THE BIAS-CORRECTED REGRESSION ESTIMATOR

Natalja Jurevitš

University of Tartu, Estonia

Abstract

It is well known that the regression estimator is biased for the total it has to estimate. The bias is small for big samples. Nevertheless, beside the sample size, the bias depends on the auxiliary variables, on their relation to the study variables and on the sampling design. It is important to know sources of the bias and in some cases to use the bias-corrected regression estimator.

The bias of the regression estimator is developed from its Taylor expansion and its main term has a general form (Musting, 2004):

$$\begin{aligned} B \approx & -vec'\left(Cov\left(\hat{\mathbf{t}}_{\mathbf{xy}}, \hat{\mathbf{t}}_{\mathbf{x}}\right)\right) vec\,\mathbf{T}^{-1} \\ & + vec'\left(Cov\left(vec\,\hat{\mathbf{T}}, \hat{\mathbf{t}}_{\mathbf{x}}\right)\right) vec\left(\mathbf{T}^{-1} \otimes \mathbf{t}'_{\mathbf{xy}} \mathbf{T}^{-1}\right), \end{aligned} \tag{1}$$

where the involved totals are:

$$\mathbf{T} = \sum_U \frac{\mathbf{x}_i \mathbf{x}'_i}{\sigma_i^2},\ \mathbf{t}_{xy} = \sum_U \frac{\mathbf{x}_i y_i}{\sigma_i^2},\ \mathbf{t}_x = \sum_U \mathbf{x}'_i,\ t_y = \sum_U y_i, \tag{2}$$

and $\hat{\mathbf{T}},\ \hat{\mathbf{t}}_{xy}, \hat{\mathbf{t}}_x, \hat{t}_y$ are their design-unbiased estimators.

In the presentation we will give the important special cases of the bias. We assume the model with one auxiliary variable both with and without intercept and the group mean model. We will consider different designs, e.g. SI and multinomial.

We construct the bias-corrected regression estimator:

$$\hat{t}_{y,corr} = \hat{t}_y - \hat{B}, \tag{3}$$

and study its properties.

In a simulation study we use the data taken from (Knottnerus, 2003). In addition to the existent real variables we simulate some different study variables, so that correlation between *y* and *x* is large, small or negative.

Practical study shows that in cases of small correlation the bias-corrected regression estimator (3) is more accurate than ordinary regression estimator. In most cases the variability of the corrected estimator is the same as of the ordinary regression estimator.

References

Musting, K. (2004) Study of the bias of generalized regression estimator. *Workshop on Survey Sampling Theory and Methodology*, June 18-22, 2004, Tartu, Estonia, p. 78-81

Knottnerus, P. (2003) *Sample Survey Theory. Some Pythagorean Perspectives.* Springer-Verlag, New York, p. 300.