Kukush, Schneeweiß, Shklyar:

Quasi Score is more efficient than Corrected Score in a general nonlinear measurement error model

Sonderforschungsbereich 386, Paper 451 (2005)

Online unter: http://epub.ub.uni-muenchen.de/

Quasi Score is more efficient than Corrected Score in a general nonlinear measurement error model

A. Kukush[1], H. Schneeweiss[2] and S. Shklyar[1]

[1] Kiev National Taras Shevchenko University,
Vladimirskaya st. 60, 01033, Kiev, Ukraine,
e-mail: shklyar@univ.kiev.ua,
alexander_kukush@univ.kiev.ua

[2] University of Munich,
Akademiestr.1, 80799 Munich, Germany
e-mail: schneew@stat.uni-muenchen.de

Abstract

We compare two consistent estimators of the parameter vector β of a general exponential family measurement error model with respect to their relative efficiency. The quasi score (QS) estimator uses the distribution of the regressor, the corrected score (CS) estimator does not make use of this distribution and is therefore more robust. However, if the regressor distribution is known, QS is asymptotically more efficient than CS. In some cases it is, in fact, even strictly more efficient, in the sense that the difference of the asymptotic covariance matrices of CS and QS is positive definite.

Key words: Measurement errors, nonlinear regression, exponential family, corrected score, quasi score, asymptotic efficiency.

1 Introduction

We consider the following nonlinear measurement-error model. Let Y be a random variable which has a p.d.f. (with respect to a σ-finite Borel measure) belonging to the exponential family:

$$\log f(y|\xi) = \frac{y\xi - C(\xi)}{\varphi} + c(y, \varphi), \tag{1}$$

where $\xi \in \mathbb{R}$ is the canonical parameter and $\varphi > 0$ a dispersion parameter, which may be known or unknown. The function $C : \mathbb{R} \to \mathbb{R}$ is supposed to be sufficiently smooth. The parameter ξ is a function of some random vector X and an unknown parameter vector β:

$$\xi = \xi(X, \beta).$$

The density of Y, as given above, is therefore a conditional density given X. Note that the conditional mean of Y given X (i.e., the regression function) is given by

$$\mathbb{E}(Y|X) = C'\left[\xi\left(X, \beta\right)\right]. \tag{2}$$

and the conditional variance (i.e., the residual variance of the regression) by

$$\mathrm{Var}(Y|X) = \varphi C''[\xi(X, \beta)]. \tag{3}$$

We assume $C''(\xi) > 0$. It is the regression function which we want to estimate, β being the parameter of interest. However, in a measurement error model, X is unobservable. Instead we observe the surrogate vector variable W, which is related to X via

$$W = X + U,$$

U being the (unobservable) measurement error. The random vector U is supposed to be independent of (X, Y).

The problem is to estimate β from a sample (Y_i, W_i), $i = 1, \cdots, n$, of observable data. We assume that the triples (Y_i, X_i, U_i), $i = 1, \cdots, n$, are independent. We also assume that β lies in the interior of a compact subspace Θ_β of Euclidian space.

For this model there exist several consistent estimators of β if additional information on the distribution of X and U is available. Functional methods only rest on the knowledge of the distribution of U. E.g., if $U \sim N(0, \sigma_u^2)$, σ_u^2 must be known. Among these methods the Corrected Score (SC) estimation method is most prominent, cf. Stefanski (1989), Nakamura (1990), Buonaccorsi (1996).

Structural methods use in addition knowledge of the distribution of X. E.g., if $X \sim N(\mu_x, \sigma_x^2)$, μ_x and σ_x^2 must be known. In principle, given the distribution of U, the distribution of X can be estimated from the observation W_i, $i = 1, \cdots, n$, without resort to the model. Here however we assume the distribution of X to be known. Among the structural methods the most prominent one is the Quasi Score (QS) estimation method, cf. Gleser (1990), Carroll et al. (1995).

It might be surmised that QS is more efficient that CS as it uses more information. But this is by no means obvious, as QS, unlike ML, does not use this information in a most efficient way. Nevertheless QS is quite popular because it is often much simpler to compute than ML. It is therefore of great interest to know whether QS is indeed more efficient than CS.

For the log-linear Poisson model this has been proved by Shklyar and Schneeweiss (2005) and for the polynomial model by Shklyar et al. (2005). Here we give a general proof.

In Section 2 and 3 we introduce the QS and CS procedures, respectively. A new, so-called Simple Score (SS), estimator is introduced in Section 4. It serves as an intermediate estimator in comparing the efficiency of QS and CS. This comparison is elaborated in Section 5. Section 6 has some examples of exponential family models. Section 7 studies a number of cases, where the efficiency comparison can be strengthened to a strict efficiency comparison. Section 8 contains some concluding remarks.

2 The quasi score (QS) estimator

Let

$$\begin{aligned} m &:= m(W,\beta) := \mathbb{E}\,(Y|W) \\ v &:= v(W,\beta,\varphi) := \mathbb{E}\left[(Y-m)^2\,|W\right] \end{aligned}$$

We assume m and v to be sufficiently smooth. We can compute m and v from the error-free regression function $\mathbb{E}\,(Y|X)$ and residual variance $\mathrm{Var}\,(Y|X)$, see (2) and (3):

$$m = \mathbb{E}\left[C'(\xi)|W\right] \tag{4}$$

$$v = \mathrm{Var}[C'(\xi)|W] + \varphi\mathbb{E}[C''(\xi)|W]. \tag{5}$$

Estimation of β is performed with the help of an unbiased vector-valued estimating (or: score) function. The QS score function is given by

$$S_Q := (Y-m)v^{-1}m_\beta. \tag{6}$$

Here and hereafter, the subscript β stands for the derivative with respect to β, i.e., $m_\beta := \frac{\partial m}{\partial \beta}$. For any scalar function of β, its derivative is a column vector of the same dimension as β. Obviously,

$$\mathbb{E}S_Q = \mathbb{E}[\mathbb{E}\,(S_Q|W)] = \mathbb{E}[\mathbb{E}\,(Y-m|W)\,v^{-1}m_\beta] = 0,$$

and so S_Q is indeed an unbiased estimating function. This score function must be supplemented by another unbiased estimating function for φ. But, for simplicity, we here assume that φ is known. The following results remain true even if φ is unknown and has to be estimated, see below.

The QS estimator $\hat{\beta}_Q$ is then found as the solution to the equation

$$\sum_{i=1}^n S_Q\left(Y_i, W_i, \hat{\beta}_Q\right) = 0, \qquad \hat{\beta}_Q \in \Theta_\beta.$$

Under natural assumptions, cf. Kukush and Schneeweiss (2005), $\hat{\beta}_Q$ exists uniquely, is strongly consistent, and asymptotically normal with asymptotic covariance matrix (ACM) Σ_Q given by the sandwich formula

$$\begin{aligned}\Sigma_Q &= A_Q^{-1} B_Q A_Q^{-\top} \\ A_Q &:= -\mathbb{E} S_{Q\beta} := -\mathbb{E}\frac{\partial S_Q}{\partial \beta^\top} \\ B_Q &:= \mathrm{Var} S_Q = \mathbb{E} S_Q S_Q^\top.\end{aligned}$$

Note that, by convention, the derivative of a column vector valued function $k(\beta)$ with respect to β is always meant to be a matrix $k_\beta = \frac{\partial k}{\partial \beta^\top}$ with (i,j)-element $\frac{\partial k_i}{\partial \beta_j}$.

An easy computation shows that $A_Q = B_Q = \mathbb{E} v^{-1} m_\beta m_\beta^\top$ and thus

$$\Sigma_Q = (\mathbb{E} v^{-1} m_\beta m_\beta^\top)^{-1}.$$

If φ is unknown and has to be estimated with the help of some score function like, e.g., $S_\varphi = Y^2 - m^2 - v$ or $S_\varphi = (Y-m)^2 - v$. Then β and φ have to be estimated simultaneously. Nevertheless, the ACM of β does not change. The proof of this assertion is essentially the same as given in Shklyar et al. (2005).

3 The corrected score (CS) estimator

We start from the likelihood score function of the error-free model (1):

$$S_{ML} := Y\xi_\beta - C'\xi_\beta.$$

The corrected score function is then given by

$$S_C := Yg - h, \tag{7}$$

where $g = g(W,\beta)$, $h = h(W,\beta)$ are (vector-valued) functions which are the solutions to the deconvolution problems

$$\begin{aligned}\mathbb{E}(g|X) &= \xi_\beta \quad (8)\\ \mathbb{E}(h|X) &= C'\xi_\beta. \quad (9)\end{aligned}$$

Obviously,

$$\mathbb{E} S_C = \mathbb{E}[\mathbb{E}(Yg - h|Y,X)] = \mathbb{E}\left(Y\xi_\beta - C'\xi_\beta\right) = 0,$$

and so S_C is indeed an unbiased estimating function.

Similar to QS, the ACM of the CS estimator of β is given by

$$\begin{aligned}
\Sigma_C &= A_C^{-1} B_C A_C^{-\top} \\
A_C &:= -\mathbb{E} S_{C\beta} := -\mathbb{E}\frac{\partial S_C}{\partial \beta^\top} \\
B_C &:= \mathrm{Var} S_C = \mathbb{E} S_C S_C^\top .
\end{aligned}$$

A_C turns out to be symmetric as will become clear during the course of proving Lemma 1.

4 The simple score (SS) estimator

The SS score function is a simplified version of the QS score function. By replacing in the latter the term $v^{-1} m_\beta$ with the function of the preceding section, g, we get

$$S_S := (Y - m) g.$$

The unbiasedness of S_S is shown in the same way as for S_Q. Under general conditions, the SS estimator is consistent and asymptotically normal with an ACM Σ_S which again is given by a sandwich formula similar to those of QS and CS.

According to the theory of score functions, cf. Heyde (1997), QS is optimal within the class of all estimators with score functions of the form $S = (Y - m)a$, where $a := a(W, \beta, \varphi)$ is an arbitrary function. As S_S is of this form, we see that

$$\Sigma_Q \le \Sigma_S$$

in the sense of the Loewner order.

5 Efficiency comparison of SS and CS

We compare the A and B terms of the sandwich formulas for SS and CS.

Lemma 1

$$A_C = A_S.$$

Proof. First start with CS. We have, by (7),

$$\begin{aligned} A_C &= -\mathbb{E}S_{C\beta} = -\mathbb{E}\left(Yg_\beta - h_\beta\right) \\ &= \mathbb{E}\mathbb{E}\left(h_\beta - Yg_\beta|W\right) \\ &= \mathbb{E}h_\beta - \mathbb{E}mg_\beta =: A_{C1} - A_{C2}. \end{aligned}$$

But, by (8) and (9),

$$\begin{aligned} \mathbb{E}\left(h_\beta|X\right) &= \frac{\partial \mathbb{E}(h|X)}{\partial \beta^\top} = (C'\xi_\beta)_\beta \\ &= C''\xi_\beta\xi_\beta^\top + C'\xi_{\beta\beta}, \\ \mathbb{E}\left(g_\beta|X\right) &= \frac{\partial \mathbb{E}\left(g|X\right)}{\partial \beta^\top} = \xi_{\beta\beta}. \end{aligned}$$

(Here $\xi_{\beta\beta}$ is short for $\frac{\partial \xi_\beta}{\partial \beta^\top} = \frac{\partial^2 \xi}{\partial \beta \partial \beta^\top}$). Therefore,

$$\begin{aligned} A_{C1} &= \mathbb{E}h_\beta = \mathbb{E}\mathbb{E}(h_\beta|X) \\ &= \mathbb{E}\left(C''\xi_\beta\xi_\beta^\top + C'\xi_{\beta\beta}\right) \end{aligned}$$

and

$$\begin{aligned} A_{C2} &= \mathbb{E}mg_\beta = \mathbb{E}[\mathbb{E}(C'|W)g_\beta] \\ &= \mathbb{E}\mathbb{E}\left(C'g_\beta|W\right) = \mathbb{E}C'g_\beta \\ &= \mathbb{E}\mathbb{E}\left(C'g_\beta|X\right) = \mathbb{E}[C'\mathbb{E}\left(g_\beta|X\right)] \\ &= \mathbb{E}C'\xi_{\beta\beta}. \end{aligned}$$

Thus

$$A_C = \mathbb{E}C''\xi_\beta\xi_\beta^\top.$$

As to SS,

$$\begin{aligned} A_S &= -\mathbb{E}S_{S\beta} = \mathbb{E}gm_\beta^\top \\ &= \mathbb{E}[g\mathbb{E}\left(C'_\beta|W\right)^\top] = \mathbb{E}gC''\xi_\beta^\top \\ &= \mathbb{E}\mathbb{E}\left[gC''\xi_\beta^\top|X\right] = \mathbb{E}\mathbb{E}\left(g|X\right)C''\xi_\beta^\top \\ &= \mathbb{E}\xi_\beta C''\xi_\beta^\top = \mathbb{E}C''\xi_\beta\xi_\beta^\top \\ &= A_C, \end{aligned}$$

which proves the lemma.$\diamond$

Lemma 2

$$B_S \leq B_C.$$

Proof. For CS we have

$$\begin{aligned} B_C &= \mathbb{E} S_C S_C^\top \\ &= \mathbb{E}\mathbb{E}\left[\{(Y-m)g+(mg-h)\}\{(Y-m)g+(mg-h)\}^\top|W\right] \\ &= \mathbb{E}vgg^\top + \mathbb{E}\,(mg-h)\,(mg-h)^\top. \end{aligned} \tag{10}$$

For SS we have

$$\begin{aligned} B_S &= \mathbb{E} S_S S_S^\top = \mathbb{E}\,(y-m)^2\,gg^\top \\ &= \mathbb{E}\mathbb{E}\left[(Y-m)^2 gg^\top|W\right] \\ &= \mathbb{E}vgg^\top. \end{aligned}$$

Obviously $B_S \leq B_C$, which proves the lemma.$\diamond$

Lemma 1 and Lemma 2 imply that $\Sigma_S \leq \Sigma_C$. Together with the result of Section 4 we can now maintain

Theorem

$$\Sigma_Q \leq \Sigma_S \leq \Sigma_C.$$

The theorem states that QS is more efficient than CS, and SS is intermediate between the two.

6 Examples

In this section we present a few examples of exponential-family models with measurement errors and show how CS and QS estimators can be constructed, see also Carrol et al. (1995). Explicit solutions can often be found under the following normality assumption:

(**N**) X and U are independent random variables and $X \sim N(\mu_x, \sigma_x^2)$, $U \sim N(0, \sigma_u^2)$.

6.1 Gaussian regression model

Consider $Y \sim N(\mu, \sigma^2)$ with $\mu = q(X, \beta)$, q being the regression function.

Then $\xi = \mu$ and $\varphi = \sigma^2$. We have $C(\xi) = \frac{1}{2}\xi^2,\ C'(\xi) = \xi$.

For QS we need to construct, according to (4) and (5),

$$\begin{aligned} m(W,\beta) &= \mathbb{E}[q(X,\beta)|W] \\ v(W,\beta,\sigma_\epsilon^2) &= \mathbb{V}[q(X,\beta|W] + \sigma_\epsilon^2, \end{aligned}$$

and for CS we need to find functions g and h, see (8) and (9), such that

$$\begin{aligned} \mathbb{E}[g(X,\beta)|X] &= \frac{\partial q(X,\beta)}{\partial \beta} \\ \mathbb{E}[h(X,\beta)|X] &= q(X,\beta)\frac{\partial q(X,\beta)}{\partial \beta}. \end{aligned}$$

For a polynomial model (which includes the linear model as a special case), all these functions are easy to compute under (**N**), cf. Kukush et al. (2005). An efficiency comparison for polynomial models has been carried out by Shklyar et al. (2005).

6.2 Loglinear Poisson model

Consider $Y \sim Po(\lambda)$ with $\lambda = \exp(\beta_0 + \beta_1 X)$.

Then $\xi = \log \lambda$ and $\varphi = 1$. We have $C(\xi) = C'(\xi) = e^\xi$.

For QS we have, under (**N**),

$$\begin{aligned} m(W,\beta) &= \exp\{\beta_0 + \beta_1\mu(W) + \frac{1}{2}\beta_1^2\tau^2\} \\ v(W,\beta) &= m^2(W,\beta)\{\exp(\beta_1^2\tau^2) - 1\} + m(W,\beta), \end{aligned}$$

where

$$\mu(W) := \mathbb{E}(X|W) = W - \frac{\sigma_u^2}{\sigma_w^2}(W - \mu_w), \tag{11}$$

$$\tau^2 := \mathbb{V}(X|W) = \sigma_u^2 - \frac{\sigma_u^4}{\sigma_w^2}. \tag{12}$$

For CS we find, when $U \sim N(0, \sigma_u^2)$,

$$
\begin{aligned}
g &= (1, W)^\top \\
h &= -e^d (1, W - \sigma_u^2 \beta_1)^\top,
\end{aligned}
$$

where $d = \beta_0 + \beta_1 W - \frac{1}{2}\beta_1^2 \sigma_u^2$.

For details see Kukush et al. (2004), and for an efficiency comparison see Shklyar and Schneeweiss (2005). In the latter paper a different SS estimator was used than in the present paper, which led to a much more complicated proof of the superiority of SS over QS; the present proof is simpler.

6.3 Loglinear Gamma model

Consider $Y \sim G(\mu, \nu)$, i.e.,

$$
f(y) = \frac{1}{\Gamma(\nu)} (\frac{\nu}{\mu})^\nu y^{\nu-1} \exp(-\frac{\nu}{\mu} y), \quad y > 0,
$$

with $\mu = \exp(\beta_0 + \beta_1 X)$. (In the special case $\nu = 1$, we have the loglinear exponential model.)

Here $\xi = -\frac{1}{\mu}$ and $\varphi = \frac{1}{\nu}$. We have $C(\xi) = -\log(-\xi)$, $C'(\xi) = -\frac{1}{\xi}$.

For QS we need to compute

$$
\begin{aligned}
m(W, \beta) &= \mathbb{E}[\exp(\beta_0 + \beta_1 X)|W] \\
v(W, \beta, \nu) &= (1 + \frac{1}{\nu}) \mathbb{E}[\exp(2\beta_0 + 2\beta_1 X)|W] \\
&\quad -\{\mathbb{E}[\exp(\beta_0 + \beta_1 X)|W]\}^2.
\end{aligned}
$$

Under (**N**) these become (in a similar way as in Section 6.2)

$$
\begin{aligned}
m(W, \beta) &= \exp\{\beta_0 + \beta_1 \mu(W) + \frac{1}{2}\beta_1^2 \tau^2\} \\
v(W, \beta, \nu) &= (1 + \frac{1}{\nu}) \exp\{2\beta_0 + 2\beta_1 \mu(W) + 2\beta_1^2 \tau^2\} \\
&\quad - \exp\{2\beta_0 + 2\beta_1 \mu(W) + \beta_1^2 \tau^2\}.
\end{aligned}
$$

For CS we need to find functions g and h such that

$$\begin{aligned}\mathbb{E}[g(W,\beta)|X] &= \exp(-\beta_0 - \beta_1 X)(1,X)^\top \\ \mathbb{E}[h(W)|X] &= (1,X)^\top.\end{aligned}$$

Under (**N**) these become

$$\begin{aligned}g(W,\beta) &= \exp(-\beta_0 - \beta_1 W - \frac{1}{2}\beta_1^2\sigma_u^2)(1, W + \beta_1\sigma_u^2)^\top \\ h(W) &= (1,W)^\top.\end{aligned}$$

6.4 Logit model

Consider $Y \sim B(1,\pi)$, i.e.,

$$f(y) = \pi^y(1-\pi)^{1-y} \quad , y \in \{0,1\},$$

with $\pi = \{1 + \exp(-\beta_0 - \beta_1 X)\}^{-1}$.

Here $\xi = \log(\frac{\pi}{1-\pi})$ and $\varphi = 1$. We have $C(\xi) = \log(1+e^\xi)$ and $C'(\xi) = (1+e^{-\xi})^{-1}$.

For QS we need to construct

$$\begin{aligned}m(W,\beta) &= \mathbb{E}[\{1+\exp(-\beta_0 - \beta_1 X)\}^{-1}|W] \\ v(W,\beta) &= m(W,\beta)\{1 - m(W,\beta)\}.\end{aligned}$$

There is no closed form expression for $m(W,\beta)$, even under (**N**). The expected value has to be computed by numerical integration, Crouch and Spiegelman (1990), Monahan and Stefanski (1992). However, a possible way out is to use a probit model as an approximation to the logit model. Indeed, it is well-known that the logistic function $(1+e^{-x})^{-1}$ is closely approximated by $\Phi(x/c)$, where Φ is the standard normal distribution function and $c = 1.70$, cf. Johnson and Kotz (1970, Chapter 22). Thus assume that $\pi = \Phi\{\frac{1}{c}(\beta_0 + \beta_1 X)\}$. Then, under (**N**),

$$m(W,\beta) = \Phi\left(\frac{\frac{1}{c}(\beta_0 + \beta_1\mu(W)}{\sqrt{1 + \frac{1}{c^2}\beta_1^2\tau^2}}\right).$$

So the conditional model, given W, is again a probit model and can be estimated by standard methods, one possibility being that it is again approximated by a logit model.

For CS, again with π following the logit model, we need to find functions g and h such that

$$\begin{aligned}\mathbb{E}[g(W)|X] &= (1,X)^\top \\ \mathbb{E}[h(W,\beta)|X] &= \{1+\exp(-\beta_0-\beta_1 X\}^{-1}(1,X)^\top\end{aligned}$$

Obviously $g(W) = (1,W)^\top$. But, according to Stefanski (1989), $h(W,\beta)$ does not exist in general. However, if (β_0, β_1, X) can be restricted such that $\beta_0 + \beta_1 X > 0$ (sometimes known as "rare event" restriction, Buzas and Stefanski (1996)), then a corrected score function exists. It can be evaluated with the help of a Taylor series expansion of the logistic function.
Indeed, with $z = \beta_0 + \beta_1 x$,

$$(1+e^{-z})^{-1} = \sum_{k=0}^{\infty}(-1)^k e^{-kz},$$

which is absolutely convergent if, and only if, $z > 0$. The function h is then given by

$$h = \sum_{k=0}^{\infty}(-1)^k \exp\{-k(\beta_0+\beta_1 W) - \frac{k^2}{2}\beta_1^2\sigma_u^2\}\begin{pmatrix} 1 \\ W + k\beta_1\sigma_u^2 \end{pmatrix}.$$

This is a consequence of the identities

$$\begin{aligned}\mathbb{E}[e^{aW}|X] &= \exp(aX + \frac{1}{2}a^2\sigma_u^2) \\ \mathbb{E}[We^{aW}|X] &= (X + a\sigma_u^2)\exp(aX + \frac{1}{2}a^2\sigma_u^2),\end{aligned}$$

see also Buzas and Stefanski (1996).

We cannot compare CS and QS for the logit model because these two methods cannot be considered under the same assumptions. It should be mentioned that a method different from CS exists, which does not need any restrictions on (β_0, β_1, X); this is the method of conditional scores, cf. Stefanski and Caroll (1987, Section 6.4).

7 Strict inequalities

In this section we study conditions under which the order relation $\leq$ in our theorem can be supplemented by an inequality $\neq$ or replaced by a strict ordering $<$, in the sense that the difference of the ACMs is positive definite and not only positive semi-definite. We focus on a comparison of Σ_S and Σ_C. The relations $\Sigma_S \neq \Sigma_C$ and $\Sigma_S < \Sigma_C$ imply the corresponding relations $\Sigma_Q \neq \Sigma_C$ and $\Sigma_Q < \Sigma_C$, respectively. We investigate some special cases, where such relations can be derived. Except for Corollary 1, we adopt assumption (**N**) of Section 6 throughout.

7.1 Inequality under the generalized polynomial model

From the proof of Lemma 2 (cf. (10)) it is clear that $\Sigma_S = \Sigma_C$ if, and only if, $mg - h = 0$ as a (vector-valued) function of W.

The generalized polynomial model is a special case of our exponential family model with

$$\xi = \beta_0 + \beta_1 X + \ldots + \beta_k X^k = \beta^\top \rho(X),$$

where $\beta := (\beta_0, \beta_1, \ldots, \beta_k)^\top$ and $\rho(X) := (1, X, \ldots, X^k)^\top$.

The functions g and h of (8) and (9) are given by

$$\mathbb{E}(g|X) = \rho(X) \quad (13)$$

$$\mathbb{E}(h|X) = C'(\xi)\rho(X). \quad (14)$$

Let $\beta_{-0} = (\beta_1, \ldots, \beta_k)^T$. We have the following corollaries to our theorem.

Corollary 1
In the generalized polynomial model, $\beta_{-0} = 0$ implies $\Sigma_S = \Sigma_C$.

Proof. $\beta_{-0} = 0$ implies $\xi = \beta_0$ and therefore $m = C'(\beta_0)$ and $h = C'(\beta_0)g$. Hence $mg - h = 0$ and $\Sigma_S = \Sigma_C$.$\diamond$

Corollary 2
In the generalized polynomial model under (**N**), the condition $\beta_{-0} \neq 0$ implies $\Sigma_S \neq \Sigma_C$ under the additional condition that $|C'(t)| \leq c_0(1+|t|)^q$ for some constant c_0 and some $q > 0$.

Proof. Suppose $\Sigma_S = \Sigma_C$. Then $mg-h = 0$, which implies $\mathbb{E}(mg-h)|X) \equiv 0$. Consider the first component, $\alpha_0 = \alpha_0(X)$, of $\mathbb{E}(mg-h|X)$, which then must also be zero. The first components of g and $\mathbb{E}(h|X)$ are, respectively, 1 and $C'(\xi)$, see (13) and (14). Therefore, by the definition of m, see (4),

$$\alpha_0 = \mathbb{E}\left[\mathbb{E}\{C'(\xi)|W\}|X\right] - C'(\xi).$$

To evaluate $\mathbb{E}[C'(\xi)|W] = \mathbb{E}[C'(\beta^T\rho(X))|W]$, recall that under (**N**)

$$X|W \sim N(\mu(W), \tau^2)$$

with $\mu(W)$ and τ^2 from (11) and (12). For simplicity write $\mu(W) = a+bW$, where $0 < b < 1$. Then

$$X = a + bW + \tau V_0,$$

where $V_0 \sim N(0,1)$ and V_0 is independent of W. Thus

$$\mathbb{E}[C'(\xi)|W] = \mathbb{E}[C'\{\beta^\top\rho(a+bW+\tau V_0)\}|W].$$

Clearly, we can replace V_0 in this expression by a variable V_1 with the same distribution as V_0 but being independent not only of W but also of (W, X). We can then write

$$\mathbb{E}[C'(\xi)|W] = \mathbb{E}[C'\{\beta^\top\rho(a+bW+\tau V_1)\}|W, X]$$

and consequently

$$\mathbb{E}[\mathbb{E}\{C'(\xi)|W\}|X] = \mathbb{E}[C'\{\beta^\top\rho(a+bW+\tau V_1)\}|X].$$

With $W = X + U$, we finally get

$$\alpha_0 = \mathbb{E}[C'\{\beta^\top\rho(a+bX+V)\}|X] - C'(\beta^\top\rho(X)), \tag{15}$$

where $V := bU + \tau V_1 \sim N(0, b^2\sigma_u^2 + \tau^2)$. Now by Lemma A of the appendix $\alpha_0 \equiv 0$ implies, under the condition of the corollary, that $C'(\xi)$ is a constant. But, because of $\beta_{-0} \neq 0$, ξ varies on an interval and therefore $C''(\xi) = 0$, which contradicts the model assumption $C''(\xi) > 0$. Therefore $\Sigma_S = \Sigma_C$ cannot be true.$\diamond$

7.2 Strict inequality in the generalized linear model

From the proof of Lemma 2 it follows that $\Sigma_S < \Sigma_C$ if, and only if, the components of $mg - h$ are linearly independent functions of W.

We note that for a Gaussian polynomial model, see Section 6.1, the function m and the components of g and h are polynomials in W. By comparing their degrees one can show that if $\beta_{-0} \neq 0$, the components of $mg - h$ are linearly independent, proving that $\Sigma_S < \Sigma_C$, see Shklyar et al. (2005).

We now specialize to the case of a generalized linear model (GLM), which is characterized by $\rho(X) = (1, X)^\top$ and $\xi = \beta_0 + \beta_1 X$.

Corollary 3
In a GLM under (**N**) with $C(\xi)$ being a polynomial in ξ of degree $p \geq 2$, $\beta_1 \neq 0$ is necessary and sufficient for $\Sigma_S < \Sigma_C$.

Proof. Necessity follows from Corollary 1.

In general, a sufficient condition for $\Sigma_S < \Sigma_C$ is that the components of $\mathbb{E}[(mg - h)|X]$ are linearly independent functions of X.

In the GLM, we have, by (13) and (14), $g = (1, W)^T$ and $\mathbb{E}(h|X) = C'(\xi)(1, X)^\top$ and hence, by (4),

$$\begin{aligned}\mathbb{E}[mg - h|X] &= \mathbb{E}[\mathbb{E}\{C'(\beta_0 + \beta_1 X)|W\}(1, W)^\top|X] \\ &\quad -C'(\beta_0 + \beta_1 X)(1, X)^\top.\end{aligned}$$

The first component of this vector is just α_0 of (15), which here becomes

$$\alpha_0 = \mathbb{E}[C'\{\beta_0 + \beta_1(a + bX + V)\}|X] - C'(\beta_0 + \beta_1 X). \tag{16}$$

Similarly, the second component can be evaluated as

$$\alpha_1 = \mathbb{E}[(X + U)C'\{\beta_0 + \beta_1(a + b(X + U) + \tau V_1)\}|X] - XC'(\beta_0 + \beta_1 X). \tag{17}$$

The first term on the r.h.s can be further reduced with the help of what is sometimes called Stein's Lemma, Stein (1981):

If $U \sim N(0, \sigma_u^2)$ and f is any function and f' its derivative, then $\mathbb{E}[Uf(U)] = \sigma_u^2 \mathbb{E}f'(U)$, provided the expectations exist.

Applying this lemma to (17), we get

$$\begin{aligned}\alpha_1 &= X\{\mathbb{E}[C'(\beta_0 + \beta_1(a + bX + V))|X] - C'(\beta_0 + \beta_1 X)\} \\ &\quad + \sigma_u^2 \beta_1 b \mathbb{E}[C''\{\beta_0 + \beta_1(a + bX + V)\}|X], \end{aligned} \tag{18}$$

where we used $V = bU + \tau V_1$. Now suppose C is a polynomial in ξ of degree $p \geq 2$ and $\beta_1 \neq 0$. Then α_0 is a nonvanishing polynomial in X (see the proof of Corollary 2) of degree r, say, and α_1 is a polynomial of degree $r+1$. Therefore α_0 and α_1 are linearly independent as functions of X.$\diamond$

Another GLM is the log-linear Poisson model, see Section 6.2, which is characterized by $C(\xi) = e^\xi$ and $\xi = \exp(\beta_0 + \beta_1 X)$. For this model we can state the following corollary, see also Shklyar and Schneeweiss (2002).

Corollary 4
In the log-linear Poisson model under (**N**), $\Sigma_S < \Sigma_C$ if $\beta_1 \neq 0$.

Proof. With $C(\xi) = e^\xi$ and $V \sim N(0, \sigma_v^2)$ we have

$$\mathbb{E}[C'\{\beta_0 + \beta_1(a + bX + V)\}|X] = e^{\beta_0 + \beta_1(a+bX)} e^{\frac{\beta_1^2 \sigma_v^2}{2}}$$

and the same for C'' in place of C'. We therefore find for the two components of $\mathbb{E}(mg - h|X)$, see (16) and (18):

$$\begin{aligned}\alpha_0 &= D_1 e^{\beta_1 bX} - D_2 e^{\beta_1 X} \\ \alpha_1 &= X D_1 e^{\beta_1 bX} - X D_2 e^{\beta_1 X} + D_3 e^{\beta_1 bX}\end{aligned}$$

with some constants $D_1, D_2, D_3 \neq 0$. As the functions $e^{\beta_1 bx}, e^{\beta_1 x}, xe^{\beta_1 bx}, xe^{\beta_1 x}$ are linearly independent it follows that α_0 and α_1, are also linearly independent. This implies $\Sigma_S < \Sigma_C$.$\diamond$

7.3 Strict inequality in an exponential polynomial model

In this subsection we investigate an exponential polynomial model of the following kind:

$$Y = \sum_{j=0}^{k} \beta_j e^{\lambda_j X} + \epsilon, \tag{19}$$

$\epsilon \sim N(0, \sigma_\epsilon^2), \quad \lambda_0 = 0 < \lambda_1 < \ldots < \lambda_k.$

The λ_i are real numbers. (Imaginary numbers λ_i lead to a trigonometric polynomial and can be treated in a similar way.) Here $C(\xi) = \frac{1}{2}\xi^2$, and $\xi = \beta^T \rho(X)$, where now $\rho(X) := (e^{\lambda_0 X}, \ldots, e^{\lambda_k X})^\top$ and $\beta = (\beta_0, \ldots, \beta_k)^\top$ as before.

By an extension of Corollary 1 to the present model, $\beta_{-0} = 0$ implies $\Sigma_S = \Sigma_C$. The converse is stated in the following corollary.

Corollary 5
In the exponential polynomial model (19) under (**N**), $\Sigma_S < \Sigma_C$ if $\beta_{-0} \neq 0$.

Proof. We have to show that the components of $\mathbb{E}[mg - h|X]$ are linearly independent if $\beta_{-0} \neq 0$. Let us find g, which has to satisfy (13) with the new ρ. Due to the identity

$$\mathbb{E}[\exp(\lambda_i W - \frac{1}{2}\lambda_i^2 \sigma_u^2)|X] = \exp(\lambda_i X), \quad i = 0, \ldots, k, \tag{20}$$

we can satisfy (13) if we take

$$g(W) = K\rho(W),$$

where $K = \text{diag}(k_0, k_1, \ldots, k_k), \quad k_i := e^{-\frac{1}{2}\lambda_i^2 \sigma_u^2}$. Therefore, by (2),

$$\mathbb{E}[mg|X] = K\mathbb{E}[\rho(W)\mathbb{E}(\sum_0^k \beta_j e^{\lambda_j X}|W)|X].$$

By similar arguments as in the derivation of (15), we can write

$$\mathbb{E}[mg|X] = K\mathbb{E}[\rho(X+U)\sum_0^k \beta_j \exp\{\lambda_j(a + bX + bU + \tau V_1)\}|X],$$

where $V_1 \sim N(0,1)$ is independent of (X,U). As in Section 7.1, let $V := bU + \tau V_1$. Then, owing to the joint normality of (U,V),

$$U = \gamma_0 V + \gamma_1 U_1,$$

where $U_1 \sim N(0,1)$ is independent of V. We thus can write

$$\mathbb{E}(mg|X) = KE[\rho(X + \gamma_0 V + \gamma_1 U_1)\sum_0^k \beta_j \exp\{\lambda_j(a + bX + V)\}|X].$$

Taking expectations with respect to U_1, keeping V and X fixed, we get for the i-th component a_i of $\mathbb{E}[mg|X]$

$$a_i = k_i e^{\frac{1}{2}\lambda_i^2\gamma_1^2} e^{\lambda_i X}\mathbb{E}[e^{\lambda_i\gamma_0 V}\sum_0^k \beta_j \exp\{\lambda_j(a + bX + V)\}|X].$$

Taking expectation with respect to V, we get

$$\begin{aligned} a_i &= k_i e^{\frac{1}{2}\lambda_i^2\gamma_1^2} e^{\lambda_i X}\sum_0^k \beta_j \exp\{\lambda_j(a + bX) + (\lambda_i\gamma_0 + \lambda_j)^2\frac{\sigma_v^2}{2}\} \\ &= \sum_0^k \beta_j r_{ij} e^{(\lambda_i + b\lambda_j)X}, \end{aligned} \tag{21}$$

where

$$r_{ij} = k_i \exp\{\frac{\lambda_i^2}{2}\gamma_1^2 + \lambda_j a + \frac{\sigma_v^2}{2}(\lambda_i\gamma_0 + \lambda_j)^2\}.$$

The vector $\mathbb{E}(h|X)$ satisfying (14) is given by

$$\mathbb{E}[h[X] = \{\beta^\top \rho(X)\}\rho(X)$$

and its i-th component by

$$b_i := \sum_{j=0}^k \beta_j e^{(\lambda_i + \lambda_j)X} \quad, i = 0, \ldots, k. \tag{22}$$

The components of $E(mg - h|X)$ are $a_i - b_i,\ i = 0, \ldots, k$. Consider a linear combination of these components

$$l = \sum_{i=1}^k z_i(a_i - b_i)$$

with at least one $z_i \neq 0$. Let m be the maximum index such that $z_m \neq 0$, $0 \leq m \leq k$. As $\beta_{-0} \neq 0$, there is a maximum index p such that $\beta_p \neq 0$,

$1 \leq p \leq k$. By (21) and (22), l is a linear combination of terms of the form $e^{(\lambda_i+\lambda_j)X}$ and $e^{(\lambda_i+b\lambda_j)X}$. The term with the largest exponent is

$$z_m \beta_p e^{(\lambda_m+\lambda_p)X}, \quad \lambda_p > 0,$$

which cannot be cancelled by any of the other terms of l because all of them have smaller exponents. Therefore $l \not\equiv 0$ for any vector $(z_0, z_1, \ldots, z_k) \neq 0$, and the $a_i - b_i$ are linearly independent. This proves the corollary.$\diamond$

8 Conclusion

We proved for a general nonlinear measurement-error model of the exponential family type that the quasi score estimator is more efficient than the corrected score estimator in the sense that $\Sigma_C - \Sigma_Q$ is positive semi-definite. In a number of important cases we can even say that this difference is positive definite.

This does not mean that QS is to be preferred under all circumstances. QS uses the distribution of X, which should be known, whereas CS does not need the knowledge of this distribution. The simplicity of the QS procedure is only guaranteed if the distribution of X is in some sense simple, e.g., if it is a Gaussian distribution as assumed in parts of this paper or a mixture of Gaussians. But more importantly, QS is a biased method if the wrong distribution for X has been used, cf. Schneeweiss and Cheng (2003). CS does not have that kind of bias. In this sense CS is more robust than QS and might therefore be preferred to QS when it is uncertain what distribution for X can be reasonably assumed.

Another drawback of QS is that even if the type of distribution of X is known, there are still a number of unknown parameters of this distribution, which have to be estimated before QS can be applied. Here we assumed these nuisance parameters known. But in case they have to be estimated, the ACM of QS will be larger than given in this paper, thus reducing its relative efficiency.

Furthermore, let us note that for small error variance σ_u^2, CS and QS (and also SS) have approximately the same ACM even if nuisance parameters have to be estimated. More precisely, $\Sigma_Q - \Sigma_C = O(\sigma_u^4)$, cf. Kukush and Schneeweiss (2005).

Finally let us refer to an unsolved problem. We know (and we used this fact) that QS is optimal within a special linear class of score functions, cf. Heyde (1997). It seems plausible that a similar result might hold for CS.

Appendix

Lemma A
Suppose $f : \mathbb{R} \to \mathbb{R}$ is a continuous function such that $|f(x)| \leq C_1(1+x^2)^q$ with some constant C_1 and some positive q. Let $V \sim N(0, \alpha^2)$ and let $a, b \in \mathbb{R}$, $0 < b < 1$. Then the equation

$$\mathbb{E} f(a + bx + V) \equiv f(x) \tag{23}$$

implies that $f(x)$ is independent of x.

The condition on f implies that $f \in S'(\mathbb{R})$, the space of slowly growing generalized functions, cf. Vladimirov (1979).

Proof. Equation (23) can be written in the form

$$f(x) * \varphi(x) = f(\frac{x-a}{b}) =: f_{ab}(x),$$

where $\varphi(x)$ is the density of $N(0, \alpha^2)$ and $*$ is the convolution sign. (Note that $\varphi \in S(\mathbb{R})$, the space of quickly decreasing basic functions). Taking Fourier transforms, this translates into

$$f^*(y)\varphi^*(y) = f^*_{ab}(y),$$

where the Fourier transforms f^* and f^*_{ab} are generalized functions from $S'(\mathbb{R})$, cf. Vladimirov (1979). Owing to the identities

$$\begin{aligned} \varphi^*(y) &= e^{-\frac{\alpha^2 y^2}{2}} \\ f^*_{ab}(y) &= b f^*(by) e^{iay}, \end{aligned}$$

we have

$$f^*(y)e^{-\frac{\alpha^2 y^2}{2}} = bf^*(by)e^{iay}$$

or, using the substitution $t = by$,

$$f^*(t) = \frac{1}{b}f^*(\frac{t}{b})e^{-\frac{\alpha^2 t^2}{2b^2}}e^{-i\frac{at}{b}}.$$

Repeated substitution of the l.h.s. in the r.h.s. yields for $m = 1, 2, \ldots$,

$$\begin{aligned} f^*(t) &= \frac{1}{b^m}f^*(\frac{t}{b^m}) \cdot e^{-\frac{\alpha^2 t^2}{2}(\frac{1}{b^2}+\cdots+\frac{1}{b^{2m}})} \cdot e^{-iat(\frac{1}{b}+\cdots+\frac{1}{b^m})} \\ &= \frac{1}{b^m}f^*(\frac{t}{b^m})\exp\{c_1\frac{t^2}{b^{2m}}(1-b^{2m}) + c_2\frac{t}{b^m}(1-b^m)\}, \end{aligned}$$

where $c_1 := -\frac{\alpha^2}{2(1-b^2)}$ and $c_2 := -\frac{ia}{1-b}$ are constants. Let $g \in S(\mathbb{R})$ and let f^* act on g:

$$\begin{aligned} &\langle f^*(t), g(t)\rangle \\ =\ &\langle \frac{1}{b^m}f^*(\frac{t}{b^m})\exp\{c_1\frac{t^2}{b^{2m}}(1-b^{2m}) + c_2\frac{t}{b^m}(1-b^m)\}, g(t)\rangle \\ =\ &\langle f^*(z)\exp\{c_1 z^2(1-b^{2m}) + c_2 z(1-b^m)\}, g(b^m z)\rangle, \end{aligned}$$

where we used the transformation $t = b^m z$. Letting m tend to infinity, the last expression becomes

$$\langle f^*(z)\exp\{c_1 z^2 + c_2 z\}, g(0)\rangle = C_2 g(0)$$

with some constant C_2. Thus $f^*(t) = C_2\delta(t)$, δ being the Dirac (generalized) function.

But this means that $f(x)$ is a constant, cf. Vladimirov (1979).

References

Buonaccorsi, J.P. (1996), A modified estimating equation approach to correcting for measurement error in regression. *Biometrika* **83** 433-440.

Buzas, J.S., Stefanski, L.A. (1996), A note on corected-score estimation. *Statistics & Probability Letters* **28** 1-8.

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

Crouch, E.A., Spiegelman, D. (1990), The evaluation of integrals of the form $\int_{-\infty}^{\infty} f(t)\exp(-t^2)\;dt$: application to logistic-normal models. *JASA* **85** 464-467.

Gleser, L.J.(1990), Improvement of the naive estimation in nonlinear errors-in-variables regression. *Statistical Analysis of Measurement Error Models and Application* (P. J. Brown, W. A. Fuller eds.). *Contemporary Mathematics* **112** 99–114.

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

Johnson, N.L., Kotz, S. (1970), *Distributions in Statistics: Continuous Univariate Distributions*, Vol.2, Wiley, New York.

Kukush, A., Schneeweiss, H. (2005), Comparing Different Estimators in a Nonlinear Measurement Error Model. I and II. *Mathematical Methods of Statistics* **14** 53–79 and 203–223.

Kukush, A., Schneeweiss, H., Wolf, R. (2004), Three estimators for the Poisson regression model with measurement errors. *Statistical Papers* **45** 351–368.

Kukush, A., Schneeweiss, H., Wolf, R. (2005), Relative efficiency of three estimators in a polynomial regression with measurement errors. *Journal of Statistical Planning and Inference* **127** 179–203.

Monahan, J., Stefanski, L.A. (1992), Normal scale mixture approximations to $F^*(z)$ and computation of the logistic-normal integral. In: Balakrishnan, N. (ed.): *Handbook of the Logistic Distribution*, Marcel Dekker, New York, 529-540.

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

Schneeweiss, H., Cheng, C.-L. (2003), *Bias of the quasi score estimator of a measurement error model under misspecification of the regressor distribution.* Discussion Paper 340/SFB 386, Universität München. To be published in Journal of Multivariate Analysis.

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

Shklyar, S., Schneeweiss, H., Kukush, A. (2005), *Quasi Score is more efficient than Corrected Score in a polynomial measurement error model.* Discussion Paper 445/SFB 386 Universität München.

Stefanski, L.A. (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.

Stefanski, L.A., Carroll, R.J. (1987), Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74** 703-716.

Stein, C.M. (1981), Estimation of the mean of a multivariate normal distribution. *Annals of Statistics* **9** 1135 - 1151.

Vladimirov, V.S. (1979), *Generalized Functions* in *Mathematical Physics* (translated from the second Russian edition by G. Yankovski), Mir, Moscow.