Proceedings of the 8[th] International Conference on Applied Informatics
Eger, Hungary, January 27–30, 2010. Vol. 1. pp. 247–253.

Nonparametric Regression and Measurement Error*

István Fazekas[a], Alexander Kukush[b]

[a] University of Debrecen, Faculty of Informatics
e-mail: fazekasi@inf.unideb.hu

[b] Kyiv National Taras Shevchenko University, Mechanical and Mathematical Department
e-mail: Alexander_Kukush@univ.kiev.ua

Abstract

Nonparametric regression estimations are considered when the variables are measured with error. SIMEX type modifications of certain function approximation procedures are studied. Asymptotic properties of the estimator are obtained.

Keywords: Non-linear regression, measurement error, nonparametric estimation, SIMEX estimator, asymptotic variance.

MSC: 62G08, 62G20.

1. Measurement error models

Consider the simplest model (linear regression)

$$Y_t = \beta_0 + \beta_1 U_t + \varepsilon_t, \quad t = 1, \dots, n.$$

Usually the least squares estimator $\widehat{\beta}_1$ of β_1 is applied. However, if we can observe the variable U_t with error, i.e. instead of U_t we observe

$$X_t = U_t + \delta_t,$$

then $\widehat{\beta}_1$ will not be unbiased (see [8]).

If the original estimator is used without modification in the case of measurement error, then we call it naive estimator. In most cases one has to modify the naive estimator. To this end additional knowledge is necessary about the error. E.g. the

*Partially supported by the Hungarian Foundation of Scientific Researches under Grant No. OTKA T-079128.

least squares estimator was modified by a deconvolution method by Fazekas and Kukush [6].

The total least squares (TLS) method is often used in errors-in-variables models (see, e.g. [12]). The TLS method is a very useful tool of applied statistics. However, Fazekas, Kukush and Zwanzig [7] proved the inconsistency of the TLS method, moreover a correction of the TLS was suggested.

This paper is devoted to the study of the asymptotic behaviour of the non-parametric SIMEX estimator. In Section 2 we describe the well-known parametric SIMEX estimator. That estimator was introduced by Cook and Stefanski [5] for parametric regression problems to handle measurement errors. The idea of the SIMEX estimator was applied for nonparametric models by Carroll, Maca and Ruppert [3]. In Section 3 we present the nonparametric SIMEX estimator. The main result of the paper is Theorem 4.1. It describes the asymptotic behaviour of the nonparametric SIMEX estimator. In Section 4 a complete proof of Theorem 4.1 is presented.

2. The parametric SIMEX estimator

First consider the parametric SIMEX estimator proposed in [5]. SIMEX means SIMulation and EXtrapolation. Consider the parametric regression model

$$Y = f(V, U, \theta) + \varepsilon$$

where Y is the response variable, V is the covariate measured without error, U is the predictor, ε is the random error term. f is a known function, but the parameter is unknown. We have to estimate θ.

The problem is that we can observe U with error:

$$X = U + \sigma Z$$

can be measured instead of U. Here Z is standard normal and it is independent of Y, U and V. We assume that σ is either known or well estimated.

Denote $Y_i, V_i, U_i, X_i,\ i = 1, \dots, n$, the sample. We assume that there is an estimation procedure T with good statistical properties:

$$\widehat{\theta}_{\mathrm{TRUE}} = T(\{Y_i, V_i, U_i \,:\, i = 1, \dots, n\}).$$

However this procedure is not feasible because $U_i,\ i = 1, \dots, n$, are not available. If we use in the original estimator for the variables measured with error, we obtain the so called naive estimator

$$\widehat{\theta}_{\mathrm{NAIVE}} = T(\{Y_i, V_i, X_i \,:\, i = 1, \dots, n\}).$$

However, this estimator can have unpleasant statistical properties.

Now generate additional random variables $Z_{b,i}$, $i = 1, \ldots, n$. We assume that $Z_{b,i}$, $i = 1, \ldots, n$, are independent standard normal random variables and they are independent of Y_i, V_i, U_i, X_i, $i = 1, \ldots, n$. For $\lambda > 0$, define

$$X_{b,i}(\lambda) = X_i + \lambda^{1/2}\sigma Z_{b,i}, \quad i = 1, \ldots, n.$$

Apply our original estimation procedure to the sample $\{Y_i, V_i, X_{b,i} : i = 1, \ldots, n\}$. We obtain

$$\widehat{\theta}_b(\lambda) = T(\{Y_i, V_i, X_{b,i} : i = 1, \ldots, n\}).$$

We need the conditional expectation

$$\widehat{\theta}(\lambda) = E(\widehat{\theta}_b(\lambda)|\{Y_i, V_i, X_i : i = 1, \ldots, n\}).$$

However the exact calculation of $\widehat{\theta}(\lambda)$ is not feasible. We have to approximate it by an appropriate arithmetic mean.

Generate independent standard normal random variables $Z_{b,i}$, $b = 1, \ldots, B, i = 1, \ldots, n$, being independent of the original sample Y_i, V_i, U_i, X_i, $i = 1, \ldots, n$. Then

$$\tilde{\theta}(\lambda) = \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}_b(\lambda)$$

is a good approximation of $\widehat{\theta}(\lambda)$. As the variance of the error in the observed predictor variables is $(1+\lambda)\sigma^2$, we consider $\lambda = -1$ as the predictor without error.

So the SIMEX estimator is the extrapolation of $\tilde{\theta}(\lambda)$ back to $\lambda = -1$.

Cook and Stefanski [5] presented numerical (stochastic simulation) results for the asymptotic behaviour of the SIMEX estimator. Carroll, Küchenhoff, Lombard and Stefanski [2] obtained partial results and numerical evidence for the asymptotic normality of the SIMEX estimator. Gontar and Küchenhoff [9] gave an expansion of the SIMEX estimator with small measurement errors.

3. The nonparametric SIMEX estimator

Assume that we have a relationship between y and $\boldsymbol{\xi}$ of the form

$$y = f(\boldsymbol{\xi}) + \varepsilon \tag{3.1}$$

where y, $\boldsymbol{\xi}$, and ε are random, ε is the unobservable error term and f is a fixed but unknown deterministic function. Our aim is to estimate f.

Assume that when y and $\boldsymbol{\xi}$ are observable then we can estimate f consistently. That is when the observations $(y_1, \boldsymbol{\xi}_1), \ldots, (y_n, \boldsymbol{\xi}_n)$ for $(y, \boldsymbol{\xi})$ are given, then the estimator $\widehat{f}_n$ based on $(y_1, \boldsymbol{\xi}_1), \ldots, (y_n, \boldsymbol{\xi}_n)$ can be calculated and

$$\lim_{n\to\infty} \widehat{f}_n = f \quad \text{in probability.} \tag{3.2}$$

$\widehat{f}_n$ can be produced by a known method (an estimating procedure, a learning algorithm) that can be applied for any input data.

However, we can not observe the precise $\boldsymbol{\xi}$. We observe it with error, that is we observe

$$\boldsymbol{x} = \boldsymbol{\xi} + \boldsymbol{\delta}. \tag{3.3}$$

Here $\boldsymbol{x}$, $\boldsymbol{\xi}$, and $\boldsymbol{\delta}$ are q-dimensional random vectors. $\boldsymbol{\delta}$ is an unobservable measurement error. We assume that $\boldsymbol{\delta}$ is independent of $(\boldsymbol{\xi}, \varepsilon)$, $\mathbb{E}\boldsymbol{\delta} = \mathbf{0}$, $\mathrm{var}(\boldsymbol{\delta}) = \sigma^2 I_q$ (where I_q denotes the $q \times q$ unit matrix). Assume that $\sigma^2 > 0$ is known. We deal with the case when y is onedimensional.

Let the observations $(y_1, \boldsymbol{x}_1), \dots, (y_n, \boldsymbol{x}_n)$ for $(y, \boldsymbol{x})$ be given. It means that $(y_i, \boldsymbol{x}_i)$ has the same distribution as $(y, \boldsymbol{x})$ for each i. (More precisely, $(y_i, \boldsymbol{\xi}_i, \varepsilon_i)$ has the same distribution as $(y, \boldsymbol{\xi}, \varepsilon)$ and $\boldsymbol{x}_i = \boldsymbol{\xi}_i + \varepsilon_i$ for each i. We shall not use explicitly that the sample consists of i.i.d. observations because our conditions will be given in terms of $\widehat{f}_n$.) Our starting point is the so called naive estimator. That is the original estimation procedure is applied to the data with error. Our estimating procedure offers $\widehat{f}_n(t, \sigma^2)$ that is based on the observations with error. More precisely

$$\widehat{f}_n(t, \sigma^2) = \widehat{f}_n[\boldsymbol{x}_1, \dots, \boldsymbol{x}_n; y_1, \dots, y_n](t) \tag{3.4}$$

denotes the estimator based on the sample with error having variance σ^2 that is on $(y_1, \boldsymbol{x}_1), \dots, (y_n, \boldsymbol{x}_n)$.

In the particular case of $\sigma^2 = 0$ we would obtain the true estimator

$$\widehat{f}_n(t, 0) = \widehat{f}_n[\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n; y_1, \dots, y_n](t)$$

based on the (unavailable) values of $(y_1, \boldsymbol{\xi}_1), \dots, (y_n, \boldsymbol{\xi}_n)$. The consistency condition (3.2) is equivalent to

$$\lim_{n \to \infty} \widehat{f}_n(t, 0) = f(t) \quad \text{in probability for each } t \in \mathbb{R}. \tag{3.5}$$

Carroll, Maca and Ruppert [3] proposed the SIMEX estimator for certain nonparametric problems. One has to modify the naive estimator. Generate an additional sample independent of our original observations, i.e. let

$$\{\boldsymbol{\delta}_{i,b} \,:\, i = 1, \dots, n, b = 1, \dots, B\}$$

be i.i.d. q-dimensional standard normal vectors. Here $B > 0$ is a fixed integer. Let $\Lambda = \{\lambda_0, \lambda_1, \dots, \lambda_m\}$ be given pairwise distinct numbers with $\lambda_0 = 0$ and $\lambda_1 > 0, \dots, \lambda_m > 0$. Here m is a fixed positive integer. Now we 'increase' the variance of the error in the following way. For each $\lambda \in \Lambda$ let

$$\boldsymbol{x}_{i,b}(\lambda) = \boldsymbol{x}_i + \sqrt{\lambda}\sigma\boldsymbol{\delta}_{i,b},$$

$i = 1, \dots, n$, $b = 1, \dots, B$. We obtain the estimation $\widehat{f}_{n,b}(t, (1+\lambda)\sigma^2)$ applying our original estimator to the sample with increased error variance, i.e. to $(y_1, \boldsymbol{x}_{1,b}(\lambda), \dots, (y_n, \boldsymbol{x}_{n,b}(\lambda))$:

$$\widehat{f}_{n,b}(t, (1+\lambda)\sigma^2) = \widehat{f}_n[\boldsymbol{x}_{1,b}(\lambda), \dots, \boldsymbol{x}_{n,b}(\lambda); y_1, \dots, y_n](t). \tag{3.6}$$

Now create the average of these estimators

$$\bar{f}_n(t,(1+\lambda)\sigma^2) = \frac{1}{B}\sum_{b=1}^{B} \widehat{f}_{n,b}(t,(1+\lambda)\sigma^2). \tag{3.7}$$

For each fixed t we fit by least squares method a polynomial of degree m to the values $\bar{f}_n(t,(1+\lambda_0)\sigma^2),\ldots,\bar{f}_n(t,(1+\lambda_m)\sigma^2)$. That is let $\boldsymbol{\varrho}(\lambda) = (1,\lambda,\ldots,\lambda^m)^\top$ where $^\top$ denotes the transpose.

Then

$$\boldsymbol{\gamma}^\top \boldsymbol{\varrho}(\lambda), \quad \boldsymbol{\gamma} \in \mathbb{R}^{m+1},$$

is a polynomial of degree m. Let

$$\widehat{\boldsymbol{\gamma}}(t) = \arg \min_{\boldsymbol{\gamma}\in\mathbb{R}^{m+1}} \sum_{\lambda\in\Lambda} \left(\bar{f}_n(t,(1+\lambda)\sigma^2) - \boldsymbol{\gamma}^\top \boldsymbol{\varrho}(\lambda)\right)^2. \tag{3.8}$$

Finally

$$\widetilde{f}(t) = (\boldsymbol{\varrho}(-1))^\top \widehat{\boldsymbol{\gamma}}(t) \tag{3.9}$$

is the SIMEX estimator of $f(t)$.

4. Asymptotic properties of the SIMEX estimator

To obtain asymptotic properties of $\widetilde{f}(t)$, we need the following condition on $\widehat{f}$. Let $\lambda_{\max} = \max_{1\le i\le m}\lambda_i$. Denote C^k the set of k-times continuously differentiable functions. Assume that there exists a $\sigma_0^2 > 0$ and a deterministic function $f_\infty(t,u)$, $0 \le u \le (1+\lambda_{\max})\sigma_0^2$ such that $f_\infty(t,.) \in C^{l+1}[0,(1+\lambda_{\max})\sigma_0^2]$ for each t and such that

$$\lim_{n\to\infty} \widehat{f}_{n,b}(t,(1+\lambda)\sigma^2) = f_\infty(t,(1+\lambda)\sigma^2) \quad \text{in probability} \tag{4.1}$$

for each t, $\lambda\in\Lambda$, and $0\le\sigma^2\le\sigma_0^2$.

Let $g(s) = \mathrm{O}(h(s))$ denote that $\limsup_{s\to 0} g(s)/h(s) < \infty$, while $\mathrm{o}_P(1)$ denotes a quantity that converges to 0 in probability, as $n\to\infty$.

Theorem 4.1. *Let $l \le m$. Assume that conditions* (3.2) *and* (4.1) *are satisfied. Then for the SIMEX estimator defined by* (3.9) *we have*

$$\widetilde{f}(t) = f_S(t,\sigma^2) + \mathrm{o}_P(1), \quad \textit{as} \quad n\to\infty, \tag{4.2}$$

for each t and $0\le\sigma^2\le\sigma_0^2$ where

$$f_S(t,\sigma^2) = f(t) + \mathrm{O}(\sigma^{2l+2}), \quad \textit{as} \quad \sigma\to 0. \tag{4.3}$$

Proof. Let t, $\lambda\in\Lambda$, and $0\le\sigma^2\le\sigma_0^2$ be fixed. Then, by (4.1), as $n\to\infty$,

$$\widehat{f}_{n,b}(t,(1+\lambda)\sigma^2) = f_\infty(t,(1+\lambda)\sigma^2) + \mathrm{o}_P(1). \tag{4.4}$$

Using Taylor's expansion, we have

$$\widehat{f}_{n,b}(t,(1+\lambda)\sigma^2) = f_\infty(t,0) + \sum_{j=1}^{l} \frac{1}{j!} f_\infty^{(j)}(t,0)(1+\lambda)^j\sigma^{2j} + \mathrm{O}(\sigma^{2l+2}) + \mathrm{o}_P(1). \quad (4.5)$$

Here $f_\infty^{(j)}(t,0)$ denotes the j-th partial derivative with respect to the second argument at point $(t,0)$, moreover the asymptotic behaviour is meant that first $n\to\infty$, then $\sigma\to 0$. Now, by assumptions (3.2) and (4.1), $f_\infty(t,0)=f(t)$. By (4.5) and (3.7)

$$\bar{f}_n(t,(1+\lambda)\sigma^2) = f(t) + \sum\nolimits_{j=1}^{l} \frac{a_j}{j!}(1+\lambda)^j\sigma^{2j} + \mathrm{O}(\sigma^{2l+2}) + \mathrm{o}_P(1). \quad (4.6)$$

Consider now the optimization problem (3.8). It is equivalent to the ordinary least squares estimation in the linear model

$$\boldsymbol{f} \approx A\boldsymbol{\gamma}$$

where

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & \lambda_1 & \lambda_1^2 & \dots & \lambda_1^m \\ 1 & \lambda_2 & \lambda_2^2 & \dots & \lambda_2^m \\ \vdots & & & & \\ 1 & \lambda_m & \lambda_m^2 & \dots & \lambda_m^m \end{pmatrix},$$

$$\boldsymbol{f} = \begin{pmatrix} \bar{f}_n(t,(1+\lambda_0)\sigma^2) \\ \bar{f}_n(t,(1+\lambda_1)\sigma^2) \\ \vdots \\ \bar{f}_n(t,(1+\lambda_m)\sigma^2) \end{pmatrix} = \quad (4.7)$$

$$= \begin{pmatrix} f(t) \\ f(t) \\ \vdots \\ f(t) \end{pmatrix} + \sum_{j=1}^{l} \frac{a_j}{j!}\sigma^{2j} \begin{pmatrix} (1+\lambda_0)^j \\ (1+\lambda_1)^j \\ \vdots \\ (1+\lambda_m)^j \end{pmatrix} + \mathrm{O}(\sigma^{2l+2}) + \mathrm{o}_P(1).$$

The solution is $\widehat{\boldsymbol{\gamma}} = (A^\top A)^{-1}A^\top\boldsymbol{f}$. Actually it gives the coefficients of the approximating polynomial (having degree at most m). The operation $(A^\top A)^{-1}A^\top$ can be applied for each summand in the above expression of $\boldsymbol{f}$. As $m\ge l\ge j$, the approximation to the polynomial $(1+\lambda)^j$ is precise, therefore the coefficients are

$$\binom{j}{0}, \binom{j}{1}, \dots, \binom{j}{j}, 0, \dots, 0.$$

Its combination with the coordinates of $\boldsymbol{\varrho}(-1)$ that is with $+1$'s and -1's is zero. Therefore if we write (4.7) into the SIMEX estimator

$$\widetilde{f}(t) = (\boldsymbol{\varrho}(-1))^\top\widehat{\boldsymbol{\gamma}}(t) = (\boldsymbol{\varrho}(-1))^\top(A^\top A)^{-1}A^\top\boldsymbol{f}, \quad (4.8)$$

the terms of the sum $\sum_{j=1}^{l}\dots$ disappear. So we obtain (4.2)–(4.3). □

References

[1] Berényi, P., Valyon, J., Horváth, G. *Neural Modeling of an Industrial Process with Noisy Data*. Lecture Notes in Computer Science. Springer Berlin/Heidelberg, 2001.

[2] Carroll, R. J., Küchenhoff, H., Lombard, F., Stefanski, L. A. Asymptotics for the SIMEX estimator in nonlinear measurement error models. *J. Am. Stat. Assoc.* Vol. 91, No.433, 242–250 (1996).

[3] Carroll, R. J., Maca, J. D., Ruppert, D. Nonparametric regression in the presence of measurement error. *Biometrika*, Vol. 86, No.3, 541–554 (1999).

[4] Carroll, R. J., Ruppert, D., Stefanski, L, A., Crainiceanu, C. M. *Measurement error in nonlinear models: a modern perspective*. 2nd ed. Monographs on Statistics and Applied Probability 105. Boca Raton, FL: Chapman & Hall/CRC, (2006).

[5] Cook, J.R., Stefanski, L.A. Simulation-extrapolation estimation in parametric measurement error models. *J. Am. Stat. Assoc.* Vol. 89, No.428, 1314–1328 (1994).

[6] Fazekas, I., Kukush, A. G. Asymptotic properties of an estimator in nonlinear functional errors-in-variables models with dependent error terms. *Comput. Math. Appl.* Vol. 34 (1997), no. 10, 23–39.

[7] Fazekas, I., Kukush, A., Zwanzig, S. Correction of nonlinear orthogonal regression estimator. *Ukraïn. Mat. Zh.* Vol. 56 (2004), no. 8, 1101–1118; translation in *Ukrainian Math. J.* Vol. 56 (2004), no. 8, 1308–1330.

[8] Fuller, W. A. *Measurement error models*. Wiley Series in Probability and Statistics; Wiley-Interscience. Hoboken, NJ: John Wiley & Sons, (2006).

[9] Gontar, O., Küchenhoff, H. The expansion of SIMEX estimator in nonlinear errors-in-variables model with small measurement errors. *Theory of Stochastic Processes* Vol. 14 (2008), no. 1, 39–48.

[10] Van Gorp, J., Schoukens, J., Pintelon, R. Learning neural networks with noisy inputs using the errors-in-variables approach. *IEEE Transactions on Neural Networks*, Vol. 11, 2 (2000), 402–414.

[11] Stefanski, L.A., Cook, J.R. Simulation-extrapolation: The measurement error jackknife. *J. Am. Stat. Assoc.* Vol. 90, No.432, 1247–1256 (1995).

[12] Van Huffel, S., Lemmerling, P. (editors). *Total least squares and errors-in-variables modeling.* Analysis, algorithms and applications. Including papers from the 3rd International Workshop on TLS and EIV Modeling held in Leuven, August 27–29, 2001. Kluwer Academic Publishers, Dordrecht, 2002.

István Fazekas
P.O. Box 12, 4010 Debrecen, Hungary,

Alexander Kukush
Volodymyrska st. 60, 01033 Kyiv, Ukraine