Theory of Stochastic Processes
Vol.13 (29), no.4, 2007, pp.69–81

ALEXANDER KUKUSH, ANDRII MALENKO, AND HANS SCHNEEWEISS

COMPARING THE EFFICIENCY OF ESTIMATES IN CONCRETE ERRORS-IN-VARIABLES MODELS UNDER UNKNOWN NUISANCE PARAMETERS

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. 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. 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. Under unknown nuisance parameters we derive conditions under which the QS estimator is strictly more efficient than the CS estimator. We focus on the loglinear Poisson, the Gamma, and the logit model.

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:

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

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*

Alexander Kukush is supported by the Swedish Institute grant SI-01424/2007.
Invited lecture.
2000 *Mathematics Subject Classifications*: 62J05, 62J12, 62F12, 62F10, 62H12, 62J10.
Key words and phrases: Mean-variance model, measurement error model, quasi score estimator, corrected score estimator, nuisance parameter, optimality property.

model, cf. Carroll *et al.* (1995). We want to estimate θ on the basis of an i.i.d. sample (x_i, y_i), $i = 1, \dots, n$.

General statements and results on the polynomial EIVM can be found in Shklyar *et al.* (2007) for known nuisance parameter α and in Kukush *et al.* (2006) for unknown α. Here we consider other special cases of the models, which can be treated as mean-variance model (1), i.e. the loglinear Poisson, the Gamma, and the logit model. We focus on the case of unknown mean and variance of the latent variable. The case of known nuisance parameters is considered in Kukush and Schneeweiss (2006).

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 Lowener order, i.e., for symmetric matrices A and B of equal size, $A < B$ and $A \le B$ means that $B - A$ is positive definite and $B - A$ is positive semidefinite, respectively.

The paper is organized as follows. Section 2 contains general results on mean-variance models and measurement error model. In Section 3 special cases of Poisson, Gamma, and logit EIVM are treated, and Section 4 concludes.

2. General results

The estimation of θ in the mean-variance model (1) cannot be performed by 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 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):

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

where unbiasedness means that $\forall\, \theta \in \Theta : \quad \mathbf{E}\, S_L(x, y; \theta) = 0$. 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 (this is the identifiability condition).

A quasi score functions is defined as follows, see Kukush *et al.* (2006):

$$S_Q(x, y; \theta) := \frac{(y-m)m_\theta}{v} + l_\theta, \quad l := \log \rho(x, \theta). \tag{4}$$

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

$$\sum_{i=1}^{n} S_Q(x_i, y_i, \hat{\theta}_Q) = 0. \tag{5}$$

The quasi-score function (4) belongs to $\mathcal{L}$, therefore the estimator $\hat{\theta}_Q$ is consistent and asymptotically normal under regularity conditions.

Theorem 2.1 (Optimality of QS) *Let S_L be a score function from the class $\mathcal{L}$ and S_Q be the quasi-score function (4). 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 2.2 (Strict Optimality of QS) *Under the conditions of Theorem 2.1*

$$\operatorname{rank}(\Sigma_L - \Sigma_Q) = \operatorname{rank}\left[\begin{pmatrix} mg_i - h_i \\ vg_i \end{pmatrix}, \begin{pmatrix} l_{\theta_i} \\ m_{\theta_i} \end{pmatrix}, i = 1, \ldots, d\right] - d, \tag{6}$$

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 (6) are linearly independent.

If

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

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

$$\{(mg - h)_i,\ i = 1, \ldots, d,\ l_{\theta_j}, j \in B_\theta\} \tag{7}$$

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

A reader can find the proofs of Theorems 2.1 and 2.2 in Kukush *et al.* (2006).

2.1 Measurement error model

A measurement error model is a model where the response variable y depends on a latent (unobservable) variable ξ with distribution $p(\xi, \alpha)$. Here θ is split into two subvectors,

$$\theta = (\beta^\top, \alpha^\top)^\top, \quad \beta \in \mathbb{R}^k, \quad \alpha \in \mathbb{R}^{d-k}. \tag{8}$$

In such a case we call β the unknown parameter of interest and α – the unknown nuisance parameter.

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

$$x = \xi + \delta, \tag{9}$$

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 known σ_δ^2.

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:

$$\mathbf{E}\,(y|\xi) = m^*(\xi, \beta), \quad \mathbf{V}\,(y|\xi) = v^*(\xi, \beta). \tag{10}$$

Note that m^* and v^* do not depend on α. From (10) we can derive the conditional mean and variance functions of y given x:

$$m(x, \beta, \alpha) := \mathbf{E}\,(y|x) = \mathbf{E}\,[m^*(\xi, \beta)|x] \tag{11}$$

$$v(x, \beta, \alpha) := \mathbf{V}\,(y|x) = \mathbf{E}\,[v^*(\xi, \beta)|x] + \mathbf{V}\,[m^*(\xi, \beta)|x]. \tag{12}$$

To compute these, we need to know the conditional distribution of ξ given x, which we can derive from the unconditional distribution of ξ, $p(\xi, \alpha)$, and the measurement equation (9).

The quasi-score function (4) takes the form

$$S_Q = \begin{pmatrix} (y-m)v^{-1}m_\beta \\ (y-m)v^{-1}m_\alpha + l_\alpha \end{pmatrix}. \tag{13}$$

An important special case for $p(\xi, \alpha)$ 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

$$\begin{aligned} \mu(x) &= Kx + (1-K)\mu, & (14) \\ \tau^2 &= K\sigma_\delta^2, & (15) \end{aligned}$$

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

$$l_\alpha = (l_\mu, l_\sigma)^\top = \left(\frac{x-\mu}{\sigma^2}, \frac{(x-\mu)^2}{\sigma^3} - \frac{1}{\sigma} \right)^\top. \tag{16}$$

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

$$\begin{aligned} \mathbf{E}[g|\xi] &= v^{*-1}m_\beta^* & (17) \\ \mathbf{E}[h|\xi] &= m^* v^{*-1} m_\beta^*. & (18) \end{aligned}$$

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, see Stefanski (1989).

2.2 Pre-estimation

In the measurement error model 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 (13) 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 2.1 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

$$m_\alpha = Am_\beta \tag{19}$$

with some nonrandom matrix A (which may depend on θ).

Corollary 2.2 *Suppose in a model with nuisance parameters as described by (8, 9) condition* (19) *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 2.2, 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 2.2.

Corollary 2.3 *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*

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

and

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

are both linearly independent.

The proof of Corollaries 2.2 and 2.3 can be found in Kukush *et al.* (2006).

3. Special cases

Consider the mean-variance measurement error model of Section 2.1 and assume that the error free mean function m^* is a function of a linear predictor in ξ:

$$m^*(\xi, \beta) = \tilde{m}(\beta_0 + \beta_1\xi), \quad \beta = (\beta_0, \beta_1)^\top. \tag{20}$$

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

$$m = \mathbf{E}\,(m^*|x) = \mathbf{E}\,[\tilde{m}\{\beta_0 + \beta_1(Kx + (1-K)\mu + \tau\gamma)\}|x]\,, \tag{21}$$

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

This is a Generalized Linear Model (GLM).

We have $m_\mu = \beta_1(1-K)m_{\beta_0}$ and thus by Corollary 2.3 the QS estimator of μ is just empirical mean, $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Now suppose that in GLM

$$\tilde{m}'' = c_0\tilde{m}' \tag{22}$$

with some constant c_0. Then by Corollary 2.3 we obtain the QS estimator of σ^2 is just empirical variance. This property holds for the Poisson and the Gamma models, but it does not hold for Logit one. We give indirect proof of the fact that in the Logit model QS estimator of σ^2 is not empirical variance and σ has to be estimated together with other unknown parameters.

3.1 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),

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

with $\mu(x)$ and τ^2 from (14) and (15), respectively. The β-component of the CS function is, cf. Shklyar and Schneeweiss (2005),

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

We know that μ and σ^2 can be pre-estimated and therefore $\Sigma_C - \Sigma_Q$ is of the form

$$\Sigma_L - \Sigma_Q = \begin{pmatrix} \Sigma_L^{(\beta)} - \Sigma_Q^{(\beta)} & 0 \\ 0 & 0 \end{pmatrix}. \tag{23}$$

We can apply Corollary 2.2. 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 2.2, $\Sigma_Q^{(\beta)} < \Sigma_C^{(\beta)}$ under $\beta_1 \neq 0$.

3.2 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 Kukush *et al.* (2006), 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

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

cf. Kukush *et al.* (2005). As in Section 3.1, we can apply Corollary 2.2. 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 3.1, m_{β_0} and m_{β_1} are linearly independent under $\beta_1 \neq 0$. Thus by Corollary 2.2, $\Sigma_Q^{(\beta)} < \Sigma_C^{(\beta)}$ under $\beta_1 \neq 0$.

3.3 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

$$m = \mathbf{E}\left[\{1 + \exp(-\beta_0 - \beta_1(Kx + (1 - K)\mu + \tau\gamma)\}^{-1}|x\right], v = m(1 - m), \tag{24}$$

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

We can then construct the quasi score function (13) for $\theta = (\beta_0, \beta_1, \mu, \sigma)^\top$ with l_α from (16). As y is binary, the QS estimator of θ is just the ML estimator. Note that, according to properties of GLM, 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

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

Due to complex zeros in the denominator one cannot solve the deconvolution problem

$$\mathbf{E}(S_C^{(\beta)}|y,\xi) = S_M^{(\beta)}.$$

Therefore 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

$$g_c = (1+e^{a-\beta_1 x},\ x + (x+\beta_1\sigma_\delta^2)e^{a-\beta_1 x})^\top, \quad h_c = (1,\ x)^\top, \tag{25}$$

where $a = -\beta_0 - \beta_1^2\sigma_\delta^2/2$. Function $S_{C^*}^{(\beta)}$ 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 $\hat{\sigma}^2_{C^*} = s_x^2$.

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 2.1,

$$\Sigma_Q \leq \Sigma_{C^*} \quad \text{and} \quad \Sigma_Q \leq \Sigma_D. \tag{26}$$

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 (26),

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

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

Theorem 3.1 *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 $\hat{\sigma}_{C^*}^2 = \hat{\sigma}_D^2 = s_x^2$.

3.4 Proof of Theorem 3.1

The first statement is a direct consequence of Theorem 2.1. 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

$$l_\mu \propto x - \mu, \quad l_\sigma \propto (x-\mu)^2 - \sigma^2.$$

By Corollary 2.3 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 (24), we have, as $x \to -\infty$,

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

Together with (25) 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 (24),

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

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

$$\begin{pmatrix} m_{\beta_0} \\ m_{\beta_1} \\ m_\sigma \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ (1-K)\mu & K & \tau^2 \\ -\beta_1 K_\sigma \mu & \beta_1 K_\sigma & \beta_1\tau\tau_\sigma \end{pmatrix} \begin{pmatrix} \mathbf{E}\,[H'|x] \\ x\ \mathbf{E}\,[H'|x] \\ \mathbf{E}\,[H''|x] \end{pmatrix} \tag{27}$$

Because of $\tau^2 = K\sigma_\delta^2$, see (15), and $K_\sigma \neq 0$, the transformation matrix on the right hand side of (27) 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 (27) 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

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

4. Conclusions

We studied the Poisson, the Gamma, and the logit errors-in-variance models with unknown nuisance parameters. For the Poisson and the Gamma models, we showed that the Quasi-Score estimator for β is strictly more efficient than the Corrected Score one for β. For the logit model, we proved that the compound Quase-Score estimator for β and σ is strictly more efficient than both Corrected Score and Conditional Score estimators for β and σ. In particular in the logit model the Quasi-Score estimator for σ is different from the empirical variance of x. For the Gamma and the Poisson models the Quasi-Score estimators for σ coincides with the empirical variance of x. All the three models are GLMs, therefore the Quasi-Score estimator of μ is just the empirical mean of x.

References

1. Carroll, R. J., Ruppert, D., and Stefanski, L.A., *Measurement Error in Nonlinear Models*, Chapman and Hall, London, (1995).
2. Heyde, C. C., *Quasi-Likelihood And Its Application*, Springer, New York, (1997).
3. Kukush, A. and Schneeweiss, H., *Comparing different estimators in a nonlinear measurement error model. I.*, Mathematical Methods of Statistics, (2005), **14**, 53–79.
4. Kukush, A., Schneeweiss, H., and Shklyar, S., *Quasi Score is more efficient than Corrected Score in a general nonlinear measurement error model*, Discussion Paper 451, SFB 386, Universität München, (2005).
5. Kukush., A. and Schneeweiss, H., *Asymptotic optimality of the quasi-score estimator in a class of linear score estimators*, Discussion Paper 477, SFB 386, Universität München, (2006).
6. Kukush, A., Malenko, A., and Schneeweiss, H., *Optimality of the quasi-score estimator in a mean-variance model with applications to measurement error models*, Discussion Paper 494, SFB 386, Universität München, (2006).
7. Schneeweiss, H., and Kukush, A., Comparing the efficiency of structural and functional methods in measurement error models. Submitted (a).
8. Schervish, M. J., *Theory of Statistics*, Springer, New York, (1995).
9. Shklyar, S., and Schneeweiss H. (2005). *A comparison of asymptotic covariance matrices of three consistent estimators in the Poisson regression model with measurement errors*, Journal Multivariate Analysis, (2005), **94** (2), 250–270.
10. Shklyar, S., Schneeweiss, H., and Kukush, A., *Quasi Score is more efficient than Corrected Score in a polynomial measurement error model*, Metrika, (2007), **65**, 275–295.
11. Stefanski, L., *Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models*, Communications in Statistics, Part A - Theory and Methods, (1989), **18**, 4335–4358.

Department of Mathematical Analysis, Kyiv National Taras Shevchenko University, Kyiv, Ukraine.
E-mail address: alexander_kukush@univ.kiev.ua

Department of Probability Theory and Mathematical Statistics, Kyiv National Taras Shevchenko University, Kyiv, Ukraine
E-mail address: exipilis@yandex.ru

University of Muenchen, Germany
E-mail address: Hans.Schneeweiss@stat.uni-muenchen.de