Theory of Stochastic Processes
Vol.13 (29), no.1-2, 2007, pp.57-65

OLENA GONTAR AND ANDRII MALENKO

SIMEX ESTIMATOR FOR POLYNOMIAL ERRORS-IN-VARIABLES MODEL

For polynomial errors-in-variables model, the Simex estimator is constructed in such way that it is consistent, as the samples size grows and the size of auxiliary sample is fixed. Then the estimator is modified in such a way that it shows good results for small samples without losing its asymptotic properties for large samples. Simulation studies corroborate the theoretical findings.

1. Introduction

We consider polynomial measurement error model

$$\begin{cases} y_i = \sum_{j=0}^{m} \beta_j \xi_i^j + \varepsilon_i, \\ x_i = \xi_i + \delta_i, \end{cases} \tag{1}$$

where y_i, x_i are observed and ξ_i are unobservable independent random variables, $i = \overline{1,n}$. Suppose that δ_i are i.i.d. normal random variables and their variance σ_δ^2 is known.

It is well known that the naive estimator of regression parameters β_0, $\beta_1, \ldots, \beta_m$, which ignores measurement errors is inconsistent. Cheng and Schneeweiss (1998) proposed the adjusted least squares $\hat{\beta}_{ALS}$ estimator in the model (1) which is consistent. This estimator can be viewed as resulting from the principle of corrected score due to Stefanski (1989) and Nakamura (1990). A small sample modification of $\hat{\beta}_{ALS}$ estimator was proposed in Cheng et al. (2000), such that it shows good results for small samples without losing its asymptotic properties for large samples.

Another estimator was introduced by Cook and Stefanski (1994) and is called Simex. The key idea underlying Simex is the fact that the effect of measurement error on an estimator can be determined experimentally via

2000 *Mathematics Subject Classification* 62J02, 62F12, 62-07.

Key words and phrases. Simex estimator, errors-in-variables models, Hermite polynomials

simulation. This is achieved by studying the naive regression estimator as a function f of measurement error variance in the regressors.

The purpose of this paper is to construct the **consistent** Simex estimator of the regression parameter. The observed variables are used for modeling the function f. This idea is close to the idea of Polzehl and Zwanzig (2005). Simulation studies show that for finite sample the Simex estimator in polynomial regression can sometimes produce extremely large estimating errors as well as the ALS estimator. It is proposed how to modify this estimator for small samples still preserving its asymptotic properties.

The paper is organized as follows. In the next section the polynomial errors-in-variables model is introduced and auxiliary lemmas are proved. Section 3 is devoted to construction of Simex estimator and the proof of its consistency. The small sample modification is proposed in Section 4. Section 5 gives some simulation results and shows the effect of modification, and Section 6 concludes. In the paper expectation is denoted as $\mathbf{E}$, the almost sure convergence as $\xrightarrow{P1}$, and the convergence in probability as $\xrightarrow{P}$.

2. Model and additional lemmas

We consider the polynomial errors-in-variables model of order $m \geq 1$,

$$\begin{cases} y_i = \beta_0 + \beta_1 \xi_i + \ldots + \beta_m \xi_i^m + \varepsilon_i, \\ x_i = \xi_i + \delta_i, \end{cases} \qquad i = \overline{1, n}.$$

Here $\{\xi_i,\ i \geq 1\}$, $\{\varepsilon_i,\ i \geq 1\}$, $\{\delta_i,\ i \geq 1\}$ are i.i.d. and mutually independent sequences. We assume that $\mathbf{E}|\xi_1|^m < \infty$, $\delta_1 \sim N(0, \sigma_\delta^2)$, σ_δ^2 is known, $\mathbf{E}\varepsilon_1 = 0$, $\mathbf{E}\varepsilon_1^2 < \infty$. The variances of ξ_1, ε_1, and δ_1 are supposed to be positive.

Denote $X_i = (1, x_i, \ldots, x_i^m)^t$. The naive, or ordinary least squares estimator of β is $\widehat{\beta}_{naive} = M_{XX}^{-1} M_{XY}$, where $M_{XX} := \overline{XX^t}$, $M_{XY} := \overline{Xy}$. Here the bar means averaging over n.

To introduce $\widehat{\beta}_{ALS}$ estimator consider the Hermite polynomials $h_k(x,t)$ of x which possess the following properties:

$$h_{-1}(x,t) = h_0(x,t) = 1, \quad h_{k+1}(x,t) = x h_k(x,t) + t k h_{k-1}(x,t), \quad k \geq 1,$$

and let $H(x,t)$ be the matrix of the following structure: $H_{rs}(x,t) = h_{r+s}(x,t)$, $r, s = 0, \ldots, m$. Denote the matrix $\frac{1}{n} \sum\limits_{i=1}^{n} H(x_i, -\sigma_\delta^2)$ as M_H and the vector $\{h_0(x_i, -\sigma_\delta^2), h_1(x_i, -\sigma_\delta^2), \ldots, h_m(x_i, -\sigma_\delta^2)\}^t$ as h_i. Then $\widehat{\beta}_{ALS}$ is defined as a solution to a linear equation:

$$M_H \widehat{\beta}_{ALS} = \frac{1}{n} \sum_{i=1}^{n} h_i. \tag{2}$$

To construct the simex estimator fix a number B. Consider standard normal i.i.d. sequence $\{\eta_{ib},\ i \geq 1,\ b = \overline{1,B}\}$, which is independent of other random variables in the model. Denote $x_{ib}(\lambda) = x_i + \eta_{ib}\sqrt{\lambda}$, $i \geq 1$, $b = \overline{1,B}$, and $X_{ib}(\lambda) = \{1, x_{ib}(\lambda), \ldots, x_{ib}^m(\lambda)\}^t$.

Introduce $M_{XX}(\lambda) = \overline{X(\lambda)X^t(\lambda)}$, $M_{XY}(\lambda) = \overline{X(\lambda)y}$. Hereafter the bar means averaging over n and b, e.g., $\overline{X(\lambda)y} = \frac{1}{nB}\sum_{i=1}^{n}\sum_{b=1}^{B} X_{ib}(\lambda)y_i$. The corresponding naive estimate of β is $\widehat{\beta}_{naive}(\lambda) = M_{XX}^{-1}(\lambda)M_{XY}(\lambda)$. For each λ introduce the matrix $M_H(\lambda) = \frac{1}{n}\sum_{i=1}^{n} H(x_i, \lambda)$. From Lemma1 below it follows that

$$M_{XX}(\lambda) = M_H(\lambda) + o(1), \text{ as } n \to \infty, \text{ a.s.} \tag{3}$$

Lemma 1. *Let ξ and δ be independent random variables with $\delta \sim N(0, \sigma_\delta^2)$. Then*

$$\boldsymbol{E}((\xi + \delta)^n | \xi) = h_n(\xi, \sigma_\delta^2).$$

Proof. To prove the next equality one should use partial integration

$$\mathbf{E}((\xi+\delta)^{n+1}|\xi) = \xi\mathbf{E}((\xi+\delta)^n|\xi) + E(\delta(\xi+\delta)^n|\xi) =$$

$$= \xi\mathbf{E}((\xi+\delta)^n|\xi) + (n-1)\sigma_\delta^2\mathbf{E}(\delta(\xi+\ \delta)^{n-1}|\xi).$$

Then induction is used. □

Lemma 2. *Let $X = (1, x, \ldots, x^m)^t$ and $h(x,t) = (h_0(x,t), \ldots, h_m(x,t))^t$, and T be a transition matrix: $h(x,t) = T(t)X$. Then $T(t+s) = T(t)T(s)$ and $T(-t) = T^{-1}(t)$, $t, s \in \mathbb{R}$.*

Proof. Assume that s and t are positive real numbers. Let $x = \xi + \delta + \gamma$, where $\delta \sim N(0, s)$, $\gamma \sim N(0, t)$, $s > 0$, $t > 0$, and ξ, δ, γ are mutually independent. Let

$$\rho = (1, \xi, \ldots, \xi^m)^t,\ \psi = (1, \xi+\delta, \ldots, (\xi+\delta)^m)^t.$$

By Lemma 1 we can write $\mathbf{E}(X|\xi) = h(\xi, s+t) = T(s+t)\rho$. But

$$\mathbf{E}(X|\xi) = \mathbf{E}(\mathbf{E}(X|\xi,\delta),\xi) = \mathbf{E}(h(\xi+\delta,t)|\xi) =$$
$$= \mathbf{E}(T(t)\psi|\xi) = T(t)h(\xi,s) = T(t)T(s)\rho.$$

Thus for positive real numbers we proved that $T(t+s) = T(t)T(s)$. This equality is extended for arbitrary real numbers, because the Hermite polynomials can be constructed for any real parameter t and entries of $T(t)$ are polynomials on t. Then $T(-t)T(t) = T(0) = I$. □

Now using Lemmas 1 and 2 we get that $\mathbf{E}(M_{XY}(\lambda)|X,y) = T(\lambda)M_{XY}$, therefore a.s.

$$M_{XY}(\lambda) = T(\lambda)M_{XY} + o(1), \text{as } n \to \infty. \tag{4}$$

3. Simex estimator

The following model is proposed for fitting the naive estimators:

$$\widehat{\beta}(\lambda,\theta) = M_H^{-1}(\lambda)T(\lambda)\theta.$$

Let $K \geq 1$, $0 = \lambda_1 < \lambda_2 < \ldots < \lambda_K$. The parameter θ is estimated by least squares method as $\widehat{\theta} = \operatorname*{argmin}_\theta \sum\limits_{k=1}^{K} ||\widehat{\beta}_{naive}(\lambda_k) - \widehat{\beta}(\theta,\lambda_k)||^2$. Thus $\widehat{\theta}$ equals

$$\widehat{\theta}^t = \left(\sum_{k=1}^{K} M_{XY}^t(\lambda_k)M_{XX}^{-1}(\lambda_k)M_H^{-1}(\lambda_k)T(\lambda_k)\right)\left(\sum_{k=1}^{K} T^t(\lambda_k)M_H^{-2}(\lambda)T(\lambda_k)\right)^{-1}$$

Using (3) and (4) it is easy to see that a.s.

$$\widehat{\theta} = M_{XY} + o(1), \text{ as } n \to \infty. \tag{5}$$

We define the Simex estimator as

$$\widehat{\beta}_{Simex} := \widehat{\beta}(-\sigma_\delta^2, \widehat{\theta}) = M_H^{-1}(-\sigma_\delta^2)T(-\sigma_\delta^2)\widehat{\theta}.$$

Theorem 1. *Under the model assumptions, the Simex estimator is strongly consistent:*

$$\widehat{\beta}_{Simex} \xrightarrow{P1} \beta, \text{ as } n \to \infty.$$

Proof. Using Lemmas 1 and 2 we can prove that

$$\frac{1}{n}\sum_{i=1}^{n} h_k(x_i,\lambda) \xrightarrow{P1} \mathbf{E}h_k(x,\lambda) = \mathbf{EE}((x+\sqrt{\lambda}\varepsilon)^k|\ x) =$$
$$= \mathbf{E}(x+\sqrt{\lambda}\varepsilon)^k = \mathbf{E}(\xi+\delta+\sqrt{\lambda}\varepsilon)^k = \mathbf{EE}((\xi+\delta+\sqrt{\lambda}\varepsilon)^k|\ \xi) = \mathbf{E}h_k(\xi,\lambda+\sigma_\delta^2).$$

Thus substituting $(-\sigma_\delta^2)$ for λ we obtain

$$\frac{1}{n}\sum_{i=1}^{n} h_k(x_i,-\sigma_\delta^2) \to \mathbf{E}h_k(\xi,-\sigma_\delta^2+\sigma_\delta^2) = \mathbf{E}h_k(\xi,0) = \mathbf{E}\xi^t.$$

Hence $M_H(-\sigma_\delta^2) \xrightarrow{P1} \mathbf{E}\rho\rho^t$, where $\rho := (1,\xi,\ldots,\xi^m)^t$. Using (5) and Lemma1 again, we obtain

$$\widehat{\theta} \xrightarrow{P1} \mathbf{E}M_{XY} = \mathbf{EE}(M_{XY}|\ \xi) = \mathbf{E}T(\sigma_\delta^2)\rho\rho^t\beta = T(\sigma_\delta^2)\mathbf{E}\rho\rho^t\beta.$$

The consistency of Simex estimator is obvious from Lemma 2:

$$\widehat{\beta}_{Simex} = \widehat{\beta}(-\sigma_\delta^2, \widehat{\theta}) \xrightarrow{P1} (\mathbf{E}\rho\rho^t)^{-1} T(-\sigma_\delta^2) T(\sigma_\delta^2) \mathbf{E}\rho\rho^t \beta = \beta. \qquad \square$$

Remark. In the special case $K = 1$ we have $\lambda_1 = 0$ and transition matrix $T(\lambda_1) = T(0) = I_m$ (the identity matrix). The matrix $M_H(\lambda_1) = M_H(0) = = M_{XX}$. We notice that $M_{XX}(\lambda_1) = M_{XX}(0) = M_{XX}$, and $M_{XY}(\lambda_1) = = M_{XY}(0) = M_{XY}$. Hence we obtain that $\widehat{\theta} = M_{XY}$, $\widehat{\beta}_{Simex} = M_H^{-1}(-\sigma_\delta^2)T(-\sigma_\delta^2)M_{XY}$. Therefore

$$M_H(-\sigma_\delta^2)\widehat{\beta}_{Simex} = T(-\sigma_\delta^2)M_{XY}. \tag{6}$$

Thus $\widehat{\beta}_{Simex}$ is the solution to the equation (6). But this equation (in current notations) is the same as the equation (2) for the ALS estimator of β. So in the case $K = 1$ the Simex estimator coincides with the ALS estimator.

4. Small sample modification

From $\widehat{\beta}_{Simex} = M_H^{-1}(-\sigma_\delta^2)T(-\sigma_\delta^2)\widehat{\theta}$ we can write that $\widehat{\beta}_{Simex}$ is the solution to the following equation:

$$M_H(-\sigma_\delta^2)\widehat{\beta}_{Simex} = T(-\sigma_\delta^2)\widehat{\theta}. \tag{7}$$

We have $M_H(-\sigma_\delta^2) \xrightarrow{P1} E\rho\rho^t$, as $n \to \infty$, and therefore it is positive definite for $n \geq n_0(w)$ a.s. But for small samples, however, $M_H(-\sigma_\delta^2)$ can be indefinite and this can cause significant bias for the Simex estimator. Introduce $V_i = h(x_i, -\sigma_\delta^2)h^t(x_i, -\sigma_\delta^2) - H(x_i, -\sigma_\delta^2)$. Taking average over n we can write $\overline{V} = \overline{h(-\sigma_\delta^2)h^t(-\sigma_\delta^2)} - M_H(-\sigma_\delta^2)$. Using this relation, the estimation equation (7) can be rewritten as

$$\left(\overline{h(-\sigma_\delta^2)h^t(-\sigma_\delta^2)} - \overline{V}\right)\widehat{\beta}_{Simex} = T(-\sigma_\delta^2)\widehat{\theta}. \tag{8}$$

Define λ as the smallest positive root of the equation $\det(A - \lambda B) = 0$, where

$$A = \begin{pmatrix} \overline{y^2} & \overline{yh^t(-\sigma_\delta^2)} \\ \overline{h(-\sigma_\delta^2)y} & \overline{h(-\sigma_\delta^2)h^t(-\sigma_\delta^2)} \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 0 & \overline{V} \end{pmatrix}.$$

We assume that A is positive definite.

To construct small sample modification of Simex estimator we use the same approach as in Cheng et al. (2000) is used. Proofs of the next two theorems are similar to that paper.

The modified Simex estimator can be found as a solution to the equation:

$$\left(\overline{h(-\sigma_\delta^2)h^t(-\sigma_\delta^2)} - a\overline{V}\right)\widehat{\beta}_{MSimex} = T(-\sigma_\delta^2)\widehat{\theta}. \tag{9}$$

Here a is defined as

$$\begin{cases} a = (n-\alpha)/n, & \text{if } \lambda > 1+\frac{1}{n}, \\ a = \lambda(n-\alpha)/(n+1), & \text{if } \lambda \le 1+\frac{1}{n}, \end{cases} \tag{10}$$

with some $\alpha < n$ to be chosen so that the resulting estimator possesses better small sample properties. The number $\alpha = m+1$ is the lowest α that one should choose, see the discussion in Cheng et al.(2000).

Theorem 2. *The following inequality holds a.s.:*

$$\overline{h(-\sigma_\delta^2)h^t(-\sigma_\delta^2)} - a\overline{V} \ge \frac{\alpha+1}{n+1}\overline{h(-\sigma_\delta^2)h^t(-\sigma_\delta^2)} > 0. \tag{11}$$

(Hereafter inequalities for matrices are understood in Lowener order.)

Proof. As A is positive definite it can be decomposed as $A = CC^t$ with a nonsingular matrix C. Define $\widetilde{B} = C^{-1}BC^{-t}$. Let d be the largest eigenvalue of $\widetilde{B}$. As the second diagonal element of $\overline{V}$, $\overline{h_1^2(-\sigma_\delta^2)} - \overline{h_2(-\sigma_\delta^2)} = \sigma_\delta^2$, is positive, $\widetilde{B}$ has at least one positive eigenvalue, and therefore $d > 0$. It follows that $\lambda = \frac{1}{d}$. Let D be the diagonal matrix of eigenvalues of $\widetilde{B}$ and E be a matrix, the columns of which are the corresponding normalized eigenvectors. Then $\widetilde{B} = EDE^t, EE^t = I$. It follows that $B = CEDE^tC^t$, with $T = CE$ we have $A = TT^t$, and $B = TDT^t$. Hence for any scalar c,

$$A - cB = T(I - cD)T^t, \tag{12}$$

with a nonsingular matrix T.

In the first case, when $\lambda > 1+\frac{1}{n}$, we see that $d < n(n+1)$ and therefore $D < n(n+1)I$. Hence

$$I - aD = I - \frac{n-\alpha}{n}D > \frac{\alpha+1}{n+1}I.$$

In the second case $\lambda \le 1+\frac{1}{n}$. In general $d^{-1}D \le I$. This implies that

$$I - aD = I - \frac{\lambda(n-\alpha)}{n+1}D = I - \frac{(n-\alpha)}{n+1}d^{-1}D \ge \frac{\alpha+1}{n+1}I.$$

Thus in both cases we obtain $A - aB \ge \frac{\alpha+1}{n+1}A > 0$. Deleting the first row and column of these matrices results in (11). □

Theorem 3. *The modified estimator* $\widehat{\beta}_{MSimex}$ *is asymptotically equivalent to unmodified one* $\widehat{\beta}_{Simex}$*:*

$$\sqrt{n}(\widehat{\beta}_{MSimex} - \widehat{\beta}_{Simex}) \xrightarrow{P} 0, \ \textit{as } n \to \infty.$$

Proof. First, prove that $P(\lambda > 1)$ converges to 1, as $n \to \infty$. Condition $\lambda > 1$ is equivalent to $d < 1$ or $D < I$. According to (12) this is equivalent to $A > B$. As in the proof of Theorem 2, it can be shown that $A > B$ is equivalent to $\overline{h(-\sigma_\delta^2)h^t(-\sigma_\delta^2)} - \overline{V} > 0$.

Since $\overline{h(-\sigma_\delta^2)h^t(-\sigma_\delta^2)} - \overline{V} = M_H(-\sigma_\delta^2)$ converges to the matrix $E(\rho\rho^t)$, which is positive definite with probability 1, one can state that $P(A > B) =$ $= P(\lambda > 1)$ which converges to 1, as $n \to \infty$. Now for $\lambda > 1$ we have, by the definition of a, that

$$\frac{n-\alpha}{n+1} < a \le \frac{n-\alpha}{n},$$

and after some algebra

$$\frac{\alpha+1}{\sqrt{n}} > (1-a)\sqrt{n} \ge \frac{\alpha}{\sqrt{n}}.$$

This inequality holds with probability tending to 1, as $n \to \infty$. Since outer parts of this inequality converge to 0, we have $(1-a)\sqrt{n} \xrightarrow{P} 0$, as $n \to \infty$.

By subtracting equation (9) from (8) we derive after some algebra

$$(\overline{h(-\sigma_\delta^2)h^t(-\sigma_\delta^2)} - a\overline{V})(\widehat{\beta}_{Simex} - \widehat{\beta}_{MSimex})\sqrt{n} = (1-a)\sqrt{n}\ \overline{V}\widehat{\beta}_{Simex}.$$

The right-hand side converges to 0, whereas $\overline{h(-\sigma_\delta^2)h^t(-\sigma_\delta^2)} - a\overline{V} > 0$, therefore

$$\sqrt{n}(\widehat{\beta}_{MSimex} - \widehat{\beta}_{Simex}) \xrightarrow{P} 0, \text{ as } n \to \infty. \qquad \square$$

5. Simulation results

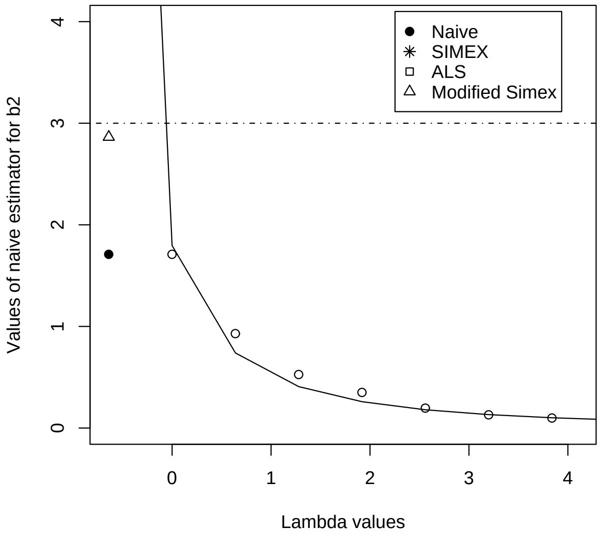
Simulation was made in R-package. We studied the quadratic model

$$y_i = b_0 + b_1\xi_i + b_2\xi_i^2 + \varepsilon_i,\ x_i = \xi_i + \delta_i.$$

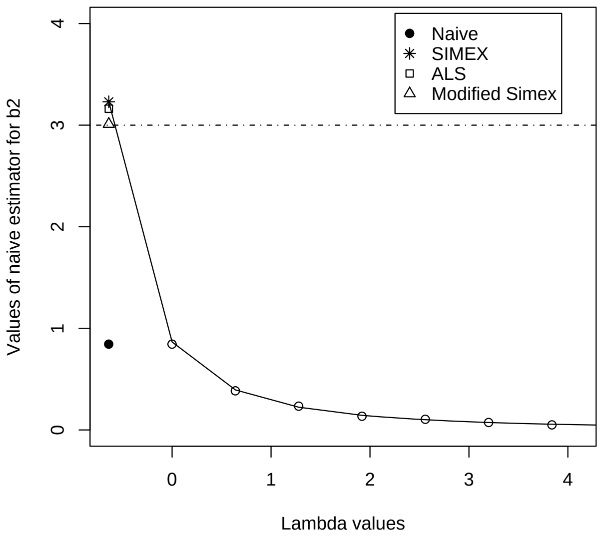
We specified ε_i and δ_i as normally distributed variables with $\mathbf{E}\varepsilon_i\delta_i = 0$ and $\sigma_\delta^2 = \sigma_\varepsilon^2 = 0.25$ and $\sigma_\xi^2 = 1$. The sample size n equals 20. For Simex estimator the following values were used: $B = 100$, $K = 11$, $\lambda_k = k\sigma_\delta^2$, $k = \overline{0,10}$. True values were $b_0 = 5, b_1 = 6, b_2 = 3$.

The simulation results are plotted below for the parameter b_2. The naive estimator is denoted by solid circle, the ALS by square, the Simex by star, and the modified Simex by triangle. Circles correspond to naive estimators with larger variance. Solid line describes the behavior of fitted model and dashed line denotes the true value of the parameter.

In the first picture $M_H(-\sigma_\delta^2)$ is not positive definite, and as a result the Simex estimator has extremely large estimating error ($\widehat{\beta}_{Simex}$=35.03, while $\widehat{\beta}_{MSimex}$=2.91).



In the second picture $M_H(-\sigma_\delta^2)$ is positive definite, and the Simex estimator is a good one ($\widehat{\beta}_{Simex}$=3.19 and $\widehat{\beta}_{MSimex}$=3.04).



6. Conclusion

In the article the Simex estimator for polynomial errors-in-variables model is constructed. It differs from the classical Simex estimator proposed by Cook and Stefanski(1995) due to the fact that the observed variables are used to model the naive estimator as a function of extra variance. The consistency of constructed Simex estimator is proved. Then this estimator is modified such that it shows good results for small samples without losing its asymptotic properties for large samples. Simulation studies made in statistical package *R* corroborate the theoretical result.

Bibliography

1. Cheng, C.-L., and Schneeweiss, H. (1998). *Polynomial regression with errors in the variables.* J.R. Statist. Soc. B, 60, 189-199.
2. Cheng, C.-L., Schneeweiss, H., and Thamerus, M. (2000). *A small sample estimator for a polynomial regression with errors in the variables.* J.R. Statist. B, 62, 699-709.
3. Cook, J., and Stefanski, L.A. (1995). *A simulation extrapolation method for parametric measurement error models.* Journal of the American Statistical Association, 89, 1314-1328.
4. Nakamura, T. (1990). *Corrected Score functions for errors-in-variables models: methodology and application to generalized linear models.* Biometrika, 77, 127-137.
5. Polzehl, J., and Zwanzig, S.(2005) *Simex and TLS: an equivalence result.* WIAS, Technical Report 999, Berlin.
6. Stefanski, L. A. (1989). *Unbiased estimation of a nonlinear function of a normal mean with application to measurement error model.* Computation in Statistics, Series A, 18, 4335-4358.

Department of Probability Theory and Mathematical Statistics, Kyiv National Taras Shevchenko University, Kyiv, Ukraine
E-mail address: gontaro@ukr.net.

Department of Probability Theory and Mathematical Statistics, Kyiv National Taras Shevchenko University, Kyiv, Ukraine
E-mail address: exipilis@yandex.ru.