Improvement of estimators in a linear regression problem in the case of violation of some classical assumptions

A.I.Sakhanenko, Yu.Yu.Linke

Sobolev Institute of Mathematics

Suppose we have observations $Y_1, Y_2, \ldots$ with $\mathbf{E}Y_i = \beta x_i$ for each i, where β is the main unknown parameter to be estimated and $\{x_{ni}\}$ are some numbers. It is well known that in the classical case, where $\{x_{ni}\}$ are known and $\{\mathbf{D}Y_i\}$ are equal, the best unbiased linear estimator for β_n exists and coincides with the least squares estimator $\beta^*_{no}(x_\bullet)$. But in practise it is more natural to assume that $\{x_i\}$ are unknown and that we are given some additional observations $\{X_i\}$ with $\{\mathbf{E}X_i = x_i\}$. At a first glance in these situations it is natural to replace the coefficients $\{x_{ni}\}$ by the observations $\{X_{ni}\}$ and use estimator $\beta^*_{no}(X_\bullet)$. But under the new conditions the estimator $\beta^*_{no}(X_\bullet)$ are non consistent even in the simplest case where $\{X_i\}$ are normal with the same variances.

With the aim to bypass this fundamental difficulty, the authors constructed in [2] and studied a new class of the explicit two-step estimates designed specifically for working in the presence of random errors in $\{x_i\}$. But this estimates have a downside: in their construction we assume that the distributions of $\{X_{ni}\}$ are known precisely. This requirement is difficult to meet in practice. We suggest in [3] the following assumption which holds considerably oftener.

Assumption: we observe independent random vectors $(Y_i, X_{ki},\ k = 1, \ldots, m_i)$, where $i = 1, 2 \ldots,$ and $m_i \geq 2$, consisting of independent components with $\{\mathbf{E}Y_i = \beta x_i\}$ Moreover, for all i the random variables $\{X_{ki},\ k = 1, \ldots, m_i\}$ have identical normal distributions with means x_i.

In such a situation authors suggest to use the following estimators

$$\beta^{**}_n = \frac{\sum_{i=1}^n \gamma_i(\beta^*_n, \overline{X}_i) Y_i}{\sum_{i=1}^n \gamma_i(\beta^*_n, \overline{X}_i)\overline{X}_i - \sum_{i=1}^n S_i^2 \gamma_i'(\beta^*_n, \overline{X}_i)}, \tag{1}$$

where $\overline{X}_i = \frac{1}{m_i}\sum_{k=1}^{m_i} X_{ki}$, $S_i^2 = \frac{1}{m_i(m_i-1)}\sum_{k=1}^{m_i}\left(X_{ki} - \overline{X}_i\right)^2$.

It is proved that estimators β^{**}_n are consistent and asymptotically normal for a wide class of functions $\{\gamma_i(\cdot,\cdot)\}$ and estimators β^*_n of the first step.

For example, if $\gamma_i(\cdot, \overline{X}_i) = \overline{X}_i$ and $\gamma_i(\cdot, \overline{X}_i) \equiv 1$ then the estimators $\beta^{**}_n = \beta^*_n$ improve the "classical" estimator $\beta^*_{no}(X_\bullet)$.

Estimators (1) with functions $\{\gamma_i(\beta, \overline{X}_i)\}$ depends on β may be useful in the geteroscedastic case where the variances $\{\mathbf{D}Y_i\}$ depend on the main parameter β.

[1] Linke Yu. Yu. and Sakhanenko A. I. *Asymptotically optimal estimation in a linear regression problem with random errors in coefficients.* Siberian Math. J. 2010 V.51 N 1 pp. 104–118.

[2] Sakhanenko A. I. and Linke Yu. Yu. *Improvement of estimators in a linear regression problem with random errors in coefficients* Siberian Math. J. 2011 V.52, N 1 pp. 113-126.

[3] Sakhanenko A. I. and Linke Yu. Yu. *Consistent estimation in a linear regression problem with random errors in coefficients* Siberian Math. J. 2011 V.52, N 4.