CrossMark

Overlapping component and inequality decomposition: a simulation study for the Gini index

Michele Costa[1] (ORCID)

Abstract This paper analyses the effects of overlapping subgroups on inequality decompositions. Within the framework of the Gini inequality index is performed a simulation study which allows to compare and evaluate alternative decompositions with respect to different degrees and situations of overlapping.

1 Introduction

Overlapping subgroups play a particular role in inequality decomposition since they give a complex and not easily assessable contribution to total inequality. With respect to the measurement of the traditional components of inequality decomposition, the inequality within subgroups and the inequality between subgroups, it is possible to observe relevant effects related to the presence of overlapping units.

First, overlapping units lead to an increase of the importance of inequality within subgroups on total inequality.

Second, they introduce some non trivial problems into the measurement of the inequality between subgroups.

Third, they pose relevant questions on the interpretation of inequality between subgroups.

Given a poor subgroup A and a non-poor subgroup B, we need to interpret the contribution to the inequality between subgroups given by a unit belonging to A, which is richer than a unit belonging to B. Or, analogously, we need to interpret the contribution to the inequality between subgroups given by a unit belonging to B, which is poorer than a unit belonging to A.

✉ Michele Costa
michele.costa@unibo.it

[1] Department of Statistical Sciences, University of Bologna, 40126 Bologna, Italy

🖄 Springer

A related point to be addressed is whether overlapping units belonging to A are to be classified as "poor" and whether overlapping units belonging to B are to be considered as "non-poor".

It follows that overlapping component requires a more detailed framework for both the measurement and the interpretation of inequality. In particular, in order to properly evaluate and explain total inequality, overlapping units have to be analyzed separately from non overlapping units.

Furthermore, the aim of inequality decompositions is generally related to the identification of the relevant factors determining the inequality structure. Given a possible inequality factor (such as gender, working condition, education level, area of residence, etc.) and having decomposed the inequality in terms of such factor, high levels of overlapping indicate that the factor only slightly contributes to total inequality, while low levels of overlapping suggest a stronger contribution. Therefore, overlapping analysis provides a powerful insight on inequality structure and also allows to assess the importance of the various factors on total inequality.

The case of overlapping subgroups has been deeply analyzed, using the term "transvariation", by Gini [13,14] and Dagum [4]. Their studies link transvariation to inequality measurement achieved by means of the Gini index [6,10,12,17]. The effects of overlapping component on inequality decomposition have been the focus of the interest for many other researchers (see [12,13,17]). The different ways to analyze and interpret these effects represent the main motivation of the various approaches to the Gini index decomposition.

The aim of the paper is to evaluate and compare alternative proposals for the measurement of overlapping component in the Gini index decomposition. Furthermore, we want to analyse the properties of these proposals with respect to different degrees of overlapping.

To this purpose we develop a Monte Carlo study which allows to observe and compare the various Gini index decompositions as well as to illustrate the effects related to increasing levels of overlapping.

The paper is organized as follows. Section 2 addresses the measurement of overlapping. Section 3 illustrates three different methodologies for the decomposition of the Gini index. Section 4 develops the simulation study and Sect. 5 concludes.

2 The measurement of overlapping

The general case of a population of n units disaggregated into k subgroups of size n_j, with $\sum_{j=1}^{k} n_j = n$, refers to subgroups characterized by some degree of overlapping. Let be y_{ji} the value of character y in the i-th unit of the j-th subgroup and, accordingly, y_{hr} the value of y in the r-th unit of the h-th subgroup.

Overlapping occurs when at least one of the differences $(y_{ji} - y_{hr})$, for $i = 1,\ldots, n_j$ and $r = 1,\ldots, n_h$, shows opposite sign to the difference $(\lambda_j - \lambda_h)$, where λ_j and λ_h are mean values (usually the arithmetic mean or the median) respectively for subgroups j and h.

In order to evaluate the relevance of overlapping, Gini introduced two indexes: the probability of transvariation and the intensity of transvariation [14].

Probability of transvariation refers to overlapping which occurs when $y_{ji} < y_{hr}$ and the median of the j-th subgroup, $_{me}\bar{y}_j$, is greater or equal to the median of the h-th subgroup, $_{me}\bar{y}_h$.

Null differences $(y_{ji} - y_{hr})$ are equally divided between overlapping and non-overlapping differences. The probability of transvariation between subgroups j and h, $p(t_{jh})$, is then computed as the ratio between the number of overlapping differences and its maximum; that is:

$$p(t_{jh}) = (_1nt_{jh} +_2 nt_{jh})/n_j n_h \tag{1}$$

where $_1nt_{jh}$ and $_2nt_{jh}$ are, respectively, the number of negative and the number of null differences $(y_{ji} - y_{hr})$ for $i = 1,\ldots, n_j$ and $r = 1,\ldots, n_h$, with $_{me}\bar{y}_j \geq_{me} \bar{y}_h$.

The maximum of $p(t_{jh})$ is reached when the two subgroups completely overlap and the median of the j-th subgroup equals the median of the h-th subgroup. The probability of transvariation ranges between 0, when there is no transvariation and no difference is overlapping, and 1, when overlapping component reaches its maximum.

The second index introduced by Gini, the so called intensity of transvariation, refers to overlapping which occurs when $y_{ji} < y_{hr}$ while the mean of the j-th subgroup, $\bar{y}_j$, is greater or equal to the mean of the h-th subgroup, $\bar{y}_h$.

With respect to the probability of transvariation, which takes the number of overlapping differences into account, the intensity of transvariation is based on their size, that is on the quantities $|y_{ji} - y_{hr}|$ for $y_{ji} < y_{hr}$, $\bar{y}_j \geq \bar{y}_h$, $i = 1,\ldots, n_j$ and $r = 1, n_h$.

Let T_{jh} be the sum of overlapping differences between subgroups j and h: $T_{jh} = \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}|$, for $y_{ji} < y_{hr}$ and $\bar{y}_j \geq \bar{y}_h$.

It is immediate to notice that T_{jh} increases as the difference $(\bar{y}_j - \bar{y}_h)$ is decreasing. Furthermore, for $\bar{y}_j = \bar{y}_h$, T_{jh} reaches its maximum, that is

$$_{\max}T_{jh} = \frac{1}{2} \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}| \quad .$$

The intensity of transvariation is obtained as the ratio between the sum of overlapping differences $|y_{ji} - y_{hr}|$ and its maximum; that is:

$$i(t_{jh}) = 2T_{jh}/\sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}|. \tag{2}$$

Intensity of transvariation ranges between 0, when no difference is overlapping, and 1, when the two subgroups completely overlap.

By jointly using intensity and probability of transvariation it is possible to obtain a wide information set about the relevance and the extent of the overlapping component.

3 Overlapping and the Gini index decomposition

The role of the overlapping component is analyzed within the framework of the Gini index which, in a population disaggregated into k subgroups, can be expressed as

$$G = \frac{1}{2n^2\bar{y}} \sum_{j=1}^{k} \sum_{h=1}^{k} \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}| \tag{3}$$

where $\bar{y}$ is the arithmetic mean of y in the overall population.

In the extant literature on the Gini index decomposition, the component of inequality within subgroups is generally measured as

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

where G_{jj} is the Gini index in the j-th subgroup, $p_j = n_j/n$ is the share of subgroup j in the overall population and $s_j = (\sum_{i=1}^{n_j} y_i n_i / \sum_{i=1}^{n} y_i n_i)$ is the share of subgroup j in the overall y.

The measurement of inequality within subgroups based on (4) is not exempt from criticism [12]. However, in the following the topic of inequality within will not addressed. Conversely, we will focus on the problem which mainly motivates the controversial and articulate debate on the Gini index decomposition: the measurement of the inequality between subgroups.

3.1 Bhattacharia and Mahalanobis's Gini index decomposition

Bhattacharia and Mahalanobis's [1] seminal proposal for the measurement of inequality between G_b, given by

$$_{BM}G_b = \sum_{j=1}^{k} \sum_{\substack{h=1 \\ j \neq h}}^{k} p_j p_h |\bar{y}_j - \bar{y}_h|/2\bar{y}$$

starts a tradition in which differences between subgroups are measured on the basis of the difference between subgroups means.

Following the contribution by Bhattacharia and Mahalanobis, many authors (see [15], and references therein) suggested further measures for evaluating the inequality between subgroups. In all cases, however, the core of each expression is still represented by the difference between subgroups means, as in Bhatthacharia and Mahalanobis.

This approach to the measurement of G_b does not present particular disadvantages when the subgroups are not overlapping, but, in the general case of overlapping units, it leads to the presence of a third term, $_{BM}G_t$, which acts as a "residual" in the Gini index decomposition:

$$G = G_w +_{BM} G_b +_{BM} G_t \qquad (5)$$

The minimum value of $_{BM}G_b$ is 0 and it occurs when $\bar{y}_j = \bar{y}$ $\forall j$. The component $_{BM}G_t$ ranges between 0, when the subgroups are not overlapping, and $(G - G_w)$, when $\bar{y}_j = \bar{y}$ $\forall j$ and $_{BM}G_b = 0$.

By using (5), Shorrocks [19] classifies the Gini index as a non-additively decomposable measure, thus shifting the preferences of many researchers toward generalized entropy indexes, which are additively decomposable.

3.2 Yitzhaki and Lerman's Gini index decomposition

The topic of overlapping subgroups in the Gini index decomposition is addressed in 1991 by Yitzhaki and Lerman [21], who extensively contributed to the development of the studies about the Gini index and its decomposition. In 1991 they introduce a specific overlapping index, O_{jh}, which measures the degree according to which the distribution of the j-th subgroup is included in the range of the distribution of the h-th subgroup.

O_{jh} is defined as

$$O_{jh} = cov_h(y, F_j(y))/cov_h(y, F_h(y))$$

where $F_j(y)$ is the cumulative distribution, which, in the sample, is estimated by the rank of the observation. The denominator of O_{jh} is the covariance between incomes of subgroup h and their rank, while the numerator of O_{jh} is the covariance between incomes of subgroup h and their rank, as belonging to subgroup j.

Following Yitzhaki and Lerman [21] and Yitzhaki [20], the inequality between subgroups is measured as

$$_{YL}G_b = 2cov(\bar{y}_j, \bar{F}_j)/\bar{y}$$

where $\bar{F}_j = \bar{R}_j/n$ and $\bar{R}_j$ is the average rank of group j in the overall population.

In summary, the Gini index decomposition proposed by Yitzhaki and Lerman is

$$G = G_w + _{YL}G_b + \sum_{h=1}^{k} s_h G_h \sum_{j=1, j \neq h}^{k} p_j O_{jh} = G_w + _{YL}G_b + _{YL}G_t \qquad (6)$$

The minimum value of $_{YL}G_b$ is 0 and it occurs when $\bar{y}_j = \bar{y}$ $\forall j$; the component $_{YL}G_t$ ranges between 0, when there is no overlapping, and $(G - G_w)$, when $\bar{y}_j = \bar{y}$ $\forall j$ and $_{YL}G_b$ = 0.

The approach by Yitzhaki and Lerman is further developed by Frick et al. [11], who obtained new results on the Gini index decomposition and overlapping measurement by providing an interesting interpretation of overlapping in terms of the inverse of stratification.

3.3 Dagum's Gini index decomposition

A second contribution to the analysis of overlapping subgroups in the Gini index decomposition is proposed by Dagum [8,9], who extends a previous study by Mehran [18]. In Dagum's proposal, both the inequality between subgroups and the contribution of overlapping units can be evaluated on the basis of two quantities: the Gini index between subgroups j and h, G_{jh}, and the economic relative distance D_{jh}. First, starting from the Gini index for the subgroup j, G_{jj},

$$G_{jj} = \frac{1}{2n_j^2 \bar{y}_j} \sum_{i=1}^{n_j} \sum_{r=1}^{n_j} |y_{ji} - y_{jr}|,$$

the Gini index between subgroup j and h, G_{jh}, can be simply expressed as

$$G_{jh} = \frac{1}{n_j n_h (\bar{y}_j + \bar{y}_h)} \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}|.$$

Second, the economic relative distance D_{jh} [5,7] is

$$D_{jh} = (d_{jh} - p_{jh})/(d_{jh} + p_{jh})$$

where d_{jh} and p_{jh} are, respectively, the gross economic affluence and the first order moment of transvariation between the j-th and the h-th subgroups. Given $\bar{y}_j \geq \bar{y}_h$, the gross economic affluence between the j-th and the h-th subpgroups, d_{jh}, is a weighted average of the differences $(y_{ji} - y_{hr})$ for all $y_{ji} > y_{hr}$. Following Dagum [5,7], d_{jh} can be expressed as

$$d_{jh} = E_j(y F_h(y)) + E_h(y F_j(y)) - E_h(y)$$

where the subscripts indicate the subgroups j and h, $F(y)$ is the cumulative distribution function and E stands for the mathematical expectation operator. Furthermore, the first-order moment of transvariation between the j-th and the h-th subgroups, p_{jh}, is a weighted average of the differences $(y_{ji} - y_{hr})$ for all $y_{ji} < y_{hr}$. Following Dagum [5,7], p_{jh} can be expressed as

$$p_{jh} = E_j(y F_h(y)) + E_h(y F_j(y)) - E_j(y).$$

In this framework, inequality between subgroups is measured as

$$_DG_b = \sum_{j=1}^{k} \sum_{h=1, j\neq h}^{k} G_{jh} D_{jh} p_j s_h$$

while the contribution to total inequality given by overlapping units is evaluated by means of

$$_DG_t = \sum_{j=1}^{k} \sum_{h=1, j\neq h}^{k} G_{jh}(1 - D_{jh}) p_j s_h.$$

When $\bar{y}_j = \bar{y} \;\; \forall j$, the differences $(y_{ji} - y_{hr})$ for all $y_{ji} > y_{hr}$ equal the differences $(y_{ji} - y_{hr})$ for all $y_{ji} < y_{hr}$, that is, when $\bar{y}_j = \bar{y} \;\; \forall j$, the component $_DG_b$ equals the component $_DG_t$. The minimum value of $_DG_b$ is then $(G - G_w)/2$ and the component $_DG_t$ ranges between 0 and $(G - G_w)/2$.

On the whole, Dagum's Gini index decomposition is given by

$$G = G_w + {}_DG_b + {}_DG_t \tag{7}$$

It is straightforward to note how both G_{jh} and D_{jh} involve an heavy computational effort. In order to overcome this problem a simplified version of Dagum's contribution is developed by Costa [2,3] as

$$_DG_b = {}_DG_b^* + 0.5(G - G_w - {}_DG_b^*)$$

and

$$_DG_t = 0.5(G - G_w - {}_DG_b^*)$$

where

$$G_b^* = \sum_{j=1}^{k-1} \sum_{h=j+1}^{k} \frac{p_{hj}^* - s_{hj}^*}{p_{hj}^* s_{jh}^* + p_{jh}^* s_{hj}^*} (p_j s_h + p_h s_j),$$

$$p_{hj}^* = p_h/(p_h + p_j),$$

$$s_{hj}^* = s_h/(s_h + s_j).$$

It is also worth noting how the intuition behind Dagum's Gini index decomposition is both extremely simple and appealing: each difference $|y_{ji} - y_{hr}|$ in expression (3) is attributed to inequality within subgroups for $j = h$, to inequality between subgroups for $j \neq h$, $\bar{y}_j \geq \bar{y}_h$ and $y_{ji} \geq y_{hr}$, and to transvariation for $j \neq h$, $\bar{y}_j \geq \bar{y}_h$ and $y_{ji} < y_{hr}$.

3.4 The comparison among Gini index decompositions

The main difference among Gini index decompositions is the measurement of inequality between G_b and it can be illustrated by analyzing the case of perfectly overlapping subgroups.

Following Bhattacharia and Mahalanobis [1], and Yitzhaki and Lerman [21], complete overlapping implies a null inequality between, $G_b = 0$, while all differences among subgroups distributions are evaluated by means of the transvariation component,

$G_t = G - G_w$.

Conversely, within Dagum's approach, a complete overlapping leads to equally assign $(G - G_w)$ to inequality between G_b and to transvariation component G_t:

$G_t = G_b = (G - G_w)/2$.

It is possible to identify a direct relationship between the decompositions introduced by Dagum and by Bhattacharia-Mahalanobis. The link between the two proposals is given by the measurement of the inequality between subgroups: in particular, $_{BM}G_b$ is obtained as the sum of all differences $(y_{ji} - y_{hr})$, $\forall j, h, i, r$, with $\bar{y}_j > \bar{y}_h$, while $_D G_b$ corresponds to the sum of all differences $|y_{ji} - y_{hr}|$, $\forall j, h, i, r$, with $\bar{y}_j > \bar{y}_h$. If differences $(y_{ji} - y_{hr})$ are non negative—that is, if there is no transvariation—$_{BM}G_b = {_D}G_b$, while, in the case of overlapping subgroups, $_{BM}G_b < {_D}G_b$ and $(_D G_b - {_{BM}}G_b)$ is given by the negative differences $(y_{ji} - y_{hr})$, that is by $_D G_t$. It follows that $_D G_b = {_{BM}} G_b + {_D} G_t$, which implies $_D G_t = 0.5_{BM} G_t$.

Given the characteristics of the different decompositions with respect to the overlapping component measurement, the approaches by Bhattacharya and Mahalanobis [1] and by Yitzhaki and Lerman [21] could be more useful in an exploratory context, where we need to assess the existence of inequality factors. In contrast, Dagum's [8] proposal could be more adequate in a confirmative framework, where the inequality factors are already known but it is necessary to assess their contribution to total inequality.

In order to evaluate the behaviour of the various Gini index decompositions with respect to the presence of overlapping we propose a two step procedure. First, we normalize the transvariation components $_{BM}G_t$, $_{YL}G_t$ and $_D G_t$ by dividing each quantity by its maximum, thus obtaining

$$_{BM}G_t^* = {_{BM}}G_t/(G - {-}G_w),$$
$$_{YL}G_t^* = {_{YL}}G_t/(G - {-}G_w),$$
$$_D G_t^* = 2_D G_t/(G - {-}G_w).$$

Second, we compare the normalized transvariation components to the probability and the intensity of transvariation by calculating the absolute errors $e_{p(t)}$ and $e_{i(t)}$ where

$$e_{p(t)} = |G_t^* - p(t)| \tag{8}$$

and

$$e_{i(t)} = |G_t^* - i(t)| \tag{9}$$

By means of expressions (8) and (9) we are able to evaluate the goodness of fit of the various decompositions with respect to overlapping component. High values of $e_{p(t)}$ and $e_{i(t)}$ indicate the existence of a relevant distance between G_t^* and the measure of transvariation, thus suggesting that the decomposition fails to correctly evaluate the overlapping component. In contrast, low values of $e_{p(t)}$ and $e_{i(t)}$ point towards a good performance of the decomposition in the measurement of the overlapping component.

Within the analysis of (8) and (9), it is also relevant to observe how the result $_D G_t = 0.5_{BM} G_t$ implies $_D G_t^* = {_{BM}} G_t^*$ and, therefore, $_D e_{p(t)} = {_{BM}} e_{p(t)}$ and $_D e_{i(t)} = {_{BM}} e_{i(t)}$.

In the following we develop a simulation study which provides m replications of the Gini index decomposition. In this case, starting from $e_{p(t)}$ and $e_{i(t)}$, it is possible to obtain the mean absolute errors $E_{p(t)}$ and $E_{i(t)}$ as

$$E_{p(t)} = \frac{1}{m} \sum_{j=1}^{m} |G_{tj}^* - p(t)_j| \quad and \tag{10}$$

Table 1 Frequency distribution of the probability of transvariation in the simulated samples by number k of subgroups

p(t)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Beta										
$k = 2$	0.01	0.04	0.07	0.09	0.11	0.13	0.14	0.15	0.15	0.12
$k = 3$	0.00	0.01	0.04	0.08	0.14	0.18	0.20	0.18	0.13	0.04
$k = 4$	0.00	0.00	0.01	0.06	0.16	0.23	0.25	0.19	0.08	0.01
Gamma										
$k = 2$	0.43	0.11	0.08	0.07	0.06	0.06	0.05	0.05	0.05	0.03
$k = 3$	0.25	0.20	0.16	0.12	0.10	0.07	0.05	0.03	0.01	0.01
$k = 4$	0.11	0.24	0.23	0.17	0.11	0.07	0.04	0.02	0.01	0.01

$$E_{i(t)} = \frac{1}{m} \sum_{j=1}^{m} |G_{tj}^* - i(t)_j|. \tag{11}$$

Furthermore, the comparison of the Gini index decompositions is performed also on the basis of the ratios $_{BM}G_t/G$, $_{YL}G_t/G$, and $_DG_t/G$, which are analyzed with respect to the ratio G_w/G, and to the probability and the intensity of transvariation.

4 The simulation study

The aim of the Monte Carlo study developed in this section is to evaluate and compare the alternative Gini index decompositions with respect to the presence of transvariation.

Both probability and intensity of transvariation, together with three different Gini index decompositions [(expressions (5) by Bhattacharya and Mahalanobis, (6) by Yitzhaki and Lerman and (7) by Dagum], are computed on simulated samples, which are randomly extracted from beta or gamma distribution with the purpose to cover a wide range of different situations.

We consider the cases of $k = 2, 3, 4$ for the number of subgroups and $n = 100, 500, 1000, 5000$ and 10,000 for the sample size. Subgroup's size n_j, $j = 1, \ldots, k$, is randomly selected with the constraints $\sum_{j=1}^{k} n_j = n$ and $n_j \geq 50$. Only for sample size $n = 100$, the minimum value of n_j is set to 10.

For each subgroup, beta distribution parameters $B(a, b)$ are randomly selected with $0.5 \leq a \leq 2$ and $0.5 \leq b \leq 4$, while gamma distribution parameters $\Gamma(c, d)$ are randomly selected with $0.5 \leq c \leq 10$ and $0.5 \leq d \leq 10$.

For each combination of k and n, 20.000 samples are randomly generated, 50 % from the beta distribution and 50 % from the gamma distribution, for a total of $m = 300.000$ samples.

Table 1 reports the frequency distribution of the probability of transvariation by number k of subgroups. Each row illustrates the results related to 50000 samples. For example, from the first row it is possible to observe how, within the 50000 samples generated from the beta distribution and divided in $k = 2$ subgroups, we obtain a probability of transvariation lesser than 0.2 only in the 5 % of cases, while a probability of transvariation greater than 0.8 occurs in the 27 % of cases. Results are summarized for the 5 sample sizes used in the simulation study ($n = 100, 500, 1000, 5000$ and 10.000), since different values of n lead to the same classification of the simulated samples by both probability and intensity of transvariation.

Table 2 Frequency distribution of the intensity of transvariation in the simulated samples by number k of subgroups

$i(t)$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Beta										
$k = 2$	0.02	0.08	0.11	0.12	0.12	0.12	0.12	0.12	0.11	0.07
$k = 3$	0.00	0.03	0.09	0.15	0.18	0.19	0.16	0.12	0.06	0.01
$k = 4$	0.00	0.01	0.06	0.16	0.24	0.24	0.18	0.10	0.03	0.01
Gamma										
$k = 2$	0.53	0.11	0.08	0.06	0.05	0.04	0.04	0.03	0.03	0.02
$k = 3$	0.44	0.19	0.13	0.09	0.06	0.04	0.03	0.01	0.01	0.00
$k = 4$	0.31	0.28	0.18	0.11	0.06	0.03	0.02	0.01	0.01	0.00

It is possible to observe how the simulated samples extracted from gamma distribution are concentrated on the lower values of probability of transvariation, while in the simulated samples extracted from beta distribution medium and high values of probability of transvariation are more frequent. Furthermore, as expected, by increasing the number of subgroups k, frequencies shift towards higher levels of transvariation.

The frequency distribution of the intensity of transvariation in the simulated samples by number k of subgroups is reported on Table 2, where it is possible to find a pattern only slightly different from Table 1.

As expected, probability and intensity of transvariation have a similar effect on the distribution of the simulated samples.

Results related to the Gini index decompositions are summarized on Table 3 for probability of transvariation and on Table 4 for intensity of transvariation by means of the ratios $G_w/G, _{BM}G_t/G, _{YL}G_t/G$ and $_DG_t/G$. Each row refers to 100,000 samples divided in k subgroups.

The ratio G_w/G evaluates the weight of inequality within in total inequality: for example, for $k = 2$, G_w/G starts from 40 % when overlapping is weak, and it increases to over 60 % for high levels of overlapping. The number k of subgroups strongly influence G_w/G, which decreases by increasing k.

The ratios $_{BM}G_t/G, _{YL}G_t/G$ and $_DG_t/G$ illustrate the dynamic of transvariation component by increasing levels of overlapping. In the first column of Table 3, when overlapping is absent or weak, all decompositions lead obviously to the same results, with $_{BM}G_t \approx _{YL}G_t \approx _DG_t \approx 0$. The main difference between the decomposition can be clearly observed in the last column of Table 3: when overlapping is complete (or almost complete), we have, for Bhattacharya-Mahalanobis and Yitzhaki-Lerman, $G_w/G + _{BM}G_t/G \approx G_w/G + _{YL}G_t/G \approx 1$, while, for Dagum, it holds $G_w/G + 2*_DG_t/G \approx 1$. From Table 3, by comparing $_{BM}G_t/G$ and $_DG_t/G$, it is also possible to find the result $_DG_t = 0.5 _{BM}G_t$ derived in paragraph 3.4. Furthermore, it can be observed how $_{BM}G_t/G < _{YL}G_t/G$, and, given $_DG_t = 0.5 _{BM}G_t$, we have $_DG_t/G < _{BM}G_t/G < _{YL}G_t/G$.

The analysis of the ratios $G_w/G, _{BM}G_t/G, _{YL}G_t/G$ and $_DG_t/G$ by intensity of transvariation (Table 4) leads to different values but identical conclusion with respect to the analysis performed by referring to probability of transvariation.

Furthermore, for both probability and intensity of transvariation, sample size n does not influence the ratios $G_w/G, _{BM}G_t/G, _{YL}G_t/G$ and $_DG_t/G$.

Table 3 Decomposition by Bhattacharya and Mahalanobis, Yitzhaki and Lerman, Dagum and by probability of transvariation

$p(t)$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
G_w/G										
$k=2$	0.40	0.48	0.52	0.55	0.57	0.59	0.61	0.62	0.63	0.63
$k=3$	0.24	0.29	0.31	0.33	0.36	0.39	0.42	0.44	0.45	0.46
$k=4$	0.11	0.21	0.22	0.24	0.26	0.29	0.31	0.33	0.35	0.35
$_{BM}G_t/G$										
$k=2$	0.01	0.03	0.05	0.08	0.11	0.15	0.19	0.24	0.29	0.33
$k=3$	0.01	0.03	0.07	0.11	0.16	0.21	0.27	0.33	0.41	0.46
$k=4$	0.01	0.04	0.08	0.12	0.18	0.24	0.32	0.40	0.48	0.55
$_{YL}G_t/G$										
$k=2$	0.04	0.10	0.16	0.21	0.25	0.29	0.32	0.35	0.36	0.37
$k=3$	0.04	0.10	0.17	0.24	0.31	0.38	0.44	0.48	0.51	0.53
$k=4$	0.02	0.10	0.18	0.26	0.34	0.42	0.50	0.56	0.60	0.63
$_{D}G_t/G$										
$k=2$	0.01	0.01	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16
$k=3$	0.01	0.02	0.03	0.05	0.08	0.11	0.14	0.17	0.20	0.23
$k=4$	0.01	0.02	0.04	0.06	0.09	0.12	0.16	0.20	0.24	0.28

Table 4 Decomposition by Bhattacharya and Mahalanobis, Yitzhaki and Lerman, Dagum and by intensity of transvariation

$i(t)$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
G_w/G										
$k=2$	0.42	0.51	0.56	0.58	0.60	0.61	0.62	0.63	0.63	0.63
$k=3$	0.25	0.31	0.34	0.37	0.40	0.42	0.43	0.45	0.46	0.47
$k=4$	0.18	0.23	0.24	0.27	0.29	0.31	0.33	0.35	0.36	0.33
$_{BM}G_t/G$										
$k=2$	0.02	0.05	0.08	0.12	0.16	0.20	0.24	0.27	0.31	0.35
$k=3$	0.02	0.06	0.11	0.16	0.22	0.28	0.33	0.39	0.45	0.50
$k=4$	0.02	0.06	0.12	0.18	0.25	0.32	0.39	0.46	0.52	0.52
$_{YL}G_t/G$										
$k=2$	0.06	0.15	0.22	0.26	0.30	0.33	0.34	0.36	0.37	0.37
$k=3$	0.06	0.16	0.25	0.32	0.38	0.43	0.47	0.50	0.52	0.53
$k=4$	0.06	0.17	0.26	0.35	0.43	0.50	0.55	0.59	0.61	0.56
$_{D}G_t/G$										
$k=2$	0.02	0.05	0.08	0.12	0.16	0.20	0.24	0.27	0.31	0.35
$k=3$	0.01	0.03	0.05	0.08	0.11	0.14	0.17	0.20	0.22	0.25
$k=4$	0.01	0.03	0.06	0.09	0.12	0.16	0.20	0.23	0.26	0.26

On the whole, within the Dagum's approach the overlapping component plays a minor role both with respect to other inequalities components and in comparison to overlapping in other decompositions. On the contrary, within Yitzhaki and Lerman's decomposition,

Table 5 Results of the simulation study: mean absolute error $E_{p(t)}$ by probability of transvariation $p(t)$ and number of subgroups k

p(t)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Bhattacharya Mahalanobis										
$k = 2$	0.04	0.10	0.15	0.18	0.19	0.19	0.17	0.13	0.10	0.07
$k = 3$	0.05	0.11	0.16	0.19	0.20	0.20	0.18	0.15	0.10	0.06
$k = 4$	0.03	0.11	0.16	0.19	0.20	0.20	0.18	0.15	0.11	0.08
Yitzhaki Lerman										
$k = 2$	0.01	0.04	0.08	0.11	0.15	0.17	0.17	0.16	0.11	0.05
$k = 3$	0.01	0.03	0.05	0.08	0.10	0.12	0.13	0.13	0.10	0.06
$k = 4$	0.01	0.04	0.05	0.06	0.08	0.10	0.12	0.12	0.10	0.06
Dagum										
$k = 2$	0.04	0.10	0.15	0.18	0.19	0.19	0.17	0.13	0.10	0.07
$k = 3$	0.05	0.11	0.16	0.19	0.20	0.20	0.18	0.15	0.10	0.06
$k = 4$	0.03	0.11	0.16	0.19	0.20	0.20	0.18	0.15	0.11	0.08

Table 6 Results of the simulation study: mean absolute error $E_{i(t)}$ by intensity of transvariation $i(t)$ and number of subgroups k

i(t)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Bhattacharya Mahalanobis										
$k = 2$	0.02	0.06	0.07	0.07	0.06	0.05	0.03	0.02	0.01	0.00
$k = 3$	0.03	0.07	0.09	0.09	0.09	0.07	0.05	0.04	0.02	0.01
$k = 4$	0.03	0.07	0.09	0.10	0.09	0.08	0.06	0.04	0.02	0.01
Yitzhaki Lerman										
$k = 2$	0.06	0.17	0.24	0.28	0.29	0.28	0.25	0.20	0.13	0.05
$k = 3$	0.04	0.09	0.14	0.18	0.20	0.21	0.20	0.17	0.12	0.06
$k = 4$	0.02	0.06	0.10	0.14	0.17	0.19	0.18	0.16	0.12	0.07
Dagum										
$k = 2$	0.02	0.06	0.07	0.07	0.06	0.05	0.03	0.02	0.01	0.00
$k = 3$	0.03	0.07	0.09	0.09	0.09	0.07	0.05	0.04	0.02	0.01
$k = 4$	0.03	0.07	0.09	0.10	0.09	0.08	0.06	0.04	0.02	0.01

overlapping component becomes a major source of total inequality, while in Bhattacharya and Mahalnobis's approach it plays an intermediate role.

The comparison among alternative Gini index decompositions can be performed by means of the mean absolute errors $E_{p(t)}$ and $E_{i(t)}$, calculated as in (10) and (11), which are illustrated on Tables 5, 6 for probability and intensity of transvariation, respectively. Each row refers to 100,000 samples divided in k subgroups.

The first and the last columns of Table 5 illustrate the results related to the two extreme situations: absence of overlapping and complete overlapping, respectively. In both cases we can observe low values of $E_{p(t)}$, thus indicating a good performance of the different decompositions in the overlapping component measurement.

Furthermore, by comparing Bhattacharya and Mahalanobis's decompositions to Dagum's approach, we find the result $_De_{p(t)} = _{BM}e_{p(t)}$ demonstrated in Sect. 3.4.

Finally, it is interesting to note how, by referring to probability of transvariation, Yitzhaki and Lerman's approach leads to the lowest values of the mean absolute error.

The analysis of the mean absolute error by intensity of transvariation (Table 6) validates the indications of $E_{p(t)}$ related to the two extreme situations (first and last columns still shows really low values of $E_{i(t)}$) and to the comparison between Bhattacharya-Mahalnobis's and Dagum's approaches ($_De_{i(t)} = _{BM}e_{i(t)}$).

However, in Table 6 it is also possible to observe a relevant difference with respect to Table 5: Yitzhaki and Lerman's approach shows the highest values of $E_{i(t)}$, thus inverting the previous result based on $E_{p(t)}$.

Furthermore, the Gini index decompositions performed on the simulated data also indicates that samples extracted from beta distribution and samples extracted from gamma distribution show a similar pattern with respect to the role of the overlapping component. Given the strong diversity among sample distributions this result ensures an high level of robustness to our study.

Finally, a last result refers to the number of observations n, which seems to not influence the overlapping component within the Gini index decomposition: sample sizes from 100 to 10,000 lead to the same pattern.

5 Concluding remarks

This paper developed a simulation study aimed to analyze and compare different Gini index decompositions. Our focus is on the role of overlapping component, evaluated by means of both probability and intensity of transvariation.

The analysis of the characteristics of the different decompositions with respect to the overlapping component measurement leads to suggest that the approaches by Bhattacharya–Mahalanobis and Yitzhaki–Lerman could be more useful in an exploratory context, where we need to assess the existence of inequality factors, while Dagum's proposal could be more adequate in a confirmative framework, where the inequality factors are already known, but it is necessary to assess their contribution to total inequality.

Results of the simulation study suggest that the number of observations does not influence the overlapping component, while the number of subgroups plays a key role. We also provide evidence about the role of inequality within in total inequality by increasing levels of overlapping and compare different methods for overlapping component measurement.

On the basis of probability of transvariation, that is by considering the frequency of overlapping, Yitzhaki and Lerman's proposal leads to the lowest mean absolute error, thus indicating the best performance in the overlapping component measurement. On the contrary, by referring to intensity of transvariation, that is by considering the extent of overlapping, the Dagum's approach allows to obtain better results into the measurement of G_t.

Finally, the analysis of overlapping suggests appealing results also in the context of empirical studies, where it can be extremely helpful into the assessment of the "poverty" degree of the different units and into the investigation of the inequality structure.

References

1. Bhattacharya, N., Mahalanobis, B.: Regional disparities in household consumption in India. J. Am. Stat. Assoc. **62**(317), 143–161 (1967)

2. Costa, M.: Gini index decomposition for the case of two subgroups. Commun. Stat. Simul. Comput. **37**(4), 631–644 (2008)
3. Costa, M.: Transvariation and inequality between subpopulations in the Dagum's Gini index decomposition. Metron **67**(3), 229–241 (2009)
4. Dagum, C.: Transvariazione tra più di due distribuzioni. In: Gini, C. (ed.): Memorie di Metodologia Statistica. Libreria Goliardica, Roma (1959)
5. Dagum, C.: Inequality measures between income distributions with applications. Econometrica **48**(7), 1791–1803 (1980)
6. Dagum, C.: Gini ratio. In: Eatwell, J., et al. (eds.) The New Palgrave Dictionary of Economics. Mac Millian Press, London (1987a)
7. Dagum, C.: Measuring the economic affluence between populations of income receivers. J. Bus. Eco. Stat. **5**(1), 5–12 (1987b)
8. Dagum, C.: A new decomposition of the Gini income inequality ratio. Emp. Eco. **22**(4), 515–531 (1997a)
9. Dagum, C.: Decomposition and interpretation of Gini and Theil entropy inequality measures. In: American Statistical Association (ed.): Proceedings of the Business and Economic Statistics Section (1997b)
10. Deutsch, J., Silber, J.: Gini's "Transvariazione" and the measurement of distance between distributions. Empir. Econ. **22**(4), 547–554 (1997)
11. Frick, J., Goebel, J., Schechtman, E., Wagner, G., Yitzhaki, S.: Using analysis of Gini for detecting whether two subsamples represent the same universe: the German socio-economic panel study experience. Soc. Methods Res. **34**(4), 427–468 (2006)
12. Frosini, B.V.: Approximation and decomposition of Gini, Pietra-Ricci and Theil inequality measures. Empir. Econ. **43**(1), 175–197 (2012)
13. Gini, C.: Sulla misura della concentrazione e della variabilità dei caratteri. Atti R Ist Veneto Sci Lett Arti **73**(2), 1203–1248 (1914) (**English translation in Metron 63(1), 3–38 (2005)**)
14. Gini, C.: Transvariazione. Libreria goliardica, Roma (1959)
15. Giorgi, G.M.: Il rapporto di concentrazione di Gini. Genesi, evoluzione ed una bibliografia commentata. Libreria Editrice Ticci, Siena (1992)
16. Giorgi, G.M.: Gini's scientific work: an evergreen. Metron. **63**(3), 299–315 (2005)
17. Giorgi, G.M.: The Gini inequality index decomposition. An evolutionary study. In: Deutsch, J., Silber, J. (eds): The Measurement of Individual Well-Being and Group Inequalities. Routledge, London (2011)
18. Mehran, F.: A statistical analysis of income inequality based on a decomposition of the Gini index. In: International Statistical Institute. Proceedings of the 40th ISI Session, Warsaw (1975)
19. Shorrocks, A.F.: The class of additively decomposable inequality measures. Econometrica **48**(3), 613–625 (1980)
20. Yitzhaki, S.: Economic distance and overlapping distributions. J. Econ. **61**(1), 147–159 (1994)
21. Yitzhaki, S., Lerman, R.: Income stratification and income inequality. Rev. Income Wealth **37**(3), 313–329 (1991)