ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Kukush, Alexander; Malenko, Andrii; Schneeweiss, Hans

Working Paper

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

Discussion Paper, No. 494

Provided in Cooperation with:
Collaborative Research Center (SFB) 386: Statistical Analysis of discrete structures - Applications in Biometrics and Econometrics, University of Munich (LMU)

Suggested Citation: Kukush, Alexander; Malenko, Andrii; Schneeweiss, Hans (2006) : Optimality of the quasi-score estimator in a mean-variance model with applications to measurement error models, Discussion Paper, No. 494, Ludwig-Maximilians-Universität München, Sonderforschungsbereich 386 - Statistische Analyse diskreter Strukturen, München, https://doi.org/10.5282/ubm/epub.1862

This Version is available at:
https://hdl.handle.net/10419/31154

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

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.We also study a number of special measurement error models in greater detail.

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.

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:

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

Here θ is an unknown d-dimensional parameter vector to be estimated. The parameter θ belongs to an open 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.* (1995). 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. 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.* (1995), 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. 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.* (1995). 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. We study such models, among which the polynomial model is the most prominent one. Other models are the Poisson, the Gamma, and the logit model.

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 measurement error model was considered under known nuisance parameters and the efficiency of the QS estimator (in the usual sense) was compared to

the LS estimator, thereby extending a similar result of Heyde (1997). In the present paper, we study the case of unknown nuisance parameters. For that case, Schneeweiss (2005) dealt with the computation of the asymptotic covariance matrix of the estimator in a polynomial measurement error model. Shklyar (2006) proved for the log-linear Poisson model with errors in the variables and nuisance parameters that the QS estimator is more efficient than the CS estimator, where the unknown mean and variance of the latent variable were pre-estimated by the empirical mean and empirical variance. He used a direct computation of the asymptotic covariance matrices. We can derive this result by reasoning from general principles.

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.* (1995). 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 $\mathbf{E}$ to denote the expectation of random values, vectors, and matrices and $\mathbf{V}$ 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, 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 $B - A$ is positive semidefinite, respectively.

The paper is organized as follows. In Section 2, we introduce the class of linear unbiased scores. In Section 3, we introduce our new QS estimator. Section 4 contains general results on the comparison of QS and LS estimators. Section 5 comments on the case of an unknown dispersion parameter. In Section 6, 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 7 deals with cases where pre-estimation of the nuisance parameters is possible. Section 8 studies strict optimality of QS in various measurement eror models. In Section 9, we present some examples where QS and CS are equally efficient under restricted parameter points. Section 10 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 typical example of such an estimating function is the quasi score function to be introduced in Section 3. It is a member of a general class of estimating functions. Let $\mathcal{L}$ be the class 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 $\forall\, \theta \in \Theta: \ \mathbf{E}\, S_L(x, y; \theta) = 0$. 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.

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 $\hat{\theta}_L$ 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),

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

A_L is supposed to be nonsingular (identifiability condition).

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

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

We want to derive an identity from (4), which we will need later on. Equation (4) can be written as

$$\int_{\mathbb{R}} (mg - h)\rho \, d\nu = 0,$$

Differentiating with respect to θ yields

$$\int_{\mathbb{R}} (mg - h)_\theta \cdot \rho \, d\nu + \int_{\mathbb{R}} (mg - h)\rho_\theta^{\top} \, d\nu = 0.$$

We thus have the identity

$$\mathbf{E}\,(mg - h)_\theta + \mathbf{E}\,(mg - h) l_\theta^{\top} = 0, \tag{5}$$

where $l := \log \rho(x, \theta)$.

3 Heuristic derivation of the QS estimator

We want to construct a special linear score function, which we will call the quasi-score function and which will play a fundamental role in the following. It is based on the mean and variance functions (1). To this purpose, we suppose that $y|x \sim N(m, v)$. We need this assumption only for the heuristic derivation of the quasi-score function and will discard it later on after it has served its purpose. The joint density of x and y then equals

$$f(x, y) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{(y-m)^2}{2v}} \rho,$$

and the log-likelihood function is

$$L(x, y) := \log f(x, y) = -\frac{1}{2} \log v - \frac{(y - m)^2}{2v} + l + const.$$

To construct a score function for θ, we compute

$$\frac{\partial L}{\partial \theta} = \frac{(y - m) m_\theta}{v} + l_\theta - \frac{1}{2} \frac{\partial}{\partial \theta} \log v - \frac{(y - m)^2}{2} \frac{\partial}{\partial \theta} \frac{1}{v}.$$

Note that the expectation of the sum of the last two terms is zero. So by omitting these two terms (and thus neglecting the dependence of v on θ), we derive a quasi score function S_Q as follows:

$$S_Q(x, y; \theta) := \frac{(y - m)m_\theta}{v} + l_\theta. \tag{6}$$

It differs from the usual quasi-score function as exemplified, e.g., in Heyde (1997), by the term l_θ. It is obviously unbiased even without the provisional assumption of normality, which we will discard from now on. We assume that

$$\mathbf{E}\, S_Q S_Q^\top = \mathbf{E}\, v^{-1} m_\theta m_\theta^\top + \mathbf{E}\, l_\theta l_\theta^\top$$

is positive definite. This (identifiability) condition is equivalent to the condition that the d two-dimensional random vectors

$$\begin{pmatrix} l_{\theta_i} \\ m_{\theta_i} \end{pmatrix}, \quad i = 1, \ldots, d, \tag{7}$$

are linearly independent.

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

$$\sum_{i=1}^{n} S_Q(x_i, y_i, \hat{\theta}_Q) = 0. \tag{8}$$

The quasi-score function (6) belongs to $\mathcal{L}$, therefore the estimator $\hat{\theta}_Q$ is consistent and asymptotically normal under regularity conditions.

4 Comparison of QS to LS

We want to derive alternative formulas for the ACMs of the LS estimator $\hat{\theta}_L$ and of the QS estimator $\hat{\theta}_Q$. These formulas will be useful in comparing Σ_Q to Σ_L. To this purpose, we show that

$$\mathbf{E}\, S_{L\theta} = -\,\mathbf{E}\, S_L S_Q^\top. \tag{9}$$

We first have

$$\mathbf{E}\, S_{L\theta} = \mathbf{E}\,(m g_\theta - h_\theta). \tag{10}$$

On the other hand,

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

From (10), (11), and (5) we obtain

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

which yields (9).

Now, (9) implies that the ACM of $\hat{\theta}_L$, given by (3), can be written as

$$\Sigma_L = \left(\mathbf{E}\, S_L S_Q^\top\right)^{-1} \mathbf{E}\, S_L S_L^\top \left(\mathbf{E}\, S_L S_Q^\top\right)^{-\top}. \tag{12}$$

As S_Q belongs to $\mathcal{L}$, we can apply (12) and obtain for the ACM of $\hat{\theta}_Q$

$$\Sigma_Q = \left(\mathbf{E}\, S_Q S_Q^\top\right)^{-1}. \tag{13}$$

We now can state the following theorems.

Theorem 4.1 (Optimality of QS) *Let S_L be a score function from the class $\mathcal{L}$ and S_Q be the quasi-score function (6). 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.

Theorem 4.2 (Strict Optimality of QS) *Under the conditions of Theorem 4.1*

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

where $\operatorname{rank}[\cdot]$ *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

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

then

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

As an immediate consequence, we have the following corollary:

Corollary 4.1 *A sufficient condition for $\Sigma_Q < \Sigma_L$ is that the random variables*

$$\{(mg - h)_i,\ i = 1, \ldots, d,\ l_{\theta_j}, j \in B_\theta\} \tag{15}$$

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 1. The inequality $\Sigma_Q \leq \Sigma_L$ of Theorem 4.1 can also be obtained as a direct consequence of identity (9) and Heyde's (1997) criterion for asymptotic optimality.

5 Estimation of an unknown dispersion parameter

In this section we assume additionally that the conditional variance depends also on an unknown parameter $\varphi \in \mathbb{R}^+$, $v = v(x, \theta, \varphi)$, while $m(x, \theta)$ does not depend on φ, and the distribution of x does not depend on φ either.

This holds, e.g., when the conditional distribution of $y|x$ has a density belonging to an exponential family with dispersion parameter $\varphi > 0$.

We introduce an unbiased score function for φ,

$$S^{(\varphi)} = (y - m)^2 - v.$$

(Another option could be the function $y^2 - m^2 - v$). The estimators $\hat{\theta}_Q$, $\hat{\varphi}_Q$ are generated by the compound score function $S_Q^{(\theta,\varphi)} = \left(S_Q^{(\theta)\top}, S^{(\varphi)}\right)^\top$, where $S_Q^{(\theta)}$ is the same as S_Q given in (6). Similarly for the score function $S_L = S_L^{(\theta)}$ given in (2) with $g = g(x, \theta, \varphi)$ and

$h = h(x, \theta, \varphi)$, we introduce the compound score function $S_L^{(\theta,\varphi)} = \left(S_L^{(\theta)\top}, S^{(\varphi)}\right)^\top$ and the corresponding estimators $\hat{\theta}_L$, $\hat{\varphi}_L$. By the sandwich formula, the ACM of $(\hat{\theta}_L, \hat{\varphi}_L)^\top$ equals

$$\Sigma_L^{(\theta,\varphi)} = \left(A_L^{(\theta,\varphi)}\right)^{-1} B_L^{(\theta,\varphi)} \left(A_L^{(\theta,\varphi)}\right)^{-\top} \tag{16}$$

Since m does not depend on φ, we have

$$\mathbf{E}\,\frac{\partial}{\partial\varphi} S_L^{(\theta)} = \mathbf{E}\,(yg_\varphi - h_\varphi) = \mathbf{E}\,(mg - h)_\varphi = 0.$$

The last equation follows from the identity $\mathbf{E}_\theta(mg - h) = 0$ for all φ. Thus the matrix

$$A_L^{(\theta,\varphi)} = -\,\mathbf{E}\,\frac{\partial}{\partial(\theta,\varphi)} S_L^{(\theta,\varphi)} =: \begin{pmatrix} A_L^{(\theta)} & 0 \\ * & A_L^{(\varphi)} \end{pmatrix}. \tag{17}$$

is lower block-triangular. In addition, we have

$$B_L^{(\theta,\varphi)} = \mathbf{E}\, S_L^{(\theta,\varphi)} S_L^{(\theta,\varphi)\top} =: \begin{pmatrix} B_L^{(\theta)} & * \\ * & B_L^{(\varphi)} \end{pmatrix}, \tag{18}$$

$$\Sigma_L^{(\theta,\varphi)} =: \begin{pmatrix} \Sigma_L^{(\theta)} & * \\ * & \Sigma_L^{(\varphi)} \end{pmatrix}.$$

From (16), (17), and (18) we obtain that

$$\Sigma_L^{(\theta)} = \left(A_L^{(\theta)}\right)^{-1} B_L^{(\theta)} \left(A_L^{(\theta)}\right)^{-\top},$$

which coincides with the ACM (3) of $\hat{\theta}_L$ under known φ. Therefore the optimality properties of $\hat{\theta}_Q$, which were established in the previous section, still hold true under unknown φ. For that reason, we assume in the subsequent sections that φ is known.

Note, however, that this result depends on the fact that the same $S^{(\varphi)}$ was chosen both for QS and for LS. For an example of a different choice see Shklyar *et al.* (2006).

6 Estimation of a nuisance parameter in a regression model

6.1 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 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 (6) takes the form

$$S_Q = \begin{pmatrix} (y - m)v^{-1}m_\beta \\ (y - m)v^{-1}m_\alpha + l_\alpha \end{pmatrix}. \tag{19}$$

Such a model arises naturally in the context of measurement error models, see Section 6.2. All the previous results hold true.

It seems reasonable in this case to also restrict the competing linear score estimator to an appropriate subclass of $\mathcal{L}$. Assume that S_L is of the form

$$S_L = \begin{pmatrix} yg(x,\beta) - h(x,\beta) \\ \{\log \rho(x,\alpha)\}_\alpha \end{pmatrix}, \tag{20}$$

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 $\mathbf{E}\, l_\alpha = 0$ anyway. Note that S_Q is not a member of this restricted class. Nevertheless, we can still apply Theorems 4.1 and 4.2 to compare Σ_L to Σ_Q. In particular, Theorem 4.2 takes the form:

Theorem 6.1 *If $\theta = (\beta^\top, \alpha^\top)^\top$ and $\rho = \rho(x,\alpha)$ and S_L is of the form* (20)*, then*

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

6.2 Measurement error model

The model of Subsection 6.1 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

$$x = \xi + \delta, \tag{21}$$

where the measurement error δ is independent of ξ and y and $\mathbf{E}\,\delta = 0$. 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:

$$\mathbf{E}\,(y|\xi) = m^*(\xi,\beta), \quad \mathbf{V}(y|\xi) = v^*(\xi,\beta). \tag{22}$$

Note that m^* and v^* do not depend on α. From (22) we can derive the conditional mean and variance functions of y given x:

$$m(x,\beta,\alpha) := \mathbf{E}\,(y|x) = \mathbf{E}\,[m^*(\xi,\beta)|x] \tag{23}$$

$$v(x,\beta,\alpha) := \mathbf{V}(y|x) = \mathbf{E}\,[v^*(\xi,\beta)|x] + \mathbf{V}[m^*(\xi,\beta)|x]. \tag{24}$$

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 (21). An important special case for $\rho(\xi,\alpha)$ is the normal distribution $\xi \sim N(\mu_\xi, \sigma_\xi^2)$, $\sigma_\xi^2 > 0$. In this case, $x \sim N(\mu,\sigma^2)$, $\mu = \mu_\xi$, $\sigma^2 = \sigma_\xi^2 + \sigma_\delta^2$, $\alpha = (\mu,\sigma)^\top$, and $\xi|x \sim N(\mu(x),\tau^2)$ with

$$\mu(x) = Kx + (1-K)\mu \tag{25}$$

$$\tau^2 = K\sigma_\delta^2, \tag{26}$$

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

The subvector l_α in the score function S_Q takes the special form

$$l_\alpha = (l_\mu, l_\sigma)^\top = \left(\frac{x-\mu}{\sigma^2}, \frac{(x-\mu)^2}{\sigma^3} - \frac{1}{\sigma} \right)^\top. \tag{27}$$

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

$$\begin{aligned} \mathbf{E}\,[g|\xi] &= v^{*-1}m^*_\beta \qquad (28)\\ \mathbf{E}\,[h|\xi] &= m^* v^{*-1}m^*_\beta. \qquad (29) \end{aligned}$$

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

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

is a linear score function within the class $\mathcal{L}$. It is called the corrected score function of the measurement error model. In a number of important cases such functions g and h can be found in closed form. But there are also cases where g and h do not exist, Stefanski (1989).

6.3 Regression model without measurement errors

In the typical (linear or nonlinear) regression model without measurement errors we have a complete separation of $\theta^\top = (\beta^\top,\ \alpha^\top)$, β only appearing in the mean and variance functions and α appearing only in the density function of x: $m = m(x, \beta), v = v(x, \beta), \rho = \rho(x, \alpha)$. The identifiability condition (7) simplifies to the condition that the two systems of random variables

$$[m_{\beta_i},\ i = 1, \ldots, k] \quad \text{as well as} \quad [l_{\alpha_j},\ j = 1, \ldots, d - k] \qquad (30)$$

are both linearly independent. The linear score functions and the quasi score function, which is a special case of the former, take the form

$$S_L = \begin{pmatrix} yg - h \\ l_\alpha \end{pmatrix}, \quad S_Q = \begin{pmatrix} (y - m)v^{-1}m_\beta \\ l_\alpha \end{pmatrix}, \qquad (31)$$

where g and h are of dimension k and are functions of x and β only. We have

$$\begin{aligned} \mathbf{E}\,S_{L\theta} &= \mathbf{E}\begin{pmatrix} mg_\beta - h_\beta & 0 \\ 0 & l_{\alpha\alpha} \end{pmatrix}\\ \mathbf{E}\,S_L S_Q^\top &= \mathbf{E}\begin{pmatrix} gm_\beta^\top & (mg - h)l_\alpha^\top \\ 0 & l_\alpha l_\alpha^\top \end{pmatrix}. \end{aligned}$$

By (9), this implies the identity

$$\mathbf{E}\,(mg - h)l_\alpha^\top = 0, \qquad (32)$$

which will be useful in comparing Σ_L to Σ_Q. As the ACM of $\hat{\alpha}$ is the same submatrix for all Σ_L and since $\Sigma_Q \leq \Sigma_L$, the difference of Σ_L and Σ_Q takes the form

$$\Sigma_L - \Sigma_Q = \begin{pmatrix} \Sigma_L^{(\beta)} - \Sigma_Q^{(\beta)} & 0 \\ 0 & 0 \end{pmatrix}. \qquad (33)$$

Theorems 4.1 and 4.2 now reduce to

Theorem 6.2 *In a regression model with β the regression parameter and α the nuisance parameter describing the regressor distribution, the following holds:*

a) For any linear score function (31), $\Sigma_Q^{(\beta)} \leq \Sigma_L^{(\beta)}$.

b)

$$\text{rank}\left(\Sigma_L^{(\beta)} - \Sigma_Q^{(\beta)}\right) = \text{rank}\left[\left(\begin{array}{c} mg_i - h_i \\ vg_i \end{array}\right), \left(\begin{array}{c} 0 \\ m_{\beta_i} \end{array}\right), \quad i = 1, \ldots, k\right] - k. \tag{34}$$

In particular, if $\text{span}\{vg_i,\ i = 1, \ldots, k\} \cap \text{span}\{m_{\beta_i},\ i = 1, \ldots, k\} = \{0\}$, *then*

$$\text{rank}\left(\Sigma_L^{(\beta)} - \Sigma_Q^{(\beta)}\right) = \text{rank}\left[\left(\begin{array}{c} h_i \\ g_i \end{array}\right),\ i = 1, \ldots, k\right]. \tag{35}$$

c) If the components $\{(mg - h)_i,\ i = 1, \ldots, k\}$ *are linearly independent, then* $\Sigma_Q^{(\beta)} < \Sigma_L^{(\beta)}$.

Clearly, Theorem 6.2 also holds if the distribution of the regressor is completely known, so that there is no nuisance parameter α.

7 Pre-estimation of nuisance parameters

7.1 General model

In the model of Section 6.1 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 (19) 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 4.1 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

$$m_\alpha = Am_\beta \tag{36}$$

with some nonrandom matrix A (which may depend on θ). Then, first of all, the identifiability condition (7) simplifies to the same condition as in Section 6.3, see (30). Furthermore, the quasi score function S_Q of (19) can be linearly transformed into an equivalent quasi score function S_Q^*, where the second subvector consists of l_α only:

$$S_Q^* = \left(\begin{array}{cc} I & 0 \\ -A & I \end{array}\right)^{-1} S_Q = \left(\begin{array}{c} (y - m)v^{-1}m_\beta \\ l_\alpha \end{array}\right).$$

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. Therefore $\Sigma_L - \Sigma_Q$ is of the form (33) and Theorem 6.1 reduces to

$$\begin{aligned}&\operatorname{rank}(\Sigma_L^{(\beta)} - \Sigma_Q^{(\beta)}) + d \\ &= \operatorname{rank}\left[\begin{pmatrix} mg_i - h_i \\ vg_i \end{pmatrix}, \begin{pmatrix} 0 \\ m_{\beta_i} \end{pmatrix}, i = 1, \ldots, k, \begin{pmatrix} l_{\alpha_j} \\ 0 \end{pmatrix}, j = 1, \ldots, d-k\right]. \quad (37)\end{aligned}$$

An immediate consequence of (37) is the following corollary, which corresponds to Corollary 4.1.

Corollary 7.1 *Suppose in a model with nuisance parameters as described in Section 6.1 condition* (36) *holds, then a sufficient condition for* $\Sigma_Q^{(\beta)} < \Sigma_L^{(\beta)}$ *is that the two systems of random variables*

$$\{m_{\beta_i}, i = 1, \ldots, k\} \quad \text{and} \quad \{(mg - h)_i, i = 1, \ldots, k,\, l_{\alpha_j}, j = 1, \ldots, d-k\}$$

are both linearly independent.

For later use, we formulate an extension of Corollary 7.1, 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 7.1.

Corollary 7.2 *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*. Suppose further that there exists a nonrandom nonsingular square matrix* B *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 system of random variables*

$$\{m_{\beta_i}, i = 1, \ldots, k, m_{\alpha_j}, j = 1, \ldots, r\} \quad \text{and} \quad \{(mg - h)_i, i = 1, \ldots, k,\, l_{\alpha_j}, j = 1, \ldots, d-k\}$$

are both linearly independent.

Note that $\tilde{l}_{\alpha''}$ 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 6.2, where the nuisance parameter $(\mu, \sigma)^\top$ or at least μ can be pre-estimated without loss of efficiency.

7.2 Pre-estimation of μ

Consider the mean-variance measurement error model of Section 6.2 and assume that the error free mean function m^* is a function of a linear predictor in ξ:

$$m^*(\xi, \beta) = \tilde{m}(\beta_0 + \beta_1 \xi), \quad \beta = (\beta_0, \beta_1)^\top. \quad (38)$$

The mean function $m = m(x, \beta, \alpha)$ can then be computed as follows:

$$m = \mathbf{E}(m^*|x) = \mathbf{E}\left[\tilde{m}\{\beta_0 + \beta_1(Kx + (1-K)\mu + \tau\gamma)\}|x\right], \quad (39)$$

where $\gamma \sim N(0,1)$ and γ is independent of x. From (39) we have

$$\begin{aligned} m_{\beta_0} &= \mathbf{E}[\tilde{m}'|x] \quad (40) \\ m_\mu &= \beta_1(1-K)\,\mathbf{E}[\tilde{m}'|x], \quad (41)\end{aligned}$$

where $'$ denotes the derivative. Thus

$$m_\mu = \beta_1(1-K)m_{\beta_0}. \tag{42}$$

According to Section 7.1, S_Q is thus equivalent to

$$S_Q^* = \begin{pmatrix} (y-m)v^{-1}m_\beta \\ (y-m)v^{-1}m_\sigma + l_\sigma \\ l_\mu \end{pmatrix}. \tag{43}$$

Thus 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 $\hat{\mu}_Q = \overline{x} := \frac{1}{n}\sum_{i=1}^n x_i$. Thus for such models the QS estimator of μ is the empirical mean of x.

7.3 Pre-estimation of σ

Continuing with the model of Section 7.2, we now derive conditions under which not only μ but also σ can be pre-estimated without loss of efficiency.

Starting from (39), we find, in addition to (40) and (41),

$$m_{\beta_1} = (Kx + (1-K)\mu)\,\mathbf{E}\left[\tilde{m}'|x\right] + \tau^2\,\mathbf{E}\left[\tilde{m}''|x\right], \tag{44}$$

$$m_\sigma = \beta_1 K_\sigma(x-\mu)\,\mathbf{E}\left[\tilde{m}'|x\right] + \beta_1\tau\tau_\sigma\,\mathbf{E}\left[\tilde{m}''|x\right]. \tag{45}$$

Here we used the identity

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

Indeed, by partial integration,

$$\mathbf{E}\left[\tilde{m}'(a+\tau\gamma)\gamma|x\right] = \int \tilde{m}'(a+\tau\gamma)\gamma q(\gamma)d\gamma = \tau\int \tilde{m}''(a+\tau\gamma)q(\gamma)d\gamma = \tau\,\mathbf{E}\left[\tilde{m}''(a+\tau\gamma)|x\right],$$

where $q(\gamma)$ is the density of the standard normal distribution and $a = a(x)$ is any function of x.

Now suppose that

$$\tilde{m}'' = c_0\tilde{m}' \tag{46}$$

with some constant c_0. Then by (40), (44), and (45), 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 7.1, 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(46) has the solution

$$\tilde{m}(t) = c_1 e^{c_0 t} + c_2. \tag{47}$$

In the log-linear gamma model and the log-linear Poisson model, see Sections 8.2 and 8.3, respectively, $\tilde{m}(t) = e^t$. Thus for these models μ and σ can be pre-estimated.

In the logit model, see Section 8.4, $\tilde{m}(t) = (1 + e^{-t})^{-1}$, which does not satisfy the differential equation (46). In this case, σ must be estimated along with the other parameters, although μ can still be pre-estimated by $\overline{x}$. The same is true for the probit model.

7.4 Linear-in-β and polynomial Gaussian model

In this section, we study the possibility of pre-estimating μ and σ in the context of a measurement error model, where the mean function is linear in β. It is given by the equations

$$\begin{aligned} y &= \beta^\top \Phi(\xi) + \varepsilon, \quad \beta \in \mathbb{R}^k, \quad \Phi(\xi) \in \mathbb{R}^k \\ x &= \xi + \delta, \end{aligned} \tag{48}$$

where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$, ε independent of ξ, and ξ and δ are as before. Φ is a known vector-valued function of ξ. Clearly,

$$m^*(\xi, \beta) = \beta^\top \Phi(\xi) \tag{49}$$

and $v^* = \sigma_\varepsilon^2$. ($\sigma_\varepsilon^2$ corresponds to the dispersion parameter φ, which has to be estimated along with β. For reasons given in Section 5, we can assume σ_ε^2 to be known when we are only interested in comparing the ACMs of $\hat{\beta}_C$ and $\hat{\beta}_Q$). It follows that

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

Now suppose that

$$\Phi'(\xi) = D_1 \Phi(\xi), \tag{50}$$

where D_1 is a constant square matrix (which may depend on θ). Then

$$m_\mu = (1-K)\beta^\top D_1 m_\beta.$$

Therefore, according to Section 7.1, μ can be pre-estimated and $\hat{\mu}_Q = \overline{x}$.

Relation (50) holds, e.g., when $\Phi(\xi) = (1, \xi, \ldots, \xi^k)^\top$ (polynomial model) or $\Phi(\xi) = (e^{\lambda_1 \xi}, e^{\lambda_2 \xi}, \ldots, e^{\lambda_k \xi})^\top$, $\lambda_i \neq \lambda_j$, $i, j = 1, \ldots, k$, or $\Phi(\xi) = (1, \cos(a\xi), \sin(a\xi), \ldots, \cos(ka\xi), \sin(ka\xi))^\top$, $a > 0$ (trigonometric polynomial model), or $\Phi(\xi) = e^{\lambda\xi}(1, \xi, \ldots, \xi^k)^\top$, $\lambda \in \mathbb{R}$.

Considering the nuisance parameter σ, we have (see the derivation of (45))

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

In the polynomial model, m_σ is a polynomial function of x of degree k, while the components of $\mathbf{E}\,(\Phi|x)$, i.e., $\mathbf{E}\,[\xi^j|x]$, are polynomials of degree j, $j = 0, \ldots, k$. Therefore m_σ is a linear combination of the components of $\mathbf{E}\,(\Phi|x)$ (with coefficients depending on θ). Thus $m_\sigma = b^\top m_\beta$ with some constant vector b. According to Section 7.1, 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.

8 Strict optimality of QS as compared to CS in special measurement error models

Hereafter we consider the measurement error model (21), (22), where ξ and δ are normal independent random variables as described in Section 6.2. We suppose σ_δ^2 to be known. μ and σ are nuisance parameters.

8.1 Polynomial model

In the polynomial measurement error model (see Section 7.4), $y = \beta^\top \zeta + \varepsilon$, and $x = \xi + \delta$ with $\zeta = \zeta(\xi) = (1, \xi, \ldots, \xi^k)^\top$. The variable ε is independent of ξ and δ, and $\varepsilon \sim N(0, \sigma_\varepsilon^2)$.

This model is a special case of model (48) with $\Phi(\xi) = \zeta(\xi)$, $m^* = \beta^\top \zeta$, and $v^* = \sigma_\varepsilon^2$.

We distinguish the cases of nuisance parameters known and unknown.

Under known nuisance parameters, β is the only parameter to be estimated. The QS and CS functions are constructed as follows, see Shklyar *et al.* (2006):

$$S_Q = \frac{(y-m)m_\beta}{v}, \quad S_C = yt(x) - T(x)\beta, \tag{51}$$

where $t(x) = (t_0(x), \ldots, t_k(x))^\top$ is such that $\mathbf{E}\,(t(x)|\xi) = \zeta$ and $T(x) \in \mathbb{R}^{(k+1)\times(k+1)}$ such that $T(x)_{ij} = t_{i+j}(x),\ i, j = 0, \ldots, k$. The functions $t_j(x)$ are polynomials in x of degree j with leading term x^j, $j = 0, \ldots, k$. The mean function $m = m(x, \beta)$ is given by $m = \beta^\top r(x)$, where $r(x) = r = (r_0, \ldots, r_k)^\top$, $r_j = r_j(x)$ 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. From Section 7.4, we know that μ and σ can be pre-estimated on the basis of l_μ and l_σ alone, yielding the QS as well as CS estimators $\overline{x}$ and s_x^2, respectively: $\hat{\mu}_C = \hat{\mu}_Q = \overline{x}$, $\hat{\sigma}_C^2 = \hat{\sigma}_Q^2 = s_x^2$. The β part of the CS and QS functions remain unchanged as in (51) 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 8.1 *In a polynomial measurement error model of degree k with true degree s, the following relations regarding the ACMs of CS and QS hold.*

1. *Under known nuisance parameters:*
   - *(a) if $s = 0$, then $\Sigma_Q = \Sigma_C$;*
   - *(b) if $s > 0$, then $\Sigma_Q < \Sigma_C$.*
2. *Under unknown nuisance parameters:*
   - *(a) if $s = 0$, then $\Sigma_Q = \Sigma_C$;*
   - *(b) if $s = 1$, then $\operatorname{rank}(\Sigma_C^{(\beta)} - \Sigma_Q^{(\beta)}) = k - 1$;*
   - *(c) if $s = 2$, then $\operatorname{rank}(\Sigma_C^{(\beta)} - \Sigma_Q^{(\beta)}) = k$;*
   - *(d) if $s \geq 3$, then $\Sigma_Q^{(\beta)} < \Sigma_C^{(\beta)}$.*

Remark: In particular, in case $k = s = 1$, $\Sigma_Q^{(\beta)} = \Sigma_C^{(\beta)}$, which is natural since in a linear model under unknown nuisance parameters $\hat{\beta}_C = \hat{\beta}_Q$.

8.2 Poisson model

In the loglinear Poisson measurement error model, $y|\xi \sim Po(\lambda)$ with $\lambda = \exp(\beta_0 + \beta_1\xi)$, and $x = \xi + \delta$. Here $m^* = v^* = \lambda$.

For QS, we have, cf. Shklyar and Schneeweiss(2005),

$$m(x,\theta) = \exp\left\{\beta_0 + \beta_1\mu(x) + \beta_1^2\tau^2/2\right\}, \quad v(x,\theta) = m^2(x,\theta)(e^{\beta_1^2\tau^2} - 1) + m(x,\theta).$$

with $\mu(x)$ and τ^2 from (25) and (26), respectively. The β-component of the CS function is, cf. Shklyar and Schneeweiss (2005),

$$S_C^{(\beta)} = yg - h, \quad g = (1,x)^\top, \quad h = \exp\left\{\beta_0 + \beta_1 x - \frac{1}{2}\beta_1^2\sigma_\delta^2\right\}(1, x - \sigma_\delta^2\beta_1)^\top.$$

From Section 7.3, we know that μ and σ^2 can be pre-estimated and therefore $\Sigma_C - \Sigma_Q$ is of the form (33). We can apply Corollary 7.1. For $\beta_1 \neq 0$, the variables $\{(mg-h)_0, (mg-h)_1, l_\mu, l_\sigma\}$ are linearly independent, since the functions

$$\left\{1,\ x,\ x^2,\ e^{\beta_1 Kx},\ e^{\beta_1 x},\ xe^{\beta_1 Kx},\ xe^{\beta_1 x}\right\}$$

are linearly independent. For the same reason, m_{β_0} and m_{β_1} are linearly independent under $\beta_1 \neq 0$:

$$m_{\beta_0} = e^{const} \cdot e^{\beta_1 Kx}, \quad m_{\beta_1} = const \cdot e^{\beta_1 Kx} + const \cdot xe^{\beta_1 Kx}.$$

Thus by Corollary 7.1, $\Sigma_Q^{(\beta)} < \Sigma_C^{(\beta)}$ under $\beta_1 \neq 0$.

8.3 Gamma model

In the loglinear Gamma measurement error model, $y|\xi$ follows a Gamma distribution $G(\omega, \pi)$ with $\omega = \exp(\beta_0 + \beta_1\xi)$, $\pi > 0$, and $x = \xi + \delta$:

$$f(y|\eta) = \frac{1}{\Gamma(\pi)}\left(\frac{\pi}{\omega}\right)^\pi y^{\pi-1}\exp\left(-\frac{y\pi}{\omega}\right), \quad y > 0.$$

Here $m^* = \omega$ and $v^* = \pi^{-1}\omega^2$, where π^{-1} corresponds to the dispersion parameter φ, which, according to Section 5, we can assume to be known. For QS, we have

$$m(x,\theta) = \exp\left\{\beta_0 + \beta_1\mu(x) + \beta_1^2\tau^2/2\right\},$$

$$v(x,\theta) = \left(1 + \frac{1}{\pi}\right)\exp\{2\beta_0 + 2\beta_1\mu(x) + 2\beta_1^2\tau^2\} - \exp\{2\beta_0 + 2\beta_1\mu(x) + \beta_1^2\tau^2\}.$$

The β-component of the CS function is

$$S_C^{(\beta)} = yg - h, \quad g = \exp\left\{-\beta_0 - \beta_1 x - \frac{1}{2}\beta_1^2\sigma_\delta^2\right\}(1, x + \beta_1\sigma_\delta^2)^\top, \quad h = (1,x)^\top,$$

cf. Kukush *et al.* (2005). As in Section 8.2, see also Section 7.3, we can apply Corollary 7.1. For $\beta_1 \neq 0$, the variables $\{(mg-h)_0, (mg-h)_1, l_\mu, l_\sigma\}$ are linearly independent, since the functions

$$\left\{1,\ x,\ x^2,\ e^{\beta_1(1-K)x},\ xe^{\beta_1(1-K)x}\right\}$$

are linearly independent. In addition, as in Section 8.2, m_{β_0} and m_{β_1} are linearly independent under $\beta_1 \neq 0$. Thus by Corollary 7.1, $\Sigma_Q^{(\beta)} < \Sigma_C^{(\beta)}$ under $\beta_1 \neq 0$.

8.4 Logit model

In the logit measurement error model, y is a binary variable following a binomial distribution, the mean of which is a logistic function of a linear predictor in ξ:

$$y \sim \mathcal{B}(1,\pi), \quad \pi = H(\eta) = (1+e^{-\eta})^{-1}, \quad \eta = \beta_0 + \beta_1\xi, \quad x = \xi + \delta.$$

For this model, $m^* = \pi$, $v^* = \pi(1-\pi)$.

For QS, we need the mean and variance functions of y given x, which are given by

$$m = \mathbf{E}\left[\{1+\exp(-\beta_0 - \beta_1(Kx + (1-K)\mu + \tau\gamma)\}^{-1}|x\right], \quad v = m(1-m), \tag{52}$$

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

We can then construct the quasi score function (19) for $\theta = (\beta_0, \beta_1, \mu, \sigma)^\top$ with l_α from (27). As y is binary, the QS estimator of θ is just the ML estimator. Note that, according to Section 7.2, the QS estimator of μ is the empirical mean $\overline{x}$. We cannot say the same for the QS estimator of σ^2, see below.

To find the CS estimator, we start from the maximum likelihood score function for β in the error free model, which is given by

$$S_M^{(\beta)} = \left(y - \frac{1}{1+e^{-\eta}}\right)(1,\ \xi)^\top.$$

We construct a modified corrected score (C*S) function for β, as a function $S_{C^*}^{(\beta)} = S_{C^*}^{(\beta)}(y,x,\beta)$ such that

$$\mathbf{E}\,(S_{C^*}^{(\beta)}|y,\xi) = S_M^{(\beta)}(1+e^{-\eta}) = \{y(1+e^{-\eta})-1\}(1,\xi)^\top.$$

$S_{C^*}^{(\beta)}$ is of the form $S_{C^*}^{(\beta)} = yg_c - h_c$, where g_c and h_c are functions of x and β such that

$$\mathbf{E}\,(g_c|\xi) = (1+e^{-\beta_0-\beta_1\xi})(1,\xi)^\top, \quad \mathbf{E}\,(h_c|\xi) = (1,\xi)^\top.$$

The solutions to these deconvolution problems are

$$g_c = (1+e^{a-\beta_1 x},\ x + (x+\beta_1\sigma_\delta^2)e^{a-\beta_1 x})^\top, \quad h_c = (1,\ x)^\top, \tag{53}$$

where $a = -\beta_0 - \beta_1^2\sigma_\delta^2/2$. According to (20), $S_{C^*}^{(\beta)}$ has to be supplemented by the subvector l_α, which yields the conventional estimators of the nuisance parameters μ and σ^2: $\hat{\mu}_{C^*} = \overline{x}$ and $\hat{\sigma}_{C^*}^2 = s_x^2$.

In addition to the QS and CS estimators, we also consider the conditional score (DS) estimator, cf. Carroll *et al.* (1995). Let $z = x + y\sigma_\delta^2\beta_1$, $\eta_* = \beta_0 + \beta_1 z$. Then

$$\begin{aligned} \mathbf{E}\,(y|z) &= m_* := H(\eta_* - \beta_1^2\sigma_\delta^2/2) \\ \mathbf{V}(y|z) &= v_* := H(1-H). \end{aligned}$$

The conditional score function for β is then given by, cf. Carroll *et al.* (1995), $S_D^{(\beta)} = (y-m_*)(1,z)^t$. It is obviously unbiased. By using the fact that y is binary, the conditional score function can be written as a linear function of y: $S_D^{(\beta)} = yg_d - h_d$, where

$$\begin{aligned} g_d &= \{1 - H(\beta_0 + \beta_1 x + \beta_1^2\sigma_\delta^2/2)\}(1, x+\beta_1\sigma_\delta^2)^\top + H(\beta_0 + \beta_1 x - \beta_1^2\sigma_\delta^2/2)(1,x)^\top, \\ h_d &= -H(\beta_0 + \beta_1 x - \beta_1^2\sigma_\delta^2/2)(1,x)^\top. \end{aligned}$$

If $S_D^{(\beta)}$ is supplemented by the subvector $(l_\mu, l_\sigma)^\top$, then DS is a member of the class $\mathcal{L}$ of linear score functions. The conditional score estimators of μ and σ^2 are $\hat{\mu}_D = \overline{x}$ and $\hat{\sigma}_D^2 = s_x^2$.

Now, according to Theorem 4.1,

$$\Sigma_Q \leq \Sigma_{C^*} \quad \text{and} \quad \Sigma_Q \leq \Sigma_D. \tag{54}$$

But we can also compare $\Sigma_{C^*}^{(\beta,\sigma)}$ and $\Sigma_D^{(\beta,\sigma)}$ to $\Sigma_Q^{(\beta,\sigma)}$, where these matrices are the ACMs of the corresponding estimators of $(\beta_0, \beta_1, \sigma)^\top$.

Since $\hat{\mu}_{C^*} = \hat{\mu}_D = \hat{\mu}_Q$, we have for the μ-components $\Sigma_{C^*}^{(\mu)} = \Sigma_D^{(\mu)} = \Sigma_Q^{(\mu)}$, and thus by (54),

$$\operatorname{rank}\left(\Sigma_{C^*}^{(\beta,\sigma)} - \Sigma_Q^{(\beta,\sigma)}\right) = \operatorname{rank}\left(\Sigma_{C^*} - \Sigma_Q\right),$$

$$\operatorname{rank}\left(\Sigma_D^{(\beta,\sigma)} - \Sigma_Q^{(\beta,\sigma)}\right) = \operatorname{rank}\left(\Sigma_D - \Sigma_Q\right).$$

Theorem 8.2 *In the logit model, $\Sigma_Q^{(\beta,\sigma)} \leq \Sigma_{C^*}^{(\beta,\sigma)}$ and $\Sigma_Q^{(\beta,\sigma)} \leq \Sigma_D^{(\beta,\sigma)}$. When $\beta_1 \neq 0$, the inequalities become strict inequalities.*

In particular, under $\beta_1 \neq 0$, $\Sigma_Q^{(\sigma)} < \Sigma_{C^*}^{(\sigma)}$ and $\Sigma_Q^{(\sigma)} < \Sigma_D^{(\sigma)}$. This means that in the logit model $\hat{\sigma}_Q^2$ is an asymptotically more efficient estimator of σ^2 than $\hat{\sigma}_{C^*}^2 = \hat{\sigma}_D^2 = s_x^2$.

9 ACMs at a parameter point with constant mean and variance functions

In the previous sections, we studied the behavior of the ACMs of LS and QS under any θ. In particular, the second statement of Theorem 4.1 says that if the ACMs of LS and CS are equal for all θ, then LS and QS give rise to the same estimator: $\hat{\theta}_L = \hat{\theta}_Q$. It can, however, happen that for some special $\theta = \theta^*$ the ACMs of LS and QS are equal even though the corresponding estimators $\hat{\theta}_L$ and $\hat{\theta}_Q$ do not coincide. This can happen, in particular, if the mean and variance functions are constant at $\theta = \theta^*$.

We study this possibility in the context of the measurement error model of Section 6.2 with CS in place of LS.

9.1 Rank condition under constant mean and variance functions

Let us suppose that in the measurement error model of Section 6.2 the mean and variance functions $m^*(\xi, \beta)$ and $v^*(\xi, \beta)$ are constant at some $\beta = \beta^*$:

$$m^*(\xi, \beta^*) \equiv m_0, \quad v^*(\xi, \beta^*) \equiv v_0.$$

Then, by (23) and (24), $m(\xi, \beta)$ and $v(\xi, \beta)$ are also constant at $\beta = \beta^*$:

$$m(x, \beta^*) \equiv m_0, \quad v(x, \beta^*) \equiv v_0. \tag{55}$$

As (55) holds for all μ and σ,

$$m_\mu = m_\sigma = 0$$

at $\beta = \beta^*$.

In addition, we have, at $\beta = \beta^*$, by (28), (29),

$$\begin{aligned} \mathbf{E}\left[(mg-h)|\xi\right] &= m_0\,\mathbf{E}\left(g|\xi\right) - \mathbf{E}\left(h|\xi\right) \\ &= m_0 v_0^{-1} m_\beta^* - m_0 v_0^{-1} m_\beta^* = 0 \end{aligned}$$

and thus (under regularity conditions), at $\beta = \beta^*$,

$$mg - h = m(x,\beta^*)g(x,\beta^*) - h(x,\beta^*) \equiv 0. \tag{56}$$

Now, by Theorem 6.1, (55) and (56) imply

$$\operatorname{rank}\left(\Sigma_C - \Sigma_Q\right) = \operatorname{rank}\left[\left(\begin{array}{c} 0 \\ g_i \end{array}\right), \left(\begin{array}{c} 0 \\ m_{\beta_i} \end{array}\right),\ i = 0,\ldots,k, \left(\begin{array}{c} l_\mu \\ 0 \end{array}\right), \left(\begin{array}{c} l_\sigma \\ 0 \end{array}\right)\right] - (k+2) \tag{57}$$

at $\beta = \beta^*$. But at $\beta = \beta^*$ the identifiability condition (7) simplifies to the condition that the two system of random variables $[m_{\beta_i}, i = 1,\ldots,k,]$ and $[l_\mu, l_\sigma]$ are both linearly independent (just as in Section 6.3). Consequently, (57) simplifies to

$$\operatorname{rank}\left(\Sigma_C - \Sigma_Q\right) = \operatorname{rank}\left[g_i, m_{\beta_i}, i = 1,\ldots,k\right] - k \tag{58}$$

at $\beta = \beta^*$.

9.2 Linear predictor model

We apply the result of Section 9.1 to a linear predictor model as defined in Section 7.2, i.e., $m^*(\xi,\beta) = \tilde{m}(\beta_0 + \beta_1\xi)$. In addition, we assume that $v^*(\xi,\beta) = v_0$ if $\beta_1 = 0$. We then can state the following.

Theorem 9.1 *Suppose that in a linear predictor model $\beta_1 = 0$ and that $v^*(\xi,\beta) = v_0$ at $\beta_1 = 0$, then $\Sigma_Q = \Sigma_C$.*

Since several of the special models considered in Section 8 are linear predictor models, we have the following corollary, which also summarizes the results of Sections 8.2 to 8.4.

Corollary 9.1 *1. In both Poisson and Gamma models we have $\Sigma_Q^{(\beta)} < \Sigma_C^{(\beta)}$ under $\beta_1 \neq 0$ and $\Sigma_Q = \Sigma_C$ under $\beta_1 = 0$.*

2. In the logit model we have $\Sigma_Q^{(\beta,\sigma)} < \Sigma_{C^*}^{(\beta,\sigma)}$ under $\beta_1 \neq 0$ and $\Sigma_Q = \Sigma_{C^*}$ under $\beta_1 = 0$.

10 Conclusion

When one wants to estimate a parametric regression of y on x given by a conditional mean function $\mathbf{E}\left(y|x\right) = m(x,\theta)$ and supplemented by a conditional variance function $\mathbf{V}(y|x) = v(x,\theta)$, 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.

We study a number of concrete measurement error models in greater detail and present conditions under which QS is strictly more efficient than CS and also mention when they are equally efficient.

¿From this study one might get the impression that in a measurement error model QS is always to be preferred to CS, even in the presence of nuisance parameters. But this impression can be misleading. It is only true if not only the mean-variance model but also the distribution of the regressor ξ has been correctly specified. First of all, there are cases where there is no well-defined distribution for ξ, and then CS is the estimator of ones choice. In other cases, one may be uncertain about the true distribution of ξ, in particular, that distribution need not be Gaussian. If one then still uses the QS approach on the wrong assumption of a Gaussian regressor, the QS estimator will turn out to be (asymptotically) biased, cf. Schneeweiss and Cheng (2006).

Acknowledgements. The authors are grateful to Dr. Sergiy Shklyar for fruitful discussions.

11 Appendix

11.1 Lemmas

Lemma 11.1 *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 (A) *denotes the defect of a matrix A, i.e., the dimension of its kernel* $\ker(A)$.

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 11.2 *Let f and g be two random vectors of the same dimension d, such that $\mathbf{E}\, gg^\top > 0$. Consider the matrix $M = \mathbf{E}\, ff^\top - \mathbf{E}\, fg^\top (\mathbf{E}\, gg^\top)^{-1}\, \mathbf{E}\, gf^\top$. 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) $\operatorname{rank} M = \operatorname{rank}\,[f_i, g_i,\ i = 1, \ldots, d] - d$, *where the latter rank is the maximum number of linearly independent random variables in the set* $\{f_i, g_i,\ i = 1, \ldots, d\}$.

Proof. 1) To prove the first statement, let

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

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

2) To prove the second statement, let

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

Then $M = (F - A^\top A)$ and by Lemma 11.1

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

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

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

□

11.2 Proof of Theorem 4.1

We apply the first statement of Lemma 11.2 to the random vectors $g = S_Q$ and $f = S_L$. We have

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

Due to (12) and (13) 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 $\mathbf{E}\, S_L S_Q^\top$ 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 $\sum_{i=1}^n H(\theta) S_L(x_i, y_i; \theta) = 0$, 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.

□

11.3 Proof of Theorem 4.2

We apply the second statement of Lemma 11.2 with $g = S_Q$, $f = S_L$. By (12) and (13),

$$\begin{aligned} \operatorname{rank} (\Sigma_L - \Sigma_Q) &= \operatorname{rank} M = \operatorname{rank} [(S_L)_i, (S_Q)_i,\ i = 1, \ldots, d] - d \\ &= d - \operatorname{def} [(S_L)_i, (S_Q)_i,\ i = 1, \ldots, d]. \end{aligned} \tag{59}$$

To find the defect, we form a linear combination of the components of S_L and S_Q, which equals zero a.s.:

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

or

$$\left(c_1^\top g + \frac{c_2^\top m_\theta}{v}\right) y = c_1^\top h + \frac{c_2^\top m m_\theta}{v} - c_2^\top l_\theta \quad \text{a.s.} \tag{60}$$

The defect in (59) is equal to the maximum number of linearly independent vectors $(c_1^\top, c_2^\top)^\top$ which satisfy (60). But (60) is equivalent to

$$c_1^\top g + \frac{c_2^\top m_\theta}{v} = 0 \quad \text{and} \quad c_1^\top h + \frac{c_2^\top m m_\theta}{v} - c_2^\top l_\theta = 0 \quad \text{a.s.} \tag{61}$$

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, (61) 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

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

and (14) follows from (59).

□

11.4 Proof of Corollary 4.1

Suppose the random variables (15) are linearly independent. Then because of the identifiability condition (7), 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

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

implies first $a = 0$ because of the independence of the random variables in (15) and then $b = 0$ because of (7). According to Theorem 4.2, it follows that $\Sigma_Q < \Sigma_L$.

□

11.5 Proof of Theorem 6.2

Part a) follows immediately from Theorem 4.1 and (33).

Part b): By Theorem 6.1 and because of (33), the rank equation (37) holds true. But

$$\text{span}\left[\begin{pmatrix} mg_i - h_i \\ vg_i \end{pmatrix}, \begin{pmatrix} 0 \\ m_{\beta_i} \end{pmatrix}, i = 1, \ldots, k\right] \cap \text{span}\left[\begin{pmatrix} l_{\alpha_j} \\ 0 \end{pmatrix}, j = 1, \ldots, d-k\right] = \left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right\}$$

because, for any vectors a and b, $a^\top(mg-h) = b^\top l_\alpha$ implies $\mathbf{E}\, b^\top l_\alpha l_\alpha^\top = \mathbf{E}\, a^\top (mg-h) l_\alpha^\top = 0$ by (32) and thus $b^\top l_\alpha = 0$. Therefore, since the l_{α_j} are linearly independent, see (30), equality (34) follows. (35) is proved in the same way as the last part of Theorem 4.2.

Part c) follows from part b) taking into account that, by (30), the m_{β_i} are linearly independent.

□

11.6 Proof of Theorem 8.1

1) First we assume that the nuisance parameters μ and σ^2 are known. Then $\Sigma_Q^{(\beta)} = \Sigma_Q$ and $\Sigma_C^{(\beta)} = \Sigma_C$. We have $m = \beta^\top r$, $g = t$, $h = T\beta$, and $m_\beta = r$. By Theorem 6.2, part b),

$$\text{rank } (\Sigma_C - \Sigma_Q) = \text{rank } \left[\begin{pmatrix} \{(tr^\top - T)\beta\}_i \\ vt_i \end{pmatrix}, \begin{pmatrix} 0 \\ r_i \end{pmatrix}, \quad i = 0, \ldots, k \right] - k - 1.$$

First, consider $s = 0$. Then $t\, r^\top \beta = T\beta$ and $v = \sigma_\epsilon^2$ and

$$\text{rank } (\Sigma_C - \Sigma_Q) = \text{rank } \left[\begin{pmatrix} 0 \\ x^i + \ldots \end{pmatrix}, \begin{pmatrix} 0 \\ K^i x^i + \ldots \end{pmatrix}, \quad i = 0, \ldots, k \right] - k - 1.$$

The rank of the right-hand side is $k + 1$ and therefore rank $(\Sigma_C - \Sigma_Q) = 0$, i.e., $\Sigma_Q = \Sigma_C$.

Next, consider $s \geq 1$. Then $v = a_s x^{2s-2} + \ldots$, $a_s \neq 0$ and

$$\Big((tr^\top - T)\beta\Big)_i = \beta_s (K^s - 1) x^{i+s} + \ldots, \quad \beta_s \neq 0, \quad i = 0, \ldots, k.$$

These $k+1$ functions are linearly independent, and by Theorem 6.2, part b), $\Sigma_Q < \Sigma_C$. This proves part 1 of the theorem.

2) Now we suppose μ and σ^2 are to be estimated together with β.

In order to compute the rank of $\Sigma_C^{(\beta)} - \Sigma_Q^{(\beta)}$, we apply Theorem 6.1. The system to be examined is

$$\left[\begin{pmatrix} ((tr^\top - T)\beta)_i \\ vt_i \end{pmatrix}, \begin{pmatrix} 0 \\ r_i \end{pmatrix}, \ i = 0, \ldots, k, \ \begin{pmatrix} 0 \\ m_\mu \end{pmatrix}, \begin{pmatrix} 0 \\ m_\sigma \end{pmatrix}, \begin{pmatrix} l_\mu \\ 0 \end{pmatrix}, \begin{pmatrix} l_\sigma \\ 0 \end{pmatrix} \right].$$

The polynomials m_μ and m_σ are of degree not greater than k. Therefore they can be excluded from the system, which then can be simplified to

$$\left[\begin{pmatrix} ((tr^\top - T)\beta)_i \\ vt_i \end{pmatrix}, \begin{pmatrix} 0 \\ x^i \end{pmatrix}, \ i = 0, \ldots, k, \ \begin{pmatrix} x - \mu \\ 0 \end{pmatrix}, \begin{pmatrix} (x-\mu)^2 - \sigma^2 \\ 0 \end{pmatrix} \right].$$

Consider $s = 0$. Writing only the leading terms of the polynomials, the system becomes

$$\left[\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ x^k \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ x^k \end{pmatrix}, \begin{pmatrix} x \\ 0 \end{pmatrix}, \begin{pmatrix} x^2 \\ 0 \end{pmatrix} \right]$$

and its rank is $k + 3$. Thus, according to Theorem 6.1 with $d = k + 3$, $\Sigma_Q^{(\beta)} = \Sigma_C^{(\beta)}$ and $\Sigma_Q = \Sigma_C$.

Consider $s = 1$. The leading terms of the polynomials in the system are

$$\left[\begin{pmatrix} x \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} x^{k+1} \\ x^k \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ x^k \end{pmatrix}, \begin{pmatrix} x \\ 0 \end{pmatrix}, \begin{pmatrix} x^2 \\ 0 \end{pmatrix} \right]$$

and the rank is $2k + 2$, and rank $(\Sigma_C^{(\beta)} - \Sigma_Q^{(\beta)}) = k - 1$.

Consider $s = 2$. The leading terms of the polynomials are

$$\left[\begin{pmatrix} x^2 \\ x^2 \end{pmatrix}, \ldots, \begin{pmatrix} x^{k+2} \\ x^{k+2} \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ x^k \end{pmatrix}, \begin{pmatrix} x \\ 0 \end{pmatrix}, \begin{pmatrix} x^2 \\ 0 \end{pmatrix} \right]$$

and the rank is $2k + 3$, and rank $(\Sigma_C^{(\beta)} - \Sigma_Q^{(\beta)}) = k$.

Consider $s \geq 3$. The leading terms of the polynomials are

$$\left[\begin{pmatrix} x^s \\ x^{2s-2} \end{pmatrix}, \ldots, \begin{pmatrix} x^{k+s} \\ x^{k+2s-2} \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ x^k \end{pmatrix}, \begin{pmatrix} x \\ 0 \end{pmatrix}, \begin{pmatrix} x^2 \\ 0 \end{pmatrix}\right]$$

and the rank is $2k+4$, and rank $(\Sigma_C^{(\beta)} - \Sigma_Q^{(\beta)}) = k+1$. Thus $\Sigma_Q^{(\beta)} < \Sigma_C^{(\beta)}$.

□

11.7 Proof of Theorem 8.2

The first statement is a direct consequence of Theorem 4.1. So we need only prove the strict inequalities under $\beta_1 \neq 0$.

First we prove the linear independence of $[l_\mu, l_\sigma, (mg_c - h_c)_0, (mg_c - h_c)_1]$, then the linear independence of $[l_\mu, l_\sigma, (mg_d - h_d)_0, (mg_d - h_d)_1]$, and finally the linear independence of $[m_{\beta_0}, m_{\beta_1}, m_\sigma]$, where $l_\mu \propto x - \mu$, $l_\sigma \propto (x-\mu)^2 - \sigma^2$. By Corollary 7.2 with $\alpha' = \sigma$ and $\alpha'' = \mu$, these facts will yield that $\Sigma_Q^{(\beta,\sigma)} < \Sigma_{C^*}^{(\beta,\sigma)}$ and $\Sigma_Q^{(\beta,\sigma)} < \Sigma_D^{(\beta,\sigma)}$.

Consider the case $\beta_1 > 0$ (the case $\beta_1 < 0$ can be treated similarly).

1) From (52), we have

$$m(x) = \mathbf{E}\,[H(\beta_0+\beta_1\xi)|x] \sim \exp\{\beta_0+\beta_1(Kx+(1-K)\mu)\}\,\mathbf{E}\,e^{\beta_1\tau\gamma} = const\cdot e^{\beta_1 Kx} \quad \text{as} \quad x \to -\infty.$$

Together with (53) it follows that

$$(mg_c - h_c)(x) \sim const \cdot e^{\beta_1(K-1)x}(1,x)^\top \quad \text{as} \quad x \to -\infty.$$

Thus the functions $l_\mu, l_\sigma, (mg_c - h_c)_0, (mg_c - h_c)_1$ have different asymptotic behavior, as $x \to -\infty$, and are therefore linearly independent.

2) As to the asymptotic behavior of $(mg_d - h_d)$, we have, as $x \to -\infty$,

$$(g_d)_0 \to 1, \quad (g_d)_1 \sim x$$

$$(h_d)_0 \sim const \cdot e^{\beta_1 x}, \quad (h_d)_1 \sim const \cdot xe^{\beta_1 x},$$

and thus

$$(mg_d - h_d)_0 \sim const \cdot e^{\beta_1 Kx}, \quad (mg_d - h_d)_1 \sim const \cdot xe^{\beta_1 Kx}.$$

Again the functions $l_\mu, l_\sigma, (mg_d - h_d)_0, (mg_d - h_d)_1$ have different asymptotic behavior as $x \to -\infty$ and are therefore linearly independent.

3) We have, by (52), see also (40), (44), (45),

$$\begin{aligned} m_{\beta_0} &= \mathbf{E}\,[H'|x] = \mathbf{E}\,[H'\{\beta_0 + \beta_1(Kx + (1-K)\mu + \tau\gamma)\}|x], \\ m_{\beta_1} &= (Kx + (1-K)\mu)\,\mathbf{E}\,[H'|x] + \tau^2\,\mathbf{E}\,[H''|x], \\ m_\sigma &= \beta_1 K_\sigma(x-\mu)\,\mathbf{E}\,[H'|x] + \beta_1\tau\tau_\sigma\,\mathbf{E}\,[H''|x], \end{aligned}$$

where $H^{(i)} = H^{(i)}(\beta_0 + \beta_1\xi)$. This system of equations can also be written in matrix form:

$$\begin{pmatrix} m_{\beta_0} \\ m_{\beta_1} \\ m_\sigma \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ (1-K)\mu & K & \tau^2 \\ -\beta_1 K_\sigma \mu & \beta_1 K_\sigma & \beta_1\tau\tau_\sigma \end{pmatrix} \begin{pmatrix} \mathbf{E}\,[H'|x] \\ x\,\mathbf{E}\,[H'|x] \\ \mathbf{E}\,[H''|x] \end{pmatrix} \tag{62}$$

Because of $\tau^2 = K\sigma_\delta^2$, see (26), and $K_\sigma \neq 0$, the transformation matrix on the right hand side of (62) is nonsingular if $\beta_1 \neq 0$. By the properties of the logistic function, we have

$$H' = H - H^2, \quad H'' = H' - 2(H^2 - H^3).$$

Therefore the vector on the right hand side of (62) is a nonsingular linear transformation of the vector of functions $(f_1(x), f_2(x), f_3(x))^\top$, where

$$f_1(x) = \mathbf{E}\,[H - H^2|x], \quad f_2(x) = x\,\mathbf{E}\,[H - H^2|x], \quad f_3(x) = \mathbf{E}\,[H^2 - H^3|x].$$

To prove the linear independence of $[m_{\beta_0}, m_{\beta_1}, m_\sigma]$ it thus suffices to show that $[f_1, f_2, f_3]$ are linearly independent. But this is guaranteed by the fact that these functions have different asymptotic behavior, as $x \to -\infty$. Indeed, $\mathbf{E}\,[H^r|x] \sim const \cdot e^{r\beta_1 Kx}$ and thus

$$f_1(x) \sim const \cdot e^{\beta_1 Kx}, \quad f_2(x) \sim const \cdot xe^{\beta_1 Kx}, \quad f_3(x) \sim const \cdot e^{2\beta_1 Kx}.$$

□

11.8 Proof of Theorem 9.1

According to (58), we have for $k = 2$:

$$\text{rank}\,(\Sigma_C - \Sigma_Q) = \text{rank}\,(g_0, g_1, m_{\beta_0}, m_{\beta_1}) - 2.$$

We evaluate $g_0, g_1, m_{\beta_0}, m_{\beta_1}$. For $\beta_1 = 0$,

$$\begin{aligned} m^*_{\beta_0} &= \tilde{m}'(\beta_0) =: d_0 \\ m^*_{\beta_1} &= \xi\tilde{m}'(\beta_0) = d_0\xi. \end{aligned}$$

It follows that (under regularity conditions)

$$g = v_0^{-1} d_0 (1, x)^\top$$

for $\beta_1 = 0$. As, see (39),

$$m = \mathbf{E}\,[\tilde{m}\{\beta_0 + \beta_1(Kx + (1-K)\mu + \tau\gamma)\}|x]\,,$$

we have, for $\beta_1 = 0$,

$$\begin{aligned} m_{\beta_0} &= \mathbf{E}\,\tilde{m}'(\beta_0) = d_0 \\ m_{\beta_1} &= \mathbf{E}\,\left[\{Kx + (1-K)\mu + \tau\gamma\}\tilde{m}'(\beta_0)|x\right] \\ &= \{Kx + (1-K)\mu + \tau\gamma\}\tilde{m}'(\beta_0) =: d_1 + d_2 x. \end{aligned}$$

It follows that, for $\beta_1 = 0$,

$$\text{rank}\,(g_0, g_1, m_{\beta_0}, m_{\beta_1}) - 2 = \text{rank}\,(1, x) - 2 = 0$$

and thus $\Sigma_Q = \Sigma_C$.

□

References

[1] Armstrong, B. (1985), Measurement error in the generalized linear model. *Communications in Statistics - Simulation and Computation* **14**, 529-544.

[2] Carroll, R. J., Ruppert, D., and Stefanski, L.A. (1995), *Measurement Error in Nonlinear Models*. Chapman and Hall, London.

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

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

[5] Kukush, A. Schneeweiss, H, and Shklyar, S. (2005), *Quasi Score is more efficient than Corrected Score in a general nonlinear measurement error model*, Discussion Paper 451, SFB 386, Universität München.

[6] 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.

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

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

[9] Schneeweiss, H. (2005), *The polynomial and the Poisson measurement error models: some futher results on quasi score and corrected score estimators*, Discussion Paper 446, SFB 386, Universität München.

[10] Schneeweiss, H. and Cheng, C.-L. (2006), Bias of the structural quasi-score estimator of a measurement error model under misspecification of the regressor distribution. *Journal of Multivariate Analysis* **97**, 455-473.

[11] Shklyar, S. (2006), Comparing estimators in Poisson log-linear model. To appear in *Buletin of Kyiv National University*.

[12] Shklyar, S. and Schneeweiss, H. (2005), A comparison of asymptotic covariance matrices of three consistent estimators in the Poisson regreession model with measurement errors. *Journal of Multivariate Analysis* **94**, 250-270.

[13] Shklyar, S., Schneeweiss, H., and Kukush, A. (2006), Quasi Score is more efficient than Corrected Score in a polynomial measurement error model. *Metrika* online.

[14] Stefanski, L. (1989), Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Communications in Statistics, Part A - Theory and Methods* **18**, 4335-4358.

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