Noname manuscript No.
(will be inserted by the editor)

Analyzing the effects of negative and non-negative values on income inequality. Evidence from the Survey of Household Income and Wealth of the Bank of Italy (2012)

Emanuela Raffinetti† · **Elena Siletti**† · **Achille Vernizzi**†

Received: date / Accepted: date

Abstract Generally, inequality indices play a basic role in the analysis of welfare economics, also appearing as technical tools applied to income data. A good deal of findings in this research field is provided by the Gini coefficient, typically used for non-negative income values. Even if negative income is often an unfamiliar concept, its presence in real surveys may lead to difficulty in applying the classical Gini-based inequality measures, as it lies outside their standard ranges. In this paper, the more general issue of negative values is considered and a reformulation of the main Gini-based inequality measures adjusted for the problem of negative values is adopted with the purpose of providing theoretical extensions for the income decomposition approach by both income sources and area components. Investigations about the related inferential issues, conducted thorough simulation studies based on resampling techniques, highlight how the traditional approach of removing negative income values may yield different results in terms of inequality estimation, proving that the proposed approach, based on preserving negative values, is the more appropriate practice to follow to avoid the loss of data that really provide a coherent picture of the inequality condition.

Keywords: negative incomes, normalisation of Gini-based inequality indices, income decomposition by sources and area components, bootstrap techniques

†Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano, Via Conservatorio 7, 20122 Milano, Italy
E-mail: emanuela.raffinetti@unimi.it
E-mail: elena.siletti@unimi.it
E-mail: achille.vernizzi@unimi.it
Tel.: +39 02503 21460
Fax: +39 02503 21450

1 Introduction

An active research area that generates a wide interest in many scientific disciplines, such as economics and statistics, is currently represented by the measurement and assessment of income and other transferable attribute distributions. In such a context, as a measure of income inequality (as well as inequality of taxes, wealth, and other economic attributes), the Gini coefficient is the most popular. In the literature, the Gini coefficient application is always narrowed to distributions with non-negative values. Nevertheless, the non-negativeness condition of the attribute yields relevant restrictions in all analysis of the actual situation of inequality arising among individuals. This is because, nowadays, several events may involve the presence of negative attributes for an income unit (e.g., [26]). In fact, as we remark, there are many ways in which a negative income can arise. For instance, in real surveys, we can observe negative income values when assessing financial assets. Other important factors affecting the non-negativeness of income values are that the business of a self-employed individual might incur a loss or that an individual might provide a transfer exceeding his own total income. Negative values may appear even for tax systems, as in the case of negative taxes, caused, for example, by children's allowances. The main difficulties associated with negative values involve the violation of the normalisation principle. The inclusion of negative values may lead to a standard Gini coefficient with an upper bound greater than one, as deeply discussed by [6], [7] and further mentioned in a footnote by [18]. In applied research, presented in literature, the most common approach is based on removing units having negative values from the dataset, trusting that this exclusion will have no substantial impact on the analysis of the attribute (e.g., [29]). Nevertheless, as stated by [28], such a procedure may involve the loss of relevant information in the study of inequality, especially in reference to certain periods and certain groups. In our opinion, the deletion of negative values mainly increases the risk of obtaining biased results in all the analyses focusing on a single income source. Indeed, in many cases, the single income source can play a relevant role in the interpretation of the global economic welfare framework, above all for a period marked by an international economic crisis characterised by a variety of losses that in turn result in the presence of negative attributes. We remark that, if even one source presents a relevant number of units with negative values, these units must be eliminated not only in the distribution with negative values but also in all the distributions at stake in the decomposition, including the distribution of overall income, for the decomposition to be verified. An additional convention adopted in the current research practice is to convert the negative values into zeros (e.g. [13], [4], [5], [16]). It is quite clear that this solution is preferable to the former only when that the negative values are sufficiently small in absolute terms, compared to the positive ones; otherwise, the risk of bias still remains. For the reasons cited above and with the aim of providing a more exhaustive picture of the economic welfare framework, the classical Gini coefficient, as well as the other standard Gini-based inequality measures, must be adjusted to overcome the normalisation principle violation in the presence of negative income values. To pursue this aim, [6], subsequently improved by [7], suggest that the normalisation factor should add the part of the concentration area lying below the x-axis. In terms of absolute differences, the average, in the denominator,

should include the part of the absolute average differences involving all the negative terms and the smallest positive terms, which are enough, when summed, to compensate for the whole negative amount. The zero terms, if present, are considered within this group. In a recent work by [24], a new definition of the "polarised" scenario was introduced, with the purpose of formalising a Gini coefficient suitable to the case of negative attributes[1]. Because of the great interest in literature about the study of the role of a single income source in determining the different inequality scenarios (e.g., [3]), in this paper, the normalisation approach suggested by [24] is further extended to make it suitable to both the other Gini-based inequality measures (i.e., the concentration and re-ranking coefficients) and the decomposition of the Gini coefficient by income sources and area components. To shed light on the usefulness of the proposed extensions, an empirical analysis is provided by referring to the Household Income and Wealth survey of the Bank of Italy (2012) and considering the decomposition of household overall income into financial capital gains, on one side, and the remaining sources, considered all together, on the other side. The Gini-based indices obtained through the new normalisation approach are compared with those based on the traditional normalisation of the Gini, concentration, and re-ranking coefficients, after the units with negative values have been removed. The income source's effect on inequality within and between the three main Italian geographical areas (North Italy, Central Italy, South Italy and islands) is assessed, also including Italy as a whole. Finally, to yield inferential conclusions about the impact both on the decomposition of inequality and on each considered component caused by the elimination of units with negative values, a simulation study based on bootstrap techniques was carried out to test the significance of differences between the two different approaches. Since not only the whole country's data but also the relations across units within and between the three main Italian macro-areas were taken into account, the bootstrap tests were performed on a range of quite various scenarios. The paper is structured as follows. Section 2 includes an overview of the Gini coefficient adjusted for the negative values. In addition, the extension of the concentration coefficient as well as of the re-ranking coefficient and the decomposition approach of income, by sources and area components, to the case of negative values is proposed. Section 3 focuses on the empirical application of both the new Gini-based inequality coefficients and decomposition approach to data from the 2012 Survey of Household Income and Wealth (SHIW) of the Bank of Italy ([2]), providing a discussion of both the obtained descriptive and inferential results. Section 4 concludes.

2 An extension of the inequality methodology to a negative income scenario

The focus of this section is twofold. On one hand, an overview of the existing Gini-based inequality indices is given jointly to the description of the traditional decomposition approach of income by sources and by area components. On the other hand,

[1] Negative values may also arise when dealing with non-monetary attributes. For this reason, the normalisation introduced in [24] is relevant to both methodologists interested in index construction and applied researchers interested in using derived measures. An example of its usefulness in real-life datasets can be found in a recent contribution of [19], where the normalisation is applied to negative demographic and territorial attributes.

the extension of this methodology is considered with the aim of highlighting our proposal when dealing with the presence of negative values. Subsection 2.1 addresses the illustration of the Gini coefficient computation extended both to weighted incomes and negative values; subsection 2.2 is devoted to the re-formalisation of the concentration and re-ranking indices and traces the main steps underlying the decomposition approach of income by sources, as introduced by [21]; finally, subsection 2.3 reconsiders the decomposition proposed by [10] by providing the Gini and concentration coefficients split into a within and a gross-between component. Beside these extensions, the corresponding statistical and economic interpretation will also be provided, especially when dealing with data from the Italian household income distribution study.

2.1 The Gini coefficient adjusted for negative values

Recently, to ensure that the Gini coefficient lies between 0 and 1, [24] introduced a new normalisation. Following them, the re-formalised Gini coefficient extended to both weighted incomes and negative values is expressed as

$$G^* = \frac{\Delta_Y}{2\mu_Y^*} = \frac{1}{2\mu_Y^* N^2} \sum_{i=1}^{H} \sum_{j=1}^{H} |y_i - y_j| p_i p_j, \tag{1}$$

where Y is the vector of incomes (included the negative values), H is the total number of considered income units, p_i and p_j are weights[2] associated with y_i and y_j, such that $\sum_{i=1}^{H} p_i = N$, and μ_Y^* is the new normalisation term.

The new normalisation term μ_Y^* in Equation (1) corresponds to $\mu_Y^* = (T_Y^+ + T_Y^-)/N$, with $T_Y^+ = \sum_{i=1}^{H} max(0, y_i) p_i$ (i.e., the overall amount of the positive income values) and $T_Y^- = |\sum_{i=1}^{H} min(0, y_i) p_i|$ (i.e., the overall amount of the absolute negative income values)[3].

As stressed by [24], the new considered reference distribution, corresponding to the polarised situation and hence to the maximum inequality scenario when there are negative values, is obtained by generalising the conditions required for distributions without negative $y_1, \ldots, y_H$, where one assigns to one unit the total amount of the attribute, $T_Y = \sum_{i=1}^{H} y_i p_i$ (maintained fixed). To achieve the maximum value of Δ_Y in the case of negative income values, they do keep fixed not only the average, but also the overall amount of the positive values of the attribute, that is, T_Y^+, and the amount of the negative values of the attribute in absolute terms, that is T_Y^-. Additional basic considerations regarding the Gini coefficient behavior and role in the decomposition approach of income by sources, with and without the presence of negative values, will be further deeply discussed in the following subsection when dealing with the extension of the other main Gini-based inequality measures used for the study of income distributions.

[2] Note that the weights p_i and N may be non-integers.

[3] We observe that $T_Y^+ + T_Y^- = \sum_{i=1}^{H} |y_i| p_i$.

2.2 Re-formalising the decomposition approach by income sources for negative values

To shed light on our proposal based on a new approach to household income decomposition by sources and mainly to better appreciate the obtained results in terms of both the Gini and the other basic Gini-based inequality measures, this subsection is organised into two specific topics. First, an overview of the traditional decomposition approach framework in the classical scenario (without negative values) is given with the aim of retracing the underlying steps. Second, the same framework is revisited with the purpose of extending the household income decomposition approach to the case of negative income values. This is done to suggest the best practice to follow for the study of income distributions if the classical condition without negative values is not achieved. In addition, some crucial issues that arise when considering or avoiding negative income values in the analysis are relieved and discussed.

2.2.1 The income decomposition approach by income sources in the traditional scenario (without negative values)

Because one of the main topics illustrated in the content of this subsection addresses the decomposition approach of income by sources, let us now express the total income Y, owned by an income unit, as composed of different income sources. For the sake of simplicity, let us now suppose that the total income Y is split into two income sources, which we denote by F (corresponding to a certain income source, such as financial capital gain as considered in the application reported in Section 3) and by $D = (Y - F)$, respectively. Here, the income source D represents exactly the remaining income source, such that $Y = F + D$. Thus, for the h-th household, the total income is defined as $y_h = f_h + d_h$. To better appreciate the effect related to the proposed adjustments when negative income values are involved, let us first consider the assumption that no negative values arise for Y, F, and D. In such a scenario, the Gini coefficient of the source Z, with $Z = \{F, D\}$, trivially becomes

$$G_Z = \frac{\Delta_Z}{2\mu_Z}, \text{ where } \Delta_Z = \frac{\frac{1}{N^2}\sum_{i=1}^{H}\sum_{j=1}^{H}(z_i - z_j)p_i p_j I_{i-j}^{Z}}{2\mu_Z}, \tag{2}$$

where μ_Z represents the source Z mean value and I_{i-j}^{Z} is the indicator function such that

$$I_{i-j}^{Z} = \begin{cases} 1, \text{ if } z_i \geq z_j \\ -1, \text{ if } z_i < z_j. \end{cases}$$

Analogously, if we align the incomes of source Z according to Y, the concentration coefficient of Z can be defined as

$$C_{Z|Y} = \frac{\Delta_{Z|Y}}{2\mu_Z}, \text{ where } \Delta_{Z|Y} = \frac{1}{N^2}\sum_{i=1}^{H}\sum_{j=1}^{H}(z_i - z_j)p_i p_j I_{i-j}^{Z|Y} \tag{3}$$

and $I_{i-j}^{Z|Y}$ is the indicator function such that

$$I_{i-j}^{Z|Y} = \begin{cases} 1, \text{ if } y_i > y_j \\ -1, \text{ if } y_i < y_j \\ I_{i-j}^{Z}, \text{ if } y_i = y_j. \end{cases}$$

The concentration coefficient in (3) is based on ranking the Z values according to the values of the total income Y sorted in turn in non-decreasing order. In the literature, it is also referred to as "pseudo-Gini" because it mimics the Gini coefficient behavior except for the income source Z re-ordering criterion based on the total income Y ordering. Some references to the concentration coefficient, "pseudo-Gini", can be found in [12], [23], and [27], among others. The concentration index $C_{Z|Y}$ plays a relevant role in the Gini coefficient decomposition approach. The decomposition of the Gini coefficient by income sources was originally a result of a contribution by [25] and subsequently developed by [11], [23], [21], and [22]. More precisely, according to both [21] and [22], when considering the two sources F and D, the total income is the results from the weighted sum of the concentration indices of F and D, respectively; both income sources are ranked with respect to the distribution of total income Y:

$$G_Y = \frac{\mu_F}{\mu_Y} C_{F|Y} + \frac{\mu_D}{\mu_Y} C_{D|Y}. \tag{4}$$

From Equation (4), as $\mu_F + \mu_D = \mu_Y$, it holds that

$$\frac{\mu_F}{\mu_Y}(C_{F|Y} - G_Y) + \frac{\mu_D}{\mu_Y}(C_{D|Y} - G_Y) = 0. \tag{5}$$

As [22] stress, if the difference $(C_{Z|Y} - G_Y)$ is negative, the component Z (in such a case the component F or D) has an inequality-reducing effect. In contrast, [22] write that, if $(C_{Z|Y} - G_Y)$ is positive, the presence of income from source Z makes the total inequality higher than it would be in the absence income from the source.

If we introduce the Atkinson-Plotnick-Kakwani re-ranking index

$$R_{F|Y} = G_Z - C_{Z|Y} = \frac{1}{2\mu_Z N^2} \sum_{i=1}^{H} \sum_{j=1}^{H} (z_i - z_j) p_i p_j (I_{i-j}^{Z} - I_{i-j}^{Z|Y}), \tag{6}$$

Equation (4) can be re-expressed as

$$G_Y = (G_F - R_{F|Y}) \frac{\mu_F}{\mu_Y} + (G_D - R_{D|Y}) \frac{\mu_D}{\mu_Y}. \tag{7}$$

$R_{Z|Y}$ measures the lack of co-graduation between Z and Y and it lies in the close range $[0, 2G_Z]$, whence it derives that $0 \leq \frac{R_{Z|Y}}{G_Z} \leq 2$. By splitting G_Y into $(\mu_F/\mu_Y)G_Y$ and $(\mu_D/\mu_Y)G_Y$, after some manipulations, from (7), one can derive the following expression[4]:

$$G_Y - G_D = \frac{\mu_F}{\mu_D}(C_{F|Y} - G_Y) - R_{D|Y}. \tag{8}$$

[4] This decomposition is also used by [17] to assess the redistribution effects of taxes.

An expression in (8) illustrates the overall effect on inequality caused by source F. From Equation (8), we can note that, even if $(C_{F|Y} - G_Y)$ is positive, the overall effect of source F can be inequality reducing if $R_{D|Y}$ is greater than $\frac{\mu_F}{\mu_D}(C_{F|Y} - G_Y)$, that is, if in summing F to D, the ranking of D significantly differs from that of the resulting distribution $Y = D + F$. If the distribution of F is counter-graduated with respect to that of D, the highest values in F can compensate for the lowest ones in D and vice-versa so that a significant positive $(C_{F|Y} - G_Y)$ can be overcome by subtracting the component $R_{D|Y}$, especially when the coefficient (μ_F/μ_D) is small. Obviously, the exercise of totally removing a source may make sense only when the source at stake is not a relevant one.

2.2.2 The income decomposition approach by income sources in the negative value scenario

In the previous subsection, an overview of the traditional household income decomposition approach by income sources was provided and enriched with several references to the different contributions in the literature. All the basic considerations involved the classical hypothesis of no negative income were checked for any source. Let us now suppose that the assumption of non-negative values is not fulfilled, meaning that, for instance, the total income Y and/or the income source Z present some negative values. In this scenario, two different methods of proceeding may be considered. The first proposal is to erase all negative values. Note that, even though it appears to be the more common solution to the problem, this method deserves appropriate devices, especially for the decomposition approach. $G_Y = \frac{\Delta_Y}{2\mu_Y}$ can be expressed as $\frac{\Delta_Y}{2\mu_Y} = \frac{\Delta_{F|X}}{2\mu_Y} + \frac{\Delta_{D|X}}{2\mu_Y}$ and consequently $\Delta_Y = \Delta_{F|X} + \Delta_{D|X}$, meaning that

$$\frac{1}{N^2}\sum_{i=1}^{H}\sum_{i=1}^{H}(y_i - y_j)p_i p_j I_{i-j}^{Y} = \frac{1}{N^2}\sum_{i=1}^{H}\sum_{i=1}^{H}(f_i - f_j)p_i p_j I_{i-j}^{F|Y} + \frac{1}{N^2}\sum_{i=1}^{H}\sum_{i=1}^{H}(d_i - d_j)p_i p_j I_{i-j}^{D|Y}. \quad (9)$$

Such a relation requires that the same pairs be simultaneously considered in both the term on the left side and the two terms on the right side of Equation (9). Therefore, if we delete units with negative F values, the same units must be deleted both from D and Y. Generally, if the h-th household presents a negative value for one source, the h-th household should be erased to verify the decompositions in Equations (4), (7), and (8). An alternative procedure that we suggest when focusing on the household income decomposition approach by sources is not to delete any negative value from the analysis, why we resort to crucial adjustments. This allows us to avoid any drop-out, preserving invariant the Gini and concentration coefficients in the absolute mean difference formulas. According to [24] presented in subsection 2.1 the normalisation terms μ_F and μ_D must be replaced by $\mu_F^* = (T_F^+ + T_F^-)/N = (\sum_{i=1}^{H}|f_i|p_i)/N$ and $\mu_D^* = (T_D^+ + T_D^-)/N = (\sum_{i=1}^{H}|d_i|p_i)/N$. As a consequence, if we think of Δ_Y, as the sum of $\Delta_{F|Y}$ and $\Delta_{D|Y}$, when no compensation happens between the F and D

distribution, the maximum for Δ_Y should be equal to the maximum for Δ_F plus the maximum for Δ_D, so $\mu_Y^* = \mu_F^* + \mu_D^*$. It is worth stressing that, when negative values are not erased from the analysis, the distribution of Y and, consequently, the associated Δ_Y are invariant regardless of the sources considered. Only μ_Y^* depends on the sources into which Y is split. Conversely, when units presenting negative values for one or more sources are removed, these units must necessarily also be deleted for Y so that both μ_Y and Δ_Y depend on the sources at stake[5].

2.3 The income decomposition approach by area components in the traditional and negative values scenarios

Beside the decomposition approach by income sources and the related estimation of the contribution of each component to the total inequality, a further basic decomposition approach of inequality focuses on the area components, defined as the territorial macro-areas that are part of a country. Through such decomposition, an exhaustive picture of the regions that mainly affect the country's inequality scenario can be provided. The decomposition approach by area components presented here aims at retracing the same steps underlying the proposal of [10], both for the classical situation of non-negative income values and for the case where even negative values are involved in the analysis. Let us consider two pairs of macro-regions, a and b. For the sake of coherence with the terminology used by [10], when territorial areas are taken into account, the Gini coefficient for income source Z is defined as[6]

$$\begin{aligned} G_Z = & G_Z^a \frac{N_a^2 \nu_Z^a}{(N_a+N_b)(N_a\nu_Z^a + N_b\nu_Z^b)} + G_Z^b \frac{N_b^2 \nu_Z^b}{(N_a+N_b)(N_a\nu_Z^a + N_b\nu_Z^b)} \\ & + \frac{N_a N_b (\nu_Z^a + \nu_Z^b)}{(N_a+N_b)(N_a\nu_Z^a + N_b\nu_Z^b)} G_Z^{GB}, \end{aligned} \tag{10}$$

where G_Z^a and G_Z^b represent the Gini coefficients of income source Z *within* the macro-areas a and b, G_Z^{GB} denotes the Gini coefficient *gross-between* for income source Z, N_a and N_b are such that $N_a = \sum_{i=1}^{H_a} p_i$ and $N_b = \sum_{i=1}^{H_b} p_i$ (with H_a and H_b being the total number of income units belonging to the macro-areas a and b, respectively), and ν_Z^a and ν_Z^b are the normalisation terms for macro-areas a and b. Note that, when assuming the classical condition of non-negative income values, ν_Z^a and ν_Z^b exactly correspond to μ_Z^a and μ_Z^b; when negative values are included, ν_Z^a and ν_Z^b should be μ_Z^{a*} and μ_Z^{b*}, respectively.

The *gross-between* component G_Z^{GB} evaluates the inequalities between income units belonging to different macro-areas, in the same manner as the decomposition proposed by [25]. It can be computed as

[5] The distribution of Y and its average also change when the negative values of a sources are replaced by zeros.

[6] If the total income Y is considered, it would be enough to replace the subscript Z with the subscript Y in Equation (10).

$$G_Z^{GB} = \frac{\Delta_Z^{GB}}{\nu_Z^a + \nu_Z^b} \text{ with } \Delta_Z^{GB} = \frac{1}{N_a N_b} \sum_{i=1}^{H_a} \sum_{i=1}^{H_b} (z_{a,i} - z_{b,j}) p_{a,i} p_{b,j} I_{i-j}^Z. \tag{11}$$

As stated by [20], the expression "gross-between component" singles out G^{GB} from the traditional between-group measures, which are based only on mean incomes[7]. Analogously, the concentration coefficient $C_{Z|Y}$ can be split both into a within and a gross-between component. By retracing the same steps for the Gini of the income source Z decomposition into macro-areas (Equation 10), the concentration coefficient $C_{Z|Y}$ is obtained as follows

$$\begin{aligned} C_{Z|Y} =& C_{Z|Y}^a \frac{N_a^2 \nu_Z^a}{(N_a + N_b)(N_a \nu_Z^a + N_b \nu_Z^b)} + C_{Z|Y}^b \frac{N_b^2 \nu_Z^b}{(N_a + N_b)(N_a \nu_Z^a + N_b \nu_Z^b)} \\ &+ \frac{N_a N_b (\nu_Z^a + \nu_Z^b)}{(N_a + N_b)(N_a \nu_Z^a + N_b \nu_Z^b)} C_{Z|Y}^{GB}, \end{aligned} \tag{12}$$

where

$$C_{Z|Y}^{GB} = \frac{\Delta_{Z|Y}^{GB}}{\nu_Z^a + \nu_Z^b} \text{ with } \Delta_{Z|Y}^{GB} = \frac{1}{N_a N_b} \sum_{i=1}^{H_a} \sum_{i=1}^{H_b} (z_{a,i} - z_{b,j}) p_{a,i} p_{b,j} I_{i-j}^{Z|Y}. \tag{13}$$

From Equations (11) and (13), it follows that

$$R_{Z|Y}^{GB} = G_Z^{GB} - C_{Z|Y}^{GB}. \tag{14}$$

Even in this case, if the assumption of non-negative income values is not fulfilled, in Equations (12) and (13), the normalisation terms denoted by ν_Z^a and ν_Z^b must be substituted by the normalisation terms μ_Z^{a*} and μ_Z^{b*}, adjusted for negative values. Therefore, when the normalisation factor is μ_Z^*, the Gini, concentration, and re-ranking indices will be remarked by the apex "*". As a consequence, G_Y, G_Z, $C_{Z|Y}$, and $R_{Z|Y}$ become G_Y^*, G_Z^*, $C_{Z|Y}^*$, and $R_{Y|Z}^*$, respectively.

3 Application to the SHIW data of the Bank of Italy (2012)

The income inequality theoretical approach proposed in the current paper needs to be addressed through an illustrative application to real data. To provide an interpretation of results in all case studies involving the analysis of income distributions, even those characterised by negative values. For this purpose, data collected by the Bank of Italy's 2012 Survey of Household Income and Wealth (SHIW) ([2]) were taken into account. The SHIW began in the 1960s with the aim of gathering data on the incomes and savings of Italian households. The 2012 survey covered 8,151 households

[7] As [10] shows, the term Δ_Z^{GB} could be further split into the *between* and *transvariation* components. The former depends on the averages of a and b, and the latter (e.g., [15], [8], and [9]) arises from the fact that the income differences are of opposite signs compared to the difference in their corresponding mean incomes.

Table 1 Descriptive statistics for equivalent total net income (Y)

Total net income (Y)	North	Centre	South and Islands	Italy
Number of households	3,512	1,720	2,919	8,151
Sum of weights	8,415.31	3,523.93	6,166.01	18,105.25
Households with negative Y values	0.057%	-	-	0.025%
Households with zero Y values	0.028%	0.116%	0.171%	0.098%
Households with positive Y values	99.915%	99.884%	99.829%	99.877%
MIN	-543.15	0	0	-543.15
MAX	158,019.97	201,517.86	105,077.00	201,517.86
Average	15,832.06	14,833.12	10,075.66	13,677.20
CV	0.6940	0.6626	0.7520	0.7359
Skewness	3.38	2.79	4.15	3.36
Kurtosis	26.60	24.82	40.34	27.80

and 20,022 persons, distributed over about 300 Italian municipalities. To take into account the lack of homogeneity among household caused by the different numbers of components, their ages, and the number of income earners per family, the relative equivalence scale suggested by [17] was applied to the sum of monetary incomes in household h. The scale is given by the expression

$$sd_i = (ad_i + 0.2ch_{1,i} + 0.4ch_{2,i} + 0.7ch_{3,i})^{0.8} + 0.1w_i, \text{ where } i = 1,2,...,H, \quad (15)$$

where H is the number of families (i.e., $8,151$), ad is the number of adults within the family, ch_1 is the number of children age 5 years or under, ch_2 is the number of children between the ages of 6 and 14 years, ch_3 is the number of children between the ages of 15 and 17 years, w is the number of employees or self-employed people within the families, and 0.8 is the parameter that indicates the economies of scale attached to the equivalence scale. The equivalent income, yielded by dividing the total nominal income by the scale in (15), is the equivalent income for one adult whose income is perceived without any working activity. Each equivalent income is then associated with a weight that is given by multiplying its scale and the weight given in SHIW.[8]

Since our aim is to shed light on the real attitude of our proposed approach in providing a more coherent interpretation of the actual inequality scenario, we focused on the total net income Y, defined as the sum of two sources: financial capital gain F and all other income sources considered together and hence denoted by $D = Y - F$. The usefulness of our proposed income inequality decomposition approach by sources finds grounds in the presence of negative values for total income and the two composing sources.

Looking at Figures 1, 2 and 3 we can see how the three types of income [9] (i.e., the total net income, the financial capital gain, and the remaining income sources) are distributed differently both for Italy as a whole and for the three macro-areas (North, Centre, South and islands). Tables 1, 2, and 3 report the main descriptive statistics of

[8] In so doing, the weight represents the number of equivalent components in the family (given by the scale) and the representativeness of the sampled family with respect to the Italian population. $N = \sum_{i=1}^{H} p_i = 18,105.25$ is lower than the total number of persons in SHIW, which is $20,022$.

[9] The density functions are plotted by the Gaussian smoothing kernel. Incomes are expressed in equivalent terms.

Fig. 1 Density plot of equivalent total income

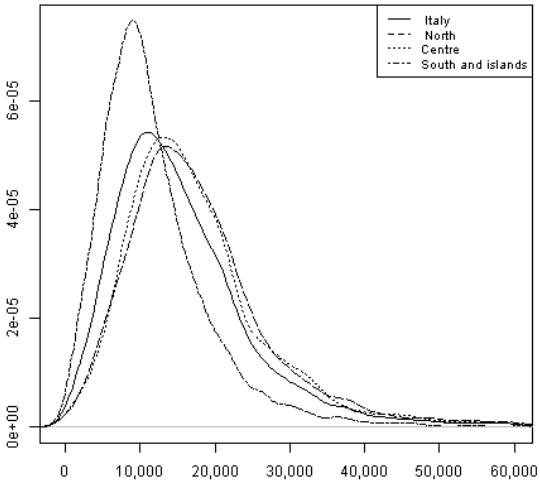
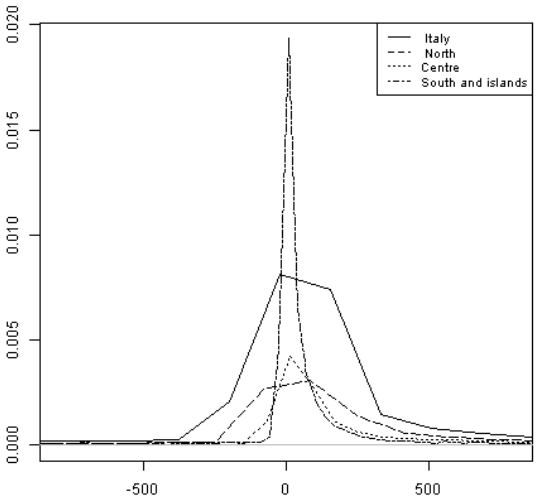


Fig. 2 Density plot of equivalent financial gain

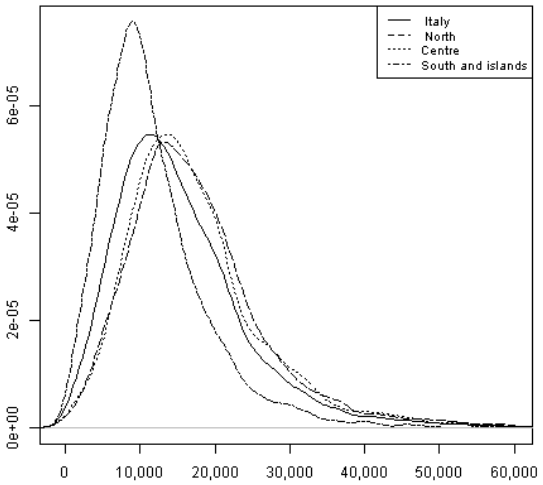


the three equivalent income items. The impact of the negative values, with regard to the three sources, is quite different according to the income at stake. Even if the per-

Table 2 Descriptive statistics for equivalent financial capital gain (F)

Financial capital gain (F)	North	Centre	South and Islands	Italy
Number of households	3,512	1,720	2,919	8,151
Sum of weights	8,415.31	3,523.93	6,166.01	18,105.25
Households with negative F values	11.133%	12.849%	7.023%	10.023%
Households with zero F values	11.674%	11.337%	31.483%	18.697%
Households with positive F values	77.192%	75.814%	61.494%	71.280%
MIN	-11,033.83	-17,123.16	-7,245.66	-17,123.16
MAX	73,127.16	23,640.83	9,235.74	73,127.16
Average	148.62	-66.91	19.12	62.57
CV	11.2774	15.8256	34.3651	20.7054
Skewness	16.53	0.25	7.12	17.08
Kurtosis	562.30	79.50	108.79	745.97

Fig. 3 Density plot of equivalent remaining income



centages of Italian households with both a total negative income (0.025%) and negative remaining income sources (0.037%) are indeed insignificant percentages, the same consideration does not apply to the percentage of households with a negative financial capital gain (10.023%). Moreover, the percentages are different within the three areas, especially in the South and islands area. The three areas present different kurtosis and skewness indices, especially in relation to F (Table 2). If, to compute the standard Gini-based inequality measures, the procedure of removing all these households from the dataset is chosen, a non-negligible loss of information occurs. This is because, as discussed in Section 2, when the analysis, as in this case, concerns sources and not just overall incomes, the same families must be erased when considering their overall income. The need to preserve all the data to ensure a non-biased quantification of the actual income inequality scenario may drive researchers in the economics and

Table 3 Descriptive statistics for equivalent remaining income sources ($D = Y - F$)

Remaining income sources (D)	North	Centre	South and Islands	Italy
Number of households	3,512	1,720	2,919	8,151
Sum of weights	8,415.31	3,523.93	6,166.01	18,105.25
Households with negative D values	0.085%	-	-	0.037%
Households with zero D values	0.114%	0.116%	0.206%	0.147%
Households with positive D values	99.801%	99.884%	99.794%	99.816%
MIN	-2,271.82	0	0	-2,271.82
MAX	140,260.60	198,363.64	109,010.85	198,363.64
Average	15,683.44	14,900.03	10,056.53	13,614.63
CV	0.6584	0.6507	0.7418	0.7099
Skewness	3.01	2.74	4.31	3.10
Kurtosis	22.02	24.66	44.67	24.68

statistics fields to favor the use of the proposal illustrated in this paper. Commonly, negative incomes or even low positive incomes present a functional form that is not symmetric to that of larger positive ones. Thus, they cannot be represented with a linear functional form. This issue also arises in the distributions of the total income and the two considered income sources F and D. Indeed, the three income distributions are asymmetric. Consequently, the study of the relationship between the total income and the single income sources cannot be correctly assessed by employing symmetric measures, such as Pearson's correlation coefficient. The indices proposed in the current paper and used for both the income decomposition into income sources and in terms of macro-areas employ the concentration coefficient, which is an asymmetrical measure. Indeed, given two variables, it is based on the values of one variable ranked according to the values of the other variable. Such an asymmetry is an intrinsic condition in regression analysis. Thus, imposing a symmetric correlation on the data may in some cases affect the sign of the correlation. To avoid misleading results in econometric studies, where the key assumptions can be modified, a symmetric approach can be taken into account if supported by a sensitivity analysis (e.g., [30] and [31]), whose role is to allow researchers to show how their findings vary with changes in specification or functional form (e.g., [1]). An example of sensitivity analysis in the income distribution field study can be found in [14].

By comparing the re-ranking index (Equation (7)) $R^*_{F|Y}$ with $R_{F|Y}$ and $R^*_{D|Y}$ with $R_{D|Y}$, the former of each of the two pairs (labeled by "*") being calculated without erasing any observations and the latter considering only non-negative values, the loss of information caused by the elimination of negative values is evident. Table 4 reports the four indices both at levels and as percentages of the related Gini coefficients for the whole country, within each of the three macro-regions, and gross-between the three macro-regions, considered two by two. Let us focus on the percentages, which do not depend on the normalisation factors but only on the average differences, Δ's. We can see that the $R^*_{F|Y}$'s are never less than two and a half times the corresponding $R_{F|Y}$ (North) and that it is more than four and a half times for the gross-between when we consider the Centre with the South and islands area. Even if the percentages calculated for $R^*_{D|Y}$ and $R_{D|Y}$ are much lower, $R^*_{D|Y}/G^*_D$, when compared to the corresponding $R_{D|Y}/G_D$, is similarly never lower than two and a half times, and it can be even greater than four (within Centre and gross-between Centre/South and islands).

Table 4 $R_{F|Y}$, $R_{D|Y}$, $R^*_{F|Y}$, and $R^*_{D|Y}$ at levels and as percentages of G_F and G_D

	Whole dataset		**Non-negative values**	
	$R^*_{F\|Y}$	$R^*_{D\|Y}$	$R_{F\|Y}$	$R_{D\|Y}$
	Whole country			
Italy	0.5540	0.0011	0.1564	0.0003
	Within macro-regions			
North	0.4395	0.0012	0.1666	0.0005
Centre	0.7319	0.0014	0.2162	0.0003
South and islands	0.6231	0.0006	0.1708	0.0002
	Gross-between macro-regions			
North/Centre	0.5684	0.0013	0.1847	0.0004
North/South and islands	0.5359	0.0008	0.1189	0.0002
Centre/South and islands	0.8001	0.0010	0.1627	0.0002
	$(R^*_{F\|Y}/G^*_F)*100$	$(R^*_{D\|Y}/G^*_D)*100$	$(R_{F\|Y}/G_F)*100$	$(R_{D\|Y}/G_D)*100$
	Whole country			
Italy	61.7	0.31	18.18	0.10
	Within macro-regions			
North	50.12	0.39	19.99	0.16
Centre	82.58	0.44	26.07	0.10
South and islands	67.98	0.18	19.45	0.05
	Gross-between macro-regions			
North/Centre	64.07	0.42	21.91	0.13
North/South and islands	58.29	0.22	13.28	0.07
Centre/South and islands	87.31	0.26	18.74	0.06

Hence, an investigation of the related inferential issues is basic to confirm such differences. Inferential issues are examined here through a simulation study based on bootstrap resampling techniques. First, 1,000 randomly selected samples for Y, D, and F were drawn, with replacement, from the whole dataset (to include negative values). Second, 1,000 randomly selected samples for Y, D, and F were drawn, with replacement, from the whole dataset, removing observations presenting negative values for Y, D, or F. Through the obtained simulation findings, inferential conclusions about the impact of the income sources, with and without negative values, on the Gini-based inequality measures are derived. Simulations were carried out within and between macro-areas, with different distributions for the same income item, to yield more reliable indications. The bootstrap results for the Gini-based inequality measures presented in the paper with respect to Italy and the macro-regions are shown in Table 5, where, in addition to the Gini-based inequality measure bootstrap estimates, the pseudo-t (expressed as the ratio between the bootstrap estimate and its related standard error) is also reported, within parentheses, for the purpose of validating the reliability of the obtained estimates. Figure 4 summarises the bootstrap point estimates of the indices reported in Table 5. Typically, if all the income units are included, the estimates associated with the considered Gini-based measures are lower than those obtained by erasing units presenting losses, except for the Gini coefficients of the financial capital gain F. Another striking finding is that the distribution of F, when ranked according to Y, yields a concentration coefficient that is lower than that computed excluding negative values. These two simultaneous effects explain the reason that the $R^*_{F|Y}$ indices are greater than the $R_{F|Y}$ ones, as we previously observed.

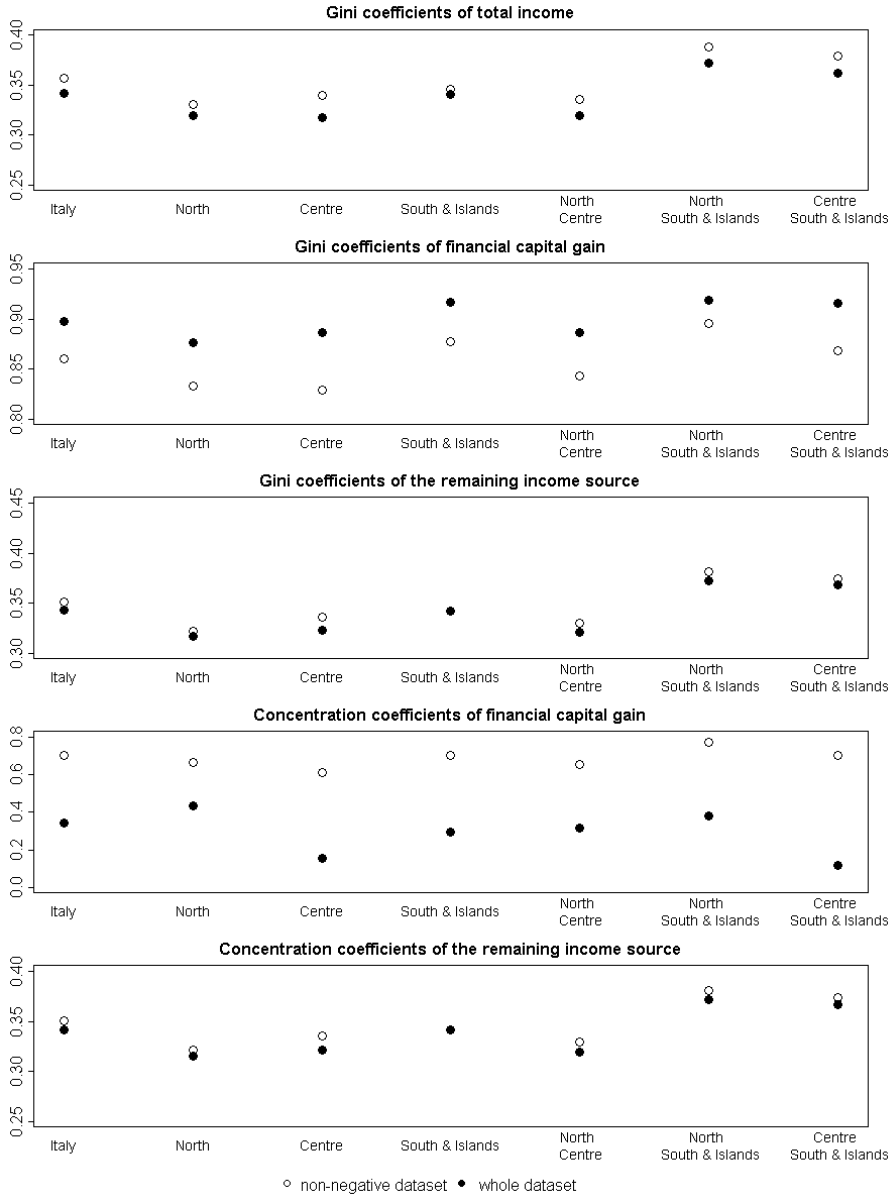
Table 5 Bootstrap results for random sampling from the non-negative dataset and from the whole dataset for Italy, within macro-regions, and gross-between macro-regions

Gini-based inequality measures (non-negative dataset)	***Estimate (pseudo-t)***	***Gini-based inequality measures (whole dataset)***	***Estimate (pseudo-t)***
	Italy		
G_Y	0.3572 (72.90)	G_Y^*	0.3419 (74.33)
G_F	0.8601 (128.37)	G_F^*	0.8979 (80.59)
G_D	0.3509 (74.66)	G_D^*	0.3429 (76.20)
$C_{F\|Y}$	0.7037 (41.15)	$C_{F\|Y}^*$	0.3439 (11.06)
$C_{D\|Y}$	0.3505 (74.57)	$C_{D\|Y}^*$	0.3418 (75.96)
	North		
G_Y	0.3301 (47.16)	G_Y^*	0.3191 (50.65)
G_F	0.8333 (84.17)	G_F^*	0.8769 (171.94)
G_D	0.3221 (45.37)	G_D^*	0.3165 (48.69)
$C_{F\|Y}$	0.6667 (30.87)	$C_{F\|Y}^*$	0.4374 (13.17)
$C_{D\|Y}$	0.3216 (45.30)	$C_{D\|Y}^*$	0.3152 (47.76)
	Centre		
G_Y	0.3393 (28.28)	G_Y^*	0.3174 (30.52)
G_F	0.8292 (78.97)	G_F^*	0.8863 (138.48)
G_D	0.3357 (29.45)	G_D^*	0.3233 (32.33)
$C_{F\|Y}$	0.6130 (18.69)	$C_{F\|Y}^*$	0.1544 (2.39)
$C_{D\|Y}$	0.3354 (26.62)	$C_{D\|Y}^*$	0.3219 (28.24)
	South and islands		
G_Y	0.3461 (40.24)	G_Y^*	0.3405 (37.83)
G_F	0.8779 (34.84)	G_F^*	0.9167 (100.74)
G_D	0.3424 (29.45)	G_D^*	0.3420 (38.00)
$C_{F\|Y}$	0.7071 (10.96)	$C_{F\|Y}^*$	0.2936 (2.94)
$C_{D\|Y}$	0.3422 (40.74)	$C_{D\|Y}^*$	0.3413 (36.70)
	North/Centre		
G_Y	0.3361 (46.04)	G_Y^*	0.3191 (52.31)
G_F	0.8430 (105.38)	G_F^*	0.8871 (216.37)
G_D	0.3301 (47.16)	G_D^*	0.3205 (53.42)
$C_{F\|Y}$	0.6582 (32.58)	$C_{F\|Y}^*$	0.3187 (8.54)
$C_{D\|Y}$	0.3297 (48.49)	$C_{D\|Y}^*$	0.3191 (53.18)
	North/South and islands		
G_Y	0.3887 (61.70)	G_Y^*	0.3722 (65.30)
G_F	0.8954 (144.42)	G_F^*	0.9194 (270.41)
G_D	0.3814 (60.54)	G_D^*	0.3728 (62.13)
$C_{F\|Y}$	0.7765 (44.63)	$C_{F\|Y}^*$	0.3835 (10.71)
$C_{D\|Y}$	0.3811 (59.55)	$C_{D\|Y}^*$	0.3719 (61.98)
	Centre/South and islands		
G_Y	0.3788 (49.84)	G_Y^*	0.3616 (51.66)
G_F	0.8684 (87.72)	G_F^*	0.9164 (213.12)
G_D	0.3748 (51.34)	G_D^*	0.3682 (53.36)
$C_{F\|Y}$	0.7057 (24.43)	$C_{F\|Y}^*$	*0.1163 (1.81)*
$C_{D\|Y}$	0.3746 (50.62)	$C_{D\|Y}^*$	0.3673 (51.73)

As shown in Table 5, all the Gini-based inequality measures present great pseudo-t, except for the concentration coefficient $C_{F|Y}^*$ referring to the macro-region Centre/South and island, where the pseudo-t value is 1.81; these results provide evidence on the reliability of the estimates.

To better highlight the differences in value between the Gini-based inequality measures computed on the whole dataset and on the dataset without negative values, the bootstrap sampling technique was extended to the quantities $G_Y^* - G_Y$, $G_F^* - G_F$,

Fig. 4 Bootstrap results for random sampling from the non-negative dataset and the whole dataset reported in Table 5 for Italy, within macro-regions, and gross-between macro-regions



$G_D^* - G_D$, $C_{F|X}^* - C_{F|X}$, and $C_{D|X}^* - C_{D|X}$. Through this analysis, we want to stress that the traditional methodology, based on erasing units with negative values, may yield estimates that are significantly different from those obtained by preserving all

the data. This conclusion is validated by most of the results expressed in Tables 6 and 7 and depicted in Figures 5, 6, 7, 8, and 9.

Table 6 Differences in value between the Gini indices computed on the whole dataset and the non-negative dataset

	$G^*_Y - G_Y$	*Pseudo-t*	$G^*_F - G_F$	*Pseudo-t*	$G^*_D - G_D$	*Pseudo-t*
	Whole country					
Italy	-0.0153	-7.58	0.0378	9.00	-0.0080	-4.04
	Within macro-regions					
North	-0.0109	-4.35	0.0437	8.31	-0.0056	-2.23
Centre	-0.0218	-5.12	0.0571	6.38	-0.0124	-3.16
South and islands	-0.0056	*-1.13*	0.0388	2.39	-0.0004	*-0.08*
	Gross-between macro-regions					
North/Centre	-0.0170	-6.73	0.0441	8.28	-0.0097	-4.14
North/South and islands	-0.0165	-6.20	0.0240	6.71	-0.0086	-3.15
Centre/South and islands	-0.0172	-5.55	0.0480	6.55	-0.0066	*-1.89*

Table 7 Differences in value between the concentration indices computed on the whole dataset and the non-negative dataset

	$C^*_{F\|Y} - C_{F\|Y}$	*Pseudo-t*	$C^*_{D\|Y} - C_{D\|Y}$	*Pseudo-t*
	Whole country			
Italy	-0.3598	-13.74	-0.0087	-4.28
	Within macro-regions			
North	-0.2292	-8.69	-0.0064	-2.54
Centre	-0.4586	-6.54	-0.0135	-3.54
South and islands	-0.4135	-7.32	-0.0009	*-0.17*
	Gross-between macro-regions			
North/Centre	-0.3395	-9.20	-0.0106	-4.41
North/South and islands	-0.3930	-12.74	-0.0092	-3.42
Centre/South and islands	-0.5894	-10.07	-0.0073	-2.18

The differences $G^*_Y - G_Y$, $G^*_F - G_F$, and $G^*_D - G_D$ are reported in Table 6, while $C^*_{F|Y} - C_{F|Y}$ and $C^*_{F|D} - C_{D|Y}$ are displayed in Table 7. Along with these differences, the tables report the associated bootstrap pseudo-*t* values. For $C^*_{F|Y} - C_{F|Y}$ and $G^*_F - G_F$, we observe that, in the former case, the pseudo-*t* varies from -6.54 to -13.74; in the latter it varies from 2.39 to 9.00. These findings are reflected in what arises from Figures 6 and 8, where the corresponding 95% bootstrap confidence intervals (*CI*) are plotted for $G^*_F - G_F$ and $C^*_{F|Y} - C_{F|Y}$, respectively. More precisely, in Figure 6, all the *CI* ranges are positive, whilst, in Figure 8, they are all negative. This means that we must reject both the hypothesis that $G^*_F = G_F$ and that $C^*_{F|Y} = C_{F|Y}$, at the 5% significance level in all the considered cases, as no *CI* encompasses the value of zero.

For $G^*_Y - G_Y$, all the *CI*s, except the one associated with the South and islands macro-region, include only negative values. By looking at Table 6, we can see that this macro-region presents a pseudo-*t* that is relatively low (in absolute terms), being equal to -1.13, whilst, in the remaining six cases, it ranges from -4.35 to -7.58.

The South and islands macro-region presents *CI* plots that include both positive and negative values even for $G_D^* - G_D$ and $C_{D|Y}^* - C_{D|Y}$ (Figures 7 and 9), which is somewhat confirmed by the corresponding pseudo-t values, which are -0.08 (Table 6) and -0.17 (Tables 7), respectively. Borderline cases are represented by the gross-between $G_D^* - G_D$ and $C_{D|Y}^* - C_{D|Y}$ in the relation between the Centre and the South and islands. In the former case, the *CI* includes a small segment of positive values (the pseudo-t is -1.89), and here, we cannot reject the hypothesis $G_D^* = G_D$; in the latter, its higher limit is a bit lower than zero (the pseudo-t is -2.18), so we reject the condition $C_{D|Y}^* = C_{D|Y}$ at the 5% significance level. In the remaining five cases, the *CI*s lie entirely under the zero line, and consequently, we can reject the hypothesis that it is neutral to include or erase units with negative values in at least one source.

Fig. 5 Confidence intervals for the $G_Y^* - G_Y$ bootstrap estimate reported in Table 6

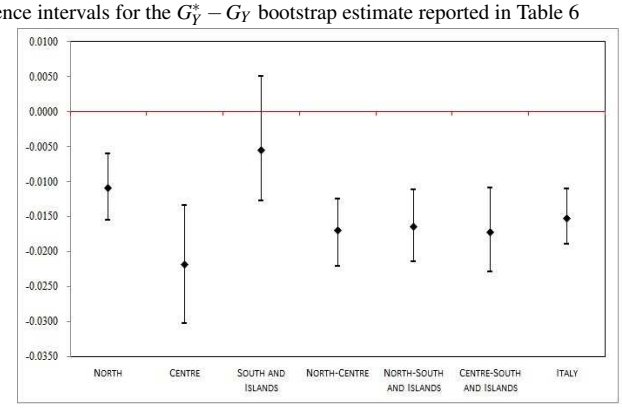
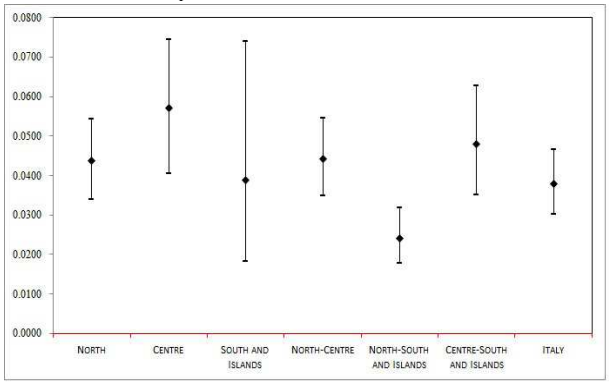


Fig. 6 Confidence intervals for the $G_F^* - G_F$ bootstrap estimate reported in Table 6



The presence of non-significant pseudo-t tests and overlapping confidence intervals in some macro-areas, both for the total net income Y and the income source D,

Fig. 7 Confidence intervals for the $G_D^* - G_D$ bootstrap estimate reported in Table 6

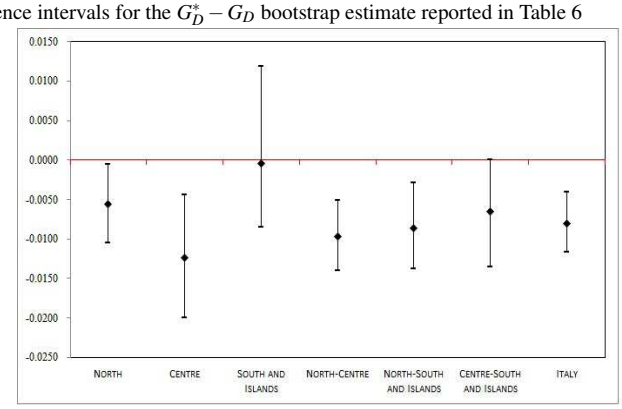


Fig. 8 Confidence intervals for the $C_{F|Y}^* - C_{F|Y}$ bootstrap estimate reported in Table 7

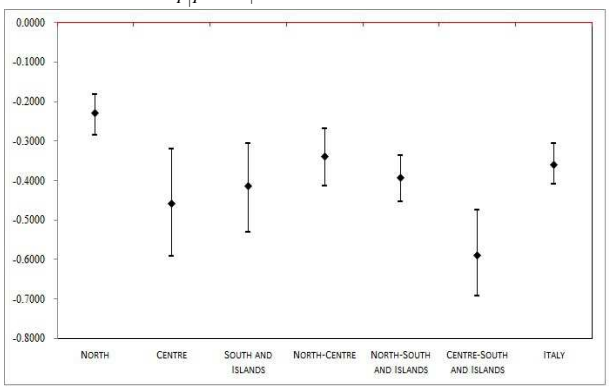
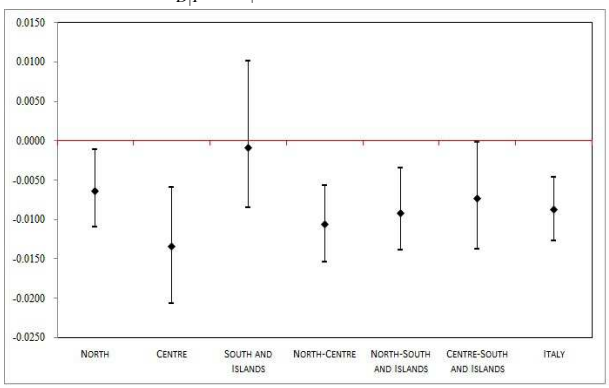


Fig. 9 Confidence intervals for the $C_{D|Y}^* - C_{D|Y}$ bootstrap estimate reported in Table 7



is caused by the very narrow percentage of negative values. Non-significant results never occur for the income source F, which actually covers over 10% of the nega-

Table 8 Bootstrap results for $G_Y - G_D$ and $G_Y^* - G_D^*$ - Italy, within macro-regions and gross-between macro-regions

Gini-based inequality measures (non-negative dataset)	***Estimate***	***Gini-based* inequality measures (whole dataset)**	***Estimate***
Italy			
$G_Y - G_D$	0.0063	$G_Y^* - G_D^*$	-0.0010
North			
$G_Y - G_D$	0.0079	$G_Y^* - G_D^*$	0.0027
Centre			
$G_Y - G_D$	0.0036	$G_Y^* - G_D^*$	-0.0059
South and islands			
$G_Y - G_D$	0.0037	$G_Y^* - G_D^*$	-0.0014
North/Centre			
$G_Y - G_D$	0.0060	$G_Y^* - G_D^*$	-0.0014
North/South and islands			
$G_Y - G_D$	0.0074	$G_Y^* - G_D^*$	-0.0005
Centre/South and islands			
$G_Y - G_D$	0.0040	$G_Y^* - G_D^*$	-0.0067

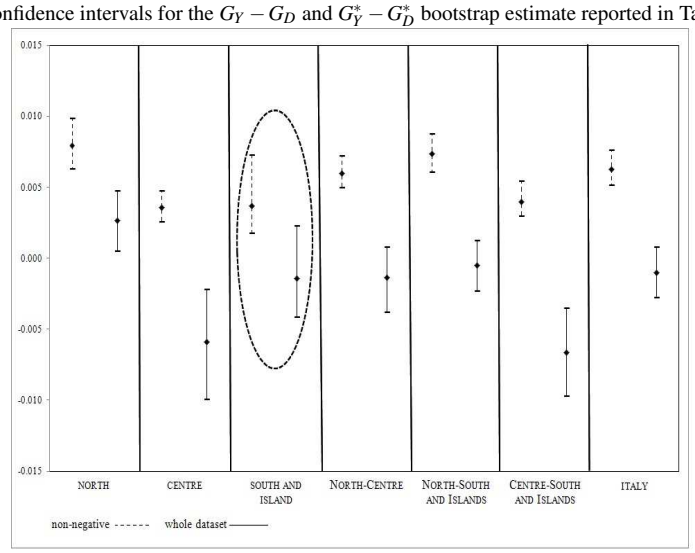
tive values for Italy as a whole. Therefore, we believe that a further investigation of the impact of the income source F on income inequality may be interesting, especially when comparing the performance of two different approaches in dealing with the presence of negative values. To determine the overall contribution provided by income source F on income inequality, a bootstrap simulation study was carried out on both the difference between the Gini coefficient of the total net income Y and the Gini coefficient of the income source D, when resampling from the non-negative dataset (i.e., $G_Y - G_D$), and on the difference between the same Gini coefficients when resampling from the whole dataset (i.e., $G_Y^* - G_D^*$). The obtained estimates are displayed in Table 8, while in Figure 10, the plots of the corresponding confidence intervals are shown.

The estimates in Table 8 indicate that the difference $G_Y^* - G_D^*$ is always smaller in value than the difference $G_Y - G_D$, leading to the conclusion that the original income source F as it is (that is, including the negative values) plays a non-negligible role in determining the inequality income scenario. The need to preserve all the data to ensure that an unbiased estimation of the existing inequality situation also arises from the analysis of the $G_Y - G_D$ and $G_Y^* - G_D^*$ confidence intervals in Figure 10. In fact, all the confidence intervals do not overlap[10], except for the macro-area South and islands, proving that the approach illustrated in this paper is the more appropriate practice for the study of income inequality in the presence of negative values[11].

[10] If the confidence intervals do not overlap, the null hypothesis that the two differences are equal is *a fortiori* rejected according to the test size of the complement at one of the confidence levels.

[11] We stress that preliminary analyses on the Δ's of the considered indices allow us to further reject the equality hypothesis of the differences for the ratio and in most cases for the simple differences. These results are available on request.

Fig. 10 Confidence intervals for the $G_Y - G_D$ and $G_Y^* - G_D^*$ bootstrap estimate reported in Table 8



4 Conclusions

In the current paper, the main features of the primary inequality measures, such as the Gini, concentration, and re-ranking coefficients, and the income decomposition approach, adjusted for income distributions that include even negative values, are discussed. The problem that arises when dealing with negative income values is associated with the violation of the normalisation principle. For this reason, in most of applied research in the economic and statistical fields, the more common procedure involves removing all negative values from the analysis. However, this method yields a relevant loss of information that may heavily affect the actual income inequality scenario. Thus, a more suitable methodology is needed. Based on a recent contribution for the Gini normalisation term, we propose an extension of this normalisation to all the Gini-based inequality measures characterising the income decomposition approach, in terms of both income sources and area components, with the aim of obtaining an unbiased description of the inequality situation. In addition to the new formalisation of the income decomposition approach from the theoretical point of view, an application on empirical data based on the Survey of Household Income and Wealth (SHIW) conducted by the Bank of Italy (2012) is illustrated. Specifically, we first considered and decomposed by source the total net income distribution. The total net income is given as the sum of two sources: the financial capital gain and the remaining income sources, expressed as the difference between the total net income and the financial capital gain. Second, we analysed the effects of income sources on the overall inequality within and gross-between the three main Italian geographical areas (North Italy, Central Italy and South Italy and Islands). The investigations extended to the within and between macro-region components help to fully assess the significance of differences arising in the results yielded by removing part of the

data information and those yielded by using the whole data information. Tests were conducted through simulation studies based on bootstrap resampling techniques and carried out both by considering Italy as a whole, and within and between the three Italian macro-areas. The differences in the distribution of the selected income items with regard to the three macro-areas allow us to improve the reliability of simulations. Inferential findings highlight how the two considered approaches (removing and preserving the negative incomes) lead to different results in terms of inequality estimation, proving that the proposed approach is the more appropriate practice to follow to avoid the loss of data that really provide a coherent picture of the inequality condition.

Acknowledgements We are thankful to Maria Giovanna Monti for her support and suggestions at an early stage of the research proposed in the paper.

References

1. Angrist, J., & Pischke, J.S. (2010). The credibility revolution in empirical economics: how better research design is taking the con out of econometrics, *Journal of Economic Perspectives*, 24(2), 3-30.
2. Banca d'Italia (2012). Survey of Household Income and Wealth (SHIW) of the Bank of Italy in 2012. `http://www.bancaditalia.it/statistiche/indcamp/bilfait/dismicro/annuale/ascii/ind12_ascii.zip`.
3. Bellú, L. G., & Liberati, P. (2006). Policy Impact on Inequality. Decomposition of Income Inequality by Income Sources, EASYPol for the Food and Agriculture Organization of the United Nations, FAO. `http://www.fao.org/docs/up/easypol/446/decomp_inequlty_by_source_053en.pdf`.
4. Burkhauser, R.V., Larrimore, J. & Simon, K. (2013). Measuring the Impact of Health Insurance on Levels and Trends in Inequality and how the affordable care act of 2010 could affect them, *Contemporary Economic Policy*, 3(4), 779-794.
5. Cavalcanti Ferreira, P. & Pereira Gomes, D.B. (2015). Heterogeneity of Initial Assets and Wealth Inequality, *Economics Working Papers*, n. 768, (Ensaios Economicos da EPGE) from FGV/EPGE Escola Brasileira de Economia e Finanças, Getulio Vargas Foundation (Brazil), available at `http://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/13886/Heterogeneity-of-Initial-Assets-and-WealthInequality.pdf;jsessionid=A378E0327C44851685C4E2E38624B853?sequence=3`.
6. Chen, C. N., Tsaur, T. W. & Rhai, T. S. (1982). The Gini coefficient and negative income. *Oxford Economic Papers*, 34(3), 473-478.
7. Berebbi, Z. M. & Silber, J. (1985). The Gini coefficient and negative income: a comment. *Oxford Economic Papers*, 37(3), 525-526
8. Dagum, C. (1959). Transvariazione fra più di due distribuzioni (in Italian). In: C. Gini. (Ed.), *Memorie di metodologia statistica*, II. Libreria Goliardica, Roma.
9. Dagum, C. (1961). Transvariacion en la hipotesis de varaibles aleatorias normales multidimensionales (in Spanish). *Proceedings of the International Statistical Institute*, 38(4), 473-486, Tokyo.
10. Dagum, C. (1997). A new approach to the decomposition of the Gini income inequality ratio. *Empirical Economics*, 22, 515-531.
11. Fei, J. C. H., Ranis, G., & Kuo, S. W. Y. (1978). Growth and the family distribution of income by factor components. *The Quarterly Journal of Economics*, 92(1), 17-53.
12. Fei, J. C. H., Ranis, G., & Kuo, S. W. Y. (1979). *Growth with Equity: The Taiwan Case*. Oxford University Press, London.
13. Feng, S., Burkhauser, R. V., & Butler, J. S.(2006). Levels and Long-Term Trends in Earnings Inequality: Overcoming Current Population Survey Censoring Problems Using the GB2 Distribution, *Journal of Business & Economic Statistics*, 24(1), 57-62.
14. Frick, J.R., Goebel, J., Schechtman, E., Wagner, G.G., & Yitzhaki, S (2004). Using Analysis of Gini (ANoGi) for Detecting Whether Two Sub-Samples Represent the Same Universe: The SOEP Experience, IZA DP No. 1049, Discussion Paper Series, 1-27, available at `http://repec.iza.org/dp1049.pdf`.

15. Gini, C. (1916). Il concetto di transvariazione e le sue prime applicazioni (in Italian). *Giornale degli Economisti e Rivista di Statistica*, 21-44.
16. Gornick, J.C. & Milanovic, B. (2015).Income Inequality in the United States in Cross-National Perspective: Redistribution Revisited, *LIS Center Research Brief*, n.1, available at `https://www.gc.cuny.edu/CUNY_GC/media/CUNY-GraduateCenter/PDF/Centers/LIS/LIS-Center-Research-Brief-1-2015.pdf`.
17. Kakwani, N., & Lambert, P. J. (1998). On measuring inequality in taxation: a new approach. *European Journal of Political Economy*, 14, 369-380.
18. Lambert, P. J., & Yitzhaki, S. (2013). The inconsistency between measurement and policy instruments in family income taxation. *FinanzArchiv: Public Finance Analysis*, 69(3), 241-255.
19. Malý, J. (2015). Impact of Polycentric Urban Systems on Intra-regional Disparities: A Micro-regional Approach. *European Planning Studies*, 1-23.
20. Mussard, S., Alperin, P. M. N., Seyte, F., & Terraza, M. (2006). Extensions of Dagum's Gini decomposition. *Statistica & Applicazioni*, IV, 5-29.
21. Podder, N. (1993). The Disaggregation of the Gini Coefficient by Factor Components and its Applications to Australia. *Review of Income and Wealth*, 39(1), 51-61.
22. Podder, N., & Chatterjee, S. (2002). Sharing the national cake in post reform New Zealand: income inequality trends in terms of income sources. *Journal of Public Economics*, 86(1), 1-27.
23. Pyatt, G., Chen, C., & Fei, J. (1980). The Distribution of Income by Factor Components. *The Quarterly Journal of Economics*, 95(3), 451-473.
24. Raffinetti, E., Siletti, E., & Vernizzi, A. (2015) On the Gini coefficient normalization when incomes with negative values are considered. *Statistical Methods & Applications*, 24(3), 507-521.
25. Rao, V. M. (1969). Two Decompositions of Concentration Ratio. *Journal of the Royal Statistical Society-Series A (General)*, 132(3), 418-425.
26. Sandovall, H. H., & Urzúa, C. M. (2009). Negative net incomes and the measurement of poverty: a note. *Journal of Management, Finance and Economics*, 3(1), 29-36.
27. Silber, J. (1989). Factor Components, Population Subgroups and the Computation of the Gini Index of Inequality. *The Review of Economics and Statistics*, 71(1), 107-115.
28. Schutz, R. R. (1951). On the measurement of the income inequality. *The American Economic Review*, 41(1), 107-122.
29. Van De Ven, J., & Creedy, J. (2005). Taxation, reranking and equivalence scales. *Bulletin of Economic Research*, 57(1), 13-36.
30. Yitzhaki, S. (2003). Gini's mean difference: a superior measure of variability for non-normal distributions, *Metron*, 61(2), 285-316.
31. Yitzhaki, S. (2015). Gini's mean difference offers a response to Leamer's critique, *Metron*, 73(1), 31-43.