Theory of Stochastic Processes
Vol. 14 (30), no. 1, 2008, pp. 39–48

UDC 519.21

OLENA GONTAR AND HELMUT KÜCHENHOFF

THE EXPANSION OF A SIMEX ESTIMATOR IN THE NONLINEAR ERRORS-IN-VARIABLES MODEL WITH SMALL MEASUREMENT ERRORS

The nonlinear structural errors-in-variables model is investigated. We consider a Simex estimator with polynomial extrapolation function. The expansion of a Simex estimator is based on the asymptotic expansion of a naive estimator for small measurement errors. It is shown that the Simex estimator has an asymptotic deviation from a true value of the unknown parameter which is negligible compared with a measurement error variance, while the deviation of the naive estimator is proportional to the measurement error variance.

Introduction

The regression problems where predictors are measured with additive errors are considered. We denote the response variable by y and the d-dimensional predictor by ξ which cannot be observed. Instead, we observe $x = \xi + \sigma_\delta\gamma$, $\sigma_\delta > 0$, where γ is a standard normal vector in $\mathbb{R}^d$ independent of ξ. The term $\sigma_\delta\gamma$ is the measurement error. We will estimate the unknown parameter vector β_0 related to the distribution of (y, ξ). We consider the structural case, so that $\{y_i, x_i, \xi_i\}$, $i = \overline{1, n}$, are independent and identically distributed. If one could observe xi_i, then we suppose that one could consistently estimate β_0 by solving the estimating equation $\frac{1}{n}\sum_{i=1}^{n}\psi(y_i, \xi_i, \beta) = 0$. This estimator is usually called a naive estimator, when it is used in spite of measurement errors in regressors, i.e. the naive estimator is a solution to

$$\frac{1}{n}\sum_{i=1}^{n}\psi(y_i, x_i, \beta) = 0. \tag{1}$$

Denote the naive estimator by $\widehat{\beta}_{naive}(\sigma_\delta^2)$. Naive estimators are used to obtain a Simex estimator. Simex is the simulation-based method of estimating and reducing the bias due to a measurement error. The technique was proposed by Cook and Stefanski (1994). The Simex procedure contains two main steps.

Simulation. Let $\Lambda = \{0, \lambda_1, \dots, \lambda_M\}, \lambda_k > 0$, $k = \overline{1, M}$. For all $i = \overline{1, n}$, $b = \overline{1, B}$, where B is the number of additional samples, the independent standard normal variables $\varepsilon_{i,b}^*$ are generated. For each $\lambda \in \Lambda$, an additional measurement error is added to regressors: $x_{i,b}(\lambda) = x_i + \sqrt{\lambda}\sigma_\delta\varepsilon_{i,b}^*$, $\sigma_\delta > 0$. Using $x_{i,b}(\lambda)$ instead of x_i and taking the average over b, a set of averaged naive estimators $\widehat{\beta}_{naive}((1 + \lambda_k)\sigma_\delta^2), k = \overline{1, M}$ is constructed.

2000 *AMS Mathematics Subject Classification.* Primary 62J02, 62F10, 62F12.

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

We would like to thank Prof. A. Kukush for the helpful and fruitful discussions of the Simex expansion and the proofs of theorems, as well as for his valuable comments.

Extrapolation. Let a parametric model which describes the dependence of the naive estimator on the extra measurement error variance be $G(\lambda, \Gamma)$, and let $\widehat{\Gamma}$ be an estimator of Γ. The Simex estimator is defined as $\widehat{\beta}_{Simex} = G(-1, \widehat{\Gamma})$.

It was shown that the Simex estimator is consistent if the model $G(\lambda, \Gamma)$ corresponds to the true model asymptotically [1]. Typically one uses the model $G(\lambda, \Gamma)$ which is different from the true model, but the Simex estimator shows better numerical results for small and medium samples as compared with the consistent estimator Corrected Score. It has been proposed to use the quadratic functions $G_Q(\lambda, \Gamma) = \gamma_1 + \gamma_2\lambda + \gamma_3\lambda^2$ for extrapolation. Carroll et al. (1996) proved asymptotical the normality of Simex estimators using the assumption that the exact dependence of a naive estimator from the additional measurement error variance is known:

$$E\widehat{\beta}_{naive}((1+\lambda)\sigma_\delta^2) = G(\lambda, \Gamma)$$

for a certain parameter value Γ. We consider the Simex estimator with polynomial extrapolation function of any fixed degree. Using the asymptotic expansion of the naive estimator, we will show that the Simex estimator has an asymptotic deviation from the true value of the unknown parameter which is negligible as compared with a measurement error variance, while the deviation of the naive estimator is proportional to the measurement error variance.

We denote the Euclidean norm of a vector ξ by $\|\xi\|$, the neighborhood of x of radius r by $B(x, r)$, the closure of $B(x, r)$ by $\overline{B(x, r)}$, the identity matrix of order n by by I_n, and a transposed vector x by x^t.

The paper is organized as follows. We start with assumptions about the model and the expansion of a naive estimator. Then the main result of the paper - the expansion of the Simex estimator - is proved and applied to the exponential family and mean-variance models, and we conclude by a discussion.

Assumptions

We assume that the regressors $\xi_i \in \mathbb{R}^d$ are independent identically distributed random vectors and $\forall\lambda \in \mathbb{R}$, $Ee^{\lambda\|\xi\|} < \infty$. Suppose that the regressors are measured with error, and $x_i = \xi_i + \sigma_\delta\gamma_i$, $\sigma_\delta > 0$, rather than ξ_i, are observed, where $\gamma_i \simeq N(0, I_d)$, and σ_δ^2 is known. The predictors y_i are scalar variables. The variables ξ_i, γ_i are mutually independent. We assume that $\beta^0 \in \text{int}K$ is a true value of the parameter β, where K is a convex compact set in $\mathbb{R}^p$. The function $\psi : \mathbb{R} \times \mathbb{R}^d \times K \to \mathbb{R}^p$ is a vector function which is smooth enough. The naive estimator $\widehat{\beta}_{naive}$ is defined as a solution of the estimating equation

$$S_n(\beta) := \frac{1}{n}\sum_{i=1}^{n} \psi(y_i, x_i, \beta) = 0. \tag{2}$$

We assume that $\|\psi(y, \xi, \beta)\| \leq k_1 e^{k_2\|\xi\|}$ and the derivative $\|\psi^\beta(y, \xi, \beta)\| \leq k_3 e^{k_4\|\xi\|}$, where $k_i \in \mathbb{R}, i = \overline{1, 4}$ are constants. The next convergence takes place:

$$P\{S_n(\beta) \to E\psi(y, x, \beta), \text{ uniformly in } \beta \in K, \text{ as } n \to \infty\} = 1. \tag{3}$$

An analogous convergence was demonstrated by Schneeweiss and Kukush (2006) in the proof of Theorem 4.1.

Denote $F(\beta, \sigma_\delta) = E\psi(y, x, \beta),\ \beta \in K$. We obtain

$$F(\beta, \sigma_\delta) = E\psi(y, \xi + \sigma_\delta\gamma, \beta),\ \beta \in K. \tag{4}$$

Here, we allow σ_δ to be negative or equal to 0 since the distributions of $\sigma_\delta\gamma$ and $-\sigma_\delta\gamma$ are identical.

Expansion of a naive estimator

Hereafter, we use the assumptions of the general model stated above. The next theorem makes it possible to expand a naive estimator for small measurement errors.

Theorem 1. *Assume that the following conditions hold:*

1 . The function $\psi(y,\xi,\beta) \in C^1(\mathbb{R}\times\mathbb{R}^d\times U \to \mathbb{R}^p)$, $U \supset K$, where U is open.

2 . For the function $F(\beta,0) = E\psi(y,\xi,\beta)$, there exists the unique solution β_0 of the equation $F(\beta,0)=0$ on the convex compact set K.

3 . The matrix $V = E\psi^{\beta}(y,\xi,\beta_0)$ is nonsingular.

Then there exists $\sigma > 0$ such that, for all $\sigma_\delta \in B(0,\sigma)$, the equation $F(\beta,\sigma_\delta) = 0$ has the unique solution $\beta_{naive}(\sigma_\delta)$ in K. Moreover, the function $\beta_{naive}(\sigma_\delta)$ is an even function of $\sigma_\delta \in B(0,\sigma)$.

Proof. 1. The first and third conditions of the theorem allow us to use the implicit function theorem (see Appendix A). According to this theorem, there exist $\delta_1 > 0$ and $\rho > 0$ such that the equation $F(\beta,\sigma_\delta)=0$ has the unique solution

$$\beta_{naive}(\sigma_\delta)\colon B(0,\delta_1) \to B(\beta_0,\rho).$$

The neighborhood $B(\beta_0,\rho) \subset K$. The implicit function theorem states that $\beta_{naive}(\sigma_\delta) \in C^1(B(0,\delta_1) \to K)$. Note that the function $F(\beta,\sigma_\delta)$ is an even function of the second variable. Indeed,

$$F(\beta,-\sigma_\delta) = E\psi(y,\xi-\sigma_\delta\gamma,\beta) = E\psi(y,\xi+\sigma_\delta\gamma,\beta) = F(\beta,\sigma_\delta).$$

Consider the equation $F(\beta,-\sigma_\delta)=0$ which is equivalent to $F(\beta,\sigma_\delta)=0$. The solution to the equation $F(\beta,\sigma_\delta)=0$ is unique for $\sigma_\delta \in B(0,\delta_1)$ on K, so this implies that $\beta_{naive}(\sigma_\delta) = \beta_{naive}(-\sigma_\delta)$, and the function $\beta_{naive}(\sigma_\delta)$ is an even function for $\sigma_\delta \in B(0,\delta_1)$.

2. Consider the set $K_1 := K \setminus B(\beta_0,\rho)$. The set K_1 is a compact set as well. We will prove that the function β_{naive} is the unique solution over the whole compact set K. Consider the function $F(\beta,0)$. According to the second condition, it has the unique solution β_0 on the compact set K. Due to the continuity of $F(\beta,0)$, this means that there exists a constant $c>0$ such that $\|F(\beta,0)\| > c$, for all $\beta \in K_1$.

3. Admit that the function $F(\beta,\sigma_\delta)$ is continuous on the compact set $K_1 \times \overline{B(0,\delta_1)}$. Then the function $F(\beta,\sigma_\delta) \rightrightarrows F(\beta,0)$ uniformly in β over K_1 as $\sigma_\delta \to 0$. This means that $\forall\varepsilon>0\ \exists\delta_2>0$ such that $\forall\sigma_\delta \in B(0,\delta_2)$ and $\forall\beta\in K_1$, and the following inequality holds: $\|F(\beta,\sigma_\delta) - F(\beta,0)\| \le \varepsilon$. This implies that

$$|\ \|F(\beta,\sigma_\delta)\| - \|F(\beta,0)\|\ | \le \|F(\beta,\sigma_\delta) - F(\beta,0)\| \le \varepsilon.$$

We state that, for all $\beta \in K_1$ and for all $\sigma_\delta \in B(0,\delta_2)$, $\|F(\beta,\sigma_\delta)\| \ge \|F(\beta,0)\| - \varepsilon \ge c-\varepsilon$ holds. As ε can be chosen arbitrary, we set $\varepsilon = \dfrac{c}{2}$. Then $\|F(\beta,\sigma_\delta)\| \ge \dfrac{c}{2} > 0, \sigma_\delta \in B(0,\delta_2)$.

4. We now set $\delta = \min(\delta_1,\delta_2)$. Then, for all $\sigma_\delta \in B(0,\delta)$, there exists the unique solution $\beta_{naive}(\sigma_\delta)$ to the equation $F(\beta,\sigma_\delta)=0$ on the compact set K and $\|\beta_{naive}(\sigma_\delta) - \beta_0\| \le \rho$. Theorem 1 is proved.

Theorem 2. *Assume that conditions 2 and 3 of Theorem 1 hold and, for fixed $l \ge 1$, the function $\psi(y,\xi,\beta)$ satisfies the following conditions:*

1 . The score function $\psi(y,\xi,\beta) \in C^{2l+2}(\mathbb{R}\times\mathbb{R}^d\times U \to \mathbb{R}^p)$ with respect to ξ and β, and $U \supset K$, where U is open.

2 . For any partial derivative $D_q\psi(y,\xi,\beta)$ of order $q \le 2l+2$ with respect to components of ξ and components of β, $\|D_q\psi(y,\xi,\beta)\| \le c_1 e^{c_2\|\xi\|}$, where c_1, c_2 are constants.

Then there exists $\sigma > 0$ *such that, for all* $\sigma_\delta \in B(0, \sigma)$,

$$\widehat{\beta}_{naive}(\sigma_\delta) = \beta_{naive}(\sigma_\delta) + o(1) \text{ a.s., as } \ n \to \infty, \tag{5}$$

where

$$\beta_{naive}(\sigma_\delta) = \beta_0 + \sum_{j=1}^{l} \frac{\beta_{naive}^{(2j)}(0)}{(2j)!} \sigma_\delta^{2j} + O(\sigma_\delta^{2l+2}), \text{ as } \ \sigma_\delta \to 0. \tag{6}$$

Remark 1. Below, we will write relations like (5) and (6) as

$$\widehat{\beta}_{naive}(\sigma_\delta) = \beta_0 + \sum_{j=1}^{l} \frac{\beta_{naive}^{(2j)}(0)}{(2j)!} \sigma_\delta^{2j} + O(\sigma_\delta^{2l+2})_{\sigma_\delta \to 0} + o(1)_{n \to \infty}. \tag{7}$$

Remark 2. Expansion (7) resembles the expansion of an orthogonal regression estimator for the functional model from Fazekas et al. (2002). But the expansion (7) is much simpler, since the naive estimator converges a.s., as $n \to \infty$, see (5), while the orthogonal regression estimator for the functional model need not converge.

Proof. First, we note that Theorem 1 holds. The idea is to apply a Taylor expansion to the function $\beta_{naive}(\sigma_\delta)$. From the first condition of Theorem 2, it follows that $F(\beta, \sigma_\delta) \in C^{2l+2}(\mathbb{R}^d \times U \to \mathbb{R}^p)$, and this implies that $\beta_{naive}(\sigma_\delta) \in C^{2l+2}(\mathbb{R} \to \mathbb{R}^p)$. The second condition of Theorem 2 and the condition $Ee^{\lambda\|\xi\|} < \infty$ yield the boundedness of $\beta_{naive}^{(2l+2)}(\sigma_\delta)$ for all $\sigma_\delta \in B(0, \delta)$, where δ is defined from Theorem 1. Now we can use the Taylor expansion for $\beta_{naive}(\sigma_\delta)$. As $\beta_{naive}(\sigma_\delta)$ is an even function, the Taylor expansion of this function will have only summands of even powers. Thus,

$$\beta_{naive}(\sigma_\delta) = \beta_0 + \sum_{j=1}^{l} \frac{\beta_{naive}^{(2j)}(0)}{(2j)!} \sigma_\delta^{2j} + O(\sigma_\delta^{2l+2}), \text{as } \ \sigma_\delta \to 0.$$

Finally, it follows from (3) and Lemma 1 from Appendix A that $\widehat{\beta}_{naive}(\sigma_\delta) = \beta_{naive}(\sigma_\delta) + o(1)$ a.s., as $n \to \infty$, and this proves the theorem. Theorem 2 is proved.

Remark 3. If $l = 2$, then $\beta_{naive}(\sigma_\delta) = \beta_0 - V^{-1}K\sigma_\delta^2 + O(\sigma_\delta^4)$, as $\sigma_\delta \to 0$, where V is defined in Theorem 1 and $K = E\psi^{\xi}(y, \xi, \beta_0)$.

Simex with polynomial extrapolant function

We introduce the polynomial extrapolant function for a Simex estimator. We show that the Simex with polynomial extrapolant has an asymptotic deviation from the true value, which is negligible as compared with a measurement error variance.

Supplementary sample generation. Let $\Lambda = \{0, \lambda_1, \ldots, \lambda_M\}$, $\lambda_k > 0, k = \overline{1, M}$. Let B be a large fixed natural number. For all $i = \overline{1, n}$ and for all $b = \overline{1, B}$, standard normal variables $\varepsilon_{i,b}^* \simeq N(0, I_d)$ are generated. For each $\lambda \in \Lambda$, an additional variance is added to regressors $x_{i,b}(\lambda) = x_i + \sqrt{\lambda}\sigma_\delta \varepsilon_{i,b}^*$. Here, σ_δ is a true standard deviation of the measurement error. Now $x_{i,b}(\lambda)$ are used as new regressors.

Estimation. For each $\lambda \in \Lambda$ averaged over b, naive estimators $\widehat{\beta}_{naive}(\lambda)$ are calculated.

Parametric model for naive estimators. Let the j-th coordinate of $\widehat{\beta}_{naive}(\lambda)$ depends on λ by a polynomial law $g_j(\lambda) := \gamma_{j0} + \gamma_{j1}\lambda + \cdots + \gamma_{jm}\lambda^m, j = \overline{1, p}$, and $m \geq 1$ is fixed. Denote the extrapolant function $G(\lambda, \Gamma) := (g_1(\lambda), \ldots, g_p(\lambda))^t$. We estimate the unknown parameter Γ by the method of least squares:

$$\widehat{\Gamma} = arg \min_{\Gamma \in \mathbb{R}^{(m+1) \times p}} \frac{1}{M+1} \sum_{k=0}^{M} \|\widehat{\beta}_{naive}(\lambda_k) - G(\lambda_k, \Gamma)\|^2.$$

Extrapolation. A Simex estimator is defined as $\widehat{\beta}_{Simex} = G(-1, \widehat{\Gamma})$.

Theorem 3. *Let Theorem 2 hold and $l \le m \le M$. Then the following expansion of Simex is true:*

$$\widehat{\beta}_{Simex} = \beta_{Simex}(\sigma_\delta) + o(1) \ \textit{a.s.}, \ \textit{as } n \to \infty,$$

where $\beta_{Simex}(\sigma_\delta) = \beta_0 + O(\sigma_\delta^{2l+2})$ as $\sigma_\delta \to 0$.

Proof. 1. It is not difficult to find the explicit form of $\widehat{\Gamma}$. We have $G(\lambda, \Gamma) = s(\lambda)\Gamma$, where the $p \times (m+1)p$ matrix $s(\lambda)$ is equal to

$$s(\lambda) = \begin{pmatrix} 1 & \lambda & \dots & \lambda^m & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & \lambda & \dots & \lambda^m & 0 & 0 & \dots & 0 \\ & & \vdots & & & & \ddots & & & & \vdots & \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & \lambda & \dots & \lambda^m \end{pmatrix},$$

and $\Gamma = (\gamma_{10}, \gamma_{11}, \dots, \gamma_{1m}, \gamma_{20}, \gamma_{21}, \dots, \gamma_{2m}, \dots, \gamma_{p0}, \gamma_{p1}, \dots, \gamma_{pm})^t$. The least squares estimator of Γ equals

$$\widehat{\Gamma} = \left(\frac{1}{M+1} \sum_{k=0}^{M} s^t(\lambda_k) s(\lambda_k) \right)^{-1} \frac{1}{M+1} \sum_{k=0}^{M} s^t(\lambda_k) \widehat{\beta}_{naive}(\lambda_k).$$

2. Note that the matrix inverse to the $(m+1)p \times (m+1)p$ matrix $\frac{1}{M+1} \sum_{k=0}^{M} s^t(\lambda_k) s(\lambda_k)$ exists for all $m \le M$. To prove this, consider a discrete random variable ζ which takes the values $\{0, \lambda_1, \dots, \lambda_M\}$ with equal probabilities $1/(M+1)$. For the vector $\psi = (1, \zeta, \dots, \zeta^m)^t$, let us consider the matrix $E\psi\psi^t$. The matrix $\frac{1}{M+1} \sum_{k=0}^{M} s^t(\lambda_k) s(\lambda_k)$ is block-diagonal with $(m+1) \times (m+1)$ block $E\psi\psi^t$ on diagonal p times. So the existence of the matrix inverse to $E\psi\psi^t$ is sufficient for the existence of the matrix inverse to $\frac{1}{M+1} \sum_{k=0}^{M} s^t(\lambda_k) s(\lambda_k)$. The Gram matrix $E\psi\psi^t$ is nonsingular. Indeed, let us suppose that there exists $a \ne 0$, $a \in \mathbb{R}^{m+1}$, such that $a^t E\psi\psi^t a = E(a^t\psi)^2 = 0$. This means that $a^t\psi = a_m\zeta^m + a_{m-1}\zeta^{m-1} + \dots + a_0 = 0$ a.s. Therefore, for each $k = 0, 1, \dots, M, a_m\lambda_k^m + a_{m-1}\lambda_k^{m-1} + \dots + a_0 = 0$. But this is impossible since, by the conditions of Theorem 3, $M + 1 > m$. This contradiction proves that $E\psi\psi^t$ is nonsingular.

3. Consider the $(m+1)p \times p$ matrix $J_r, r = \overline{1, m+1}$: the elements

$$(J_r)_{(j-1)(m+1)p+r,j} = 1, \qquad j = 1, \dots, p,$$

and other elements are zero. The transpose matrix equals

$$J_r^t := \begin{pmatrix} 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ & & \vdots & & & & \ddots & & & & \vdots & \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \end{pmatrix}.$$

Denote $s(-1) = s(\lambda)|_{\lambda=-1}$. The following equalities hold:

(8)
$$\frac{1}{M+1} \sum_{k=0}^{M} s^t(\lambda_k) s(\lambda_k) J_{r+1} = \frac{1}{M+1} \sum_{k=0}^{M} s^t(\lambda_k) \lambda_k^r, \ s(-1) J_{r+1} = (-1)^r I_p, \ r = \overline{0, m}.$$

4. Consider $\widehat{\beta}_{naive}((1+\lambda)\sigma_\delta^2) = \sum_{j=0}^{l} \frac{\beta_{naive}^{(2j)}(0)}{(2j)!}\sigma_\delta^{2j}\sum_{r=0}^{j} C_j^r \lambda^r + O(\sigma_\delta^{2j+2})_{\sigma_\delta^2\to 0} + o(1)_{n\to\infty}$. Remember that $l \le m$, and, therefore, we can use (8) for $r \le l$.

We have

$$
\begin{aligned}
\widehat{\beta}_{Simex} &= s(-1)\widehat{\Gamma} = s(-1)\left(\frac{1}{M+1}\sum_{k=1}^{M} s^t(\lambda_k)s(\lambda_k)\right)^{-1}\frac{1}{M+1}\sum_{k=0}^{M} s^t(\lambda_k)\widehat{\beta}_{naive}(\lambda_k)\\
&= s(-1)\left(\frac{1}{M+1}\sum_{k=0}^{M} s^t(\lambda_k)s(\lambda_k)\right)^{-1}\frac{1}{M+1}\sum_{k=0}^{M} s^t(\lambda_k)\sum_{j=0}^{l}\frac{\beta_{naive}^{(2j)}(0)}{(2j)!}\sigma_\delta^{2j}\sum_{r=0}^{j} C_j^r\lambda_k^r\\
&\quad + O(\sigma_\delta^{2l+2})_{\sigma_\delta^2\to 0} + o(1)_{n\to\infty}\\
&= \sum_{j=0}^{l}\frac{\beta_{naive}^{(2j)}(0)}{(2j)!}\sigma_\delta^{2j}\sum_{r=0}^{j} C_j^r s(-1)\left(\frac{1}{M+1}\sum_{k=0}^{M} s^t(\lambda_k)s(\lambda_k)\right)^{-1}\frac{1}{M+1}\sum_{k=0}^{M} s^t(\lambda_k)\lambda_k^r\\
&\quad + O(\sigma_\delta^{2l+2})_{\sigma_\delta^2\to 0} + o(1)_{n\to\infty}\\
&= \sum_{j=0}^{l}\frac{\beta_{naive}^{(2j)}(0)}{(2j)!}\sigma_\delta^{2j}\sum_{r=0}^{j} C_j^r(-1)^r + O(\sigma_\delta^{2l+2})_{\sigma_\delta^2\to 0} + o(1)_{n\to\infty}\\
&= \sum_{j=0}^{l}\frac{\beta_{naive}^{(2j)}(0)}{(2j)!}\sigma_\delta^{2j}(1-1)^j + O(\sigma_\delta^{2l+2})_{\sigma_\delta^2\to 0} + o(1)_{n\to\infty}\\
&= \beta_0 + O(\sigma_\delta^{2l+2})_{\sigma_\delta^2\to 0} + o(1)_{n\to\infty},
\end{aligned}
$$

and this proves the theorem.

Exponential Family

The regression model is described by a conditional distribution of y given ξ and given an unknown parameter vector θ. We assume this distribution to be represented by a probability density function

$$
f(y|\xi,\beta,\varphi) = \exp\left(\frac{y\eta - c(\eta)}{\varphi} + a(y,\varphi)\right) \text{ with } \eta = \eta(\xi,\beta), \tag{9}
$$

where β is a regression parameter vector, and φ is a scalar dispersion parameter such that $\theta = (\beta^t,\varphi)^t$, and a, c, and η are known functions. The function $c(\cdot)$ is smooth enough, and $c''(\cdot) > 0$. We assume that $\beta^0 \in \mathrm{int}K$ is a true value of the parameter β, where K is a convex compact set in $\mathbb{R}^p$, and $\varphi_0 \in [a_1, b_1]$, $a_1 > 0, b_1 < \infty$, φ_0 is a true value of the parameter φ. If the variable ξ would be observable, one could estimate β and φ by maximum likelihood. Consider the corresponding likelihood-score function for β in two cases:

1. The dispersion parameter φ is known:

$$
\psi_1(y,\xi,\beta) = (y - c'(\eta))\eta^\beta. \tag{10}
$$

2. The dispersion parameter φ is unknown:

$$
\psi_2(y,\xi,\beta,\varphi) = \begin{pmatrix} (y - c'(\eta))\eta^\beta \\ (y - c'(\eta))^2 - c''(\eta)\varphi \end{pmatrix}. \tag{11}
$$

The score function in both cases is unbiased. This implies, under natural regularity conditions, that the estimators of the unknown parameters, obtained as solutions to the equation $\frac{1}{n}\sum_{i=1}^{n}\psi_1(y_i,\xi_i,\beta)=0$ or $\frac{1}{n}\sum_{i=1}^{n}\psi_2(y_i,\xi_i,\beta,\varphi)=0$, respectively, are consistent. But we observe not ξ, but x that differs from the latent variable ξ by a measurement error $\sigma_\delta\gamma$ which is independent of ξ and y. We assume that $\gamma \simeq N(0,I_d)$, and σ_δ^2 is known. The naive estimator is obtained as a solution to the equation $\frac{1}{n}\sum_{i=1}^{n}\psi_1(y_i,x_i,\beta)=0$ in the case of the known dispersion parameter and $\frac{1}{n}\sum_{i=1}^{n}\psi_2(y_i,x_i,\beta,\varphi)=0$ in the case of the unknown dispersion parameter. The next corollary is true in both cases.

Corollary 1. *Consider model* (9). *The naive estimator is calculated with the help of* (10) *if the dispersion parameter φ is known and with the help of* (11) *if the dispersion parameter is unknown. Let the following conditions hold for fixed $l \ge 1$:*

1. *The function $c(\eta)\in C^{2l+3}(\mathbb{R})$ and, for all $q\le 2l+3$, the derivative $|c^{(q)}(\eta)|\le c_1e^{c_2\|\xi\|}$ for some constants c_1,c_2.*
2. *The function $\eta(\xi,\beta)\in C^{2l+3}(\mathbb{R})$ with respect to β, and, for all $q\le 2l+3$, any partial derivative of order q with respect to components of β $|\eta^{(q)}(\xi,\beta)|\le c_1e^{c_2\|\xi\|}$ for some constants c_1,c_2.*
3. *The identifiability condition for the error-free model: the equation $E(c'(\eta_0)-c'(\eta))\eta^\beta = 0$, $\beta\in K$, where $\eta_0=\eta(\xi,\beta_0)$, has the unique solution $\beta=\beta_0$.*
4. *The matrix $E\eta_0^\beta(\eta_0^\beta)^t$ is nonsingular, and, for each ξ,β and for some constants a_1 and a_2, the inequality $c''(\eta)\ge a_1e^{-a_2\|\xi\|}$ holds.*

Then, for $l\le m\le M$, the expansion of Simex with polynomial extrapolant function is true:

$$\widehat{\beta}_{Simex}=\beta_{Simex}(\sigma_\delta)+o(1)\ \text{a.s.},\ \text{as } n\to\infty,$$

where $\beta_{Simex}(\sigma_\delta)=\beta_0+O(\sigma_\delta^{2l+2})$, as $\sigma_\delta\to 0$. If φ is estimated, then

$$\widehat{\varphi}_{Simex}=\varphi_{Simex}(\sigma_\delta)+o(1)\ \text{a.s.},\ \text{as } n\to\infty,$$

where $\varphi_{Simex}(\sigma_\delta)=\varphi_0+O(\sigma_\delta^{2l+2})$, as $\sigma_\delta\to 0$.

Proof. In the case of the known dispersion parameter, the corollary follows directly from Theorem 3. So we consider only the case of the unknown dispersion parameter. Introduce a new parameter $\theta=(\beta,\varphi)^t\in K\times[a_1,b_1]$, and its true value $\theta_0=(\beta_0,\varphi_0)^t$. The naive estimator for this parameter is obtained via the estimating function:

$$\psi_2(y,x,\theta)=\begin{pmatrix}(y-c'(\eta))\eta^\beta\\ (y-c'(\eta))^2-c''(\eta)\varphi\end{pmatrix}.$$

The first two conditions of the corollary correspond to the conditions of Theorem 2. We need to check the second and third conditions of Theorem 1. The second condition of Theorem 1 states that the equation $E\psi_2(y,\xi,\theta)=0$ has the unique solution $\theta=\theta_0$ on $K\times[a_1,b_1]$. The equation

$$E\psi_2(y,\xi,\theta)=\begin{pmatrix}E(c'(\eta_0)-c'(\eta))\eta^\beta\\ E(c''(\eta_0)\varphi_0-c''(\eta)\varphi)+E(c'(\eta_0)-c'(\eta))^2\end{pmatrix}=0$$

has the unique solution $\theta_0=(\beta_0,\varphi_0)^t$, if the third condition of Corollary 1 holds. The last condition to be checked is the third condition of Theorem 1. It requires $E\psi_2^\theta(y,\xi,\theta_0)$ to be nonsingular. Consider

$$E\psi_2^\theta(y,\xi,\theta_0)=-\begin{pmatrix}Ec''(\eta_0)\eta_0^\beta(\eta_0^\beta)^t & 0\\ Ec'''(\eta_0)\eta_0^\beta\varphi_0 & Ec''(\eta_0)\end{pmatrix}.$$

It is nonsingular, if the fourth condition of Corollary 1 holds. Then the next expansion of the Simex estimator of θ is true: $\widehat{\theta}_{Simex} = \theta_{Simex} + o(1)$ a.s., as $n \to \infty$, where $\theta_{Simex} = \theta_0 + O(\sigma_\delta^{2l+2})$, as $\sigma_\delta \to 0$. And this is the statement of Corollary 1.

Application for the Gaussian model. Consider a nonlinear errors-in-variables model:

$$\begin{cases} y_i = g(\xi_i, \beta^0) + \varepsilon_i; \\ x_i = \xi_i + \delta_i; \end{cases} i = \overline{1,n}.$$

The regressors $\xi_i \in \mathbb{R}^d$ are independent identically distributed random vectors, and $\forall \lambda \in \mathbb{R}$, $Ee^{\lambda\|\xi\|} < \infty$. The errors in regressor $\delta_i \in \mathbb{R}^d$ are normal identically distributed random vectors with zero expectation and the variance of $\sigma_\delta^2 I_d$. The errors in predictors ε_i are normal independent identically distributed with zero expectation. The random variables $\xi_i, \delta_i, \varepsilon_i$ are mutually independent.

Here, $\eta(\xi, \beta) = g(\xi, \beta),\ c(\eta) = \frac{1}{2}\eta^2,\ \varphi = \sigma_\varepsilon^2$. The corollary holds if the function $g(\xi, \beta)$ satisfies conditions 1 to 4 from Corollary 1, where $g(\xi, \beta)$ is used instead of $\eta(\xi, \beta)$.

Application for the Loglinear Poisson model. For the variable ξ, define $\lambda = \exp(\xi^t\beta)$. Then the loglinear Poisson model is defined as $y \sim P_0(\lambda)$, where $P_0(\lambda)$ stands for the Poisson distribution with intensity λ. Here, $\eta = \log\lambda, c(\eta) = e^\eta$, and $\varphi = 1$. The first two conditions of Corollary 1 hold for any $l \geq 1$. The third condition is equivalent to the existence of a unique solution to the equation $E\xi\xi^t(\beta_0 - \beta) = 0$, and this holds, while $E\xi\xi^t$ is nonsingular as a covariance matrix. The fourth condition requires $E\exp(\xi^t\beta_0)\xi\xi^t$ to be nonsingular, which is also true. Thus, Corollary 1 holds.

Corollary 1 holds also for the Gamma and Logit models. All the considered models are with errors in the variables (concerning these models, see [2]).

Mean-Variance model

Suppose that a relation between the response variable y and the regressor ξ is given by the conditional mean and the conditional variance:

$$E(y|\xi) = m(\xi, \beta),\ \operatorname{var}(y|\xi) = v(\xi, \beta, \varphi), \tag{12}$$

where φ is the scalar dispersion parameter. It is supposed that $v(\xi, \beta, \varphi) > 0$ for all $\xi,\ \beta,\ \varphi$. We assume that the true value of the parameter $\beta^0 \in \operatorname{int}K$, where K is a convex compact set in $\mathbb{R}^p$ and $\varphi_0 \in [a_1, b_1],\ a_1 > 0, b_1 < \infty$, φ_0 is the true value of the parameter φ. We assume that the regressors $\xi_i \in \mathbb{R}^d$ are independent identically distributed random vectors and $\forall \lambda \in \mathbb{R}$, $Ee^{\lambda\|\xi\|} < \infty$. The specification of only the mean and the variance in model (12) allows one to construct the consistent estimator of β and φ. The conditionally unbiased estimating function is

$$\psi(y, \xi, \beta, \varphi) = \begin{pmatrix} (y - m(\xi, \beta))(v(\xi, \beta, \varphi))^{-1} m^\beta(\xi, \beta) \\ (y - m(\xi, \beta))^2 - v(\xi, \beta, \varphi). \end{pmatrix} \tag{13}$$

If the dispersion parameter is known, we omit the second line in (13). In the case of measurement errors, when we do not observe the latent ξ, but observe x, which equals $x = \xi + \sigma_\delta\gamma$, the naive estimator is obtained from (1). We assume that $\gamma \simeq N(0, I_d)$, and σ_δ^2 is known.

Let us introduce a new parameter $\theta = (\beta, \varphi)^t \in K \times [a_1, b_1]$, and let its true value $\theta_0 = (\beta_0, \varphi_0)^t$.

Corollary 2. *Assume that, for model* (12)*, the next conditions hold for fixed* $l \geq 1$*:*
1. *The function* $m(\xi, \beta) \in C^{2l+3}(\mathbb{R}^d \times \mathbb{R}^p)$*, and, for all* $q \leq 2l+3$*, any partial derivative of order* q *satisfies* $\|m^{(q)}(\xi, \beta)\| \leq c_1 e^{c_2\|\xi\|}$ *for some constants* c_1, c_2*.*

2. *The function* $v(\xi,\theta) \in C^{2l+3}(\mathbb{R}^d \times \mathbb{R}^{p+1})$ *with respect to components of* ξ *and* θ*, and, for all* $q \le 2l+3$*, any partial derivative of order* q *satisfies* $\|v^{(q)}(\xi,\theta)\| \le c_1 e^{c_2\|\xi\|}$ *for some constants* c_1, c_2*, and* $v(\xi,\theta) \ge c_3 e^{c_4\|\xi\|}$ *for some constants* c_3, c_4.
3. *The identifiability condition for the error-free model is as follows: the system of equations*

$$\begin{cases} E(m(\xi,\beta_0) - m(\xi,\beta))(v(\xi,\theta))^{-1} m^\beta(\xi,\beta) = 0 \\ E(v(\xi,\theta_0) - v(\xi,\theta)) + E(m(\xi,\beta_0) - m(\xi,\beta))^2 = 0 \end{cases}, \ \theta \in K \times [a_1, b_1]$$

has the unique solution $\theta = \theta_0$.
4. *The matrices* $Em^\beta(\xi,\beta_0)(m^\beta(\xi,\beta_0))^t$ *and* $Ev^\varphi(\xi,\theta_0)$ *are nonsingular.*

Then, for $l \le m \le M$*, the expansion of the Simex estimator with polynomial extrapolant function is true:*

$$\widehat{\theta}_{Simex} = \theta_{Simex}(\sigma_\delta) + o(1), \ \text{a.s., as} \ \ n \to \infty,$$

where $\theta_{Simex}(\sigma_\delta) = \theta_0 + O(\sigma_\delta^{2l+2})$*, as* $\sigma_\delta \to 0$.

Proof. We need only to check the third condition of Theorem 1. Consider

$$E\psi^\theta(y,\xi,\theta_0) = -\begin{pmatrix} Em^\beta(\xi,\beta_0)(m^\beta(\xi,\beta_0))^t(v(\xi,\theta_0))^{-1} & 0 \\ v^\beta(\xi,\theta_0) & v^\varphi(\xi,\theta_0) \end{pmatrix}.$$

This matrix is nonsingular due to the fourth condition of Corollary 2.

Discussion

We have given some theoretical reasons for good performance of the Simex estimator for the polynomial extrapolant function. Our key idea was to use the Taylor expansion of the estimator in $\sigma_\delta = 0$. Therefore, one limitation of our result is that it can be applied only to the case of small measurement errors. To illustrate this point, let us consider the simplest linear model $y = \beta_1 + \beta_2\xi$ with $\beta_1 = 0$ and $\beta_2 = 1$. We assume the variance of ξ is known and equal to 1. In this case, the function $\beta_{naive}(\sigma_\delta)$ is obtained explicitly: $\beta_{naive}(\sigma_\delta) = \dfrac{1}{1+\sigma_\delta^2}$. We take the Taylor expansion of the function $\beta_{naive}(\sigma_\delta)$ as an extrapolation model for Simex, and the value of the y-intercept gives the Simex estimator. In Fig. 1, the Taylor expansion of $\beta_{naive}(\sigma_\delta)$ up to the forth and sixth powers of σ_δ was used. It can be seen that it works well for small values of σ_δ and does not if measurement errors are not small. But the function $\beta_{naive}(\sigma_\delta)$ still can be well approximated by a polynomial. Figure 2 shows such an approximation of $\beta_{naive}(\sigma_\delta)$ by the polynomial $1.20769 - 0.923077\sigma_\delta + 0.215385\sigma_\delta^2$, and Simex still works.

So the above theory is good for small measurement errors, and what is the behavior of the Simex estimator in the case of large measurement errors should be investigated.

Appendix A

The Implicit Function Theorem. *Let* $x \in \mathbb{R}^n, y \in \mathbb{R}^m$*,* A *be an open set in* $\mathbb{R}^{m+n}$*, and* $(x^0, y^0) \in A$*. Consider the function* $F : A \to \mathbb{R}^n$ *with the properties:*

1. $F(x^0, y^0) = 0$.
2. $F \in C^1(A \to \mathbb{R}^n)$.
3. $F^y(x^0, y^0)$ *is non-singular.*

Then $\exists\sigma > 0, \exists\rho > 0$*, and there exists the unique function* $f : B(x^0,\sigma) \to B(y^0,\rho)$ *with the following properties:*

1.$B(x^0,\sigma) \times B(y^0,\rho) \subset A$. *2.*$f(x^0) = y^0$.

3.$f \in C^1(B(x^0,\sigma) \to \mathbb{R}^n)$ *and* $f' = -(F^y(x, f(x)))^{-1}F^x(x, f(x)), \forall x \ in B(x^0,\sigma)$.

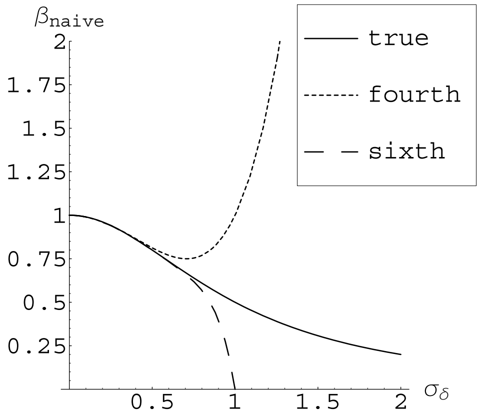

FIGURE 1

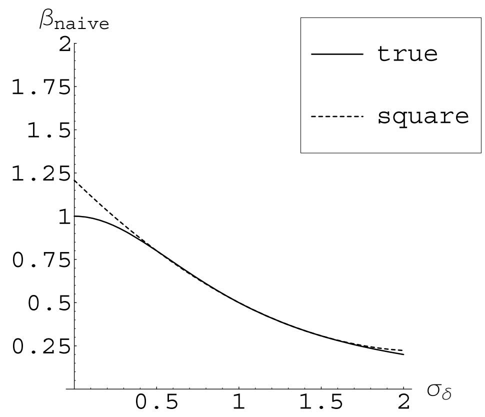

FIGURE 2

Lemma 1. *Let K be a compact set in $\mathbb{R}^n$ and $F_n : K \to \mathbb{R}^n$ be nonrandom continuous functions, $F : K \to \mathbb{R}^n$. Uniformly for θ in K, $F_n(\theta) \to F(\theta)$ as $n \to \infty$. Suppose that $F_n(\theta_n) = 0$, and $\theta^\star$ is the unique solution to $F(\theta) = 0$ on K. Then $\theta_n \to \theta^\star$, as $n \to \infty$.*

BIBLIOGRAPHY

1. R.J.Carroll, H.Kuechenhoff, F.Lombard, and L.A.Stefanski, *Asymptotics for the SIMEX Estimator in Nonlinear Measurement Error Models*, Journal of the American Statistical Association **91** (1996), no. 4, 242-250.
2. R.J.Carroll, D.Ruppert, L.A.Stefanski, and C.M.Crainiceanu,, *Measurement Error in Nonlinear Models*, New York: Chapman and Hall, 2006.
3. J.R.Cook and L.A.Stefanski, *Simulation-extrapolation estimation in parametric measurement error models*, Journal of the American Statistical Association **89** (1994), 1314-1328.
4. I.Fazekas, A.Kukush, and S.Zwanzig, *Correction of nonlinear orthogonal regression estimator*, Ukrainian Mathematical Journal **56** (2004,), no. 8, 1308-1329.
5. A.Kukush and H.Schneeweiss, *Comparing different estimators in a nonlinear measurement error model I and II*, Mathematical Method of Statistics **14** (2005), no. 1-2, 53-79, 203-223.
6. H.Schneeweiss and A.Kukush, *Comparing the efficiency of structural and functional methods in measurement error models*, Discussion Paper 491, SFB 386, Ludwig - Maximilians - University of Munich (2006).

47, LOMONOSOV STR., KYIV, UKRAINE
E-mail: gontaro@ukr.net; kuechenhoff@stat.uni-muenchen.de