Silvia De Nicolò, Maria Rosaria Ferrante,
Silvia Pacei

Mind the Income Gap: Bias Correction of Inequality Estimators in Small-Sized Samples

Quaderni di Dipartimento

Serie Ricerche 2022, n. 1
ISSN 1973-9346

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Dipartimento di Scienze Statistiche "Paolo Fortunati"

Mind the Income Gap: Bias Correction of Inequality Estimators in Small-Sized Samples

Silvia De Nicolò* Maria Rosaria Ferrante Silvia Pacei

University of Bologna

Abstract

Income inequality estimators are biased in small samples, leading generally to an underestimation. This aspect deserves particular attention when estimating inequality in small domains. After investigating the nature of the bias, we propose a bias correction framework for a large class of inequality measures comprising Gini Index, Generalized Entropy and Atkinson index families by accounting for complex survey designs. The proposed methodology is based on Taylor's expansions and does not require any parametric assumption on income distribution, being very flexible. Design-based performance evaluation of our proposal has been carried out using data taken from the EU-SILC survey, showing a noticeable bias reduction for all the measures. Lastly, a small area estimation exercise shows the risks of ignoring prior bias correction in a basic area-level model, determining model misspecification.

Keywords — Bias Correction, Complex Surveys, Income Inequality, Small Sample Inference

1 Introduction

Inequality estimators are known to be biased in small samples (Breunig and Hutchinson, 2008; Deltas, 2003). The bias may depend on the dispersion of the variable of interest and, for some specific measures, also on the skewness of its distribution (Breunig, 2001). This aspect deserves attention given that estimates of inequality measures are used for comparisons across time and location. Neglecting such bias may bring out discrepancies that, rather than being true inequality gaps, are completely due to not comparable sample sizes or to different underlying distributions of the variable of interest (Breunig and Hutchinson, 2008).

The problem of observations scarcity arises when dealing with inequality in specific sub-populations, such as age-sex-race groups, as well as in case of inequality mapping at great geographical levels of disaggregation. The interest in reliable economic inequality estimates is growing due to the observed increment in the income gap and social exclusion among regions and to their potential contribution to policies planning and regional studies

*silvia.denicolo@unibo.it

(Cavanaugh and Breau, 2018). In this context, the small area estimation is the main field of application: consider that small area models specified at "area" or "domain" level require the assumption of (approximate) unbiasedness of survey estimators. Avoiding such an issue in case of inequality estimators may lead to model misspecification, resulting in misleading inference.

When measuring economic inequality, disposable income is generally adopted as the variable of interest. Income data are collected through household surveys with a complex sampling design, involving stratification and selection of sampling units in more than one stage. Thus, the sample selection process, together with ex-post treatment procedures such as calibration and imputation, invariably introduces a complex correlation structure in the data that has to be taken into account. This makes the development of a theoretically valid bias correction challenging, in contrast to classical *iid* settings. Furthermore, the bias issue is even exacerbated in income data applications, traditionally affected by extreme values (Van Kerm, 2007), since inequality measures are known to be unrobust to outliers (Cowell and Victoria-Feser, 1996). This aspect depends clearly on the type of measure we are dealing with and it becomes even more cumbersome to handle in case of small samples.

The inequality concept does not have a unique definition but diverse ones, spanning from objective to subjective viewpoints. The statistical indicators used for inequality measurement are several, incorporating diverse properties and sensitivities to income transfers. In light of this, the concurrent estimation of various indicators, alternative to the most common Gini index, may enable a wider picture on inequality phenomenon.

Concerning the Gini index, a large body of literature faces the small sample bias issue, such as Jasso (1979), Lerman and Yitzhaki (1989), Deltas (2003), Davidson (2009), Van Ourti and Clarke (2011) in *iid* samples. The application context is large, spanning from economic inequality to crime (Mohler et al., 2019) or scholar citations concentration (Kim et al., 2020). Fabrizi and Trivisano (2016) tackle it in the complex survey case and their correction is indeed considered within a small area estimation framework. However, concerning alternative measures such as Atkinson Indexes and the Generalized Entropy (GE) measures, the literature on bias is very scarce even in the *iid* case: some contributions are provided by Giles (2005), Schluter and van Garderen (2009) and Breunig and Hutchinson (2008) by adopting different methodological approaches of correction.

The aim of the paper is to investigate the nature of the bias and to propose a methodological framework for bias correction in the specific setting of finite populations and complex sampling design. We consider a large set of measures, from the Gini index to two parametric families of measures: the Atkinson and the Generalized Entropy family. This may foster further studies to focus on alternative inequality measures and to provide a concurrent estimation of diverse indicators in small domains. To our purpose, we take advantage from the methodology based on Taylor's expansions, we show that the same analytical results can be obtained through other types of linearization, such as the one proposed by Graf (2011). An advantage of our proposal is that any parametric assumption on income distribution is not required, providing a very flexible framework. Our bias correction proposal is evaluated via simulations showing a noticeable bias reduction for all the measures and leading, in some cases, to approximately unbiased estimators. An in-depth analysis of measure sensitivities confirms the great impact outliers have on the magnitude of estimators bias and error.

The paper is organized as follows. The considered inequality measures are defined in

Section 2, while the bias correction strategy is set out in Section 3 and the bias-correction estimation steps are detailed in Section 4. A design-based simulation study involving the European Survey of Income and Living Condition (EU-SILC) income data (Guio, 2005) is provided in Section 5 to evaluate the magnitude of the bias and the efficacy of our proposal. Lastly, a small area estimation exercise is carried out in Section 6, in order to highlight the utility of our proposal in practice. Conclusions are drawn in Section 7.

2 Inequality Measures

The most famous inequality measure is, indeed, the Gini concentration index, employed in social sciences for measuring concentration in the distribution of a positive random variable. There are several equivalent definitions of Gini index (Ceriani and Verme, 2015), we will use the formulation of Sen (1997). Suppose we have a finite population $\mathcal{U}$ of $N(<\infty)$ elements labelled as $\{1, \ldots, N\}$. Let y_i be a characteristic of interest, in our case income, for the i-th unit of the finite population, where $y_i \in \mathbb{R}^+$, $i = 1, \ldots, N$, and a sample s_{iid} of size n_{iid} is picked through simple random sampling. The Gini index estimator is defined as

$$\theta_G = \frac{2}{\hat{\mu} n_{iid}^2} \sum_{i \in s_{iid}} n_i y_i - \frac{n_{iid}+1}{n_{iid}},$$

with n_i denoting the rank of i-th unit and $\hat{\mu}$ the sample mean.

However, the estimation of alternative measures, in addition to the Gini index, may enable a more meaningful assessment of different aspects of economic inequality. Gini index does not allow the decomposition of inequality in within groups and between groups components, moreover, it is positional (weakly) transfer sensitive, namely index variations induced by income transfers between individuals depend on the ranks of donor and recipient. Lastly, it constitutes a stochastic dominance measure based on a partial ordering of probability distributions: two very different distributions - one having more inequality amongst the poor, the other amongst the rich can have the same index value.

When the distributional dominance fails, welfare-based measures, such as Atkinson Indexes, may provide for a *complete* ranking among alternative distributions at the expense of more stringent assumptions as to how to represent social welfare (Bellu and Liberati, 2006). Atkinson index has support [0,1] and is defined as

$$\theta_A(\varepsilon) = \begin{cases} 1 - \frac{1}{\hat{\mu}} \left(\frac{1}{n_{iid}} \sum_{i \in s_{iid}} y_i^{1-\varepsilon} \right)^{1/(1-\varepsilon)} & \text{for } \varepsilon \neq 1 \\ 1 - \frac{1}{\hat{\mu}} \left(\prod_{i \in s_{iid}} y_i \right)^{1/n_{iid}} & \text{for } \varepsilon = 1. \end{cases}$$

The parameter ε expresses the level of inequality aversion: as ε increases, the index becomes more sensitive to changes at the lower end of the income distribution.

Besides, an additive decomposable family of inequality measures is the Generalized Entropy class. As opposed to the measures seen before, this class has the advantage to be strongly transfer-sensitive, meaning that it reacts to transfers depending on donor and recipient income levels. It is based on the concept of entropy which applied to income distributions has the meaning of deviations from perfect equality:

$$\theta_{GE}(\alpha) = \begin{cases} \frac{1}{n_{iid}\alpha(\alpha-1)} \sum_{i\in s_{iid}} \left[\left(\frac{y_i}{\hat{\mu}}\right)^{\alpha} - 1 \right] & \alpha \neq 0, 1, \\ \frac{1}{n_{iid}} \sum_{i\in s_{iid}} \frac{y_i}{\hat{\mu}} \ln \frac{y_i}{\hat{\mu}} & \alpha \to 1, \\ -\frac{1}{n_{iid}} \sum_{i\in s_{iid}} \ln \frac{y_i}{\hat{\mu}} & \alpha \to 0. \end{cases}$$

The parameter α sets the sensitivity of the index: a large α induces the index to be more sensitive to the upper tail, and vice versa a small α to the lower tail. $\theta_{GE}(0)$ is the Mean Log Deviation, while $\theta_{GE}(1)$ is the well known Theil index. Atkinson and Generalized Entropy are two interrelated parametric families of measures, as a transformation of the Atkinson Index is a member of the GE class:

$$\theta_A(\varepsilon) = 1 - [\varepsilon(\varepsilon - 1) \cdot \theta_{GE}(1 - \varepsilon) + 1]^{1/(1-\varepsilon)}.$$

In this paper, we consider the estimation of both classes separately, since common parameter values used in one family do not correspond deterministically to parameter values commonly used for the other family. Lastly, we consider as inequality measure the coefficient of variation (CV), which is linked with a member of the GE family, namely $\theta_{GE}(2) = \mathrm{CV}^2/2$. Its square has been used in some income distribution analyses, including OECD (2011), even though it seems to be very sensitive to top outliers (Atkinson, 2015).

3 Bias Correction Proposal

The bias of inequality estimators in small samples can be due to the structure of inequality measures as a non-linear function of estimators. The bias can be either positive or negative, depending on the characteristics of the reference variable distribution, except for the Mean Log Deviation which has structurally negative bias; this aspect is made clearer in the following. Among the measures with non-predictable bias direction, Breunig (2001) shows that the bias of CV and GE(2) is negatively related to income skewness. This aspect could be analyzed in-depth by imposing a distributional assumption on the income variable, but this is out of scope. For GE and Atkinson measures, the limiting behavior of their bias is described in the following proposition.

Proposition 1. *For the measures belonging to the GE and Atkinson families, the expectation of their sample estimator $\hat{\theta}$, considering its true population value as θ, can be expressed as:*

$$\mathbb{E}[\hat{\theta}] = \theta + O\left(\frac{1}{n_{iid}}\right),$$

with n_{iid} denoting the sample size in the iid case.

Proof. In appendix. □

Our bias-correction proposal constitutes a generalization of the framework of Breunig and Hutchinson (2008), developed for *iid* observations, to the finite population and full design-based setting. At the same time, we extend the proposal to a wider set of measures comprising Gini Index. We provide a closed-form bias correction for complex designs which allows us to avoid the use of resampling techniques and can be applied in a distribution-free setting at once. This generalization has been developed considering Horvitz-Thompson type estimators, and the ultimate clusters technique for variances and covariances estimation.

We are interested in a variety of non-linear functions of income values as inequality measures are. Let denote with s a sample of size n, drawn using a complex sampling design, with $p(s)$ the probability of selecting the particular sample $s \subset \mathcal{U}$ out of the set of all possible samples $\mathcal{Q}$, thus $p(s) \geq 0$ and $\sum_{s \in \mathcal{Q}} p(s) = 1$. The inclusion probability of unit k is denoted with π_k, being $\pi_k = \sum_{s \in \mathcal{Q}_k} p(s)$ with $\mathcal{Q}_k$ the set of all possible samples including unit k.

We consider the generic inequality measure written as a function of the mean μ and $\gamma = \mathbb{E}[g(y)]$, with $g(\cdot)$ a generic monotone transformation of the income variable. The population value for the generic inequality measure is

$$\theta = f(\mu, \gamma), \tag{1}$$

with $f(\cdot)$ a twice-differentiable function. The related estimator in our complex survey framework is $\hat{\theta} = f(\hat{\mu}, \hat{\gamma})$ in which Horvitz-Thompson estimators of μ and γ are plugged in, i.e.

$$\hat{\mu} = \frac{\sum_{i \in s} w_i y_i}{N} \quad \text{and} \quad \hat{\gamma} = \frac{\sum_{i \in s} w_i g(y_i, w_i)}{N}, \tag{2}$$

where $w_i = 1/\pi_i$ or a treated and calibrated version of it and N is the population size. Note that results in this section hold also when employing Hájek type estimator, i.e. with denominator $\hat{N} = \sum_{i=1}^{n} w_i$, since it is approximately unbiased (Särndal et al., 2003, pg. 182). Kakwani (1990) uses a similar approach to express inequality indices to derive their asymptotic standard error. By simply applying a second order Taylor's series expansion of the sample estimator around the population values and evaluating its expected value, the bias can be expressed as

$$\begin{aligned}\mathbb{E}[\hat{\theta} - \theta] = &\frac{\partial f(\gamma, \mu)}{\partial \gamma} \mathbb{E}[\hat{\gamma} - \gamma] + \frac{1}{2} \frac{\partial^2 f(\gamma, \mu)}{\partial \gamma^2} (\mathbb{V}[\hat{\gamma}] + \mathbb{E}^2[\hat{\gamma} - \gamma]) + \\ &+ \frac{\partial^2 f(\gamma, \mu)}{\partial \gamma \partial \mu} (Cov[\hat{\gamma}, \hat{\mu}] - \mu \, \mathbb{E}[\hat{\gamma} - \gamma]) + \frac{1}{2} \frac{\partial^2 f(\gamma, \mu)}{\partial \mu^2} \mathbb{V}[\hat{\mu}] + O(n^{-2}),\end{aligned} \tag{3}$$

notice that $\hat{\mu}$ is unbiased.

We detail the design-based estimators for each inequality measure and we provide their explicit bias formulation based on Equation 3, defining all the relevant quantities, in Table 1. The complex survey estimators of Atkinson and Generalized Entropy measures come from Biewen and Jenkins (2006), while as regards the Gini index, we employ the alternative formulation defined by Sen (1997) and the complex survey estimator proposed by Langel and Tillé (2013). Let denote with $\sqrt{n/(n-1)}$ the standard bias-correction adjustment for the weighted variance; $F(\cdot)$ denotes the cumulative distribution function of the variable of interest and lastly consider $\hat{N}_i = \sum_{k \in s} w_k \mathbb{1}(n_k \leq n_i)$. The notation $\mathbb{1}(A)$ defines an indicator function, assuming value 1 if A is observed and 0 otherwise.

Notice that the bias expression results of Table 1 can be reached also in a different way, by applying the linearization proposed by Graf (2011) and extended by Vallée and

Tillé (2019), as made explicit in the Appendix. However, the methodology we adopted allows to explicit the bias in closed form for a generic inequality measure that comprises the entire set of considered measures, as in Equation 3, isolating its components and easing a general interpretation. In contrast, the methodology of Graf (2011) requires a separate derivation for each measure, not enabling a general formulation.

Concerning the Gini index estimator $\hat{\theta}_G$, let us consider γ and $\hat{\gamma}$ defined in Table 1 for Gini, the approximate bias in small sample is

$$\begin{aligned}\mathbb{E}[\hat{\theta}_G - \theta_G] &\approx \frac{2}{\mu}\,\mathbb{E}[\hat{\gamma}-\gamma] + \frac{2\gamma}{\mu^3}\mathbb{V}[\hat{\mu}] - \frac{2}{\mu^2}(Cov[\hat{\mu},\hat{\gamma}] - \mu\,\mathbb{E}[\hat{\gamma}-\gamma]) \\ &= \frac{4}{\mu}\,\mathbb{E}[\hat{\gamma}-\gamma] + \frac{2\gamma}{\mu^3}\mathbb{V}[\hat{\mu}] - \frac{2}{\mu^2}Cov[\hat{\mu},\hat{\gamma}],\end{aligned} \tag{4}$$

where θ_G is the true value. The derivation of the approximate bias related to the weighted estimator $\hat{\gamma}$ is not trivial. As explained by Langel and Tillé (2013), its numerator is not composed of two simple sums. Indeed the quantity $\hat{N}_k$, an estimator of the rank of unit k, is random since its value depends on the selected sample. A solution is to consider the approximate bias of the corresponding *iid* estimator, i.e. $\mathbb{E}[\hat{\gamma}-\gamma] = -1/n(\gamma - \mu/2)$ as derived by Davidson (2009), so that:

$$\mathbb{E}[\hat{\theta}_G - \theta_G] = \frac{-2\theta_G}{n} + \frac{2\gamma}{\mu^3}\mathbb{V}(\hat{\mu}) - \frac{2}{\mu^2}Cov(\hat{\mu},\hat{\gamma}). \tag{5}$$

This correction is in line with Davidson (2009) and Fabrizi and Trivisano (2016) proposals. However these are based on a first-order Taylor's expansions and thus limited to the first term of the right-hand side equation (4), ours extends it to a second-order expansion. This translates into the fact that, while Jasso (1979), Deltas (2003) and Davidson (2009) proposals identify the adjusted Gini in *iid* context as $n(n-1)^{-1}\hat{\theta}_G$, our correction reconsiders the shape of the adjusted estimator with a further order of approximation as

$$\frac{n}{n-2}(\hat{\theta}_G - a), \tag{6}$$

with a equals the sum of the second and third terms of (5).

As clear from Table 1, the bias correction of GE(2) does not include the coefficient of skewness of the income distribution, as showed by Breunig (2001). Actually, a reliable estimation of that quantity, while being straightforward in the *iid* case, appears cumbersome in case of weighted data being defined on a discrete grid of values. This leads to the non-applicability of the bias formula derived by Breunig (2001) in our case.

4 Bias Estimation

In this section, we detail the estimation of the approximate bias defined in Table 1 for each measure. Such estimation is not trivial considering that the mentioned expressions depend on design variances and covariances $\mathbb{V}[\hat{\mu}]$, $\mathbb{V}[\hat{\gamma}]$ and $Cov[\hat{\mu},\hat{\gamma}]$. We consider a complex survey design involving stratification and multi-stage selection, with both Self-Representing (SR), included at the first stage with probability one, and Non-Self-Representing (NSR) strata. This design is consistent with the majority of income survey designs and, in general, with official statistics household surveys.

Measure	$\gamma = \mathbb{E}[g(y)]$	Design Estimator	$\hat{\gamma}$	$f(\hat{\mu}, \hat{\gamma})$	Approximate Bias
Gini	$\mathbb{E}[y \cdot F(y)]$	$\frac{2\sum_{i\in s} w_i y_i (\hat{N}_i - w_i/2)}{N^2 \hat{\mu}} - 1$	$\frac{\sum_{i\in s} w_i y_i (\hat{N}_i - \frac{w_i}{2})}{N^2}$	$\frac{2\hat{\gamma}}{\hat{\mu}} - 1$	$\frac{4}{\mu}\,\mathbb{E}[\hat{\gamma} - \gamma] + \frac{2\gamma}{\mu^3}\mathbb{V}[\hat{\mu}] - \frac{2}{\mu^2} Cov[\hat{\mu}, \hat{\gamma}]$
GE(α) $\alpha \neq 0, 1$	$\mathbb{E}[y^\alpha]$	$\frac{n(n-1)^{-1}}{\alpha(\alpha-1)}\left[\frac{\sum_{i\in s} w_i y_i^\alpha}{N\hat{\mu}^\alpha} - 1\right]$	$\frac{\sum_{i\in s} w_i y_i^\alpha}{N}$	$\frac{n(n-1)^{-1}}{\alpha(\alpha-1)}\left[\frac{\hat{\gamma}}{\hat{\mu}^\alpha} - 1\right]$	$\frac{n(n-1)^{-1}}{\mu^{\alpha+1}(\alpha-1)}\left[\frac{\gamma(\alpha+1)}{2\mu}\mathbb{V}[\hat{\mu}] - Cov[\hat{\gamma}, \hat{\mu}]\right]$
GE(0)	$\mathbb{E}[\log y]$	$\frac{1}{N}\sum_{i\in s} w_i \log\frac{\hat{\mu}}{y_i}$	$\frac{\sum_{i\in s} w_i \log y_i}{N}$	$\log(\hat{\mu}) - \hat{\gamma}$	$-\frac{1}{2\mu^2}\mathbb{V}[\hat{\mu}]$
GE(1)	$\mathbb{E}[y(\log y)]$	$\frac{1}{N}\sum_{i\in s} w_i \frac{y_i}{\hat{\mu}} \log\frac{y_i}{\hat{\mu}}$	$\frac{\sum_{i\in s} w_i y_i \log y_i}{N}$	$\frac{\hat{\gamma}}{\hat{\mu}} - \log(\hat{\mu})$	$\left[\frac{\gamma}{\mu^3} + \frac{1}{2\mu^2}\right]\mathbb{V}[\hat{\mu}] - \frac{1}{\mu^2} Cov[\hat{\mu}, \hat{\gamma}]$
A(ε) $\varepsilon \neq 1$	$\mathbb{E}[y^{1-\varepsilon}]$	$1 - \frac{1}{\hat{\mu}}\left[\frac{1}{N}\sum_{i\in s} w_i y_i^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}$	$\frac{\sum_{i\in s} w_i y_i^{1-\varepsilon}}{N}$	$1 - \frac{\hat{\gamma}^{\frac{1}{1-\varepsilon}}}{\hat{\mu}}$	$\frac{\gamma^{\frac{\varepsilon}{1-\varepsilon}}}{\mu}\left[\frac{Cov[\hat{\gamma},\hat{\mu}]}{\mu(1-\varepsilon)} - \frac{\gamma}{\mu^2}\mathbb{V}[\hat{\mu}] - \frac{\varepsilon}{2\gamma(1-\varepsilon)^2}\mathbb{V}[\hat{\gamma}]\right]$
A(1)	$\mathbb{E}[\log y]$	$1 - \frac{1}{\hat{\mu}}\prod_{i\in s} y_i^{w_i/N}$	$\frac{\sum_{i\in s} w_i \log y_i}{N}$	$1 - \frac{\exp\{\hat{\gamma}\}}{\hat{\mu}}$	$\frac{\exp\{\gamma\}}{\mu^2}\left[Cov[\hat{\gamma}, \hat{\mu}] - \frac{\mu}{2}\mathbb{V}[\hat{\gamma}] - \frac{1}{\mu}\mathbb{V}[\hat{\mu}]\right]$
CV	$\mathbb{E}[y^2]$	$\left[\frac{n}{N(n-1)}\sum_{i\in s} w_i \frac{y_i^2}{\hat{\mu}^2} - 1\right]^{\frac{1}{2}}$	$\frac{\sum_{i\in s} w_i y_i^2}{N}$	$\left[\frac{n}{n-1}\frac{\hat{\gamma} - \hat{\mu}^2}{\hat{\mu}^2}\right]^{\frac{1}{2}}$	$\sqrt{\frac{n}{n-1}}\frac{1}{\mu^3}\left(\frac{\gamma}{\mu^2} - 1\right)^{-\frac{3}{2}}\left[\mathbb{V}[\hat{\mu}]\frac{\gamma}{2\mu}\left(\frac{2\gamma}{\mu^2} - 3\right) - -Cov[\hat{\gamma}, \hat{\mu}]\left(\frac{\gamma}{2\mu^2} - 1\right) - \frac{1}{8\mu}\mathbb{V}[\hat{\gamma}]\right]$

Table 1: Relevant quantities for each measure including the approximate bias.

We define the shape of an unbiased estimator for the variance of Horvitz-Thompson estimators, such as $\hat{\mu} = \sum_{i \in s} w_i y_i / N$, when $w_i = 1/\pi_i$, as stated by Arnab (2017, pg. 30)

$$\hat{\mathbb{V}}[\hat{\mu}] = \frac{1}{N^2}\left(\sum_{i \in s} y_i^2 \frac{1-\pi_i}{\pi_i^2} + \sum_{i \in s}\sum_{k \in s, i \neq k} y_i y_k \frac{\pi_{ik} - \pi_i \pi_k}{\pi_{ik}}\right), \tag{7}$$

with π_{ik}, $\forall i, k \in \mathcal{U}, i \neq k$ denoting the second-order inclusion probabilities i.e. the probability that the sample includes both i-th and k-th units. However generally (a) $w_i \neq 1/\pi_i$ and (b) π_{ik}, $\forall i, k \in \mathcal{U}, i \neq k$ are difficult to calculate under complex sampling designs.

Therefore, the variance estimator to be considered constitutes an approximation relying on simplified assumptions. Firstly, we assume that Primary Sampling Units (PSU) are sampled with replacement, and secondly we reduce multi-stage sampling into a single-stage process by relying on the Ultimate Clusters technique (Kalton, 1979). Moreover, we take into account the hybrid nature of the probability scheme, blending a variance estimator for stratified design associated with the SR strata, including a finite population correction factor, and a typical Ultimate Cluster variance estimator for multi-stage schemes associated with the NSR strata. The latter one is widely used in official statistics, see Osier et al. (2013) for Eurostat procedures. Therefore, without loss of generality, consider a two-stage scheme, let $\hat{\mu} = \sum_h \sum_d \sum_i \tilde{w}_{hdi} y_{hdi}$ and $\tilde{w}_{hdi} = w_{hdi}/N$ with h stratum indicator, d Primary Sampling Unit (PSU) indicator and i Secondary Sampling Unit indicator (SSU), be a linear estimator for μ, its standard error estimate is as follows:

$$\begin{aligned}\hat{\mathbb{V}}[\hat{\mu}] &= \sum_{h=1}^{H_{SR}} \mathbb{V}[\hat{\mu}_h] + \sum_{h=1}^{H_{NSR}} \mathbb{V}[\hat{\mu}_h] \\ &= \sum_{h=1}^{H_{SR}} M_h^2 (1-f_h) \frac{s_h^2}{m_h} + \sum_{h=1}^{H_{NSR}} n_h s_{\hat{\mu}_h}^2 \\ &= \sum_{h=1}^{H_{SR}} M_h \frac{M_h - m_h}{m_h(m_h-1)} \sum_{i=1}^{m_h} (y_{hi} - \bar{y}_h)^2 + \sum_{h=1}^{H_{NSR}} \frac{n_h}{n_h - 1} \sum_{d=1}^{n_h} (\hat{\mu}_{hd} - \bar{\mu}_h)^2, \end{aligned} \tag{8}$$

with H_{SR} self-representative and H_{NSR} non self-representative strata, M_h the number of resident households in strata h, m_h the number of sample households in strata h, $f_h = m_h/M_h$ a finite population correction factor, n_h the number of PSUs in strata h. Consider, moreover, that $\bar{y}_h = \sum_{i=1}^{m_h} y_{hi}/m_h$, $\hat{\mu}_{hd} = \sum_{i=1}^{m_d} \tilde{w}_{hdi} y_{hdi}$ with i denoting the household label and m_d the number of sample households in PSU d, lastly $\bar{\mu}_h = \sum_{d=1}^{n_h} \hat{\mu}_{hd}/n_h$, with n_h being the number of PSU in stratum h. Obviously if $n_h = 1$ for some strata, the estimator (8) cannot be used. A solution is to collapse strata to create "pseudo-strata" so that each pseudo-stratum has at least two PSUs. Common practice is to collapse a stratum with another one that is similar with respect to some survey target variables (Rust and Kalton, 1987).

An estimator of $\mathbb{V}[\hat{\gamma}]$ can be obtained by adopting the same strategy used for $\mathbb{V}[\hat{\mu}]$ in (8). Whereas, as regards the estimation of the design covariance, consider that

$$Cov[\hat{\gamma}, \hat{\mu}] = \frac{1}{2}\left(\mathbb{V}[\hat{\gamma} + \hat{\mu}] - \mathbb{V}[\hat{\gamma}] - \mathbb{V}[\hat{\mu}]\right). \tag{9}$$

Thus, a possible estimator $\hat{Cov}[\hat{\gamma}, \hat{\mu}]$ would be simply obtained by plugging in the variance estimators previously mentioned, while $\mathbb{V}[\hat{\gamma} + \hat{\mu}]$ could be estimated by considering $\hat{\gamma} + \hat{\mu} =$

$\sum_{i \in s} w_i(g(y_i) + y_i)/N$. The estimation procedure is completed by replacing μ and γ with $\hat{\mu}$ and $\hat{\gamma}$.

5 Design-Based Simulation

A design-based simulation study has been carried out to evaluate our bias correction proposal. In this simulation, the cross-section Italian EU-SILC sample (2017 wave) has been assumed as pseudo-population and the 21 NUTS-2 regions have been considered as target domains. The study is based on real income data, in order to check whether this specific framework can work with close-to-reality data, affected by peculiar problems (e.g. extreme values, skewness).

For comparison purposes, two simulation scenarios have been carried out. In the first one, the original income data are employed as pseudo-population. In the second one, an extreme values treatment is performed, with reference to both upper and lower tails, to circumvent non-robustness problems and the resulting dataset is specified as an alternative pseudo-population. Subsequently, we compare the results obtained after the treatment with the ones before treatment to isolate the effect of outliers when evaluating bias-correction performances (Table 2).

The issue of robust estimation of economic indicators based on a semi-parametric Pareto upper tail model is well-established in literature. See Brzezinski (2016) for a review and Alfons et al. (2013) for a specification suitable for survey data. On the contrary, the issue of robust treatment of outliers in the lower tail of income distribution appears less established (Masseran et al., 2019; Van Kerm, 2007). Concerning the upper tail, we operated a semi-parametric Pareto-tail modelling procedure using the Probability Integral Transform Statistic Estimator (PITSE) proposed by Finkelstein et al. (2006), which blends very good performances in small samples and fast computational implementation, as suggested by Brzezinski (2016). As regards the lower tail, we used an inverse Pareto modification of the PITSE estimator, suggested by Masseran et al. (2019). In our simulations, the treatment has been done at a regional level to the original EU-SILC sample and the detection of outliers has been carried out following Safari et al. (2018) by using a Generalized Boxplot outlier detection procedure. We expect that, when outlying observations are representative, this procedure would highly bias the outcome and thus we do not recommend it.

From the pseudo-populations, we repeatedly select 1,000 two-stage stratified samples, mimicking the sampling strategy adopted in the survey itself: in the first stage, SR strata are always included in the sample, while a stratified sample of PSU in NSR strata is selected; in the second stage, a systematic sample of households is drawn from each PSU included at the first stage. We repeated the drawing for both scenarios involving different sampling rates, 1,5% and 3% respectively. The Relative Bias (RB) and the Absolute Relative Error (ARE) in percentage has been calculated for each region r using

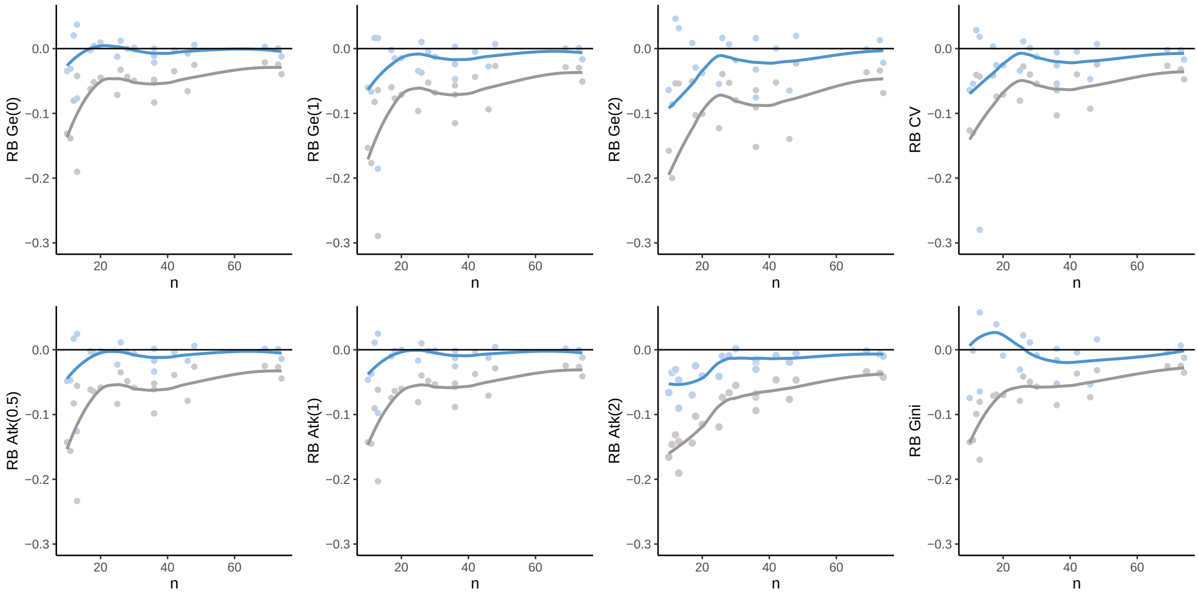

Figure 1: Relative Bias of non-corrected measures (grey line), and of corrected measures (blue line) in 3% samples after extreme value treatment versus the (average) sample size.

the 1,000 iterations as:

$$\mathtt{RB}_r = \frac{1}{1,000}\sum_{p=1}^{1,000}\left(\frac{\hat{\theta}_{p,r}}{\theta_r} - 1\right),$$

$$\mathtt{ARE}_r = \frac{1}{1,000}\left(\sum_{p=1}^{1,000}\left|\frac{\hat{\theta}_{p,r}}{\theta_r} - 1\right|\right),$$

where θ_r is the population value for region r and $\hat{\theta}_{i,r}$ is its estimate for the generic iteration p. In our simulation setting, the regional sample size ranges from 6 to 96 individuals (from 6 to 32 households) for the 1,5% sampling rate, and from 11 to 196 individuals (10 to 74 households) for the 3% sampling rate.

Concerning the treated pseudo-population scenario, Figure 1 illustrates the relative bias for each domain of non-corrected measures (gray line) of corrected measures (blue line) in 3% samples versus the (average) sample size. The negative relation between sample size and average relative bias is clear for both the design-based estimator $\hat{\theta}$ and the bias corrected estimator $\hat{\theta}_{corr}$. This confirm the nature of the bias as a small sample bias and shows the effectiveness of the correction, even if based on Taylor's expansion, known to be a large-n approximation. The bias reduction is noticeable for all measures, leading to slightly biased estimates depending on the measure. Notice that the bias correction works well for measures not particularly sensitive to extreme observations such as Gini index, GE(0), Atk(0.5) and Atk(1). In case of CV and GE(2), the correction provides good results, but it seems, however, not to capture all the bias components. This may confirm the results of Breunig (2001), suggesting that the coefficient of variation squared and GE(2) bias depends on the coefficient of skewness of the income distribution, not considered in our bias correction.

		CV	GE(0)	GE(1)	GE(2)	A(0.5)	A(1)	A(2)	Gini
with Extreme Value treatment									
1.5%									
$\hat{\theta}$	$\bar{\text{RB}}$	-11.9	-13.9	-16.0	-17.6	-14.9	-15.0	-19.0	-14.6
	$\bar{\text{ARE}}$	25.8	44.1	42.9	47.9	41.9	40.9	38.7	24.5
$\hat{\theta}_{corr}$	$\bar{\text{RB}}$	-5.6	-4.1	-6.8	-9.0	-5.4	-5.2	-9.4	0.5
	$\bar{\text{ARE}}$	25.9	46.4	44.9	50.2	43.9	42.8	39.6	34.3
3.0%									
$\hat{\theta}$	$\bar{\text{RB}}$	-7.4	-6.6	-8.6	-10.6	-7.6	-7.3	-9.7	-7.2
	$\bar{\text{ARE}}$	19.8	32.0	32.0	38.0	30.6	29.6	28.2	16.6
$\hat{\theta}_{corr}$	$\bar{\text{RB}}$	-2.8	-0.6	-2.3	-3.6	-1.5	-1.2	-2.9	0.3
	$\bar{\text{RB}}$ $(n \geq 20)$	-1.4	-0.4	-1.2	-1.6	-0.8	-0.6	-1.4	-0.9
	$\bar{\text{ARE}}$	20.4	33.4	33.7	40.6	32.0	30.4	29.6	19.7
without Extreme Value treatment									
1.5%									
$\hat{\theta}$	$\bar{\text{RB}}$	-18.2	-12.7	-17.5	-23.3	-15.3	-15.6	-48.0	-14.9
	$\bar{\text{ARE}}$	30.0	52.9	46.4	53.5	45.9	47.4	56.8	25.5
$\hat{\theta}_{corr}$	$\bar{\text{RB}}$	-12.1	-3.9	-8.7	-15.0	-6.3	-5.9	-41.6	0.04
	$\bar{\text{ARE}}$	29.3	54.4	48.0	55.7	47.5	49.4	54.6	34.7
3.0%									
$\hat{\theta}$	$\bar{\text{RB}}$	-12.7	-6.8	-10.5	-15.8	-8.7	-8.4	-38.1	-7.9
	$\bar{\text{ARE}}$	24.5	39.4	36.0	46.2	34.3	35.6	49.0	17.7
$\hat{\theta}_{corr}$	$\bar{\text{RB}}$	-7.8	-1.2	-3.9	-8.0	-2.5	-2.0	-32.4	0.06
	$\bar{\text{RB}}$ $(n \geq 20)$	-8.3	-1.2	-3.6	-8.7	-2.3	-1.7	-30.0	-1.2
	$\bar{\text{ARE}}$	24.8	40.4	37.9	49.4	35.7	37.0	48.2	20.9

Table 2: Percentage RB and ARE averaged on the 21 regions explicited for each inequality estimator and scenario.

Bias and error averaged across all areas for each scenario, sampling rate and estimator are shown in Table 2. By focusing again on treated population results, the correction induces a reduction of the `RB` spanning from 5% (CV, 3% rate) to 14% (Gini, 1.5% rate) approximately by considering both sampling rates. When the sample size is greater than 20 individuals ($n \geq 20$), the bias-corrected estimators seems to be approximately unbiased. Furthermore, the bias correction induces a slight but negligible error (`ARE`) increase on average, except for the Gini index case which presents a relevant increase. This is due to the shape of the unbiased estimators, as described by (6), where a sum of estimators is multiplied by a factor $n/(n-2)$, which inherently inflates the variance by its square.

Moving to the non-treated population results, the bias and error increase dramatically both for $\hat{\theta}$ and for $\hat{\theta}_{corr}$. In particular, the bias is high for some measures estimated on non-treated data due to the non-robustness properties to extreme values. It is the case of Atk ($\varepsilon = 2$), extremely sensitive to low-income values (under 100 euro per year) which is -48% biased on average for the scenario with the smallest sample sizes. Also GE with α equal to 1 and 2 are highly sensitive to high-income values being -18% and -23% biased. Moreover, the bias correction seems not to change in magnitude depending on the sample size and the presence of extreme values and no relevant changes in the comparison of $\hat{\theta}$ and $\hat{\theta}_{corr}$ are recorded.

Summarizing, in case of population not affected by income extreme values, the bias correction may provide approximately unbiased estimates for a large class of measures. Vice versa, it becomes necessary to restrict the attention to the most robust measures such as GE with $\alpha = 0$, Atkinson index with $\varepsilon = 1$ and Gini Index. Another important aspect to point out is that, in certain countries, the EU-SILC is based on registers that better capture top incomes, thus, a cross-country comparison of income inequality by effects on a tail-sensitive measure must be another reason for caution (Atkinson, 2015). Such results may constitute a reference when measuring inequality in small samples, however, since the simulation scenario uses very specific data, reflections that has been drawn cannot be general or conclusive.

5.1 Monte Carlo Distributions of Corrected Estimators

Lastly, a brief analysis of the distribution of inequality estimators is carried out, considering samples of increasing size, in order to evaluate how quickly their distribution tends to become symmetric. We consider regions as target domains and we keep the same simulation setting with different sampling rates, i.e. 10% (from 36 to 607 individuals, 28 to 254 households), 5% (from 16 to 337 individuals, 14 to 131 households) and 3% (from 11 to 196 individuals, 10 to 74 households). The coefficient of skewness η_3 and excess kurtosis η_4 empirical values are set out in Table 3 while such empirical distributions are depicted for each region in Figure 2. As clear from Table 3 and Figure 2, the empirical distributions tend to become more positively skewed and leptokurtic at decreasing sample sizes. This is quite evident for the General Entropy measures, similarly but to a lesser extent for the other measures.

			CV	GE(0)	GE(1)	GE(2)	A(0.5)	A(1)	A(2)	Gini
10%	$\hat{\theta}_{corr}$	$\hat{\eta}_3$	0.81	0.66	0.88	1.30	0.71	0.54	0.26	0.38
		$\hat{\eta}_4$	1.16	0.91	1.60	2.87	1.06	0.57	-0.01	0.34
5%	$\hat{\theta}_{corr}$	$\hat{\eta}_3$	1.00	0.94	1.19	1.80	0.96	0.74	0.40	0.54
		$\hat{\eta}_4$	1.88	1.73	2.81	6.09	1.81	1.01	0.08	0.56
3%	$\hat{\theta}_{corr}$	$\hat{\eta}_3$	1.00	1.07	1.27	1.90	1.04	0.83	0.49	0.58
		$\hat{\eta}_4$	1.66	1.94	2.87	6.13	1.87	1.03	0.03	0.60

Table 3: Coefficients of skewness and excess kurtosis of the Monte Carlo distributions related to corrected inequality estimators.

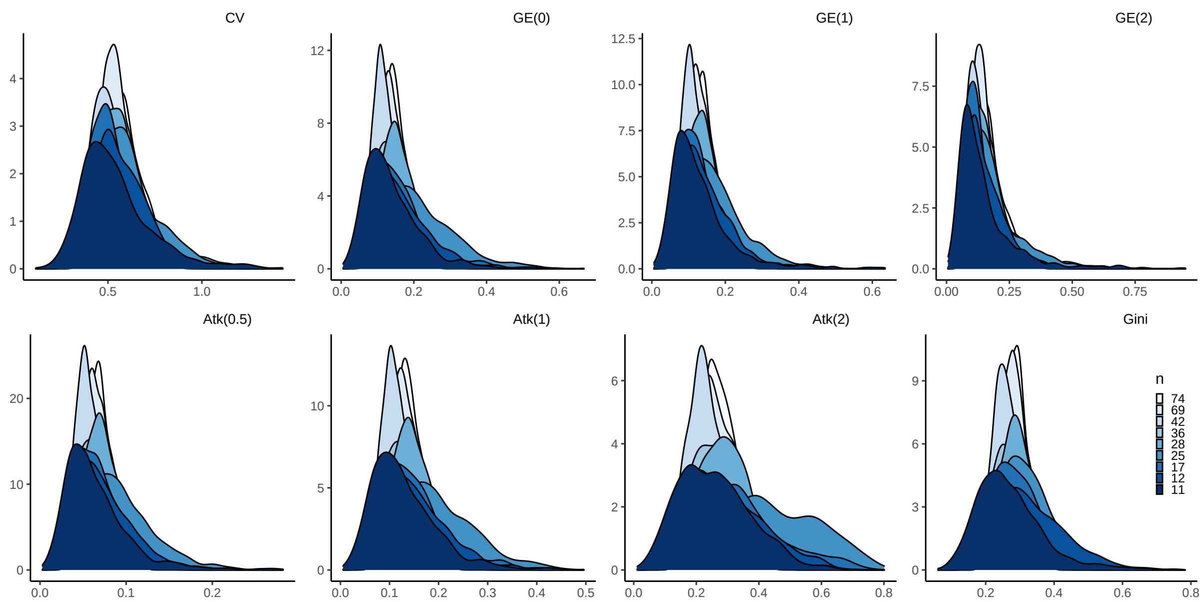

Figure 2: Monte Carlo distributions of corrected inequality measure with 3% sampling rate.

6 A Small Area Estimation Exercise

In the previous sections, we propose a method to correct small sample bias of inequality estimators in complex surveys. Even if bias-corrected, such estimators are still unreliable due to the high variability induced by the small sample size: this means that their estimates cannot be released or used for further inference. As a consequence, when measuring inequality at a fine-grained level, it becomes necessary to rely on Small Area Estimation (SAE) techniques. Such estimation techniques exploit available auxiliary information to produce estimates with acceptable uncertainty. In particular, the model-based techniques employ hierarchical models which can be defined both at area level, linking area-defined survey estimates with areal covariates, or at unit (individual) level, linking individual income data with individual covariates. See Tzavidis et al. (2018) for an up-to-date review.

In this context, area level models appear to be less demanding in terms of data requirement as they exploit only areal covariates, whereas unit level models may require individual observations pertaining to the whole population. Moreover, area level models enable the incorporation of design-based properties. Such models constitute a typical framework of application of our bias-correction proposal, as they assume unbiasedness of the survey estimators used as input. As a consequence, their applicability to the estimation of inequality measures is inevitably tied to a preliminary bias correction, in contrast with unit level models that do not handle survey estimators.

In this section, we perform a SAE exercise by using the sample related to the first iteration of the simulation detailed in Section 5. The purpose is not to propose a small area estimation strategy nor to provide a real application of inequality mapping, but rather to illustrate the framework of application of our bias-correction proposal and, especially, to underline the risk of avoiding bias-correction when estimating inequality in small domains. Such exercise is carried out by applying the well known Fay-Herriot model, implemented through the package `sae` (Molina and Marhuenda, 2015) to both the uncorrected and corrected survey estimators of a selected subset of measures: the Theil index (Generalized Entropy with $\alpha = 1$), the Atkinson index with $\varepsilon = 1$ and the Gini index.

Specifically, let us consider $\hat{\theta}_1, \dots, \hat{\theta}_M$ as the set of survey estimators referring to a generic inequality measure in M small areas, with corresponding population values $\theta_1, \dots, \theta_M$, and $\boldsymbol{x}_m$ the set of p areal covariates for area m, $m = 1, \dots, M$. The classical area-level model is the Fay and Herriot (1979) one, with the following assumptions:

$$\hat{\theta}_m \sim \mathcal{N}(\theta_m, D_m), \tag{10}$$

$$\theta_m \sim \mathcal{N}(\boldsymbol{x}_m^T \boldsymbol{\beta}, \sigma^2), \qquad m = 1, \dots, M. \tag{11}$$

where D_m denotes the sampling variance of the survey estimator, usually assumed to be known in order to allow for identifiability, $\boldsymbol{\beta}$ the set of regression coefficients and σ^2 the model variance. This clearly implies $\mathbb{E}(\hat{\theta}_m) = \theta_m \ \forall m$, i.e. the unbiasedness of survey estimators. As a consequence, neglecting bias correction of survey estimators effectively leads to model misspecification.

As mentioned above, the sampling variance is separately estimated from the data and given as input to the small area model. Since our exercise is merely illustrative, we adopt the Monte Carlo variances of the design-based simulation in Section 5 as sampling variances and simulated covariates for both estimators. However, in real application, variance estimation is the crux of a SAE procedure. In the case of raw inequality estimators, it may be easily carried out via linearization. Linearized variables for each measure could be derived consistently with Langel and Tillé (2013) for Gini index and Biewen and Jenkins (2006) for Generalized Entropy and Atkinson indexes. On the other hand, the variance of bias corrected estimators adds a new level of complexity since the estimator formula is no longer the classical one. Indeed, it comprises a bias correction component that appears cumbersome to estimate via linearization since it is inherently a result of several linearizations. Therefore, in real applications, we recommend relying on resampling methods, an example is the design-aware bootstrap procedure developed before by Fabrizi et al. (2011, 2020). A comprehensive review on the use of bootstrap methods for survey data can be found in Lahiri (2003) and an interesting comparison between variance estimation techniques for poverty and inequality measures has been carried out by De Santis et al. (2020).

The comparison between raw and corrected survey estimates for all the three measures is displayed in Figure 3. Raw estimates show lower values of inequality in comparison with the corrected ones for all areas and all the measures considered. This is in accordance with the underestimation highlighted by simulation results of Section 5. The sampling coefficient of variations of both estimators are high, ranging from 0.24 to 3.48 for the Theil index, from 0.18 to 4.45 for the Atkinson index and, lastly from 0.11 to 0.92 for the Gini index, with slightly higher values in case of corrected estimators, as the bias correction induces mild variance inflation. Such values point out the need for SAE techniques.

The model-based (or EBLUP) estimates in both cases are compared in Figure 4, the inequality levels estimated by the misspecified model shows noticeable lower values, resulting in a misleading inference. In particular, this is quite evident for the Gini index case, where the divergence seems to increase at increasing levels of inequality. By focusing only on the EBLUP results based on corrected estimates, the decrease in terms of estimates error induced by the model is depicted in Figure 5. The reduction is great and testifies that the variance reduction procedure, put in place by the SAE model, is effective. As a consequence, model-based estimates result to be reliable and ready to be used for further analysis or mapping.

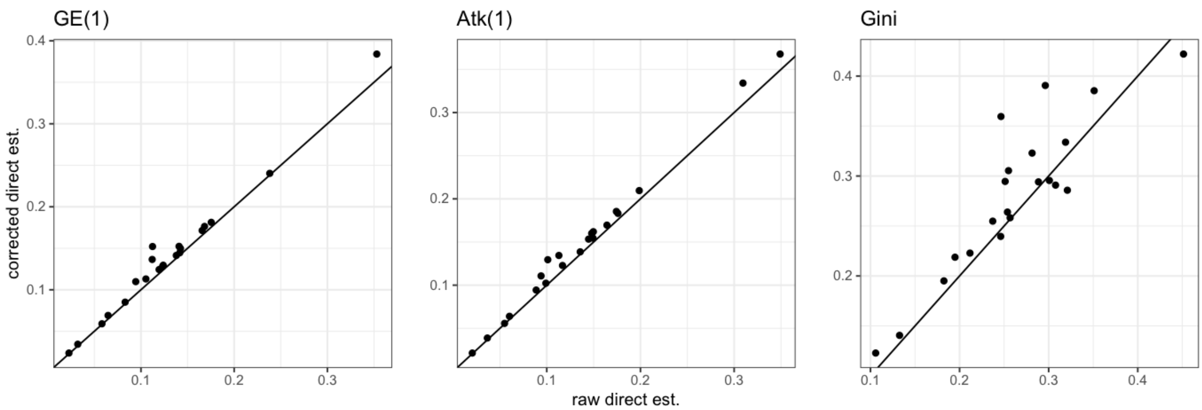

Figure 3: Bias corrected survey estimates versus raw survey estimates. Bisector line in black.

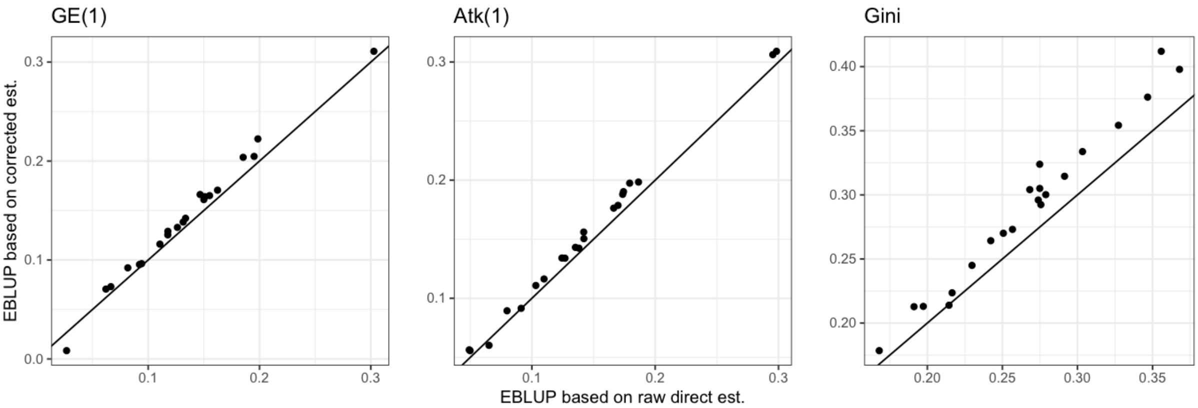

Figure 4: Model-based estimates based on bias corrected survey estimates versus model-based estimates based on raw survey estimates. Bisector line in black.

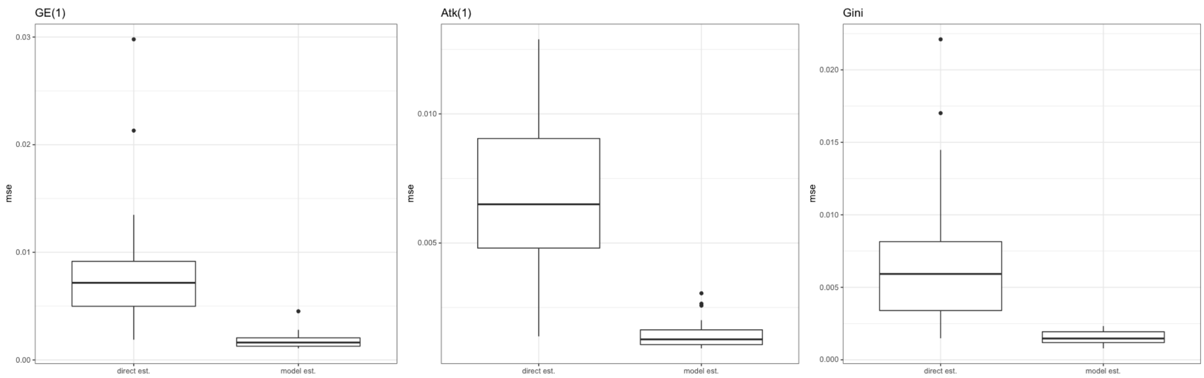

Figure 5: MSE of bias corrected survey estimators versus their related model-based estimators.

7 Conclusions

A strategy based on Taylor's expansion has been proposed to correct the small sample bias of inequality estimators. The sets of inequality measures considered is large, as the comparison of diverse measures may enable to enlighten the peculiar point of view each measure provides, as single tiles in a mosaic. Indeed, the well-known Gini and Theil indexes are widely applied in several fields for inequality and concentration estimation.

A sensitivity analysis and simulation has also been conducted to study the estimator behaviour to extreme values and the magnitude of the proposed correction. Results show that survey-based estimator may be biased in small samples, inducing often an underestimation of inequality. Such underestimation is greater in case of populations affected by extreme values. Moreover, simulation results validate the correction proposal as effective, consistently reducing the bias and leading in some cases to approximately unbiased estimators.

An underlined heterogeneity of sensitivities and bias is recorded across measures. As a consequence, our results may help in choosing the most suitable inequality measure depending on the context. The measures that are structurally more sensible to extreme values on the tails appear to be more biased, in particular GE($\alpha = 2$) and Atkinson($\epsilon = 2$). Therefore, in case of samples without extreme income values, the bias correction may provide approximately unbiased estimates. On the other hand, if extreme values are observed, it becomes necessary to focus on the most robust measures such as Mean Log Deviation, Atkinson index with $\varepsilon = 1$ and Gini Index to be corrected. Furthermore, the estimator distributions show increasing positive skewness and lepto-kurtosis at decreasing sample sizes.

An illustrative small area application has been carried out. The results obtained shows that neglecting the bias issue translates into a misleading inference. This is particularly evident for the popular Gini index. In such an application, we use a basic area level model, the Gaussian one. Indeed, the skewed distribution of inequality estimators and the unit-interval support of Gini and Atkinson estimators might urge a more refined model, which may lead to model-based estimators with increased performances: this constitutes an interesting future direction of research. Further directions include also the extension of this framework to other widely used inequality measures, such as those based on quintiles.

References

A. Alfons, M. Templ, and P. Filzmoser. Robust estimation of economic indicators from survey samples based on Pareto tail modelling. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 62(2):271–286, 2013. ISSN 00359254. doi: 10.1111/j.1467-9876.2012.01063.x.

R. Arnab. *Survey sampling theory and applications.* Academic Press, 2017.

A. Atkinson. *Handbook of Income Distribution*, volume 2. 2015. ISBN 9780444594303. doi: 10.1016/B978-0-444-59428-0.09990-2.

L. G. Bellu and P. Liberati. Social welfare, social welfare functions and inequality aversion. *Food and Agriculture Organization of the United Nations*, 2006.

M. Biewen and S. P. Jenkins. Variance estimation for generalized entropy and atkinson inequality indices: The complex survey data case. *Oxford Bulletin of Economics and Statistics*, 68(3):371–383, 2006. ISSN 03059049. doi: 10.1111/j.1468-0084.2006.00166.x.

R. Breunig. An almost unbiased estimator of the coefficient of variation. *Economics Letters*, 70(1):15–19, 2001.

R. Breunig and D. L.-A. Hutchinson. Small sample bias corrections for inequality indices. In W. N. Toggins, editor, *New Econometric Modelling Research*, pages 61–83. Nova Science Publishers, 2008.

M. Brzezinski. Robust estimation of the Pareto tail index: a Monte Carlo analysis. *Empirical Economics*, 51(1):1–30, 2016. ISSN 03777332. doi: 10.1007/s00181-015-0989-9.

A. Cavanaugh and S. Breau. Locating geographies of inequality: publication trends across oecd countries. *Regional Studies*, 52(9):1225–1236, 2018.

L. Ceriani and P. Verme. Individual diversity and the gini decomposition. *Social Indicators Research*, 121(3):637–646, 2015.

F. A. Cowell and M.-P. Victoria-Feser. Robustness properties of inequality measures. *Econometrica: Journal of the Econometric Society*, 64(1):77–101, 1996.

R. Davidson. Reliable inference for the Gini index. *Journal of Econometrics*, 150(1): 30–40, 2009. ISSN 03044076. doi: 10.1016/j.jeconom.2008.11.004. URL `http://dx.doi.org/10.1016/j.jeconom.2008.11.004`.

R. De Santis, L. Barabesi, and G. Betti. Variance estimation techniques for poverty and inequality measures from complex surveys: a simulation study. Technical report, Department of Economics, University of Siena, 2020.

G. Deltas. The small-sample bias of the Gini coefficient: Results and implications for empirical research. *Review of Economics and Statistics*, 85(1):226–234, 2003. ISSN 00346535. doi: 10.1162/rest.2003.85.1.226.

E. Fabrizi and C. Trivisano. Small area estimation of the gini concentration coefficient. *Computational Statistics & Data Analysis*, 99:223–234, 2016.

E. Fabrizi, M. R. Ferrante, S. Pacei, and C. Trivisano. Hierarchical bayes multivariate estimation of poverty rates based on increasing thresholds for small domains. *Computational Statistics & Data Analysis*, 55(4):1736–1747, 2011.

E. Fabrizi, M. R. Ferrante, and C. Trivisano. A functional approach to small area estimation of the relative median poverty gap. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183(3):1273–1291, 2020.

R. E. Fay and R. A. Herriot. Estimates of income for small places: an application of james-stein procedures to census data. *Journal of the American Statistical Association*, 74(366a):269–277, 1979.

M. Finkelstein, H. G. Tucker, and J. Alan Veeh. Pareto Tail Index Estimation Revisited. *North American Actuarial Journal*, 10(1):1–10, 2006. ISSN 10920277. doi: 10.1080/10920277.2006.10596236.

D. E. Giles. The bias of inequality measures in very small samples: some analytic results. Technical report, Department of Economics, University of Victoria, 2005.

M. Graf. Use of survey weights for the analysis of compositional data. *Compositional data analysis: theory and applications*, 2011.

A.-C. Guio. Income poverty and social exclusion in the eu25. *Statistics in Focus: Population and Social Conditions*, (13):1–7, 2005.

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

N. Kakwani. Large sample distribution of several inequality measures: with application to cote d'ivoire. In *Contributions to Econometric Theory and Application*, pages 50–81. Springer, 1990.

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

L. Kim, C. Adolph, J. D. West, and K. Stovel. The influence of changing marginals on measures of inequality in scholarly citations: Evidence of bias and a resampling correction. *Sociological Science*, 7:314–341, 2020.

P. Lahiri. On the impact of bootstrap in survey sampling and small-area estimation. *Statistical Science*, pages 199–210, 2003.

M. Langel and Y. Tillé. 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, 2013. ISSN 09641998. doi: 10.1111/j.1467-985X.2012.01048.x.

R. I. Lerman and S. Yitzhaki. Improving the accuracy of estimates of gini coefficients. *Journal of Econometrics*, 42(1):43–47, 1989.

N. Masseran, L. H. Yee, M. A. M. Safari, and K. Ibrahim. Power law behavior and tail modeling on low income distribution. *Mathematics and Statistics*, 7(3):70–77, 2019.

G. Mohler, P. J. Brantingham, J. Carter, and M. B. Short. Reducing bias in estimates for the law of crime concentration. *Journal of quantitative criminology*, 35(4):747–765, 2019.

I. Molina and Y. Marhuenda. sae: An r package for small area estimation. *The R Journal*, 7(1):81, 2015.

OECD. *Divided We Stand: Why Inequality Keeps Rising*. OECD Publishing Paris, 2011.

G. Osier, Y. G. Berger, and T. Goedeme. Standard error estimation for the eu-silc indicators of poverty and social exclusion. *Eurostat Methodologies and Working Papers Series*, 2013.

K. Rust and G. Kalton. Strategies for collapsing strata for variance estimation. *Journal of Official Statistics*, 3(1):69, 1987.

M. A. M. Safari, N. Masseran, and K. Ibrahim. Outliers detection for pareto distributed data. In *American Institute of Physics Conference Series*, volume 1940, 2018.

C.-E. Särndal, B. Swensson, and J. Wretman. *Model Assisted Survey Sampling*. Springer Science & Business Media, 2003.

C. Schluter and K. J. van Garderen. Edgeworth expansions and normalizing transforms for inequality measures. *Journal of Econometrics*, 150(1):16–29, 2009. ISSN 03044076. doi: 10.1016/j.jeconom.2008.12.022. URL `http://dx.doi.org/10.1016/j.jeconom.2008.12.022`.

A. K. Sen. *On Economic Inequality*. Oxford: Clarendon Press, 1997.

N. Tzavidis, L.-C. Zhang, A. Luna, T. Schmid, and N. Rojas-Perilla. From start to finish: a framework for the production of small area official statistics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(4):927–979, 2018.

A.-A. Vallée and Y. Tillé. Linearisation for variance estimation by means of sampling indicators: Application to non-response. *International Statistical Review*, 87(2):347–367, 2019.

P. Van Kerm. Extreme incomes and the estimation of poverty and inequality indicators from eu-silc. Technical report, IRISS Working Paper Series 2007-01, CEPS/INSTEAD, 2007.

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

Appendix

Proof of Proposition 1

Proof. Let us consider a sample with *iid* elements $\{y_1, \dots, y_{n_{iid}}\}$, drawn from a population via simple random sampling, where y_i is the variable for the i-th unit with expected value μ and variance σ^2. Let us consider also $\{g(y_1), \dots, g(y_{n_{iid}})\}$ with $g(y)$ a generic monotone transformation of the income variable, induced by $g(\cdot) : \mathbb{R}^+ \to \mathbb{R}$, that changes for each measure, having expected value γ and variance ϕ^2. Considering that a generic inequality measure can be expressed as $\theta = f(\mu, \gamma)$ with $f(\cdot)$ a twice-differentiable function, $\hat{\mu} = \sum_{i=1}^{n_{iid}} y_i/n_{iid}$ and $\hat{\gamma} = \sum_{i=1}^{n_{iid}} g(y_i)/n_{iid}$, we can easily obtain estimator moments as $\hat{\mu} \sim [\mu, \sigma^2/n_{iid}]$ and $\hat{\gamma} \sim [\gamma, \phi^2/n_{iid}]$. Consider moreover that

$$Cov[\hat{\mu}, \hat{\gamma}] = \mathbb{E}[\hat{\mu}\hat{\gamma}] - \mu\gamma = \frac{1}{n_{iid}}(\mathbb{E}[y \cdot g(y)] - \mu\gamma) = \frac{Cov[y, g(y)]}{n_{iid}}.$$

Let us define the population value of a generic inequality measure θ as $f(\mu, \gamma)$, with $f(\cdot)$ a generic twice-differentiable function. By expanding the inequality measure estimator $\hat{\theta}$ as $f(\hat{\mu}, \hat{\gamma})$, via Taylor's expansion around the population values and considering its expected value:

$$\begin{aligned}\mathbb{E}[\hat{\theta}] &= \theta + \frac{1}{2} f_{\gamma,\gamma}(\gamma, \mu)\mathbb{V}[\hat{\gamma}] + f_{\gamma,\mu}(\gamma, \mu) Cov[\hat{\gamma}, \hat{\mu}] + \frac{1}{2} f_{\mu,\mu}(\gamma, \mu)\mathbb{V}[\hat{\mu}] + O(n_{iid}^{-2}) \\ &= \theta + O(n_{iid}^{-1}) + O(n_{iid}^{-1}) + O(n_{iid}^{-1}) + O(n_{iid}^{-2}) \\ &= \theta + O(n_{iid}^{-1}),\end{aligned}$$

where $f_\gamma = \frac{\partial f(\gamma,\mu)}{\partial \gamma}$ and $f_{\gamma\mu} = \frac{\partial^2 f(\gamma,\mu)}{\partial\gamma\partial\mu}$. □

Reaching results of Table 1 in a different way

By considering the linearization proposed by Graf (2011) and extended by Vallée and Tillé (2019), we can easily reach the same results set out in Table 1 in a different way.

Let us recall the notation as $\mathcal{U}$ denoting a finite population of $N(< \infty)$ elements. Let y_i be the income of the i-th unit, where $y_i \in \mathbb{R}^+$, $\forall i = 1, \dots, N$ and $\boldsymbol{y} = (y_1, \dots, y_N)$ its population vector. A sample s of size n is drawn with a complex sampling design with probability of selection $p(s)$ such that $p(s) \geq 0$ and $\sum_{s \subset \mathcal{U}} p(s) = 1$. Let us define $\mathbb{1} = (\mathbb{1}_1, \dots, \mathbb{1}_N)$ the vector of sampling indicator of unit i, taking value 1 if unit i is in the sample and 0 otherwise. The first order inclusion probability of unit i is π_i, where $\pi_i = \mathbb{E}(\mathbb{1}_i)$, i.e. the expectation with respect to the sampling design and its population vector $\boldsymbol{\pi} = (\pi_1, \dots, \pi_N)$. π_{ij} denotes the second order inclusion probability for $i \neq j$.

Consider $\hat{\theta} = \hat{\theta}(\mathbb{1}, \boldsymbol{y})$ an estimator of $\theta = \theta(\boldsymbol{y})$ with $\hat{\theta}(\mathbb{1}, \boldsymbol{y})$ twice differentiable with respect to $\mathbb{1}$. Graf (2011) shows that an approximation for $\hat{\theta}$ is

$$\hat{\theta} \approx \theta + \sum_{i \in \mathcal{U}} (\mathbb{1}_i - \pi_i) \frac{\partial \hat{\theta}}{\partial \mathbb{1}_i}\Big|_{\mathbb{1}=\boldsymbol{\pi}} + \frac{1}{2} \sum_{i \in \mathcal{U}} \sum_{j \in \mathcal{U}} (\mathbb{1}_i - \pi_i)(\mathbb{1}_j - \pi_j) \frac{\partial^2 \hat{\theta}}{\partial \mathbb{1}_i \partial \mathbb{1}_j}\Big|_{\mathbb{1}=\boldsymbol{\pi}}.$$

We can derive the approximation of the bias as

$$\mathbb{E}[\hat{\theta}] \approx \theta + 0 + \frac{1}{2} \sum_{i \in \mathcal{U}} \sum_{j \in \mathcal{U}} (\pi_{ij} - \pi_i\pi_j) \frac{\partial^2 \hat{\theta}}{\partial \mathbb{1}_i \partial \mathbb{1}_j}\Big|_{\mathbb{1}=\boldsymbol{\pi}}. \tag{12}$$

In the following, we derive the bias for GE(α) with $\alpha \neq 0, 1$ but such result can be extended to all the measures pertaining to Generalized Entropy and Atkinson indexes families.

Let us recall from the manuscript the survey estimator of GE(α) as a function of two Horvitz-Thompson type estimators which, under the assumption of $w_i = 1/\pi_i$, can be rewritten as

$$\hat{\mu} = \frac{1}{N} \sum_{i \in \mathcal{U}} \frac{\mathbb{1}_i y_i}{\pi_i} \quad \text{and} \quad \hat{\gamma} = \frac{1}{N} \sum_{i \in \mathcal{U}} \frac{\mathbb{1}_i y_i^\alpha}{\pi_i},$$

with GE(α) estimator explicited as

$$\hat{\theta}_{GE}(\alpha) = \frac{n(n-1)^{-1}}{\alpha(\alpha-1)} \left(\frac{\hat{\gamma}}{\hat{\mu}^\alpha} - 1 \right).$$

By applying (12), its bias may be expressed with an approximate result as

$$\mathbb{E}[\hat{\theta}_{GE}(\alpha)] - \theta_{GE}(\alpha) \approx \frac{n(n-1)^{-1}}{2\mu^{\alpha+1}(\alpha-1)} \frac{1}{N^2} \sum_{i \in \mathcal{U}} \sum_{j \in \mathcal{U}} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) \left(y_i y_j \frac{\gamma(\alpha+1)}{\mu} - y_i y_j^\alpha - y_i^\alpha y_j \right). \tag{13}$$

Considering that variance and covariance of Horvitz-Thompson estimator in population are defined as

$$\mathbb{V}[\hat{\mu}] = \frac{1}{N^2} \sum_{i \in \mathcal{U}} \sum_{j \in \mathcal{U}} y_i y_j \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right), \tag{14}$$

$$Cov[\hat{\mu}, \hat{\gamma}] = \frac{1}{N^2} \sum_{i \in \mathcal{U}} \sum_{j \in \mathcal{U}} y_i y_j^\alpha \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right), \tag{15}$$

as stated by Arnab (2017, pg. 30), (13) can be easily rewritten as

$$\mathbb{E}[\hat{\theta}_{GE}(\alpha)] - \theta_{GE}(\alpha) \approx \frac{n(n-1)^{-1}}{\mu^{\alpha+1}(\alpha-1)} \left(\mathbb{V}[\hat{\mu}] \frac{\gamma(\alpha+1)}{2\mu} - Cov[\hat{\mu}, \hat{\gamma}] \right). \tag{16}$$

Such analytical result coincides with the corresponding formula in Table 1 of the manuscript.