j} \sum_{h=1}^{n_j} |y_i - y_h| = \frac{\Delta_j}{2\mu_j}.$$

Mookherjee and Shorrock (M.-S., hereafter) (1982) show that the Gini index can be written as

$$G = \sum_j p_j^2 \lambda_j G_j + \frac{1}{2} \sum_j \sum_h p_j p_h |\lambda_j - \lambda_h| + R \tag{1}$$

where $p_j = n_j / n$ and $\lambda_j = \mu_j / \mu$.

Denoting

$$G_W = \sum_j p_j^2 \lambda_j G_j$$

and

$$G_B = \frac{1}{2}\sum_j \sum_h p_j p_h \left|\lambda_j - \lambda_h\right|$$

expression (1) rewrites as

$G = G_W + G_B + R$.

Remarking that this form of decomposition is due to Bhattacharya and Mahalanobis (1967), Rao (1969) and Pyatt (1976), M.-S. observe that G_W is some kind of average of inequality values within each group. The term G_B corresponds to the value of the Gini index replacing the incomes of all individuals with the mean income of the group to which they belong. Being individuals considered different with respect to income and to another characteristic, M-S maintain that this term evaluates inequality in average incomes due only to the different values assumed by the characteristic used to form groups. The term R is defined as "interaction effect" among groups. The authors point out that this depends upon the frequency and magnitude of overlaps between the incomes in different groups. They remark that it is impossible to interpret R with any precision except to say that this term is the residual necessary to preserve the identity. Furthermore, they underline that the way in which it reacts to changes in the group characteristics is so obscure that it can cause the overall Gini value to respond perversely to such changes. One has to note that M.-S. (1982) do not give a specific formula to compute the value of R: this term is really a residual term calculated by difference between the Gini index and its two first components.

In 1993, Lambert and Aronson (L.-A., hereafter) reconsider the Gini index decomposition suggesting a simple geometric approach interpreting all three components of overall inequality Gini index directly and explicitly in terms of areas on the Lorenz diagram. On considering a population partitioned into k groups, in a Lorenz diagram the line of perfect equality and the Lorenz curve $L(p)$ are drawn. Between them, two particular concentration curves $L_B(p)$ and $C(p)$ are considered. In so doing, the area delimitated by the perfect equality line and by the Lorenz curve, results partitioned into three different components (A_W, A_B, A_0) whose values are calculated using the formula of the Gini index expressed in area terms [1]. To construct the first concentration curve $L_B(p)$, the groups are lined up according to the non-decreasing order of their means. The total income of each group is then redistributed in such a way that each person of the group

---

[1] $G = 2\int_0^1 [p - L(p)]dp = 1 - 2\int_0^1 L(p)\,dp$

gets the mean income μ_i. In the second concentration curve $C(p)$ the groups are lined up as in the first concentration curve, but the units are ordered within each group following the ascending order of their incomes. One has to note that, using this lexicographic order, the richest person of a group finds himself standing next to the poorest person in the following group. Starting from the perfect equality line, the first concentration curve records the introduction of between group inequality, the second takes into account the between and within inequality neglecting overlapping and the Lorenz curve is used to represent the total inequality. Then evaluating inequality in area terms, one has

$$G=2\int_0^1\left[p-L(p)\right]dp=2\int_0^1\left[p-L_B(p)\right]dp+2\int_0^1\left[L_B(p)-C(p)\right]dp+2\int_0^1\left[C(p)-L(p)\right]dp.$$

L.-A. show the following correspondences

$$A_B=2\int_0^1\left[p-L_B(p)\right]dp=G_B;$$

$$A_W=2\int_0^1\left[L_B(p)-C(p)\right]dp=G_W;$$

$$A_0=2\int_0^1\left[C(p)-L(p)\right]dp=R. \qquad (2)$$

They point out that the overlapping term is at once a between groups and within groups effect measuring a between groups phenomenon, the overlapping, that is generated by inequality within groups. It is interesting to note that notwithstanding the new interpretation of the Gini index components that decompositions (2) brings, L.-A. think that the Gini index is not "rehabilitated for use in analysing source of inequality change, e.g. through time or as result of a change in tax policy…The generalised entropy measures, which disaggregate into solely between group and within group components, have been purpose designed to facilitate such analysis." (L.-A., 1993, p. 1225).

In the Gini index decomposition, the overlapping term can be very significant but what is known about it is very little. It seems clear that L.-A. think that the presence of this term in the Gini decomposition does not permit an effective evaluation of inequality changes.

1.1. Dagum decomposition

Within the framework of the Italian statistical school, Dagum (1997) proposes his decomposition of the Gini index. In this decomposition, the two tools appears particularly important: the transvariation and the between group Gini ratio extended concept. The transvariation concept was previously defined by Gini (1916, 1959) and by Dagum (1959, 1960, 1961) (for bibliographic references, see Dagum (1997)). The extended Gini ratio between two subpopulation groups is defined by Dagum (1980).

Dagum (1997) decomposes the Gini index as follows:

$$G = G_w + G_{nb} + G_t$$

where G_w is the within group inequality index, G_{nb} represents the measure of the contribution to the overall inequality deriving from the inequality existing among the group affluence (net relative economic distance), and G_t is linked to the transvariation among groups.

We think that the meaning both of the Dagum decomposition and of the used symbols may be better understood when dividing the decomposition in two stages. Firstly, we suppose the Gini index decomposed in two parts

$$G = G_w + G_{gb} \tag{3}$$

where G_w is defined as

$$G_w = \sum_{j=1}^{k} G_{jj} p_j s_j \tag{4}$$

with

$$p_j = n_j / n; \quad s_j = n_j \mu_j / n\mu; \quad \sum_{j=1}^{k} p_j = \sum_{j=1}^{k} s_j = 1; \quad \sum_{j=1}^{k} \sum_{h=1}^{k} p_j s_h = 1$$

and

$$G_{jj} = \frac{1}{2 n_j^2 \mu_j} \sum_{i=1}^{n_j} \sum_{r=1}^{n_j} \left| y_{ji} - y_{jr} \right| = \frac{\Delta_{jj}}{2\mu_j}$$

is the Gini index within the j^{th} subpopulation. The symbol Δ_{jj} indicates the mean difference within the j^{th} group.

The term G_{gb} in (3) represents the gross Gini ratio between groups. It is defined as

$$G_{gb} = \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} \quad p_j s_h + p_h s_j \tag{5}$$

where the term

$$G_{jh} = \frac{1}{n_j n_h(\mu_j + \mu_h)} \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} \left| y_{ji} - y_{hr} \right| = \frac{\Delta_{jh}}{\mu_j + \mu_h} \tag{6}$$

is the extended Gini ratio between the j^{th} and the h^{th} groups as defined by Dagum (1980) and Δ_{jh} is the absolute mean difference between the considered groups.

The introduction of the concept of relative economic affluence (REA), linked to Dagum's idea of economic distance (Dagum 1980), leads to the second stage of the decomposition and we will discuss it in the next section. Using REA, G_{gb} is decomposed into two parts

$$G_{gb} = G_{nb} + G_t$$

where, as said above, G_{nb} represents contribution to gross inequality between groups deriving from the inequality existing among the group affluence (net relative economic distance) and G_t represents the contribution to G_{gb} due to the transvariations. The formal expressions of G_{nb} and G_t are

$$G_{nb} = \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} \;\; p_j s_h + p_h s_j \;\; D_{jh} \tag{7}$$

$$G_t = \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} \;\; p_j s_h + p_h s_j \;\; 1 - D_{jh} \tag{8}$$

with D_{jh} standing for the relative economic affluence between two groups. The term D_{jh} will be defined in the follow-up to this section.

In the third section, it will be clear that transvariations represent overlapping, here we stress that in (4) the sum of the weighting factors adds to one $\sum_{j=1}^{k} \sum_{h=1}^{k} p_j s_h = 1$.

1.2. The Dagum relative economic affluence

One of the most interesting aspects of the Dagum decomposition is the introduction of the relative economic affluence (REA). In order to define the relative economic affluence between the j^{th} and the h^{th} groups, one has to analyze two concepts: the former is termed by Dagum *gross economic affluence* d_{jh}, the latter by *first moment of the transvariation* p_{jh}.

Let us consider the j^{th} and the h^{th} groups with incomes Y_j and Y_h. In Dagum decomposition (1997) only positive incomes are considered, then supposing that the income distribution functions are continuous on $[0, \infty)$, the income distribution functions are symbolized respectively with $F_j(y)$ and $F_h(y)$, and the density functions with $f_j(y)$ and $f_h(y)$. We denote the two groups average incomes with

$$M_j(Y) = \int_0^\infty y f_j(y)\, dy = \mu_j,$$

$$M_h(Y) = \int_0^\infty y f_h(y)\, dy = \mu_h.$$

Moreover, it is supposed

$$M_j(Y) > M_h(Y).$$

The gross economic affluence d_{jh} between the j^{th} and the h^{th} groups is defined as "the weighted average of the income difference y_{ji}-y_{hr} for all incomes y_{ji} of the members belonging to the *j-th* subpopulation with incomes greater than y_{hr} of all members belonging to the h-th subpopulation, such that, $\mu_j > \mu_h$". (Dagum, 1997, p. 522). Dagum (1997) writes (see expression (19) in Dagum, 1997):

$$d_{jh} = M_j\left[F_h(y)Y\right] + M_h\left[F_j(y)Y\right] - M_h(Y). \tag{9}$$

The first moment of the transvariation, p_{jh}, is represented by "the weighted average of the income difference y_{hr}-y_{ji} for all pairs of economic units, one taken from the h-th subpopulation group and the other from the j-th such that $y_{hr} > y_{ji}$ and $\mu_j > \mu_h$" (Dagum, 1997, p. 522). Dagum (1997) writes (see expression (21) in Dagum, 1997):

$$p_{jh} = M_j\left[F_h(y)Y\right] + M_h\left[F_j(y)Y\right] - M_j(Y). \tag{10}$$

Both in d_{jh} and in p_{jh}, the weighting factor is the joint density function $f_j(y) f_h(y)$.

Using symbol z for the j^{th} group incomes and the symbol x for the h^{th} group incomes, in appendix we show that [2]

$$d_{jh} = \int_0^\infty f_j \; z \; dz \int_0^z \; z - x \; f_h \; x \; dx = M_j \left[F_h \; x \; Z \right] - M_h(X) + M_h \left[F_j \; z \; X \right] \tag{11}$$

$$p_{jh} = \int_0^\infty f_j \; z \; dy \int_z^\infty \; x - z \; f_h \; x \; dx = M_j \left[F_h \; x \; Z \right] + M_h \left[F_j \; z \; X \right] - M_j(Z) . \tag{12}$$

Then, being

$$\Delta_{jh} = \int_0^\infty \int_0^\infty |z - x| f_j \; z \; f_h \; x \; dzdx =$$

$$\int_0^\infty f_j \; z \; dy \int_0^z \; z - x \; f_h \; x \; dx + \int_0^\infty f_j \; z \; dy \int_z^\infty \; x - z \; f_h \; x \; dx .$$

one observes immediately that

$$\Delta_{jh} = d_{jh} + p_{jh} . \tag{13}$$

Substituting (9) and (10) into (13), one has the following expression for Δ_{jh}

$$\Delta_{jh} = 2M_j \left[F_h \; y \; Y \right] + 2M_h \left[F_j \; y \; Y \right] - \; M_h(Y) + M_j(Y) \; .$$

Dagum terms the difference between d_{jh} and p_{jh} net economic affluence.

The maximum value for $d_{jh} - p_{jh}$ is $d_{jh}=\Delta_{jh}$, obtained when $p_{jh}=0$, that is when the two distributions do not overlap.

The *relative economic affluence* (REA) between the j^{th} and the h^{th} groups with $\mu_j > \mu_h$ is defined as

$$D_{jh} = \; d_{jh} - p_{jh} \; / \Delta_{jh} \; .$$

---

[2] To define p_{jh}, Dagum uses the expression $p_{jh} = \int_0^\infty f_h \; y \; dy \int_0^y \; y - x \; dF_j \; x$ (Dagum,1997, p. 522). We think that (12) is more clear, obviously all the results of Dagum are confirmed.

REA is a normalized measure of the difference in average economic affluence between two groups and takes values in the closed interval $[0,1]$, needless to say, D_{jh}=0 when the average incomes of the two groups are the same, and D_{jh}=1 when there is no overlapping.

We remark that, on deriving REA, no particular hypothesis is introduced on the income density functions of the two groups.

2. COMPARISONS AMONG GINI INEQUALITY INDEX DECOMPOSITIONS AND THE ALTERNATIVE EXPRESSION FOR THE OVERLAPPING TERM PROPOSED IN THIS PAPER

It is our opinion that the Dagum decomposition clarifies the L.-A. first remark. Moreover, we think that the alternative expression for the overlapping term proposed in this section could further enhance the interpretation of R.

Dagum defines the average economic affluence of a population as the income mean μ (Dagum, 1980), using (9), (10), (11) and (12) we observe the net economic affluence between two groups may be represented by the difference between the average economic affluences of the considered groups.

$$d_{jh} - p_{jh} = M_j(Y) - M_h(Y) \tag{14}$$

Then, using (14), we rewrite D_{jh} as

$$D_{jh} = \left[M_j(Y) - M_h(Y)\right] / \Delta_{jh} \,. \tag{15}$$

We note that, following the definition of D_{jh}, the difference (1-D_{jh}) measures the difference between the relative economic affluence variation and its maximum value. The difference is due to the presence of overlapping between the groups.

Using the definition of Δ_{jh} as given in (13) and remembering (14), one can write (1-D_{jh}) as

$$1 - D_{jh} = 2p_{jh} / \Delta_{jh} \tag{16}$$

That is (1-D_{jh}) is equal to the intensity of transvariation.

As said above, p_{jh}=0 when, given M_j(Y)>M_h(Y), all the h^{th} group incomes are smaller than all the h^{th} group incomes. In other words, p_{jh}=0 if and only if the two income density functions do not overlap.

Therefore, we may use the "first moment" of the transvariation to represent the overlapping between the two income distributions.

We stress that this "measure" takes into account a part of the income differences between the $n_j \times n_h$ binary combinations of economic units belonging to the j^{th} and the h^{th} subpopulations. More precisely, given $M_j(Y)>M_h(Y)$, the "measure" is the weighted sum of the absolute values of the negative differences (y_{ji}-y_{hr}), the weighting factor being the joint density function f_j $(y)f_h(y)$. To construct p_{jh} one has to consider two elements: (*i*) the differences between incomes of the individuals, belonging to the more affluent group, poorer than individuals belonging to the less affluent group, (*ii*) the relative number of these differences. One can say that p_{jh} represents the "amount of the overlapping" between two groups.

The value of p_{ij} *may* then change when the two groups income values or their frequency distributions vary. More precisely the value of p_{ij} varies if, changing income values inside the overlapping range or their frequencies, the negative income differences change in a no compensative way. It follows that changes in variance and in asymmetry of the two groups income distributions may or not modify p_{jh}. Changes in variance and in asymmetry alter p_{jh} when they are also originated by no compensative variations in p_{jh} components. It follows that, if inequality in one of the two groups changes, p_{ij} may or may not change.

2.1. An alternative expression for the overlapping term

We want to show firstly that G_t=R=A_0 and then we will rewrite G_t in an alternative way.

Let us compare the Gini inequality index decompositions. We recall the Mookherjee and Shorrock (1982) decomposition written as in (1)

$$G=\sum_j p_j^2\lambda_j G_j+\frac{1}{2}\sum_j\sum_h p_j p_h\left|\lambda_j-\lambda_h\right|+R=G_W+G_B+R$$

with $p_j=n_j/n$ and $\lambda_j=\mu_j/\mu$,

and the decomposition of the Gini index proposed by Dagum (1997) as written in the first section:

$$G = G_w + G_{nb} + G_t \,.$$

It is easy to see that the contribution of the Gini inequality within groups to the total Gini ratio, G_w, defined as in (4)

$$G_w = \sum_{j=1}^{k} G_{jj} p_j s_j \qquad p_j = n_j / n; \quad s_j = n_j \mu_j / n\mu; \quad \sum_{j=1}^{k} p_j = \sum_{j=1}^{k} s_j = 1$$

has the same formal expression of G_W in M.- S. (1982) and then the same expression of A_W in L-A (1993).

The same holds for G_{nb} and G_B. In fact, reconsidering expressions (7) and (6) we have

$$G_{nb} = \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} \;\; p_j s_h + p_h s_j \;\; D_{jh}$$

and

$$G_{jh} = \frac{1}{n_j n_h (\mu_j + \mu_h)} \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} \left| y_{ji} - y_{hr} \right| = \frac{\Delta_{jh}}{\mu_j + \mu_h} \,.$$

Then substituting (6) in (7) and remembering the expression (15) given for D_{jh}, we obtain

$$G_{nb} = \sum_{j=2}^{k} \sum_{h=1}^{j-1} \frac{\mu_j - \mu_h}{\mu_j + \mu_h} \;\; p_j s_h + p_h s_j \;\; .$$

Observing that

$$p_j s_h + p_h s_j \;\; = \left[n_j n_h \;\; \mu_j + \mu_h \;\; \right] / n^2 \mu$$

one has

$$G_{nb} = \sum_{j=2}^{k} \sum_{h=1}^{j-1} \frac{n_j n_h \;\; \mu_j - \mu_h}{n^2 \mu} = G_B \,. \tag{17}$$

However, we stress that there is a substantial difference between G_B and G_{nb}. The component G_B is obtained starting from the hypothesis of income equidistributed in the two compared groups. G_{nb} is one of the two parts of the gross between inequality component of the Gini due to the net difference in average affluence (represented by income mean) existing between each pair of groups. Dagum derives this component introducing a particular index (REA) without any hypothesis on income distributions.

Being G_w=G_W and G_{nb}=G_B, must hold the equality G_t=R. The residual component R may be directly calculated using the expression (8) proposed by Dagum

$$G_t = \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} \quad p_j s_h + p_h s_j \quad 1 - D_{jh} \quad .$$

We observe that substituting in (8) expressions (6) and (15) given respectively for G_{jh} and D_{jh} one has

$$R = G_t = \sum_{j=2}^{k} \sum_{h=1}^{j-1} \frac{n_j n_h}{n^2 \mu} \quad \Delta_{jh} - \mu_j + \mu_h \quad . \tag{18}$$

Remembering expressions (13) and (14), we observe that

$$\Delta_{jh} = \left[\mu_j - \mu_h \right] + 2 p_{jh}$$

and we rewrite expression (18) as

$$R = 2 \sum_{j=2}^{k} \sum_{h=1}^{j-1} \frac{n_j n_h}{n^2 \mu} \quad p_{jh} \quad . \tag{19}$$

Summing up, Mookherjee and Shorrocks (1982) evaluate the overlapping term R by the difference between the overall Gini index and the sum of Gini within and Gini between components. Lambert and Aronson (1993) obtain the value of R considering an area. Here, we observe that, using the Dagum decomposition, R may be expressed by the second component of the gross between inequality G_t. From this, we derive that overlapping term may be written as twice the ratio between the weighted sum of the absolute measures of overlapping p_{jh} and the population average income. As said above, p_{jh} is a function of both a part of the compared groups incomes and their distribution functions. As L.-A. (1993) maintain, the overlapping term is at once a between groups and within groups effect measuring a between groups phenomenon, the overlapping, that is generated by inequality within groups. However, changes in within group inequality do not necessarily influence R.

It happens if and only if at least a component of the "measure" of the overlapping between groups p_{jh} is involved in the change.

3. AN APPLICATION TO POLISH DATA

Through the expression (2), the value of the residual term can be directly calculated as the difference between two curves: (*i*) the concentration curves obtained ranking all incomes according to the order of the groups average incomes, and (*ii*) the Lorenz curve. The proposed expression (19) computes the residual term R involving directly the transvariations among groups. We can say that this term evaluates the importance of the intersections among the income group sets, and then, it evaluates the degree of the homogeneity among the income values, enlightening on the power that the adopted classification criterion has in forming groups. Using expression (19) we are able to decompose the residual term and to assess the contribute given by each pair of groups. Thanks to this procedure, we can evaluate the effects of a tax system (*i*) on the within group inequality (G_W), (*ii*) on the between group average economic affluence (G_B), and (*iii*) on the overlapping term (R).

As it is known, in the present Polish personal tax system, no family allowances are scheduled: the only distinction is made between singles and couples, being indifferent for the latter having or not having children. Couples with or without children may add their incomes and then apply the tax schedule reported in table 1 to each separate half of total income: the resulting tax is doubled. Such taxation treats a couple, with or without children, as if there were two singles with income equal to half of the couple's income. In a recent study (Vernizzi, Monti, Kosny, 2006) two different theoretical tax system are proposed in order to take into account the family composition in a more fair way.

Table 1

Present income tax schedule

Income bracket [PLN]		Tax rate
0	2790	0%
2790	37024	19%
37024	74048	30%
74048		40%

Source: Polish Ministry of Finance (http://www.mf.gov.pl)

Here we analyze the inequality changes induced by the present tax system in the income distribution of the year 2001. The performed analysis is an elaboration of the summary statistics published in Vernizzi, Monti, Kosny (2006). These statistics concern a sub-sample of 20430 families of the Polish Central Statistical Office sample (Household Budget Survey 2001). In the analysis, single households and married couples with no more than three children are considered.

Table 2

Basic statistics for nominal income (PLN)

	Single (S)	Couple (C)	Couple with 1 child (C+1)	Couple with 2 children (C+2)	Couple with 3 children (C+3)
BT nominal income					
Mean	11231.56	21745.50	24552.25	24110.62	20751.76
Number of households	4728	5775	4033	4339	1555
Relative frequency	0.141	0.259	0.217	0.272	0.111
Minimum	0	0	0	0	0
Maximum	240000.00	222600.00	216000.00	240000.00	168670.00
AT nominal income					
Mean	9555.62	18630.52	20861.00	20485.37	17732.99
Number of households	4728	5775	4033	4339	1555
Relative frequency	0.141	0.259	0.217	0.272	0.111
Minimum	0	0	0	0	0
Maximum	156007.50	157575.00	153615.00	168015.00	125216.80

Source: Vernizzi, Monti, Kosny (2006)

Table 2 presents the nominal average incomes for each family type calculated before (BT) and after taxation (AT) and some other general characteristics of the used sample. The data was implicitly used but not published in the above-cited paper. We observe that couples without children present an average income, which is roughly twice that of the singles. Couples with three children are on average in a worse position: 20751.76 PLN, 4.6% lower than couples without children. On average, the highest level is related to couples with one child (24552.25 PLN) followed by couples with two children (1.8% lower).

Vernizzi, Monti, Kosny (2006) chose the single as reference type and expression (20) transforms the income y_i, giving to the i-type household a certain welfare level, into the income ($y_{S,i}$) ensuring to the single the same welfare level:

$$y_{S,i} = S_{S,i}\ y_i\ = \gamma_S + \frac{1}{m_i}\ y_i - \gamma_i\ . \tag{20}$$

In (20), the parameter γ_i represents the minimum survival income level for *i*-type household, while parameter m_i takes into account the needs of household *i* over the subsistence level.

Table 3

Family coefficients and exemptions for the equivalent income function

		S	C	C+1	C+2	C+3	C+n
OECD scale coefficients	m_i	1	1.5	1.8	2.1	2.4	1.5 + n·0.3
Minimum survival incomes	γ_i	4000	6700	10000	13300	16600	6700 + n·3300

Source: Vernizzi, Monti, Kosny (2006)

The modified OECD scale suggests the values of the parameters m_i. This scale assigns 1 for first adult, 0.5 for consecutive adults and 0.3 for each child. The information published yearly by the Institute of Labour and Social Matters (IPiSS) for selected family types (see IPiSS, 2001) allows evaluating the minimum survival income. Table 3 shows the values of parameters m_i and γ_i based on these sources.

As the minimum value of incomes is zero (see table 2), the function (20) will yield negative incomes. We observe that the minimum negative income will be referred to the family with three children when its nominal income is equal to zero:

$$S_{S,C+3} = \gamma_s + \frac{1}{m_{C+3}}\ 0 - \gamma_{C+3}\ .$$

Then, to avoid the negative income problem, this family is chosen as reference type and (20) becomes

$$S_{C+3,i} = 16600 + \frac{1}{m_i / 2.4}\ y_i - \gamma_i\ .$$

To improve the understanding of the obtained results, in our calculations we will use the *per capita* equivalent income referred to the family with three children:

$$y_{C+3,i} = \frac{1}{2.4}\left(16600 + \frac{1}{m_i / 2.4}\, y_i - \gamma_i\right). \tag{21}$$

Through (21) we obtain the equivalent average incomes presented in table 4.

Table 4

BT and AT average equivalent incomes

	S	C	C+1	C+2	C+3	Population
BT average income	14148.23	16947.00	15001.25	12064.58	8646.57	13877.09
AT average income	12472.28	14870.35	12951.09	10338.27	7388.75	12049.29
Absolute differences	1675.94	2076.65	2050.16	1726.31	1257.82	1827.80
Percentage variations	11.86	12.25	13.67	14.32	14.55	13.17

Source: own calculations

The ranking of average incomes is modified by the equivalent income function. In decreasing order, and dealing with nominal incomes, the before tax ranking is

$C + 1,\ C + 2,\ C,\ C + 3,\ S.$

Conversely, dealing with equivalent incomes the ranking becomes

$C,\ C + 1,\ S,\ C + 2,\ C + 3.$

In the BT average equivalent incomes survey the couple with three children have income which is far lower than the maximum: 51% less than the couple without children. Moreover, the average income for the single rises to 94% with respect to the income of the couple with a child and to 83% of the maximum income (couple without children).

As is well known, one cannot carry on the transformation of nominal incomes into equivalent incomes without ambiguity. The results in table 4 depend on the welfare function chosen to transform nominal incomes into the equivalent ones; if another welfare function were proposed different results could be obtained.

The analysis performed in the pursue deals with the equivalent incomes obtained by (21).

We observe that taxation does not modify the average incomes order. Table 4 shows that average equivalent income of the population decreases by

13.17% after tax. The welfare loss induced by taxation is quite different considering population groups and we observe that the welfare loss increases when the number of household members increases.

The structure of the changes induced by the tax system is first analyzed comparing the before and after tax Gini indices and evaluating the variations in its components; then, the taxation effects on family groups will be analyzed. Table 5 reports the Gini indices both before and after tax.

Table 5

BT and AT Gini indices and their components

(Absolute values and percentages. Equivalent incomes)

	Gini within (G_W)	**Gini between (G_B)**	**Overlapping term (R)**	**Gini index (G)**
BT	0.06492	0.10210	0.14504	0.31206
Percentage composition (1)	20.803	32.718	46.479	100
AT	0.06050	0.10656	0.12858	0.29564
Percentage composition (2)	20.464	36.043	43.492	100
Difference (2)-(1).	-0.339	3.325	-2.986	

Source: own calculations

After taxation the Gini inequality index decreases by 5.27% (table 6).

Table 6

Decomposition of the Gini index percentage variation

$(ATG_W\text{-}BTG_W)/BTG$	$(ATG_B\text{-}BTG_B)/BTG$	$(ATR - BTR)/BTG$	$(ATG\text{-}BTG)/BTG$
-1.415	1.429	-5.274	-5.259

Source: own calculations

Table 6 suggests two remarks about the inequality change. First, the reduction in Gini within component induced by the progressivity of the tax system is compensated by an increase of the in-between inequality, that is, by an increase of average affluence differences among groups: the within group income distributions are less unequal after tax (-1.415% variation with respect to BT Gini index), but the groups are more unequal when we consider their average equivalent incomes (+1.429% variation with respect to BT Gini index). Second, the inequality change is essentially due to the overlapping term variation (-5.274% with respect to BT Gini index). As far as the equivalent incomes are concerned, this means that the importance of the intersection among group income sets is less relevant: the incomes values are more clustered after taxation than before.

Deeper insight into the structure of the inequality change is obtained analyzing the taxation effects on the Gini index components with reference to each group and pair of groups. Let us begin considering the Gini within components for each family type.

Table 7 reports the before and after tax weighted Gini within indices, the percentage of these values with respect to the overall Gini and the absolute variations of the percentages, i.e. the differences between before and after tax percentages. Table 7 confirms that taxation reduces within inequality both for the overall population and for each family type group.

Table 7

Gini within index: weighed components for family type

(Equivalent incomes)

Family type	BT Gini within	AT Gini within	(1) % BT with respect to Gini index	(2) % AT with respect to Gini index	Variation (2)-(1)
C	0.01943	0.01809	6.22489	6.12054	-0.10435
C+1	0.01586	0.01477	5.08208	4.99716	-0.08492
S	0.00523	0.00486	1.67704	1.64298	-0.03406
C+2	0.02143	0.01999	6.86824	6.76148	-0.10676
C+3	0.00297	0.00279	0.95068	0.94223	-0.00844
Gini within	0.06492	0.06050	20.80293	20.46440	-0.33853

Source: Vernizzi, Monti, Kosny (2006) and own calculations

Table 8 reports the group Gini within indices without weights.

Table 8

Gini index within each family type

(Values without weights. Equivalent incomes)

Family type	BT Gini. within	AT Gini within	% variation
C.	0.2378	0.2192	-7.822
C+1	0.3123	0.2926	-6.308
S	0.2576	0.2355	-8.579
C+2	0.3331	0.3148	-5.494
C+3	0.3835	0.3659	-4.589

Source: own calculations

Table 8 shows that, although taxation reduces overall inequality, when each family type is considered the inequality reduction appears to be inversely related to the family size.

In table 9, the Gini between index is analyzed.

Table 9

Gini between: weighted components for family type pairs

(Equivalent incomes)

Group pairs	BT Gini between x100	AT Gini between x 100	(1) BT % with respect to BT Gini index	(2) AT % with respect to AT Gini index	difference (2)-(1)
C, C+1	0.78597	0.89287	2.51867	3.02009	0.50141
C, S	0.73631	0.72659	2.35955	2.45767	0.09812
C, C+2	2.47548	2.64642	7.93282	8.95140	1.01857
C, C+3	1.72369	1.78932	5.52365	6.05230	0.52865
C+1, S	0.18807	0.12158	0.60267	0.41122	-0.19144
C+1, C+2	1.24778	1.27858	3.99857	4.32474	0.32617
C+1, C+3	1.10588	1.11483	3.54386	3.77087	0.22701
S, C+2	0.57661	0.68013	1.84778	2.30051	0.45273
S, C+3	0.62357	0.66358	1.99827	2.24453	0.24627
C+2, C+3	0.74662	0.74201	2.39257	2.50983	0.11726
Gini between	10.20997	10.65592			

Source: own calculations

As immediately appears from expression (17)

$$G_B = \sum_{j=2}^{k} \sum_{h=1}^{j-1} \frac{n_j n_h}{n^2 \mu} \frac{\mu_j - \mu_h}{}$$

the Gini between index is a weighted sum of the Gini between each pair of groups: in the first two columns of table 9, we report both G_B and its components (multiplied by 100). The third and fourth columns report the percentages of the Gini between each pair of groups, calculated with respect to the overall Gini. The percentages are referred both to the before (third column) and after (fourth column) tax situation. Focusing on the childless couple, before taxation, the Gini between this typology and the couple with one child represent 2.52% of the overall Gini index. After tax, the percentage raises to 3.02%: taxation enlarges the average welfare difference between these two groups and then the importance of this inequality factor increases evaluating the overall inequality. The last column of table 9 shows the variations of the percentages. On considering the adopted welfare measure and remembering the meaning of the Gini between component, positive variations make evident that the tax system augments the average welfare differences for the related pair of family types. The negative sign in the fifth row shows that after tax the average welfare is less unequal only for the pair single - couple with one child.

Table 10 presents the pairs of family types ranked according to the increasing order of their own Gini between.

Table 10

Increasing order of the Gini between for each pair of family types

Before tax order	**After tax order**	**Order of % variation**
C+1, S	C+1, S	C+1, S
S, C+2	S, C+3	C, S
S, C+3	S, C+2	C+2, C+3
C, S	C, S	C+1, C+3
C+2, C+3	C+2, C+3	S, C+3
C, C+1	C, C+1	C+1, C+2
C+1, C+3	C+1, C+3	S, C+2
C+1, C+2	C+1, C+2	C, C+1
C, C+3	C, C+3	C, C+3
C, C+2	C, C+2	C, C+2

Source: own calculations

The table shows that the changes induced by taxation do not modify the Gini between group order in a relevant way. We observe that the lowest values are measured comparing the single with each of the other family types both before and after taxation. In fact, in table 10, the first four positions of each column involve the single matched with the other family types. Before taxation, the minimum value of the Gini between is observed considering the single and the couple with one child, we remark that taxation reinforces this position. The inequality in average equivalent incomes due only to the different composition of the households assumes the highest values comparing the couple without children with the couples with two or three children. The last column of table 10 shows that the tax system enlarges these differences being the most important variations referred to these groups. We can conclude that taxation strengthens the before tax ranking among the groups when the ranking is based on the importance of the average welfare differences.

Table 11 analyzes the behaviour of the overlapping term R.

The table shows the before and after tax values of the overlapping terms between each group pair, both in absolute values and in percentages of the overall Gini indices, together with the differences of the percentages themselves. Analyzing the data, we see immediately that taxation induces changes in overlapping term with opposite sign with respect to the Gini between changes. Taxation reduces R in almost all pairs of family types. We observe only a positive variation, which is referred to the single matched with the couple with one child. On considering these two groups, taxation has the following consequences: first, it reduces the difference between the average welfare measures and second, it augments the homogeneity among the equivalent income values of the two groups. The tax effects are opposite considering all the other pair of family types: the tax system augments the differences of the average welfare level and reduces the intersection between equivalent incomes sets associated with family types.

Table 11

Overlapping term R: weighted components for family type pairs

(Equivalent incomes)

Groups pair	**BT R x 100**	**AT R x 100**	**(1) BT % with respect to Gini index**	**(2) AT % with respect to Gini index**	**difference (2)-(1)**
C, C+1	2.83767	2.51633	9.09346	8.51136	-0.58210
C, S	1.38623	1.25593	4.44226	4.24811	-0.19415
C, C+2	2.14382	1.80658	6.86998	6.11069	-0.75929
C, C+3	0.42047	0.33728	1.34742	1.14083	-0.20658
C+1, S	1.67213	1.60781	5.35844	5.43833	0.07989
C+1, C+2	2.57624	2.31294	8.25571	7.82340	-0.43231
C+1, C+3	0.55261	0.47665	1.77087	1.61224	-0.15863
S, C+2	1.61750	1.39724	5.18336	4.72610	-0.45726
S, C+3	0.33107	0.27328	1.06093	0.92437	-0.13656
C+2, C+3	0.96620	0.87422	3.09624	2.95701	-0.13922
R	14.50393	12.85825	46.47866	43.49244	-2.98621

Source: own calculations

Table 12 ranks in the first two columns the pair of family types following the decreasing order of the R values and in the last column following the decreasing order of the percentage differences. It is easy to see that taxation does not change the ranking among the pair of family types and that the smallest values for the overlapping terms refer to the couple with three children compared with all the other groups. This observation suggests that applying the proposed welfare measures, the set of equivalent incomes referred to the couples with three children almost does not intersect the sets of the equivalent incomes related to the other family types.

Table 12

Decreasing order of the overlapping term R for each pair of family types

Before tax order	After tax order	Percentage variation order
C, C+1	C, C+1	C, C+2
C+1, C+2	C+1, C+2	C, C+1
C, C+2	C, C+2	S, C+2
C+1, S	C+1, S	C+1, C+2
S, C+2	S, C+2	C, C+3
C, S	C, S	C, S
C+2, C+3	C+2, C+3	C+1, C+3
C+1, C+3	C+1, C+3	C+2, C+3
C, C+3	C, C+3	S, C+3
S, C+3	S, C+3	C+1, S

Source: own calculations

Now recalling expression (19) for the overlapping term R:

$$R = 2\sum_{j=2}^{k}\sum_{h=1}^{j-1}\frac{n_j n_h}{n^2\mu}\; p_{jh}$$

together with expression (17) for the Gini between index:

$$G_B = \sum_{j=2}^{k}\sum_{h=1}^{j-1}\left[n_j n_h \;\; \mu_j - \mu_h \;\;\right]/n^2\mu$$

and expression (4) $G_W = \sum_{j=1}^{k} G_{jj} p_j s_j$ for the Gini within index, we can evaluate the extent to which the overall population inequality is determined by each group inequality.

To obtain this result we proceed as follows. Let us indicate the weighted Gini within the i-th group as ${}_iG_W$, the weighted Gini between the i^{th} and j^{th} group as ${}_{i,j}G_B$ and the weighted overlapping term between the same groups as ${}_{i,j}R$.

Then, for the five groups taken into consideration here, we can rewrite the overall Gini index $G = G_W + G_B + R$ as

$$G = \left[{}_iG_W + \sum_{\substack{j=1 \\ j \neq i}}^{5} {}_{i,j}G_B + \sum_{\substack{i=1 \\ j \neq i}}^{5} {}_{i,j}R \right] + Z. \qquad (22)$$

In (22), G is divided into two parts. The former, within the square brackets, is referred to the i^{th} group inequality both considering the group alone and matching it with all the other ones. The latter, Z, represents the within and between groups inequality that does not involve the i^{th} group.

Being ${}_iR = \sum_{\substack{j=1 \\ j \neq i}}^{5} {}_{i,j}R$ and ${}_iG_B = \sum_{\substack{j=1 \\ j \neq i}}^{5} {}_{i,j}G_B$, we define the ratio

$${}_iIC = \frac{{}_iG_W + {}_iG_B + {}_iR}{G} \qquad (23)$$

as the extent to which the i^{th} group inequality determines the overall population inequality measured by the Gini index. We term (22) Inequality Contribution of the i^{th} group (${}_iIC$).

For example, let us evaluate the before tax inequality contribution for the couple without children, ${}_CI.C.$

The before tax weighted Gini within for the couples without children, ${}_CG_W = 0.019$, is written in the first column of table 7. The Gini between index (${}_CG_B$= 0.057) is obtained looking at table 9. In the first column of the table we have

$100_{C.S}G_B$	$100_{C.C+1}G_B$	$100_{C.C+2}G_B$	$100_{C.C+3}G_B$	$Sum=100_CG_B$
0.73631	0.78597	2.47548	1.72369	5.72144

With reference to the overlapping term, looking at table 11 we obtain ${}_CR$=0.068. In the first column of the table we have

$100_{C.S}R$	$100_{C.C+1}R$	$100_{C.C+2}R$	$100_{C.C+3}R$	$Sum=100_CR$
1.38623	2.83767	2.14382	0.42047	6.78818

Then, remembering that the before tax Gini index is 0.312 (see table 5), we obtain ${}_CIC$=0.475

100 ${}_CIC$= (1.943+ 5.721+6.788)/0.312=47.52

Table 13 reports ${}_iG_W$, ${}_iG_B$ and ${}_iR$ calculated for each family type.

Table 13

Inequality per family type

Family type	Gini within (${}_iG_W$) Weighted components x 100		Gini between (${}_iG_B$) Weighted components x 100		Overlapping term ${}_iR$ Weighted components x 100	
	BT	AT	BT	AT	BT	AT
S	0.52333	0.48574	2.12456	2.19188	5.00693	4.53426
C	1.94251	1.80950	5.72144	6.05521	6.78818	5.91612
C+1	1.58589	1.47738	3.32769	3.40786	7.63865	6.91372
C+2	2.14327	1.99899	5.04649	5.34715	7.30375	6.39098
C+3	0.29666	0.27856	4.19975	4.30975	2.27034	1.96143

Source: own calculation

Table 14 shows the percentage inequality contributes 100 *IC* for each family type. Moreover, in the same table are reported the percentages of ${}_iG_W$, ${}_iG_B$ and ${}_iR$, calculated with respect to the Gini index for each family type.

Table 14

Percentage Inequality Contribution per family type

	S	C	C+1	C+2	C+3
Before tax BT					
${}_iG_W$ **percentage w.r.t** G	1.67704	6.224892	5.08208	6.86824	1.00345
${}_iG_B$ **percentage w.r.t** G	6.80826	18.33469	10.66377	16.17175	13.45835
${}_iR$ **w.r.t** G	16.04498	22.96072	24.47848	23.40528	7.27545
IC percentage	24.53029	47.52030	40.22433	46.44527	21.73725
After tax AT					
${}_iG_W$ **percentage w.r.t** G	1.64298	6.12054	4.99716	6.76148	0.94223
${}_iG_B$ **percentage w.r.t** G	7.41394	20.48145	11.52691	18.08647	14.57753
${}_iR$ **w.r.t** G	15.33691	20.01099	23.38534	21.61720	6.63445
IC percentage	24.39383	46.61298	39.90941	46.46516	22.15422
Differences (AT-BT)					
${}_iG_W$ **percentage w.r.t** G	-0.03406	-0.10435	-0.08492	-0.10676	-0.64099
${}_iG_B$ **percentage w.r.t** G	0.60567	2.14676	0.86314	1.91472	1.11918
${}_iR$ **w.r.t** G	-0.70808	-2.94973	-1.09314	-1.78808	-0.06122
IC percentage	-0.13646	-0.90732	-0.31492	0.01989	0.41697

Source: own calculations

Let us focus again on the couple without children. Taking into account the before tax situation, it appears that considering this group and matching it with all the others, we obtain 47.52% of the population inequality measured by the Gini index.

Moreover, the weighted inequality within this group and the gross between inequality (${}_CG_B+{}_CR$) obtained comparing the group with all the others, represent, respectively, 6.23% and 41.29% of the overall population inequality. Dividing the gross between inequality, we observe that the difference in average welfare among this group and all the others explains 18.33% of the overall inequality. The percentage value assumed by ${}_CR$, 22.96%, evaluates the degree of homogeneity between the couple and all the other typologies.

We conclude that taxation induces the following inequality changes: (*i*) the within the group inequality index, ${}_CG_W$, decreases (from 6.23% to 6.12%); (*ii*) ${}_CR$, which measures how much couple incomes overlap with those related to other family types (that is the homogeneity income degree), decreases too (from 22.96% to 20.01%); (*iii*) the difference between couples' average welfare and other family types' average welfare levels increases after taxation (from 18.33% to 20.48%); (*iv*) after tax the inequality contribution of the couple without children to the overall population inequality is lower (46.61%) than before (47.52%): this result is essentially due to the sensible reduction in overlapping (2.95 percent points), which overrides the increases in ${}_CG_B$ (2.15 percentage points) and in ${}_CG_W$ (0.11 percentage points).

Using IC, we are able to rank the family type with respect to their Gini index share or, which is the same, with respect to their part of overall inequality. From table 14 we have the following ranking

C, C+2, C+1, S, C+3.

The ranking does not vary on considering the before and after tax situation.

If we go back to formulae (4), (17) and (19), we notice that the *IC* percentages reported in table 13 depend on family types' relative frequencies: so, when considering that *IC* is minimum for the couple with three children, 21.73%, we have to keep in mind that couples with three children are just 7.6% of the total number of families, whilst couples without children are 28.26%.

Nevertheless, this consideration has no effect on the analysis of changes in ${}_iG_W$, ${}_iG_B$, ${}_iR$ and IC induced by taxation. Looking at the third section of table 14, we can observe that ${}_iG_W$, ${}_iG_B$, ${}_iR$ variations present the same signs for all family types: however the IC variations are not in the same direction for all family types. Taxation reduces IC indices for the single, the couple and the couple with one child; the contrary happens for couples with two and three children.

These results seem to suggest the following conclusions. Given the equivalent income function (21), which makes family welfare depend on a minimum survival income for each member in addition to a coefficient which depends on the number of members, we can outline some conclusions. The present tax system reduces the contribution to overall inequality for couples without children, singles, couples with one child, while on the contrary, the contribution for couples with two and with three children increases (table 14): taxation seems to increase distances among families with a greater number of members and the rest; then, given the lower average equivalent incomes they show before tax (table 4), their relative situation is worsened by taxation. Moreover, singles and couples with one child present features more similar after then before taxation: Gini between these two typologies decreases (see table 9) whilst their overlapping term increases (see table 11). Finally, we observe that Gini within the income distributions of these family types decreases according to smaller percentages (see table 8) than for the others.

CONCLUSIONS

In this paper the Dagum decomposition of the Gini index (1997) is discussed and compared with Mookherjee and Shorrock (1982) and Lambert and Aronson (1993) decompositions. In so doing, a deeper insight into the meaning of residual term R is given and an alternative expression to calculate this term is proposed. We suggest that the weighted sum of transvariations may be used to evaluate the overlapping between two groups and we show that the residual term R may be written as twice the ratio between the weighted sum of the overlapping evaluations and the population average income. The weighted sum of transvariations between two groups is a function of two elements: the first being a subset of the incomes belonging to the compared groups and the latter is their joint distribution function. As L.-A. (1993) maintain, the overlapping term is at once a between groups and within groups effect measuring a between groups phenomenon, the overlapping, that is generated by inequality within groups. However, changes in within groups inequality do not necessarily influence R. It happens if and only if components of the weighted sum of transvariations between the groups are involved in the change in a non-compensatory way. The expression proposed here allows to calculate the overlapping term R referred to each pair of groups and then referred to the overall population in a relatively simple way. We use it analysing a particular source of income inequality changes, that is the taxation. In the last section of the paper, we consider a Polish population subset composed by families with a different number of members and we try to evaluate the effects that the present Polish tax system has on the income inequality with reference to the whole population and to each group of families. A particular equivalent income function is used to take into account the different composition of the families. Considering equivalent income as a measure of welfare level, we observe that the present tax system is unfair larger families. The inequality within the income distributions of these family types decreases by smaller amount and taxation increases the percentage of the total inequality ascribed to these groups. Moreover, only considering the singles with respect to couples with one child taxation reduces the difference between the average welfare measures and augments the homogeneity among the equivalent incomes values. The tax system induces opposite changes considering all the other pairs of family types: differences among average welfare levels augment and family type income clusters result to be more separate, that is less homogeneous.

Acknowledgments

I wish to thank Achille Vernizzi, Bruno Bosco, Walenty Ostasiewicz, Alessandro Santoro and an anonymous referee for helpful comments and suggestions. Needless to say none of the above mentioned persons should be responsible for the remaining deficiencies.

Appendix

We show that the definitions of d_{jh} and p_{jh} given by Dagum (1997) lead to the following formal expressions:

$$d_{jh}=M_j\left[F_h \ \ y \ \ Y\right]+M_h\left[F_j \ \ y \ \ Y\right]-M_h(Y)$$

$$p_{jh}=M_j\left[F_h \ \ y \ \ Y\right]+M_h\left[F_j \ \ y \ \ Y\right]-M_j(Y)$$

quoted as (19) and (21) in Dagum (1997).

To avoid misunderstanding, we denote:
x the h^{th} group incomes with density function $f_h(x)$,
z the j^{th} group incomes with density function $f_j(z)$,
and we remember that $f(z,x) = f_j(z)f_h(x)$.
Then, applying the definition of d_{jh}, we have

$$d_{jh} = \int_0^\infty f_j(z)dz \int_0^z (z-x) f_h(x)dx$$

$$\int_0^\infty f_j(z)dz \int_0^z (z-x) f_h(x)dx = \int_0^\infty f_j(z)dz\left[\int_0^z z f_h(x)dx - \int_0^z x f_h(x)dx\right] =$$

$$\int_0^\infty zF_h(x) f_j(z)dz - \int_0^\infty f_j(z)dz \int_0^z x f_h(x)dx = M_j\left[F_h(x)Z\right] - \int_0^\infty f_j(z)dz \int_0^z x f_h(x)dx .$$

Remembering that for the integration order inversion one has (see Amerio (1997), pag. 386)

$$\int_T f(z,x)dzdx = \int_0^h dz \int_0^z f(z,x)dx = \int_0^h dx \int_x^h f(z,x)dz$$

we write

$$\int_0^\infty f_j(z)dz \int_0^z x f_h(x)dx = \int_0^\infty f_h(x)dx \int_x^\infty x f_j(z)dz = \int_0^\infty x f_h(x)dx \int_x^\infty f_j(z)dz =$$

$$\int_0^\infty x\left(1-F_j(z)\right) f_h(x)dx = M_h(X) - M_h\left[F_j(z)X\right].$$

Then

$$d_{jh} = M_j\left[F_h(x)Z\right] - M_h(X) + M_h\left[F_j(z)X\right].$$

Denoting all incomes with y and then writing as Dagum (1997) in (19) page 522, one has

$$d_{jh} = M_j\left[F_h(Y)\right] - M_h(Y) + M_h\left[F_j(Y)\right].$$

In the paper the above expression is reported as expression (9).
Now, we show that

$$p_{jh} = M_j\left[F_h(y)Y\right] + M_h\left[F_j(y)Y\right] - M_j(Y).$$

Remembering that

$$\int_0^\infty \int_0^\infty (z-x) f_j(z) f_h(x) dzdx = \int_0^\infty f_j(z)dz\left[\int_0^z (z-x) f_h(x)dx + \int_z^\infty (z-x) f_h(x)dx\right]$$

=

$$= \int_0^\infty f_j \;\; z \; dz \int_0^z \;\; z-x \;\; f_h \;\; x \;\; dx + \int_0^\infty f_j \;\; z \; dz \int_z^\infty \;\; z-x \;\; f_h \;\; x \;\; dx = M_j(Z) - M_h \;\; X$$

we can rewrite $M_j(Z) - M_h \;\; X$ as

$$d_{jh} + \int_0^\infty f_j \;\; z \; dz \int_z^\infty \;\; z-x \;\; f_h \;\; x \;\; dx = M_j(Z) - M_h \;\; X \;\; .$$

Then, substituting the obtained expression for d_{jh}, one has

$$\int_0^\infty f_j \;\; z \; dy \int_z^\infty \;\; z-x \;\; f_h \;\; x \;\; dx = M_j(Z) - M_j\left[F_h \;\; x \;\; Z\right] - M_h\left[F_j \;\; z \;\; X\right].$$

Dagum defines p_{jh} as “the weighted average of the income difference y_{hr} - y_{ji} for all pair of economic units, one taken from h subpopulation the other from the j subpopulation such that $y_{hr} > y_{ji}$ and $\mu_j > \mu_h$” (Dagum, 1997, p. 522). Here we have

$$\int_0^\infty f_j \;\; z \; dy \int_z^\infty \;\; x-z \;\; f_h \;\; x \;\; dx = +M_j\left[F_h \;\; x \;\; Z\right] + M_h\left[F_j \;\; z \;\; X\right] - M_j(Z) \, .$$

Substituting the y instead of z and x as done by Dagum, we obtain the Dagum expression (21) on p. 522 reported in the paper as expression (10).

$$p_{jh} = +M_j\left[F_h \;\; y \;\; Y\right] + M_h\left[F_j \;\; y \;\; Y\right] - M_j(Y) \, ,$$

q.e.d.

REFERENCES

Amerio, L., *Analisi matematica*, [*Mathematical Analysis*], Vol.2, UTET, Torino, 1977 (In Italian).

Bhattacharya, N., Mahalanobis, B., *Regional Disparities in Household Consumption in India*, “Journal of the American Statistical Association”, Vol. 62, pp. 143-161, 1967.

Blackorby, C., Donaldson, D., Auersperg, M., *A New Procedure for the Measurement of Inequality Within and Among Population Subgroups: Ethical Consistent Subindices*, “Canadian Journal of Economics”, Vol. 14, pp 665-85, 1981.

Bourgignon, F., *Decomposable Income Inequality Measure*, “Econometrica”, Vol. 47, pp 901-920, 1979.

Cowell, F., *On the Structure of Additive Inequality Measures*, “Review of Economic Studies”, Vol. 47, pp. 521-531, 1980.

Dagum, C., *Inequality Measures between Income Distributions with Applications,* “Econometrica”, Vol. 48, pp. 1791-180, 1980.

Dagum, C., *A New Approach to the Decomposition of the Gini Income Inequality Ratio*, “Empirical Economics”, Vol. 22, pp. 551-531, 1997.

Ebert, U., *Dual Decomposable Inequality Measures*, “The Canadian Journal of Economics”, Vol. 32, pp. 234-246, 1999.

Giorgi, G., M., *Bibliographic portrait of the Gini concentration ratio*, "Metron", Vol. 48, pp. 183-221, 1990.

Giorgi, G., M., *A fresh look at the topical interest of the Gini concentration ratio*" "Metron", Vol. 51, pp.83-98, 1993.

IPiSS, *Wysokość i struktura minimum egzystencji w 2001 roku [Level and Structure of Minimum Survival Income in 2001]*, Instytut Pracy i Spraw Socjalnych, 2001.

Lambert, P., Aronson, J., R., *Inequality decomposition Analysis and the Gini Coefficient Revisited*, "The Economic Journal", Vol.103, pp. 1221-1227, 1993.

Mookherjee, D., Shorrock, A., *A Decomposition Analysis of the Trend in UK Income Inequality*, "The Economic Journal", Vol.92, pp. 886-902, 1982.

Pyatt, G., *On the Interpretation and Disagregations of Gini Coefficient*, "Economic Journal", Vol. 86, pp.243-255, 1976.

Rao, V., *The Decomposition of the Concentration Ratio*, "Journal of the Royal Statistical Society", Vol. 132, pp. 418-25, 1969.

Russel, R., *A Note on Additive Decomposable Inequality Measures*, "Review of Economic Studies", Vol. 52, pp. 347-352, 1985.

Shorrock, A., *The Class of Additively Decomposable Measures*, "Econometrica", Vol. 48, pp. 613-625, 1980.

Shorrock, A., *On the Distance between Income Distributions*, "Econometrica", Vol. 50, pp. 1337-1340, 1982.

Shorrock, A., *Inequality Decomposition by Population Subgroup*, "Econometrica", Vol. 52, pp. 1369-86, 1984.

Silber, J., *Factor Components, Population Subgroups and the Computation of the Gini Index of Inequality*, "The Review of Economics and Statistics", Vol. 71, pp 107-115, 1989.

Theil, H., *Economics and Information Theory*, Amsterdam, North-Holland, 1967.

Vernizzi, A., Monti, M., Kosny, M., *An Overall Inequality Reducing and Equitable Tax System with Application to Polish Data* in: Ostasiewicz, W. (ed.), "Toward Quality of Life Improvement". The Publishing House of Wrocław University of Economics, pp. 33-90, 2006.

Yitzahaki, S., Lerman, R., *Income Stratification and Income Inequality*, "Review of Income and Wealth", Vol. 37, pp. 313-329, 1991.

Yitzahaki, S., *Economic Distance and Overlapping of Distributions*, "Journal of Econometrics", Vol. 61, pp. 147-159, 1994.

Received: May 2007, revised version: September 2007