The Gini index decomposition and the overlapping between population subgroups

Published Version:
Costa, M. (2021). The Gini index decomposition and the overlapping between population subgroups. London : Routledge CRC Press.

Availability:
This version is available at: https://hdl.handle.net/11585/835203 since: 2025-01-10

Published:
DOI: http://doi.org/

The Gini index decomposition and the overlapping between population subgroups

Michele Costa
Department of Economics, University of Bologna,
Bologna, Italy

Abstract: This paper illustrates two contributions given by Gini to the study of inequality: the famous Gini inequality index and the analysis of overlapping between subgroups. The link between these topics is provided by the decomposition of the Gini index and the measurement of the contribution to total inequality attributable to the differences between subgroups of the population. These differences can be evaluated on the basis of the subgroup means or by referring to all the characteristics of the subgroup distributions: we evaluate and compare the two approaches, stressing strengths and weaknesses of both. The usefulness of the joint use of Gini index decomposition and overlapping evaluation is demonstrated in the analysis of gender gap, complemented with a case study related to the Italian personal income by gender.

Keywords: Gender income inequality; Gini index; Inequality between subgroups; Inequality decomposition; Overlapping; Transvariation.

Subject Classifications: 91B82; 91C99; 62A01; 62P20.

1. INTRODUCTION

A century after the introduction of the Gini index, the topics related to inequality measurement are central in the political and economic debate. Financial crises are over quite all around the world, aggregate production and income are at the pre-crisis levels, stock markets indexes hit new records, but the mood is far distant from the enthusiasm which usually characterizes post crises periods. A persistent gloom hampers entire countries, insecurity is a global and dominant feeling, protectionisms condition world trade, nationalisms are at the rising, countries like UK prefer individual to common path.

Inequality, a rising inequality, is the key to reconcile these two situations. Production, income and wealth are grown, but their growth is unequally shared. If we look only at the GDP, or at the DowJones index, we don't get a fully informative picture, while the basic information set should be improved by adding also an inequality measure.

The debate about GDP, the contributions by the Sarkozy commission, or similar committees, the proposals related to alternative GDP evaluation systems are an important, but still insufficient step. Starting from the daily public information provided by newspapers and televisions, and moving to all the actors of the society, up to the targets of the national and supranational political strategies and to the data elaborated by the statistical agencies, we need to include also inequality at the first positions of our priorities.

An increase of the GDP is a good news only if its growth is not too unequally shared, while the combination of an increase of both GDP and inequality would represent a bad news for most people.

Address correspondence to M. Costa, Department of Economics, University of Bologna, Piazza Scaravilli 2, 40126 Bologna, Italy; E-mail: michele.costa@unibo.it

If the overall inequality level represents a fundamental reference, the information in it included can be successfully exploited by means of inequality decomposition. Inequality decomposition allows to provide insights about the inequality structure, ranking the different inequality factors and assessing their relevance. At this regard the aspect of interest refers to the measurement of the inequality between subgroups, for which the literature presents many contributions, thus requiring a critical comparison.

An issue strictly connected to the inequality between subgroups is the overlapping between subgroups, a topic which Gini studied with his usual strong interest and determined motivation. While in the absence of overlapping we do not encounter particular difficulties, its presence complicates the measure of the contribution to the overall inequality related to the differences between the subgroups.

In the following we propose a joint analysis of Gini index decomposition and overlapping, by developing a comparison between two different approaches: from one side the methods based on the subgroup means, and from the other the evaluations performed on the basis of all the characteristics of the subgroup distributions. The combination of the two approaches allows to overcome the disadvantages of the single proposals, thus avoiding the risk of underestimating the real effect of the inequality factors under study.

The next Section outlines the issues related to overlapping and its measurement, while Section 3 presents the main aspects of the Gini index decomposition. Section 4 illustrates a development of overlapping and Gini index decomposition focused on gender gap analysis, with a case study related to the Italian personal income. From Section 2 to Section 4 an illustrative example is carried out, which works as a guideline for the reader, attempting to make clearer the indicators and the measures presented in the paper. Section 5 concludes.

2. OVERLAPPING

Overlapping represents a real challenge when measuring inequality, its presence increases the complexity and the difficulties of the measurement process and makes the interpretation of the final results more slippery. However, as frequently happens when challenges are involved, it also leads to innovative proposals and original solutions.

How to deal with the presence of overlapping is the main argument in the debate on the inequality decomposition, it represents one of the pillars of the classification into additively and non additively decomposable indices (Shorrocks (1980)), it is the motivation of the plurality of contributions on inequality decomposition.

Overlapping has a direct effect on the interpretation of the inequality structure and, consequently, on the inequality decomposition.

When we divide the total population in subgroups (for example by gender, educational level, etc.) the case of non overlapping subgroups allows an immediate evaluation of the contribution to overall inequality attributable to the factor used to partition the population into subgroups. The absence of overlapping indicates in itself a strong inequality factor. Depending on the size of the differences between the subgroups, we can rank the related factors and thus achieve the inequality structure. In this case overall inequality can be successfully decomposed in only two components, the inequality within and the inequality between subgroups, without any question on the additivity of the decomposition.

Factors able to originate non overlapping subgroups, or real data situations in which this is possible, are extremely rare, while a certain degree of overlapping represents the most usual and consolidated case. Increasing levels of overlapping indicate a weaker influence of the factor on inequality until, when the subgroups completely overlap, the influence of the factor on inequality reaches its minimum. In this case, the two previous components, the inequality within and the inequality between, are no longer sufficient to decompose overall inequality, and we need to introduce a third term, able to take into account the presence of overlapping units and their effect on inequality.

In the decompositions based on only two components, the third terms is obtained as a residual, while, in the decompositions developed on three components, the third term evolves accordingly to the expressions proposed by the different Authors.

Overlapping is also strictly connected to the effectiveness of the economic policy actions implemented to reduce inequality.

The absence of overlapping allows to operate directly on the poorest (or the richest) subgroup, within a simplified framework where the effects of any economic policy are quite easily detectable.

From the other side, in presence of overlapping, it becomes fundamental to disentangle the effects related to the overlapping units, which could strongly influence the final results of the policy action. The overlapping units of the richer subgroup would suffer a disadvantage compared to the overlapping units of the poorer subgroup, reducing the effectiveness of the intervention.

Being within the celebrations of the Gini's centenary, a last, but not least relevant feature of overlapping refers to the interest given to this topic by Gini. During his long polyhedral research activity, Gini dealt with numerous subjects, among which there is also the overlapping, which Gini called "transvariazione" in 1916 (Gini (1959)). An anthology of the Gini's works on overlapping was published in 1959 (Gini (1959)), almost unnoticed being written in Italian (Deutsch and Silber (1997)).

2.1. The Measurement of Overlapping

Given the importance of overlapping, it follows that its measurement plays a key role.

We need, first of all, to define the notion of overlapping. Given two subgroups j and h, with n_j and n_h respectively the sizes of the two subgroups, let λ_j and λ_h be two synthetic indicators (usually the mean or the median) of the two subgroups. Overlapping occurs when $\lambda_j \geq \lambda_h$ and, at least for one pair of units, we have $y_{ji} < y_{hr}$; following the Gini definition of transvariation, we have that the differences between the λ and the units have opposite sign:

$$(\lambda_j - \lambda_h) \geq 0$$

$$(y_{ji} - y_{hr}) < 0$$

Besides the concept of overlapping (or transvariation), Gini (1959) introduced also the notion of ipertransvariation ("ipertransvariazione" in Italian), which occurs when $\lambda_j \geq \lambda_h$ and two further conditions are verified: for at least one unit of the j-th subgroup, $y_{ji} < \lambda_h$, and, for at least one unit of the h-th subgroup, $y_{hr} > \lambda_j$:

$$(\lambda_j - \lambda_h) \geq 0$$

$$(y_{ji} - \lambda_h) < 0$$

$$(y_{hr} - \lambda_j) > 0.$$

Ipertransvariation represents a special case of overlapping, a situation which so far has not been given much importance, but which deserves more attention and could lead to interesting results (Deutsch and Silber (1997)).

In order to introduce a measure of overlapping, it is useful to start from the minimum and the maximum between which the measure ranges.

The minimum of any overlapping measure is naturally identified as zero and it corresponds to the case of absence of overlapping. The evaluation of the differences between two non overlapping subgroups is quite simple and the subgroups means are the first and most natural candidate. If the subgroup distributions are skewed and characterized by fat tails (as usually happens for income and wealth distributions), it can be necessary to appropriately weight the subgroups means. Overall inequality is then perfectly decomposable into two components, inequality within and inequality between, where the inequality between captures the contribution to overall inequality attributable to the differences between the subgroups, that is the relevance of the underlying inequality factor.

The maximum of overlapping is not so easily assessed. When two subgroups completely overlap, with $\lambda_j = \lambda_h$, we have two possible interpretations.

First, we can consider the two subgroups as perfectly equivalent, without any difference between them. Second, we can still calculate the differences $(y_{ji} - y_{hr})$ and attribute the positive differences to the inequality between and the negative ones to the overlapping.

The first solution is more suitable in an explorative framework, where we need to assess the existence of the two (or more) subgroups, that is, when we do not know a priori the inequality factors and we analyse the data looking for clusters or subgroups of specific interest. In this case, a complete overlapping implies that the underlying inequality factor no longer affects the inequality, which is completely explained by the inequality within and by the overlapping (o residual) component, while the inequality between is equal to zero.

If we already know the subgroups, that is when we have a priori knowledge of the underlying inequality factor, we work in a confirmative framework, where we want to assess the relevance and rank the inequality factors, and where the second approach can be more useful. In this case, the subgroups and the underlying inequality factor maintain a minimum effect on overall inequality also when overlapping reaches its maximum, and the inequality between still plays a role in the inequality decomposition.

It follows that, in order to evaluate the situation of complete overlapping, we need to deal also with the measurement of the inequality between the subgroups, deciding if maximum overlapping implies a zero or positive inequality between. We must be aware that this choice is not neutral, but has consequences on the evaluation and the ranking of the inequality factors, and it may be appropriate to evaluate inequality in both cases. The two possible interpretations of the case of maximum overlapping play an important role also in the Gini index decomposition and will represent a key issue in the next Sections.

Among his contributions to the study of overlapping, Gini introduced two indicators, the probability and the intensity of transvariation, which are briefly summarized in the following and which still now represent the basic indicators for overlapping measurement.

2.1.1. The Probability of Transvariation

The probability of transvariation refers to the case in which λ is the median, that is to the overlapping which occurs when

$$(\bar{y}_{mej} - \bar{y}_{meh}) \geq 0$$

$$(y_{ji} - y_{hr}) < 0$$

Given two subgroups j and h, of size respectively n_j and n_h, we have that the the number of possible differences $|y_{ji} - y_{hr}|$ is $n_j n_h$. Within the $n_j n_h$ possible differences $(y_{ji} - y_{hr})$, we can determine

${}_1n_{jh}$, the number of positive differences,

${}_2n_{jh}$, the number of null differences, and

${}_3n_{jh}$, the number of negative differences, such as

$${}_1n_{jh} + {}_2n_{jh} + {}_3n_{jh} = n_j n_h.$$

The evaluation of overlapping is naturally related to the number of negative differences ${}_3n_{jh}$, while the ${}_2n_{jh}$ null differences are assigned by Gini for one half to the negative differences and for one half to the positive differences.

In order to derive a normalized indicator, Gini proposed to obtain the probability of transvariation as the ratio

$$pt_{jh} = \frac{2 * ({}_2n_{jh}/2 + {}_3n_{jh})}{n_j n_h} = \frac{{}_2n_{jh} + 2 * {}_3n_{jh}}{n_j n_h}.$$

Probability of transvariation ranges between 0, when the subgroups do not overlap, and 1, when the overlapping is perfect.

Probability of transvariation reflects the frequentist view of probability, based on the ratio between favourable to possible cases, and it allows a first evaluation of the degree of the overlapping between the subgroups j and h.

2.1.2. The Intensity of Transvariation

The intensity of transvariation refers to the case in which λ is the arithmetic mean, that is to the overlapping which occurs when

$$(\bar{y}_j - \bar{y}_h) \geq 0$$

$$(y_{ji} - y_{hr}) < 0$$

With respect to the probability of transvariation, which takes into account the number of overlapping differences, the intensity of transvariation is based on their size. Gini proposed to calculate the sum of overlapping differences T_{jh}, with

$$T_{jh} = \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}| \qquad (y_{ji} - y_{hr}) < 0,$$

and to derive the intensity of transvariation as the ratio of T_{jh} to the sum of all differences $|y_{ji} - y_{hr}|$

$$it_{jh} = \frac{2 * T_{jh}}{\sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}|}$$

As the probability of transvariation, also the intensity of transvariation ranges between 0, when the subgroups do not overlap, and 1, when the overlapping is perfect.

Usually intensity of transvariation is lower than probability of transvariation, since overlapping mainly occurs for observations located in a narrow range. By increasing overlapping, the difference between the two indicator decreases, up to reach the same value, the unit, when overlapping is maximum.

The joint combination of probability and intensity of transvariation allows to evaluate both the frequency and the relevance of overlapping, thus providing a powerful insight on this aspect of inequality.

2.2. An Illustrative Example

In order to illustrate the different cases related to overlapping, we include an example where we observe a variable y which assumes values from 1 to 10. The overall $n = 90$ observations are divided in 2 subgroups, A and B, with $n_A = 60$ and $n_B = 30$.

Four different situations, representing the four main overlapping typologies, are reported. In the first case, the subgroups A and B are not overlapping, in the second there is a low overlapping between A and B, in the third we observe an high overlapping level, while, in the fourth case, the overlapping is complete.

The y values for the four cases are presented in Table 1 and in Figure 1.

Table 2 shows the mean, the median, the coefficient of variation $cv_j = \sigma_j / \bar{y}_j$, the probability and the intensity of transvariation for the four cases illustrated in Table 1 and in Figure 1.

Given $n_A n_B = 60 * 30 = 1800$ possible differences $(y_{Bi} - y_{Ar})$, in the first case, where $\bar{y}_{meB} = 7$, $\bar{y}_{meA} = 2$ and $(\bar{y}_{meB} - \bar{y}_{meA}) = (7 - 2) = 5$, all 1800 differences are positive and, since both the null and the negative differences are zero, ${}_2n_{BA} = 0$ and ${}_3n_{BA} = 0$, also the probability of transvariation is zero:

$$pt_{BA} = \frac{{}_2n_{BA} + 2 *_3 n_{BA}}{n_B n_A} = \frac{0 + 0 * 2}{60 * 30} = 0.00.$$

Figure 1. Example of overlapping: 90 observations are divided in two non overlapping subgroups (I), two weakly overlapping subgroups (II), two highly overlapping subgroups (III) and two perfectly overlapping subgroups (IV).

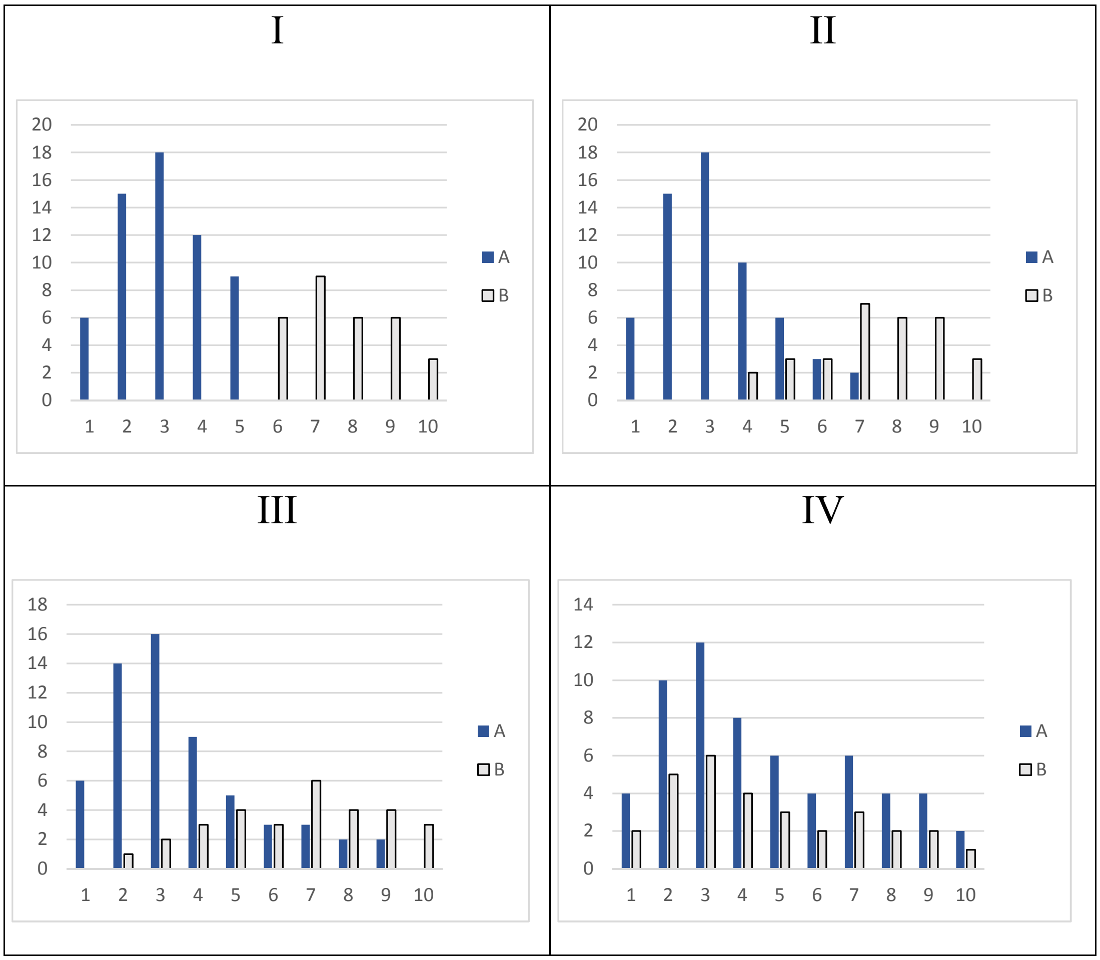

Table 1. Example of overlapping: 90 observations are divided in two non overlapping subgroups (I), two weakly overlapping subgroups (II), two highly overlapping subgroups (III) and two perfectly overlapping subgroups (IV).

tot		I				II				III				IV			
		A		B		A		B		A		B		A		B	
y_i	n_i	y_i	n_i	y_i	n_i	y_i	n_i	y_i	n_i	y_i	n_i	y_i	n_i	y_i	n_i	y_i	n_i
1	6	1	6	1	0	1	6	1	0	1	6	1	0	1	4	1	2
2	15	2	15	2	0	2	15	2	0	2	14	2	1	2	10	2	5
3	18	3	18	3	0	3	18	3	0	3	16	3	2	3	12	3	6
4	12	4	12	4	0	4	10	4	2	4	9	4	3	4	8	4	4
5	9	5	9	5	0	5	6	5	3	5	5	5	4	5	6	5	3
6	6	6	0	6	6	6	3	6	3	6	3	6	3	6	4	6	2
7	9	7	0	7	9	7	2	7	7	7	3	7	6	7	6	7	3
8	6	8	0	8	6	8	0	8	6	8	2	8	4	8	4	8	2
9	6	9	0	9	6	9	0	9	6	9	2	9	4	9	4	9	2
10	3	10	0	10	3	10	0	10	3	10	0	10	3	10	2	10	1
	90		60		30		60		30		60		30		60		30

The intensity of transvariation is based on the difference between the subgroup's means, that is $(\bar{y}_B - \bar{y}_A) = (7.70 - 3.05) = 4.65$, and on T_{BA}, that is the sum of the negative differences $(y_{Bi} - y_{Ar})$: without negative differences, since ${}_3n_{BA} = 0$, also $T_{BA} = 0$ and consequently the intensity of transvariation is zero

$$it_{BA} = \frac{2 * T_{BA}}{\sum_{i=1}^{n_B} \sum_{r=1}^{n_A} |y_{Bi} - y_{Ar}|} = \frac{2 * 0}{2587.50} = 0.00$$

In the second case, still with $\bar{y}_{meB} = 7$, $\bar{y}_{meA} = 2$ and $(\bar{y}_{meB} - \bar{y}_{meA}) = 5$, the presence of some overlapping units leads to ${}_2n_{BA} = 61$ null and to ${}_3n_{BA} = 43$ negative differences $(y_{Bi} - y_{Ar})$; the corresponding probability of transvariation results

$$pt_{BA} = \frac{{}_2n_{BA} + 2 *_3 n_{BA}}{n_B n_A} = \frac{61 + 2 * 43}{60 * 30} = 0.08.$$

The difference between the subgroup's means is $(\bar{y}_B - \bar{y}_A) = (7.40 - 3.20) = 4.20$ and T_{BA}, that is the sum of the ${}_3n_{BA} = 43$ negative differences is equal to 63; the related intensity of transvariation is calculated as

$$it_{BA} = \frac{2 * T_{BA}}{\sum_{i=1}^{n_B} \sum_{r=1}^{n_A} |y_{Bi} - y_{Ar}|} = \frac{2 * 63}{3262.80} = 0.04.$$

In the third case, an higher level of overlapping between the subgroups A and B leads to ${}_2n_{BA} = 136$ null and to ${}_3n_{BA} = 226$ negative differences $(y_{Bi} - y_{Ar})$, with a probability of transvariation given by

$$pt_{BA} = \frac{{}_2n_{BA} + 2 *_3 n_{BA}}{n_B n_A} = \frac{136 + 2 * 226}{60 * 30} = 0.33.$$

The difference between the subgroup's means is $(\bar{y}_B - \bar{y}_A) = (6.60 - 3.60) = 3$ and the sum of the ${}_3n_{BA} = 226$ negative differences is $T_{BA} = 515$; the intensity of transvariation results

$$it_{BA} = \frac{2 * T_{BA}}{\sum_{i=1}^{n_B} \sum_{r=1}^{n_A} |y_{Bi} - y_{Ar}|} = \frac{2 * 515}{4320} = 0.24.$$

The fourth case refers to the situation of maximum overlapping, where the subgroup's medians have the same value $\bar{y}_{meB} = \bar{y}_{meA} = 3$ as holds for the subgroup's means $\bar{y}_B = \bar{y}_A = 4.60$.

Table 2. Sample size, mean, median, coefficient of variation, probability of transvariation and intensity of transvariation for the subgroups A and B in the four cases (zero, low, high, maximum overlapping) reported in Table 1.

	tot	first		second		third		fourth	
		A	B	A	B	A	B	A	B
n_j sample size	90	60	30	60	30	60	30	60	30
$\bar{y}_j$ mean	4.60	3.05	7.70	3.20	7.40	3.60	6.60	4.60	4.60
$\bar{y}_{mej}$ median	4.00	2.00	7.00	2.00	7.00	2.00	6.00	3.00	3.00
cv_j	0.55	0.42	0.16	0.53	0.20	0.61	0.31	0.55	0.55
${}_1n_{BA}$ n. positive differences		1800		1733		1212		788	
${}_2n_{BA}$ n. null differences		0		61		136		224	
${}_3n_{BA}$ n. negative differences		0		43		226		788	
pt_{BA} prob. transvariation		0		0.08		0.33		1.00	
T_{BA}		0		63		515		2540	
it_{BA} int. transvariation		0		0.04		0.24		1.00	

In this extreme case, the number of positive differences $(y_{Bi} - y_{Ar})$, ${}_1n_{BA}$, is 788, while the null differences are ${}_2n_{BA} = 224$ and the negative differences ${}_3n_{BA} = 788$: when the overlapping is maximum, with $\bar{y}_{meB} = \bar{y}_{meA}$, it follows that the number of positive differences is equal to the number of negative differences: ${}_1n_{BA} ={}_3 n_{BA}$. The property ${}_1n_{BA} ={}_3 n_{BA}$ represents the basis for assigning, in the case of maximum overlapping, the difference between the subgroups in equal parts to the inequality between and to the overlapping component.

Starting from ${}_2n_{BA} = 224$, ${}_3n_{BA} = 788$, $n_A = 60$ and $n_B = 30$, the probability of transvariation results

$$pt_{BA} = \frac{{}_2n_{BA} + 2 *_3 n_{BA}}{n_B n_A} = \frac{224 + 2 * 788}{60 * 30} = 1.00.$$

The sum of the ${}_3n_{BA} = 788$ negative differences is $T_{BA} = 2540$, which is the same value as the sum of the ${}_1n_{BA} = 788$ positive differences, thus leading to the maximum of the intensity of transvariation:

$$it_{BA} = \frac{2 * T_{BA}}{\sum_{i=1}^{n_B} \sum_{r=1}^{n_A} |y_{Bi} - y_{Ar}|} = \frac{2 * 2540}{5080} = 1.00.$$

As expected, probability of transvariation pt_{BA} is always lower than the intensity of transvariation it_{BA}, except in case IV, when the two indicators are equal to one.

3. THE GINI INDEX DECOMPOSITION

The Gini index (Gini (1914)) during its over 100 years of life has known many different expressions and applications.

For the case of a population disaggregated into k subgroups of size n_j, with $\sum_{j=1}^{k} n_j = n$, the Gini index G can be expressed as follows

$$G = \frac{1}{2n^2\bar{y}} \sum_{j=1}^{k} \sum_{h=1}^{k} \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}| \tag{3.1}$$

where $\bar{y}$ is the arithmetic mean of y in the overall population, y_{ji} is the value of y in the i-th unit of the j-th subgroup and, accordingly, y_{hr} is the value of y in the r-th unit of the h-th subgroup.

A further expression of the Gini index which is extremely useful for the index decomposition is

$$G = \sum_{j=1}^{k} \sum_{h=1}^{k} G_{jh} p_j s_h \tag{3.2}$$

where

G_{jj} is the Gini index of the j-th subgroup,

G_{jh} is the Gini index between subgroup j and subgroup h, with

$$G_{jh} = \frac{1}{n_j n_h (\bar{y}_j + \bar{y}_h)} \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}|, \tag{3.3}$$

$p_j = n_j/n$ is the population share of the j-th subgroup, and

$s_j = (n_j \bar{y}_j)/(n\bar{y})$ is the income share of the j-th subgroup.

The literature on the Gini index (Dagum (1987), Dagum and Zenga (1990)) and on its decomposition is quite extensive and cannot be easily summarized, however for a detailed discussion of the Gini index we can look at the studies by Giorgi (see, e.g., Giorgi (2005), Giorgi (2011b), Giorgi and Gubbiotti (2017)), a researcher living in the same city of Gini, working at the same university of Gini and who has devoted a great portion of his studies to the dissemination, and in some cases to the rediscovery, of Gini's contributions.

Also the decomposition of the Gini index is the object of uncountable studies and researches, which belong to the more general discussion about the decomposition of inequality measures. Among the different proposals, Podder (1993) analyses the contribution of each subgroup to total inequality, avoiding the traditional distinction between inequality within and inequality between, but focusing on the Gini index elasticities with respect to the subgroups distributions.

In the following we do not propose an exhaustive review of the methods of decomposing the Gini index, but we suggest to distinguish the various proposals in two classes, one based on subgroup means, and which does not explicitly consider the overlap, and one based on all the characteristics of the subgroup distributions, including the overlap. For each class a proposal is identified and a comparison between the two approaches is provided.

3.1. Inequality Within

The measurement of the inequality within, although not being free of criticism (see, e.g. Frosini (2012)), is the object of a certain consensus among the different contributions to the Gini index decomposition.

Our focus is on the inequality between and the overlapping component, while for the inequality within we directly use the measure usually adopted in the literature, where G_w is obtained as a weighted sum of the Gini index of each subgroup and its expression is

$$G_w = \sum_{j=1}^{k} G_j p_j s_j \tag{3.4}$$

The component of inequality within G_w allows to evaluate the relevance of the variability of the subgroups on the overall inequality: more variable, that is unequal, subgroups lead to an higher G_w, while less variable, that is more equidistributed, subgroups indicate lower values of the inequality within.

3.2. Inequality Between and Overlapping Component

Most of the debate about inequality decomposition, and specifically about the Gini index decomposition, revolves around the measurement of the effects on overall inequality related to the differences between subgroups. In particular, the key point refers to the evaluation of the effects related to the overlapping between subgroups and most of the differences among the many contributions to the Gini index decomposition can be traced to this issue.

3.2.1. Mean-Based Evaluations

A milestone in the literature on the Gini index decomposition is represented by the pioneering contribution by Bhatthacharya and Mahalanobis (1967) who proposed to evaluate the inequality between on the basis of the subgroup means as

$$G_b = \sum_{j=1}^{k} \sum_{h=1}^{k} p_j p_h |\bar{y}_j - \bar{y}_h| / 2\bar{y} \qquad j \neq h \tag{3.5}$$

Overlapping is not explicitly evaluated by Bhattacharya and Mahalanobis, their decomposition is exact when the subgroups do not overlap, while the presence of overlapping requires to add to the decomposition a residual third term, so that

$$G - G_w = G_b + R.$$

When the subgroups completely overlap, with $\bar{y}_j - \bar{y}_h, \quad i, j = 1, ..., k$, the inequality between G_b is equal to zero and $G = G_w + R$. Maximum overlapping is therefore addressed in the framework of the first of the two interpretations outlined in Section 2.1.

The proposal by Bhattacharya and Mahalobis has had a great impact on the discussion about the Gini index decomposition. Many empirical researches and theoretical works (see e.g. the famous contribution by Rao (1969) and the interesting decomposition proposed by Pyatt (1976), where the Gini index is treated as a statistical game) refer to Bhattacharya and Mahalanobis decomposition, which can be assumed as a benchmark for the decompositions where the measurement of the inequality between is based on the subgroup means.

3.2.2. Distribution-Based Evaluations

A significant step forward in the Gini index decomposition is made in 1991 thanks to Yitzhaky and Lerman (Yitzhaki and Lerman (1991), Yitzhaki (1994)), who introduced a decomposition of the Gini index where overlapping is explicitly taken into account. Moreover, their contribution is not constrained to use only the subgroup means, but it moves towards the use of the whole distribution. Within their proposal, the inequality between subgroups is evaluated as

$$G_b = 2cov(\bar{y}_j, \bar{F}_j)/\bar{y} \tag{3.6}$$

where F indicates the cumulative distribution, estimated in the sample by the rank of the observation; therefore $\bar{F}_j$ is calculated as $\bar{F}_j = \bar{R}_j/n$ and $\bar{R}_j$ is the average rank of the j-th subgroup in the overall population.

Yitzhaky and Lerman also introduce an overlapping component, which is measured on the basis of

$$O_{jh} = cov_h(y, F_j(h))/cov_h(y, F_h(y)) \tag{3.7}$$

where the numerator is the covariance over the h-th subgroup between y and the rank of a unit of subgroup h in the subgroup j, while the denominator is the covariance over the h-th subgroup between y and the ranking in the h-th subgroup.

The overlapping component is then included in the decomposition as

$$G = \sum_{h=1}^{k} s_h G_h \sum_{j=1}^{k} p_j O_{jh} + G_b \tag{3.8}$$

When the subgroups completely overlap, with $\bar{y}_j = \bar{y}_h, \quad j,h = 1,...,k$, the inequality between G_b is equal to zero. As in the decomposition by Bhattacharya and Mahalanobis, maximum overlapping is addressed following the first of the two interpretations of Section 2.1.

To Yitzhaki and Lerman, and to their school (see, e.g., Frick et al. (2006)), we owe many papers on the topic of the Gini index decomposition, which allow a strong progress of the inequality studies and start a line of research followed by several other authors.

Still in the framework of the three-terms decompositions of the Gini index, and of the distribution-based evaluation of the inequality between, in 1997 Dagum proposed a further contribution (Dagum (1997)), building on a previous work by Mehran (1975). The Dagum's proposal is characterized by a specific attention given to the overlapping between subgroups; the attention paid to overlapping in the framework of the Dagum's decomposition is certainly not an accident, but it represents the result of a precise choice and derives from the studies made by Dagum on this topic during the fifties (Dagum (1959)), when he worked in Rome with Gini.

Dagum suggests to evaluate the contribution to overall inequality related to the differences between the subgroups as

$$G - G_w = G_b + G_t = \sum_{j=1}^{k}\sum_{h=1}^{k} G_{jh} p_j s_h \qquad j \neq h$$

The original proposal for the expressions of G_b and G_t requires an intensive computational effort, but it is possible (Costa (2008), Costa (2009), Costa (2016)) to derive simplified formulas for both the components. More specifically, we can express G_b and G_t as

$$G_b = G_b^* + 0.5(G - G_w - G_b^*) \tag{3.9}$$

and

$$G_t = 0.5(G - G_w - G_b^*) \tag{3.10}$$

where

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

$$p_{hj}^* = p_h/(p_h + p_j), \qquad s_{hj}^* = s_h/(s_h + s_j).$$

In this way we are able to obtain G_b and G_t on the basis only of the G_j, p_j and s_j, overcoming most of the computational efforts related to the 1997 version.

The Dagum's contribution is able to be both simple and intuitive at the same time, since, by referring to the basic expression (3.1) of the Gini index, it essentially suggests to divide the sum of the differences $|y_{ji} - y_{hr}|$ into three parts:

first, the sum of the differences in which $j = h$,

second, the sum of the differences in which $j \neq h$, $\bar{y}_j > \bar{y}_h$, $y_{ji} > y_{hr}$,

third, the sum of the differences in which $j \neq h$, $\bar{y}_j > \bar{y}_h$, $y_{ji} < y_{hr}$.

The first part is interpreted as the inequality within G_w, the second as the inequality between G_b, and the third as the overlapping component G_t.

The Dagum's decomposition differs from many of the other contributions as regards the situation of maximum overlap. In this case, in fact, the inequality between is not equal to zero, as usually happens, but has the same value of the overlapping component, and it results $G = Gw + G_b + G_t$, with $G_b = G_t$. Maximum overlapping is therefore addressed in the framework of the second of the two interpretations presented in Section 2.1.

The work by Dagum inspired many other Authors, in particular Mussard (Mussard (2004), Mussard et al. (2006)) extensively studied the decomposition, linked (Mussard and Richard (2012)) the Dagum's and Yitzhaki's decompositions, with Mornet et al. (2013) proposed a multi-level generalization, and also stressed the role of overlapping (Mussard and Savard (2012)). Building on the contributions by Mussard, Mussini (2013) developed a matrix approach which allows to investigate the role of interaction between subgroups. A further contribution is given by Giudici and Raffinetti (2011), who propose a Gini index decomposition in terms of concordance and discordance shares. Still in the framework of distribution-based evaluation, and building on the Dagum's contribution, Ogwang (2014), Ogwang (2016) decomposes the Gini index by using an interesting regression approach.

Overall, the contributions developed within the framework of distribution-based evaluations are strictly linked to the overlapping measurement and stress this point as fundamental in inequality analysis. In the following, the Dagum's decomposition is used as the reference for the decompositions developed in a distribution-based framework.

3.3. The Comparison of Decompositions

The analysis of the links between different decomposition methods represents the topic of interest for several Authors (see, e.g., Mussard and Richard (2012) and Radaelli (2010)) who investigate the common aspects and the differences which characterize the many decompositions proposed over more than 50 years.

Mean-based decompositions, such as the Bhattacharya and Mahalanobis's proposal, are naturally equivalent to distribution-based decompositions, such as the Dagum's method, in absence of overlapping.

Starting from the simplest case of two non overlapping subgroups, with $\bar{y}_1 < \bar{y}_2$, we have for the decomposition by Bhattacharya and Mahalanobis

$$G_{bBM} = 2p_1p_2(\bar{y}_2 - \bar{y}_1)/2\bar{y} = p_2p_1\bar{y}_2/\bar{y} - p_1p_2\bar{y}_1/\bar{y} = p_1s_2 - p_2s_1 =$$

$$= p_1(1 - s_1) - (1 - p_1)s_1 = p_1 - p_1s_1 - s_1 + p_1s_1 = p_1 - s_1$$

and for the decomposition by Dagum

$$G_{bD} = G_{12}p_1s_2 + G_{21}p_2s_1 = G_{12}(p_1s_2 + p_2s_1) = \frac{p_1s_2 + p_2s_1}{n_1n_2(\bar{y}_1 + \bar{y}_2)} \sum_{j=1}^{n_1} \sum_{h=1}^{n_2} (y_{2j} - y_{1h}) =$$

$$= \frac{p_1s_2 + p_2s_1}{n_1n_2(\bar{y}_1 + \bar{y}_2)} (\bar{y}_2 - \bar{y}_1) = \frac{p_1p_2(\bar{y}_2 - \bar{y}_1)}{\bar{y}} = p_1p_2\bar{y}_2/\bar{y} - p_1p_2\bar{y}_1/\bar{y} =$$

$$= p_1(1 - s_1) - (1 - p_1)s_1 = p_1 - s_1.$$

In absence of overlapping, as intuitive, the inequality between has the same value in both decompositions, and it results $G_b = (p_1 - s_1)$.

The extension to the case of k non overlapping subgroups, with $\bar{y}_1 < \bar{y}_2 < ... < \bar{y}_k$, is available by using to the quantities

$$p^*_{hj} = p_h/(p_h + p_j), \qquad s^*_{hj} = s_h/(s_h + s_j),$$

which allow to normalize the population share and the income share with respect to the sizes of the subgroups h and j.

By means of p^*_{hj} and s^*_{hj} we have that, for both the decompositions, the inequality between G_b can be obtained from the differences $(p^*_{hj} - s^*_{hj})$ as

$$G_b = \sum_{j=1}^{k} \sum_{h=1}^{k} p_jp_h|\bar{y}_j - \bar{y}_h|/2\bar{y} = \sum_{j=1}^{k-1} \sum_{h=j+1}^{k} \frac{p^*_{hj} - s^*_{hj}}{p^*_{hj}s^*_{jh} + p^*_{jh}s^*_{hj}} (p_js_h + p_hs_j) \tag{3.11}$$

When the subgroups overlap, to the mean-based decompositions, such as the one by Bhattacharya and Mahalanobis, is added a third residual term, while the measure of G_b is not modified.

From the other side, the distribution-based decompositions, such as the proposal by Dagum, have a different development, the overlapping component assumes a positive value, which increases by increasing levels of overlapping, and also the measurement of G_b can be influenced.

Starting again from the case of $k = 2$ subgroups, for Bhattacharya and Mahalanobis we have

$$G - G_w = G_{bBM} + R = (p_1 - s_1) + R$$

while in the Dagum's decomposition it results

$$G - G_w = G_{bD} + G_t = (p_1 - s_1 + G_t) + G_t.$$

The third residual term in the decomposition by Bhattacharya and Mahalanobis is then equal to twice the overlapping component in the Dagum's decomposition, with the relevant difference that in the Dagum decomposition also the inequality between is influenced by the presence of overlapping.

The extension to the case of k overlapping subgroups leads, for Bhattacharya and Mahalanobis, to

$$G - G_w = G_{bBM} + R = (\sum_{j=1}^{k} \sum_{h=1}^{k} p_j p_h |\bar{y}_j - \bar{y}_h| / 2\bar{y}) + R$$

while, for the Dagum's decomposition, we have

$$G - G_w = G_{bD} + G_t = (\sum_{j=1}^{k-1} \sum_{h=j+1}^{k} \frac{p^*_{hj} - s^*_{hj}}{p^*_{hj} s^*_{jh} + p^*_{jh} s^*_{hj}} (p_j s_h + p_h s_j) + G_t) + G_t$$

where G_t is obtained as in (3.10).

The comparison between the two approaches is still based on the relation (3.11) and the result

$$R = 2 * G_t$$

holds also for the case of k overlapping subgroups, thus allowing to easily compare the two decompositions of the Gini index.

The relation between R and G_t can also be exploited in order to derive a relation between the expressions of the inequality between subgroups in the two decompositions. Since

$$G - G_w = G_{bBM} + R = G_{bD} + G_t$$

by using $R = 2G_t$, we obtain

$$G_{bBM} + 2G_t = G_{bD} + G_t$$

$$G_{bBM} = G_{bD} + G_t - 2G_t = G_{bD} - G_t$$

which provides a further element of comparison both between the two decompositions and between the two approaches.

In absence of overlapping, the subgroups means, appropriately weighted, are fully informative, while the presence of overlapping could reduce the relevance of the subgroups means as a reference for the inequality between subgroups, depending on how we consider and treat the overlapping units.

Table 3. Population share, income share, Gini index for the subgroups A and B in the four cases (zero, low, high, maximum overlapping) reported in Table 1.

	tot	first		second		third		fourth	
		A	B	A	B	A	B	A	B
p_j populations share	1.00	0.67	0.33	0.67	0.33	0.67	0.33	0.67	0.33
s_j character share	1.00	0.44	0.56	0.46	0.54	0.52	0.48	0.67	0.33
G_j Gini index	0.31	0.22	0.09	0.25	0.13	0.30	0.19	0.31	0.31

3.4. An Illustrative Example

In this Section we follow the example previously introduced, where four different overlapping typologies are analysed and compared. Our purpose here is to study the Gini index decomposition with respect to different degrees of overlapping.

Table 3 reports the population share p_j, tha character share s_j and the Gini index for the subgroups A and B in the four cases analysed. The p_j values are constant, as n_A and n_B are always the same (60 and 30 respectively) in all 4 cases. The s_j, on the contrary, show the usual dynamic which characterizes the overlapping: initially, when overlapping is absent or low, the s_j are quite different from the p_j, while increasing levels of overlapping lead the s_j to converge to the p_j, until, when the overlapping is complete, the character share s_j equal the population share p_j. A similar behaviour interests the Gini indexes, which are different in absence or low overlapping, while are coincident in the fourth case.

The Gini index decomposition for our four cases is illustrated in Table 4, for the mean-based evaluations we adopt the Bhattacaria and Mahalanobis's proposal, while for the distribution-based evaluations we resort to the Dagum's contribution.

Without overlapping, in case I, the Gini index is perfectly decomposable as the sum of G_w and G_b, with $R = G_t = 0$, and the two decompositions lead to the same results.

The presence of overlapping introduces some differences between the two decompositions, and the results shown in Table 4 allow to highlight three effects related to increasing levels of overlapping between the subgroups.

The first obvious effect refers to R and G_t, which, reflecting the importance on overall inequality related to overlapping units, steadily increase from case I to case IV. Furthermore, by comparing R with G_t, we can observe the relation $R = 2 * G_t$ illustrated in the Section 3.3.

The second usual effect related to an increasing overlapping is a growing inequality within, as it is clearly shown in Table 4: in the first case, G_w is the 26.8% of total inequality ($G_w/G = 0.0821/0.3067 = 0.2677$), value which increases to 32.8% in the second case, to 43.8% in the third, up to the 55.6% in the fourth last case.

The third effect, consequence of a greater inequality within and an higher overlapping, is a reduced importance of inequality between on overall inequality. The decrease is not the same for the two decompositions, as it is fully detectable by comparing the third and the sixth column of Table 4. From Table 4 we can also observe the relation $G_{bBM} = G_{bD} - G_t$, shown in Section 3.3, and therefore obtain the third column as the difference between the sixth and the seventh.

The case IV, that is the situation of maximum overlapping, allows to highlight one of the main differences between the two decompositions and the two possible ways to address perfect overlapping highlighted in Section 2.1.

Following Bhattacaria and Mahalanobis, in the fourth case the inequality between reaches its minimum, which is equal to zero, and all differences between the two subgroups are attributed to the overlapping component, in their case the residual term R.

In the framework of Dagum's decomposition, the maximum overlapping is evaluated by giving the same

Table 4. Gini index decomposition by Bhattacaria-Mahalanobis and by Dagum for the subgroups A and B in the four cases (zero, low, high, maximum overlapping) reported in Table 1.

	Bhattacaria Mahalanobis			Dagum		
	Gw	Gb	R	Gw	Gb	Gt
I	0.0821	0.2246	0	0.0821	0.2246	0
II	0.1005	0.2029	0.0034	0.1005	0.2046	0.0017
III	0.1342	0.1449	0.0276	0.1342	0.1587	0.0138
IV	0.1704	0	0.1363	0.1704	0.0682	0.0682

importance to the negative (overlapping) and positive (non overlapping) differences $(y_{Bi} - y_{Ar})$, which have the same number ${}_1n_{BA} ={}_3 n_{BA} = 788$, and also the same intensity, $T_{BA} = 2540$, as shown in the last column of Table 2. It follows that the differences between the two subgroups are equally assigned to the inequality between and to the overlapping component, $G_b = G_t = 0.0682$, and the minimum of G_b is not zero but G_t.

4. INEQUALITY DECOMPOSITION, OVERLAPPING AND POLITICAL ECONOMY: THE ANALYSIS OF GENDER GAP

The joint combination of inequality decomposition and overlapping measurement allows many interesting developments. Among the many possibilities, in the following it is illustrated how both the Gini index decomposition and the overlapping evaluation can improve gender gap analysis.

Gender gaps are the focus of many political and social actions and represent worldwide one of the main concern of national and supranational agencies. Interest for gender gaps motivated many researches, which in recent years have grown at an ever faster rate, producing an impressive quantity of studies and reports (see, e.g., Gender Equality Index Report (2013) and references therein).

Issues related to gender inequalities are particularly suitable to illustrate the necessity to move from mean-based measures to distribution-based evaluations.

Traditionally gender inequalities are firstly evaluated on the basis of the difference between the mean of the female subgroup $\bar{y}_f$ and the mean of the male subgroup $\bar{y}_m$, and the size of the gender gap is evaluated by referring to this difference, or absolute

$$gap_1 = (\bar{y}_m - \bar{y}_f),$$

or in some relative form, such as

$$gap_2 = (\bar{y}_m - \bar{y}_f)\bar{y}_f$$

or

$$gap_3 = (\bar{y}_m - \bar{y}_f)/\bar{y}.$$

Besides the difference $(\bar{y}_m - \bar{y}_f)$, we can resort to the difference between the population share and the income share of the female subgroup

$$gap_4 = (p_f - s_f),$$

which refers to the equidistribution case $(p_f = s_f)$ and still evaluates gender gap in a mean-based framework where overlapping is not taken into account.

The informative content of the subgroup means becomes here a key point, and the obvious risk is to underestimate the real size of gender gap if the subgroup means are not fully informative.

At this regard it is crucial to assess the degree of overlapping, since the informative content of the subgroup means is inversely related to the degree of overlapping: increasing levels of overlapping reduce the information provided by the subgroup means, thus leading to a possible underestimation of the gender gap.

By resorting to the Gini index decomposition, and by dividing the total population in two subgroups by gender, we are able to evaluate the relevance of gender on overall inequality, thus providing a further measurement of gender gap.

With respect to the general case, here we have $k = 2$ subgroups, constituted by the female (f) and male (m) units respectively. By using formula (2) we can express the Gini index as

$$G = G_f p_f s_f + G_m p_m s_m + G_{fm} p_f s_m + G_{mf} p_m s_f$$

where the first two terms refer to the component of inequality within

$$G_w = G_f p_f s_f + G_m p_m s_m.$$

By referring to the proposal by Bhattacharya and Mahalanobis, the component of inequality between (see Section 3.3) is calculated as

$$G_b = 2p_f p_m(\bar{y}_m - \bar{y}_f)/2\bar{y} = p_f - s_f$$

and the related decomposition of the Gini index results

$$G = G_w + G_b + R = (G_f p_f s_f + G_m p_m s_m) + (p_f - s_f) + R.$$

In the framework of the Dagum's decomposition, the component of inequality between can be expressed as

$$G_b = p_f - s_f + G_t$$

and we can obtain the overlapping component by using the simplified version of the Dagum's decomposition applied to the case of k=2 subgroups

$$G = G_w + G_b + G_t = (G_f p_f s_f + G_m p_m s_m) + (p_f - s_f + G_t) + G_t$$

where G_t is calculated on the basis of G_f, G_m, p_f, p_m, s_f and s_m as

$$G_t = (G - G_w - p_f - s_f)/2.$$

Within the Gini index decomposition we can evaluate the gender gap on the basis of the ratio

$$gap_5 = G_b/G$$

which measures the contribution of gender as inequality factor to overall inequality.

The approach developed on the basis of the inequality decomposition also allows to detect the contribution of overlapping to the evaluation of gender gap by comparing $(p_f - s_f + G_t)/G$ to $(p_f - s_f)/G$, that is the Dagum's proposal to the contribution by Bhattacharya and Mahalanobis.

In the evaluation of gender gap by means of inequality decomposition, the ratio of the inequality between to overall inequality, that is gap_5, calculated with respect to different measures of G_b, allows to achieve a variety of further indicators about the gender gap.

From gap_1 to gap_3 we have indicators exclusively based on the subgroup means, which are simple to calculate, easy to understand and immediate to communicate. With gap_4 we enter the analysis of inequality, without considering the presence of overlapping, and with gap_5 we add the information related to the decomposition of the Gini index, where the overlapping component can have different roles depending on the type of decomposition.

Table 5. Gender gap evaluation for the four cases illustrated in Table 1.

	gap_1 $(\bar{y}_B - \bar{y}_A)$	gap_2 $(\bar{y}_B - \bar{y}_A)/\bar{y}_A$	gap_3 $(\bar{y}_B - \bar{y}_A)/\bar{y}$	gap_4 $(p_A - s_A)$	gap_5 G_{bBM}/G	gap_5 G_{bD}/G
I	4.65	1.53	1.01	0.23	0.73	0.73
II	4.20	1.31	0.91	0.20	0.66	0.67
III	3.00	0.83	0.65	0.15	0.47	0.52
IV	0.00	0.00	0.00	0.00	0.00	0.22

Overall we obtain a series of indicators, all useful: in the case that they provide different indications, it is opportune to consider them all, so as to avoid to leave out important information on the structure of the inequality.

The literature on gender gap, although still largely focused on average values, starts to take an interest in distribution-based evaluations (see, e.g., Bonnet, Meurs and Rapport (2018), Larraz (2015), Goraus et al. (2017), Selezneva and Van Kerm (2016) and references therein), pointing to this direction as a promising line of future development.

4.1. An Illustrative Example

In the following we apply the different evaluations of gender gap introduced in Section 4 to the four overlapping typologies illustrated in Table 1 and in Figure 1. Subgroup B, the most affluent, acts as the male subgroup, while the subgroup A acts as the female subgroup; that is, gender gap evaluation starts from the difference $(\bar{y}_B - \bar{y}_A)$.

In Table 5 are reported the values of the different gap_i calculated for the four cases analysed in our example. In the first case, all methods indicate the highest gap: the values on the first row are always the largest. Furthermore, it is interesting to observe how, in absence of overlapping, mean-based and distribution-based indicators lead to the same results: $gap_5 = G_{bBM}/G = G_{bD}/G = 0.73$.

Increasing levels of overlapping lead to smaller gaps, up to reach the minimum in the fourth case. For all methods the minimum is equal to zero, except for gap_5 in the Dagum version, which, consistently with the evaluation of G_{bD} shown in Table 4, is not zero even in case IV. Maximum overlapping is addressed, from gap_1 to gap_5 in the Bhattacaria and Mahalanobis version, in the framework of the first interpretation proposed in Section 2.1: the two subgroups are considered as perfectly equivalent and the gap is 0. On the contrary, gap_5 in the Dagum version refers to the second interpretation of maximum overlapping, where the inequality factor still plays a role and the gap reflects the importance of the inequality factor.

The efficient comparison of the different measures can be achieved by referring to the dynamic observable with respect to increasing levels of overlapping.

Gap_1 shows, from case I to case II, a decrease of 9.7%, and, from II to III, a reduction of 28.6%. The strongest decreases (-14.4% and -36.6%, respectively) are indicated by gap_2, while gap_3 and gap_4 provide intermediate evaluations between gap_1 and gap_2.

Gap_5, in the Bhattacaria and Mahalanobis's version, shows a dynamic almost equal to gap_1 and gap_3, perfectly aligned with mean-based evaluations. The lesser decreases are suggested by gap_5 calculated in the Dagum's version, thus indicating that, by using a distribution-based method, the decrease of the gap is not so strong as predicted by mean-based methods.

4.2. A Case Study: the Italian Personal Income by Gender

As a case study to illustrate the evaluation of gender gap, we analyse the Italian personal income and compare the approach based on all the characteristics of the subgroup distributions to the approach related to the

Table 6. Population share, income share, mean, median, coefficient of variation and Gini index for the Italian personal income by gender, 2006 and 2016.

	2006			2016		
	female	male	tot	female	male	tot
n_j sample size	6234	7194	13428	5826	6064	11890
p_j population share	0.462	0.538	1.000	0.483	0.517	1.000
s_j income share	0.345	0.655	1.000	0.399	0.601	1.000
$\bar{y}_j$	13887	22701	18628	15742	22140	19052
$\bar{y}_{mej}$	12200	18341	15333	14560	19002	16900
cv_j	0.739	1.097	1.076	0.699	0.820	0.812
G_j	0.351	0.352	0.372	0.345	0.350	0.358

Table 7. Probability and intensity of transvariation for the Italian personal income by gender, 2006 and 2016.

	2006	2016
pt_{mf}	0.592	0.725
it_{mf}	0.422	0.564

subgroup means. The data are provided by the Bank of Italy by means of a biannual multidimensional survey on the income and wealth of the Italian households. In the following are analysed the waves related to 2006 and 2016.

The basic descriptive statistics for Italian personal income by gender are reported in Table 6. The population share of the female subpopulation moves from 46.2% to 48.7%, with an increase of 2.1 points, while the related income share changes from 34.5% to 39.9%, with an increase of 5.4 points, which indicates a stronger similarity to the case of equidistribution, where $p_f = s_f$.

As typical for income distribution, the mean and the median are quite different, thus suggesting that the income distribution is characterized by high levels of skewness.

Besides the skewness, a further element of interest refers to the coefficient of variation $cv_j = \sigma_j/\bar{y}_j$, which indicates that male and female subgroups have a different variability.

Overall, high skewness and different variability suggest that mean-based indicators should not be a preferred tool for the analysis of our data, and point to the need to combine mean-based indicators with other methods, such as the Gini index, robust with respect to skewness and different levels of variability.

Given the strategic role played by overlapping, it becomes relevant to include in our analysis also some measures of overlapping, namely the probability and the intensity of transvariation, whose results for Italian personal income divided by men and women are reported in Table 7.

Both indicators show, from 2006 to 2016, an increase of around 10 points and both indicate, during the period analysed, a stronger overlap between female and male subgroups. As expected, the probability of transvariation is higher than the intensity of transvariation, thus suggesting that overlapping mainly occurs in areas of the income subgroups distributions, where the differences between incomes are contained.

The analysis of the data shown in Tables 6 and 7 confirms how the income distributions of women and men differ in many aspects, but also share a large area of overlap. It follows that in our analysis it could be useful, if not necessary, to explicitly take into account also the presence of overlapping units.

The results of the Gini index decomposition related to the partition in 2 subgroups by gender are illustrated in Tables 8 and 9. As outlined in Section 3, we use the proposal by Bhattacharya and Mahalanobis for the mean-based methods, and the Dagum's decomposition for the distribution-based evaluations. Table 8 reports the basic ingredients for deriving the components of the two Gini index decompositions, which

Table 8. Information set for the Gini index decomposition by gender, Italian personal income 2006 and 2016.

	2006			2016		
	f	m	tot	f	m	tot
$p_j s_j G_j$	$p_f s_f G_f = 0.056$	$p_m s_m G_m = 0.124$	$G_w = 0.180$	$p_f s_f G_f = 0.066$	$p_m s_m G_m = 0.109$	$G_w = 0.175$
$\frac{\lvert\bar{y}_j - \bar{y}_h\rvert}{2\bar{y}}$	$\frac{\lvert\bar{y}_f - \bar{y}_m\rvert}{2\bar{y}} = 0.237$	$\frac{\lvert\bar{y}_m - \bar{y}_f\rvert}{2\bar{y}} = 0.237$		$\frac{\lvert\bar{y}_f - \bar{y}_m\rvert}{2\bar{y}} = 0.168$	$\frac{\lvert\bar{y}_m - \bar{y}_f\rvert}{2\bar{y}} = 0.168$	
$\frac{p_j p_h \lvert\bar{y}_j - \bar{y}_h\rvert}{2\bar{y}}$	$p_f p_m 0.237 = 0.059$	$p_m p_f 0.237 = 0.059$	$G_b = 0.118$	$p_f p_m 0.168 = 0.042$	$p_m p_f 0.168 = 0.042$	$G_b = 0.084$
G_{jh}	$G_{fm} = 0.393$	$G_{mf} = 0.393$		$G_{fm} = 0.368$	$G_{mf} = 0.368$	
$p_j s_h G_{jh}$	$p_f s_m G_{fm} = 0.119$	$p_m s_f G_{mf} = 0.073$	$G - G_w = 0.192$	$p_f s_m G_{fm} = 0.107$	$p_m s_f G_{mf} = 0.076$	$G - G_w = 0.183$

Table 9. Personal income inequality decomposition by gender, Italy 2006 and 2016.

	Bhattacaria Mahalanobis			Dagum		
	Gw	Gb	R	Gw	Gb	Gt
2006	0.180	0.118	0.074	0.180	0.155	0.037
2016	0.175	0.084	0.099	0.175	0.133	0.049

we are analysing. The first row refers to the inequality within and allows to distinguish the contributions of the female and male subgroups. Initially, the contribution of the mal subgroup is predominant, with $p_m s_m G_m / G_w = 0.124/0.180 = 0.689$ in 2006, but we can observe, from 2006 to 2016, a substantial rebalancing, with $p_m s_m G_m / G_w = 0.109/0.175 = 0.623$ in 2016.

The second and the third rows of Table 8 refer to the measurement of the inequality between following Bhattacaria and Mahalanobis, while the last two rows provide the elements to obtain $(G - G_w)$ according to the Dagum approach.

By using the data of Table 8, we can calculate the components of the Gini index decompositions by Bhattacharya and Mahalanobis and by Dagum, which are shown in Table 9. The inequality within G_w seems quite stable from 2006 to 2016 and accounts for around 49% of total inequality.

The remaining 51% of inequality can be attributed to the differences between the female and male subgroups. Within this 51%, we can observe, for the proposal by Bhattacharya and Mahalanobis, a decrease of 3.4 points of the inequality between and an increase of 2.5 points of the residual caused by the presence of overlapping. For the Dagum's proposal, the inequality between decreases of 2.2 points, while the overlapping component increases of 1.2 points. By comparing R to G_t it is possible to observe the relation $R = 2 * G_t$, derived in Section 3.3 and which links the two decompositions. It follows that the two evaluations of G_b differ for the quantity G_t, that is we can obtain the third column of Table 9 as the difference between the sixth and the seventh column: $G_{bBM} = G_{bD} - G_t$.

The general decrease of G_b represents a positive signal, since it indicates a reduction of the gender gap, but it becomes crucial to assess the relevance of the decrease, that is the reduction of the gender gap.

Our evaluations of gender gap, obtained by means of the different measures introduced in Section 4, are shown in Table 10. For the gap_5 two versions are calculated, one for the proposal by Bhatthacharya and Mahalanobis and the other for the Dagum's decomposition. All indicators suggest a decreasing income gender gap from 2006 to 2016, but the decreasing rate is not uniform across the gap_i. The greatest reduction is indicated by gap_2 and is equal to 36%; in all other cases we observe decreases of more than 25%, with the

Table 10. Gender personal income gap evaluation, Italy 2006 and 2016.

	gap_1 $(\bar{y}_m - \bar{y}_f)$	gap_2 $(\bar{y}_m - \bar{y}_f)/\bar{y}_f$	gap_3 $(\bar{y}_m - \bar{y}_f)/\bar{y}$	gap_4 $(p_f - s_f)$	gap_5 G_{bBM}/G	gap_5 G_{bD}/G
2006	8813	0.635	0.473	0.118	0.316	0.416
2016	6398	0.406	0.336	0.084	0.234	0.372
Δ0616	-27.4%	-36.1%	-29.0%	-28.8%	-25.9%	-10.6%

only exception of gap_5 for the Dagum's contribution, which indicates a variation of -10.6%. By comparing the last two columns of Table 10 we are able to perceive and to quantify the effect of overlapping in the evaluation of gender gap: its inclusion leads to a strong reduction of the improvement in terms of gender equality.

The indicators gap_1, gap_2 and gap_3, exclusively based on the subgroup means, suggest the strongest decreases in the gender gap. As part of the analysis of inequality, and referring to measures such as gap_4 and gap_5 calculated following Bhattacharya and Mahalanobis, that is without considering overlapping, the reduction of the gender gap is slightly less accentuated, while, if we consider the role of overlapping, the decrease in the gender gap is extremely reduced.

In presence of overlapping, mean-based indicators could lead to an underestimation of gender gap, and an increasing role of overlapping points to an increase of the underestimation. In order to avoid the risk to underestimate the effect of the inequality factor, in our study the gender, it is extremely useful to include in the analysis also evaluations, as gap_5 in the Dagum version, which exclude the possibility to underestimate the source of inequality.

5. CONCLUSIONS

If Gini's most famous contribution is certainly the measure of inequality that bears his name, this does not exhaust the incredible wealth of ideas and proposals formulated by Gini in the course of his scientific activity. Among Gini's other contributions, a non-secondary place is occupied by the study of overlapping, a topic to which Gini worked for several decades, starting from 1916.

Inequality and overlapping are closely interconnected and the evaluation of only one of them cannot be separated from the analysis of the other. Overlapping greatly influences the inequality structure and can be extremely useful into the assessment and the ranking of the different inequality factors.

In the work we schematize two different approaches to the study of inequality and overlapping: on the one hand the methods based on the subgroup means, on the other hand those developed on all the characteristics of the subgroup distributions. As representative examples of the two approaches we use two decompositions of the Gini index, the one proposed by Bhattacharya and Mahalanobis and the one proposed by Dagum, respectively. The comparison between the two decompositions and the analysis of one of the possible applications of the decomposition of the Gini index, the study of the gender gap, allow us to underline the usefulness of these methods and the importance of the overlapping component.

Depending on our choice about the role of overlapping units, we can get different evaluations of the inequality between the subgroups, and, consequently, of the importance of the underlying inequality factors. It is therefore necessary to be fully aware of the fact that the choice is not neutral, but can lead to an even significant underestimation of the inequality factors, as illustrated in the case study on the gender gap evaluation illustrated in Section 4.

The contributions by Gini to both inequality and overlapping measurement are still relevant and surprisingly actual, and building on them it is possible to derive an impressive quantity of information about the sources of the inequality, thus providing a solid base for any social, economic and politic action aimed at alleviate poverty and reduce social inequalities.

REFERENCES

Bhattacharya, B. and Mahalanobis, B. (1967). Regional disparities in household consumption in India, *Journal of the American Statistical Association* 62: 143–161.

Bonnet, C., Meurs, D. and Rapport, B. (2018). Gender inequalities in pensions: different components similar levels of dispersion, *Journal of Economic Inequality* 16: 527–552.

Costa, M. (2008). Gini index decomposition for the case of two subgroups, *Communications in Statistics* 37: 631–644.

Costa, M. (2009). Transvariation and inequality between subpopulations in the Dagum's Gini index decomposition, *Metron* 67: 134–120.

Costa, M. (2016). Overlapping component and inequality decomposition: a simulation study for the Gini index, *Metron* 74: 193–205.

Dagum, C. (1959). Transvariazione tra piú di due distribuzioni, *Transvariazione* C. Gini, ed., pp. 608–647, Roma: Libreria goliardica.

Dagum, C. (1987). Gini ratio, *The New Palgrave Dictionary of Economics*, London: Mac Millian Press.

Dagum, C. and Zenga, M. (1990). *Income and Wealth Distribution, Inequality and Poverty* Berlin: Springer-Verlag.

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

Deutsch, J. and Silber, J. (1997). Gini's "Transvariazione" and the measurement of distance between distributions, *Empirical Economics* 22: 547–554.

Elbers, C., Lanjouw, P., Mistiaen, J.A. and Ozler, B. (2008). Reinterpreting Between-Group Inequality, *Journal of Economic Inequality* 6: 231–245.

Frick,J., Goebel,J., Schechtman, E., Wagner, G. and Yitzhaki, S. (2006). Using analysis of Gini for detecting whether two subsamples represent the same universe: the German socio-economic panel study experience, *Sociological Methods and Research* 34: 427–468.

Frosini, B.V. (2012). Approximation and decomposition of Gini, Pietra-Ricci and Theil inequality measures, *Empirical Economics* 43: 175–197.

Gender Equality Index Report (2013). European Institute for Gender Equality.

Gini, C. (1914). Sulla misura della concentrazione e della variabilit dei caratteri, *Atti R Ist Veneto Sci Lett Arti* 73: 1203–1248; English translation in (2005) *Metron* 63: 3–38.

Gini, C. (1916). Il concetto di transvariazione e le sue prime applicazioni, reprinted in Gini, C.(1959), *Transvariazione*, pp. 1–55, Roma: Libreria goliardica.

Gini, C. (1959), *Transvariazione*. Roma: Libreria goliardica.

Giorgi, G.M. (2005). Gini's scientific work: an evergreen, *Metron* 63: 299–315.

Giorgi, G.M. (2011a). The Gini inequality index decomposition. An evolutionary study, *The measurement of individual well-being and group inequalities* J. Deutsch and J., Silber, eds., pp. 185-218, London: Routledge.

Giorgi, G.M. (2011b). Corrado Gini: the man and the scientist, *Metron* 69: 1–28.

Giorgi, G.M. and Gubbiotti, S. (2017). Celebrating the memory of Corrado Gini: a personality out of the ordinary, *International Statistical Review* 85: 325–339.

Giudici, P. and Raffinetti E. (2011). On the Gini measure decomposition, *Statistics and Probability Letters* 81: 133–139.

Goraus, K., Tyrowicz, J. and van der Velde, L. (2017). Which gender gap estimates to trust? A comparative analysis, *The Review of Income and Wealth* 63: 118–146.

Larraz, S. (2015). Decomposing the Gini inequality index: an expanded solution with survey data applied to analyze income inequality, *Sociological Methods and Research* 44: 508–531.

Mehran, F. (1975). A statistical analysis of income inequality based on a decomposition of the Gini index. A case study: Italy. in: International Statistical Institute. Proceedings of the 40th ISI Session, Warsaw.

Mornet, P., Zoli, C., Mussard, S., Sadefo-Kamdem, J., Seyte, F. and Terraza, M. (2013). The α, β -multi-level α-Gini decomposition with an illustration to income inequality in France 2005, *Economic Modelling* 35: 944–963.

Mussard, S. (2004). The bidimensional decomposition of the Gini ratio. A case study: Italy, *Applied Economics Letters* 11: 503–505.

Mussard, S., Pi Alperin, M.N., Seite, F. and Terraza, M. (2006). Extensions of Dagum's Gini decomposition, *Statistica and Applicazioni* 4: 5–30.

Mussard, S. and Richard, P. (2012). Linking Yitzhaki's and Dagum's Gini decompositions, *Applied Economics* 44: 2997–3010.

Mussard, S. and Savard, L. (2012). The Gini multi-decomposition and the role of Gini's transvariation: application to partial trade liberalization in the Philippines, *Applied Economics* 44: 1235–1249.

Mussida, C. and Picchio, M. (2014). The gender wage gap by education in Italy, *Journal of Economic Inequality* 12: 117–147.

Mussini, M. (2013). A matrix approach to the Gini index decomposition by subgroup and by income source, *Applied Economics* 45: 2457–2468.

Ogwang, T. (2014). A convenient method of decomposing the Gini index by population subgroups, *Journal of Official Statistics* 30: 91–105.

Ogwang, T. (2016). The marginal effects in subgroup decomposition of the Gini index, *Journal of Official Statistics* 32: 733–745.

Podder, N. (1993). The disaggregation of the Gini coefficient by factor components and its application to Australia, *Review of Income and Wealth* 39: 51–61.

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

Radaelli, P. (2010). On the decomposition by subgroups of the Gini index and Zenga's uniformity and inequality indexes, *International Statistical Review* 78: 81–101.

Rao, V.M. (1969). Two decompositions of concentration ratio, *Journal of the Royal Statistical Society Series A* 132: 418–425.

Selezneva, E. and Van Kerm, P. (2016). A distribution-sensitive examination of the gender wage gap in Germany, *Journal of Economic Inequality* 14: 21–40.

Shorrocks, A.F. (1980). A class of additively decomposable inequality measures, *Econometrica* 48: 613–625.

Yitzhaki, S. and Lerman, R. (1991). Income stratification and income inequality, *Review of Income and Wealth* 37: 313–329.

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