

Small-Sample Bias Correction of Inequality Estimators in Complex Surveys

Journal of Official Statistics

1–24

© The Author(s) 2024

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0282423X241244920

journals.sagepub.com/home/jof



**Silvia De Nicolò, Maria Rosaria Ferrante,
and Silvia Pacci**

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 and performing small area estimation at the area level. We propose a bias correction framework for a large class of inequality measures comprising the Gini Index, the Generalized Entropy, and the Atkinson index families by accounting for complex survey designs. The proposed methodology does not require any parametric assumption on income distribution, being very flexible. Design-based performance evaluation of our proposal has been carried out using EU-SILC data, their results show a noticeable bias reduction for all the measures. Lastly, an illustrative example of application in small area estimation confirms that ignoring ex-ante bias correction determines model misspecification.

Keywords

complex surveys, finite populations, income inequality, small area estimation

1. Introduction

The interest in reliable local estimates of economic inequality is growing due to the observed increment in the income gap and social exclusion among regions. Specifically, inequality estimates for specific sub-populations—such as areas at a fine level of geographical disaggregation or rather specific socio-demographic groups—are increasingly in demand (Márquez et al. 2019). Policymakers and stakeholders need these to formulate and implement policies, distribute resources, and measure the effect of policy actions at local levels. In addition, their contribution to regional studies is valuable in the process

Department of Statistical Sciences “P.Fortunati,” University of Bologna, Bologna, Italy

Corresponding author:

Silvia De Nicolò, Department of Statistical Sciences “P.Fortunati,” University of Bologna, Via Belle Arti 41, Bologna 40126, Italy.

Email: silvia.denicolo@unibo.it



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

of decomposing spatial spillovers and identifying local areas that drive inequality at national levels (Cavanaugh and Breau 2018).

When dealing with inequality estimation in specific groups or local scales, a problem of observations scarcity typically arises. Disposable income is generally adopted as the variable of interest and the primary source of data collection is through household surveys. However, since such surveys are not planned for the estimation of target quantities in specific domains, they result in small sample sizes. In this context, small area estimation techniques are applied, integrating survey data with auxiliary data to “borrow strength” across areas and, in this way, improve the reliability of estimates.

The small area models can be specified at the unit (individual or household) level; previous proposals dealing with inequality estimation are provided by Tzavidis and Marchetti (2016) and Marchetti and Tzavidis (2021) by means of robust methods. However, such models require a large amount of data as, generally, the auxiliary variables have to be known for each unit of the population and linked to survey data. This may be hard to get as administrative archives are not publicly accessible at individual level, cross-linked and associated with survey data (Harmening et al. 2023). On the other hand, small area models defined at the area level are less demanding in terms of data requirements, needing only survey (direct) estimates endowed with related measures of uncertainty and areal covariates (Rao and Molina 2015). An application of such models to inequality estimation can be found in Benedetti and Crescenzi (2023).

Area-level models in their classical specification, the Fay-Herriot model (Fay and Herriot 1979), have the strict assumption of (approximate) unbiasedness of the survey estimators given as input (Rao and Molina 2015). However, inequality estimators are biased in small samples, often underestimating inequality (Breunig and Hutchinson 2008; Deltas 2003). In this paper, we focus on such bias which may depend on the non-linear nature of inequality indicators, on the characteristic of the distribution of the variable of interest, that is, the income variable (Breunig 2001), and on the uncertainty induced by the sample selection scheme.

Unfortunately, such an issue is typically neglected when measuring inequality with area-level models, leading to model misspecification and thus to a possible misleading inference. Note that this aspect deserves attention given that estimates of inequality measures are often used for comparisons across time and locations. Neglecting it may bring out discrepancies that, rather than being true inequality gaps, may be due to disparate sample sizes or to different underlying distributions of the variable of interest (Breunig and Hutchinson 2008). In this vein, we propose a bias correction strategy for a large set of inequality measures and we adopt it in an illustrative small area estimation exercise.

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 context of application is varied, spanning from economic inequality to crime or concentration of scholarly citations (Kim et al. 2020; Mohler et al. 2019). Fabrizi and Trivisano (2016) tackle such an issue 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), Breunig and Hutchinson (2008), Schluter and van Garderen (2009) by adopting different methodological approaches of correction.

Note that income data are collected through household surveys with complex sampling designs that adopt stratification and/or 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 highly unrobust to them (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 the case of small samples.

We investigate the nature of the bias and propose a methodological framework for bias correction. Our proposal constitutes a generalization of the framework of Breunig and Hutchinson (2008), developed for *iid* observations, to the finite population and design-based setting. At the same time, we extend the proposal to a wider set of measures from the Gini index to two parametric families of measures: the Atkinson and the Generalized Entropy family, commonly used to measure inequality (Daly and Valletta 2006). We consider a wide variety of measures as the concurrent estimation of alternative indicators—as opposed to the more commonly used Gini Index—may bring to light a wider picture of the inequality phenomenon. This is motivated by their interesting properties such as the additive decomposability, for Generalized Entropy measures, and the explicit social welfare representation, for Atkinson measures. Moreover, all the measures considered pertain to the class of dispersion-based measures, sharing common features that enable the development of a general bias correction framework. To the best of our knowledge, this is the first proposal of bias correction for the Atkinson and Generalized Entropy indexes in the complex survey case, whereas it provides an extension for the Gini index case with respect to existing proposals as it is made clear in Section 4.

To our purpose, we take advantage of a methodology based on Taylor's expansions, even if the same analytical results can be obtained through other types of linearization, such as the one proposed by Graf (2011). Our extension for complex designs is based on the introduction in the estimation strategy of (i) sampling weights, as to consider the unequal probabilities of selection and (ii) relevant design information, such as strata and clusters, to control for possible correlation among units. Other limitations associated with household surveys are related to non-sampling issues, such as non-response and non-representativeness, which may significantly impact the accuracy of estimates. The incorporation of sampling weights, if properly treated for non-response and calibrated to known population totals, may also protect against such issues. This is the case of the Italian EU-SILC survey data we employ in this paper (ISTAT 2021).

By considering a combination of stratified and multistage cluster sampling, the incorporation of weights is made explicit by adopting Horvitz-Thompson type estimators and the ultimate clusters technique for design variances and covariances estimation. 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. Results under different simulation scenarios confirm that the presence of extreme values does not seem to compromise the bias correction process. Lastly, we provide a small area estimation exercise that shows the risk of ignoring ex-ante bias correction.

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 Statistics on Income and Living Condition (EU-SILC) income data 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, 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. Suppose we have a finite population U of $N < \infty$ elements labeled as $\{1, \dots, N\}$. Let $y \in \mathbb{R}^+$ be a continuous random variable denoting a characteristic of interest, in our case income, for all the units of the finite population and $F(y)$ its cumulative distribution function. The Gini index can be defined as

$$\theta_G = \frac{2}{\mu} \int_0^{+\infty} yF(y)dF(y) - 1,$$

with $\mu = \int_0^{+\infty} ydF(y)$ (Davidson 2009). However, the estimation of alternative measures, in addition to the Gini index, may enable a more meaningful assessment of different aspects of economic inequality. The Gini index is decomposable within and between groups only in very specific cases (Mookherjee and Shorrocks 1982). Moreover, it is positional (weakly) transfer sensitive, namely income transfers induce index variations depending on the ranks of the transfer donor and recipient.

An explicit incorporation of social welfare in inequality measurement is given by Atkinson indexes, which 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

$$\theta_A(\varepsilon) = \begin{cases} 1 - \frac{1}{\mu} \left(\int_0^{+\infty} y^{1-\varepsilon} dF(y) \right)^{1/(1-\varepsilon)} & \text{for } \varepsilon \neq 1 \\ 1 - \frac{1}{\mu} \exp \left(\int_0^{+\infty} \log(y) dF(y) \right) & \text{for } \varepsilon = 1. \end{cases}$$

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 of being 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, when applied to income distributions, has the meaning of deviation from perfect equality:

$$\theta_{GE}(\alpha) = \begin{cases} \frac{1}{\alpha(\alpha-1)} \left(\frac{1}{\mu^\alpha} \int_0^{+\infty} y^\alpha dF(y) - 1 \right), & \alpha \neq 0 | 1, \\ \frac{1}{\mu} \int_0^{+\infty} y \log(y) dF(y) - \log \mu, & \alpha \rightarrow 1, \\ \log \mu - \int_0^{+\infty} \log(y) dF(y), & \alpha \rightarrow 0. \end{cases}$$

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 the coefficient of variation (CV) as an inequality measure, being linked with a member of the GE family, namely $\theta_{GE}(2) = CV^2/2$. Its square has been used in some income distribution analyses, including OECD (2011), even though it seems to be very sensitive to top outliers (Cowell and Victoria-Feser 1996).

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 a structurally negative bias as shown further on in this section. Among the measures with non-predictable bias direction, Breunig (2001) shows that the bias of CV and GE (2) is negatively related to the skewness of income distribution. This aspect could be analyzed in-depth by imposing a distributional assumption on the income variable, but this is beyond the scope of this paper. 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:

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

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

Proof. In appendix.

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 $s \subset U$ out of the set of all possible samples Q thus $p(s) \geq 0$ and $\sum_{s \in Q} p(s) = 1$. The inclusion probability of unit k is denoted with π_k , being $\pi_k = \sum_{s \in Q_k} p(s)$ with 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 $\gamma = E[g(y)]$, with $g(\cdot)$ a generic monotone transformation of the income variable. The population value for the generic inequality measure is

$$\theta = f(\mu, \gamma),$$

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, that is,

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

where N is the population size and $\mathbf{w} = (w_1, \dots, w_n) = (1/\pi_1, \dots, 1/\pi_n)$ or a transformation of it after a non-response treatment and calibration. Note that $\hat{\mu}$ is unbiased and that the results of this section hold also for Hájek type estimators, that is, with denominator $\hat{N} = \sum_{i=1}^n w_i$, since it is approximately unbiased (Särndal et al. 2003). 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

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

In Table 1, we detail the survey estimators for each inequality measure and their bias formulation based on Equation (1) along with all relevant quantities. The complex survey estimators of Atkinson and Generalized Entropy measures come from Biewen and Jenkins (2006), while as for the Gini index, we employ the alternative formulation defined by Sen (1997) and the complex survey estimator proposed by Langel and Tillé

Table 1. Relevant Quantities for Each Measure Including the Approximate Bias.

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

(2013). Let denote with $\sqrt{n/(n-1)}$ the standard bias-correction adjustment for the weighted variance; lastly consider $\widehat{N}_i = \sum_{k \in s} w_k 1(n_k \leq n_i)$ with n_i denoting the rank of i -th unit. The notation $1(A)$ defines an indicator function, assuming value 1 if A is observed and 0 otherwise.

Let us denote the Gini index estimator with $\widehat{\theta}_G$, its approximate bias in small samples is

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

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

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

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 (2), 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 $n(n-1)^{-1}\widehat{\theta}_G$, our correction reconsiders the shape of the adjusted estimator with a further order of approximation as

$$\frac{n}{n-2} (\widehat{\theta}_G - a), \tag{4}$$

with a equals the sum of the second and third terms of (3).

Note that the bias formulas of Table 1 can also be reached differently, namely by applying the linearization proposed by Graf (2011) and extended by Vallée and Tillé (2019), as made explicit in the Appendix. Graf’s methodology requires a separate derivation for each measure. In contrast, Equation (1) defines a general bias formulation of the bias that applies to the entire set of considered measures, isolating its components and easing a general interpretation. This is one of the pros of our methodology, together with the fact that it is a distribution-free procedure, not requiring any parametric assumption on income distribution. Another pro that is worth mentioning is that Taylor’s expansion in Equation (1) relies on design variance and covariances. Since uncertainty estimation of complex design estimators is of great interest, such quantities can safely be estimated in a complex survey context as several variance estimation techniques have been proposed and tested in literature. On the contrary, other methodologies, such as small-sigma or Edgeworth expansions, may require parametric assumptions on income distribution and/or the estimation of moments up to higher orders, which may be unreliable in the case of complex design data.

As is clear from Table 1, the bias correction of GE(2) does not include the coefficient of skewness of the income distribution, as shown by Breunig (2001). A reliable estimation of that quantity, while being straightforward in the *iid* case, appears cumbersome in the case of weighted data being defined on a discrete grid of values. This adds up to another aspect: their estimators may be particularly unstable in small samples (Joanes and Gill 1998). This leads to the non-applicability of Breunig (2001) result 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 $V[\hat{\mu}]$, $V[\hat{\gamma}]$, and $Cov[\hat{\mu}, \hat{\gamma}]$. We consider a complex survey design involving stratification and multi-stage selection, with both Self-Representing (SR) strata, that is, included at the first sampling 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.

We define an unbiased estimator for the variance of Horvitz-Thompson estimators, such as $\hat{\mu}$, when $w_i = 1/\pi_i$ as

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

with π_{ik} , $\forall i, k \in U, i \neq k$ denoting the second-order inclusion probabilities that is, the probability that the sample includes both i -th and k -th units (Arnab 2017). However generally (a) $w_i \neq 1/\pi_i$ and (b) π_{ik} , $\forall i, k \in U, i \neq k$ are difficult to calculate under complex sampling designs.

Therefore, the variance estimator to be considered constitutes an approximation that relies 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. Without loss of generality, let us consider a two-stage scheme, where $\hat{\mu} = \sum_h \sum_d \sum_i w_{hdi} y_{hdi} / N$ is a linear estimator of μ , with h the stratum indicator, d the Primary Sampling Unit (PSU) indicator, and i the secondary sampling unit (household) indicator. Its variance estimator is as follows:

$$\begin{aligned} \widehat{V}[\hat{\mu}] &= \sum_{h=1}^{H_{SR}} V[\hat{\mu}_h] + \sum_{h=1}^{H_{NSR}} V[\hat{\mu}_h] \\ &= \sum_{h=1}^{H_{SR}} M_h \frac{M_h - m_h}{m_h(m_h - 1)} \sum_{i=1}^{m_h} (y_{hi} - \bar{y}_h)^2 + \sum_{h=1}^{H_{NSR}} \frac{n_h}{n_h - 1} \sum_{d=1}^{n_h} (\hat{\mu}_{hd} - \bar{\mu}_h)^2, \end{aligned} \quad (5)$$

with H_{SR} Self-Representing and H_{NSR} non Self-Representing 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 $\bar{y}_h = \sum_{i=1}^{m_h} y_{hi} / m_h$, $\hat{\mu}_{hd} = \sum_{i=1}^{m_d} w_{hdi} y_{hdi} / N$ 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 (5) 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 $V[\hat{\gamma}]$ can be obtained by adopting the same strategy used for $V[\hat{\mu}]$ in (5). Whereas, regarding the estimation of the design covariance, consider that

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

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

The Gini index estimator differs from the other indexes since $\hat{\gamma}$ is a non-linear statistic. Thus, a linearization of $\hat{\gamma}$ is needed to make it tractable and carry on variance estimation with the procedure described above. We consider again the linearization proposed by Graf (2011) with the practical adaptation of Graf and Tillé (2014) for inequality estimators. In such adaptation, the linearized variable is merely a function of the partial derivatives with respect to the weights, that in the case of $\hat{\gamma}$ defined for Gini index in Table 1 is

$$v_k = \frac{\partial \hat{\gamma}}{\partial w_k} = \frac{1}{N^2} \left[y_k (\hat{N}_k - w_k) + \sum_{i \in S_k} w_i y_i \right],$$

for a generic unit k where $S_k = \{i \in s, n_i > n_k\}$. In this way, the estimator can be re-expressed through a linear approximation, namely $\hat{\gamma} \approx \sum_{i \in s} w_i v_i$, and it becomes possible to perform variance estimation of linear statistics.

In this section, we have detailed the estimation of each quantity that contributes to the definition of the bias-corrected estimator of inequality measures. Note that the issues related to the sampling variance of bias-corrected estimators and its estimation are addressed later on in Section 6.

5. Design-Based Simulation

A design-based simulation study has been conducted 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 twenty-one 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 works with close-to-reality data, affected by peculiar problems, for example, extreme values and 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 concerning both upper and lower tails, to circumvent non-robustness problems. The issue of robust estimation of economic indicators through an extreme values treatment in the upper tail of income distribution is well-established in the literature. See Brzezinski (2016) for a review and Alfons et al. (2013) for a suitable specification for survey data. On the contrary, the issue of treatment of extreme values in the lower tail of income distribution appears less established (Hlasny et al. 2022; Masseran et al. 2019; Van Kerm 2007).

The treatment is done at a regional level to the original EU-SILC sample and the detection of outlier is carried out by using the Generalized Boxplot procedure for skewed or heavy-tailed distributions (Bruffaerts et al. 2014). Outliers are defined as the observed values that exceed certain bounds computed by directly taking into account the skewness and tail heaviness of the distribution. Such outliers, once identified, are randomly replaced by draws from Pareto or inverse Pareto tails. On the upper tail, we operate a semi-parametric Pareto-tail modeling 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 use an inverse Pareto modification of the PITSE estimator suggested by Masseran et al. (2019). The resulting dataset is specified as an alternative (hereafter, *treated*) pseudo-population. The number of treated observations, together with some summary statistics about survey data, can be found in the Appendix.

From both pseudo-populations, we repeatedly select 1,000 two-stage stratified samples, mimicking the sampling strategy adopted in the survey itself. In the EU-SILC survey, the first-stage is characterized by a stratified sampling of municipalities according to NUTS-2 region and population sizes. In the second-stage, households are selected within each PSU through systematic sampling. The simulation study mimics this design by approximating strata to NUTS-2 regions. We repeated the drawing for both scenarios involving different sampling rates, 1.5% and 3% respectively. Results before/after treatment are compared to isolate the effect of extreme values when evaluating bias-correction performances.

The Relative Bias (RB), Mean Square Error (MSE), its variance component percentage (%VAR) and the Root Mean Square Error (RMSE) are calculated for each region r using the one thousand iterations as:

$$\begin{aligned} \text{RB}_r &= \frac{1}{1,000} \sum_{p=1}^{1,000} \left(\frac{\hat{\theta}_{p,r}}{\theta_r} - 1 \right), \\ \text{MSE}_r &= \frac{1}{1,000} \sum_{p=1}^{1,000} (\hat{\theta}_{p,r} - \theta_r)^2, \\ \text{VAR}\%_r &= \frac{1}{1,000} \sum_{p=1}^{1,000} (\hat{\theta}_{p,r} - \bar{\theta}_r)^2 \times \frac{1}{\text{MSE}_r}, \\ \text{RMSE}_r &= \sqrt{\text{MSE}_r}, \end{aligned}$$

where θ_r is the population value for region r , $\hat{\theta}_{p,r}$ its estimate at iteration p and $\bar{\hat{\theta}}_r$ is its mean value over all iterations. In our simulation setting, the regional sample sizes range from six to ninety-six individuals (from six to thirty-two households) on average over the simulated samples for the 1.5% sampling rate, and from 11 to 196 individuals (ten to seventy-four 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) and 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 survey estimator $\hat{\theta}$ and the bias-corrected estimator $\hat{\theta}^{corr}$. This confirms the nature of the bias as a small sample bias and shows the effectiveness of the correction, even if based on a large- n approximation as the Taylor's expansion. 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 that are not particularly sensitive to extreme observations such as the Gini index, GE (0), Atk (0.5) and Atk (1). In the 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. A general recommendation, therefore, is to avoid the use of such measures when facing strong income skewness.

Bias and error averaged across all areas for each scenario, sampling rate and estimator are shown in Table 2. By still focusing 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 ($n \geq 20$), the bias-corrected estimators seem to be approximately unbiased. Furthermore, it is important to note that the bias correction induces a slight but negligible error (RMSE) increase for every measure, except for the Gini index which presents a relevant increase. This exception may be explained by the shape of the unbiased estimators, as described by (4), where a sum of estimators is multiplied by a factor $n / (n - 2)$, which inherently inflates the variance by its square.

Note that the variance component percentage of MSE (%VAR) for the corrected estimators is always greater than for the non-corrected counterparts, and it reaches the 100% of MSE in some cases. This means that the error is largely due to estimators variance, while the bias has a minimal component. The price to pay for a bias correction procedure is an increase in variance; this bias-variance trade-off pushes us to a reflection. Since we are in a small sample context, both corrected and uncorrected estimators are strongly unreliable, requiring a variance reduction step. To undertake this step, the corrected estimators are preferable as their error is largely due to estimator variability ($> 95\%$). Instead, variance reduction techniques could not be applied to uncorrected estimates as a great source of error is their bias. In that case, such techniques may induce further bias that, depending on the sign, can lead to deteriorated estimates.

Let us focus on comparing the treated population scenario with the non-treated one. In the latter case, bias and error increase dramatically both for $\hat{\theta}$ and for $\hat{\theta}^{corr}$. In particular, the bias is great for some measures estimated on the non-treated scenario due to their non-robustness properties to extreme values. It is the case of Atk ($\varepsilon = 2$), extremely

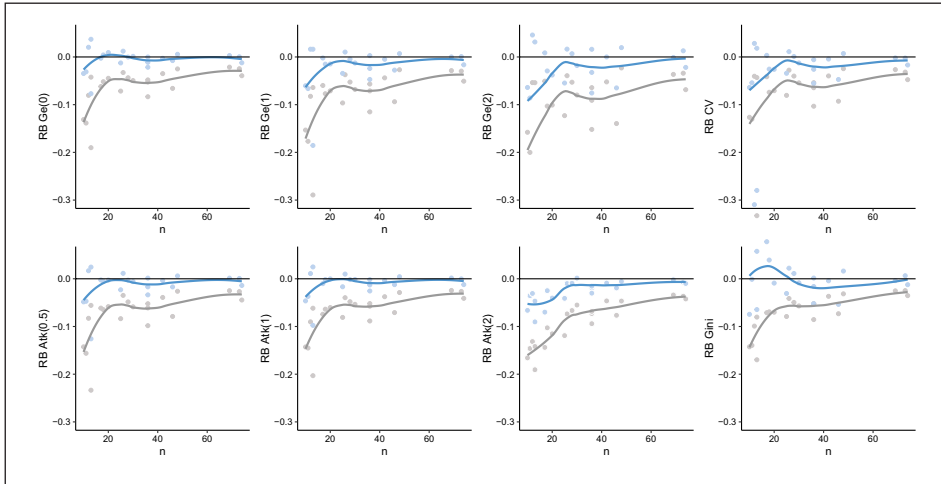


Figure 1. Relative Bias of non-corrected measures (gray line), and corrected measures (blue line) in 3% samples after extreme value treatment versus the (average) sample size.

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. However, the correction leads to a bias reduction comparable to the one discussed for the treated pseudo-population: it seems not to change in magnitude with respect to sample size and extreme values presence. Therefore, the presence of extreme values does not seem to compromise the bias correction process, implying only a slight error increase in resulting estimators.

To summarize, our results highlight that in the case of populations that are not affected by income extreme values, the bias correction may provide approximately unbiased estimates for a large class of measures at the expense of, in most cases, only a slight error increase. Vice versa, it might be necessary to restrict the attention to the most robust measures such as GE with $\alpha = 0$, Atkinson index with $\varepsilon = 1$, and Gini Index to obtain estimates affected by a negligible bias.

6. A Small Area Estimation Exercise

In the previous sections, we propose a method to correct the 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 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 take advantage of available auxiliary information to produce estimates with acceptable uncertainty. More specifically, the model-based SAE 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.

Table 2. Percentage RB and RMSE Averaged on the Twenty-One Regions for Each Inequality Estimator and Scenario.

	CV	GE(0)	GE(1)	GE(2)	A(0.5)	A(1)	A(2)	Gini
<i>With extreme values treatment</i>								
1.5%								
$\hat{\theta}$								
\overline{RB}	-11.9	-13.9	-16.0	-17.6	-14.9	-15.0	-19.0	-14.6
\overline{RMSE}	0.189	0.086	0.077	0.106	0.037	0.071	0.136	0.085
$\overline{VAR \%}$	78.3	93.6	90.7	90.7	91.6	90.9	82.7	75.8
$\hat{\theta}^{corr}$								
\overline{RB}	-5.6	-4.1	-6.8	-9.0	-5.4	-5.2	-9.4	0.5
\overline{RMSE}	0.192	0.095	0.086	0.117	0.041	0.078	0.142	0.131
$\overline{VAR \%}$	91.8	99.5	98.4	96.9	99.0	99.0	95.7	98.1
3.0%								
$\hat{\theta}$								
\overline{RB}	-7.4	-6.6	-8.6	-10.6	-7.6	-7.3	-9.7	-7.2
\overline{RMSE}	0.186	0.062	0.059	0.086	0.027	0.052	0.101	0.059
$\overline{VAR \%}$	91.7	97.4	95.5	94.9	96.1	96.1	92.3	88.2
$\hat{\theta}^{corr}$								
\overline{RB}	-2.8	-0.6	-2.3	-3.6	-1.5	-1.2	-2.9	0.3
$\overline{RB} (n \geq 20)$	1.4	0.4	1.2	1.6	0.8	0.6	1.4	0.9
\overline{RMSE}	0.190	0.068	0.065	0.097	0.030	0.056	0.108	0.073
$\overline{VAR \%}$	95.5	100.0	99.5	98.5	99.8	99.8	99.4	97.9

(Continued)

Table 2. (Continued)

	CV	GE(0)	GE(1)	GE(2)	A(0.5)	A(1)	A(2)	Gini
<i>Without extreme values treatment</i>								
1.5%								
$\hat{\theta}$	-18.2	-12.7	-17.5	-23.3	-15.3	-15.6	-48.0	-14.9
$\overline{\text{RMSE}}$	0.233	0.129	0.096	0.155	0.046	0.097	0.380	0.092
$\overline{\text{VAR}}\%$	74.9	96.3	91.0	88.3	92.8	92.6	37.8	76.7
$\overline{\text{RB}}$	-12.1	-3.9	-8.7	-15.0	-6.3	-5.9	-41.6	0.04
$\overline{\text{RMSE}}$	0.238	0.140	0.107	0.176	0.050	0.107	0.369	0.140
$\overline{\text{VAR}}\%$	87.7	99.7	97.9	94.1	98.9	99.0	47.8	98.1
3.0%								
$\hat{\theta}$	-12.7	-6.8	-10.5	-15.8	-8.7	-8.4	-38.1	-7.9
$\overline{\text{RMSE}}$	0.195	0.093	0.076	0.144	0.035	0.073	0.332	0.065
$\overline{\text{VAR}}\%$	84.2	98.2	95.0	93.3	96.1	96.5	48.7	87.5
$\overline{\text{RB}}$	-7.8	-1.2	-3.9	-8.0	-2.5	-2.0	-32.4	0.06
$\overline{\text{RB}} (n \geq 20)$	-8.3	-1.2	-3.6	-8.7	-2.3	-1.7	-30.0	-1.2
$\overline{\text{RMSE}}$	0.210	0.099	0.087	0.172	0.038	0.080	0.328	0.081
$\overline{\text{VAR}}\%$	93.9	100.0	99.4	98.2	99.7	99.9	58.5	98.5

In this context, area-level models appear to be less demanding in terms of data requirements and enable the incorporation of design-based properties. Such models constitute a typical framework for the application of our bias-correction proposal, as they assume the unbiasedness of 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 involve survey estimators.

In this section, we perform an SAE exercise by measuring inequality in specific small domains through the 2017 Italian EU-SILC data, already employed in Section 5. The domains considered are the interaction between five NUTS-1 regions (North-East, North-West, Center, South, Insular), three DEGURBA classification types (Urban, Peri-urban, and Rural), and six household types (one-member households, lone parents with one or two dependent children, lone parents with three or more dependent children, couples with one or two dependent children, couples with three or more dependent children, households without dependent children). As dependent children, we mean sons/daughters aged less than twenty-five. This allows the estimation of inequality for geographics domains and specific sub-population of interest such as household types.

The purpose is not to propose a small area estimation strategy 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 Fay-Herriot model (Fay and Herriot 1979), a landmark model in the small area literature, implemented through the package *sae* (Molina and Marhuenda 2015) to both uncorrected and corrected survey estimators. The objective is to check whether the inclusion of biased or bias-corrected survey estimates in the model may lead to different results. We perform the exercise on the most popular estimators among the ones previously considered: 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 \mathbf{x}_m the set of p areal covariates for area m , $m = 1, \dots, M$. The classical area-level model is the Fay-Herriot one, specified as follows:

$$\hat{\theta}_m | \theta_m \sim N(\theta_m, D_m), \quad (6)$$

$$\theta_m \sim N(\mathbf{x}_m^T \boldsymbol{\beta}, \sigma^2), \quad m = 1, \dots, M, \quad (7)$$

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

As mentioned above, the design variance is separately estimated from the data and given as input to the small area model: its estimation in real application is the crux of an SAE procedure. In the case of uncorrected 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 the Gini index and Biewen and Jenkins (2006) for the Generalized Entropy and the Atkinson indexes. On the other hand, the variance of

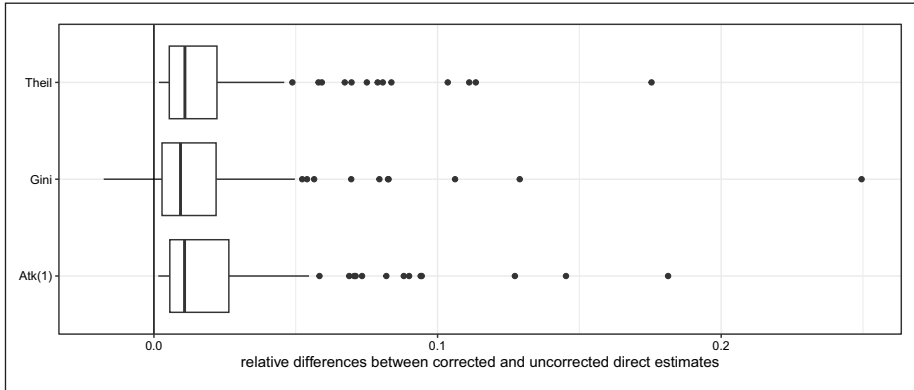


Figure 2. Relative differences between corrected and uncorrected direct estimates for each measure.

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, we recommend relying on resampling methods; a comprehensive review of the use of bootstrap methods for survey data can be found in Lahiri (2003). In this case, we employed the design-aware bootstrap procedure as presented by Fabrizi et al. (2011, 2020). The algorithm involves both a drawing procedure that considers the multi-stage selection process and a calibration procedure, applied to each bootstrap sample, that adjusts weights with respect to known totals. A similar process is performed to the original EU-SILC sample by national statistical offices.

The comparison between uncorrected and corrected survey estimates for all three measures is displayed in relative differences, that is, $(\hat{\theta}_m^{corr} - \hat{\theta}_m) / \hat{\theta}_m, \forall m$, in Figure 2. Corrected estimates show, in most cases, higher values of inequality in comparison with the uncorrected ones; the highest increase reaches almost 25%. This is in accordance with the underestimation highlighted by simulation results of Section 5. The sampling coefficients of variation of both estimators range from 0.06 to 0.65 for the Theil index, from 0.05 to 0.50 for the Atkinson index and, lastly from 0.03 to 1.35 for the Gini index depending on the domain. Such values point out the need for SAE techniques.

We separately fit the Fay-Herriot model for both corrected and uncorrected estimators by using the same set of covariates. We consider only covariates defined for the geographical area of interest of the domains: the aged dependency ratios from census data and the average values and incidence by income source from tax forms data. Indeed, due to data disclosure issues, it is not possible to retrieve the disaggregated information by household type. Such covariates are subjected to variable selection to avoid multicollinearity and to neglect irrelevant regressors. The final set includes (i) the age dependency ratio, measuring the population aged between zero and fourteen over the total population, being positively related to inequality, (ii) the average income declared by entrepreneurs, and (iii) the employee income incidence, both negatively related to inequality. The model-based (or EBLUP, Empirical Best Linear Unbiased Predictor)

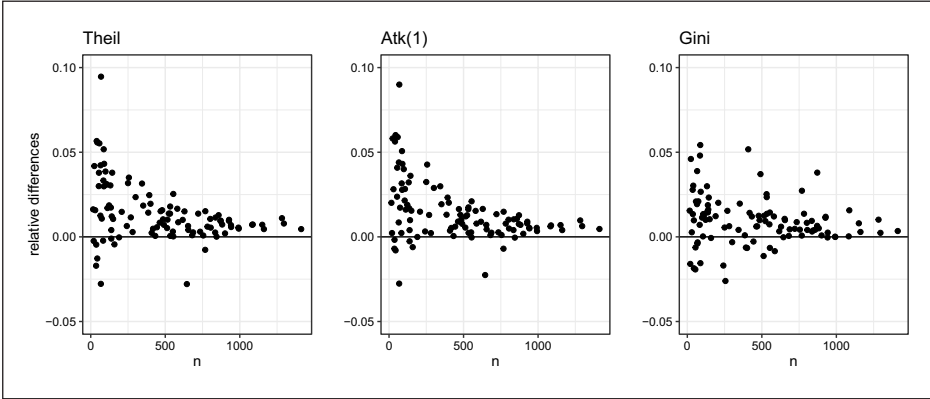


Figure 3. Sample sizes versus relative differences between model-based estimates based on corrected survey estimates and model-based estimates based on uncorrected ones.

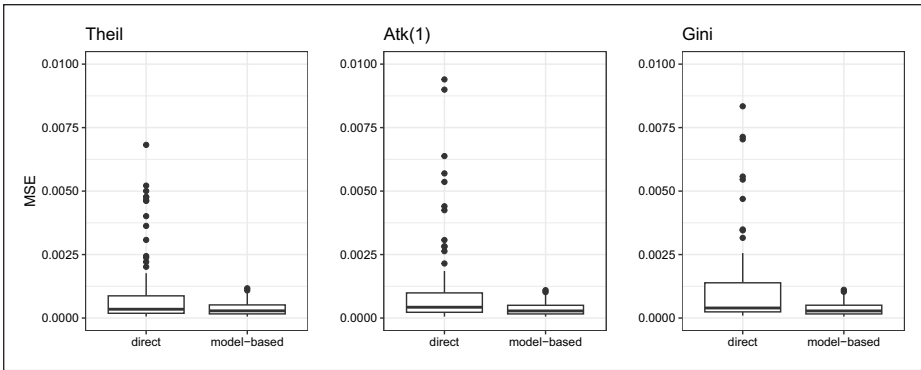


Figure 4. MSE of inequality estimators. Bias-corrected direct estimators versus model-based estimators.

estimates in both cases are compared in terms of relative differences in Figure 3. The inequality levels estimated by the misspecified model are lower in most of the cases, resulting in a misleading inference. The greatest divergences show that the model-based estimate resulting from an ex-ante bias correction is 11.4% higher than the one without correction. This confirms the risk of underestimation of inequality when neglecting such an issue. As expected, the divergences between model-based estimates in both cases decrease to zero at increasing sample sizes.

By focusing only on EBLUP results based on corrected estimates, the decrease in terms of error induced by the model is depicted in Figure 4. The reduction is relevant and testifies that the variance reduction procedure, put in place by the SAE model, is effective. As a consequence, such model-based estimates result to be reliable and ready to be used for further analysis or mapping.

7. Conclusions

A strategy based on Taylor's expansion has been proposed to correct the small sample bias of inequality estimators. The inequality measures considered are several, as the comparison of diverse measures may enable us to enlighten the specific point of view that each measure provides, like 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 with respect to outliers and a simulation study have been conducted to study the estimator behavior to extreme values and the performance of the proposed correction. Results show that survey-based estimators may be biased in small samples, inducing an underestimation that is even greater in the 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 sensitive to extreme values appear to be more biased, in particular, GE with $\alpha = 2$ and Atkinson with $\varepsilon = 2$. Therefore, in the 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.

An illustrative small-area application has been carried out to evaluate the effect of disregarding bias in a typical small-sized sample context. The results obtained show that neglecting it translates into a misleading inference and inequality underestimation. In such an application, we use a basic area-level model, the Gaussian one. Indeed, the possibly not-Gaussian sampling distributions 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 suggests an interesting direction for future research. Further directions also include the extension of this framework to other widely used inequality measures, such as those based on quintiles and the development of a multivariate SAE framework.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work of Silvia De Nicolò was partially funded by the ALMA IDEA 2022 grant (title: "Social exclusion and territorial disparities: poverty and inequality mapping through advanced methods of small area estimation," project J45F21002000001) and by the PNRR PE10 ONFOOD project (title: "Research and innovation network on food and nutrition Sustainability, Safety and Security – Working ON Foods," project J33C22002860001). Part of the European Union—NextGenerationEU funding. The work of Maria Rosaria Ferrante was funded by the European Union - NextGenerationEU, in the framework of the "GRINS -Growing Resilient, INclusive and Sustainable project" (PNRR - M4C2 - I1.3 - PE00000018 – CUP J33C22002910001). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

References

- Alfons, A., M. Templ, and P. Filzmoser. 2013. "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): 271–86.
- Arnab, R. 2017. *Survey Sampling Theory and Applications*. Cambridge, MA: Academic Press. DOI: <http://doi.org/10.1016/B978-0-12-811848-1.00001-7>.
- Bellu, L. G., and P. Liberati. 2006. "Social Welfare, Social Welfare Functions and Inequality Aversion." *Food and Agriculture Organization of the United Nations*.
- Benedetti, S., and F. Crescenzi. 2023. "The Role of Income Poverty and Inequality Indicators at Regional Level: An Evaluation for Italy and Germany." *Socio-Economic Planning Sciences* 87: 101540.
- Biewen, M., and S. P. Jenkins. 2006. "Variance Estimation for Generalized Entropy and Atkinson Inequality Indices: The Complex Survey Data Case." *Oxford Bulletin of Economics and Statistics* 68 (3): 371–83.
- Breunig, R. 2001. "An Almost Unbiased Estimator of the Coefficient of Variation." *Economics Letters* 70 (1): 15–9.
- Breunig, R., and Hutchinson, D. L.-A. 2008. "Small Sample Bias Corrections for Inequality Indices." In *New Econometric Modelling Research*, edited by W. N. Toggins, 61–83. New York, NY: Nova Science Publishers.
- Bruffaerts, C., V. Verardi, and C. Vermandele. 2014. "A Generalized Boxplot for Skewed and Heavy-Tailed Distributions." *Statistics & Probability Letters* 95: 110–7.
- Brzezinski, M. 2016. "Robust Estimation of the Pareto Tail Index: A Monte Carlo Analysis." *Empirical Economics* 51 (1): 1–30.
- Cavanaugh, A., and S. Breau. 2018. "Locating Geographies of Inequality: Publication Trends Across OECD Countries." *Regional Studies* 52 (9): 1225–36.
- Cowell, F. A., and M.-P. Victoria-Feser. 1996. "Robustness Properties of Inequality Measures." *Econometrica: Journal of the Econometric Society* 64 (1): 77–101.
- Daly, M. C., and R. G. Valletta. 2006. "Inequality and Poverty in United States: The Effects of Rising Dispersion of Men's Earnings and Changing Family Behaviour." *Economica* 73 (289): 75–98.
- 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): 226–34. DOI: <https://doi.org/10.1162/rest.2003.85.1.226>.
- Fabrizi, E., M. R. Ferrante, S. Pacei, and C. Trivisano. 2011. "Hierarchical Bayes Multivariate Estimation of Poverty Rates Based on Increasing Thresholds for Small Domains." *Computational Statistics & Data Analysis* 55 (4): 1736–47.
- Fabrizi, E., M. R. Ferrante, and C. Trivisano. 2020. "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): 1273–91.
- Fabrizi, E., and C. Trivisano. 2016. "Small Area Estimation of the Gini Concentration Coefficient." *Computational Statistics & Data Analysis* 99: 223–34.
- Fay, R. E., and R. A. Herriot. 1979. "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data." *Journal of the American Statistical Association* 74 (366a): 269–77.
- Finkelstein, M., H. G. Tucker, and J. Alan Veeh. 2006. "Pareto Tail Index Estimation Revisited." *North American Actuarial Journal* 10 (1): 1–10.

- Giles, D. E. 2005. "The Bias of Inequality Measures in Very Small Samples: Some Analytic Results." Technical Report, Department of Economics, University of Victoria.
- Graf, E., and Y. Tillé. 2014. "Variance Estimation Using Linearization for Poverty and Social Exclusion Indicators." *Survey Methodology* 40 (1): 61–79.
- Graf, M. 2011. "Use of Survey Weights for the Analysis of Compositional Data." In *Compositional Data Analysis: Theory and Applications*, chapter 9, 114–127. Hoboken, NJ: John Wiley and sons. DOI: <https://doi.org/10.1002/9781119976462>.
- Harmening, S., A.-K. Kreutzmann, S. Schmidt, N. Salvati, and T. Schmid. 2023. "A Framework for Producing Small Area Estimates Based on Area-Level Models in R." *R Journal* 15 (1): 1–26.
- Hlasny, V., L. Ceriani, and P. Verme. 2022. "Bottom Incomes and the Measurement of Poverty and Inequality." *Review of Income and Wealth* 68 (4): 970–1006.
- ISTAT. 2021. L'indagine Eu-Silc: Innovazione Nella Metodologia Di Rilevazione E Di Stima. Istituto Nazionale di Statistica.
- Jasso, G. 1979. "On Gini's Mean Difference and Gini's Index of Concentration." *American Sociological Review* 44 (5): 867–70.
- Joanes, D. N., and C. A. Gill. 1998. "Comparing Measures of Sample Skewness and Kurtosis." *Journal of the Royal Statistical Society: Series D (The Statistician)* 47 (1): 183–9.
- Kakwani, N. 1990. "Large Sample Distribution of Several Inequality Measures: With Application to Cote d'Ivoire." In *Contributions to Econometric Theory and Application*, edited by R. A. L. Carter, J. Dutta, and A. Ullah, 50–81. New York, NY: Springer.
- Kalton, G. (1979). "Ultimate Cluster Sampling." *Journal of the Royal Statistical Society: Series A (General)* 142 (2): 210–22.
- Kim, L., C. Adolph, J. D. West, and K. Stovel. 2020. "The Influence of Changing Marginals on Measures of Inequality in Scholarly Citations: Evidence of Bias and a Resampling Correction." *Sociological Science* 7: 314–41.
- Lahiri, P. 2003. "On the Impact of Bootstrap in Survey Sampling and Small-Area Estimation." *Statistical Science* 18 (2): 199–210.
- Langel, M., and Y. Tillé. 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): 521–40.
- Lerman, R. I., and S. Yitzhaki. 1989. "Improving the Accuracy of Estimates of Gini Coefficients." *Journal of Econometrics* 42 (1): 43–7.
- Marchetti, S., and N. Tzavidis. 2021. "Robust Estimation of the Theil Index and the Gini Coefficient for Small Areas." *Journal of Official Statistics* 37 (4): 955–79.
- Márquez, M. A., E. Lasarte, and M. Lufin. 2019. "The Role of Neighborhood in the Analysis of Spatial Economic Inequality." *Social Indicators Research* 141 (1): 245–73.
- Masseran, N., L. H. Yee, M. A. M. Safari, and K. Ibrahim. 2019. "Power Law Behavior and Tail Modeling on Low Income Distribution." *Mathematics and Statistics* 7 (3): 70–7.
- Mohler, G., P. J. Brantingham, J. Carter, and M. B. Short. 2019. "Reducing Bias in Estimates for the Law of Crime Concentration." *Journal of Quantitative Criminology* 35 (4): 747–65.
- Molina, I., and Y. Marhuenda. 2015. "sac: An R Package for Small Area Estimation." *The R Journal* 7 (1): 81.
- Mookherjee, D., and A. Shorrocks. 1982. "A Decomposition Analysis of the Trend in UK Income Inequality." *The Economic Journal* 92 (368): 886–902.
- OECD. 2011. *Divided We Stand: Why Inequality Keeps Rising*. Paris: OECD Publishing.
- Osier, G., Y. G. Berger, and T. Goedeme. 2013. "Standard Error Estimation for the EU-SILC Indicators of Poverty and Social Exclusion." Eurostat Methodologies and Working Papers Series.

- Rao, J. N., and I. Molina. 2015. *Small Area Estimation*. Hoboken, NJ: John Wiley & Sons. DOI: <https://doi.org/0.1002/9781118735855>.
- Rust, K., and G. Kalton. 1987. "Strategies for Collapsing Strata for Variance Estimation." *Journal of Official Statistics* 3 (1): 69–81.
- Särndal, C.-E., B. Swensson, and J. Wretman. 2003. *Model Assisted Survey Sampling*. New York: Springer Science & Business Media. DOI: <https://doi.org/10.1007/978-1-4612-4378-6>.
- Schluter, C., and K. J. van Garderen. 2009. "Edgeworth Expansions and Normalizing Transforms for Inequality Measures." *Journal of Econometrics* 150 (1): 16–29. DOI: <https://doi.org/10.1016/j.jeconom.2008.12.022>.
- Sen, A. K. 1997. *On Economic Inequality*. Oxford: Clarendon Press.
- Tzavidis, N., and S. Marchetti. 2016. "Robust Domain Estimation of Income-Based Inequality Indicators." In *Analysis of Poverty Data by Small Area Estimation*, edited by M. Pratesi, 171–86. Wiley Online Library.
- Tzavidis, N., L.-C. Zhang, A. Luna, T. Schmid, and N. Rojas-Perilla. 2018. "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): 927–79.
- Vallée, A.-A., and Y. Tillé. 2019. "Linearisation for Variance Estimation by Means of Sampling Indicators: Application to Non-Response." *International Statistical Review* 87 (2): 347–67.
- Van Kerm, P. 2007. "Extreme Incomes and the Estimation of Poverty and Inequality Indicators from EU-SILC." Technical Report, IRISS Working Paper Series 2007-01, CEPS/INSTEAD.
- Van Ourti, T., and P. Clarke. 2011. "A Simple Correction to Remove the Bias of the Gini Coefficient Due to Grouping." *Review of Economics and Statistics* 93 (3): 982–94.

Received: March 14, 2023

Accepted: December 28, 2023

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 $g(\cdot) : \mathbb{R}^+ \rightarrow \mathbb{R}$, 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

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

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:

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

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 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 $\mathbf{y} = (y_1, \dots, y_N)$ 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 $\sum_{s \subset U} p(s) = 1$. Let us define $\mathbf{1} = (1_1, \dots, 1_N)$ 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 $\pi_i = E(1_i)$, that is, 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 $\hat{\theta} = \hat{\theta}(\mathbf{1}, \mathbf{y})$ an estimator of $\theta = \theta(\mathbf{y})$ with $\hat{\theta}(\mathbf{1}, \mathbf{y})$ twice differentiable with respect to $\mathbf{1}$. Graf (2011) shows that an approximation for $\hat{\theta}$ is

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

We can derive the approximation of the bias as

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

In the following, we derive the bias for $GE(\alpha)$ with $\alpha \neq 0, 1$, however such a result can be extended to all the other measures considered.

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

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

with $GE(\alpha)$ estimator defined as

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

By applying (A.1), its bias may be expressed with an approximate result as

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

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

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

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

as stated by Arnab (2017), (A.2) can be easily rewritten as

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

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

Additional Information on the Simulation Study

Table A1. Summary Statistics on 2017 Italian EU-SILC Data Related to the Equivalized Disposable Income (Min, Max, Weighted Median, and Mean), Sample Sizes, and Nr. of Treated Observations After Extreme Values Treatment by Region.

NUTS-I region	Equivalized disposable income				Sample size	Nr. treated obs.
	Min	W. Median	W. Mean	Max		
North-west	-6,383	18,857	21,293	848,644	11,898	318
North-east	-4,567	19,845	21,679	547,646	11,751	277
Central	-12,809	17,519	20,017	439,366	11,713	344
South	-2,138	12,829	14,598	178,318	9,570	321
Insular	-5,397	12,030	13,678	107,178	3,818	155