Inequality measures and the issue of negative incomes

Emanuela Raffinetti, Elena Siletti, Achille Vernizzi

Abstract Income distribution studies have a long history in economic and statistical literature. Many results in such research area are provided by the standard inequality measure Gini coefficient, traditionally defined for non-negative incomes. In this paper the issue of negative incomes is faced and a specific reformulation of the Gini coefficient is introduced. More precisely, a new Gini coefficient normalization, held by the Pigou-Dalton transfers principle fulfillment, is presented.

Key words: Gini coefficient, negative incomes, Pigou-Dalton transfers principle, normalization term.

1 Introduction

The measurement and assessment of income distribution represents an active research area which achieves a great interest in a wide set of scientific disciplines, such as economics and statistics. In such a context, the Gini coefficient appears as the common and most popular measure of inequality of income or wealth, as supported by its several developments and applications in literature. However, such applications are mainly restricted to the case of incomes with non-negative values, in order to fulfill the classical Gini coefficient formalization. For this reason, as stated by [4], removing the negative values from the analysis is basic as otherwise the need of resorting to a more complex methodology arises. Omission of negative incomes typically represents an usual procedure which finds a wide validation in many research papers (see e.g. [5, 7]). The main troubles associated to the treatment of negative incomes regards the violation of the normalization principle. In fact, the inclusion of incomes taking negative values implies that the standard Gini coefficient formula can achieve values greater than 1. To overcome this disadvantage, the Gini coefficient has to be adjusted in order to assure that its range is bounded between 0 and +1. [1] attempted to reformulate and normalize the Gini coefficient to make comparability between the distributions without negative incomes and the distributions with some negative incomes be attained. The work presented in [1] was subsequently finished off by [2], who provided a correct expression for the Gini

Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano, Via Conservatorio 7, Milano, Italy, e-mail: emanuela.raffinetti@unimi.it, elena.siletti@unimi.it, achille.vernizzi@unimi.it

coefficient normalization term. In this paper the issue of negative incomes is further on stressed and a new reformulation of the Gini coefficient suitable for such purpose proposed. We believe that even if negative incomes can appear as an unfamiliar concept, it is worth noting the ways in which they can arise. Typically, in real surveys beside many positive incomes one can observe also negative ones. It happens when assessing families financial assets such as, for instance, capital gains.

The paper is structured as follows: Section 2 is addressed to an overview of the existing contributions dealing with the extension of inequality measures when negative incomes appear and Section 3 focuses on a reformulation of the Gini coefficient shedding the light on a new Gini coefficient normalization proposal.

2 Background

A first attempt in providing an appropriate normalization term for the Gini coefficient when negative incomes are involved was given by [1]. Let N be the number of considered income units and Y_i the income of the $i-th$ unit ordered in a non-decreasing sense. Let us denote with y_i the income share of the $i-th$ unit, that is $y_i = \frac{Y_i}{N\mu_Y}$, with μ_Y corresponding to the average of Y. By resorting to mean difference, the normalized Gini coefficient G_{CTR}[1] can be expressed as

$$G_{CTR} = \frac{1+\left(\frac{2}{N}\right)\sum_{i=1}^{k} iy_i - \left(\frac{1}{N}\right)\sum_{i=k+1}^{N} y_i(1+2(N-i))}{1+\left(\frac{2}{N}\right)\sum_{i=1}^{k} iy_i}, \tag{1}$$

with k defined in such a way: $\sum_{i=1}^{k} y_i = 0$. It is worth noting that the previous condition is rather uncommon since it results very unlike that in a sequence of incomes the first k values provide a null sum. Actually, [1] consider also the more general scenario when $\sum_{i=1}^{k} y_i < 0$ and $\sum_{i=k+1}^{N} y_i > 0$ even if a correct formulation of the corresponding Gini coefficient was finally accomplished by [2]. According to [2] the normalized Gini coefficient G_{BS}[2] in (1) for the more general case can be written as

$$G_{BS} = \frac{\frac{2}{N}\sum_{i=1}^{N} iy_i - \sum_{i=1}^{N} y_i \frac{N+1}{N}}{1+\left(\frac{2}{N}\right)\sum_{i=1}^{k} iy_i + \frac{1}{N}\sum_{i=1}^{k} y_i \left[\frac{\sum_{i=1}^{k} y_i}{y_{k+1}} - (1+2k)\right]}, \tag{2}$$

where y_{k+1} represents the first income share such that the $\sum_{i=1}^{k+1} y_i > 0$. Even if [1], and subsequently [2], provided the Gini coefficient normalization resorting to the mean difference-based formula, at the same time they built the normalization term by taking into account the geometrical construction made by linking the Gini coefficient with the concentration area.

For a discussion about the properties of the Gini coefficient adjusted for negative incomes see [6].

---

[1] CTR is the acronym of Chen, Tsaur and Rhai.

[2] BS is the acronym of Berebbi and Silber.

3 Our contribution: extension and normalization of the Gini coefficient

The purpose of this section is threefold. First, in Subsection 3.1 we aim at extending the Gini coefficient computation when beside negative incomes also weights are taken into account. Furthermore, in Subsection 3.2 we illustrate some empirical examples shedding the light on some abnormal behaviors of the Gini normalization proposed by Berebbi and Silber. Finally, Subsection 3.3 deals with our proposal based on providing a new Gini coefficient normalization adjusted for the presence of negative values.

3.1 The Gini coefficient extension for weighted data

The classical Gini coefficient of an attribute Y with non-negative values can be translated into the below mean difference-based expression:

$$G = \frac{1}{2\mu_Y N^2} \sum_{i=1}^{H} \sum_{j=1}^{H} |Y_i - Y_j| p_i p_j, \tag{3}$$

where H is the total number of considered income units, p_i and p_j are weights associated to Y_i and Y_j such that $\sum_{i=1}^{H} p_i = N$ and μ_Y is the Y average value[3]. Typically, weights are introduced (and provided, for instance, by the Bank of Italy) to bring to the entire world or to use the equivalence scales (obtaining equivalent incomes) with the aim of making incomes own by income units with different size comparable. By extending (3) to the case of also negative income values, the Berebbi and Silber Gini coefficient (hereafter denoted by G^*_{BS}) becomes

$$G^*_{BS} = \frac{1}{2\mu^*_Y N^2} \sum_{i=1}^{H} \sum_{j=1}^{H} |Y_i - Y_j| p_i p_j, \tag{4}$$

where the normalization term μ^*_Y results as

$$\mu^*_Y = \mu_Y + \frac{1}{2N^2} \sum_{i=1}^{k} \sum_{j=1}^{k} |Y_i - Y_j| p_i p_j + \frac{1}{N^2} \sum_{i=1}^{k} |Y_i - Y_{k+1}| p_i p^*_{k+1} \tag{5}$$

with $p^*_{k+1} = p_{k+1} \frac{|\sum_{j=1}^{k} Y_j p_j|}{Y_{k+1} p_{k+1}}$. Expression in (5) is an extension of the Berebbi and Silber normalization when also non-integer weights are taken into account.

Moreover, (4) can be developed as in (6). Indeed, by dividing (4) by μ_Y and denoting by $y_i = \frac{Y_i p_i}{\sum_{i=1}^{H} Y_i p_i}$ and $N_k = \sum_{i=1}^{k} p_i$, (4) can be expressed as

[3] Note that weights p_i and N are non-integer.

$$G^*_{BS} = \frac{\frac{2}{N}\sum_{i=1}^{H}\sum_{j=1}^{i} y_i p_j - \sum_{i=1}^{H} y_i \frac{N+p_i}{N}}{1+\frac{2}{N}\sum_{i=1}^{k}\sum_{j=1}^{i} y_i p_j + \frac{1}{N}\sum_{i=1}^{k} y_i \left[\frac{p_{k+1}}{y_{k+1}}\sum_{i=1}^{k} y_i - 2N_k\right] - \frac{1}{N}\sum_{i=1}^{k} y_i p_i}, \quad (6)$$

which coincides with (2) in case $p_i = p_j = 1, \forall i, j = 1, \ldots, H$. For the sake of brevity, the proof of such equivalence is not reported here but it is available on request.

3.2 Some abnormal behaviors of G^*_{BS}

The Gini coefficient normalization firstly introduced by [1] and subsequently completed by [2] presents some abnormal behaviors in detecting the existing inequality between income distributions. Let us consider the three different income scenarios, regarding ten income units, reported in Table 1). For the sake of simplicity, let us suppose that $p_i = p_j = 1, \forall i = j$.

Scenarios	**Income vector** Y									
Scenario (a):	-5	-5	-5	-5	-5	-5	-5	-5	-5	45.01
Scenario (b):	-45	0	0	0	0	0	0	0	0	45.01
Scenario (c):	-15	-10	-8	-7	-5	0	0	0	0	45.01

Table 1 Income distribution scenarios

Case (a) describes an income distribution characterized by almost all negative incomes except for the last one which is positive, case (b) is representative of a distribution where only two income units own an income which takes on one hand a negative value and on the other hand a positive value and finally in case (c) some income units have negative incomes, some others null incomes and only one has a positive income. It would be then rational expecting that the Gini coefficient computed in the three scenarios varies in order to take into account income inequalities. This does not occur, in fact as highlighted in Table 2, the normalized Gini coefficient G^*_{BS} is similar in all the considered situations and close to one, missing in measuring the existing income inequalities.

	Scenario (a)	**Scenario (b)**	**Scenario (c)**
G^*_{BS}	0.999999995	0.999999996	0.999999996

Table 2 G^*_{BS} results in scenarios (a), (b) and (c)

Such a result validates our proposal in reconsidering a new normalization for the Gini coefficient when negative incomes are involved. The main features of our proposed Gini coefficient normalization are discussed in the following subsection.

3.3 The reconsidered Gini coefficient normalization for negative incomes

The purpose of this subsection is introducing a new proposal for the Gini coefficient normalization when income distribution involves also negative values. More precisely, due to the drawbacks related to the contribution presented by [2], a more proper normalization is provided.

Let us take into account the specific scenario where, given H income units, the total positive (T^+) and negative $(-T^-)$ incomes are assigned only to two single units and all the others have no income, i.e. $Y = \{-T^-, 0, 0, 0, \ldots, 0, 0, 0, T^+\}$. Such a context corresponds to scenario illustrated in (b). Furthermore, let $p_1 = p_H = 1$, while in all the other $H-2$ cases p_i are non-integer and such as $\sum_{i=2}^{H-1} p_i = N - 2$. The income inequality maximization, here denoted by Δ_{max}, is then computed according to the mean difference formula as follows:

$$\Delta_{max} = \frac{1}{N^2}\sum_{i=1}^{H}\sum_{i=1}^{H}|Y_i - Y_j|p_i p_j = 2\frac{(N-1)(T^+ + T^-)}{N^2} = 2\mu_Y^{RSV}, \tag{7}$$

where T^- points out the absolute value of negative incomes and μ_Y^{RSV}[4] corresponds to $\frac{(N-1)(T^+ + T^-)}{N^2}$.

The same result can be reached by linking the Gini coefficient with the concentration area A obtained as

$$\begin{aligned} A =& \frac{1}{2} - \frac{1}{2N}\frac{T^+ - T^-}{T^+} + \frac{1}{2N(T^+ - T^-)}T^- + \frac{N-2}{N(T^+ - T^-)}T^- \\ &+ \frac{1}{2N(T^+ - T^-)}\frac{T^-}{T^+}T^-. \end{aligned} \tag{8}$$

One can prove that (7) and (8) coincide when (8) is multiplied by $(4/N)(T^+ - T^-)$. A comparison between the concentration area computed by [1] and [2] and that provided in (8) shows that the Gini coefficient normalization given by [2] does not take into account the term $-\frac{1}{2N(T^+ - T^-)}\frac{T^+ - T^-}{T^+}$ which for $N \to \infty$, becomes close to zero. However, this does not happen for a finite number N of observations.

It is worth noting, by fulfilling the following three conditions,

1. the income average value $\mu_Y = (T^+ - T^-)/N$ has to be preserved in any income redistribution process;
2. T^+ is the maximum positive value that can not be exceeded in any income redistribution process;
3. T^- is the minimum negative value that can not be exceeded in any income redistribution process,

[4] RSV is the acronym of Raffinetti, Siletti and Vernizzi.

that any other income redistribution, according to the "Pigou-Dalton transfers principle[5]" (see [3]) and based on income transfers among $N > 2$ units, provides a mean difference smaller than $2[(N-1)/N^2](T^+ + T^-)$. Therefore, the Gini coefficient can be normalized by the term $2\mu_Y^{RSV} = 2\frac{(N-1)(T^+ + T^-)}{N^2}$, or for $N \to \infty$, by $2\frac{(T^+ + T^-)}{N}$[6]. Alternatively one can resort also to the normalization term provided in (8), which however asymptotically holds[7].

To validate the new introduced Gini coefficient normalization, let us reconsider the three different income distribution examples presented in Subsection 3.2. Results are displayed in Table 3.

	Scenario (a)	**Scenario (b)**	**Scenario (c)**
G_{RSV}	0.555604932	1	0.834586280

Table 3 G_{RSV} results in scenarios (a), (b) and (c)

Findings in Table 3 show the attitude of the proposed normalized Gini coefficient G_{RSV}[8] in detecting the real existing inequalities among the different scenarios. We remark that the Berebbi and Silber normalization in (2) holds for strictly positive income average, whereas our contributed normalization holds also in case of negative income average.

References

1. Chen, C. N., Tsaur, T. W. and Rhai, T. S.: The Gini coefficient and negative income. *Oxford Economic Papers*, **34**, 473–478 (1982)
2. Berrebi, Z. M., Silber, J.: The Gini coefficient and negative income: a comment, *Oxford Economic Papers*, **37**, 525–526 (1985)
3. Dalton, H.: The Measurement of the Inequality of Income. *Economic Journal*, **30**, 348–361 (1920)
4. Greselin, F., Pasquazzi, L., Zitikis, R.: Zenga's New Index of Economic Inequality, Its Estimation, and an Analysis of Incomes in Italy. *Journal of Probability and Statistics*, 718905 (2010)
5. Mussini, M.: A matrix approach to the Gini index decomposition by subgroup and by income source. *Applied Economics*, **45**, 2457–2468 (2013)
6. Stich, A.: Inequality and negative income. *Statistical Methods & Applications*, **5**, 297–305 (1996)
7. Zenga, M. M., Radaelli, P., Zenga, M.: Decomposition of Zenga's inequality index by sources. *Statistica & Applicazioni*, **X**, 3–31 (2012)

[5] The "Pigou-Dalton transfers principle" requires inequality measures decreasing as a consequence of any progressive transfer from a richer to a poorer person, preserving rank-order of incomes.

[6] We remark that when $N \to \infty$, the simplifying assumption $p_1 = p_H = 1$ can be released.

[7] If on one hand, as stated by [1], the normalized Gini coefficient in (4) reaches value +1 if $N \to \infty$, on the other hand our normalized Gini coefficient achieves value +1 also when N is very small, as shown in case (b) of Table 3.

[8] $G_{RSV} = \frac{1}{2\mu_{RSV} N^2} \sum_{i=1}^{H} \sum_{j=1}^{H} |Y_i - Y_j| p_i p_j$.