Biostatistics and Biometrics
Open Access Journal
ISSN: 2573-2633

Mini Review
Volume 9 Issue 3 - April 2019
DOI: 10.19080/BBOAJ.2019.09.555765

Biostat Biometrics Open Acc J

Small Sample Bias Corrections for Entropy Inequality Measures

Maria Rosaria Ferrante[1] and Silvia Pacei[2]*

[1]*Department of Economics, University of Bologna, Italy*

[2]*Department of Statistical Sciences, University of Bologna, Italy*

Submission: February 13, 2019; **Published:** April 17, 2019

***Corresponding author:** Silvia Pacei, Department of Statistical Sciences, University of Bologna, Italy

Abstract

In this mini review, we discuss the main results obtained so far by us and other authors on the matter of bias correction for entropy inequality measures in small samples.

Keywords: Small sample inference; Coefficient of variation squared; Mean log deviation; Complex surveys

Abbreviations: GE: Generalized Entropy Measures; AARE: Average Absolute Relative Error; AARB: Average Absolute Relative Bias

Introduction

Design based estimators of many of the commonly used inequality measures are known to be biased for small sample sizes. The reason is that those inequality measures can be represented as ratios of random variables, both of which are estimated from the sample data. The expected value of a ratio of random variables is not generally equal to the ratio of the expected values. Therefore, estimates of such inequality measures are biased in small samples. The bias of the sample measure is generally $O\left(\frac{1}{n}\right)$, where n is the sample size [1].

Despite inequality measures are frequently used to compare regions or subpopulations, fairly little attention has been paid in the literature to the mentioned problem. In this paper we focus on the class of generalized entropy measures (GE), because they satisfy some important axioms that make them particularly useful in practice [2]. For example, these measures satisfy the decomposability axiom, particularly useful for territorial comparisons, as it allows to decompose the total inequality into the part due to inequality within areas and the part due to differences between areas. A few papers propose a small sample correction for those measures (see [1, 3-5]), and only in the simple random sample context. Regarding the small sample bias correction for simple random sample, Breunig [3] focuses on one entropy measure, the coefficient of variation squared $\left(CV^2\right)$ and, without imposing any restrictions on the form of the distribution of the economic variable considered, he provides a neat analytical condition under which the bias will be positive or negative.

In particular he finds that, when the variable has a positive skewed distribution (as it is often the case for income distributions and in general for size distributions [6]) small sample $\widehat{CV}^2$ tends to underestimate the true CV^2, with possible consequences on the interpretation of the results. He provides a large-n approximation for the bias of the estimator and obtain an almost unbiased estimator of the coefficient of variation squared by subtracting the estimate of the bias to the biased estimator

$$\left(\widehat{CV}^2_{corr} = \widehat{CV}^2 - \widehat{Bias}\left(\widehat{CV}^2\right)\right).$$

Breunig [4] applies that corrected estimator for the coefficient of variation squared to income data for Cina and Kenya, and compare that estimator to an alternative bias-corrected estimator, $\widehat{CV}^2_{jkn}$, obtained using the leave-one-out jackknife estimator. Through a simulation study he shows that, when the sample is small or moderately large and the skewness coefficient is greater than $\frac{3}{2}CV$, the use of one of the two corrected estimators for CV^2 is advisable. Moreover, he finds that among the two corrected estimators he considers, $\widehat{CV}^2_{corr}$ works much better in terms of mean squared error.

Giles [5] and Breunig & Hutchinson [1] consider different approaches to derive corrections for any member of the generalized entropy measure (GE) class. The GE class of measures can be expressed as

$$GE(\alpha) = \frac{1}{\alpha(\alpha-1)}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_j}{\bar{y}}\right)^{\alpha} - 1\right]; \quad j = 1,\ldots,n; \quad \alpha \in [0,\infty) \quad (1)$$

where $\bar{y}$ denotes the sample mean. Specific special members of this family include Theil's mean log deviation $(\alpha = 0)$, Theil's Index $(\alpha = 1)$ and half the squared coefficient of variation $(\alpha = 2)$.

Giles [5] derives an approximation for the bias that is increasingly accurate as the sampling error becomes smaller, regardless of the sample size. Breunig & Hutchinson [1] write the GE measures as functions of the population mean, μ, and some other population functions, and then derive corrections for the GE measures, based on a second-order Taylor's series expansion of the sample estimates around the population values. They carry out a simulation study to compare this method with the estimation of the approximate bias obtained considering two re-sampling methods: Jackknife and Bootstrap. They find that the Jackknife produces the lowest average bias for all the GE measures considered, but also the highest average mean squared error. Therefore, they suggest to consider the correction based on the Taylor's series expansion, that provides a significant reduction of the bias and the most reliable estimates amongst the three bias correction methods.

Small sample bias correction for complex surveys

Ferrante & Pacei [7] face the problem of small sample bias in a complex survey context. They aim to obtain a small sample bias correction for the Theil's mean log deviation of the individual equivalized income, calculated using data taken from the EU-SILC sample survey for Italy in 2013

$$ge(0) = \frac{1}{\hat{N}}\left[\sum_{j=1}^{n} W_j \log\left(\frac{\bar{y}}{y_j}\right)\right] \quad (2)$$

where $\bar{y}$ denotes the small sample mean calculated using sample weight, and $\hat{N} = \sum_{j=1}^{n} W_j$. The estimator corrected for the bias could then be used to obtain "small area estimates" through appropriate models. They consider the approximate bias obtained by Breunig & Hutchinson [1] for $ge(0)$ in the simple random sample case, $ABias\left(ge(0)\right) = \frac{1}{2}\mu^{-2}Var(\hat{\mu})$, and propose to extend this result to a complex sample by considering an heuristic solution, that is by substituting μ with the weighted sample mean and $Var(\hat{\mu})$ with the correspondent estimate obtained using the standard procedure used by Eurostat for a two-stage stratified sample [8].

They carry out a simulation study using the EU-SILC sample as target population and then repeatedly select 1,000 random samples from the 21 administrative regions, mimicking the sample strategy adopted in the EU-SILC itself. They consider two overall sampling rates, 1.5 and 3%, leading to a regional sample size that ranges from a minimum of 6 to a maximum of 28 for the 1.5% sample, and almost twice for the 3% sample. The comparison between $ge(0)$ and its bias corrected version, $geCorr(0)$, is carried out in terms of bias and accuracy using, as usual, the average absolute relative bias (AARB) and the average absolute relative error (AARE) indices, where the average in that case is taken over the Regions. Results show that the correction considered greatly reduces the bias of the non-corrected estimator on average (AARB% is equal to 15.9 and 4.0% respectively for $ge(0)$ and $geCorr(0)$, in 1.5% sample, while it is equal to 7.9 and 2.6% respectively for $ge(0)$ and $geCorr(0)$, in 3% sample). Moreover, the reduction of the overall reliability of the estimates is negligible on average (AARE% reaches 51.8 and 52.3% respectively for $ge(0)$ and $geCorr(0)$, in 1.5% sample, while it is equal to 37.8 and 38.2% respectively for $ge(0)$ and $geCorr(0)$, in 3% sample).

Conclusion

The overall results of the simulation studies mentioned highlight the importance of correcting the estimates of inequality measure in small samples. The use of the corrections proposed is particularly advisable when the goal is to estimate regional inequality measures, which can help to plan policies to reduce such inequality.

References

1. Breunig R, Hutchinson DLA (2008) Small sample bias corrections for inequality indices. In: New Econometric Modeling Research, William N. Toggins (eds), Nova Science Publishers: New York, USA.
2. Cowell FA (1985) Measures of Distributional Change: An Axiomatic Approach. Review of Economic Studies 52: 135-151.
3. Breunig R (2001) An Almost Unbiased Estimator of the Coefficient of Variation. Economic Letters 70(1): 15-19.
4. Breunig R (2002) Bias Correction for Inequality Measures: An Application to Cina and Kenia. Applied Economics Letters 9: 783-786.
5. Giles DE (2005) The bias of inequality measures in very small samples: some analytic results, Econometric Working Paper EWP0514, ISSN 1485-6441, University of Victoria, Department of Economics, Canada.
6. Kleiber C, Kotz S (2003) Statistical Size Distributions in Economics and Actuarial Sciences, New Jersey: John Wiley and Sons, USA.
7. Ferrante MR, Pacei S (2018) Small Area Estimation of Inequality Measures. In: Book of Short Papers SIS 2018, Pearson pp. 1276-1280.
8. Eurostat (2013), Standard error estimation for the EU-SILC indicators of poverty and social exclusion, EUROSTAT methodologies and working papers.

How to cite this article: Maria R F, Silvia P. Small Sample Bias Corrections for Entropy Inequality Measures. Biostat Biometrics Open Acc J. 2019; 9(3): 555765. DOI: 10.19080/BBOAJ.2019.09.555765

DOI: 10.19080/BBOAJ.2019.09.555765

How to cite this article: Maria R F, Silvia P. Small Sample Bias Corrections for Entropy Inequality Measures. Biostat Biometrics Open Acc J. 2019; 9(3): 555765. DOI: 10.19080/BBOAJ.2019.09.555765