Received: 10 December 2019 | Accepted: 28 March 2021

DOI: 10.1111/rssa.12701

ORIGINAL ARTICLE

Linearization and variance estimation of the Bonferroni inequality index

Ziqing Dong[1] | **Yves Tillé**[1] | **Giovanni M. Giorgi**[2] | **Alessio Guandalini**[3]

[1]Institut de Statistique, Université de Neuchâtel, Bellevaux 51, Neuchâtel, 2000, Switzerland

[2]Dipartimento Scienze Statistiche, Università "La Sapienza" di Roma, Piazzale Aldo Moro, 5, Roma, 00185, Italy

[3]Istituto Nazionale di Statistica (ISTAT), Via Cesare Balbo, 16, Roma, 00184, Italy

Correspondence
Ziqing Dong, Institut de Statistique, Université de Neuchâtel. Bellevaux 51, 2000 Neuchâtel
Email: ziqing.dong@unine.ch

Abstract

The study of income inequality is important for predicting the wealth of a country. There is an increasing number of publications where the authors call for the use of several indices simultaneously to better account for the wealth distribution. Due to the fact that income data are usually collected through sample surveys, the sampling properties of income inequality measures should not be overlooked. The most widely used inequality measure is the Gini index, and its inferential aspects have been deeply investigated. An alternative inequality index could be the Bonferroni inequality index, although less attention on its inference has been paid in the literature. The aim of this paper is to address the inference of the Bonferroni index in a finite population framework. The Bonferroni index is linearized by differentiation with respect to the sample indicators which allows for conducting a valid inference. Furthermore, the linearized variables are used to evaluate the effects of the different observations on the Bonferroni and Gini indices. The result demonstrates once for all that the former is more sensitive to the lowest incomes in the distribution than the latter.

KEYWORDS

Bonferroni, Gini, inequality measures, inference, influence function

1 | INTRODUCTION

Nobel Prize-winning economist, Joseph Stiglitz, stated that income inequality is an important measure for forecasting the wealth of a country (Stiglitz, 2012). The most widely used inequality measure is the Gini index (Gini, 1912,1914). Since Corrado Gini suggested the index, it has been the subject of numerous publications. Its use is not restricted only to the economic field. It is surprising to note that even after a century, different applications of the Gini index pop up in new fields (see, e.g. Giorgi, 2019).

Recent studies encourage the use of more than one inequality index simultaneously to better catch the inequality in different parts of the income distribution and thereby to better understand the socio-economic reality and political significance of inequality (see, e.g. Osberg, 2017; Piketty, 2015). The most suitable candidate to place side by side with the Gini index could be the Bonferroni inequality index (Bonferroni, 1933). In fact, Pundir et al. (2005) show that both can be derived from the Lorenz Curve (Lorenz, 1905). Indeed, the two indices share several properties while maintaining some very interesting peculiarities.

The opposition between the Bonferroni index and the Gini index is rooted when Carlo Emilio Bonferroni proposed his index in 1930. In the beginning, the Bonferroni index was fought by Corrado Gini and his followers who were very fond of the Gini index and who tried to avoid the use of any other measures that took the Gini index down the line (Giorgi, 1998). Only in the last 40 years, the Bonferroni index has been rediscovered by Piesch (1975) and Nygård and Sandström (1981). Several extensions and interpretations proposed for the Gini index (see, e.g. Giorgi, 2005, for a comprehensive review) have then been extended to the Bonferroni index, disclosing even more similarities and differences between the two indices. Among them, welfare implication (see, e.g. Aaberge, 2000; Bárcena-Martín & Silber, 2013, 2017; Benedetti, 1986; Chakravarty, 2007; Chakravarty & Muliere, 2004), socio-economic aspects (Bárcena & Imedio, 2008; Bárcena-Martín & Silber, 2011; Bárcena-Martín & Silber, 2013; Imedio-Olmedo et al., 2012; Silber & Son, 2010), application in fuzzy and reliability frameworks (Giordani & Giorgi, 2010; Giorgi & Crescenzi, 2001b) and decomposition by sources or by groups (Bárcena-Martín & Silber, 2013; Giorgi & Guandalini, 2018; Tarsitano, 1990) have been studied.

Because income data are usually collected through sample surveys, the sampling properties of the Bonferroni and Gini indices should not be overlooked. Inference on the Gini index is a tricky problem which has generated a large number of publications (see, e.g. Giorgi & Gigliarano, 2017; Graf & Tillé, 2014; Langel & Tillé, 2013, and reference therein). However, mainly for the reasons stated previously, less attention has been paid on the inference of the Bonferroni index. Giorgi and Mondani (1994, 1995) derive the sampling distribution of the Bonferroni index from exponential population, while Giorgi and Crescenzi (2001a) propose Bayes estimators of it from a Pareto-type I population. Furthermore, Giorgi and Nadarajah (2010) explicate the Bonferroni index for several distributions. Aaberge (2007) studies the Bonferroni index as an inequality measure belonging to the Gini's nuclear family, and Pundir et al. (2005) provide asymptotically distribution-free statistical inference for it. Finally, Giorgi and Guandalini (2013) study a sampling estimator less biased for small samples than the plug-in estimator.

In the present paper, the linearization technique developed by Graf (2011) has been applied for estimating the variances of the Bonferroni index estimators. Moreover, the obtained linearized variables are interpreted as influence function in the sense of Hampel (1974) and Hampel et al. (1985). The comparison with the linearized variables of the Gini index helps to evaluate the effects of the different observations on the Bonferroni and Gini indices. The relation between the Bonferroni index and the Gini index stated several times in the literature (De Vergottini, 1950; Pizzetti, 1951) is confirmed, and it is demonstrated once for all that the former is more sensitive to the lowest incomes in the distribution than the latter.

The paper is organized as follows: In Section 2, the Bonferroni index is defined in an infinite population. Section 3 provides the estimation of the Bonferroni index in a finite population framework. In Section 4, the linearization method developed by Graf is applied to the Bonferroni index estimators. In Section 5, the procedure for deriving the sampling variances of the Bonferroni index estimators is described. The accuracy of the Bonferroni index estimators and their associated variance estimators is tested in Section 6. In Section 7, the influence function of the Bonferroni index is derived and analysed along with the influence function of the Gini index. The relation between the influence function and the proposed linearized variables is discussed, and through an empirical demonstration, it is shown that the Bonferroni index is more sensitive to the lowest incomes in the distribution than the Gini index. Finally, Section 8 contains some concluding remarks.

2 | THE BONFERRONI INDEX

Consider a non-negative continuous random variable Y with probability density function $f(y)$ and finite mean $\mu = \int_0^\infty yf(y)\,dy \neq 0$.

Let Y denote income (or, in some cases, wage, turnover, profit or consumption expenditure). Its cumulative distribution function is given by

$$F(y) = \int_0^y f(t)dt.$$

Assume also that $F(y)$ is absolutely continuous and at least twice differentiable. Such F will be interpreted as income distribution and $p = F(y)$ as proportion of incomes less than or equal to y, $p \in [0,1]$. Let

$$\mu(y) = \frac{\int_0^y tdF(t)}{F(y)}$$

be the partial mean and $Q(p) = F^{-1}(p) = \inf_y \{y\colon F(y) \geq p\}$ be the quantile function. The Lorenz curve (1905) can be expressed as

$$L(p) = \frac{1}{\mu}\int_0^p Q(\alpha)\,d\alpha$$

(see, e.g. Gastwirth, 1971; Pietra, 1915).

Among the several ways of defining the Gini index (Xu, 2003; Yitzhaki, 1998), one can write it as function of $L(p)$:

$$G = 1 - 2\int_0^1 L(p)dp. \tag{1}$$

Indeed, G is equal to the area between the 45-degree diagonal segment (line of perfect equality) and the Lorenz curve divided by the whole area under the diagonal.

The Bonferroni index can also be written as function of the Lorenz curve (Pundir et al., 2005):

$$B = 1 - \int_0^1 B(p)dp, \tag{2}$$

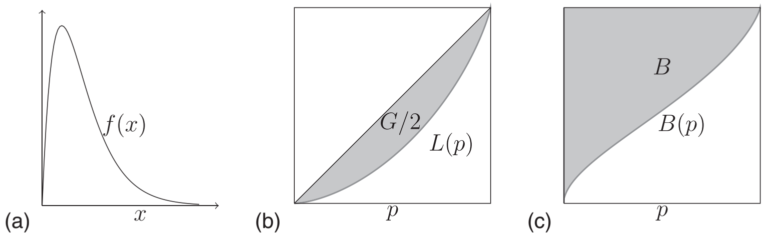

FIGURE 1 (a) Probability density function of a gamma random variable with shape $\alpha = 2$ and rate $\beta = 10$; (b) Lorenz curve, the hatched portion is equal to half of the Gini index; (c) Bonferroni curve, the hatched portion represents the Bonferroni index

where $B(p) = L(p)/p$ is the ordinate of the Bonferroni curve (Figure 1). The Bonferroni curve, $[p, L(p)/p]$, is defined on the orthogonal plane within a unit square. It does not always start from the origin of the orthogonal plane because when p goes to 0, $B(p)$ takes the form 0/0. Furthermore, it is strictly increasing and it can be convex in some parts and concave in others (Giorgi & Crescenzi, 2001b).

For B, the line of perfect equality is the line which joins the coordinate points (0,1) and (1,1). The Bonferroni index is equal to the area enclosed by the axis of ordinates, the line of perfect equality and the Bonferroni curve. That is, the concentration area coincides with B.

Except for extreme cases, maximum or null concentration, $B > G$ (De Vergottini, 1950). The difference between the two inequality indices becomes more clear by writing (1) and (2) in terms of the relative difference between the total mean and the partial mean, $r(y) = (\mu - \mu(y))/\mu$. That is,

$$G = \int_0^\infty r(y) \left[\frac{F(y)}{\int_0^\infty F(y)\ dF(y)} \right] dF(y)$$

and

$$B = \int_0^\infty r(y)\ dF(y).$$

Hence, G is a weighted mean of the $r(y)$'s, while B is their simple mean (Tarsitano, 1990, p. 230). The difference, $B-G$, increases in the first stretch up to a threshold where the weights in G are large enough. It then decreases until G and B achieve their values (Pizzetti, 1951, p. 302).

3 | DEFINITION AND ESTIMATION IN A FINITE POPULATION

The inequality measures are usually estimated by means of a sample survey. Consider a population U of size N, which could be individuals, households or enterprises. Let $y_1, \ldots, y_i, \ldots, y_N$ denote the incomes of the N population units sorted in ascending order. The total, the mean and the partial mean in the finite population are respectively defined by

$$Y = \sum_{i \in U} y_i, \ \overline{Y} = \frac{Y}{N}, \ \overline{Y}_k = \frac{1}{N_k} \sum_{i \in U} y_i \mathbb{1}[y_i \leq y_k],$$

where

$$N_k = \sum_{i \in U} \mathbb{1}[y_i \leq y_k].$$

The Bonferroni inequality index in the finite population according to its original definition (Bonferroni, 1933, p. 58) is

$$B = \frac{1}{(N-1)\,\overline{\overline{Y}}} \sum_{k \in U} \left(\overline{Y} - \overline{Y}_k \right). \tag{3}$$

As explained in Giorgi (1998), Expression (3) is the discrete analogue of Expression (2) defined in Section 2.

The finite-population definition of the Gini index, written in terms of mean difference (Gini, 1914), is

$$G = \frac{\sum_{i \in U} \sum_{j \in U} |y_i - y_j|}{2NY}. \tag{4}$$

Consider a random sample of size n selected from U according to a sampling design $p(s) = \Pr(S = s)$, for all $s \in S$. Let $a_1, \ldots, a_i, \ldots, a_N$ be the indicator Bernoulli random variables for the presence of the units in the random sample. Let $\pi_i = \mathrm{E}[a_i]$ denote the first-order inclusion probability of unit i in U and its inverse, $d_i = 1/\pi_i$, the Horvitz and Thompson (1952) weight. Let $\pi_{ij} = \mathrm{E}[a_i a_j]$ be the second-order inclusion probability.

A sampling weight w_i is associated with each sampling unit. The w_i could be equal to the inverse of the first-order inclusion probability $d_i = 1/\pi_i$ used in the Horvitz–Thompson estimator. The weights could also be the result of a more complex estimation procedure. For instance, the weights could be obtained through a calibration on known marginal totals (Deville & Särndal, 1992; Särndal, 2007). The weights could also contain a reweighting factor to compensate questionnaire nonresponse (Särndal & Lundström, 2005).

The estimators of the population size, the total, the mean and the partial mean are respectively defined by

$$\widehat{N} = \sum_{i \in U} w_i a_i, \; \widehat{Y} = \sum_{i \in U} w_i y_i a_i, \; \widehat{\overline{Y}} = \frac{\widehat{Y}}{\widehat{N}}, \; \widehat{\overline{Y}}_k = \frac{1}{\widehat{N}_k} \sum_{i \in U} w_i y_i a_i \mathbb{1}[y_i \leq y_k],$$

where

$$\widehat{N}_k = \sum_{i \in U} w_i a_i \mathbb{1}[y_i \leq y_k].$$

For the Bonferroni index, the plug-in estimator of (3) is

$$\widehat{B}_r = \frac{1}{\left(\widehat{N} - 1\right) \widehat{\overline{Y}}} \sum_{k \in U} a_k \left[w_k \left(\widehat{\overline{Y}} - \widehat{\overline{Y}}_k \right) \right], \tag{5}$$

while an alternative estimator (Giorgi & Guandalini, 2013, p. 154) is

$$\widehat{B}_t = \frac{1}{\left(\widehat{N}-1\right)\widehat{Y}} \sum_{k\in U} a_k \left\{ w_k \left(\widehat{Y} - \frac{\widehat{Y}_k + \widehat{Y}_{k-1}}{2} \right) \right\}. \tag{6}$$

Since both estimators are non-linear functions of Horvitz–Thompson estimators, they are slightly biased. The former decomposes the concentration area in rectangles (for this reason, the notation $\widehat{B}_r$ is used, where r refers to *rectangles*), while the latter uses trapezoids (for this reason, the notation $\widehat{B}_t$ is used, where t refers to *trapezoids*) in order to reduce the bias of $\widehat{B}_r$ for samples of small sizes in particular (see simulation results in Section 6).

For the Gini index, the plug-in estimator of (4) is

$$\widehat{G} = \frac{\sum_{i\in U}\sum_{j\in U} a_i a_j w_i w_j \,|\, y_i - y_j \,|}{2\widehat{N}\widehat{Y}} \tag{7}$$

(Langel & Tillé, 2013, p. 524).

4 | LINEARIZATION

In order to address the problem of inference for the Bonferroni index, a linearization method is proposed. Linearization includes a range of techniques used to approximate the variance of a non-linear statistic. With the linearization techniques, a non-linear or a complex statistic, such as B, is approximated by a sum of terms. The interest lies basically in finding linearized variables $z_{\bullet i}$'s such that

$$\widehat{B}_{\bullet} - B \approx \sum_{i\in S} w_i z_{\bullet i} - \sum_{i\in U} z_{\bullet i},$$

where $\widehat{B}_{\bullet}$ stands for $\widehat{B}_r$ or $\widehat{B}_t$. The variance of $\widehat{B}_{\bullet}$ can thereby be simply approximated by the variance of the related total estimator

$$\widehat{Z} = \sum_{i\in S} w_i z_{\bullet i}. \tag{8}$$

In general, linearized variables $z_{\bullet i}$'s must be estimated because they depend on population parameters. They can be estimated from the sample and can be used to construct an estimator of the variance by plugging $\widehat{z}_{\bullet i}$'s into the variance estimator of the total estimator under the given sampling design.

There are several ways of deriving linearized variables. For smooth functions of the totals, it is possible to linearize by performing a Taylor series expansion with respect to these totals (Woodruff, 1971). However, for parameters that are not functions of the totals, alternative methods must be used. Most of these parameters can be seen as the solution of an estimating equation. Following the estimating equations methodology developed in Binder (1983, 1991) and Binder and Patak (1994), linearized variables can be derived for estimating the sampling variance.

Deville (1999) proposes the use of influence function, already known in the field of robust statistics (see, e.g. Hampel, 1974; Hampel et al., 1985), as artificial variables for estimating the variance of complex estimators (i.e. calibration type estimators) and non-linear statistics. Furthermore, Demnati and Rao (2004) propose to use the Deville influence function on the estimated measure of mass equal to w_i.

In the present paper, in order to estimate the variances of the estimators presented in the previous section, the linearization method developed by Graf (2011) is used. This method consists of computing the derivatives of the estimator with respect to the a_i indicator variables of the presence of the units in the sample (see also Vallée & Tillé, 2019). It can be applied to almost all sampling designs as long as the expression of the variance estimator of the total estimator under the sampling design is known.

Result 1 The linearized variable of $\widehat{B}_r$ is

$$\begin{aligned}\widehat{z}_{ri} := \left(\frac{\partial \widehat{B}_r}{\partial a_i}\right)/w_i = \frac{1}{\left(\widehat{N}-1\right)\widehat{\widehat{Y}}} \Bigg\{ & \frac{1}{\widehat{N}}\left(y_i - \widehat{\widehat{Y}}\right)\widehat{B}_r - y_i\widehat{B}_r + \left(y_i - \widehat{\widehat{Y}}_i\right) \\ & - y_i \sum_{k\in U} \frac{w_k \mathbb{1}[y_i \le y_k] a_k}{\widehat{N}_k} + \sum_{k\in U} \frac{w_k \mathbb{1}[y_i \le y_k]\widehat{\widehat{Y}}_k a_k}{\widehat{N}_k} \Bigg\}.\end{aligned} \tag{9}$$

Result 2 The linearized variable of $\widehat{B}_t$ is

$$\begin{aligned}\widehat{z}_{ti} := \left(\frac{\partial \widehat{B}_t}{\partial a_i}\right)/w_i = \frac{1}{\left(\widehat{N}-1\right)\widehat{\widehat{Y}}} \Bigg\{ & \frac{1}{\widehat{N}}\left(y_i - \widehat{\widehat{Y}}\right)\widehat{B}_t - y_i\widehat{B}_t + \left(y_i - \frac{\widehat{\widehat{Y}}_i + \widehat{\widehat{Y}}_{i-1}}{2}\right) \\ & - y_i \sum_{k\in U} \frac{w_k a_k}{2}\left(\frac{\mathbb{1}[y_i \le y_k]}{\widehat{N}_k} + \frac{\mathbb{1}[y_i \le y_{k-1}]}{\widehat{N}_{k-1}}\right) \\ & + \sum_{k\in U} \frac{w_k a_k}{2}\left(\frac{\mathbb{1}[y_i \le y_k]\widehat{\widehat{Y}}_k}{\widehat{N}_k} + \frac{\mathbb{1}[y_i \le y_{k-1}]\widehat{\widehat{Y}}_{k-1}}{\widehat{N}_{k-1}}\right) \Bigg\}.\end{aligned} \tag{10}$$

The proofs of Result 1 and Result 2 are given in Appendix.

The linearized variable of $\widehat{G}$ can be found in Langel and Tillé (2013, see also references therein):

$$\widehat{z}_{Gi} := \left(\frac{\partial \widehat{G}}{\partial a_i}\right)/w_i = \frac{1}{\widehat{N}\widehat{Y}}\left\{2\widehat{N}_i\left(y_i - \widehat{\widehat{Y}}_i\right) + \widehat{Y} - \widehat{N}y_i - \widehat{G}\left(\widehat{Y} + y_i\widehat{N}\right)\right\}. \tag{11}$$

5 | VARIANCE ESTIMATION

5.1 | Variance estimation formulas

Results 1 and 2 presented in Section 4 can be used to approximate the variances of $\widehat{B}_r$ and $\widehat{B}_t$. In fact, the proposed linearization method can be easily implemented as long as the expression of the variance estimator of the total estimator under the given sampling design is known.

Consider a general sampling design whose first-order (π_i) and second-order (π_{ij}) inclusion probabilities are all positive, a generalized Horvitz–Thompson formulation for the variance estimator of the total estimator is

$$\widehat{var}(\widehat{Y}) = \sum_{i\in S}\sum_{j\in S} \frac{y_i}{\pi_i}\frac{y_j}{\pi_j}\frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}}. \tag{12}$$

In the set-up when there is no weight adjustment for non-response or calibration, the variance of $\widehat{B}_r$ (resp. $\widehat{B}_t$) can be estimated by simply replacing y_i with $\widehat{z}_{ri}$ (resp. $\widehat{z}_{ti}$) in Expression (12) as a general formulation or in the following Expressions (13), (14), (15) or more other expressions of the variance estimator of the total estimator for each specific sampling design.

Under simple random sampling without replacement with fixed sample size (*SRSWOR*), Expression (12) can be written as:

$$^{SRSWOR}\widehat{var}(\widehat{Y}) = N^2 \frac{N-n}{Nn(n-1)} \sum_{i \in S} \left(y_i - \widehat{\overline{Y}}\right)^2. \tag{13}$$

Notice that for estimating the variance of $\widehat{B}_r$ (resp. $\widehat{B}_t$) under *SRSWOR*, $\widehat{\overline{Y}}$ is calculated as the sample mean of the linearized variables $\widehat{z}_{ri}$ (resp. $\widehat{z}_{ti}$).

Under stratified simple random sampling (*StrSRS*), the population is stratified and *SRSWOR* is used within each stratum. Because the samples are independent from stratum to stratum, the variance estimator of the total estimator is obtained by summing up the strata variances estimated using Expression (13) within each stratum:

$$^{StrSRS}\widehat{var}(\widehat{Y}) = \sum_{h=1}^{H} N_h^2 \frac{N_h - n_h}{N_h n_h (n_h - 1)} \sum_{i \in S_h} \left(y_i - \widehat{\overline{Y}}_h\right)^2, \tag{14}$$

where h is the label for each stratum ($h = 1, \ldots, H$), and $\widehat{\overline{Y}}_h$ is the sample mean calculated within stratum h:

$$\widehat{\overline{Y}}_h = \frac{1}{n_h} \sum_{i \in S_h} y_i.$$

For estimating the variance of $\widehat{B}_r$ (resp. $\widehat{B}_t$) under *StrSRS*, $\widehat{\overline{Y}}_h$ is calculated as the sample mean of the linearized variables $\widehat{z}_{ri}$ (resp. $\widehat{z}_{ti}$) within stratum h.

Under stratified two-stage sampling, a double sampling procedure is implemented: one for the primary sampling units (PSUs) and one for the secondary sampling units (SSUs). As PSUs are usually selected within each stratum with probabilities proportional to their sizes, second-order (π_{ij}) inclusion probabilities must be known to obtain an exact variance estimate. This is, however, generally unfeasible, and therefore, the ultimate cluster approximation (for details, see Kalton, 1979; Wolter, 2007) is used. In practice, it is assumed that PSUs are sampled with replacement even when in reality it is not the case. In such way, the leading contribution to the variance of the total estimator would come from the estimated PSUs totals, and the contribution from the second stage to the variance is disregarded. Within each stratum, the variance formula for cluster sampling selected with probability proportional to size with replacement is used, and by summing up the estimated strata variances due to independency between strata:

$$^{Str-Two}\widehat{var}\left(\widehat{Y}\right) = \sum_{v \in V} \frac{m_v}{(m_v - 1)} \sum_{\ell=1}^{m_v} \left(\widehat{Y}_{v\ell} - \frac{\widehat{Y}_v}{m_v}\right)^2, \tag{15}$$

where V is the set of strata composed of PSUs, m_v is the number of PSUs selected within stratum v, while $\widehat{Y}_{v\ell}$ is the sample total of Y in the ℓ^{th} PSU in stratum v:

$$\widehat{Y}_{v\ell} = \sum_{i \in S_{v\ell}} w_i y_i,$$

and $\widehat{Y}_v$ is the sample total of Y in stratum v:

$$\widehat{Y}_v = \sum_{l=1}^{m_v} \widehat{Y}_{v\ell}.$$

This approximation provides conservative variance estimates with an upward bias that becomes negligible as long as the sampling fractions of PSUs are very small. In order to estimate the variance of $\widehat{B}_r$ (resp. $\widehat{B}_t$) under stratified two-stage sampling, it is also sufficient to substitute $\widehat{z}_{ri}$ (resp. $\widehat{z}_{ti}$) for y_i in the variance estimator of the total estimator $\widehat{Y}$, where $\widehat{Y}_v$ is calculated as the sample total of $\widehat{z}_{ri}$ (resp. $\widehat{z}_{ti}$) in stratum v.

When the calibration estimator is used, a two-step procedure can be implemented for the variance estimation based on the proposed linearized variables. In the first step, the linearized variable $\widehat{z}_{\bullet i}$ would be calculated as if we were dealing with the Horvitz–Thompson estimator. In the second step, residual of this linearized variable is computed by performing a regression of this linearized variable on the calibration variables:

$$\check{\widehat{z}}_{\bullet i} = \widehat{z}_{\bullet i} - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_z,$$

where $\mathbf{x}_i$ is the vector of the auxiliary variables on which the sample is calibrated and $\widehat{\boldsymbol{\beta}}_z$ is the regression coefficients estimated by

$$\widehat{\boldsymbol{\beta}}_z := \left(\sum_{i \in S} \frac{q_i \mathbf{x}_i \mathbf{x}_i^\top}{\pi_i} \right)^{-1} \sum_{i \in S} \frac{q_i \mathbf{x}_i \widehat{z}_{\bullet i}}{\pi_i}.$$

The parameters q_i's make it possible to take into account the potential heteroscedasticity problem. The linearized variance estimator of $\widehat{B}_r$ (resp. $\widehat{B}_t$) in the case of calibration can be obtained by substituting $\check{\widehat{z}}_{ri}$ (resp. $\check{\widehat{z}}_{ti}$) for y_i in Expression (12) as a general formulation or more other expressions of the variance estimator of the total estimator for each specific sampling design. This procedure and some further knowledge on how to estimate sampling variance in the presence of non-response is, for example, described in Vallée and Tillé (2019).

5.2 | Variance estimation through R packages

The proposed methodology can be easily implemented for almost all sampling designs using the `R` packages on survey estimation available online, such as `survey` (Lumley, 2011) and `ReGenesees` (Zardetto, 2015). Both packages are devoted to sampling estimation.

In this paper, `ReGenesees`[1] is used. The variance estimation of $\widehat{B}_r$ (resp. $\widehat{B}_t$) can be carried out in three steps. The first step consists in computing $\widehat{z}_{ri}$ (resp. $\widehat{z}_{ti}$) from Expression (9) (resp. Expression (10)). In the second step, through the function `e.svydesign`, the sampling design adopted must be declared by identifying the variables (i.e. elementary units, primary sampling units, secondary sampling units, stratification variables, Horvitz–Thompson sampling weights and self-representative strata) in the sample data. If calibration is concerned and if the calibration weights have not yet been

[1]The package and the user manual are available online at https://www.istat.it/it/metodi-e-strumenti/metodi-e-strumenti-it/elaborazione/strumenti-di-elaborazione/regenesees. Last access, 31 January 2021.

calculated, the functions `pop.template` and `e.calibrate` should be used for defining the calibration totals and the calibration model. However, if the Horvitz–Thompson sampling weights have already been calibrated through other statistical software, instead of using the function `e.svydesign` at the beginning, the `ext.calibrated` function should be used for identifying the sampling design, the calibrated weights and the calibration model. The third step consists in estimating the variance of the total estimator of the linearized variables $\hat{z}_{ri}$ (resp. $\hat{z}_{ti}$) from Expression (8) using the function `svystatTM`. If calibration is concerned, the `svystatTM` function automatically computes the residuals of the linearized variables, and therefore, besides the computation of $\hat{z}_{ri}$ (resp. $\hat{z}_{ti}$), no further computation is needed.

6 | SIMULATION STUDIES

In this section, the empirical demonstrations on the consistency of the Bonferroni index estimators and the accuracy of their variance estimation using the proposed linearization technique are carried out on a series of simulations based on real data from the 2015 Italian component of the European Statistics on Income and Living Conditions (IT-SILC).

6.1 | IT-SILC data

The IT-SILC is included in the European framework of surveys yearly carried out in several countries to collect microdata on income, poverty, social exclusion and living conditions and provides both cross-sectional and longitudinal indicators on poverty and social cohesion in the European Union.

The IT-SILC (Istat, 2015) implements a stratified two-stage sampling design. The municipalities are the PSUs and the households are the SSUs. Municipalities are stratified according to their sizes and are split into self-representative (SR) municipalities and non self-representative (NSR) municipalities. The former are included directly in the sample, and therefore, there is only one single stage of selection for the households. The latter are selected with probabilities proportional to their sizes, and thus, there are two stages of selection, one for the municipalities and one for the households. In both cases, households are selected under a systematic sampling design in each selected municipality. All individuals older than 16 years old inside each selected household are surveyed. The sample size is determined for constructing reliable estimates of the main parameters both at the regional and at the national level. The Italian sample in 2015 consists of 615 PSUs, in which there are 116 SR municipalities and 499 NSR municipalities. Finally, a total number of 17985 households for 49987 individuals are selected.

The Bonferroni index is estimated by its two estimators introduced in Section 3, namely $\hat{B}_r$ from Expression (5) and $\hat{B}_t$ from Expression (6). The variances of the two estimators are approximated using the aforementioned `ReGenesees` package in `R` and the confidence intervals are constructed using Expression (16) in Section 6.2. Both estimators, $\hat{B}_r$ and $\hat{B}_t$, yield the same result with the same length of confidence interval. The estimate of the Bonferroni index is equal to 0.490 with a 95% confidence interval [0.484, 0.495]. Similarly, the estimate of the Gini index is computed using the estimator $\hat{G}$ from Expression (7). Its variance and the associated confidence interval are also estimated. The estimate of the Gini index is equal to 0.365 with a 95% confidence interval [0.358, 0.371]. The Bonferroni index, as demonstrated in Section 2, assumes by definition a higher value than the Gini index.

6.2 | Simulation scheme and results

Simulation studies are performed based on the 2015 IT-SILC survey data. The obtained estimate of the Bonferroni index presented in the previous subsection is treated as the true value, B.

SRSWOR and StrSRS are designed for the simulation studies. For SRSWOR, the population of households has been reconstructed by duplicating the original sample based on the calibration survey weights. For StrSRS, a pseudo population is constructed using the multivariate hypergeometric distribution, that is, within one stratum, each calibration survey weight of the household income is viewed as a sub-population from which a sub-sample is drawn and the allocated sample size in that stratum (proportional allocation with a minimum of two sampling units guaranteed in each stratum) is the total number of draws. The multivariate hypergeometric distribution is applied to every stratum and the whole sample can be retrieved by aggregating the samples inside all strata. The strata are determined according to the regions of Italy. There are 21 strata in total and they are the Italian region Piedmont, Aosta Valley, Lombardy, Veneto, Friuli Venezia Giulia, Liguria, Emilia-Romagna, Tuscany, Umbria, Marche, Lazio, Abruzzo, Molise, Campania, Apulia, Basilicata, Calabria, Sicily, Sardinia, plus Province of Bolzano and Province of Trento from the Trentino-Alto Adige/Südtirol region. Samples with a large range of sample sizes are considered, that is, $n = 100$, 500, 1000, 2000, 5000 and 10,000. The sample selection is repeated R = 10,000 times for each sample size.

First, behaviours of the two Bonferroni index estimators are investigated. The relative bias and the normalized root-mean-square error (NRMSE) of $\widehat{B}_r$ and $\widehat{B}_t$ are computed for different sample sizes under SRSWOR and StrSRS. The empirical relative bias of $\widehat{B}_r$ is defined as

$$RelativeBias_{sim}(\widehat{B}_r) = \frac{\frac{1}{R}\sum_{i=1}^{R}\widehat{B}_r^{(i)} - B}{B},$$

while the empirical NRMSE of $\widehat{B}_r$ is

$$NRMSE_{sim}(\widehat{B}_r) = \frac{\sqrt{\frac{1}{R}\sum_{i=1}^{R}\left(\widehat{B}_r^{(i)} - B\right)^2}}{\frac{1}{R}\sum_{i=1}^{R}\widehat{B}_r^{(i)}}.$$

The empirical relative bias and the empirical NRMSE of $\widehat{B}_t$ are similar by simply replacing $\widehat{B}_r^{(i)}$ in the previous two expressions with $\widehat{B}_t^{(i)}$.

The results are presented in Tables 1 and 2. They show that both estimators slightly underestimate the Bonferroni index when $n = 100$. They are both asymptotically approximately unbiased as the relative biases of the two estimators are merely negligible when n reaches 10,000. Among the two estimators, $\widehat{B}_t$ is much less biased than $\widehat{B}_r$. Under SRSWOR, the relative bias of $\widehat{B}_t$ is around one half of the relative bias of $\widehat{B}_r$ when $n \leq 5000$, and the bias decreases drastically when $n = 10{,}000$. Under StrSRS, the relative bias of $\widehat{B}_t$ is around one half of the relative bias of $\widehat{B}_r$ when $n \leq 2000$, and the bias drops massively when n reaches 5000. In terms of NRMSE, $\widehat{B}_t$ also performs better than $\widehat{B}_r$ for small sample sizes, that is, $n \leq 1000$ under SRSWOR and $n \leq 500$ under StrSRS. For large sample sizes, it seems that $\widehat{B}_r$ outperforms $\widehat{B}_t$ regarding NRMSE; however, the differences between the two estimators are negligible in those cases. The relative bias (in absolute value) and NRMSE of both estimators decrease when the sample size increases.

Furthermore, the performance of the variance estimator $\widehat{var}_{lin}(\widehat{B}_r)$ (resp. $\widehat{var}_{lin}(\widehat{B}_t)$) constructed using the linearized variable $\widehat{z}_{ri}$ (resp. $\widehat{z}_{ti}$) computed from Expression (9) (resp. (10)) is assessed.

TABLE 1 Relative bias and NRMSE of $\widehat{B}_r$ and $\widehat{B}_t$ for different sample sizes under SRSWOR

	sample size					
	100	**500**	**1000**	**2000**	**5000**	**10,000**
Relative bias(%)						
$\widehat{B}_r$	−2.059	−0.478	−0.203	−0.105	−0.029	−0.024
$\widehat{B}_t$	−1.265	−0.229	−0.085	−0.054	−0.016	0.001
NRMSE(%)						
$\widehat{B}_r$	6.872	2.941	2.085	1.461	0.922	0.656
$\widehat{B}_t$	6.558	2.928	2.051	1.469	0.935	0.660

Abbreviations: NRMSE, normalized root-mean-square error; SRSWOR, simple random sampling without replacement with fixed sample size.

TABLE 2 Relative bias and NRMSE of $\widehat{B}_r$ and $\widehat{B}_t$ for different sample sizes under StrSRS

	sample size					
	100	**500**	**1000**	**2000**	**5000**	**10,000**
Relative bias(%)						
$\widehat{B}_r$	−2.377	−0.440	−0.216	−0.111	−0.058	−0.006
$\widehat{B}_t$	−1.215	−0.245	−0.129	−0.050	−0.008	0.001
NRMSE(%)						
$\widehat{B}_r$	7.055	2.952	2.074	1.453	0.922	0.654
$\widehat{B}_t$	6.752	2.942	2.075	1.465	0.926	0.655

Abbreviations: NRMSE, normalized root-mean-square error; StrSRS, stratified simple random sampling.

TABLE 3 Variance of $\widehat{B}_r$ and $\widehat{B}_t$ based on linearization and the Monte Carlo method for different sample sizes (results enlarged by 10,000 times) under simple random sampling without replacement with fixed sample size

	sample size					
	100	**500**	**1000**	**2000**	**5000**	**10,000**
$E_{sim}\widehat{var}_{lin}(\widehat{B}_r)$	8.638	1.983	1.010	0.511	0.207	0.103
$var_{sim}(\widehat{B}_r)$	9.839	1.999	1.028	0.508	0.203	0.103
$E_{sim}\widehat{var}_{lin}(\widehat{B}_t)$	8.872	1.997	1.021	0.514	0.206	0.103
$var_{sim}(\widehat{B}_t)$	9.664	2.032	1.005	0.516	0.209	0.104

Tables 3 and 4 present the variance estimates of $\widehat{B}_r$ and $\widehat{B}_t$, enlarged by 10,000 times, obtained through linearization and the Monte Carlo method for different sample sizes under SRSWOR and StrSRS. For a specific sample size n, given that a variance estimate through linearization is obtained for each selected sample, $E_{sim}\widehat{var}_{lin}(\widehat{B}_r)$ is computed by averaging the 10,000 replicates of $\widehat{var}_{lin}(\widehat{B}_r)$, while

$$var_{sim}(\widehat{B}_r) = \frac{1}{R-1}\sum_{i=1}^{R}\left(\widehat{B}_r^{(i)} - \overline{\widehat{B}_r}\right)^2,$$

TABLE 4 Variance of $\widehat{B}_r$ and $\widehat{B}_t$ based on linearization and the Monte Carlo method for different sample sizes (results enlarged by 10,000 times) under stratified simple random sampling

	sample size					
	100	**500**	**1000**	**2000**	**5000**	**10,000**
$E_{sim}\widehat{var}_{lin}(\widehat{B}_r)$	9.062	1.973	1.004	0.506	0.204	0.102
$var_{sim}(\widehat{B}_r)$	10.015	2.024	1.015	0.502	0.203	0.102
$E_{sim}\widehat{var}_{lin}(\widehat{B}_t)$	9.130	1.990	1.005	0.508	0.204	0.102
$var_{sim}(\widehat{B}_t)$	10.307	2.050	1.025	0.513	0.205	0.103

where

$$\overline{\widehat{B}_r} = \frac{1}{R}\sum_{i=1}^{R}\widehat{B}_r^{(i)}.$$

Similarly, $E_{sim}\widehat{var}_{lin}(\widehat{B}_t)$ and $var_{sim}(\widehat{B}_t)$ are computed. Since the variance estimates obtained through Monte Carlo experiments, namely $var_{sim}(\widehat{B}_r)$ (resp. $var_{sim}(\widehat{B}_t)$), approximate the true variance of $\widehat{B}_r$ (resp. $\widehat{B}_t$) for each sample size, they can be considered as benchmarks.

As shown in Tables 3 and 4, the variance estimators constructed using the linearization method are approximately unbiased, at least asymptotically. Underestimation of the variances of both $\widehat{B}_r$ and $\widehat{B}_t$ is observed when the sample sizes are small and the bias could be large. For example, when $n = 100$, the relative bias of the variance estimators could reach 8–12% under SRSWOR and 9–11% under StrSRS. However, once the sample size becomes sufficiently large, that is, when n reaches 1000, the variance estimates obtained using the linearization method are almost equal to the Monte Carlo variances for both sampling designs.

The linearized variance estimators can be utilized to construct empirical confidence intervals. By assuming standard normal deviates, the empirical 95% confidence interval of $\widehat{B}_r$ is

$$\left[\widehat{B}_r - 1.96\cdot\sqrt{\widehat{var}_{lin}(\widehat{B}_r)},\ \widehat{B}_r + 1.96\cdot\sqrt{\widehat{var}_{lin}(\widehat{B}_r)}\right]. \tag{16}$$

The empirical 95% confidence interval of $\widehat{B}_t$ is similar by substituting $\widehat{B}_t$ for $\widehat{B}_r$ in Expression (16).

The empirical coverage rate is the proportion of the successful coverage of the true value of the Bonferroni index (B) within the 10,000 empirical confidence intervals for each sample size. In Tables 5 and 6, the empirical coverage rates of the 95% confidence intervals for both $\widehat{B}_r$ and $\widehat{B}_t$ under SRSWOR and StrSRS are presented. The empirical coverage rate rises with the increase in the sample size. The true coverage rates of B are mostly slightly underestimated, which could be a consequence of the underestimation of B inherited from the two estimators $\widehat{B}_r$ and $\widehat{B}_t$. The underestimation of the variance estimators and the potential failure of the normality assumption could also be factors which cause the undercoverage for small sample sizes. However, when the sample size reaches 5000, there seems no significant differences with the true 95% coverage rates in terms of a t-test at a significance level of 5% for both sampling designs. An approach recently proposed by Berger and Gedik Balay (2020) is devoted to the inference for the Gini index using the empirical likelihood method for overcoming the normality assumption. The transposition of this method to the Bonferroni index could be a promising alternative method to compute the confidence interval.

TABLE 5 Empirical coverage rates of 95% confidence intervals for $\widehat{B}_r$ and $\widehat{B}_t$ under simple random sampling without replacement with fixed sample size

	sample size					
	100	**500**	**1000**	**2000**	**5000**	**10,000**
$\widehat{B}_r$	0.889	0.934	0.942	0.944	0.949	0.950
$\widehat{B}_t$	0.917	0.940	0.947	0.945	0.946	0.948

TABLE 6 Empirical coverage rates of 95% confidence intervals for $\widehat{B}_r$ and $\widehat{B}_t$ under stratified simple random sampling

	sample size					
	100	**500**	**1000**	**2000**	**5000**	**10,000**
$\widehat{B}_r$	0.886	0.931	0.939	0.944	0.948	0.951
$\widehat{B}_t$	0.908	0.938	0.939	0.944	0.948	0.950

7 | SENSITIVITY TO DIFFERENT LEVELS OF DISTRIBUTION

7.1 | Influence function and linearized variable

Influence function (Hampel, 1974; Hampel et al., 1985) is a statistical tool used for deriving asymptotic variances and investigating local robustness properties. It is essentially the first derivative of an estimator viewed as functional. Let $T(F)$ be a functional. The influence function is

$$IF_x(T) := \lim_{\varepsilon \to 0^+} \frac{T(\varepsilon\delta_x + (1-\varepsilon)F) - T(F)}{\varepsilon},$$

where δ_x denotes the pointmass 1 at x. In this paper, $x \in R_{\geq 0}$ as it is a variable representing the point of income which receives an infinitesimal perturbation. Thus, influence function of an inequality measure shows the effect of an infinitesimal perturbation at the point of income x on the inequality measure. The linearized variable obtained with the Graf method in Section 4 could be interpreted as the discrete analogue of the influence function.

The influence function of the Lorenz curve (see among others Essama-Nssah & Lambert, 2011) is

$$IF_x\{L(p)\} = \frac{1}{\mu}\left\{x\mathbb{1}(x \leq Q(p)) + Q(p)\left[p - \mathbb{1}(x \leq Q(p))\right]\right\} - L(p)\frac{x}{\mu}.$$

Monti (1991) has shown that the influence function of the Gini index is

$$IF_x(G) = \frac{1}{\mu}\{2F(x)[x - \mu(x)] + \mu - x - (\mu + x)G\} \tag{17}$$

(see also Langel & Tillé, 2013, for a synthesis). It is important to point out that, up to a multiplicative factor, the linearized variable of $\widehat{G}$ from Expression (11) in Section 4 is the discrete analogue of the influence function of the Gini index given in (17).

The first derivative of (17) is

$$\begin{aligned}\frac{\partial IF_x(G)}{\partial x} &= \frac{1}{\mu}\left[2f(x)x + 2F(x) - 2f(x)\mu(x) - 2F(x)\frac{\partial \mu(x)}{\partial x} - 1 - G\right] \\ &= \frac{2}{\mu}f(x)x + \frac{2}{\mu}F(x) - \frac{2}{\mu}xf(x) - \frac{1}{\mu} - G\frac{1}{\mu} \\ &= \frac{1}{\mu}\left[2F(x) - 1 - G\right],\end{aligned}$$

since $\frac{\partial \mu(x)}{\partial x} = \frac{xf(x) - \mu(x)f(x)}{F(x)}$. The influence function of the Gini index is a convex function as the second derivative of (17) is positive:

$$\frac{\partial^2 IF_x(G)}{\partial x^2} = \frac{2f(x)}{\mu} > 0$$

(see also Monti, 1991, p. 566).

Result 3 The influence function of the Bonferroni index is

$$\begin{aligned}IF_x(B) &= \frac{x - \mu(x)}{\mu} - \frac{x}{\mu}\int_0^\infty \frac{\mathbb{1}(x \le y)}{F(y)}dF(y) + \frac{1}{\mu}\int_0^\infty \frac{\mu(y)\mathbb{1}(x \le y)}{F(y)}dF(y) - \frac{xB}{\mu} \\ &= \frac{x\log\, F(x)}{\mu} + \frac{x - \mu}{\mu} + \frac{1}{\mu}\int_0^\infty \frac{y\mathbb{1}(x \le y)}{F(y)}dF(y) - B\frac{x}{\mu}.\end{aligned} \tag{18}$$

The proof is given in Appendix and Supplementary Material A.

Notice that, up to a multiplicative factor, the linearized variable of the plug-in estimator of the Bonferroni index $\widehat{B}_r$ from Expression (9) is the discrete analogue of (18).

The first derivative of (18) is

$$\begin{aligned}\frac{\partial IF_x(B)}{\partial x} &= \frac{1}{\mu}\left[\log\, F(x) + x\frac{1}{F(x)}f(x)\right] + \frac{1}{\mu} - \frac{1}{\mu}\frac{x}{F(x)}f(x) - \frac{B}{\mu} \\ &= \frac{1}{\mu}\log\, F(x) + \frac{1}{\mu} - \frac{B}{\mu}.\end{aligned}$$

The influence function of the Bonferroni index is also a convex function as the second derivative of (18) is positive:

$$\frac{\partial^2 IF_x(B)}{\partial x^2} = \frac{f(x)}{\mu F(x)} > 0.$$

The comparison between the second derivatives of the influence functions of the Bonferroni and Gini indices shows that

$$\begin{cases}\dfrac{\partial^2 IF_x(G)}{\partial x^2} < \dfrac{\partial^2 IF_x(B)}{\partial x^2} & \text{if} \quad x < Q(1/2) \\ \dfrac{\partial^2 IF_x(G)}{\partial x^2} \ge \dfrac{\partial^2 IF_x(B)}{\partial x^2} & \text{if} \quad x \ge Q(1/2).\end{cases}$$

On the one hand, the curvature of the influence function of the Bonferroni index is stronger for the small x values than that of the Gini index. On the other hand, the curvature of the influence function of the Gini index is stronger for the large x values than that of the Bonferroni index. Furthermore, the curvature of the influence function of the Bonferroni index at the smallest x values is of size $1/2F(x)$ greater than that of the Gini index. That is, asymptotically, for a well-behaved income distribution, the curvature of the influence function of the Bonferroni index at the smallest x values can be much stronger than that of the Gini index. This result thus clearly confirms that the Bonferroni index is more influenced by the smallest incomes than the Gini index.

7.2 | Empirical demonstration

The significance of the relation between the linearized variables in Section 4 and the influence functions is twofold. Deville (1999) proposes a generalized linearization method based on the concept of influence function that under mild conditions provides an approximately unbiased variance estimation for non-linear statistics, for instance, $\widehat{B}_{\bullet}$ and $\widehat{G}$ in this paper. Viewing the linearized variables derived with the Graf method as discrete analogues of the influence functions, the simulation results in Section 6 provides a further demonstration of the use of influence function for variance estimation. In such view, the linearized variables in Section 4 could also be utilized for understanding and comparing the sensitivity of the Bonferroni and Gini indices to different levels of the income distribution.

In Figure 2, the values of $\widehat{z}_{ri}$'s (i.e. the linearized variables of $\widehat{B}_r$) on each observed equivalized income estimated from a sample of 100 households selected under SRSWOR from the reconstructed population of the Italian households based on the real IT-SILC data are compared with those of $\widehat{z}_{Gi}$'s (i.e. the linearized variables of $\widehat{G}$). The values of $\widehat{z}_{ti}$'s (i.e. the linearized variables of $\widehat{B}_t$) are almost

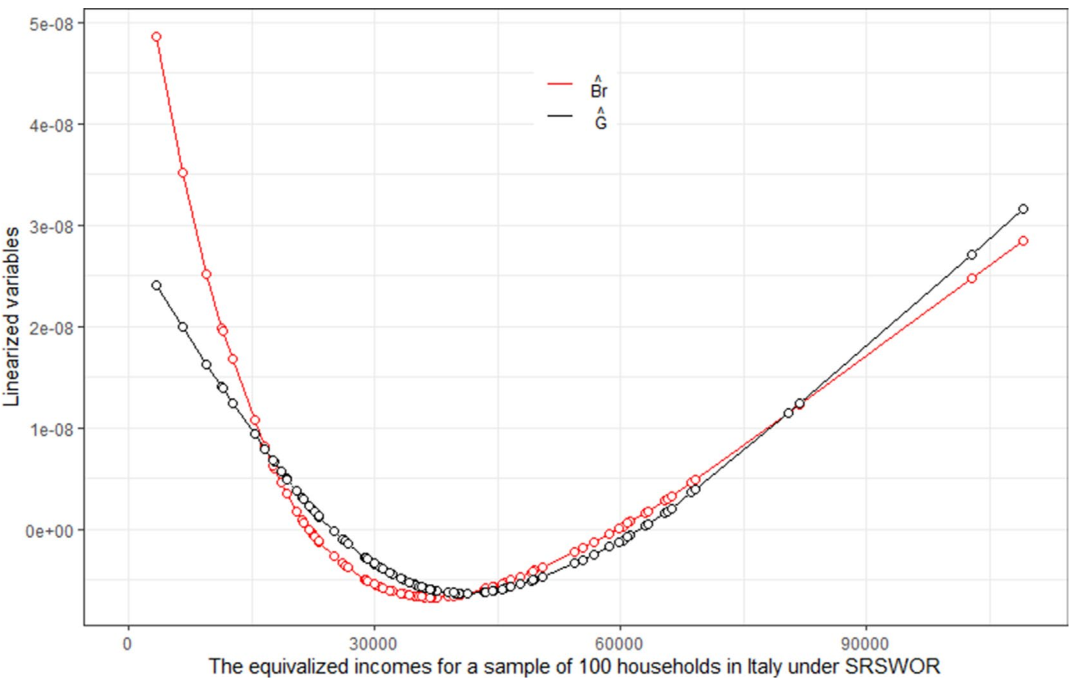

FIGURE 2 Influence of incomes on $\widehat{B}_r$ and $\widehat{G}$. SRSWOR, simple random sampling without replacement with fixed sample size

identical with those of $\hat{z}_{ri}$'s, and therefore, only $\hat{B}_r$ is represented. The contributions of the different observations to the inequality indices can thus be directly appreciated.

As shown in Figure 2, in the beginning, the curvature of the influence function of $\hat{B}_r$ is stronger than that of $\hat{G}$. This is particularly true for the smallest incomes. In the middle part of the distribution, the observations contribute approximately the same amount to the indices. Finally, in the right tail of the distribution, the curvature of the influence function of $\hat{G}$ is stronger than that of $\hat{B}_r$. Thus, the observations at the very beginning which are composed of households with the smallest incomes are more 'influential' for $\hat{B}$ than for $\hat{G}$, while the observations in the tail part of the distribution play a relatively more important role in $\hat{G}$.

Figure 2, as a typical example, confirms the theoretical results obtained in the previous subsection. They stress the different sensitivity of the Bonferroni and Gini indices to different levels of the income distribution. It demonstrates, from a different and new perspective, the evidence stated by De Vergottini (1950) and Pizzetti (1951). Furthermore, it serves to relaunch the Bonferroni index as a complementary and not only as an alternative inequality measure to the Gini index for studying income inequality.

8 | SOME CONCLUDING REMARKS

Researchers increasingly advocate the use of further inequality measures besides the Gini index. Their use, simultaneously with the evergreen Gini index, can help to better catch the inequality that lurks in different parts of the wealth distribution. The present paper focuses on the Bonferroni inequality index. Because income data are usually collected through sample surveys, attention has been paid to its inferential aspects.

As the main aim, the linearized variables for the plug-in estimator and an alternative sampling estimator of the Bonferroni index have been computed. The Graf method has been used, that is, the estimators have been linearized by differentiating them with respect to their sample indicators. A Monte Carlo simulation carried out on real income data demonstrates that both estimators are approximately unbiased, at least asymptotically, and the proposed linearization method provides a valid inference on the variances of the estimators. Furthermore, interesting results have been obtained by interpreting the linearized variable as the influence function. In this sense, the influence function provides a sensitivity measure of the inequality index to different levels of the income distribution. In particular, the curvature of the influence function of the Bonferroni index is compared with that of the Gini index. An example on a sample of real income data has been illustrated. In this paper, it has been demonstrated, from a different and new perspective, that the Bonferroni index is more sensitive to the lowest incomes in the distribution than the Gini index. Hence, the Bonferroni inequality index is relaunched not only as an alternative but also as a complement to the Gini index.

Further studies may concern some additional inequality indices more influenced by large incomes or by average incomes. In this way, it will be possible to use a set of inequality indices with different sensitivity to different parts of the wealth distribution for providing a fully comprehensive evaluation of inequality.

ACKNOWLEDGEMENTS

The authors express their gratitude for the meticulous reading and the constructive comments from the joint editor, the associate editor and two anonymous referees.

ORCID

Ziqing Dong http://orcid.org/0000-0002-2837-2485
Alessio Guandalini http://orcid.org/0000-0001-8799-6602

REFERENCES

Aaberge, R. (2000) Characterizations of Lorenz curves and income distributions. *Social Choice and Welfare*, 17(4), 639–653.

Aaberge, R. (2007) Gini's nuclear family. *The Journal of Economic Inequality*, 5(3), 305–322.

Bárcena, E. & Imedio, L.J. (2008) The Bonferroni, Gini, and De Vergottini indices. Inequality, welfare, and deprivation in the European union in 2000. In: *Inequality and opportunity: Papers from the second ECINEQ society meeting*, Bingley (UK): Emerald Group Publishing Limited, pp. 231–257.

Bárcena-Martín, E. & Silber, J. (2011) On the concepts of Bonferroni segregation index and curve. *Rivista Italiana di Economia, Demografia e Statistica*, 62(2), 57–74.

Bárcena-Martín, E. & Silber, J. (2013) On the generalization and decomposition of the Bonferroni index. *Social Choice and Welfare*, 41(4), 763–787.

Bárcena-Martín, E. & Silber, J. (2017) The Bonferroni index and the measurement of distributional change. *Metron*, 75(1), 1–16.

Benedetti, C. (1986) Sulla interpretazione benesseriale di noti indici di concentrazione e di altri. *Metron*, 44(1), 421–429.

Berger, Y. & Gedik Balay, İ. (2020) Confidence intervals of Gini coefficient under unequal probability sampling. *Journal of Official Statistics*, 36(2), 237–249.

Binder, D.A. (1983) On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, 5(3), 279–292.

Binder, D.A. (1991) Use of estimating functions for interval estimation from complex surveys. *Proceedings of the ASA Survey Research Methods Section*, 1991, 34–42.

Binder, D.A. & Patak, Z. (1994) Use of estimating functions for estimation from complex surveys. *Journal of the American Statistical Association*, 89(427), 1035–1043.

Bonferroni, C.E. (1933) *Elementi di statistica generale*. Torino: Litografia Felice Gili.

Chakravarty, S.R. (2007) A deprivation-based axiomatic characterization of the absolute Bonferroni index of inequality. *The Journal of Economic Inequality*, 5(3), 339–351.

Chakravarty, S.R. & Muliere, P. (2004) Welfare indicators: A review and new perspectives. 2. Measurement of poverty. *Metron*, 62(2), 247–281.

De Vergottini, M. (1950) Sugli indici di concentrazione. *Statistica*, 10(4), 445–454.

Demnati, A. & Rao, J.N.K. (2004) Linearization variance estimators for survey data. *Survey Methodology*, 30(1), 17–26.

Deville, J.-C. (1999) Variance estimation for complex statistics and estimators: Linearization and residual techniques. *Survey Methodology*, 25(2), 193–203.

Deville, J.-C. & Särndal, C.-E. (1992) Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418), 376–382.

Essama-Nssah, B. & Lambert, P. J. (2011) Influence functions for distributional statistics. *Society for the study of economic inequality working paper. ECINEQ WP*, 236, 2011.

Gastwirth, J.L. (1971) A general definition of the Lorenz curve. *Econometrica*, 39(6), 1037–1039.

Gini, C. (1914) Sulla misura della concentrazione e della variabilità dei caratteri. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, 73(2), 1203–1248. English translation In *Metron* 2005, 63(1), 3–38.

Giordani, P. & Giorgi, G.M. (2010) A fuzzy logic approach to poverty analysis based on the Gini and Bonferroni inequality indices. *Statistical Methods and Applications*, 19(4), 587–607.

Giorgi, G.M. (1998) Concentration index, Bonferroni. In: Kotz, S., Johnson, H.L. & Read, C.B. (eds.) *Encyclopedia of statistical sciences, update 2*, volume 2. New York: Wiley, pp. 141–146.

Giorgi, G.M. (2005) Gini's scientific work: An evergreen. *Metron*, 63(3), 299–315.

Giorgi, G.M. (2019) The Gini concentration ratio: Back to the future. *Rivista Italiana di Economia, Demografia e Statistica*, 73(2), 5–14.

Giorgi, G.M. & Crescenzi, M. (2001a) Bayesian estimation of the Bonferroni index from a Pareto-type I population. *Statistical Methods and Applications*, 10(1-3), 41–48.

Giorgi, G.M. & Crescenzi, M. (2001b) A look at the Bonferroni inequality measure in a reliability framework. *Statistica*, 61(4), 571–583.

Giorgi, G.M. & Gigliarano, C. (2017) The Gini concentration index: A review of the inference literature. *Journal of Economic Surveys*, 31(4), 1130–1148.

Giorgi, G. M. & Guandalini, A. (2013) A sampling estimator of the Bonferroni inequality index. *Rivista Italiana di Economia, Demografia e Statistica*, 67(3-4), 151–158.

Giorgi, G.M. & Guandalini, A. (2018) Decomposing the Bonferroni inequality index by subgroups: Shapley value and balance of inequality. *Econometrics*, 6(2), 1–16.

Giorgi, G.M. & Mondani, R. (1994) The exact sampling distribution of the Bonferroni concentration index. *Metron*, 52(3-4), 5–41.

Giorgi, G. M. & Mondani, R. (1995) Sampling distribution of the Bonferroni inequality index from exponential population. *Sankhyā: The Indian Journal of Statistics, Series B*, 57(1), 10–18.

Giorgi, G.M. & Nadarajah, S. (2010) Bonferroni and Gini indices for various parametric families of distributions. *Metron*, 68(1), 23–46.

Graf, M. (2011) *Use of survey weights for the analysis of compositional data*. Chichester: Wiley, pp. 114–127.

Graf, E. & Tillé, Y. (2014) Variance estimation using linearization for poverty and social exclusion indicators. *Survey Methodology*, 40(1), 61–79.

Hampel, F.R. (1974) The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346), 383–393.

Hampel, F.R, Ronchetti, E.M., Rousseeuw, P.J. & Stahel, W.A. (1985) *Robust statistics: The approach based on influence functions*, volume 196. Hoboken: John Wiley & Sons.

Horvitz, D.G. & Thompson, D.J. (1952) A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260), 663–685.

Imedio-Olmedo, L.J., Parrado-Gallardo, E.M. & Bárcena-Martín, E. (2012) Income inequality indices interpreted as measures of relative deprivation/satisfaction. *Social Indicators Research*, 109(3), 471–491.

Istat (2015) Indagine sulle condizioni di vita (UDB IT-SILC). Available from: https://www.istat.it/it/archivio/4152, (accessed on 4 December 2017).

Kalton, G. (1979) Ultimate cluster sampling. *Journal of the Royal Statistical Society, Series A (General)*, 142(2), 210–222.

Langel, M. & Tillé, Y. (2013) Variance estimation of the Gini index: Revisiting a result several times published. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), 521–540.

Lorenz, M.O. (1905) Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, 9(70), 209–219.

Lumley, T. (2011) *Complex surveys: A guide to analysis using R*, volume 565. Hoboken: John Wiley & Sons.

Monti, A.C. (1991) The study of the Gini concentration ratio by means of the influence function. *Statistica*, 51(4), 561–580.

Nygård, F. & Sandström, A. (1981) *Measuring income inequality*. Stockholm: Almqvist & Wiksell International.

Osberg, L. (2017). On the limitations of some current usages of the Gini index. *Review of Income and Wealth*, 63(3), 574–584.

Piesch, W. (1975) *Statistische Konzentrationsmaße*. Tübingen: J.B.C. Mohr (Paul Siebeck).

Pietra, G. (1915) Delle relazioni tra gli indici di variabilità. *Atti del Reale Istituto Veneto di Scienze,Lettere ed Arti*, 74(2), 775–804. English traslation In: *Metron* 2014, 72(1), 5–16.

Piketty, T. (2015) About capital in the twenty-first century. *American Economic Review*, 105(5), 48–53.

Pizzetti, E. (1951) Relazioni tra indici di concentrazione. *Statistica*, 11(3-4), 294–316.

Pundir, S., Arora, S. & Jain, K. (2005). Bonferroni curve and the related statistical inference. *Statistics and Probability Letters*, 75(2), 140–150.

Särndal, C.-E. (2007) The calibration approach in survey theory and practice. *Survey Methodology*, 33(2), 99–119.

Särndal, C.-E. & Lundström, S. (2005) *Estimation in surveys with nonresponse*. Hoboken: John Wiley & Sons.

Silber, J. & Son, H. (2010) On the link between the Bonferroni index and the measurement of inclusive growth. *Economics Bulletin*, 30(1), 421–428.

Stiglitz, J.E. (2012) *The price of inequality: How today's divided society endangers our future*. New York: W. W. Norton & Company.

Tarsitano, A. (1990) The Bonferroni index of income inequality. In: Dagum, C. & Zenga, M. (eds.) *Income and Wealth Distribution, Inequality and Poverty*. Berlin: Springer, pp. 228–242.
Vallée, A.-A. & Tillé, Y. (2019) Linearisation for variance estimation by means of sampling indicators: Application to non-response. *International Statistical Review*, 87(2), 347–367.
Wolter, K. (2007) *Introduction to variance estimation*. Berlin: Springer Science & Business Media.
Woodruff, R.S. (1971) A simple method for approximating the variance of a complicated estimate. *Journal of the American Statistical Association*, 66(334), 411–414.
Xu, K. (2003) How has the literature on Gini's index evolved in the past 80 years?*Dalhousie University, Economics Working Paper*.
Yitzhaki, S. (1998) More than a dozen alternative ways of spelling Gini. *Research on Economic Inequality*, 8, 13–30.
Zardetto, D. (2015) Regenesees: An advanced R system for calibration, estimation and sampling error assessment in complex sample surveys. *Journal of Official Statistics*, 31(2), 177–203.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Dong Z, Tillé Y, Giorgi GM, Guandalini A. Linearization and variance estimation of the Bonferroni inequality index. *J R Stat Soc Series A*. 2021;184:1008–1029. https://doi.org/10.1111/rssa.12701

APPENDIX

Proof of Result 1 First, denote u_i and $u_{i,k}$ as the derivatives of $\hat{\hat{Y}}$ and $\hat{\hat{Y}}_k$ with respect to a_i:

$$u_i = \frac{\partial \hat{\hat{Y}}}{\partial a_i} = \frac{w_i y_i \left(\sum_{j\in U} w_j a_j\right) - \left(\sum_{j\in U} w_j y_j a_j\right) w_i}{\left(\sum_{j\in U} w_j a_j\right)^2} = \frac{w_i}{\hat{N}}\left(y_i - \hat{\hat{Y}}\right) \tag{19}$$

and

$$u_{i,k} = \frac{\partial \hat{\hat{Y}}_k}{\partial a_i} = \frac{w_i y_i \mathbb{1}[y_i \le y_k] - \hat{\hat{Y}}_k w_i \mathbb{1}[y_i \le y_k]}{\sum_{i\in U} w_i \mathbb{1}[y_i \le y_k] a_i} = \frac{w_i \mathbb{1}[y_i \le y_k]}{\hat{N}_k}\left(y_i - \hat{\hat{Y}}_k\right). \tag{20}$$

One can write $\hat{B}_r = A \times P$, where $A = \mathbb{1}/[(\hat{N}-1)\hat{\hat{Y}}]$ and $P = \sum_{k\in U} a_k \left[w_k\left(\hat{\hat{Y}} - \hat{\hat{Y}}_k\right)\right]$. It is then possible to calculate

$$\frac{\partial \hat{B}_r}{\partial a_i} = P\frac{\partial A}{\partial a_i} + A\frac{\partial P}{\partial a_i}. \tag{21}$$

As

$$\frac{\partial A}{\partial a_i} = -\frac{\hat{\hat{Y}}\frac{\partial \hat{N}}{\partial a_i} + (\hat{N}-1)\frac{\partial \hat{\hat{Y}}}{\partial a_i}}{[(\hat{N}-1)\hat{\hat{Y}}]^2} = -\frac{\frac{\partial}{\partial a_i}\left(\hat{\hat{Y}}\hat{N}\right) - u_i}{[(\hat{N}-1)\hat{\hat{Y}}]^2} = \frac{u_i - w_i y_i}{[(\hat{N}-1)\hat{\hat{Y}}]^2} \tag{22}$$

and

$$\frac{\partial P}{\partial a_i} = \sum_{k\in U} \left[w_k \left(u_i - u_{i,k}\right) a_k\right] + w_i \left(\widehat{\widehat{Y}} - \widehat{\widehat{Y}}_i\right), \tag{23}$$

by plugging (22) and (23) into (21):

$$\begin{aligned}\frac{\partial \widehat{B}_r}{\partial a_i} = &\frac{u_i - w_i y_i}{\left(\widehat{N}-1\right)\widehat{\widehat{Y}}} \times \widehat{B}_r + \frac{1}{\left(\widehat{N}-1\right)\widehat{\widehat{Y}}} \times \sum_{k\in U} \left[w_k \left(u_i - u_{i,k}\right) a_k\right] \\ &+ \frac{1}{\left(\widehat{N}-1\right)\widehat{\widehat{Y}}} \times w_i \left(\widehat{\widehat{Y}} - \widehat{\widehat{Y}}_i\right).\end{aligned}$$

Through further simplification:

$$\frac{\partial \widehat{B}_r}{\partial a_i} = \frac{u_i - w_i y_i}{\left(\widehat{N}-1\right)\widehat{\widehat{Y}}} \times \widehat{B}_r + \frac{1}{\left(\widehat{N}-1\right)\widehat{\widehat{Y}}} \times w_i \left(y_i - \widehat{\widehat{Y}}_i\right) - \frac{1}{\left(\widehat{N}-1\right)\widehat{\widehat{Y}}} \times \sum_{k\in U} w_k u_{i,k} a_k. \tag{24}$$

After replacing u_i and $u_{i,k}$ with Expression (19) and (20) and dividing Expression (24) by w_i, Result 1 is obtained. □

Proof of Result 2 Similar to the proof of Result 1, one can write $\widehat{B}_t = A \times R$, where $A = \mathbb{1}/[(\widehat{N}-1)\widehat{\widehat{Y}}]$ and $R = \sum_{k\in U} a_k \left\{ w_k \left(\widehat{\widehat{Y}} - \frac{\widehat{\widehat{Y}}_k + \widehat{\widehat{Y}}_{k-1}}{2} \right) \right\}$. It is then possible to calculate

$$\frac{\partial \widehat{B}_t}{\partial a_i} = R\frac{\partial A}{\partial a_i} + A\frac{\partial R}{\partial a_i}. \tag{25}$$

As

$$\frac{\partial R}{\partial a_i} = \sum_{k\in U} \left[w_k \left(u_i - \frac{u_{i,k} + u_{i,k-1}}{2} \right) a_k \right] + w_i \left(\widehat{\widehat{Y}} - \frac{\widehat{\widehat{Y}}_i + \widehat{\widehat{Y}}_{i-1}}{2} \right), \tag{26}$$

by plugging (22) and (26) into (25):

$$\begin{aligned}\frac{\partial \widehat{B}_t}{\partial a_i} = &\frac{u_i - w_i y_i}{\left(\widehat{N}-1\right)\widehat{\widehat{Y}}} \times \widehat{B}_t + \frac{1}{\left(\widehat{N}-1\right)\widehat{\widehat{Y}}} \times \sum_{k\in U} \left\{ w_k \left(u_i - \frac{u_{i,k} + u_{i,k-1}}{2} \right) a_k \right\} \\ &+ \frac{1}{\left(\widehat{N}-1\right)\widehat{\widehat{Y}}} \times w_i \left(\widehat{\widehat{Y}} - \frac{\widehat{\widehat{Y}}_i + \widehat{\widehat{Y}}_{i-1}}{2} \right).\end{aligned}$$

Through further simplification:

$$\frac{\partial \widehat{B}_t}{\partial a_i} = \frac{u_i - w_i y_i}{\left(\widehat{N}-1\right)\widehat{\overline{Y}}} \times \widehat{B}_t + \frac{1}{\left(\widehat{N}-1\right)\widehat{\overline{Y}}} \times w_i \left(y_i - \frac{\widehat{Y}_i + \widehat{Y}_{i-1}}{2} \right) - \frac{1}{\left(\widehat{N}-1\right)\widehat{\overline{Y}}} \times \sum_{k \in U} w_k \left(\frac{u_{i,k} + u_{i,k-1}}{2} \right) a_k. \tag{27}$$

After replacing u_i and $u_{i,k}$ with Expression (19) and (20) and dividing Expression (27) by w_i, Result 2 is obtained. □

Proof of Result 3

$$\begin{aligned} IF_x(B) &= IF_x \left\{ 1 - \frac{1}{\mu} \int_0^\infty \mu(y) dF(y) \right\} = -IF_x \left\{ \frac{1}{\mu} \int_0^\infty \mu(y) dF(y) \right\} \\ &= -\frac{1}{\mu} \left\{ \mu(x) - \int_0^\infty \mu(y) dF(y) \right\} - \frac{1}{\mu} \int_0^\infty IF_x\{\mu(y)\} dF(y) + \frac{1}{\mu^2} \int_0^\infty \mu(y) dF(y)(x-\mu) \\ &= -\frac{\mu(x)}{\mu} + (1-B) - \frac{1}{\mu} \int_0^\infty \frac{(x-\mu(y))\mathbb{1}(x \le y)}{F(y)} dF(y) + \frac{1}{\mu}(1-B)(x-\mu) \\ &= -\frac{\mu(x)}{\mu} - (B-1) - \frac{x}{\mu} \int_0^\infty \frac{\mathbb{1}(x \le y)}{F(y)} dF(y) + \frac{1}{\mu} \int_0^\infty \frac{\mu(y)\mathbb{1}(x \le y)}{F(y)} dF(y) + \frac{1}{\mu}(1-B)x \\ &= \frac{x-\mu(x)}{\mu} + \frac{x}{\mu} \log F(x) + \frac{1}{\mu} \int_0^\infty \frac{\mu(y)\mathbb{1}(x \le y)}{F(y)} dF(y) - \frac{xB}{\mu}. \end{aligned}$$

Furthermore,

$$\begin{aligned} \int_0^\infty \frac{\mu(y)\mathbb{1}(x \le y)}{F(y)} dF(y) &= \int_x^\infty \frac{\int_0^y t dF(t)}{F^2(y)} dF(y) \\ &= \int_0^\infty \int_{\max(x,t)}^\infty \frac{1}{F^2(y)} dF(y) t dF(t) = \int_0^\infty -\left\{ 1 - \frac{1}{F(\max(x,t))} \right\} t dF(t) \\ &= -\mu + \int_0^\infty \frac{1}{F(\max(x,t))} t dF(t) = -\mu + \int_0^x \frac{t}{F(x)} dF(t) + \int_x^\infty \frac{t}{F(t)} dF(t) \\ &= -\mu + \mu(x) + \int_x^\infty \frac{t}{F(t)} dF(t). \end{aligned}$$

Consequently,

$$\begin{aligned} IF_x(B) &= \frac{x-\mu(x)}{\mu} + \frac{x}{\mu} \log F(x) + \frac{-\mu + \mu(x) + \int_x^\infty \frac{t}{F(t)} dF(t)}{\mu} - \frac{xB}{\mu} \\ &= \frac{x-\mu}{\mu} + \frac{x}{\mu} \log F(x) + \frac{1}{\mu} \int_0^\infty \frac{y\mathbb{1}(x \le y)}{F(y)} dF(y) - \frac{xB}{\mu}. \end{aligned}$$