Journal of Numerical and Applied Mathematics
Vol.1(96), 2008, pp.139–148

ALEXANDER KUKUSH,[1] ELENA USOLTSEVA[1]

OPTIMALITY OF CS ESTIMATORS IN NONLINEAR ERRORS-IN-VARIABLES MODEL WHEN NON-INTERCEPT TERMS ARE SMALL

In nonlinear structural measurement error model a regression parameter is estimated under unknown nuisance parameters. A class L of linear-in-y estimating functions is considered, which are conditionally unbiased given the latent variable. We search for the estimating function in L which is asymptotically optimal when the true non-intercept terms are vanishing. In polynomial model, the Corrected Score estimator is optimal in this sense, while in models with exponents or trigonometric polynomial, optimal estimator is infeasible. Using the pre-estimation of nuisance parameters we construct a new estimator, which is as asymptotically efficient as the infeasible estimator.

1. Introduction

We consider a nonlinear-in-x structural measurement error model

$$y = \xi(x) + \varepsilon, \quad \xi(x) := \beta^t \rho(x), \quad w = x + u, \tag{1}$$

where $\rho(x) := (\varphi_0(x), \varphi_1(x), \ldots, \varphi_k(x))^t$, $x \in \mathbb{R}$ with $k \geq 1$ and $\varphi_0 \equiv 1$, and all the rest functions are Borel measurable; $\beta := (\beta_0, \beta_1, \ldots, \beta_k)^t$ is regression parameter. We assume that x, u, ε are mutually independent and $\mathbf{E}\varepsilon = 0$, $\mathbf{E}\varepsilon^2 = \sigma_\varepsilon^2 < \infty$, $u \sim N(0, \sigma_u^2)$, $x \sim N(\mu, \sigma_x^2)$. All the variances are positive. We suppose that σ_u^2 is known, while nuisance parameters σ_ε^2, μ, and σ_x^2 are unknown. Let (y_i, w_i, x_i), $i \geq 1$ be i.i.d. copies of (1). We observe y_i, w_i, $i = \overline{1,n}$ and want to estimate β.

For this model, the Quasi-Likelihood (QL) estimator is asymptotically optimal in a class of estimators generated by a linear-in-y unbiased estimating function, see Shklyar et al. (2007) and Kukush et al. (2007). The QL

[1] Supported by the Swedish Institute grant SI-01424/2007.
2000 *Mathematics Subject Classifications*: 62J05, 62F12.
Key words and phrases. Nonlinear errors-in-variables model, polynomial errors-in-variables model, Corrected Score estimator, asymptotic covariance matrix.

estimator relies on the assumption of normality of the regressor x, while the Corrected Score (CS) estimator, see Carroll et al. (2006), does not use this condition and therefore is robust though less efficient in the normal model.

We study optimality of the CS estimator within subclass L of estimators generated by a linear-in-y, conditionally unbiased given x estimating functions. In fact we compare in Loewner order the asymptotic covariance matrices (ACMs) Σ_L and Σ_{CS} of the latter estimators and the CS estimator. It is wrong that for any possible β, $\Sigma_{CS}(\beta) \le \Sigma_L(\beta)$, and we try to show this inequality only in case when the true non-intercept terms in model (1) are vanishing.

The paper is organized as follows. In Section 2 we introduce the class L and find an estimator in this class with the least ACM in case $\beta_{-0} = 0$,

$$\beta_{-0} := (\beta_1, \dots, \beta_k)^t. \tag{2}$$

In general this estimate is infeasible since it depends of unknown μ and σ_x^2. In Section 3 we study the polynomial model and prove that the CS estimator is optimal in the sense of Section 2, while Section 4 shows that this is not the case for the models with exponents or trigonometric polynomial. Section 5 presents a new estimator, out of class L, which has the same ACM as the infeasible optimal estimator, and Section 6 concludes.

2. Optimality within class L in general model

We use the following notations. **E** is expectation sign, **Var** stands for the variance, and **Cov** for the variance-covariance matrix. I denotes unit matrix. Symmetric matrices A and B of equal size are compared in Loewner order, i.e., $A < B$ and $A \le B$ means that $B - A$ is positive definite and positive semidefinite, respectively. E.g., $A > 0$ means that A is positive definite. For a nonsingular matrix A, $A^{-t} := (A^{-1})^t$.

2.1. Definition of L

For the model (1), consider an estimation function

$$S_L(w, y, \beta) = p(w)y - q(w, \beta) \tag{3}$$

such that

$$\mathbf{E}(S_L|x) = 0 \quad \text{a.s.} \tag{4}$$

Here p and q are Borel measurable functions with values in $\mathbb{R}^{k+1}$. We allow p and q to depend on the nuisance parameters. The estimator $\widehat{\beta}_L$ is a measurable solution to the equation

$$\sum_{i=1}^{n} S_L(w_i, y_i, \beta) = 0, \quad \beta \in \mathbb{R}^{k+1}. \tag{5}$$

Due to the theory of estimating equations, see Carroll et al. (2006), under regularity conditions $\widehat{\beta}_L$ is consistent and asymptotically normal with ACM given by the sandwich formula

$$\Sigma_L = A_L^{-1} B_L A_L^{-t}, \quad A_L := -\mathbf{E}\frac{\partial S_L}{\partial \beta^t}, \quad B_L := \mathbf{E} S_L S_L^t. \tag{6}$$

In particular the regularity conditions demand that A_L is nonsingular and $B_L > 0$. Hereafter the expectation is taken for the functions at the true β point, e.g.,

$$\mathbf{E}\frac{\partial S_L}{\partial \beta^t} := \mathbf{E}_\beta \frac{\partial S_L}{\partial \beta^t}(w, y, \beta).$$

We mention that $\widehat{\beta}_L$ is still consistent without the assumption of the normality of x, while the QL estimator (see Section 1) can lose the consistency under the violation of normality. Thus $\widehat{\beta}_L$ is robust in this sense.

First we simplify the estimating function (3), (4). Relations (4) and (1) imply that $\mathbf{E}(p(w)|x)\rho^t\beta = \mathbf{E}(q(w,\beta)|x)$, and $q(w,\beta) = Q(w)\beta$, $Q(w) \in \mathbb{R}^{(k+1)\times(k+1)}$. Thus the function S_L has a form

$$S_L(w, y, \beta) = p(w)y - Q(w)\beta \tag{7}$$

and

$$\mathbf{E}(p|x)\rho^t = \mathbf{E}(Q|x). \tag{8}$$

Therefore, equation (5) is linear w.r.t. β.

We deal only with such functions $p(w)$ that the deconvolution problem (8) for Q has a solution. If otherwise is not specified, we assume that all the considered functions of x and w elementwisely belong to the space $S'(\mathbb{R})$ of generalized slowly varying functions. Then due to the Fourier transform argument the function $Q(w)$ is unique given $p(w)$. Thus the class L consists of the estimators $\widehat{\beta}_L$ which satisfy (5), where S_L is given by (7), (8) and satisfies the regularity conditions yielding the consistency and asymptotic normality of $\widehat{\beta}_L$.

2.2. Explicit formula for ACM

We calculate the A_L and B_L given in (6). We have

$$A_L = \mathbf{E}Q(w) = \mathbf{E}[\mathbf{E}(Q(w)|x)] = \mathbf{E}[\mathbf{E}(p(w)|x)\rho(x)^t] =$$

$$= \mathbf{E}p(w)\rho(x)^t = \mathbf{E}p(w)\rho_w(w)^t, \tag{9}$$

where

$$\rho_w(w) := \mathbf{E}(\rho(x)|w). \tag{10}$$

Let $m = m(w, \beta)$ and $v = v(w, \beta, \sigma_\varepsilon^2)$ be the conditional mean and conditional variance of y given w in the model (1). Then

$$m = \mathbf{E}(\xi|w) = \beta_0 + \beta_{-0}^t \mathbf{E}(\rho_{-0}|w), \quad \rho_{-0} := (\varphi_1(x), \dots, \varphi_k(x))^t; \tag{11}$$

$$v = \sigma_\varepsilon^2 + \mathbf{Var}(\beta_{-0}^t \rho_{-0}|w) = \sigma_\varepsilon^2 + \beta_{-0}^t M \beta_{-0}, \tag{12}$$

$M = M(w) \in \mathbb{R}^{k\times k}$, $M_{ij} = \mathbf{E}(\rho_i\rho_j|w) - \rho_{wi}\rho_{wj}$, $i, j = \overline{1,k}$, where ρ_{wi} and ρ_{wj} are components of (10).

Next,

$$B_L = \mathbf{E}[(y-m)p - (mp - Q\beta)][(y-m)p + (mp - Q\beta)]^t =$$

$$= \mathbf{E}vpp^t + \mathbf{Cov}(mp - Q\beta). \tag{13}$$

Now, (6), (9), and (13) imply the representation

$$\Sigma_L = (\mathbf{E}p\rho_w^t)^{-1}(\mathbf{E}vpp^t + \mathbf{Cov}(mp - Q\beta))(\mathbf{E}p\rho_w^t)^{-t}. \tag{14}$$

Lemma 1. *Let $\widehat{\beta}_L$ belongs to the class L. Then the ACM of $\widehat{\beta}_L$ does not depend of the intercept β_0.*

Proof. Analyzing (14) and (12) it remains to show only that $\mathbf{Cov}(mp - Q\beta)$ is independent of β_0. Indeed, by (11) the terms of $mp - Q\beta$ which involve β_0 have a form: $z = \beta_0(p_0, p_1, \dots, p_k)^t - \beta_0(Q_{00}, Q_{10}, \dots, Q_{k0})^t$. But (8) implies that $\mathbf{E}(p_i|x) = \mathbf{E}(Q_{i0}|x)$, $i = \overline{0,k}$, and since the solution to the deconvolution problem is unique under our assumptions, we have $p_i = Q_{i0}$, for all i. Therefore $z = 0$. □

Recall that β_{-0} is defined in (2).

Corollary 2. *For all $\beta \in \Theta$,*

$$\Sigma_L = \Sigma_L(\beta_{-0}) \geq \Sigma_L|_{\beta_{-0}=0}, \tag{15}$$

and

$$\Sigma_L|_{\beta_{-0}=0} = \sigma_\varepsilon^2(\boldsymbol{E}p\rho_w^t)^{-1}\boldsymbol{E}pp^t(\boldsymbol{E}p\rho_w^t)^{-t}. \tag{16}$$

Proof is straightforward and relies on Lemma 1, (14), and (12). □

Inequality (15) means that for $\widehat{\beta}_L \in L$, the most informative values of the true regression parameter are $\beta_{-0} = 0$, $\beta_0 \in \mathbb{R}$, i.e. when the non-intercept terms in the model (1) are vanishing.

2.3. Optimal in L estimator

Consider the estimator $\widehat{\beta}^*$, which is defined by (5) if we set in (7) and (8)

$$p(w) = \rho_w(w), \quad w \in \mathbb{R}. \tag{17}$$

Theorem 3. *Assume that $\widehat{\beta}^* \in L$. Then for any $\widehat{\beta} \in L$,*

$$\Sigma_L|_{\beta_{-0}=0} \geq \Sigma^*|_{\beta_{-0}=0}, \tag{18}$$

where Σ^* is the ACM of $\widehat{\beta}^*$. The equality in (18) is attained only for $\widehat{\beta} = \widehat{\beta}^*$.

Proof. Denote that RHS of (16) as $\sigma_\varepsilon^2 \Phi(p)$. We have to minimize $\Phi(p)$ in Loewner order. Denote also $S = \mathbf{E}pp^t$, $r = S^{-\frac{1}{2}}p$. Rewrite $\Phi(p) = (\mathbf{E}r\rho_w^t)^{-1}(\mathbf{E}r\rho_w^t)^{-t}$, $\Phi(p)^{-1} = (\mathbf{E}r\rho_w^t)^t(\mathbf{E}r\rho_w^t)$. By the Cauchy-Schwartz inequality for random vectors, see Kukush et al. (2007), since $\mathbf{E}rr^t = I$ we have

$$(\mathbf{E}r\rho_w^t)^t(\mathbf{E}r\rho_w^t) \leq \mathbf{E}\rho_w\rho_w^t. \tag{19}$$

The RHS of (19) is positive definite since $\widehat{\beta}^* \in L$, and under (17) $A_L = \mathbf{E}p\rho_w^t = \mathbf{E}\rho_w\rho_w^t$ is nonsingular and symmetric. Then (19) is equivalent to

$$\Phi(p) \geq (\mathbf{E}\rho_w\rho_w^t)^{-1} = \frac{1}{\sigma_\varepsilon^2}\Sigma^*|_{\beta_{-0}=0}.$$

This implies (18). The equality in (18) is attained if, and only if, the equality is attained in the Cauchy-Schwartz inequality (19), i.e. iff $r(w) = R_1\rho_w(w)$, R_1 is a nonsingular nonrandom matrix, see Kukush et al. (2007) for the proof. This happens iff

$$S^{-\frac{1}{2}}p = R_1\rho_w, \quad \text{or} \quad p = R_2\rho_w, \tag{20}$$

where $R_2 := S^{-\frac{1}{2}}R_1$ is nonsingular nonrandom matrix. Next, let $Q^*(w)$ solve (8) with $p = \rho_w$. Then $Q_2 := R_2Q^*$ solves (8) with p given in (20). For such p and Q_2, the equation (5) takes a form

$$R_2\sum_{i=1}^n(\rho_w(w_i)y_i - Q^*(w_i)\beta) = 0, \quad \beta \in \mathbb{R}^{k+1},$$

or

$$\sum_{i=1}^n(\rho_w(w_i)y_i - Q^*(w_i)\beta) = 0, \quad \beta \in \mathbb{R}^{k+1}.$$

Thus a measurable solution $\widehat{\beta}$ to this equation coincides with $\widehat{\beta}^*$. □

Remark 4. In general the estimator $\widehat{\beta}^*$ is infeasible. Indeed, to compute (10) we need the conditional distribution of x given w, and it includes the unknown nuisance parameters μ and σ_x^2.

3. Corrected Score estimate and its optimality in polynomial model

Definition 5. Corrected Score (CS) is an estimating function

$$S_{CS}(w, y, \beta) = g(w)y - H(w)\beta \tag{21}$$

such that

$$\mathbf{E}(g|x) = \rho(x), \quad \mathbf{E}(H|x) = \rho(x)\rho(x)^t. \tag{22}$$

The CS estimator $\widehat{\beta}_{CS}$ is a measurable solution to linear equation (5) with $S_L = S_{CS}$.

Hereafter we assume that both deconvolution problems (22) have solutions (in $S'(R)$). Moreover we suppose that the regularity conditions mentioned in Section 2 hold for the function (21), and thus $\widehat{\beta}_{CS} \in L$.

We mention that a.s.

$$\mathbf{E}(S_{CS}|y, x) = \rho(x)y - (\rho(x)\rho(x)^t)\beta =: S_{ML}(x, y, \beta), \tag{23}$$

and S_{ML} is log-likelihood score for the normal regression model without measurement error

$$y = \beta^t \rho(x) + \varepsilon, \quad \varepsilon \sim N(0, \sigma_\varepsilon^2).$$

The relation (23) shows that the score (21) adjusts the log-likelihood score for the measurement errors. The CS is a good choice when the law of x is not specified, and also in functional case, where the true covariates x_i, $i = \overline{1, n}$ are nonrandom (and unknown), see Carrol et al. (2006).

Now, we pass to a polynomial errors-in-variables model. It is the model (1) with

$$\rho(x) = (1, x, \ldots, x^k)^t, \quad x \in \mathbb{R}; \quad k \geq 1. \tag{24}$$

All the assumptions and constructions from Sections 1 and 2 are valid for this model. As before we need not the normality of ε. It is known, see Carrol et al. (2006), that for the polynomial model $\widehat{\beta}_{CS} \in L$.

Theorem 6. *In the polynomial model, the optimal estimator $\widehat{\beta}^*$ (see Theorem 3) coincides with the CS estimator.*

Proof. In view of the proof of Theorem 3 it is enough to check for the function (10), (24) that

$$\rho_w(w) = Rg(w), \quad w \in \mathbb{R}, \tag{25}$$

where R is nonsingular nonrandom matrix and $g(w)$ satisfies (22), (24).

First connect ρ_w and ρ. We have under the normality assumptions, see Carroll et al. (2006), that

$$x|w \sim N(\mu_1 + Kw, \tau^2). \tag{26}$$

Here

$$K := \frac{\sigma_x^2}{\mathbf{Var}(w)} = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \tag{27}$$

is the reliability ratio, $\mu_1 := (1-K)\mu$ (remember that $\mu = \mathbf{E}x$), $\tau^2 := K\sigma_u^2$. Let $\gamma \sim N(0,1)$ be independent of w. We have for $i \geq 1$ $\mathbf{E}(x^i|w) = \mathbf{E}((\mu_1 + Kw + \tau\gamma)^i|w) = \sum_{k=0}^{i} C_i^k(\mu_1 + Kw)^k \tau^{i-k}\mathbf{E}\gamma^{i-k} =: p_i(w)$. Here p_i is a polynomial of degree i. This implies that

$$\rho_w(w) = \mathbf{E}(\rho(x)|w) = R_1\rho(w) \tag{28}$$

with a nonsingular nonrandom lower triangular matrix $R_1 = R_1(\mu, \sigma_x^2)$.

Next, in the polynomial model

$$g(w) = R_2\rho(w), \tag{29}$$

where R_2 is a nonsingular nonrandom lower triangular matrix as well, see Shklyar et al. (2007). Now, (28) and (29) imply (25) with $R = R_1R_2^{-1}$. And Theorem 6 follows. □

4. Two particular models with exponents and trigonometric polynomial

4.1. The model with exponential functions

In this subsection we violate the demand that $\varphi_i \in S'(\mathbb{R})$, $i \geq 1$. Still for the considered simple model Theorem 3 holds true.

Let $\{\lambda_i, i = \overline{0,k}\} \subset \mathbb{R}$, $\lambda_0 = 1$, $\lambda_i \neq \lambda_j$, $i \neq j$. Consider the model (1) with

$$\rho(x) = (1, e^{\lambda_1 x}, \ldots, e^{\lambda_k x})^t, \quad x \in \mathbb{R}. \tag{30}$$

We apply Theorem 3 and look more closely at the function ρ_w. As in the proof of Theorem 6 we have for $\lambda \in \mathbb{R}$, see (26) and (27), that $\mathbf{E}(e^{\lambda x}|w) = \mathbf{E}(e^{\lambda(\mu_1 + Kw + \tau\gamma)}|w) = c(\lambda)e^{K\lambda w}$, $c(\lambda) := exp(\lambda\mu_1 + \frac{\lambda^2\tau^2}{2})$.

Hence

$$\rho_w(w) = (1, c(\lambda_1)e^{K\lambda_1 w}, \ldots, c(\lambda_k)e^{K\lambda_k w})^t = R_3(1, e^{K\lambda_1 w}, \ldots, e^{K\lambda_k w})^t$$

with nonsingular diagonal matrix R_3. Thus the optimal $\widehat{\beta}^*$ is generated by (7) with

$$p(w) = (1, e^{K\lambda_1 w}, \ldots, e^{K\lambda_k w})^t. \tag{31}$$

For the functions (31) and (30), $Q = Q(w; \mu, \sigma_x^2)$ can be easily found from (8) in the exponential form. The estimator $\widehat{\beta}^*$ is infeasible because the reliability ratio K is unknown under unknown σ_x^2.

4.2. Trigonometric model

Now, in the model (1) we specify $\xi(x)$ as

$$\xi(x) = \beta_0 + \sum_{i=1}^{k} (\beta_i \cos \lambda_i x + \alpha_i \sin \lambda_i x), \quad x \in \mathbb{R}. \tag{32}$$

Here $0 < \lambda_1 < \lambda_2 < \ldots < \lambda_k$, and $\beta := (\beta_0, \beta_1, \ldots, \beta_k, \alpha_1, \ldots, \alpha_k)^t$ is the parameter of interest. Similarly to Section 4.1 the optimal $\widehat{\beta}^*$ is generated by (7) with $p(w) = (1, \cos K\lambda_1 w, \ldots, \cos K\lambda_k w, \sin K\lambda_1 w, \ldots, \sin K\lambda_k w)^t$. The $\widehat{\beta}^*$ is infeasible as well.

5. Feasible estimator for the model with exponents

Return to the model (1), (30). The estimation function for the infeasible estimator $\widehat{\beta}^*$ can be written as

$$S^{(\beta)}(w, y, \beta; \mu, \sigma_x^2) = P(w; \sigma_x^2) y - Q(w; \mu, \sigma_x^2)\beta, \tag{33}$$

where p is given in (31), (27) and Q satisfies (8), (30). For unknown nuisance parameters, we use the pre-estimation:

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} w_i, \quad \widehat{\sigma}_x^2 = \frac{1}{n}\sum_{i=1}^{n}(w_i - \widehat{\mu})^2 - \sigma_u^2. \tag{34}$$

Remark 7. The estimator $\widehat{\sigma}_x^2$ is biased. It is better to use instead an unbiased estimator

$$S_x^2 := \frac{1}{n-1}\sum_{i=1}^{n}(w_i - \widehat{\mu})^2 - \sigma_u^2.$$

But $S_x^2 - \widehat{\sigma}_x^2 = \frac{O_p(1)}{n}$, therefore using S_x^2 instead of $\widehat{\sigma}_x^2$ leads to the same ACM for estimators of β, μ, and $\widehat{\sigma}_x^2$. For theoretical analysis it is more convenient to deal with $\widehat{\sigma}_x^2$ since it contains a summand having an average form.

Now, let $\theta := (\beta^t, \mu, \sigma_x^2)^t$ be a compound parameter vector. Introduce a compound estimating function

$$S_*(w, y, \beta; \mu, \sigma_x^2) = \begin{pmatrix} S^{(\beta)} \\ S^{(\mu)} \\ S^{(\sigma_x^2)} \end{pmatrix} = \begin{pmatrix} S^{(\beta)}(w, y, \beta; \mu, \sigma_x^2) \\ w - \mu \\ (w-\mu)^2 - \sigma_x^2 - \sigma_u^2 \end{pmatrix},$$

where $S^{(\beta)}$ is given in (33). A new estimator $\widehat{\theta}_* := (\widehat{\beta}_*^t, \widehat{\mu}, \widehat{\sigma}_x^2)^t$ is a measurable solution to the equation

$$\sum_{i=1}^{n} S_*(w_i, y_i, \beta; \mu, \sigma_x^2) = 0, \quad \beta \in \mathbb{R}^{k+1}, \quad \mu \in \mathbb{R}, \quad \sigma_x^2 > 0. \tag{35}$$

Thus the new estimator $\widehat{\beta}_*$ is a solution to:

$$\sum_{i=1}^{n} S^{(\beta)}(w_i, y_i, \beta; \widehat{\mu}, \widehat{\sigma}_x^2) = 0, \quad \beta \in \mathbb{R}^k, \tag{36}$$

whereas the last two components of $\widehat{\theta}_*$ are the empirical estimators (34). Therefore, the pre-estimation procedure (36), (34) leads to the simultaneous estimator $\widehat{\theta}_*$ given by (35). Next, $\widehat{\theta}_*$ is consistent and asymptotically normal with the ACM

$$\Sigma_*^{(\theta)} = \begin{pmatrix} \Sigma_*^{(\beta)} & \bullet & \bullet \\ \bullet & \Sigma_*^{(\mu)} & \bullet \\ \bullet & \bullet & \Sigma_*^{(\sigma_x^2)} \end{pmatrix}.$$

Here $\Sigma_*^{(\beta)}$, $\Sigma_*^{(\mu)}$, and $\Sigma_*^{(\sigma_x^2)}$ are ACM of $\widehat{\beta}_*$, asymptotic variance of $\widehat{\mu}$, and asymptotic variance of $\widehat{\sigma}_x^2$, respectively.

Theorem 8. *In the model (1), (30), the infeasible estimator $\widehat{\beta}^*$ and the new feasible estimator $\widehat{\beta}_*$ have the same ACM in case*

$$\beta_{-0} = 0. \tag{37}$$

Proof. Let (37) holds. The ACM $\Sigma_*^{(\theta)}$ can be computed by the sandwich formula similar to (6). Remember that Σ^* is the ACM of $\widehat{\beta}^*$, see Theorem 3. In order to show that

$$\Sigma_*^{(\beta)} = \Sigma^*, \tag{38}$$

it is enough to check that

$$\mathbf{E}\frac{\partial S^{(\beta)}}{\partial \mu} = 0, \tag{39}$$

$$\mathbf{E}\frac{\partial S^{(\beta)}}{\partial \sigma_x^2} = 0, \tag{40}$$

cf. Shklyar et al. (2007). Now, under (37), cf. the proof of Lemma 1, we have

$$S^{(\beta)} = p(w; \sigma_x^2)(y - \beta_0).$$

Then (39) follows immediately, and (40) holds as well since $\mathbf{E}(y|w)|_{\beta_{-0}=0} = \beta_0$. Thus (38) is shown. □

Remark 9. Similar statement holds for the trigonometric model (1), (32). Moreover Theorem 8 can be easily extended to general model (1).

6. Conclusion

For the polynomial model, we proved the optimality of the CS in the class L in rather restrictive sense. Of course it would be much better to show that

$$\forall \beta \in \mathbb{R}^{k+1} \quad \forall \widehat{\beta}_L \in L: \quad \Sigma_{CS}(\beta) \le \Sigma_L(\beta). \tag{41}$$

Alas (41) is wrong. The set of all estimating functions for $\widehat{\beta}_L \in L$ forms a cone, and the criterion from Heyde (1997) states that (41) holds iff the matrix

$$(\mathbf{E}\frac{\partial S_L}{\partial \beta^t})^{-1}\mathbf{E}S_L S_{CS}^t$$

does not depend of β and of the choice of S_L from the cone. It is possible to verify that this is not true, therefore (41) is wrong.

For the two models with exponents and trigonometric polynomial, we proved that the optimal estimator is infeasible and constructed a new estimator for β, out of the class L, which has the same ACM as the optimal estimator within L. The new estimator is robust in the following sense: it is still consistent in the structural model where the latent variable is not normally distributed and has finite second moment (for the model with exponents the exponential moments should be finite). The Quasi-Likelihood estimator relies on the normality assumption and is not robust in this sense, though it is more efficient in the normal model.

References

1. Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M., *Measurement Error in Nonlinear Models. A Modern Perspective,* 2-nd ed., Chapman & Hall/CRC, London, (2006).
2. Heyde, C. C., *Quasi-Likelihood And Its Application,* Springer, New York, (1997).
3. Kukush A., Malenko A., and Schneeweiss H., *Optimality of the quasi-score estimator in a mean-variance model with application to measurement error models,* Discussion Paper 494, SFB 386, Universität München, (2006).
4. Kukush, A., Malenko A., and Schneeweiss H., *Comparing the efficiency of estimates in concrete errors-in-variables models under unknown parameters,* Theory of Stochastic Processes, 13(29), N4, (2007), 69-81.
5. Shklyar, S., Schneeweiss, H., and Kukush, A., *Quasi Score is more efficient than Corrected Score in a polynomial measurement error model,* Metrika, 65, (2007), 275-295.

Department of Mechanics and Mathematics,
National Taras Shevchenko University of Kyiv,
Volodymyrska st. 64, 01033 Kyiv, Ukraine.
E-mail: alexander_kukush@univ.kiev.ua, elena_usolceva@ukr.net