Optimality of the quasi-score estimator in a mean-variance model with applications to measurement error models

Alexander Kukush, Andrii Malenko

Kyiv National Taras Shevchenko University, Ukraine

Hans Schneeweiss

University of Munich, Germany

Abstract

We consider a regression of y on x given by a pair of mean and variance functions with a parameter vector θ to be estimated that also appears in the distribution of the regressor variable x. The estimation of θ is based on an extended quasi score (QS) function. We show that the QS estimator is optimal within a wide class of estimators based on linear-in-y unbiased estimating functions. Of special interest is the case where the distribution of x depends only on a subvector α of θ, which may be considered a nuisance parameter. In general, α must be estimated simultaneously together with the rest of θ, but there are cases where α can be pre-estimated. A major application of this model is the classical measurement error model, where the corrected score (CS) estimator is an alternative to the QS estimator. We derive conditions under which the QS estimator is strictly more efficient than the CS estimator.

Keywords: Mean-variance model, measurement error model, quasi score estimator, corrected score estimator, nuisance parameter, optimality property.

MSC 2000: 62J05, 62J12, 62F12, 62F10, 62H12, 62J10.

Abbreviated title: Optimality of Quasi-Score.

Acknowledgements. Support by Deutsche Forschungsgemeinschaft (German Research Foundation) is gratefully acknowledged. The authors are grateful to Dr. Sergiy Shklyar for fruitful discussions and to an anonymous referee for helpful suggestions to improve the paper.

1 Introduction

Suppose that the relation between a response variable y and a covariate (or regressor) x is given by a pair of conditional mean and variance functions:

$$\mathbb{E}\,(y|x) =: m(x, \theta), \quad \mathbb{V}(y|x) =: v(x, \theta). \tag{1}$$

Here θ is an unknown d-dimensional parameter vector to be estimated. The parameter θ belongs to the interior of a compact parameter set Θ. The variable x has a density $\rho(x, \theta)$ with respect to a σ-finite measure ν on a Borel σ-field on the real line. We assume that $v(x, \theta) > 0$, for all x and θ, and that all the functions are sufficiently smooth. Such a model is called a *mean-variance* model, cf. Carroll *et al.* (2006). We want to estimate θ on the basis of an i.i.d. sample (x_i, y_i), $i = 1, \ldots, n$.

The remarkable feature of this model is that the parameter θ appears not only in the mean and variance functions but also in the density function of the regressor. This may seem to be a rather artificial assumption. But note that not all components of θ need to appear in the mean-variance functions and in the density function simultaneously, and we shall see that models with partial overlap of parameters in both types of functions do appear in practice. In the meantime our general assumption of a common parameter vector θ serves as a very convenient starting point. We construct an estimator of θ that takes this feature into account. We do so by basing the estimator on an (unbiased) estimating function that depends not only on m and v, but also on ρ; it depends on m and v via the conventional quasi score function, cf. Carroll *et al.* (2006), Wedderburn (1974), Armstrong (1985), Heyde (1997), and on ρ via the log-likelihood of the distribution of x. This compound estimating function might therefore be called an *extended quasi score* (QS) function, but for simplicity, we will just call it the *quasi score* (QS) function and the corresponding estimator the QS estimator. The QS estimator turns out to be optimal within a wide class of so-called *linear score* (LS) estimators.

A very important special model is given, when θ consists of two subvectors α and β, where α is a parameter describing the distribution of x. But m and v still depend on the whole of θ, i.e., on α and β. In this case, we might be mainly interested in the estimation of β, while α is a nuisance parameter. Again the remarkable trait of this model is that the parameter α not only determines the distribution of x but also the mean and variance functions, something that does not occur in an ordinary regression model. However, a model of this type arises naturally in the context of measurement error models, Fuller (1987), Cheng and Van Ness (1999), Carroll *et al.* (2006). Measurement error models form a central part of our paper. The most important LS estimator in a measurement error model, apart from QS, is the so-called *corrected score* (CS) estimator, cf. Stefanski (1989), Nakamura (1990).

As the mean and variance functions depend on α and β, these parameters have to be estimated simultaneously within the QS approach. This is the main difference of our QS approach to the more traditional one, which consists in first estimating α separately, using only the data x_i, and then, after substituting $\hat{\alpha}$ for α in the quasi score function of β, finding an estimate of β, cf. Carroll *et al.* (2006). But there are some important models, where α (or part of α) can, in fact, be estimated in advance, without invalidating the superiority property of QS *vis-a-vis* to CS – we say α can be pre-estimated. Among such models, the polynomial model is the most prominent one.

We not only can state the optimality of QS within the class of linear scores, but we can also give conditions under which this optimality is strict in the sense that the difference of the asymptotic covariance matrices of the estimators is positive definite and not just positive semidefinite. We also give conditions under which QS and CS are equally efficient.

The present paper is a continuation of a research started in Kukush and Schneeweiss (2006), where a mean-variance model was considered under *known* nuisance parameters and the efficiency of the QS estimator (in the usual sense) was compared to the LS estimator. In the present paper, we study the much more realistic case of *unknown* nuisance parameters.

We assume regularity conditions, which make it possible to differentiate integrals with respect to parameters and which guarantee that the considered estimators, generated by unbiased scores, are consistent and asymptotically normal with asymptotic covariance matrices that are given by the sandwich formula, see Carroll *et al.* (2006). These regularity conditions are discussed in Kukush and Schneeweiss (2005) for a nonlinear measurement error model. See also the discussion concerning the sandwich formula in Schervish (1995), p. 428.

We use the symbols $\mathbb{E}$ to denote the expectation of random values, vectors, and matrices and $\mathbb{V}$ to denote the variance or the covariance matrix. We often omit the arguments of functions, e.g., instead of $\rho(x, \theta)$ we write ρ for simplicity. All vectors are considered to be column vectors. We use subscripts to indicate partial derivatives with respect to some or all of the parameters, e.g., $\rho_\theta = \frac{\partial \rho}{\partial \theta}$. For any scalar function, its derivative with respect to a vector is a column vector and for a vector it is a matrix. We compare real matrices in the Loewner order, i.e., for symmetric matrices A and B of equal size, $A < B$ and $A \leq B$ means that $B - A$ is positive definite and positive semidefinite, respectively.

The paper is organized as follows. In Section 2, we introduce the class of linear unbiased scores and our new QS estimator as a special member of this class. Section 3 contains general results on the comparison of QS and LS estimators. In Section 4, we specialize our general model to the case of a regression model with nuisance parameters. Here we also introduce the measurement error model and the corrected score (CS) estimator as a special member of the class of LS estimators. Section 5 deals with cases where pre-estimation of the nuisance parameters is possible. Section 6 concludes. Two lemmas and the proofs of the main theorems are given in the appendix.

2 Class of linear scores

The estimation of θ in the mean-variance model (1) cannot be accomplished by using the maximum likelihood (ML) approach because the conditional distribution of y given x is by assumption not known. Instead an estimator of θ is based on an unbiased estimating (or score) function, which we suppose to be given. A rather general class of estimating functions is the class $\mathcal{L}$ of all unbiased linear-in-y score functions (for short: linear score (LS) functions):

$$S_L(x, y; \theta) := yg(x, \theta) - h(x, \theta), \tag{2}$$

where unbiasedness means that $\forall\, \theta \in \Theta : \ \ \mathbb{E}\, S_L(x, y; \theta) = 0$. Here g and h are vector-valued functions of dimension d, the same dimension as θ. The expectation is meant to be carried out under the same θ as the θ of the argument. Of course, wider classes of score functions are possible, Heyde(1997), but here we restrict our discussion to the linear class.

The estimator of θ based on S_L is called *linear score* (LS) estimator $\hat{\theta}_L$ and is given as the solution to the equation $\sum_{i=1}^{n} S_L(x_i, y_i; \hat{\theta}_L) = 0$. Under general conditions, see Appendix 7.5, $\hat{\theta}_L$ exists and is consistent and asymptotically normal. The asymptotic covariance matrix (ACM) Σ_L of $\hat{\theta}_L$ is given by the sandwich formula, cf. Heyde (1997),

$$\Sigma_L = A_L^{-1} B_L A_L^{-\top}, \quad A_L = -\,\mathbb{E}\, S_{L\theta}, \quad B_L = \mathbb{E}\, S_L S_L^{\top}. \tag{3}$$

A_L is supposed to be nonsingular (*identifiability condition*).

The condition of unbiasedness of the score function amounts to the statement that $\mathbb{E}\,(yg - h) = 0$, which is equivalent to

$$\mathbb{E}\,(mg - h) = 0. \tag{4}$$

In a mean-variance model, one can construct the so-called *quasi-score* (QS) estimator as a special LS estimator. It is based on the following quasi-score function S_Q:

$$S_Q(x, y; \theta) := \frac{(y - m)m_\theta}{v} + l_\theta, \tag{5}$$

where $l := \log \rho(x, \theta)$. It differs from the usual quasi-score function as exemplified, e.g., in Heyde (1997), by the term l_θ. It is obviously unbiased (i.e., $\mathbb{E}\, S_Q = 0$), and $\mathbb{E}\, S_Q S_Q^{\top} = \mathbb{E}\, v^{-1} m_\theta m_\theta^{\top} + \mathbb{E}\, l_\theta l_\theta^{\top}$. We assume that $\mathbb{E}\, S_Q S_Q^{\top}$ is positive definite (*identifiability condition* for QS).

This *identifiability condition* is equivalent to the condition that the d two-dimensional random vectors

$$\begin{pmatrix} l_{\theta_i} \\ m_{\theta_i} \end{pmatrix}, \quad i = 1, \ldots, d, \tag{6}$$

are linearly independent.

The QS estimator $\hat{\theta}_Q$ of θ is defined as the solution to the equation

$$\sum_{i=1}^{n} S_Q(x_i, y_i, \hat{\theta}_Q) = 0. \tag{7}$$

As the quasi-score function (5) belongs to $\mathcal{L}$ with $g = g_Q = \frac{m_\theta}{v}$ and $h = h_Q = \frac{mm_\theta}{v} - l_\theta$, the estimator $\hat{\theta}_Q$ is consistent and asymptotically normal under regularity conditions (Appendix 7.5) with an ACM given by (3).

3 Comparison of QS to LS

We want to compare Σ_Q to Σ_L. To this purpose, we derive alternative formulas for the ACMs of the LS estimator $\hat{\theta}_L$ and of the QS estimator $\hat{\theta}_Q$:

Lemma 3.1

$$\Sigma_L = \left(\mathbb{E}\, S_L S_Q^\top\right)^{-1} \mathbb{E}\, S_L S_L^\top \left(\mathbb{E}\, S_L S_Q^\top\right)^{-\top} \tag{8}$$

$$\Sigma_Q = \left(\mathbb{E}\, S_Q S_Q^\top\right)^{-1}. \tag{9}$$

Proof: We first have from (2)

$$\mathbb{E}\, S_{L\theta} = \mathbb{E}\,(mg_\theta - h_\theta). \tag{10}$$

On the other hand,

$$\begin{aligned} \mathbb{E}\, S_L S_Q^\top &= \mathbb{E}\left[(mg - h) + (y - m)g\right]\left[\frac{(y-m)m_\theta}{v} + l_\theta\right]^\top \\ &= \mathbb{E}\,(mg - h)l_\theta^\top + \mathbb{E}\, g m_\theta^\top. \end{aligned} \tag{11}$$

We can derive the following identity from (4):

$$\mathbb{E}\,(mg-h)_\theta + \mathbb{E}\,(mg-h)l_\theta^\top = 0. \tag{12}$$

From (10), (11), and (12) we obtain

$$\mathbb{E}\,S_{L\theta} + \mathbb{E}\,S_L S_Q^\top = \mathbb{E}\,(mg-h)_\theta + \mathbb{E}\,(mg-h)l_\theta^\top = 0,$$

which yields

$$\mathbb{E}\,S_{L\theta} = -\,\mathbb{E}\,S_L S_Q^\top. \tag{13}$$

Now, (13) implies that the ACM of $\hat{\theta}_L$, given by (3), can be written as in (8). Finally, as S_Q belongs to $\mathcal{L}$, we can apply (8) to S_Q and obtain (9) for the ACM of $\hat{\theta}_Q$. This completes the proof.

We now can state the following theorems.

Theorem 3.1 (Optimality of QS) *Let S_L be a score function from the class $\mathcal{L}$ and S_Q be the quasi-score function (5). Then*

$$\Sigma_Q \leq \Sigma_L.$$

Moreover, $\Sigma_L = \Sigma_Q$ for all θ if, and only if, $\hat{\theta}_L = \hat{\theta}_Q$ a.s.

Remark 1. Depending on the model involved, there may be other estimators that are more efficient than QS (e.g., ML), but according to the theorem they would imply a non-linear-in-y score function.

Theorem 3.2 (Strict Optimality of QS) *Under the conditions of Theorem 3.1*

$$\operatorname{rank}(\Sigma_L - \Sigma_Q) = \operatorname{rank}\left[\begin{pmatrix} mg_i - h_i \\ vg_i \end{pmatrix}, \begin{pmatrix} l_{\theta_i} \\ m_{\theta_i} \end{pmatrix}, i = 1, \ldots, d\right] - d, \quad (14)$$

where rank $[\cdot]$ *is the maximum number of linearly independent random vectors inside the square brackets. In particular,*

$$\Sigma_Q < \Sigma_L$$

if, and only if, the random vectors in (14) are linearly independent.

If

$$\begin{aligned} &\operatorname{span}\left\{\begin{pmatrix} mg_i - h_i \\ vg_i \end{pmatrix}, i = 1, \ldots, d\right\} \\ \cap \ &\operatorname{span}\left\{\begin{pmatrix} l_{\theta_i} \\ m_{\theta_i} \end{pmatrix}, i = 1, \ldots, d\right\} = \left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right\}, \end{aligned}$$

then

$$\operatorname{rank}(\Sigma_L - \Sigma_Q) = \operatorname{rank}\left[\begin{pmatrix} h_i \\ g_i \end{pmatrix}, i = 1, \ldots, d\right].$$

Here g_i and h_i are the i-th components of the vectors g and h, respectively, $i = 1, \ldots, d$. As a consequence, we have the following corollary:

Corollary 3.1 *A sufficient condition for $\Sigma_Q < \Sigma_L$ is that the random variables*

$$\{mg_i - h_i, \ l_{\theta_j}, \ i = 1, \ldots, d, \ j \in B_\theta\} \quad (15)$$

are linearly independent, where $\{l_{\theta_j}, j \in B_\theta\}$ *is a basis of* $\operatorname{span}\{l_{\theta_j},\, j = 1, \ldots, d\}$.

Remark 2. The inequality $\Sigma_Q \leq \Sigma_L$ of Theorem 3.1 can also be obtained as a direct consequence of identity (13) and Heyde's (1997) criterion for asymptotic optimality.

Remark 3. Sometimes the conditional variance depends also on an unknown parameter $\varphi \in \mathbb{R}^+$, $v = v(x, \theta, \varphi)$, while neither $m(x, \theta)$ nor the distribution of x depend on φ. It can be shown, cf. Kukush *et al.* (2006), that this does not change the results of this paper, so that φ can be treated as if it were a known parameter.

4 Estimation of a nuisance parameter in a regression model

4.1 General regression model with nuisance parameter

In this section we deal with an important special case of our general model. We suppose that θ is split into two subvectors, $\theta^\top = (\beta^\top,\ \alpha^\top)$, $\beta \in \mathbb{R}^k$, $\alpha \in \mathbb{R}^{d-k}$, such that the density of x depends only on α: $\rho = \rho(x, \alpha)$, whereas the mean and variance functions may still depend on both β and α. In this case, β can be seen as the regression parameter and is usually the parameter of interest, while α is a nuisance parameter.

The quasi-score function (5) takes the form

$$S_Q = \begin{pmatrix} (y-m)v^{-1}m_\beta \\ (y-m)v^{-1}m_\alpha + l_\alpha \end{pmatrix}. \tag{16}$$

Such a model arises naturally in the context of measurement error models, see Section 4.2. All the previous results hold true.

We obtain more detailed results if, corresponding to the special QS function (16), we also choose a special subclass $\mathcal{L}^* \subset \mathcal{L}$, to which (16) can then be compared. The corrected score function of the next subsection will be an example of an element of $\mathcal{L}^*$. Assume that S_L is of the form

$$S_L = \begin{pmatrix} yg(x,\beta) - h(x,\beta) \\ l_\alpha \end{pmatrix}, \tag{17}$$

where now g and h are of dimension k and do not depend on α. Unbiasedness of S_L again means that $\mathbb{E}(mg-h)=0$ because $\mathbb{E}\, l_\alpha = 0$ anyway. Note that S_Q is not a member of this restricted class. Nevertheless, we can still apply Theorems 3.1 and 3.2 with $\mathcal{L}$ replaced by $\mathcal{L}^*$ to compare Σ_L to Σ_Q. In particular, the first part of Theorem 3.2 takes the form:

Theorem 4.1 *If* $\theta = (\beta^\top, \alpha^\top)^\top$ *and* $\rho = \rho(x,\alpha)$ *and* S_L *is of the form* (17), *then*

$$
\begin{aligned}
&\operatorname{rank}(\Sigma_L - \Sigma_Q) + d \\
&= \operatorname{rank}\left[\begin{pmatrix} mg_i - h_i \\ vg_i \end{pmatrix}, \begin{pmatrix} 0 \\ m_{\beta_i} \end{pmatrix}, \begin{pmatrix} 0 \\ m_{\alpha_j} \end{pmatrix}, \begin{pmatrix} l_{\alpha_j} \\ 0 \end{pmatrix} \quad \begin{matrix} i=1,\dots,k \\ j=1,\dots,\ d\text{-}k \end{matrix}\right].
\end{aligned}
$$

4.2 Measurement error model

The model of Subsection 4.1 typically arises from a measurement error model. This is a model where the response variable y depends on a latent (unobservable) variable ξ with distribution $\rho(\xi, \alpha)$. The variable ξ can be observed only indirectly via a surrogate variable x, which is related to ξ through a measurement equation of the form

$$x = \xi + \delta, \tag{18}$$

where the measurement error δ is independent of ξ and y and $\mathbb{E}\,\delta = 0$. Additionally, we assume $\delta \sim N(0, \sigma_\delta^2)$ with σ_δ^2 known.

The dependence of y on ξ is either given by a conditional distribution of y given ξ or simply by a conditional mean function supplemented by a conditional variance function:

$$\mathbb{E}\,(y|\xi) = m^*(\xi, \beta), \quad \mathbb{V}(y|\xi) = v^*(\xi, \beta). \tag{19}$$

Note that m^* and v^* do not depend on α. From (19) we can derive conditional mean and variance functions of y given x, which do depend on α:

$$m(x,\beta,\alpha) := \mathbb{E}\,(y|x) = \mathbb{E}\,[m^*(\xi,\beta)|x] \tag{20}$$

$$v(x,\beta,\alpha) := \mathbb{V}(y|x) = \mathbb{E}\,[v^*(\xi,\beta)|x] + \mathbb{V}[m^*(\xi,\beta)|x]. \tag{21}$$

To compute these, we need to know the conditional distribution of ξ given x, which we can derive from the unconditional distribution of ξ, $\rho(\xi,\alpha)$, and the measurement equation (18). An example is the normal distribution in Sections 5.2 and 5.3.

Among the linear score functions, the so-called *corrected score* (CS) function is of particular interest. It is given by special functions g and h. Suppose we can find functions $g = g(x,\beta)$ and $h = h(x,\beta)$ such that

$$\mathbb{E}\,[g|\xi] = v^{*-1}m^*_\beta \tag{22}$$

$$\mathbb{E}\,[h|\xi] = m^* v^{*-1}m^*_\beta. \tag{23}$$

Then, because of $\mathbb{E}\,(yg-h) = \mathbb{E}\;\mathbb{E}\,[(yg-h)|y,\xi] = \mathbb{E}\,(y-m^*)v^{*-1}m^*_\beta = 0$,

$$S_C := \begin{pmatrix} yg-h \\ l_\alpha \end{pmatrix}$$

is a linear score function within the class $\mathcal{L}^*$. It is called the corrected score function of the measurement error model. For this score function, Theorem 4.1 applies with S_C in place of S_L. In a number of important cases (like the Poisson, the gamma, and the Gaussian polynomial model) such functions g

and h can be found in closed form, see Sections 5.3 and 5.4. But there are also cases where g and h do not exist, Stefanski (1989).

5 Pre-estimation of nuisance parameters

5.1 General model

In the model of Section 4.1 with $\theta^\top = (\beta^\top,\ \alpha^\top)$, we could also define a modified QS estimator, which is based on a score function that instead of (16) consists of the two subvectors $(y-m)v^{-1}m_\beta$ and l_α, implying an estimator of α which uses the second subvector only. This means that α would be pre-estimated using only the data x_i, not the data y_i. We can then substitute the resulting estimator $\hat{\alpha}$ in the first subvector, $(y-m)v^{-1}m_\beta$, and use this to estimate β. We might call this estimator of β a QS estimator with pre-estimated nuisance parameters or simply pre-estimated QS estimator.

Such a two-step estimation procedure is, of course, simpler to apply than the one we propose, but according to Theorem 3.1 it is at most as efficient and often less efficient than the latter one.

There are, however, cases where pre-estimation of the nuisance parameter is in accordance with our QS approach and does not reduce the efficiency of QS. Suppose that

$$m_\alpha = Am_\beta \tag{24}$$

with some nonrandom matrix A, which may depend on θ (i.e., the α-part of m_θ is linearly related to the β-part). Then, first of all, the identifiability condition (6) simplifies to the condition that the two systems of random variables

$$[m_{\beta_i},\ i=1,\ldots,k] \quad \text{as well as} \quad [l_{\alpha_j},\ j=1,\ldots,d-k] \tag{25}$$

are both linearly independent. Furthermore, the quasi score function S_Q of (16) can be linearly transformed into an equivalent quasi score function S_Q^*, where the second subvector consists of l_α only:

$$S_Q^* = \begin{pmatrix} I & 0 \\ -A & I \end{pmatrix}^{-1} \cdot S_Q = \begin{pmatrix} (y-m)v^{-1}m_\beta \\ l_\alpha \end{pmatrix}. \tag{26}$$

The QS estimator $\hat{\theta}$ based on S_Q^* is the same as the one based on S_Q. Using S_Q^*, we see that α can be estimated independently of β from the second subvector of S_Q^* alone, i.e., it can be pre-estimated without reducing the efficiency of QS.

The QS estimator of α is the same as the LS estimator of α derived from (17). Therefore $\Sigma_L - \Sigma_Q$ is of the form

$$\Sigma_L - \Sigma_Q = \begin{pmatrix} \Sigma_L^{(\beta)} - \Sigma_Q^{(\beta)} & 0 \\ 0 & 0 \end{pmatrix} \tag{27}$$

and Theorem 4.1 reduces to

$$\begin{aligned}
&\operatorname{rank}(\Sigma_L^{(\beta)} - \Sigma_Q^{(\beta)}) + d \\
&= \operatorname{rank}\left[\begin{pmatrix} mg_i - h_i \\ vg_i \end{pmatrix}, \begin{pmatrix} 0 \\ m_{\beta_i} \end{pmatrix}, \begin{pmatrix} l_{\alpha_j} \\ 0 \end{pmatrix} \begin{array}{l} i = 1, \ldots, k \\ j = 1, \ldots, d-k \end{array}\right]
\end{aligned} \tag{28}$$

An immediate consequence of (28) is the following corollary, which corresponds to Corollary 3.1.

Corollary 5.1 *Suppose in a model with nuisance parameters as described in Section 4.1 condition* (24) *holds, then a sufficient condition for* $\Sigma_Q^{(\beta)} < \Sigma_L^{(\beta)}$ *is that the two systems of random variables*

$$\{m_{\beta_i},\, i = 1, \ldots, k\} \quad \textit{and} \quad \{mg_i - h_i,\, l_{\alpha_j},\, i = 1, \ldots, k,\, j = 1, \ldots, d-k\}$$

are both linearly independent.

For later use, we formulate an extension of Corollary 5.1, which deals with the case where only part of m_α is linearly related to m_β. It can be proved in the same way as Corollary 5.1.

Corollary 5.2 *Suppose in a model with nuisance parameters the nuisance parameter vector* α *is subdivided into two subvectors* $\alpha' \in \mathbb{R}^r$ *and* $\alpha'' \in \mathbb{R}^{(d-k-r)}$ *such that* $m_{\alpha''} = Am_\beta$ *with some nonrandom matrix* A *(which may depend on* θ*). Suppose further that there exists a nonrandom nonsingular* *square matrix B (which may depend on θ) such that $\tilde{l}_{\alpha''} := Bl_{\alpha''}$ is a function of x and α'' only. Let $\theta' = (\beta^\top, \alpha'^\top)^\top$. Then a sufficient condition for $\Sigma_Q^{(\theta')} < \Sigma_L^{(\theta')}$ is that the two systems of random variables*

$$\{m_{\beta_i}, m_{\alpha_j}, i = 1, \ldots, k, j = 1, \ldots, r\} \quad \textit{and}$$

$$\{mg_i - h_i, l_{\alpha_j}, i = 1, \ldots, k, j = 1, \ldots, d-k\}$$

are both linearly independent.

Just as with (26), the QS function S_Q is equivalent to

$$S_Q^* = \begin{pmatrix} (y-m)v^{-1}m_\beta \\ (y-m)v^{-1}m_{\alpha'} + l_{\alpha'} \\ \tilde{l}_{\alpha''} \end{pmatrix}$$

and $\tilde{l}_{\alpha''}$ can be used to pre-estimate α'', and $\hat{\alpha}''_Q = \hat{\alpha}''_L$.

In the following subsections, we study some special cases of the measurement error model of Section 4.2 with Gaussian regressor x, where the nuisance parameter $(\mu, \sigma)^\top$ or at least μ can be pre-estimated without loss of efficiency.

5.2 Pre-estimation of μ in a measurement error model

In this and the following subsections, we consider the mean-variance measurement error model of Section 4.2 with a Gaussian latent variable ξ:

$\xi \sim N(\mu_\xi, \sigma_\xi^2)$ with unknown μ_ξ and $\sigma_\xi^2 > 0$. In addition, we assume that the error free mean function m^* is a function of a linear predictor in ξ:

$$m^*(\xi, \beta) = \tilde{m}(\beta_0 + \beta_1 \xi), \quad \beta = (\beta_0, \beta_1)^\top. \tag{29}$$

In order to compute the mean function $m = \mathbb{E}(y|x)$, we need to find the conditional distribution of ξ given x. First note that $x \sim N(\mu, \sigma^2)$ with $\mu = \mu_\xi$, $\sigma^2 = \sigma_\xi^2 + \sigma_\delta^2$, and our nuisance parameter vector is $\alpha = (\mu, \sigma)^\top$. Furthermore,

$$\xi | x \sim N(\mu(x), \tau^2) \tag{30}$$

with

$$\mu(x) = Kx + (1 - K)\mu \tag{31}$$

$$\tau^2 = K\sigma_\delta^2, \tag{32}$$

where $K = \sigma_\xi^2/\sigma^2$ is the reliability ratio, $0 < K < 1$.

Because of (30) the mean function $m = m(x, \beta, \alpha)$ can now be computed as follows:

$$m = \mathbb{E}(m^*|x) = \mathbb{E}\left[\tilde{m}\{\beta_0 + \beta_1(Kx + (1 - K)\mu + \tau\gamma)\}|x\right], \tag{33}$$

where $\gamma \sim N(0, 1)$ and γ is independent of x. From (33) we have

$$m_{\beta_0} = \mathbb{E}[\tilde{m}'|x] \tag{34}$$

$$m_\mu = \beta_1(1 - K)\,\mathbb{E}[\tilde{m}'|x], \tag{35}$$

where $'$ denotes the derivative and $\tilde{m}'$ is short for $\tilde{m}'\{\beta_0+\beta_1(Kx+(1-K)\mu+\tau\gamma)\}$. Thus

$$m_\mu = \beta_1(1-K)m_{\beta_0}. \tag{36}$$

This corresponds to the equation $m_{\alpha''} = Am_\beta$ of Corollary 5.2 with $\alpha'' = \mu$, and hence μ can be pre-estimated. Indeed, S_Q is equivalent to

$$S_Q^* = \begin{pmatrix} (y-m)v^{-1}m_\beta \\ (y-m)v^{-1}m_\sigma + l_\sigma \\ l_\mu \end{pmatrix}, \tag{37}$$

where

$$l_\alpha = (l_\mu, l_\sigma)^\top = \left(\frac{x-\mu}{\sigma^2}, \frac{(x-\mu)^2}{\sigma^3} - \frac{1}{\sigma}\right)^\top. \tag{38}$$

Thus, for a linear predictor mean-variance measurement error model with Gaussian regressor, μ can be pre-estimated by using the score function l_μ, i.e., by solving the estimating equation $\sum_{i=1}^n \frac{x_i-\mu}{\sigma^2} = 0$ with the solution $\hat{\mu}_Q = \overline{x} := \frac{1}{n}\sum_{i=1}^n x_i$.

5.3 Pre-estimation of σ in a measurement error model

Continuing with the model of Section 5.2, we now derive conditions under which not only μ but also σ can be pre-estimated without loss of efficiency. Starting from (33), we find, in addition to (34) and (35),

$$m_{\beta_1} \;=\; (Kx + (1-K)\mu)\,\mathbb{E}\left[\tilde{m}'|x\right] + \beta_1\tau^2\,\mathbb{E}\left[\tilde{m}''|x\right], \tag{39}$$

$$m_{\sigma} \;=\; \beta_1 K_\sigma (x-\mu)\,\mathbb{E}\left[\tilde{m}'|x\right] + \beta_1^2\tau\tau_\sigma\,\mathbb{E}\left[\tilde{m}''|x\right]. \tag{40}$$

Here we used the identity

$$\mathbb{E}\left[\tilde{m}'(a+b\gamma)\gamma|x\right] = b\,\mathbb{E}\left[\tilde{m}''(a+b\gamma)|x\right],$$

where $a = a(x)$ and $b = b(x)$ are any functions of x. Indeed, by partial integration,

$$\begin{aligned}\mathbb{E}\left[\tilde{m}'(a+b\gamma)\gamma|x\right] &= \int \tilde{m}'(a+b\gamma)\gamma q(\gamma)d\gamma = b\int \tilde{m}''(a+b\gamma)q(\gamma)d\gamma \\ &= b\,\mathbb{E}\left[\tilde{m}''(a+b\gamma)|x\right],\end{aligned}$$

where $q(\gamma)$ is the density of the standard normal distribution.

Now suppose that the following differential equation holds for $\tilde{m}$:

$$\tilde{m}'' = c_0\tilde{m}' \tag{41}$$

with some constant c_0. Then by (34), (39), (40),and (41) and because $K > 0$,

$$m_\sigma = d_1 m_{\beta_0} + d_2 m_{\beta_1}$$

with some constants d_1 and d_2. Thus

$$m_\alpha = (m_\mu, m_\sigma)^\top = A(m_{\beta_0}, m_{\beta_1})^\top = Am_\beta$$

with some constant (2×2)-matrix A, and, according to Section 5.1, μ and σ can be pre-estimated. The QS estimates of μ and σ are simply the empirical mean and variance of the data x_i:

$$\hat{\mu}_Q = \overline{x}, \quad \hat{\sigma}_Q^2 = s_x^2 := \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

The linear differential equation (41) has the solution

$$\tilde{m}(t) = c_1 e^{c_0 t} + c_2. \tag{42}$$

An example is the log-linear Poisson model with measurement errors and Gaussian regressor. It is given by $y|\xi \sim \mathrm{Po}(\lambda)$ with $\lambda = \exp(\beta_0 + \beta_1\xi)$, and $x = \xi + \delta$. Here $m^* = \lambda$ and $\tilde{m}(t) = e^t$, which satisfies (42). For this model μ and σ can be pre-estimated. The exponential model $y|\xi \sim \mathrm{Exp}(\lambda)$ with $\lambda = \exp(\beta_0 + \beta_1\xi)$ is another example and so is the more general gamma model, Kukush *et al.*(2008).

As a further example we study the polynomial measurement error model in some detail in the next subsection, where again μ and σ can be pre-estimated, but for different reasons.

5.4 Polynomial measurement error model

The polynomial measurement error model of degree k is given by $y = \beta^\top\zeta + \varepsilon$ and $x = \xi + \delta$ with $\zeta = \zeta(\xi) = (1, \xi, \ldots, \xi^k)^\top$ and $\beta = (\beta_0, \beta_1 \ldots, \beta_k)^\top$. The variable ε is independent of ξ and δ, and all variables are Gaussian. In particular, as before, $x \sim N(\mu, \sigma^2)$, where the nuisance parameters μ and σ are supposed to be unknown.

Clearly, $m^*(\xi, \beta) = \beta^\top \zeta(\xi)$ and $v^* = \sigma_\varepsilon^2$. ($\sigma_\varepsilon^2$ is a dispersion parameter, which we can assume to be known when we are only interested in comparing the ACMs of $\hat{\beta}_C$ and $\hat{\beta}_Q$, see Remark 3). It follows that

$$m = \beta^\top \mathbb{E}\,(\zeta|x), \quad m_\beta = \mathbb{E}\,(\zeta|x), \quad m_\mu = (1-K)\beta^\top \mathbb{E}\,(\zeta'|x),$$

where ζ' is the derivative of ζ. Now there is a constant square matrix D such that

$$\zeta'(\xi) = D\zeta(\xi), \tag{43}$$

and so

$$m_\mu = (1-K)\beta^\top D m_\beta.$$

Therefore, according to Section 5.1, μ can be pre-estimated and $\hat{\mu}_Q = \overline{x}$.

Considering the nuisance parameter σ, we can show by similar arguments as those that led to (40) that

$$m_\sigma = \beta^\top \left(K_\sigma(x-\mu)\,\mathbb{E}\,[\zeta'|x] + \tau\tau_\sigma\,\mathbb{E}\,[\zeta''|x]\right).$$

We see that m_σ is a polynomial function of x of degree k, while the components of $\mathbb{E}\,(\zeta|x)$, i.e., $\mathbb{E}\,[\xi^j|x]$, are polynomials of degree j, $j = 0, \ldots, k$. Therefore m_σ is a linear combination of the components of $\mathbb{E}\,(\zeta|x)$ (with coefficients depending on μ, σ, and β). Thus $m_\sigma = b^\top m_\beta$ with some constant vector b. According to Section 5.1, this implies that not only μ but also σ can be pre-estimated, and $\hat{\sigma}_Q^2 = s_x^2 := \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2$, i.e., for the polynomial model, the estimator $\hat{\sigma}_Q^2$ is just the empirical variance.

We will now completely characterize all the cases where QS is strictly more efficient than CS and where it is just as efficient as CS.

Under known nuisance parameters, β is the only parameter to be estimated. The QS and CS functions are constructed as follows, Stefanski (1989) and Cheng and Schneeweiss (1998) and Shklyar *et al.* (2007):

$$S_Q = (y - m)v^{-1}m_\beta, \quad S_C = yt(x) - T(x)\beta, \tag{44}$$

where $t(x) = (t_0(x), \dots, t_k(x))^\top$ is such that $\mathbb{E}\,(t(x)|\xi) = \zeta$ and $T(x) \in \mathbb{R}^{(k+1)\times(k+1)}$ such that $T(x)_{ij} = t_{i+j}(x),\ i, j = 0, \dots, k$. The functions $t_j(x)$ are polynomials in x of degree j with leading term x^j, $j = 0, \dots, k$. The mean function $m = m(x, \beta)$ is given by $m = \beta^\top r(x)$, where $r(x) = r = (r_0, \dots, r_k)^\top$, $r_j = r_j(x)$ being a polynomial in x of degree j with leading term $K^j x^j$. The variance function $v = v(x, \beta, \sigma_\varepsilon^2)$ is a polynomial in x of degree $2s - 2$, except when $s = 0$ (where $v = \sigma_\varepsilon^2$). Here s is the true degree of the polynomial $\beta^\top\zeta$, i.e., $s = \max\{j : \ \beta_j \neq 0\}$; if $\beta = 0$, we set $s = 0$.

Under unknown nuisance parameters, the QS and CS functions have to be supplemented by the scores l_μ and l_σ for the nuisance parameters μ and σ. We have just seen that μ and σ can be pre-estimated on the basis of l_μ and l_σ alone. The β part of the CS and QS functions remain unchanged as in (44) except that μ and σ are replaced with their estimates.

The following theorem summarizes the various cases of an efficiency comparison between QS and CS in the polynomial model.

Theorem 5.1 *In a polynomial measurement error model of degree k with true degree s and with unknown nuisance parameters, the following relations regarding the ACMs of CS and QS hold:*

1. *if $s = 0$, then $\Sigma_Q = \Sigma_C$;*
2. *if $s = 1$, then $\operatorname{rank}(\Sigma_C^{(\beta)} - \Sigma_Q^{(\beta)}) = k - 1$;*
3. *if $s = 2$, then $\operatorname{rank}(\Sigma_C^{(\beta)} - \Sigma_Q^{(\beta)}) = k$;*
4. *if $s \geq 3$, then $\Sigma_Q^{(\beta)} < \Sigma_C^{(\beta)}$,*

where Σ_Q and Σ_C are the asymptotic covariance matrices of the QS and CS estimators of $(\mu, \sigma, \beta^\top)^\top$, respectively, and $\Sigma_Q^{(\beta)}$ and $\Sigma_C^{(\beta)}$ are the asymptotic covariance matrices of β only.

The *proof* is given in Kukush *et al.* (2006), where the case of known nuisance parameters is also treated.

Remark 4. In particular, in case $k = s = 1$, $\Sigma_Q^{(\beta)} = \Sigma_C^{(\beta)}$, which agrees with the fact that in a linear model under unknown nuisance parameters $\hat{\beta}_C = \hat{\beta}_Q$.

6 Conclusion

When one wants to estimate a parametric regression of y on x given by a conditional mean function $\mathbb{E}(y|x) = m(x, \theta)$ and supplemented by a conditional variance function $\mathbb{V}(y|x) = v(x, \theta)$, the *quasi-score* (QS) estimator is often the estimator of ones choice. In its traditonal form, it is based on the QS function $(y - m)v^{-1}m_\theta$, which is conditionally unbiased. But here we assume that the distribution of x with density $\rho(x, \theta)$ also depends on θ (or part of θ). We therefore extend the QS function above so that it incorporates the information given by $\rho(x, \theta)$. For simplicity, we call this extended QS function again the QS function. It is a member of a wide class of unconditionally unbiased linear-in-y estimating functions $S_L(x, y; \theta) = yg(x, \theta) - h(x, \theta)$, which we call *linear score* (LS) functions.

We prove that the QS estimator is most efficient within the class of LS estimators. We also state conditions under which QS is strictly more efficient than LS.

Linear score estimators appear naturally in the context of measurement error models. The so-called *corrected score* (CS) estimator is a linear score estimator. Thus for measurement error models we have as a corollary to our main result that QS is more efficient than CS.

The criteria developed in this paper can be applied to various special measurement error models, see Kukush *et al.*(2008). As a particular example, the polynomial measurement error model has been studied in the present paper.

7 Appendix

7.1 Lemmas

Lemma 7.1 *Let* $A, B \in \mathbb{R}^{d\times d}$. *Then*

$$\operatorname{def} \begin{pmatrix} B & A^\top \\ A & I_d \end{pmatrix} = \operatorname{def}\,(B - A^\top A),$$

where def (G) *denotes the defect of a matrix* G, *i.e., the dimension of its kernel* $\ker(G)$.

Proof. We have

$$\begin{pmatrix} x \\ y \end{pmatrix} \in \ker \begin{pmatrix} B & A^\top \\ A & I_d \end{pmatrix}$$

iff $Bx + A^\top y = 0$ and $y = -Ax$, which is equivalent to $x \in \ker(B - A^\top A)$ and $y = -Ax$. This implies that

$$\dim\ker \begin{pmatrix} B & A^\top \\ A & I_m \end{pmatrix} = \dim\ker(B - A^\top A).$$

Lemma 7.2 *Let* f *and* g *be two random vectors of the same dimension* d *such that* $\mathbb{E}\, gg^\top > 0$. *Consider the matrix* $M = \mathbb{E}\, ff^\top - \mathbb{E}\, fg^\top(\mathbb{E}\, gg^\top)^{-1}\,\mathbb{E}\, gf^\top$. *Then*

1) M is positive semi-definite. Moreover, M is the zero matrix if, and only if, $f = Hg$ a.s., with some nonrandom square matrix H;

2) $\operatorname{rank} M = \operatorname{rank}[f_i, g_i,\ i = 1, \dots, d] - d$, *where the latter rank is the maximum number of linearly independent random variables in the set* $\{f_i, g_i,\ i = 1, \dots, d\}$.

Proof. 1) To prove the first statement, let

$$e = f - \mathbb{E}\, fg^\top (\mathbb{E}\, gg^\top)^{-1} g.$$

Then $\mathbb{E}\, ee^\top = M \geq 0$, and $M = 0$ iff $e = 0$, that is, iff $f = Hg$ with some nonrandom square matrix H.

2) To prove the second statement, let

$$F = \mathbb{E}\, ff^\top, \quad \tilde{g} = (\mathbb{E}\, gg^\top)^{-1/2} g, \quad A = \mathbb{E}\, \tilde{g} f^\top.$$

Then $M = (F - A^\top A)$ and, by Lemma 7.1,

$$\operatorname{rank} M = \operatorname{rank}[F - A^\top A] = \operatorname{rank} \begin{bmatrix} F & A^\top \\ A & I_d \end{bmatrix} - d.$$

The latter rank is the rank of the moment matrix of the random vector $[f_1, \dots, f_d, \tilde{g}_1, \dots, \tilde{g}_d]$. It is therefore equal to the rank of this vector. But due to the definition of $\tilde{g}$,

$$\operatorname{rank}[f_1, \dots, f_d, \tilde{g}_1, \dots, \tilde{g}_d] = \operatorname{rank}[f_1, \dots, f_d, g_1, \dots, g_d].$$

7.2 Proof of Theorem 3.1

We apply the first statement of Lemma 7.2 to the random vectors $g = S_Q$ and $f = S_L$. We have

$$\mathbb{E}\, ff^\top - \mathbb{E}\, fg^\top \left(\mathbb{E}\, gg^\top\right)^{-1} \mathbb{E}\, gf^\top \geq 0.$$

Due to (8) and (9) this is equivalent to $\Sigma_L - \Sigma_Q \geq 0$. Equality between Σ_L and Σ_Q for all θ holds iff for some nonrandom square matrix $H = H(\theta)$, $f = Hg$, i.e.,

$$\forall\, \theta: \quad S_L = H(\theta) S_Q \quad \text{a.s.}$$

Because $\mathbb{E}\, S_L S_Q^\top$ is nonsingular, H is nonsingular as well. Then the equation for $\hat{\theta}_L$, $\sum_{i=1}^n S_L(x_i, y_i; \theta) = 0$, is equivalent to $\sum_{i=1}^n H(\theta) S_Q(x_i, y_i; \theta) = 0$, which is a.s. equivalent to the equation for $\hat{\theta}_Q$, $\sum_{i=1}^n S_Q(x_i, y_i; \theta) = 0$. Thus $\hat{\theta}_L = \hat{\theta}_Q$ a.s.

Vice versa, if $\hat{\theta}_L = \hat{\theta}_Q$ a.s., then $\Sigma_L = \Sigma_Q$ for all θ.

7.3 Proof of Theorem 3.2

We apply the second statement of Lemma 7.2 with $g = S_Q$, $f = S_L$. By (8) and (9),

$$\begin{aligned} \operatorname{rank}(\Sigma_L - \Sigma_Q) &= \operatorname{rank} M = \operatorname{rank}\left[(S_L)_i, (S_Q)_i,\ i = 1, \ldots, d\right] - d \\ &= d - \operatorname{def}\left[(S_L)_i, (S_Q)_i,\ i = 1, \ldots, d\right]. \end{aligned} \tag{45}$$

To find the defect, we form a linear combination of the components of S_L and S_Q, see (2) and (5), which is supposed to equal zero a.s.:

$$c_1^\top gy - c_1^\top h + \frac{c_2^\top m_\theta}{v}(y-m) + c_2^\top l_\theta = 0 \quad \text{a.s.}$$

or

$$\left(c_1^\top g + \frac{c_2^\top m_\theta}{v}\right) y = c_1^\top h + \frac{c_2^\top m m_\theta}{v} - c_2^\top l_\theta \quad \text{a.s.} \tag{46}$$

The defect in (45) is equal to the maximum number of linearly independent vectors $(c_1^\top, c_2^\top)^\top$ which satisfy (46). But (46) is equivalent to

$$c_1^\top g + \frac{c_2^\top m_\theta}{v} = 0 \quad \text{and} \quad c_1^\top h + \frac{c_2^\top m m_\theta}{v} - c_2^\top l_\theta = 0 \quad \text{a.s.} \tag{47}$$

Indeed in general, $a(x)y = b(x)$ a.s. implies $a^2(x)v(x) = 0$ and therefore $a(x) = 0$ because by assumption $v(x) > 0$. Now, (47) is equivalent to

$$c_1^\top v g + c_2^\top m_\theta = 0, \quad c_1^\top (mg - h) + c_2^\top l_\theta = 0 \quad \text{a.s.}$$

Thus

$$\begin{aligned} &\operatorname{def}\ [(S_L)_i, (S_Q)_i,\ i = 1, \ldots, d] \\ &= \operatorname{def} \left[\begin{pmatrix} mg_i - h_i \\ vg_i \end{pmatrix}, \begin{pmatrix} l_{\theta_i} \\ m_{\theta_i} \end{pmatrix}, \quad i = 1, \ldots, d \right], \end{aligned}$$

and (14) follows from (45).

7.4 Proof of Corollary 3.1

Suppose the random variables (15) are linearly independent. Then because of the identifiability condition (6), the random vectors in (14) are also linearly independent. Indeed, for any constant vectors a and $b \in \mathbb{R}^d$, the system of equations

$$\begin{aligned} a^\top(mg - h) + b^\top l_\theta &= 0 \\ a^\top vg + b^\top m_\theta &= 0 \end{aligned}$$

implies first $a = 0$ because of the independence of the random variables in (15) and then $b = 0$ because of (6). According to Theorem 3.2, it follows that $\Sigma_Q < \Sigma_L$.

7.5 Consistency and asymptotic normality of $\hat{\theta}_L$

Lemma 7.3 *Consider model* (1) *of the Introduction and assume the following conditions.*

1. *The parameter set Θ is a convex compact set in $\mathbb{R}^d$, and the true parameter value θ lies in Θ°, the interior of Θ.*

2. *The functions g,h: $\mathbb{R} \times U \to \mathbb{R}^d$ of* (2) *are Borel measurable, where U is a neighborhood of Θ, moreover, $g(x,\cdot)$ and $h(x,\cdot)$ belong to $C^2(U)$ a.s.*

3. $\mathbb{E}\,|m(x,\theta)| \cdot \|g(x,t)\| < \infty$, *for all* $\theta \in \Theta^\circ$ *and* $t \in \Theta$; $\mathbb{E}\, m^2(x,\theta) \cdot \|g(x,\theta)\|^2 < \infty$, *for all* $\theta \in \Theta^\circ$.

4. $\mathbb{E}\,|m(x,\theta)| \cdot \sup_{t\in\Theta} \left| D_t^{(j)} g_k(x,t) \right| < \infty$, *for all* $\theta \in \Theta^\circ$, $j = 1, 2$, $k = 1, \ldots, d$, *and* $\mathbb{E} \sup_{t\in\Theta} \left| D_t^{(j)} h_k(x,t) \right| < \infty$, *for all* $j = 1, 2$, $k = 1, \ldots, d$, *where* g_k *and* h_k *are the* k*'th components of* g *and* h, *and* $D_t^{(j)} g_k$, $D_t^{(j)} h_k$ *denote the partial derivatives of order* j *with respect to the variable* t *of the functions* g_k, h_k, *respectively.*

5. *For any* $\theta \in \Theta^\circ$ *the equality* $\mathbb{E}\,(m(x,\theta)g(x,t) - h(x,t)) = 0$, $t \in \Theta$, *holds true if, and only if,* $t = \theta$.

6. *The matrices* $A_L = -\mathbb{E}\, S_{L\theta}$ *and* $B_L = \mathbb{E}\, S_L S_L^\top$ *are nonsingular.*

Then:

a) There exists a Borel measurable function $\hat{\theta}_L$ *of the observations* (x_i, y_i) *such that* $\sum_{i=1}^{n} S_L(x_i, y_i, \hat{\theta}_L) = 0$ *a.s. for all* $n \geq n_0(\omega)$.

b) $\hat{\theta}_L \to \theta$ *a.s., as* $n \to \infty$.

c) $\sqrt{n}(\hat{\theta}_L - \theta)$ *converges in distribution to* $N(0, \Sigma_L)$ *with* $\Sigma_L = A_L^{-1} B_L A_L^{-\top}$.

Remarks on the proof. The existence of a solution to the equation $\sum_{i=1}^{n} S_L(x_i, y_i, t) = 0$, $t \in \Theta$, for all $n \geq n_0(\omega)$ follows from Heyde (1997). Due to Pfanzagl (1969), it is possible to select the solution in a measurable way, and statement a) follows. Statements b) and c) can be proved based on the theory of estimating equations.

References

[1] Armstrong, B. (1985), Measurement error in the generalized linear model. *Comm. Statist. Simulation Comput.* **14**, 529-544.

[2] Carroll, R. J., Ruppert, D., Stefanski, L.A., and Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models.* Chapman and Hall, London.

[3] Cheng, C.-L. and Van Ness, J. W., (1999), *Statistical Regression with Measurement Error.* Arnold, London.

[4] Cheng, C.-L. and Schneeweiss, H. (1998), Polynomial regression with errors in the variables. *J. Roy. Statist. Soc. Ser. B* **60**, 189 - 199.

[5] Fuller, W. A. (1987), *Measurement Error Models.* Wiley, New York.

[6] Heyde, C. C. (1997), *Quasi-Likelihood And Its Application.* Springer, New York.

[7] Kukush, A. and Schneeweiss H. (2005), Comparing different estimators in a nonlinear measurement error model. I. *Math. Methods Statist.* **14**, 53-79.

[8] Kukush, A. and Schneeweiss H. (2006), *Asymptotic optimality of the quasi-score estimator in a class of linear score estimators*, Discussion Paper 477, SFB 386, Universität München.

[9] Kukush, A., Malenko A., and Schneeweiss H. (2006),*Optimality of the quasi-score estimator in a mean-variance model with applications to measurement error models*. Discussion Paper 494, SFB 386, University of Munich.

[10] Kukush, A., Malenko A., and Schneeweiss H. (2007), Comparing the efficiency of estimates in concrete errors-in-variables models under unknown nuisance parameters. *Theory of Stochastic Processes* **13** (29), 69-81.

[11] Nakamura, T. (1990), Corrected score function for errors-in-variables models. *Biometrika* **77**, 127-137.

[12] Pfanzagl, G. (1969), On the measurability and consistency of minimum contrast estimates. *Metrika* **14**, 249-273.

[13] Schervish, M.J. (1995), *Theory of Statistics*. Springer, New York.

[14] Shklyar, S., Schneeweiss, H., and Kukush, A. (2007), Quasi Score is more efficient than Corrected Score in a polynomial measurement error model. *Metrika* **65**, 275-295.

[15] Stefanski, L. (1989), Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Comm. Statist. Theory Methods* **18**, 4335-4358.

[16] Wedderburn, R.W.M. (1974), Quasi likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**, 439-447.

Addresses:

Alexander Kukush: Department of Mechanics and Mathematics, Kiev National Taras Shevchenko University, Volodymyrska str. 60, 01033 Kiev, Ukraine. E-mail: alexander_kukush@univ.kiev.ua

Andrii Malenko: Department of Mechanics and Mathematics, Kiev National Taras Shevchenko University, Volodymyrska str. 60, 01033 Kiev, Ukraine. E-mail: exipilis@yandex.ru

Hans Schneeweiss: Department of Statistics, University of Munich, Akademiestr. 1, 80799 Munich, Germany. E-mail: schneew@stat.uni-muenchen.de