Exploring and correcting the bias in the estimation of the Gini measure of inequality

Paper published in

Sociological Methods and Research

Thank for your interest in this publication.

Exploring and correcting the bias in the estimation of the Gini measure of inequality

Juan F. Muñoz; Pablo J. Moya-Fernández; Encarnación Álvarez-Verdejo

Abstract

The Gini index is probably the most commonly used indicator to measure inequality. For continuous distributions, the Gini index can be computed using several equivalent formulations. However, this is not the case with discrete distributions, where controversy remains regarding the expression to be used to estimate the Gini index. We attempt to bring a better understanding of the underlying problem by regrouping and classifying the most common estimators of the Gini index proposed in both infinite and finite populations, and focusing on the biases. We use Monte Carlo simulation studies to analyse the bias of the various estimators under a wide range of scenarios. Extremely large biases are observed in heavy-tailed distributions with high Gini indices, and bias corrections are recommended in this situation. We propose the use of some (new and traditional) bootstrap- and jackknife-based strategies to mitigate this bias problem. Results are based on continuous distributions often used in the modelling of income distributions. We describe a simulation-based criterion for deciding when to use bias corrections. Various real data sets are used to illustrate the practical application of the suggested bias corrected procedures.

keywords: Gini index; Income distribution; Bootstrap; Jackknife; Survey sampling

1   Introduction

Lorenz (1905) and Gini (1912) were the first to develop measures of inequality. More than a century after these contributions, inequality analysis remains an active and essential topic in numerous fields. The Gini index, also often referred to as the Gini coefficient, is probably the most commonly used to measure inequality. This indicator ranges between 0 and 1, where 0 indicates perfect equality, and 1 the opposite. Inequality is of special interest in economic studies (Piketty, 2015; Tridico, 2018), and the Gini index is used especially to measure income inequality. Many studies indicate that income inequality has been increasing overall in recent years (Bonacini et al., 2021), and marked differences can be observed across various countries. For instance, results from the *EU-SILC* (European Union Statistics on Income and Living Conditions) survey show that Slovakia has the smallest Gini index estimate (0.21), while Turkey (0.43) has the highest of all the countries in the *EU-SILC*. At a global level, the World Bank indicates that South Africa has the highest Gini index (0.63). These results are associated with countries, but more extreme values are expected at regional level, in subpopulations, small areas, etc.

The aforementioned differences across countries indicate that institutions and policies may have an important role to play in reducing inequality. In fact, reducing inequality is one of the 17 Sustainable Development Goals of the United Nations 2030 Agenda for Sustainable Development. The Gini index is a common statistical tool employed by the 2030 Agenda for measuring inequality (see Szymańska, 2021). However, it should be noted that in order to reduce inequality,

it is crucial to be able to accurately measure this phenomenon without biases and/or errors. Indeed, Efron (1990) argued that a large bias is usually an undesirable aspect of an estimator's performance. The relevance of the Gini index has also been demonstrated by its use to describe inequality in many fields and among different socioeconomic groups, such as length of life or well-being (Wang et al., 2020), educational opportunity (Bulle, 2016), housing prices (Villar and Raya, 2015), gender inequality (Larraz, 2015; Larraz et al., 2019), input- and outcome inequality (Jasso, 2021), and horizontal and vertical inequality (Canelas and Gisselquist, 2019).

There is a formal theoretical definition of the Gini index for continuous distributions, and many equivalent formulations have been proposed in the literature. As discussed by Davidson (2009), there is no disagreement about the definition of the Gini index for continuous distributions, since the various existing expressions provide the same outcome. However, for discrete distributions, many different formulations have been suggested in the extensive literature, and there has been notable controversy surrounding the appropriate version to use in this scenario (see also Langel and Tillé, 2013). For discrete distributions, various expressions of the Gini index are plug-in formulations of theoretical definitions of the Gini index for continuous distributions. Throughout this article, we refer to this value derived from a continuous distribution as the true value of the Gini index, and formulations of the Gini index for discrete distributions are referred to as empirical versions or estimators of the true Gini index. A highly debated topic in the literature is whether or not to use a specific bias corrected estimator, which is denoted as $\hat{G}_n^c$ in this paper. Jasso (1979), Deltas (2003) and Davidson (2009) provide some arguments in favour of $\hat{G}_n^c$.

Statistical techniques can be based on infinite or finite populations. Classical statistical theory assumes that sampled units are independently selected from an infinite population, whereas survey sampling theory (see Särndal et al., 2003) considers that samples are selected from a finite population. Survey sampling has specific features, and this implies that statistical techniques designed for infinite populations must be modified so that they can be used for finite populations. For instance, the usual assumption of independence is not satisfied in finite populations when samples are selected without replacement. Note also that the use of continuous probabilistic distributions to model income and wealth distributions is common practice in many real-world applications. For instance, the Dagum, Pareto, Weibull and Gamma distributions are used, respectively, by Pérez and Alaiz (2011), Atkinson (2017), Bakar and Pathmanathan (2020) and Salem and Mount (1974). The Lognormal distribution is often used to model household income in many countries (see Clementi and Gallegati, 2005).

This paper describes, in Section 2, the most common formulations for calculating the Gini index from both discrete and continuous distributions, and in scenarios of infinite and finite populations. The first aim of this paper is to regroup and classify existing empirical versions of the Gini index, and provide a better overview of the problem of estimating this parameter. The second aim is to analyse, in Section 3, the biases of different versions of the Gini index. For this purpose, we consider a variety of Gini indices and various probabilistic distributions commonly used to model income distributions. Our results reveal that extremely large biases may appear, especially for heavy-tailed distributions and large Gini indices, and bias correction procedures are recommended in this situation. As expected, the bias problem is more serious in small samples, as is the case of rural studies (Wan, 2001), small areas (Frabrizi and Trivisano, 2016), subpopulations (Särndal et al., 2003, p. 386), etc. The third contribution is to describe, in Section 4, bias correction procedures that may reduce the aforementioned large biases. Bootstrap and jackknife methods are considered, and a novel empirical bootstrap is also adapted to the problem

of estimating the Gini index. In Section 5, the bias correction procedures are analysed using Monte Carlo simulation studies. Section 6 describes a simulation-based criterion for deciding when to use bias correction procedures, which are then illustrated, in Section 7, by application to various real data sets. Finally, a brief discussion is presented in Section 8. The supplementary material contains: (i) the selected parameters of the analysed probabilistic distributions; (ii) results from simulation studies based on large samples ($n = 500$); (iii) description of the bias functions suggested in Section 4, and information on their percentages of use; and (iv) efficiency and bias ratios of estimators of the Gini index, which are explored in Sections 5 and 6.

2   The Gini index

2.1   Definition

We assume that inequality is analysed using a variable of interest Y, which is a nonnegative continuous random variable. A popular formulation of the Gini index is defined in terms of the average absolute difference between each possible pair of individuals (Qin et al., 2010), i.e.,

$$G = \frac{1}{2\mu_Y} \int_0^{+\infty} \int_0^{+\infty} |x - y| \, dF_Y(x) dF_Y(y), \tag{1}$$

where

$$\mu_Y = E[Y] = \int_0^{+\infty} y f(y) dy = \int_0^{+\infty} y \, dF_Y(y),$$

is the mean of Y, and $F_Y(y) = P(Y \le y)$ and $f(y)$ are, respectively, the distribution function and the probability density function of Y. A formulation of G based on the distribution function is (Qin et al., 2010; Berger and Gedik-Balay, 2020):

$$G = \frac{1}{\mu_Y} \int_0^{+\infty} \{2F_Y(y) - 1\} y dF_Y(y). \tag{2}$$

Anand (1983) showed that the Gini index G can be computed as $2/\mu_Y$ times the covariance between Y and the distribution function $F_Y(y)$, i.e.,

$$G = \frac{2}{\mu_Y} cov\{Y, F_Y(y)\}. \tag{3}$$

Finally, Yitzhaki (1998) and Berger and Gedik-Balay (2020) consider the expression

$$G = 1 - \frac{\mu_Z}{\mu_Y},$$

where $\mu_Z = E(Z) = \int_0^{+\infty} \{1 - F_Z(z)\} dz$ is the expectation of the minimum $Z = \min\{Y_a, Y_b\}$, and Y_a and Y_b are two independent random variables with the same distribution as Y. For continuous

3

distributions, the Gini index can be defined in many other ways, as can be seen in Yitzhaki (1998), Giorgi and Gibliarano (2017), etc. In practice, the value of G is estimated by means of a sample S, with size n, and which can be selected from either infinite or finite populations (Langel and Tillé, 2013). The estimation of G under both scenarios is discussed in Section 2.2.

2.2 Estimation

For infinite populations, $\{Y_i : i \in S\}$ are considered as a sequence, with size n, of nonnegative random variables with the same distribution as the variable of interest Y. The Gini index is estimated using an estimator of G based on the observations of individuals selected in the sample S, and which are denoted as $\{y_i : i \in S\}$. Such estimators are usually defined as plug-in formulations derived from a theoretical definition of G. This methodology may introduce a bias in comparison to the true parameter G, especially for extreme values of the Gini index. As can be seen in Section 3, a notable example is the plug-in expression of Equation (2), which is defined as (see Qin et al., 2010; Berger and Gedik-Balay, 2020):

$$\hat{G}_n^a = \frac{1}{n\bar{y}} \sum_{i \in S} \left[\{2\hat{F}_n(y_i) - 1\} y_i \right] = \frac{2}{n\bar{y}} \sum_{i \in S} y_i \, \hat{F}_n(y_i) - 1, \tag{4}$$

where $\bar{y} = n^{-1} \sum_{i \in S} y_i$ is the sample mean, $\hat{F}_n(t) = n^{-1} \sum_{i \in S} \delta(y_i \leq t)$ is the sample (empirical) distribution function, and $\delta(\cdot)$ is the indicator variable that takes the value 1 if its argument is true and 0 otherwise. The classical empirical version of G (Giorgi and Gigliarano, 2017) is the plug-in expression of Equation (1), i.e.:

$$\hat{G}_n^b = \frac{1}{2n^2\bar{y}} \sum_{i \in S} \sum_{j \in S} |y_i - y_j|. \tag{5}$$

Note that many equivalent versions of $\hat{G}_n^b$ have been suggested in the extensive literature on the Gini index. For instance, Sen (1973) proposed the popular formulation

$$\hat{G}_n^{b1} = \frac{2}{n^2\bar{y}} \sum_{i \in S} i y_{(i)} - \frac{n+1}{n} = \frac{2}{n^2\bar{y}} \sum_{i \in S} r_i \, y_i - \frac{n+1}{n},$$

where $y_{(i)}$ are the values y_i sorted in increasing order and r_i is the rank of unit i in the sample S. Similarly, the Gini index can be defined using the regression coefficient of an ordinary least squares regression (see Ogwang, 2000). This is the idea behind

$$\hat{G}_n^{b2} = \frac{2\hat{\beta}}{n} - \frac{n+1}{n},$$

which assumes the regression model $i = \beta + u_i$, and where the heterocesdatic error u_i has variance $\sigma^2 / y_{(i)}$. The least squares estimator of β is given by

$$\hat{\beta} = \frac{\sum_{i \in S} i \, y_{(i)}}{\sum_{i \in S} y_{(i)}}. \tag{6}$$

Finally, an equivalent version of $\hat{G}_n^b$ is the empirical version of Equation (3), i.e.,

$$\hat{G}_n^{b3} = \frac{2}{n\bar{y}} cov(i, y_{(i)}),$$

where

$$cov(i, y_{(i)}) = \frac{1}{n} \sum_{i \in S} i\, y_{(i)} - \frac{n+1}{2} \bar{y}.$$

The estimator $\hat{G}_n^b$ and its equivalent versions satisfy the symmetry axiom of Sen (1973), which establishes that an estimator of G based on a set of observations, say $\{y_i : i \in S\}$, must coincide with the Gini index estimated by means of the same approach but using the sample S exactly replicated, i.e., doubled in size (see Davidson, 2009). Alternatively, the bias corrected estimator

$$\hat{G}_n^c = \frac{n}{n-1} \hat{G}_n^b, \tag{7}$$

is often used instead of $\hat{G}_n^b$. Some equivalent expressions of $\hat{G}_n^c$ are:

$$\hat{G}_n^{c1} = \frac{2}{n(n-1)\bar{y}} \sum_{i \in S} i\, y_{(i)} - \frac{n+1}{n-1};$$

$$\hat{G}_n^{c2} = \frac{1}{2\bar{y}} \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} |y_{i_1} - y_{i_2}|;$$

$$\hat{G}_n^{c3} = 1 - \frac{\bar{z}}{\bar{y}}.$$

Jasso (1979) suggested the use of $\hat{G}_n^{c1}$, Wang et al. (2016) consider $\hat{G}_n^{c2}$, and $\hat{G}_n^{c3}$ is used by Berger and Gedik-Balay (2020), where $\bar{z} = n^{-1} \sum_{i \in S} z_{i.n}$ and

$$z_{i.n} = \frac{1}{n-1} \sum_{\substack{j \in S, \\ j \neq i}} \min(y_i, y_j).$$

Giles (2004) and Davidson (2009) provide theoretical justifications for the use of $\hat{G}_n^c$ to reduce the bias of $\hat{G}_n^b$. As can be seen in Section 3, $\hat{G}_n^a$ may result in serious biases for small Gini indices, but this problem can be easily solved by replacing $\hat{F}_n(t)$ in Equation (4) with the smooth (or midpoint) distribution function $\hat{F}_n^*(t) = n^{-1} \sum_{i \in S} [\delta(y_i < t) + 0.5\delta(y_i = t)]$, and the resulting estimator coincides with $\hat{G}_n^b$ (see Berger, 2008). In addition, $\hat{G}_n^a$, $\hat{G}_n^b$ and $\hat{G}_n^c$ are related when $y_i \neq y_j$ for all $i \neq j$, since

$$\hat{G}_n^b = \hat{G}_n^a - \frac{1}{n}, \tag{8}$$

5

if this condition is satisfied, and

$$\hat{G}_n^c = \frac{n}{n-1}\left(\hat{G}_n^a - \frac{1}{n}\right) \tag{9}$$

according to Equations (7) and (8). Note that expressions (5) and (7), or their equivalent formulations, are more frequently used in practice, and practitioners must be aware of the bias of $\hat{G}_n^a$ when G is small. The use of the smooth distribution function in Equation (4) will prevent this bias problem. For empirical distributions, additional formulations of the Gini index can be seen in Giorgi and Gigliarano (2017).

For a finite population U with N individuals, $\{Y_i : i \in U\}$ denotes a sequence of nonnegative random variables with the same distribution function $F_Y(y)$, and $\{y_i : i \in U\}$ are the population values of the variable of interest. In practice, social surveys are used to estimate the Gini index, and they are generally based on complex sampling designs with unequal probabilities. Therefore, the sample S is now selected from U by using a sampling design with survey weights $w_i = \pi_i^{-1}$, where $\pi_i = P(i \in S)$ are the inclusion probabilities, with $i \in S$. The problem of estimating G from finite populations thus entails two steps. First, an empirical version of G based on the population values $\{y_i : i \in U\}$ is required. We denote the population empirical versions of G in finite populations as G_N^a, G_N^b and G_N^c, and they are defined as $\hat{G}_n^a$, $\hat{G}_n^b$ and $\hat{G}_n^c$, respectively, after substituting the sample values with the population values in Equations (4), (5) and (7). The second step is to estimate the selected population empirical version (G_N^a, G_N^b or G_N^c) using weighted estimators. Some that can be found in the literature are:

$$\hat{G}_w^a = \frac{2}{\hat{N}\overline{y}_w}\sum_{i \in S} w_i y_i \hat{F}_w(y_i) - 1; \tag{10}$$

$$\hat{G}_w^b = \frac{1}{2\hat{N}^2\overline{y}_w}\sum_{i \in S}\sum_{j \in S} w_i w_j |y_i - y_j|; \tag{11}$$

and

$$\hat{G}_w^c = 1 - \frac{\overline{z}_w}{\overline{y}_w}, \tag{12}$$

where $\hat{N} = \sum_{i \in S} w_i$, $\overline{y}_w = \hat{N}^{-1}\sum_{i \in S} w_i y_i$, $\overline{z}_w = \hat{N}^{-1}\sum_{i \in S} w_i z_{i.w}$,

$$z_{i.w} = \frac{1}{\hat{N} - w_i}\sum_{\substack{j \in S, \\ j \neq i}} w_j \min(y_i, y_j),$$

and $\hat{F}_w(t) = \hat{N}^{-1}\sum_{i \in S} w_i \delta(y_i \leq t)$. Note that Equations (10), (11) and (12) reduce, respectively, to Equations (4), (5) and (7) under simple random sampling without replacement (SRSWOR).

3 Simulation studies to analyse the bias

In this section, we analyse the bias of $\hat{G}_n^a$, $\hat{G}_n^b$ and $\hat{G}_n^c$ in comparison to the true (asymptotic) value G, and using samples selected from infinite populations. This analysis is equivalent to the problem of analysing the bias of G_N^a, G_N^b and G_N^c in comparison to the true (asymptotic) value G.

3.1 Description

We consider various continuous probabilistic distributions (Pareto, Dagum, Lognormal, Weibull and Gamma) often used in the modelling of income distributions. For each probabilistic distribution, parameters involved in the theoretical formulation of G are selected such that G takes the values $\{0.1, 0.2, \dots, 0.8\}$, thus allowing us to examine different levels of inequality. Additional parameters required in distributions are also fixed, and all of them can be seen in the supplementary material (Table A1). For the Dagum distribution, the theoretical value of G depends on both shape parameters a and p, and for this reason the values $p = \{0.5, 20\}$ are also fixed, and such distributions are denoted, respectively, as Dagum-p0.5 and Dagum-p20. The aim is to analyse the biases of the various estimators of the Gini index under the described scenarios, with these estimators being calculated using samples randomly drawn from an underlying continuous distribution with a true value G for the Gini index. This framework is also adopted by Deltas (2003), Davidson (2009), Berger and Gedik-Balay (2020), etc. We analyse both small and large sample sizes, specifically, $n = \{50, 500\}$. This study is equivalent to analysing the biases for samples, with size n, selected under SRSWOR from a large finite population ($N \to \infty$), with population values drawn from the analysed probabilistic distributions.

Let $\hat{\theta}_n$ be a given statistic for the unknown parameter θ, based on the observations $\{y_i : i \in S\}$. Throughout this article, the expected value based on R replications of $\hat{\theta}_n$ is defined as

$$E_R[\hat{\theta}_n] = \bar{\theta}_n^R = \frac{1}{R} \sum_{r=1}^{R} \hat{\theta}_n^{(r)}, \tag{13}$$

where $\hat{\theta}_n^{(r)}$ is the statistic $\hat{\theta}_n$ evaluated at the r-th pseudo original sample $S^{(r)}$, which is also selected, with size n, from the distribution function $F_Y(y)$. $R = 1000$ replications are considered in simulation studies. The empirical measures can be expressed in terms of either the true Gini index G or the expected values of estimators. We use the expected values because large biases can be obtained and G is unknown in practice. Reporting the results in terms of G makes it more difficult for empirical researchers to assess the performance of estimators for the specific data that they are analysing. Finally, note that various figures in this paper require only a customary estimator of G, and we use $\hat{G}_n^c$ because it is less biased than its competitors ($\hat{G}_n^a$ and $\hat{G}_n^b$).

In this section, we first use Monte Carlo simulations to investigate the relative bias (RB) of the various empirical versions ($\hat{G}_n^a$, $\hat{G}_n^b$ and $\hat{G}_n^c$) in comparison to the true (asymptotic) value G. For a given statistic $\hat{\theta}_n$, this measure is defined as

$$RB = 100 \times \frac{B_R[\hat{\theta}_n]}{\theta},$$

where the empirical bias is given by $B_R[\hat{\theta}_n] = E_R[\hat{\theta}_n] - \theta = \bar{\theta}_n^R - \theta$. Comparisons are based on distributions with different levels of skewness because the value of the coefficient of skewness may have an impact on the bias of estimators of the Gini index. For discrete distributions, the coefficient of skewness is defined as:

$$\hat{\gamma}_n = \frac{\mu_{n.3}}{\sigma_n^3}, \tag{14}$$

where $\sigma_n = (\mu_{n.2})^{1/2}$ is the sample standard deviation, and $\mu_{n.\alpha} = n^{-1} \sum_{i \in S} (y_i - \overline{y})^\alpha$ is the α-th central moment based on S. The aim of Figure 1 is to investigate the skewness for the probabilistic distributions considered in this paper, so this figure displays the expected values $\bar{\gamma}_n^R$ versus the expected values $\bar{G}_n^{c.R}$, where $\bar{\gamma}_n^R$ and $\bar{G}_n^{c.R}$ are calculated using Equation (13) after substituting $\hat{\theta}_n^{(r)}$ with $\hat{\gamma}_n^{(r)}$ and $\hat{G}_n^{c(r)}$, respectively, and which are computed using Equations (14) and (7) at the r-th pseudo original sample $S^{(r)}$.
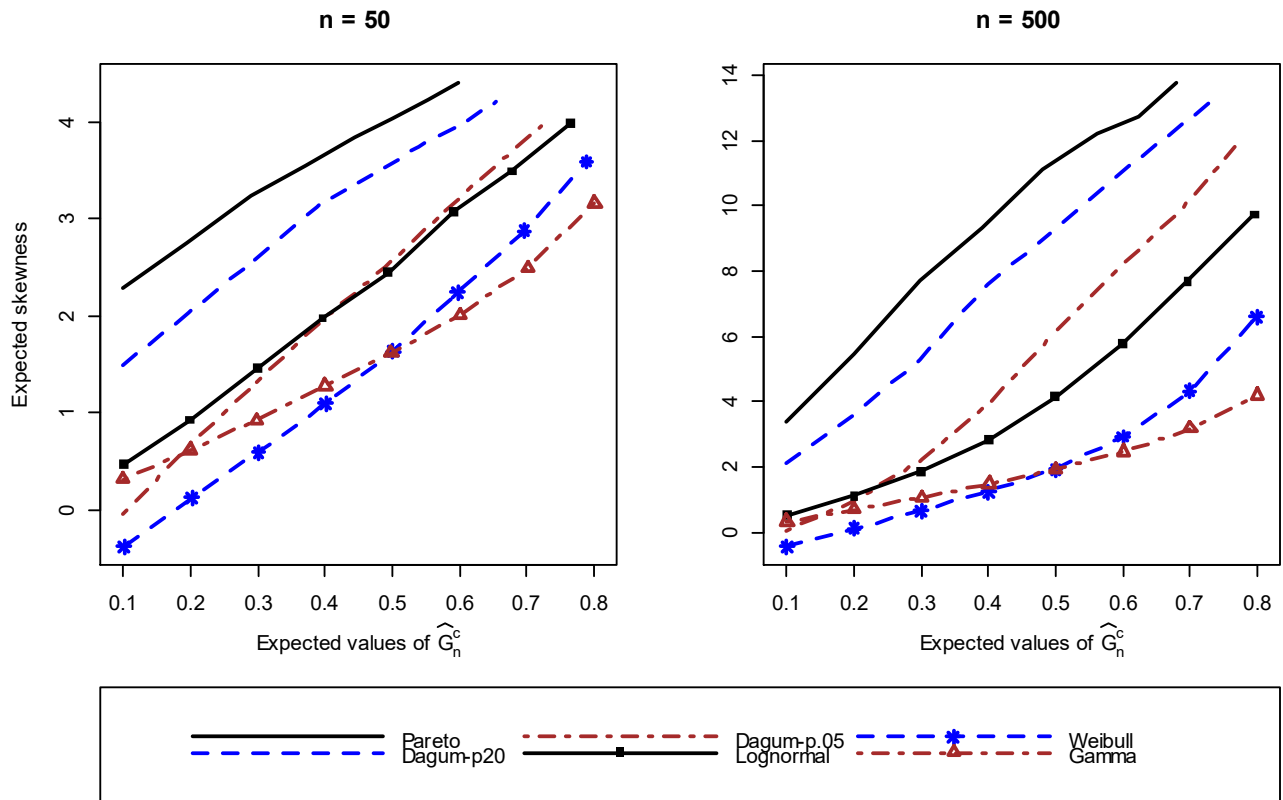

Figure 1: *Expected values of the coefficient of skewness ($\bar{\gamma}_n^R$) based on samples with sizes $n = \{50,500\}$, and randomly selected from various continuous probabilistic distributions (infinite populations). The x-axes show the expected values of the estimator $\hat{G}_n^c$ ($\bar{G}_n^{c.R}$).*

From Figure 1 we observe that the Pareto distribution is the most highly skewed distribution, followed by the Dagum-p20, Dagum-p0.5 and Lognormal distributions, in that order. The Weibull and Gamma distributions have similar values of $\bar{\gamma}_n^R$, and they are the least skewed distributions in

8

this study. For highly skewed distributions, serious biases can be observed when G is large, as a result of which the maximum value of $\bar{G}_n^{c.R}$ is far from $G = 0.8$, the maximum Gini index used in this study. This is not the case with less skewed distributions, since accurate estimates are also obtained when G is large, and the expected values $\bar{G}_n^{c.R}$ are close to the required Gini index. For the various probabilistic distributions, the expected skewness increases as the sample size rises. This can be explained by the upper bound for $\hat{\gamma}_n$ based on the sample size and suggested by Cramer (1957). This bound indicates that estimates for the coefficient of skewness may underestimate the true value when the sample size is small (Dorić et al., 2009). For the Dagum distribution, the expected skewness increases as its shape parameter p increases.

In this section we also analyse the impact of the skewness on the bias using box plots for estimates of $\hat{G}_n^c$. Thus, we illustrate this relationship between skewness and bias by comparing two different distributions in terms of skewness (Pareto and Gamma, as can be seen in Figure 1).

3.2 Results and conclusions

Figure 2 displays the RBs of $\hat{G}_n^a$, $\hat{G}_n^b$ and $\hat{G}_n^c$ when $n = 50$. First, we analyse the results from the less skewed distributions (Weibull and Gamma). The bias of $\hat{G}_n^c$ is negligible for the various expected values of estimators. The bias of $\hat{G}_n^b$ is slightly larger, in absolute terms, than that of $\hat{G}_n^c$, but lies within a reasonable range. Biases of both $\hat{G}_n^b$ and $\hat{G}_n^c$ do not seem to be affected by the value of the Gini index. $\hat{G}_n^a$ is severely biased when the expected values of estimators are small, with values of RB that can be close to 20%. This empirical version must be modified to correct this bias, and two simple solutions are discussed in Section 2.2. First, the distribution function $\hat{F}_n(t)$ can be replaced, in Equation (4), by the smooth distribution function $\hat{F}_n^*(t)$. This adjustment allows empirical versions $\hat{G}_n^a$ and $\hat{G}_n^b$ to be equivalent. Second, we can use one of the transformations described in Equations (8) and (9) when all observations are different. As the Gini index increases, the RB of $\hat{G}_n^a$ decreases and $\hat{G}_n^a$ and $\hat{G}_n^c$ have similar RBs.

For heavy-tailed distributions (Pareto, Dagum-p20, Dagum-p0.5 and Lognormal), biases of $\hat{G}_n^a$, $\hat{G}_n^b$ and $\hat{G}_n^c$ seem to be affected by the value of the Gini index, and serious negative RBs are obtained as the expected values of estimators increase (as much as -25%). We also observe a strong relationship between the RB and the coefficient of skewness. The largest RBs, in absolute terms, are produced by the Pareto distribution, which is the most skewed (see Figure 1), and biases, in absolute terms, decrease as the values of $\bar{\gamma}_n^R$ decrease.

For larger sample sizes, readers are referred to the supplementary material, where Figure A1 replicates Figure 2 for samples with size n=500. We point out that the RB, in absolute terms, decreases as the sample size increases. For less skewed distributions, the bias of $\hat{G}_n^b$ is negligible, and the RB of $\hat{G}_n^a$ can be close to 2% for the various distributions. Non-negligible biases are also observed for heavy-tailed distributions, with values of RB close to -15% when $n = 500$.
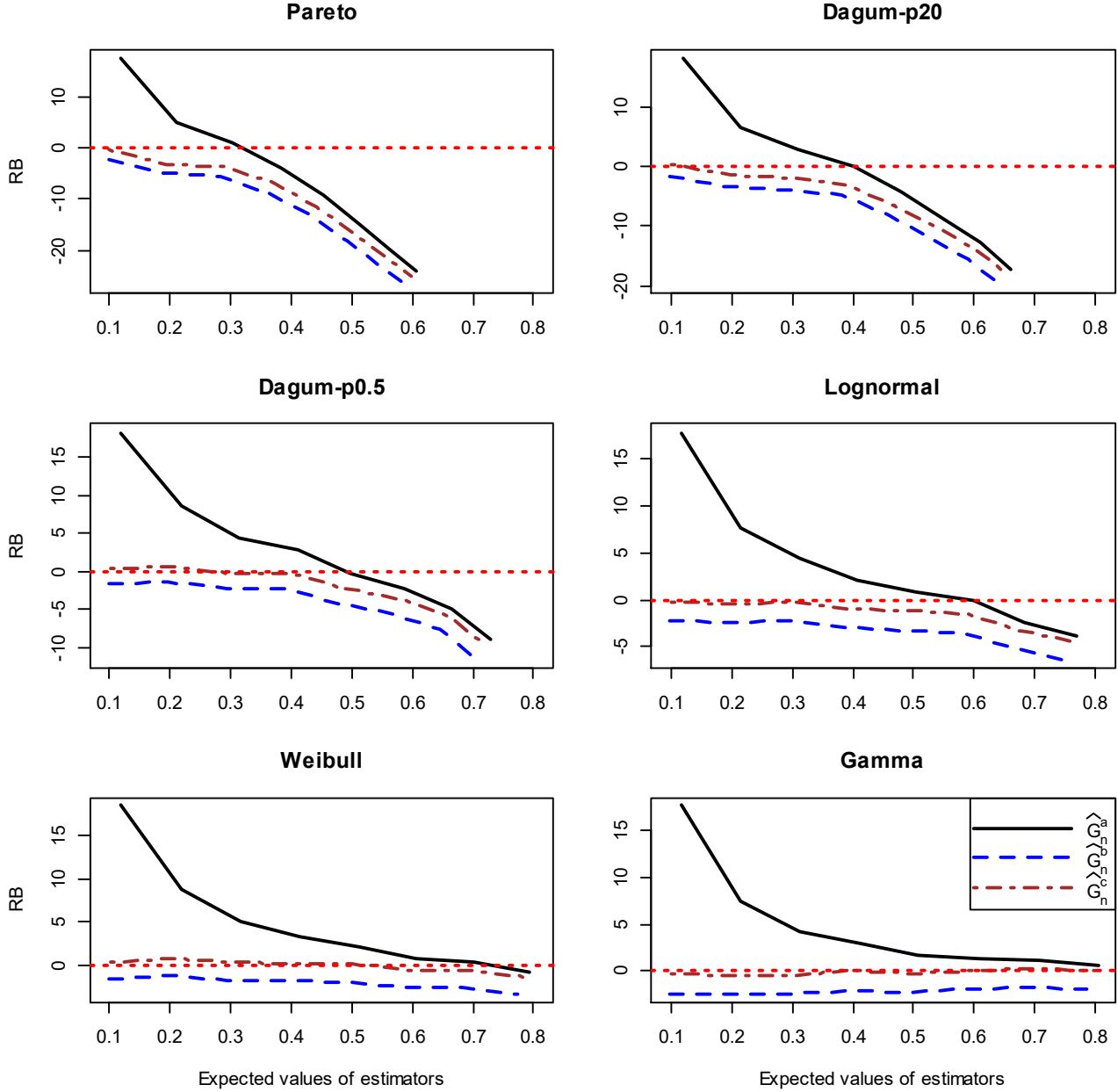
Figure 2: *Relative biases (RBs) of estimators $\hat{G}_n^a$, $\hat{G}_n^b$ and $\hat{G}_n^c$ based on samples with size $n = 50$, and randomly selected from various continuous probabilistic distributions (infinite populations).*

In Figure 3 we investigate the effect of the skewness on the bias of $\hat{G}_n^c$ using box plots and various Gini indices, with the most (Pareto) and the least (Gamma) skewed distributions from this study. From Figure 2 we observe that the bias of $\hat{G}_n^c$ is negligible for the various Gini indices when samples are selected from the Gamma distribution, while Figure 3 confirms that the estimates are concentrated, with a low variability, around the target value G. This is not the case with the Pareto distribution, which shows highly biased estimates and marked variability. From Figures 2 and 3 we observe that the bias of $\hat{G}_n^c$, in absolute terms, increases as the Gini index rises, while from Figure 3 we see that the variability of estimates also becomes higher as G increases, with values of $\hat{G}_n^c$ that can be larger than 0.9 when $G = 0.4$, or smaller than 0.3 when $G = 0.8$.
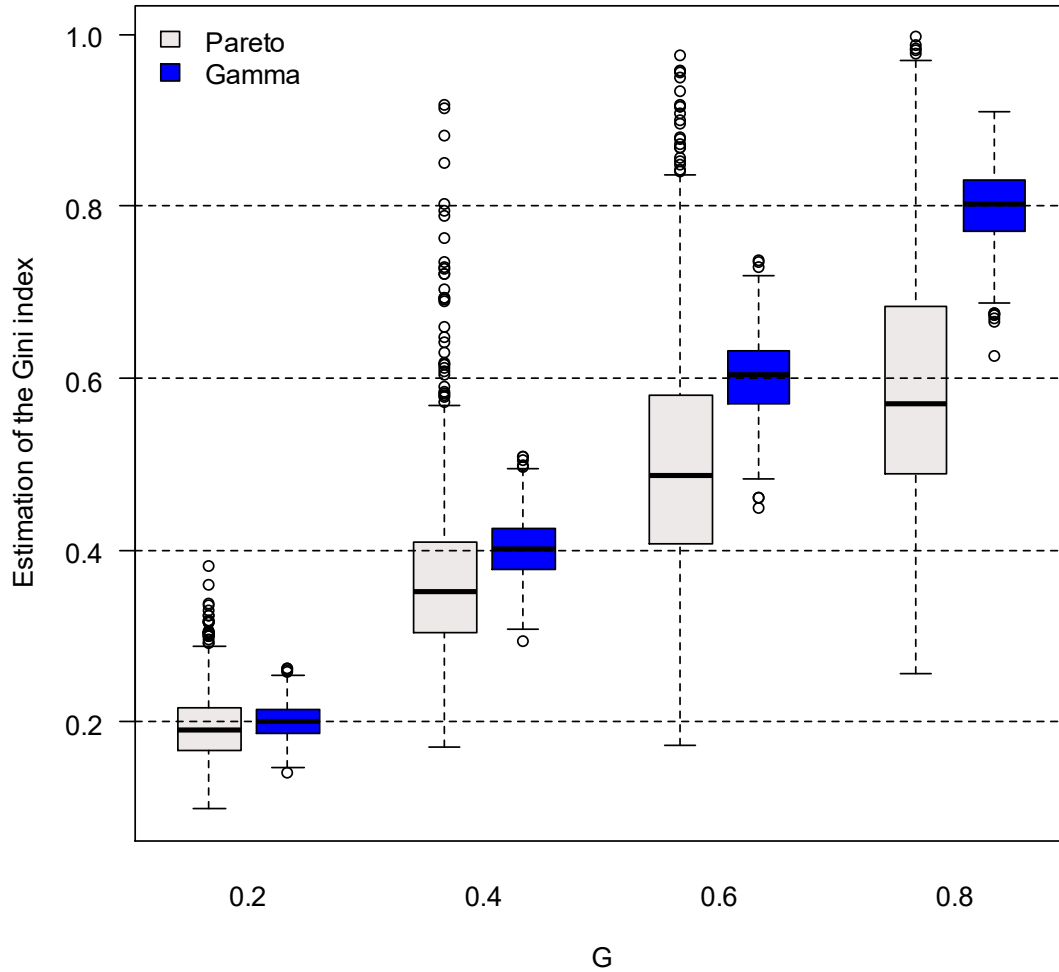
Figure 3: *Box plots for 1000 estimates of the Gini index (G) using the estimator $\hat{G}_n^c$ and various values of G. Samples, with size $n = 50$, are randomly selected from the Pareto and Gamma distributions (infinite populations).*

4 Bias correction procedures

Results from Section 3 indicate that the customary estimators of the Gini index can be severely biased, especially for heavy-tailed distributions and large Gini indices, and that the use of a bias correction procedure may alleviate this bias problem. Various bias correction procedures are presented in this section, before being analysed in Section 5, and then applied to various real data sets in Section 7. Section 6 describes a criterion for deciding when to use bias corrections.

Bootstrap and jackknife techniques (see Efron and Tibshirani, 1993, and Wolter, 2007) can be used as bias correction procedures. Some authors who have demonstrated the capacity of these methods to correct biases are Pfeffermann and Correa (2012) and Jiao and Han (2020). When it comes to the problem of estimating the Gini index, such statistical techniques have been used mainly for the construction of confidence intervals and variance estimation (see Moran, 2006, and Larraz et al., 2020, for bootstrap techniques, and Berger, 2008, and Davidson, 2009, for jackknife techniques). The estimator $\hat{G}_n^c$ emerges as the bias corrected version of $\hat{G}_n^b$ (Deltas,

2003; Davidson, 2009), but Section 3 shows that $\hat{G}_n^c$ can also be severely biased. Van Ourti and Clarke (2011) investigated a bias correction method for the Gini index, but focussing on the bias due to grouped data. We now explore correction procedures for the biases discussed in Section 3. Bias corrections are applied to $\hat{G}_n^c$ and $\hat{G}_w^c$ (defined, respectively, for infinite and finite populations) because they are less biased than the alternative empirical versions described in Section 2. However, bias correction procedures can also be applied to any other empirical version.

For an infinite population, we first suggest the jackknife technique proposed by Ogwang (2000), which can be easily implemented by a fast algorithm. Langel and Tillé (2013) showed that this method has desirable properties for the variance estimation of the Gini index. Ogwang (2000) proposed the application of the jackknife technique on $\hat{G}_n^{b2}$, with jackknife estimates defined as

$$\hat{G}_n^{b2}(k) = \hat{G}_n^{b2} + \frac{2}{n\overline{y} - y_{(k)}} \left\{ \frac{y_{(k)}\hat{\beta}}{n} + \frac{\sum_{i=1}^{n} i\, y_{(i)}}{n(n-1)} - \frac{n\overline{y} - \sum_{i=1}^{k} y_{(i)} + ky_{(k)}}{n-1} \right\} - \frac{1}{n(n-1)},$$

where $\hat{\beta}$ is the regression coefficient defined by Equation (6). Note that $\hat{G}_n^{b2}(k)$ is equivalent to applying $\hat{G}_n^{b2}$ successively to the observations $\{y_{(i)} : i \in S\}$ and after removing the $k-$th unit. The bias corrected estimator applied to $\hat{G}_n^{b2}$ and based on Ogwang's jackknife is defined as

$$\hat{G}_n^{b2.Jo} = n\hat{G}_n^{b2} - (n-1)\bar{G}_n^{b2.J},$$

where $\bar{G}_n^{b2.J} = n^{-1}\sum_{k=1}^{n} \hat{G}_n^{b2}(k)$. The bias corrected estimator applied to $\hat{G}_n^c$ and based on jackknife is given by

$$\hat{G}_n^{c.Jo} = n\hat{G}_n^c - (n-1)\bar{G}_n^{c.J}, \tag{15}$$

where $\bar{G}_n^{c.J} = n^{-1}\sum_{k=1}^{n} \hat{G}_n^c(k)$, and $\hat{G}_n^c(k)$ is the estimator $\hat{G}_n^c$ computed from the observations $\{y_{(i)} : i \in S\}$ after removing the $k-$th unit. Note that $\hat{G}_n^{c.Jo}$ is one of the two bias corrected estimators that we report in the results from infinite populations. Pfeffermann and Correa (2012) proposed an empirical bootstrap bias correction procedure based on pseudo original and bootstrap samples selected from plausible parameters. This method was used to estimate the prediction mean square error in small area estimation of proportions. We also propose the adaption of this empirical bootstrap method to the problem of estimating the bias of $\hat{G}_n^c$, thus giving rise to a novel bias corrected estimator of G.

The empirical bootstrap procedure considers a set of plausible parameters, which are randomly generated from a confidence interval for G. For each plausible parameter, a pseudo original sample is generated from the underlying distribution of the original sample data. This method uses a cross-validation procedure that splits the various pseudo original samples into two groups: training and validation. In Section 3 we observed that both the Gini index and the coefficient of skewness have an impact on the bias of $\hat{G}_n^c$ for heavy-tailed distributions. For the training group, we suggest various functions underlying the bias correction, which depend on the estimates of $\hat{G}_n^c$ and $\hat{\gamma}_n$ computed from each pseudo original sample and on the expected values of $\hat{G}_n^c$ and $\hat{\gamma}_n$ based on bootstrap samples. Efron and Tibshirani (1993) and Hall and Maiti (2006) indicate that bias corrections may increase the variance, so the validation group is used to choose the optimum bias function that minimizes the mean square error (*MSE*) of the suggested bias

12

corrected estimator. In Section 5, we also investigate the impact of using bias corrections on the *MSE*. For an infinite population, the algorithm for estimating the bias of $\hat{G}_n^c$ and for computing the suggested bias corrected estimator is described in detail as follows:

Step 1 (Plausible parameters). Select at random H plausible values for the target parameter G from a Uniform distribution, i.e., $G_h \sim Un(G_L, G_U)$, with $h = 1, \dots, H$, and where G_L and G_U are, respectively, the lower and upper limits of a confidence interval for the true Gini index G.

Step 2 (Pseudo original samples for training and validation groups). Generate a pseudo original sample S_h, with size n, from $f(y; G_h)$ and for each $h = 1, \dots, H$, where $f(y; G_h)$ is the probability density function $f(y)$ with a Gini index equal to G_h. Then, split the H samples at random into two groups, the training group Ω_T and the validation group Ω_V, such that Ω_T contains a set of T samples (S_t, with $t = 1, \dots, T$), Ω_V contains V samples (S_v, with $v = 1, \dots, V$) and $H = T + V$.

Step 3 (Estimates from the pseudo original samples). For the training group Ω_T, compute $\hat{G}_{nt}^c$ and $\hat{\gamma}_{nt}$ for each sample S_t, and using, respectively, Equations (7) and (14). Similarly, for the validation group Ω_V, compute $\hat{G}_{nv}^c$ and $\hat{\gamma}_{nv}$ using the samples S_v.

Step 4 (Training phase: expected values based on bootstrap samples). For the training group Ω_T, generate B bootstrap samples $S_t^{(b)}$, with size n, from each sample S_t, with $b = 1, \dots, B$. Compute $\hat{G}_{nt}^{c(b)}$ and $\hat{\gamma}_{nt}^{(b)}$ for each bootstrap sample $S_t^{(b)}$, using Equations (7) and (14), respectively. The expected values of $\hat{G}_n^c$ and $\hat{\gamma}_n$ based on bootstrap samples are denoted, respectively, as $\bar{G}_{nt}^{c.B}$ and $\bar{\gamma}_{nt}^B$, and are computed using $\hat{G}_{nt}^{c(b)}$ and $\hat{\gamma}_{nt}^{(b)}$ in Equation (13), after substituting r and R with b and B, respectively.

Step 5 (Training phase: expected values based on pseudo original samples). For the training group Ω_T, generate R pseudo original samples $S_t^{(r)}$, with size n, from $f(y; G_t)$ and for each $t = 1, \dots, T$, with $r = 1, \dots, R$. Compute $\hat{G}_{nt}^{c(r)}$ for each sample $S_t^{(r)}$, and using Equation (7). The expected value based on pseudo original samples is denoted as $\bar{G}_{nt}^{c.R}$, and is computed using Equation (13).

Step 6 (Training phase: coefficient estimates). For the training group Ω_T, estimate the unknown coefficients of a set of eligible bias functions $q_l(\hat{G}_{nt}^c; \bar{G}_{nt}^{c.B}; \hat{\gamma}_{nt}; \bar{\gamma}_{nt}^B)$, with $l = 1, \dots, L$, that predict the variable $D_t = \bar{G}_{nt}^{c.R} - G_t$. An example of bias function is the linear expression

$$\bar{G}_{nt}^{c.R} - G_t = a_0 + a_1(\bar{\gamma}_{nt}^B - \hat{\gamma}_{nt}). \tag{16}$$

Step 7 (Validation phase: bias corrected estimators). For the validation group Ω_V and for each function q_l, compute the suggested bias corrected estimator of G_v, defined by

$$\hat{G}_{nv}^{c.Bp}(l) = \hat{G}_{nv}^c - \hat{q}_l(\hat{G}_{nv}^c; \bar{G}_{nv}^{c.B}; \hat{\gamma}_{nv}; \bar{\gamma}_{nv}^B),$$

where $\hat{q}_l$ is the bias function q_l after substituting its coefficients with the estimates computed in Step 6.

Step 8 (Validation phase: optimum function). For the validation group Ω_V, identify the optimum function $\hat{q}_{opt}$ that minimizes the *MSE* of the estimators $\hat{G}_{nv}^{c.Bp}(l)$, and which is defined as

13

$$MSE_l = \frac{1}{V} \sum_{v=1}^{V} \left(\hat{G}_{nv}^{c.Bp}(l) - G_v \right)^2.$$

Step 9 (Bias corrected estimator). Compute the suggested bias corrected estimator

$$\hat{G}_n^{c.Bp} = \hat{G}_n^c - \hat{q}_{opt}\left(\hat{G}_n^c; \bar{G}_n^{c.B}; \hat{\gamma}_n; \bar{\gamma}_n^B \right), \tag{17}$$

where $\bar{G}_n^{c.B}$ and $\bar{\gamma}_n^B$ are, respectively, the expected values of $\hat{G}_n^c$ and $\hat{\gamma}_n$ based on bootstrap samples and derived from the original sample S. □

$\hat{G}_n^{c.Bp}$ is the second bias corrected estimator computed for infinite populations. Any method can be used to construct the confidence interval required in Step 1. Some existing confidence intervals for the Gini index are based on bootstrap (Qin et al., 2010), jackknife (Berger, 2008), linearization (Deville, 1999) or empirical likelihood (Berger and Gedik-Balay, 2020). Variance estimators for the Gini index (Langel and Tillé, 2013) can also be used to construct confidence intervals based on the normality assumption. For the sake of simplicity, we use the traditional bootstrap method with confidence interval limits given by $G_L = \hat{G}_{n(0.005)}^{c.B}$ and $G_U = \hat{G}_{n(0.995)}^{c.B}$, where $\hat{G}_{n(\alpha)}^{c.B}$ is the α-th quantile of the bootstrap estimates $\hat{G}_n^{c(b)}$. The latter are computed using the estimator $\hat{G}_n^c$ on the bootstrap sample $S^{(b)}$, which is taken, with size n, from the original sample S. As discussed by Pfeffermann and Correa (2012), the selected interval must be broad enough to contain the target parameter G. We use a confidence level of 99% because of the serious biases detected in Section 3. Pfeffermann and Correa (2012) also argue that the size of this confidence interval has no direct effect on the bound of the bias, and give a discussion on the number of parameters that should be included in the training and validation groups.

For the training group, Step 6 requires bias functions q_l that predict $D_t = \bar{G}_{nt}^{c.R} - G_t$ with the aim of estimating the bias of $\hat{G}_n^c$, i.e., $B[\hat{G}_n^c] = E[\hat{G}_n^c] - G$. In Section 3, we concluded that both the Gini index and the coefficient of skewness may have an impact on the RB of $\hat{G}_n^c$, so we suggest bias functions that depend on $\hat{G}_{nt}^c$ and $\hat{\gamma}_{nt}$. The expected values $\bar{G}_{nt}^{c.B}$ and $\bar{\gamma}_{nt}^B$ based on bootstrap samples are also considered. Table A2 from the supplementary material describes the $L = 7$ candidate bias functions considered in Step 6 of the suggested algorithm. For the sake of simplicity, we only consider multiple linear regression functions, but more complex functions and/or additional statistics can also be used, and are expected to yield more accurate results.

For an infinite population, additional bias correction procedures can be computed (See Wolter, 2007). For instance, we also calculated, in Section 5, bootstrap methods based on additive and multiplicative corrections (see Hall and Maiti, 2006, and Pfeffermann and Correa, 2012, for detailed definitions), but we omitted them because the bias correction estimators (15) and (17) are less biased. Pfeffermann and Correa (2012) also argue that the aforementioned additive and multiplicative corrections may yield non-negligible biases with small samples, meaning alternative bias correction procedures may be preferable.

Bootstrap and jackknife techniques were originally designed for infinite populations, and do not have a direct application to finite populations due to the inherent features of survey sampling. Adjustments are thus required to apply these methods to finite populations (Quatember, 2015). For finite populations, the rescaled bootstrap technique (Rao et al., 1992) can be used for bias

correction of a given empirical version of G. This method has been used in many research studies (Berger and Muñoz, 2015; Moya et al., 2020; etc.) in many areas (see Yang et al., 2010; Muñoz et al., 2018; etc.). Simplicity is the main advantage of the rescaled bootstrap over alternative bootstrap methods, which can be more computationally intensive. The rescaled bootstrap consists in computing a new set of weights (named bootstrap weights) for each bootstrap sample, which are obtained by applying a scale adjustment to the original survey weights w_i. Specifically, the bootstrap weights are given by

$$w_i^{(b)} = w_i \frac{r_i n}{n-1},$$

with $i \in S$ and $b = 1, \dots, B$, where r_i denotes the number of times that i −th unit is selected in the bootstrap sample $S^{(b)}$. For a finite population, we first consider the additive bias corrected estimator of G based on the rescaled bootstrap, which is defined as

$$\hat{G}_w^{c.Ba} = \hat{G}_w^c - \left(\bar{G}_w^{c.B} - \hat{G}_w^c \right) = 2\hat{G}_w^c - \bar{G}_w^{c.B}, \tag{18}$$

where $\bar{G}_w^{c.B}$ is the expected value of $\hat{G}_w^c$ based on the bootstrap estimates $\hat{G}_w^{c(b)}$, and which are defined as $\hat{G}_w^c$ after substituting the original survey weights w_i with the bootstrap weights $w_i^{(b)}$.

Second, we also consider the aforementioned empirical bootstrap bias correction. We now describe an extension of this method to finite populations. It requires the rescaled bootstrap technique along with confidence intervals and estimators based on survey weights. A confidence interval that can be used in Step 1 is given by the limits $G_L = \hat{G}_{w(0.005)}^{c.B}$ and $G_U = \hat{G}_{w(0.995)}^{c.B}$, where $\hat{G}_{w(\alpha)}^{c.B}$ is the α-th quantile of the weighted estimates $\hat{G}_w^{c(b)}$ derived from the rescaled bootstrap and based on $\hat{G}_w^c$. The following Step 1-b must be included between Steps 1 and 2:

Step 1-b (Pseudo original finite population). Generate a pseudo original population U_k^* with observations given by $\{y_i^* : i \in U_k^*\}$ and selected from $f(y; G_k)$, with $k = 1, \dots, K$.

The pseudo original samples of Steps 2 and 5 are selected from U_k^* instead of $f(y; G_k)$, and using the same sampling design as for the original sample S. The weighted coefficient of skewness is defined as

$$\hat{\gamma}_w = \frac{\mu_{w.3}}{\sigma_w^3}, \tag{19}$$

where $\sigma_w = (\mu_{w.2})^{1/2}$ and $\mu_{w.\alpha} = \hat{N}^{-1} \sum_{i \in S} w_i \left(y_i - \bar{y}_w \right)^\alpha$. In Step 3, $\hat{G}_{nh}^c$ and $\hat{\gamma}_{nh}$ are replaced by $\hat{G}_{wh}^c$ and $\hat{\gamma}_{wh}$, which are calculated using the sample S_h and Equations (12) and (19), respectively. In Step 4, bootstrap estimates are substituted with $\hat{G}_{wt}^{c(b)}$ and $\hat{\gamma}_{wt}^{(b)}$, which are obtained using the rescaled bootstrap method. The expected values of $\hat{G}_w^c$ and $\hat{\gamma}_w$ based on the rescaled bootstrap are denoted, respectively, as $\bar{G}_{wt}^{c.B}$ and $\bar{\gamma}_{wt}^B$, and they are computed using $\hat{G}_{wt}^{c(b)}$ and $\hat{\gamma}_{wt}^{(b)}$ in Equation (13), after substituting r and R with b and B, respectively. The same set of eligible functions are used in Step 6, but they depend on weighted quantities, i.e., $q_l\left(\hat{G}_{wt}^c; \bar{G}_{wt}^{c.B}; \hat{\gamma}_{wt}; \bar{\gamma}_{wt}^B \right)$. In Step 7, the suggested bias corrected estimators of G_v are defined by

$$\hat{G}_{wv}^{c.Bp}(l) = \hat{G}_{wv}^c - \hat{q}_l\left(\hat{G}_{wv}^c; \bar{G}_{wv}^{c.B}; \hat{\gamma}_{wv}; \bar{\gamma}_{wv}^B\right),$$

and they are used to identify, in Step 8, the optimum function $\hat{q}_{opt}$. Finally, the suggested bias corrected estimator that we compute in finite populations is given by

$$\hat{G}_w^{c.Bp} = \hat{G}_w^c - \hat{q}_{opt}\left(\hat{G}_w^c; \bar{G}_w^{c.B}; \hat{\gamma}_w; \bar{\gamma}_w^B\right). \tag{20}$$

For finite populations, additional bias correction procedures can also be computed. For instance, we also calculated, in Section 5, Campbell's (1980) jackknife (see Berger and Skinner, 2005, and Berger, 2008) and the multiplicative bootstrap method (see Hall and Maiti, 2006), but we omitted them because Campbell's jackknife is more biased than (18) and (20), and additive and multiplicative bootstrap methods give similar results.

5   Simulation studies to analyse the bias correction procedures

5.1   Description

We now evaluate various bias correction procedures by means of the *RB* measure defined in Section 3, and using the probabilistic distributions and sample sizes also described in that section. As discussed in Section 4, bias correction procedures substantially mitigate the detected biases, but the price to pay is a possible increase in the *MSE*. For this reason, we use the relative root mean square error (*RRMSE*) to investigate the effect of bias corrections on efficiency. For a given statistic $\hat{\theta}_n$, the corresponding *RRMSE* based on R replications is defined as

$$RRMSE = 100 \times \frac{\left(MSE_R[\hat{\theta}_n]\right)^{1/2}}{\theta},$$

where the empirical mean square error is given by $MSE_R[\hat{\theta}_n] = R^{-1}\sum_{r=1}^R\left(\hat{\theta}_n^{(r)} - \theta\right)^2$. Sections 6 and 8 give more detailed discussions on the importance of both bias and MSE measures. In Section 3, we observed that $\hat{G}_n^a$ yields extremely large values of RB when G is small, and $\hat{G}_n^b$ is slightly more biased than $\hat{G}_n^c$. For the sake of clarity, $\hat{G}_n^a$ and $\hat{G}_n^b$ and the corresponding bias corrected estimators are omitted from the figures in this section, but they can also be computed, as discussed in Section 4. Similarly, for finite populations, weighted estimators of G_N^a and G_N^b are omitted. For infinite populations, the percentage of the number of times that each eligible bias function is selected as the optimum function of the suggested algorithm described in Section 4 can be seen in the supplementary material (Table A3). For the various probabilistic distributions, the bias function defined in Equation (16) is the most often selected as the optimum function.

$B = 1000$ bootstrap samples are used in bootstrap methods. Following Pfeffermann and Correa (2012), we consider $H = 200$ plausible parameters, of which $T = 60$ and $V = 140$ are used for the training and validation groups, respectively. Samples are selected from finite populations with size $N = 10000$, which in turn are drawn from the investigated continuous distributions. We consider unequal inclusion probabilities by using the randomized systematic sampling design (Wu and Thompson, 2020). The effect of the design is increased by generating inclusion probabilities π_i with a correlation of 0.7 between π_i and y_i (Berger and Gedik-Balay, 2020).

5.2 Results and conclusions

For infinite populations, Figure 4 displays the RBs of estimators of G based on samples with size $n = 50$. We observe that the bias correction procedures reduce, in absolute terms, the bias of $\hat{G}_n^c$, although the empirical bootstrap ($\hat{G}_n^{c.Bp}$) performs better than the bias corrected estimator based on the jackknife ($\hat{G}_n^{c.Jo}$), especially for large Gini indices. For the Pareto distribution, the RBs of $\hat{G}_n^{c.Jo}$ and $\hat{G}_n^{c.Bp}$ are close to 0% when their expected values are smaller than 0.3 and 0.4, respectively. For the Dagum-p20 distribution, this situation comes when the expected values are smaller than 0.4, and the RBs of $\hat{G}_n^{c.Jo}$ and $\hat{G}_n^{c.Bp}$ are close to 0% when the expected values are smaller than 0.5 in the Dagum-p0.5 distribution. For the Lognormal distribution, the bias corrected estimators give negligible biases (absolute values of RB smaller than 2%) for the various Gini indices. In the extreme scenario of heavy-tailed distributions and large Gini indices, the values of RB may be as much as -25% (see the Pareto distribution at the last expected value of $\hat{G}_n^c$), and the empirical bootstrap may reduce this RB by more than 10%. The gain in accuracy of the empirical bootstrap over $\hat{G}_n^{c.Jo}$ is greater when $n = 500$ (see Figure A2 in the supplementary material). For instance, with expected values of estimators close to 0.7 in the Pareto distribution, the RBs of $\hat{G}_n^c$, $\hat{G}_n^{c.Jo}$ and $\hat{G}_n^{c.Bp}$ are -15%, -12% and -7%, respectively.

Values of RB based on samples with size $n = 50$ and derived from finite populations can be seen in Figure 5. We observe that the empirical bootstrap ($\hat{G}_w^{c.Bp}$) performs better than the rescaled bootstrap ($\hat{G}_w^{c.Ba}$), which in turn minimizes the bias problem of the customary estimator ($\hat{G}_w^c$). Biases for samples with size $n = 500$ and derived from finite populations can be seen in the supplementary material (Figure A3).

Figure 6 displays the impact on the efficiency of using corrected estimators. The efficiency of bias correction procedures lies in a reasonable range relative to the efficiency of the estimator $\hat{G}_n^c$, especially for the correction based on the empirical bootstrap. In particular, the empirical bootstrap performs better than the jackknife. In addition, the empirical bootstrap is slightly less efficient than the customary estimator $\hat{G}_n^c$, and is more efficient for large Gini indices. The larger the sample size, the lower the impact on the efficiency of both bias corrected estimators, as can be seen in Figure A4. Results derived from finite populations, and based on small and large samples, can be seen in Figures A5 and A6.
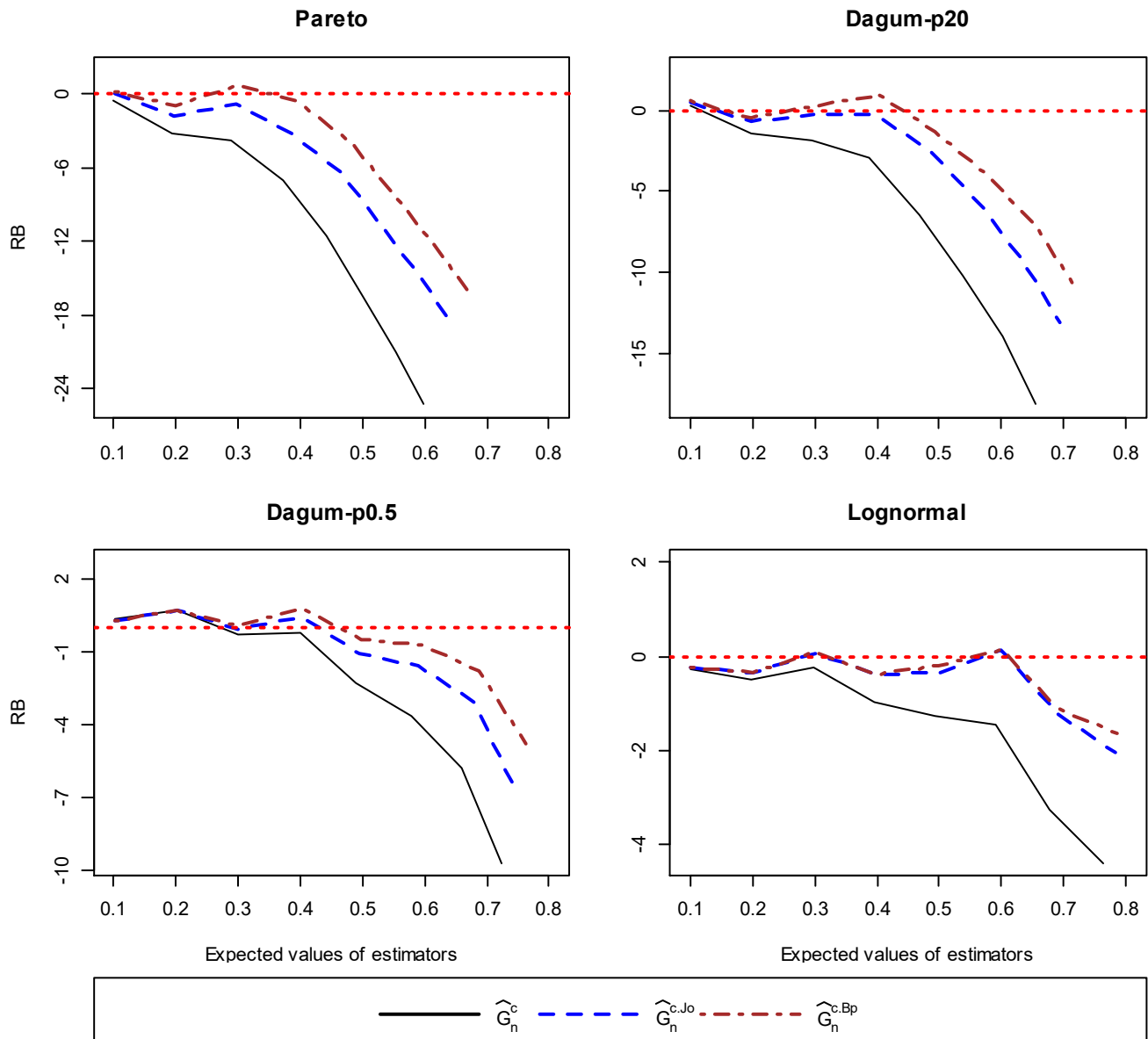
Figure 4: *Relative biases (RBs) of estimators $\hat{G}_n^c$, $\hat{G}_n^{c.Jo}$ and $\hat{G}_n^{c.Bp}$ based on samples with size $n = 50$, and randomly selected from various continuous probabilistic distributions (infinite populations).*
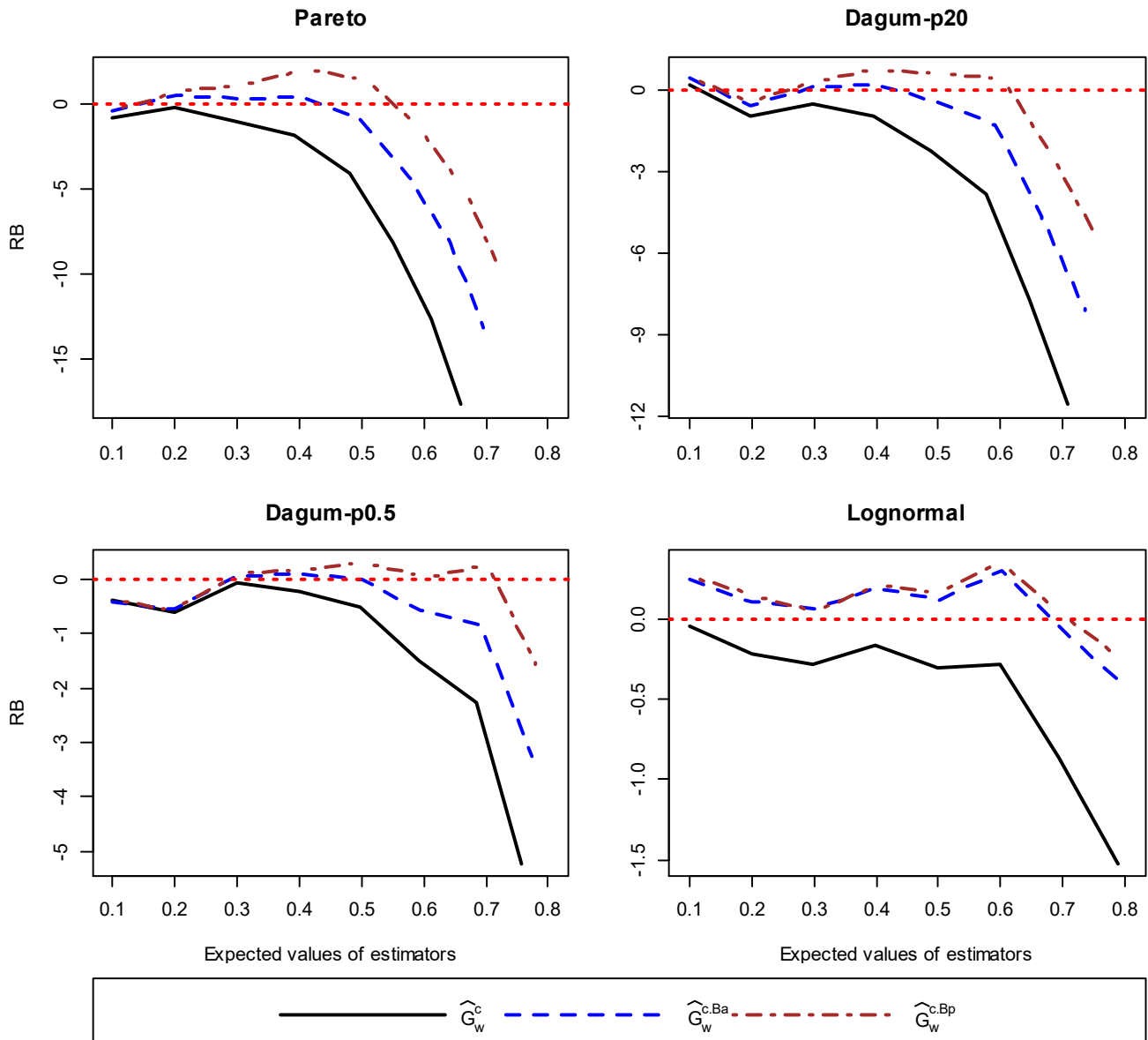
Figure 5: *Relative biases (RBs) of estimators $\hat{G}_w^c$, $\hat{G}_w^{c.Ba}$ and $\hat{G}_w^{c.Bp}$ based on samples with size $n = 50$, and randomly selected from a finite population with size $N = 10000$. Population values are randomly drawn from various continuous probabilistic distributions.*
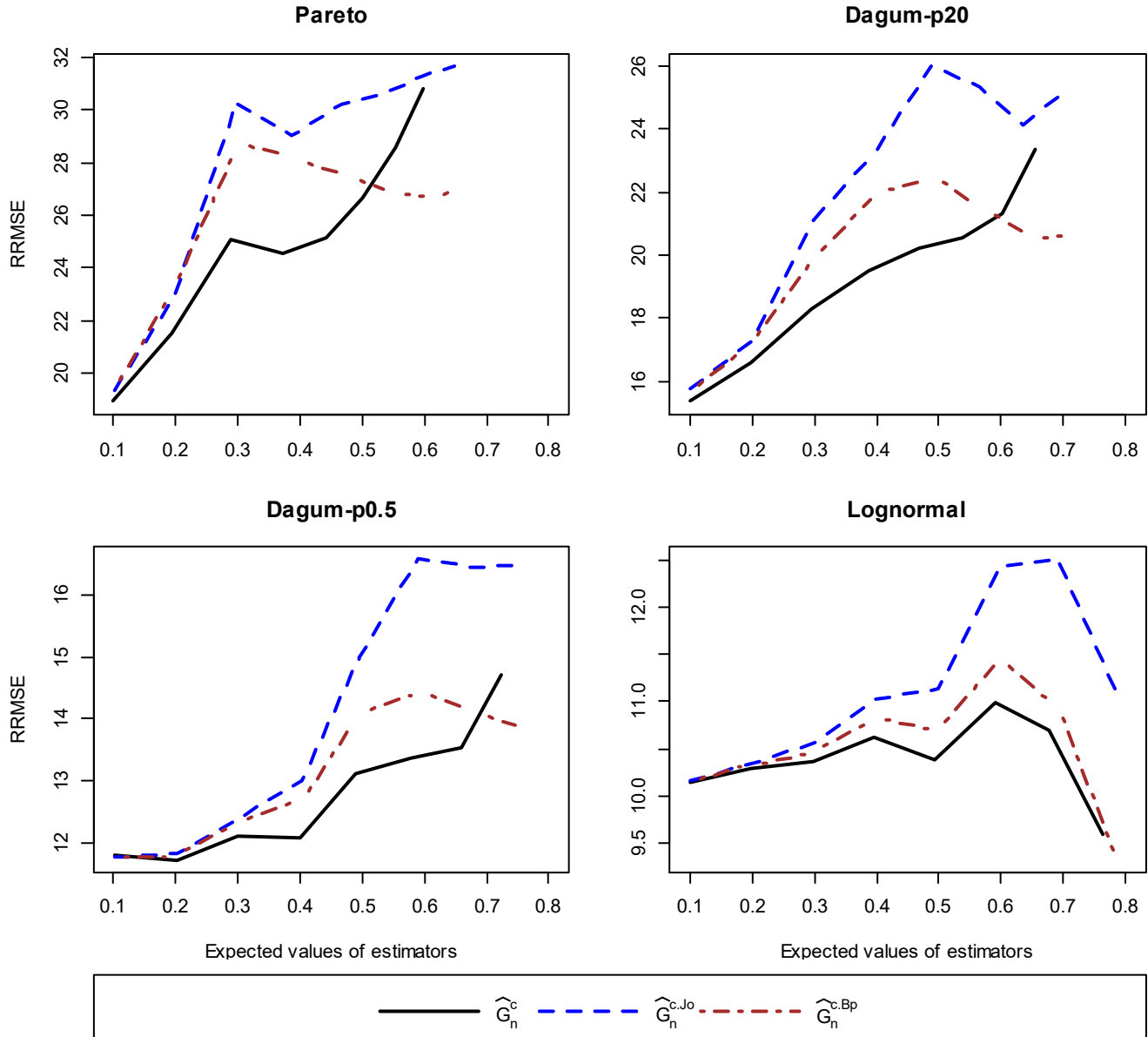
Figure 6: *Relative Root Mean Square Errors (RRMSEs) of estimators $\hat{G}_n^c$, $\hat{G}_n^{c.Jo}$ and $\hat{G}_n^{c.Bp}$ based on samples with size $n = 50$, and randomly selected from various continuous probabilistic distributions (infinite populations).*

6   A simulation-based criterion for deciding when to use bias correction

Results from Section 3 indicate that the three common empirical versions of G can be biased for heavy-tailed distributions, which may be a serious issue when the Gini index is large. An important problem that arises in practice is determining when to use bias correction procedures. Note that the bias and the *MSE* (or equivalently the *RB* and the *RRMSE*) are two relevant measures to evaluate the quality of estimators. However, as noted by Särndal et al. (2003, p. 164), the bias must also be small relative to the standard error, since failure to meet this requirement

may result in invalid confidence intervals and/or undesirable coverage probabilities. This ratio between the bias and the standard error is popularly referred to as the bias ratio. For a given statistic $\hat{\theta}_n$ and R replications, the empirical bias ratio (BR) is defined as

$$BR = \frac{B_R[\hat{\theta}_n]}{\left(V_R[\hat{\theta}_n]\right)^{1/2}},$$

where the empirical variance is given by $V_R[\hat{\theta}_n] = R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}_n^{(r)} - \bar{\theta}_n^R\right)^2$. Like the *MSE*, the *BR* also involves both bias and variance of the estimator. Särndal et al. (2003, p. 41) advise empirical researchers to avoid estimators that are considerably biased, and instead seek out estimators with small biases, and then choose one with a small variance. Following this idea, the suggested criterion consists of analysing both *RB* and *BR* measures for the estimator $\hat{G}_n^c$, and bias corrections are suggested when non-negligible biases are observed. *RRMSE* values can be used to choose the most efficient bias correction estimator. Särndal et al. (2003, p. 165) indicate that the effect of the bias ratio on the coverage probability can be ignored when $|BR| < 0.1$, and the use of bias corrected estimators is not justified here. The effect on the coverage probability is not extremely pronounced when $|BR| \leq 0.5$, but it can be a serious problem otherwise. On the other hand, absolute values of *RB* lower than 2% can be considered negligible, and bias correction procedures are not recommended if this is the case. In summary, bias corrections are suggested when the estimator $\hat{G}_n^c$ satisfies $|BR| \geq 0.1$ and $|RB| \geq 2\%$.

The aim of Figure 7 is to show that the customary estimator $\hat{G}_n^c$ can yield poor bias ratios, with bias corrections justified because they substantially minimize this problem. For infinite populations and samples with size $n = 50$, $\hat{G}_n^c$ yields absolute values of *BR* close to 1.4, and poor coverage probabilities are expected. Furthermore, the vertical lines in Figure 7 indicate the first expected value ($\bar{G}_n^{c.R}$) with non-negligible biases, i.e., with absolute values of *RB* larger than 2%. We see that the condition imposed by the bias ratio ($|BR| < 0.1$) is more demanding than the condition based on the relative bias ($|RB| < 2\%$), i.e., the first value of $\bar{G}_n^{c.R}$ with a $|BR| \geq 0.1$ is smaller than the first value of $\bar{G}_n^{c.R}$ with a $|RB| \geq 2\%$. For example, non-negligible biases are observed for the Pareto distribution when $\bar{G}_n^{c.R} \cong 0.2$, and the absolute value of *BR* is larger than 0.1 in this situation. These results reveal the presence of a mild bias problem, which can be solved using bias correction procedures, as can be seen in Figures 4 and 7. From Figure 7 we also observe that the *BR* values of the corrected estimators, in absolute terms, are generally smaller than 0.5, and are substantially smaller than those of $\hat{G}_n^c$. The desirable properties in terms of both *BR* and *RB* measures and the negligible impact on the efficiency (see the Pareto distribution in Figure 6 when $\bar{G}_n^{c.R} = 2$) indicate that correction procedures are recommended to mitigate the detected biases.
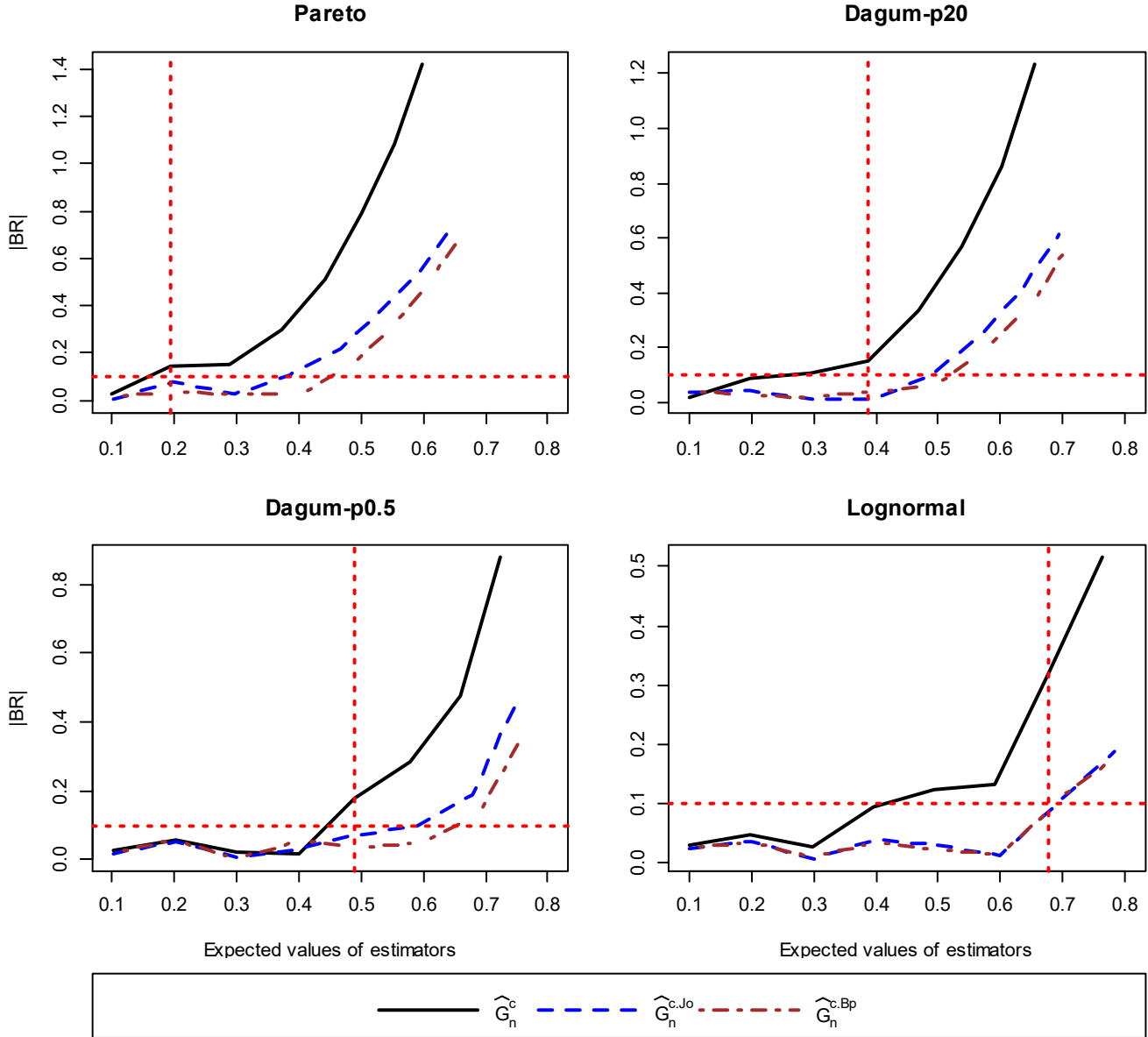
Figure 7: *Bias Ratios (BRs), in absolute terms, of estimators $\hat{G}_n^c$, $\hat{G}_n^{c.Jo}$ and $\hat{G}_n^{c.Bp}$ based on samples with size $n = 50$, and randomly selected from various continuous probabilistic distributions (infinite populations). Using the estimator $\hat{G}_n^c$, horizontal and vertical dotted lines are fixed, respectively, at $|BR| = 0.1$ and at the first expected value with $|RB| > 2\%$.*

In Figure 8 we suggest a simulation-based criterion for deciding when to use bias correction procedures. This method is based on the expected values $\bar{G}_n^{c.R}$ and $\bar{\gamma}_n^R$, since the Gini index and the skewness have a direct effect on the bias. Samples, with sizes between 50 and 1000, are drawn from the most skewed probabilistic distributions described in Section 3. Using the estimator $\hat{G}_n^c$, this criterion is based on two conditions: $|BR| \geq 0.1$ and $|RB| \geq 2\%$. A grading scale classifies the non-negligible biases into three categories: mild ($2 \leq |RB| < 5$), moderate ($5 \leq |RB| < 10$) and severe ($|RB| \geq 10$). This scale can be used to identify the scenarios where bias corrections are

either weakly or strongly recommended. Thus, while bias is not a serious issue for mild biases, the use of bias correction procedures is suggested to reduce this bias. Bias corrections are highly recommended in the case of moderate biases. The bias is a serious problem in the presence of severe biases, meaning bias corrections are strongly advised.
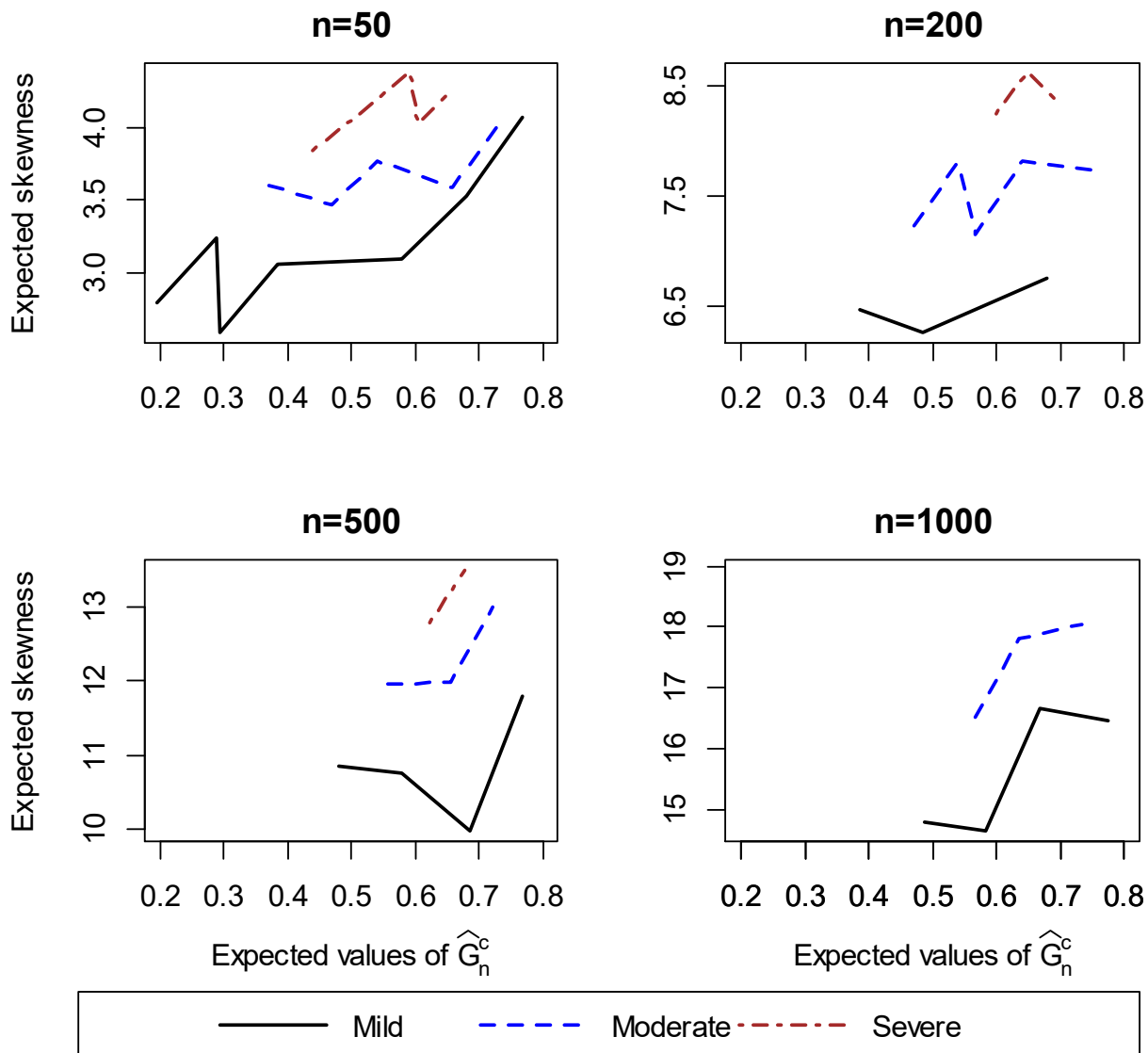

Figure 8: *Grading scale based on the Relative Biases (RBs) and Bias Ratios (BRs) of $\hat{G}_n^c$ when samples, with sizes $n = \{50,200,500,1000\}$, are randomly selected from the Pareto, Dagum-p20, Dagum-p05 and Lognormal distributions (infinite populations). Non-negligible biases with $|BR| \geq 0.1$ are considered as mild ($2 \leq |RB| < 5$), moderate ($5 \leq |RB| < 10$) and severe ($|RB| \geq 10$). The x- and y-axes show the expected values of $\hat{G}_n^c$ and $\hat{\gamma}_n$ ($\bar{G}_n^{c.R}$ and $\bar{\gamma}_n^R$, respectively).*

For small sample sizes ($n = 50$) and $\bar{\gamma}_n^R \geq 2.6$, mild biases are expected when $\bar{G}_n^{c.R} \geq 0.2$, approximately. Moderate biases are observed when $\bar{\gamma}_n^R \geq 3.5$ and $\bar{G}_n^{c.R} \geq 0.37$, and bias corrections are highly recommended in this situation. For samples with size $n = 200$, mild biases

are expected when $\bar{G}_n^{c.R} \geq 0.39$, and moderate biases are obtained when $\bar{G}_n^{c.R} \geq 0.47$. For samples with size $n = 500$, mild and moderate biases are expected when $\bar{G}_n^{c.R} \geq 0.48$ and $\bar{G}_n^{c.R} \geq 0.56$, respectively. Finally, for $n = 1000$, bias corrections could be applied when $\bar{G}_n^{c.R} \geq 0.49$, and moderate biases are observed when $\bar{G}_n^{c.R} \geq 0.57$. Severe biases are not expected when $n \geq 1000$ and $G \leq 0.8$ (the maximum true Gini index considered in this study). In summary, for small samples sizes (e.g., $n = 50$), bias correction procedures may be required when the estimates of the Gini index are greater than 0.2, and can be highly recommended when they exceed 0.37. For larger sample sizes (e.g., $n = 500$), mild biases can be expected when the estimates of the Gini index are greater than 0.48, and bias corrections are highly advisable when the estimates of the Gini index are larger than 0.56. For different sample sizes and Gini indices associated with specific data that empirical researchers are analysing, the aim of Figure 8 is to depict the values of the coefficient of skewness that would require the use of bias corrections.

7 Applications to real data sets

In this section, bias correction procedures are applied for estimating the Gini index in a total of six subpopulations with sizes between 26 and 503, and derived from three real data sets (see Table 1). A common goal in most surveys is not only to provide estimates for the whole population, but also for specific subpopulations (also named domains). For instance, estimates of unemployment in labour-force surveys are provided at national level, but this information is also of special interest at provincial and local levels. In household surveys, subpopulations are usually created on the basis of household sizes or consumption units. Age, sex and occupational groups are also often used to create subpopulations in many studies.

The first real data set consists of total net household incomes extracted from the 2019 Spanish Survey on Income and Living Conditions (ES-SILC). Subpopulations, with sizes $n = \{26,51\}$, are created using different consumption units, with the aim of using the Gini index to estimate income inequality. The second real data set is obtained from the World Bank's Enterprise Survey (WBES), which has been used extensively in international management studies (Vendrell-Herrero et al., 2022; Gomes et al., 2018; etc.). For private sector firms from over 130 developed and developing countries, the WBES contains information on a broad range of topics including competition performance, corruption, financial data, infrastructure, technology, etc. Using this survey, we estimate the Gini index of the labour productivity per hour worked in Argentinean firms for the years 2017 and 2018. The sizes of the resulting subpopulations are $n = \{61,503\}$. Finally, the third real data set (named WATER) consists of a survey on shower habits conducted in Andalusia, a region in southern Spain facing water scarcity. The interest is to analyse the inequality in time spent showering, creating subpopulations, with sizes $n = \{38,74\}$, using the number of inhabitants at provincial level.

The bias corrected estimator $\hat{G}_n^{c.Bp}$ is based on a continuous probabilistic distribution, and for this reason the Kolmogorov-Smirnov (*KS*) Goodness of Fit test is used to fit distributions to the various subpopulations used in this study. From Table 1 we observe that the Lognormal, Fisk, Dagum and Weibull distributions yield *KS* *p*-values above the usual significance level (5%), and the null hypothesis that data come from the corresponding continuous probabilistic distribution is not rejected. For the various subpopulations, the simulation-based criterion described in Section 6 indicates that the use of a bias correction procedure is recommended, since non-negligible biases and *BR*s greater than 0.1 are expected according to Figure 8. In particular, the

24

estimator $\hat{G}_n^c$ is expected to underestimate the true Gini index, with higher estimates expected from the bias corrected estimators.

Population	n	$\hat{\gamma}_n$	$\hat{G}_n^c$	$\hat{G}_n^{c.Jo}$	$\hat{G}_n^{c.Bp}$	Distribution	KS p-value
ES-SILC	26	2.81	0.518	0.534	0.540	Fisk	0.85
					0.526	Lognormal	0.65
	51	3.51	0.476	0.486	0.488	Fisk	0.99
					0.480	Lognormal	0.62
					0.499	Dagum	0.11
WBES	61	3.22	0.505	0.513	0.509	Lognormal	0.38
					0.516	Fisk	0.37
	503	15.85	0.734	0.784	0.752	Fisk	0.57
WATER	38	3.92	0.358	0.368	0.370	Dagum	0.07
					0.359	Weibull	0.06
					0.366	Fisk	0.06
					0.363	Lognormal	0.05
	74	3.09	0.435	0.439	0.442	Fisk	0.10
					0.442	Dagum	0.10

Table 1: *Estimates of the Gini index for various subpopulations derived from the ES-SILC, WBES and WATER populations. The null hypothesis that data come from a specific continuous probabilistic distribution is tested using the Kolmogorov-Smirnov (KS) Goodness of Fit test, and the corresponding p-values are provided. The bias corrected estimator $\hat{G}_n^{c.Bp}$ is based on probabilistic distributions that fit the data.*

For the subpopulation with size $n = 51$ derived from ES-SILC, we observe that estimates of the coefficient of skewness and the Gini index are, respectively, $\hat{\gamma}_n = 3.51$ and $\hat{G}_n^c = 0.476$. These results indicate that moderate biases are expected according to Figure 8. As we expected, bias corrected estimators provide higher estimates than $\hat{G}_n^c$, with values as much as 4.8% larger than $\hat{G}_n^c = 0.476$ (see the estimation of $\hat{G}_n^{c.Bp}$ based on the Dagum distribution). For $n = 503$ in the WBES population, the estimates $\hat{\gamma}_n = 15.85$ and $\hat{G}_n^c = 0.734$ indicate the presence of serious biases, and the difference with respect to the estimator $\hat{G}_n^c$ goes from 2.5% ($\hat{G}_n^{c.Bp} = 0.752$) to 6.8% ($\hat{G}_n^{c.Jo} = 0.784$). For the various subpopulations in this study, we observe that estimates derived from the bias correction procedures are larger than estimates based on $\hat{G}_n^c$, a result which coincides with the findings of Sections 3 and 5.

8   Discussion

The Gini index is a very popular indicator to measure inequality that has been used in many economic studies. For discrete distributions, the Gini index is usually estimated using a plug-in formulation of a given theoretical definition of the Gini index for continuous distributions. This methodology may introduce a serious bias in comparison to the true (asymptotic) value of the Gini index. Note that the Gini index can also be estimated using techniques such as empirical likelihood (Owen, 2001), but there is no simple application of this method to complex sampling designs. The analysis of alternative estimation methodologies is beyond the scope of this paper, i.e., we assume the classical formulations derived from theoretical definitions of the Gini index.

First, this paper attempts to provide a better overview of the problem of estimating the Gini index by regrouping and classifying the most common empirical versions proposed for discrete distributions, and defined under the two existing statistical theories (infinite and finite populations). Second, this paper identifies the scenarios where the bias may be a serious issue, and such scenarios are based on common continuous distributions often used in the modelling of income distributions. For instance, $\hat{G}_n^a$ (denoted as G_N^a in finite populations) yields large biases when the Gini index and the sample size are small, but this bias problem can be easily solved by using the midpoint distribution function in the definition of $\hat{G}_n^a$. When all the sample observations are different, another solution is to use one of the transformations described in Equations (8) and (9). In addition, results derived from this study indicate that the various empirical versions of G produce serious biases in the presence of heavy-tailed distributions and large Gini indices. Accordingly, bias correction procedures are suggested to mitigate this bias problem, and they are investigated using Monte Carlo simulation studies. We also describe a simulation-based criterion for deciding when to use bias corrections. Finally, bias corrected procedures are illustrated by application to the problem of estimating the Gini index in various real data sets.

The empirical bootstrap obtains less biased estimates than alternative bias correction procedures. With infinite populations, the traditional jackknife performs well in terms of relative bias. For finite populations, the rescaled bootstrap may reduce the bias of the existing empirical versions of G. It is important to note that the empirical bootstrap is a parametric procedure that requires generating sets of data from the probabilistic distribution fitted to the original sample. However, the use of continuous distributions in the modelling of income distributions is a common practice in many real-world applications, and the empirical bootstrap can thus be implemented if this is the case. In addition, it should be noted that for the sake of simplicity the empirical bootstrap bias correction is based only on standard regression functions, but alternative bias functions can also be used, and they may potentially improve the performance of this method. Finally, the empirical bootstrap is more computationally intensive than alternative procedures, but this is not a problem with current computing facilities.

The outcome of the grading scale described in Section 6 can help empirical researchers decide whether the specific data they are analysing have non-negligible biases and large bias ratios, meaning the use of bias corrections would therefore be recommended. For heavy-tailed distributions, non-negligible biases may appear in small samples (e.g., $n = 50$) from low estimates of the Gini index (e.g., $\hat{G}_n^c \geq 0.2$). For samples with sizes $n = 200$ and $n = 1000$, non-negligible biases can be expected for estimates of the Gini index greater than 0.4 and 0.5, respectively. Severe biases are not expected when the sample size is larger than 1000. Figure 8 gives a more precise understanding of the conditions required in practice to apply a bias correction, which depend on the sample size and estimates of both the coefficient of skewness and the Gini index.

Both bias and MSE measures are important to evaluate the quality of estimators. Numerous authors indicate that the use of bias correction procedures may have an impact on the efficiency of bias corrected estimators. This issue has also been investigated in this paper, with the results indicating that said impact is not relevant, especially as the sample size increases. The empirical bootstrap is more efficient than alternative bias correction procedures, but slightly less efficient than the customary empirical versions of G, and may even have the smallest *MSE*s for large Gini indices. Conventional advice in the literature is to avoid estimators that are considerably biased, so empirical researchers should seek estimators with smaller biases, and then choose one with a

small variance. Following this idea, the empirical bootstrap can be good choice for estimating the Gini index in the scenarios discussed in Sections 5 and 6. However, alternative bias correction procedures also perform well in terms of bias and efficiency in many situations, and they may be preferable in terms of simplicity.

For less skewed distributions (e.g., Weibull and Gamma), the bias of $\hat{G}_n^c$ is not a problem, and the bias of $\hat{G}_n^b$ lies within a reasonable range. This implies that bias correction procedures are not required for less skewed distributions. Bias corrections are applied to $\hat{G}_n^c$ because it shows the best performance in this study. However, such procedures can easily be applied to any other estimation method in the literature.

The observed biases may have an important impact on the coverage rates of confidence intervals of the Gini index, especially in the case of the large bias ratios obtained by the estimator $\hat{G}_n^c$. This implies that bias corrected estimators are highly recommended for the construction of confidence intervals, since they can be invalid and/or undesirable coverage probabilities can be obtained in the case of moderate or severe biases. Large biases are also observed by the bias corrected estimators in the case of large Gini indices and highly skewed distributions. These arguments represent promising directions for future research. For instance, the interval estimation based on bias corrected estimators can be investigated to analyse when such confidence intervals have desirable empirical coverages. Alternative estimation methodologies can also be used to improve the estimation of the Gini index. In particular, it would be interesting to reduce the biases that still remain in the aforementioned extreme situations (highly skewed distributions with large Gini indices). For instance, information from auxiliary variables can be incorporated at the estimation stage, and more accurate results are expected.

Acknowledgment

Declaration of Conflicting Interests

Supplemental Material

Supplemental material for this article is available online.

Funding

Authors' note

Supplementary material for reproduction of all analyses in this article can be seen at the OSF repository (see Muñoz et al., 2023). This repository contains two files. The document "*BiasCorrectionGiniCodes.html*" contains R scripts to conduct the Monte Carlo experiments described in this article. In addition, data sets generated from the probabilistic distributions can be obtained using such R scripts. The various real data sets used in this article can be loaded using the file "*Datasets.RData*".

References

Anand, S. (1983). *Inequality and poverty in Malaysia: Measurement and decomposition*. The World Bank.

Atkinson, A. B. (2017). Pareto and the upper tail of the income distribution in the UK: 1799 to the present. *Economica*, 84(334), 129-156.

Bakar, S. A. A., & Pathmanathan, D. (2020). Income modeling with the Weibull mixtures. *Communications in Statistics-Theory and Methods*, 1-17.

Berger, Y. G. (2008) A note on the asymptotic equivalence of jackknife and linearization variance estimation for the Gini Coefficient. *Journal of Official Statistics*, 24(4), 541-555.

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

Berger, Y. G., & Muñoz, J.F. (2015). On estimating quantiles using auxiliary information. *Journal of Official Statistics*, 31(1), 101-119.

Berger, Y. G., & Skinner, C. J. (2005). A jackknife variance estimator for unequal probability sampling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1), 79-89.

Bonacini, L., Gallo, G., & Scicchitano, S. (2021). Working from home and income inequality: risks of a 'new normal'with COVID-19. *Journal of Population Economics*, 34(1), 303-360.

Bulle, N. (2016). A Method of measuring inequality within a selection process. *Sociological Methods & Research*, *45*(1), 69-108.

Campbell, C. (1980). A different view of finite population estimation. *Proceedings of the Survey Research Methods Section, ASA,* 1980, 319-324.

Canelas, C., & Gisselquist, R. M. (2019). Horizontal inequality and data challenges. *Social Indicators Research*, 143(1), 157-172.

Clementi F., & Gallegati M. (2005) Pareto's law of income distribution: Evidence for Germany, the United Kingdom, and the United States. In: Chatterjee A., Yarlagadda S., Chakrabarti B.K. (eds) *Econophysics of Wealth Distributions. New Economic Windows.* Springer, Milano. https://doi.org/10.1007/88-470-0389-X_1

Cramer, H. (1957). Mathematical Methods of Statistics. Princeton University Press, Seventh Printing, Princeton .

Davidson, R. (2009). Reliable inference for the Gini index. *Journal of Econometrics*, 150(1), 30-40.

Deltas, G. (2003). The small-sample bias of the Gini coefficient: results and implications for empirical research. *Review of Economics and Statistics*, 85(1), 226-234.

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

Dorić, D., Nikolić-Dorić, E., Jevremović, V., & Mališić, J. (2009). On measuring skewness and kurtosis. *Quality and Quantity*, *43*(3), 481-493.

Efron, B. (1990) More efficient bootstrap computations. *Journal of the American Statistical Association*, 55, 79–89.

Efron, B., & Tibshirani, R. (1993) *An introduction to the bootstrap*. Chapman and Hall, New York, London.

Fabrizi, E., & Trivisano, C. (2016). Small area estimation of the Gini concentration coefficient. *Computational Statistics & Data Analysis*, 99, 223-234.

Gini, C. (1912). *Variabilità e mutabilità. Reprinted in Memorie di metodologica statistica*; Ed. Pizetti E.

Giles, D. E. (2004). Calculating a standard error for the Gini coefficient: some further results. *Oxford Bulletin of Economics and Statistics*, 66(3), 425-433.

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

Gomes, E., Vendrell-Herrero, F., Mellahi, K., Angwin, D., & Sousa, C. M. (2018). Testing the self-selection theory in high corruption environments: evidence from African SMEs. *International marketing review*.

Hall, P., & Maiti, T. (2006). On parametric bootstrap methods for small area prediction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2), 221-238.

Jasso, G. (1979). On Gini's mean difference and Gini's index of concentration. *American Sociological Review*, 44(5), 867-870.

Jasso, G. (2021). Linking input inequality and outcome inequality. *Sociological Methods & Research*, *50*(3), 944-1005.

Jiao, J., & Han, Y. (2020). Bias correction with jackknife, bootstrap, and taylor series. *IEEE Transactions on Information Theory*, *66*(7), 4392-4418.

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.

Larraz, B. (2015). Decomposing the Gini inequality index: An expanded solution with survey data applied to analyze gender income inequality. *Sociological Methods & Research*, 44(3), 508–533.

Larraz, B., Pavía, J. M., & Herrera-Gómez, M. (2020). Spatial aggregation and resampling expansion of big surveys: An analysis of wage inequality. *Regional Science Policy & Practice*, 13(3), 957-981.

Larraz, B., Pavía, J. M., & Vila, L. E. (2019). Beyond the gender pay gap. *Convergencia*, 81, 1-34.

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

Moran, T. P. (2006). Statistical inference for measures of inequality with a cross-national bootstrap application. *Sociological Methods & Research*, *34*(3), 296-333.

Moya, P. J., Muñoz, J. F., Álvarez-Verdejo, E., & Blanco-Encomienda, F. J. (2020). Rescaled bootstrap confidence intervals for the population variance in the presence of outliers or spikes in the distribution of a variable of interest. *Communications in Statistics-Simulation and Computation*, 1-30.

Muñoz, J. F., Álvarez-Verdejo, E., & García-Fernández, R. M. (2018). On estimating the poverty gap and the poverty severity indices with auxiliary information. *Sociological Methods & Research*, 47(3), 598-625.

Muñoz, J.F., Moya, P.J., & Álvarez-Verdejo, E. (2023). R codes for estimators of the Gini index. https://doi.org/10.17605/OSF.IO/4YNBS.

Ogwang, T. (2000). A convenient method of computing the Gini index and its standard error. *Oxford Bulletin of Economics and Statistics*, 62(1), 123-123.

Owen, A. B. (2001). *Empirical likelihood*. Chapman and Hall/CRC.

Pérez, C. G., & Alaiz, M. P. (2011). Using the Dagum model to explain changes in personal income distribution. *Applied Economics*, 43(28), 4377-4386.

Pfeffermann, D., & Correa, S. (2012). Empirical bootstrap bias correction and estimation of prediction mean square error in small area estimation. *Biometrika*, 99(2), 457-472.

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

Qin, Y., Rao, J. N. K., & Wu, C. (2010). Empirical likelihood confidence intervals for the Gini measure of income inequality. *Economic Modelling*, 27(6), 1429-1435.

Quatember, A. (2015). The bootstrap method in survey sampling. In *Pseudo-Populations* (pp. 71-84). Springer, Cham.

Rao, J. N. K., Wu, C. F. J., & Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18, 209-217.

Salem, A. B., & Mount, T. D. (1974). A convenient descriptive model of income distribution: the gamma density. *Econometrica: Journal of the Econometric Society*, 1115-1127.

Särndal, C. E., Swensson, B., & Wretman, J. (2003). *Model assisted survey sampling*. Springer Science & Business Media.

Sen, A. (1973). Poverty, inequality and unemployment: Some conceptual issues in measurement. *Economic and Political Weekly*, 1457-1464.

Szymańska, A. (2021). Reducing socioeconomic inequalities in the European Union in the context of the 2030 Agenda for Sustainable Development. *Sustainability*, 13(13), 7409.

Tridico, P. (2018). The determinants of income inequality in OECD countries. *Cambridge Journal of Economics*, 42(4), 1009-1042.

Van Ourti, T., & Clarke, P. (2011). A simple correction to remove the bias of the Gini coefficient due to grouping. *Review of Economics and Statistics*, *93*(3), 982-994.

Vendrell-Herrero, F., Darko, C. K., Gomes, E., & Lehman, D. W. (2022). Home-market economic development as a moderator of the self-selection and learning-by-exporting effects. *Journal of International Business Studies*, 1-17.

Villar, J. G., & Raya, J. M. (2015). Use of a Gini index to examine housing price heterogeneity: A quantile approach. *Journal of Housing Economics*, 29, 59-71.

Wan, G. H. (2001). Changes in regional inequality in rural China: decomposing the Gini index by income sources. *Australian Journal of Agricultural and Resource Economics*, 45(3), 361-381.

Wang, H. Y., Chou, W., Shao, Y., & Chien, T. W. (2020). Comparison of Ferguson's δ and the Gini coefficient used for measuring the inequality of data related to health quality of life outcomes. *Health and Quality of Life Outcomes*, 18, 1-9.

Wang, D., Zhao, Y., & Gilmore, D. W. (2016). Jackknife empirical likelihood confidence interval for the Gini index. *Statistics & Probability Letters*, 110, 289-295.

Wolter, K. (2007). *Introduction to variance estimation*. Springer Science & Business Media.

Wu, C., & Thompson, M. E. (2020). Simple single-stage sampling methods. In *Sampling Theory and Practice* (pp. 17-31). Springer, Cham.

Yang, Q., Liu, T., Valdez, R., Moonesinghe, R., & Khoury, M. J. (2010). Improvements in ability to detect undiagnosed diabetes by using information on family history among adults in the United States. *American Journal of Epidemiology*, 171(10), 1079-1089.

Yitzhaki, S. (1998). More than a dozen alternative ways of spelling Gini. *Research on Economic Inequality*, 8, 13–30.