Small Sample Bias Corrections for Entropy Inequality Measures

Maria Rosaria Ferrante¹ and Silvia Pacei²*

¹Department of Economics, University of Bologna, Italy
²Department 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

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

\[ \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{\bar{y}} \]

where \( \bar{y} \) is the sample mean. Specific special members denote the sample mean. Specific special members = −1 approximation for the bias of the estimator and obtain an half the squared coefficient of variation

\[ n \]

with possible

Theil's

works much better in terms of the use of one of the two corrected estimators for tends to underestimate the true, which is not generally equal to the mean squared error.

Breunig [4] applies that corrected estimator for the coefficient of variation squared to income data for China and Kenya, and compare that estimator to an alternative bias-corrected estimator, \( \bar{CV}^2_{\text{cor}} \), 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 \( \frac{1}{2} CV \), 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, \( \bar{CV}^2_{\text{cor}} \) 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

\[ GE(\alpha) = \frac{1}{\alpha} \left( \frac{1}{n} \sum_{j=1}^{n} \left( \frac{y_j}{\bar{y}} \right)^{\alpha} \right) - 1 \]

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 expressed as a function of the sampling error becomes smaller, regardless of the sample size. Breunig & Hutchinson [1] write
the GE measures as functions of the population mean, \( \mu \), 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

\[
ge(0) = \frac{1}{\tilde{N}} \left[ \sum_{j=1}^{N} W_j \log \left( \frac{y_j}{\tilde{y}_j} \right) \right] \tag{2}
\]

where \( \tilde{y} \) denotes the small sample mean calculated using sample weight, and \( \tilde{N} = \sum_{j=1}^{N} W_j \). 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, \( \text{ABias}(ge(0)) = \frac{1}{2} \mu \text{Var}(\hat{\mu}) \), and propose to extend this result to a complex sample by considering an heuristic solution, that is by substituting \( \mu \) with the weighted sample mean and \( \text{Var}(\hat{\mu}) \) 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, \( ge\text{Corr}(0) \), 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 \( ge\text{Corr}(0) \), in 1.5% sample, while it is equal to 7.9 and 2.6% respectively for \( ge(0) \) and \( ge\text{Corr}(0) \), 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 \( ge\text{Corr}(0) \), in 1.5% sample, while it is equal to 37.8 and 38.2% respectively for \( ge(0) \) and \( ge\text{Corr}(0) \), 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**
