

## COMPARING THE EFFICIENCY OF STRUCTURAL AND FUNCTIONAL METHODS IN MEASUREMENT ERROR MODELS

UDC 519.21

H. SCHNEEWEISS AND A. KUKUSH

**ABSTRACT.** The paper is a survey of recent investigations by various researchers into the relative efficiencies of structural and functional estimators of the regression parameters in a measurement error model. Structural methods, in particular the quasi-score (QS) method, take advantage of the knowledge of the regressor distribution (if available). Functional methods, in particular the corrected score (CS) method, discard such knowledge and work even if such knowledge is not available. Among other results, it has been shown that QS is more efficient than CS as long as the regressor distribution is completely known. However, if nuisance parameters in the regressor distribution are to be estimated, this no longer remains true, in general. But by modifying the QS method, the adverse effect of the nuisance parameters can be overcome. For small measurement errors, the efficiencies of QS and CS become almost indistinguishable, whether nuisance parameters are present or not. QS is (asymptotically) biased if the regressor distribution has been misspecified, while CS is always consistent and thus more robust than QS.

### 1. INTRODUCTION

In recent years, a number of results on measurement error models have been derived by various researchers that deal with the relative efficiencies of structural and functional estimation methods. The structural estimation methods take the regressor distribution into account, whereas the functional estimation methods do not rely on this distribution. Our objective is to review these results, focusing mainly on two estimators: the structural quasi-score (QS) estimator and the functional corrected score (CS) estimator. Both are consistent, but they differ in their asymptotic covariance matrices (ACMs). The most important result will be that, generally speaking, QS is more efficient than CS, but this property may become invalid when nuisance parameters describing the regressor distribution need to be estimated. The result also has to qualify when the regressor distribution is misspecified because then the QS estimator becomes biased. If the measurement errors are small, then QS and CS are essentially equally efficient.

We also mention other estimators briefly. For a recent review on the broader field of measurement error models, see [24]. Books on measurement error models are [2, 6, 9, 27, 33].

Section 2 introduces the measurement error model, and Section 3 discusses the estimators under consideration. Section 4 presents some examples. In Section 5, we introduce the asymptotic covariance matrix and discuss a few “technical” assumptions, which are

---

2000 *Mathematics Subject Classification.* Primary 62F12, 62J02.

*Key words and phrases.* Measurement errors, errors in variables, quasi score, corrected score, structural methods, functional methods, efficiency comparison.

©2010 American Mathematical Society

needed to derive the asymptotic properties of the estimators. Section 6 reviews the various efficiency results, which are interpreted in the last section, the Conclusion.

## 2. THE MODEL

The classical measurement error model consists of three parts: (1) a *regression model* relating an unobservable (generally vector-valued) regressor variable  $\xi$  to a response variable  $y$ , which is considered here to be observable without measurement errors; (2) a *measurement model* relating the unobservable  $\xi$  to an observable surrogate variable  $x$ ; and (3) a *distributional model* for  $\xi$ . We consider these three parts in some more detail.

The *regression model* can be described by a conditional distribution of  $y$  given  $\xi$  and given an unknown parameter vector  $\theta$ . We assume that this distribution is represented by a probability density function  $f(y|\xi; \theta)$  with respect to some underlying  $\sigma$ -finite measure on the Borel  $\sigma$ -field of  $\mathbb{R}$ . Here we restrict our attention to distributions that belong to the exponential class; i.e., we assume  $f$  to be of the form

$$(1) \quad f(y|\xi; \beta, \varphi) = \exp\left(\frac{y\eta - c(\eta)}{\varphi} + a(y, \varphi)\right)$$

with

$$\eta = \eta(\xi; \beta),$$

where  $\beta$  is the regression parameter vector and  $\varphi$  is a scalar dispersion parameter such that  $\theta = (\beta^\top, \varphi)^\top$  and  $a$ ,  $c$ , and  $\eta$  are known functions [13]. This class comprises the class of generalized linear models, where  $\eta = \eta(\xi^\top \beta)$ . The conditional mean and conditional variance of  $y$  given  $\xi$  are, respectively,

$$(2) \quad \mathbf{E}(y|\xi) = m^*(\xi; \beta) = c'(\eta),$$

$$(3) \quad \mathbf{V}(y|\xi) = v^*(\xi; \beta, \varphi) = \varphi c''(\eta).$$

The conditional mean function  $m^*$  is the regression function, which is to be estimated. Although we started from model (1) and derived the mean and variance functions (2) and (3), one can also start from a mean-variance model (2), (3) from the outset. Most of the following results would still hold true.

The classical *measurement model* assumes that the observed variable  $x$  differs from the latent  $\xi$  by a measurement error variable  $\delta$ , which is independent of  $\xi$  and  $y$ :

$$x = \xi + \delta$$

with  $\mathbf{E} \delta = 0$ . Here we assume that  $\delta \sim N(0, \Sigma_\delta)$  with  $\Sigma_\delta$  known.

The parameter  $\theta$  has to be estimated with the help of observable data

$$(x_i, y_i), \quad i = 1, \dots, n.$$

We assume that  $(y_i, \xi_i, \delta_i)$  are i.i.d. variables.

The *distributional model* for  $\xi$  either states that the  $\xi$  are unknown constants (functional variant) or that  $\xi$  is a random variable (structural variant) with a distribution given by a density  $h(\xi; \gamma)$ , where  $\gamma$  is a vector of nuisance parameters describing the distribution of  $\xi$ . The arguments in this paper are based on the structural variant. We typically assume that  $\xi \sim N(\mu_\xi, \Sigma_\xi)$ , although we also sometimes assume that  $\xi$  follows a finite mixture of normal distributions. Most of the time, we assume  $\gamma$  to be known. If not, it can often be estimated in advance (i.e., pre-estimated) without considering the regression model and the data  $y_i$ . For example, if  $\xi$  is normal, then  $\mu_\xi$  and  $\Sigma_\xi$  can be estimated by  $\bar{x}$  and  $S_x - \Sigma_\delta$ , respectively, where  $\bar{x}$  and  $S_x$  are the empirical mean vector and empirical covariance matrix of the data  $x_i$ .

## 3. ESTIMATORS

If the variable  $\xi$  were observable, one could estimate  $\beta$  (and also  $\varphi$ ) by the method of maximum likelihood. The corresponding likelihood-score function for  $\beta$  is given by

$$(4) \quad \psi^*(y, \xi; \beta, \varphi) = \frac{\partial \log f(y|\xi; \beta, \varphi)}{\partial \beta} = \frac{y - c'(\eta)}{\varphi} \frac{\partial \eta}{\partial \beta}$$

or, because of (2) and (3),

$$(5) \quad \psi^*(y, \xi; \beta, \varphi) = \{y - m^*(\xi; \beta)\} v^*(\xi; \beta, \varphi)^{-1} \frac{\partial m^*(\xi; \beta)}{\partial \beta}.$$

For notational simplicity, we often omit the arguments in the functions  $m^*$ ,  $v^*$ , etc. We also denote a derivative with respect to a variable  $z$ , say, by using the subscript  $z$ . Then (5) can also be written as  $\psi^* = (y - m^*)v^{*-1}m_\beta^*$ .

The score function  $\psi^*$  has to be supplemented by a score function for  $\varphi$ . For simplicity, let us assume that  $\varphi$  is known, unless otherwise stated. Thus  $\varphi$  may be omitted in the argument of  $\psi^*$ . The following results still hold true when  $\varphi$  is unknown and needs to be estimated.

The score function  $\psi^*$  is unbiased, which means that  $\mathbf{E} \psi^*(y, \xi; \beta) = 0$ , where the expectation is taken for the same  $\beta$  as the  $\beta$  in the argument. As a consequence, the estimator  $\hat{\beta}^*$  of  $\beta$  which is based on  $\psi^*$  (i.e., the solution to  $\sum_{i=1}^n \psi^*(y_i, \xi_i; \hat{\beta}^*) = 0$ ) is consistent. But as  $\xi$  is unobservable, this estimator is not feasible and cannot be used as such.

If the latent variable  $\xi$  is replaced with the surrogate  $x$ , we get an estimating function  $\psi^*(y, x; \beta)$ , which can be used to construct the *naive* (N) estimator  $\hat{\beta}_N$  as the solution to the equation  $\sum_{i=1}^n \psi^*(y_i, x_i; \hat{\beta}_N) = 0$ . Since  $\psi^*$ , as a function of  $y$  and  $x$ , is not unbiased, the resulting estimator is inconsistent. Nevertheless, one can study its (asymptotic) bias and its (asymptotic) variance [14].

In order to construct consistent estimators, we typically need to have some additional pieces of information. Here we assume that the measurement error covariance matrix  $\Sigma_\delta$  is known. We distinguish between two types of estimators, functional and structural ones, [2]. The latter type makes use of the distribution of  $\xi$ , which therefore must be given, at least up to the unknown parameter vector  $\gamma$ . The former does not need the distribution of  $\xi$  and works even when  $\xi$  is not random (functional variant).

Among the functional methods, we consider the *corrected score* (CS) estimator [22, 31]. An alternative method, particularly adapted to generalized linear models, is the Conditional Score method, [2, 32], which, however, we do not discuss in this survey. (Another one is SIMEX, which, however, is not always consistent but typically reduces the bias, [2, 7].) We want to construct an unbiased estimating function for  $\beta$  in the variables  $y$  and  $x$ . For this purpose, we need to find functions  $g_1$  and  $g_2$  of  $x$  and  $\beta$  such that

$$(6) \quad \mathbf{E}[g_1(x; \beta)|\xi] = \varphi v^{*-1} m_\beta^* = \eta_\beta,$$

$$(7) \quad \mathbf{E}[g_2(x; \beta)|\xi] = \varphi m^* v^{*-1} m_\beta^* = c'(\eta) \eta_\beta.$$

Then

$$(8) \quad \psi_C(y, x; \beta) = y g_1(x; \beta) - g_2(x; \beta)$$

is termed the corrected score function. Because  $\mathbf{E}(\psi_C|y, \xi) = \varphi \psi^*$  and  $\mathbf{E} \psi^* = 0$ ,  $\psi_C$  is unbiased and can therefore be used to construct a consistent estimator  $\hat{\beta}_C$  of  $\beta$  as the

solution of

$$\sum_{i=1}^n \psi_C(y_i, x_i; \hat{\beta}_C) = 0.$$

The functions  $g_1$  and  $g_2$  do not always exist. [31] gives the conditions for their existence and shows how to find them if they exist.

Among the structural methods, we consider the *quasi score* (QS) and the *maximum likelihood* (ML) estimators. (The *regression calibration* (RC) estimator [2] is not consistent in general, although it often reduces the bias significantly.)

For QS, we construct the (obviously unbiased) quasi-score function which is very similar to (5) but with

$$\begin{aligned} m(x; \beta) &:= \mathbf{E}(y|x), \\ v(x; \beta) &:= \mathbf{V}(y|x) \end{aligned}$$

in place of  $m^*(\xi; \beta)$  and  $v^*(\xi; \beta)$ , respectively:

$$(9) \quad \psi_Q(y, x; \beta) = [y - m(x; \beta)] v(x; \beta)^{-1} m_\beta(x; \beta).$$

Again we drop the parameter  $\varphi$ , considering it to be known. We also suppress the nuisance parameter  $\gamma$  in the argument of the functions  $m$  and  $v$ , although  $m$  and  $v$  depend on  $\gamma$ . Indeed, in order to compute  $m$  and  $v$ , we need the conditional distribution of  $\xi$  given  $x$ , which depends on the distribution of  $\xi$  with its parameter  $\gamma$ . For instance, assume that  $\xi \sim N(\mu_\xi, \Sigma_\xi)$ , where the elements of  $\mu_\xi$  and  $\Sigma_\xi$  make up the components of the parameter vector  $\gamma$ . Then  $\xi|x \sim N(\mu(x), T)$  with

$$(10) \quad \mu(x) = \mu_\xi + \Sigma_\xi(\Sigma_\xi + \Sigma_\delta)^{-1}(x - \mu_\xi),$$

$$(11) \quad T = \Sigma_\delta - \Sigma_\delta(\Sigma_\xi + \Sigma_\delta)^{-1}\Sigma_\delta;$$

see [29] and, for an extension to a mixture of normals, [25]. Very often, the first component of the vector  $x$  is the dummy variable 1 and the first component of  $\delta$  is 0. Then the first row and column of  $\Sigma_\delta$  and also of  $\Sigma_\xi$  are 0. In this case, the inverse of  $\Sigma_\xi + \Sigma_\delta$  is to be replaced with the Moore–Penrose generalized inverse. The matrix  $T$  then also has zeros in the first row and column, and the first component of  $\mu(x)$  is 1.

Given the conditional distribution of  $\xi|x$ , one can compute  $m$  and  $v$  starting from the original mean and variance functions,  $m^*$  and  $v^*$ , of the error-free model:

$$\begin{aligned} m(x; \beta) &= \mathbf{E}[m^*(\xi; \beta)|x], \\ v(x; \beta) &= \mathbf{V}[m^*(\xi; \beta)|x] + \mathbf{E}[v^*(\xi; \beta)|x]. \end{aligned}$$

The quasi-score function (9) constructed in this way is, of course, unbiased and thus produces a consistent QS estimator  $\hat{\beta}_Q$  as the solution to the estimating equation  $\sum_{i=1}^n \psi_Q(y_i, x_i; \hat{\beta}_Q) = 0$ . The equation has a unique solution for large  $n$ , but it may have multiple roots if  $n$  is not large. Heyde and Morton (1998) develop methods to deal with this case.

*Maximum likelihood* (ML) is based on the joint density of  $x$  and  $y$ , which is given (again omitting  $\varphi$  and  $\gamma$ ) by

$$q(y, x; \beta) = \int f(y|\xi; \beta)g(x|\xi)h(\xi) d\xi,$$

where  $g$  is the density of  $N(\xi, \Sigma_\delta)$ . Alternatively, one can express  $q(y, x)$  as

$$q(y, x; \beta) = \int f(y|\xi; \beta)p(\xi|x) d\xi k(x),$$

where  $k$  is the density of  $x$  and may be omitted if  $\gamma$  is known or is estimated in advance and  $p$  is the conditional density of  $\xi$  given  $x$ . Thus while QS relies only on the error-free mean and variance functions, ML relies on the whole error-free model distribution. Therefore, ML is more sensitive than QS with respect to a potential model misspecification because QS is always consistent as long as at least the mean function (along with the density of  $\xi$ ) has been correctly specified. In addition, the likelihood function is generally much more difficult to compute than the quasi-score function. This often justifies the use of the relatively less efficient QS instead of the more efficient ML method.

4. EXAMPLES

A few examples will illustrate the concepts introduced in the previous sections [2, 19].

4.1. **Polynomial model.** The polynomial model for the scalar  $\xi$  is given by, [4],

$$y = \beta_0 + \beta_1\xi + \dots + \beta_k\xi^k + \varepsilon,$$

where  $\varepsilon \sim N(0, \sigma_\varepsilon^2)$  and  $\varepsilon$  is independent of  $\xi$ . Here  $\eta = \sum_0^k \beta_r \xi^r$ ,  $c(\eta) = \frac{1}{2}\eta^2$ , and  $\varphi = \sigma_\varepsilon^2$ .

To construct the CS function, we first need to find functions  $t_r(x)$  such that

$$E[t_r(x)|\xi] = \xi^r.$$

It turns out that  $t_r(x)$  is a polynomial of degree  $r$  under the normal  $\delta$  given by [23]

$$t_r(x) = \sum_{j=0}^r \binom{r}{j} \mu_j^+ x^{r-j}$$

with

$$\mu_j^+ := \begin{cases} 0 & \text{if } j \text{ is odd,} \\ (j-1)!! (-1)^{j/2} \sigma_\delta^j & \text{if } j \text{ is even,} \end{cases}$$

where  $(j-1)!!$  is a notation for  $1 \cdot 3 \cdot \dots \cdot (j-1)$  and  $(-1)!! = 1$ . Alternatively,  $t_r(x)$  can be computed recursively by the recursion formula [31]

$$t_{r+1}(x) = t_r(x)x - r t_{r-1}(x) \sigma_\delta^2, \quad t_0(x) = 1, \quad t_{-1}(x) = 0.$$

In fact,  $t_r(x) = H_r(x/\sigma_\delta) \sigma_\delta^r$ , where the  $H_r(x)$  are the Hermite polynomials. For computing  $t_r$  under nonnormal  $\delta$ , see [3]. The CS function can now be set up as follows:

$$\psi_C(y, x; \beta) = yt(x) - T(x)\beta,$$

where  $t(x) = (t_0(x), \dots, t_k(x))^T$  and  $T(x)$  is a  $(k+1) \times (k+1)$  matrix with

$$T_{rs}(x) = t_{r+s}(x), \quad r, s = 0, \dots, k.$$

Thus for the polynomial model,

$$g_1(x; \beta) = t(x), \quad g_2(x; \beta) = T(x)\beta.$$

For constructing the QS function, we first need to find  $\mu_r(x) := E(\xi^r|x)$ . We find for normal  $\xi$  that [23]

$$\mu_r(x) = \sum_{j=0}^r \binom{r}{j} \mu_j^* \mu_1(x)^{r-j},$$

with

$$\mu_j^* = \begin{cases} 0 & \text{if } j \text{ is odd,} \\ (j-1)!! \tau^j & \text{if } j \text{ is even,} \end{cases}$$

where  $\mu_1(x) = \mu(x)$  of (10) and  $\tau^2 = T$  of (11), both for scalar  $x$ . Alternatively,  $\mu_{r+1}(x)$  can be computed recursively by

$$\mu_{r+1}(x) = \mu_r(x)\mu_1(x) + r\mu_{r-1}(x)\tau^2, \quad \mu_0(x) = 1, \quad \mu_{-1}(x) = 0.$$

The mean and variance functions are now given by

$$\begin{aligned} m(x; \beta) &= \tilde{\mu}(x)^\top \beta, \\ v(x; \beta) &= \sigma_\varepsilon^2 + \beta^\top (M(x) - \tilde{\mu}(x)\tilde{\mu}(x)^\top) \beta, \end{aligned}$$

where  $\tilde{\mu}(x) = (\mu_0(x), \dots, \mu_k(x))^\top$  and  $M(x)$  is a  $(k+1) \times (k+1)$  matrix with

$$M_{rs}(x) = \mu_{r+s}(x), \quad r, s = 0, \dots, k.$$

The QS function can then be written as

$$\psi_Q(y, x; \beta) = (y - \tilde{\mu}(x)^\top \beta) v^{-1} \tilde{\mu}(x).$$

**4.2. Loglinear Poisson model.** Let  $y \sim Po(\lambda)$  with  $\lambda = \exp(\xi^\top \beta)$ . Then  $\eta = \log \lambda$ ,  $c(\eta) = e^\eta$  and  $\varphi = 1$ . The CS function is given by [29]

$$\psi_C(y, x; \beta) = yx - \exp\left(x^\top \beta - \frac{1}{2}\beta^\top \Sigma_\delta \beta\right) (x - \Sigma_\delta \beta),$$

with

$$g_1(x; \beta) = x, \quad g_2(x; \beta) = \exp\left(x^\top \beta - \frac{1}{2}\beta^\top \Sigma_\delta \beta\right) (x - \Sigma_\delta \beta).$$

For the QS function  $\psi_Q$ , we use the mean and variance functions [29]:

$$\begin{aligned} m(x; \beta) &= \exp\left(\beta^\top \mu(x) + \frac{1}{2}\beta^\top T \beta\right), \\ v(x; \beta) &= m(x; \beta) + [\exp(\beta^\top T \beta) - 1] m^2(x; \beta), \end{aligned}$$

where  $\mu(x)$  and  $T$  are given by (10) and (11).

**4.3. Loglinear Gamma model.** Let  $y \sim G(\mu, \nu)$ , i.e.,

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

with  $\mu = \exp(\beta_0 + \beta_1 \xi)$  and  $\xi$  scalar. (In the special case  $\nu = 1$ , we have the loglinear exponential model.) Here  $\eta = -\mu^{-1}$  and  $\varphi = \nu^{-1}$ . We have  $c(\eta) = -\log(-\eta)$ . For CS, we need to derive the functions  $g_1$  and  $g_2$  of (6) and (7). We find [16]

$$\begin{aligned} g_1(x; \beta) &= \exp\left(-\beta_0 - \beta_1 x - \frac{1}{2}\beta_1^2 \sigma_\delta^2\right) (1, x + \beta_1 \sigma_\delta^2)^\top, \\ g_2(x) &= (1, x)^\top. \end{aligned}$$

For QS, we need to compute  $m$  and  $v$ :

$$\begin{aligned} m(x; \beta) &= \exp\left\{\beta_0 + \beta_1 \mu_1(x) + \frac{1}{2}\beta_1^2 \tau^2\right\}, \\ v(x; \beta, \nu) &= \left\{\left(1 + \frac{1}{\nu}\right) \exp(\beta_1^2 \tau^2) - 1\right\} m(x; \beta)^2. \end{aligned}$$

4.4. **Logit model.** Let  $y \sim B(1, \pi)$ , i.e.,

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

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

Here

$$\eta = \log\left(\frac{\pi}{1 - \pi}\right) = \beta_0 + \beta_1\xi, \quad \varphi = 1,$$

and  $c(\eta) = \log(1 + e^\eta)$ .

For CS, we need to find functions  $g_1$  and  $g_2$  such that

$$\begin{aligned} \mathbb{E}[g_1(x)|\xi] &= (1, \xi)^\top, \\ \mathbb{E}[g_2(x; \beta)|\xi] &= \{1 + \exp(-\beta_0 - \beta_1\xi)\}^{-1}(1, \xi)^\top. \end{aligned}$$

Obviously  $g_1(x) = (1, x)^\top$ . But, according to [31],  $g_2(x; \beta)$  does not exist in general. However, if  $(\beta_0, \beta_1, \xi)$  can be restricted so that  $\beta_0 + \beta_1\xi > 0$  (sometimes known as a “rare event” restriction, [1]), then a corrected score function exists. It can be evaluated with the help of a Taylor series expansion of the logistic function.

Indeed, with  $z = \beta_0 + \beta_1x$ ,

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

which is absolutely convergent if, and only if,  $z > 0$ . The function  $g_2$  is then given by

$$g_2(x) = \sum_{k=0}^{\infty} (-1)^k \exp\left\{-k(\beta_0 + \beta_1x) - \frac{k^2}{2}\beta_1^2\sigma_\delta^2\right\} \begin{pmatrix} 1 \\ x + k\beta_1\sigma_\delta^2 \end{pmatrix}.$$

For QS, we need to construct

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

There is no closed form expression for  $m(x; \beta)$ . The expected value has to be computed by numerical integration, [8, 21]. 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

$$(1 + e^{-\eta})^{-1}$$

is closely approximated by  $\Phi(\eta/c)$ , where  $\Phi$  is the standard normal distribution function and  $c = 1.70$  [12, Chapter 22]. Thus assume that  $\pi = \Phi\{c^{-1}(\beta_0 + \beta_1\xi)\}$ . Then

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

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

## 5. ASYMPTOTIC COVARIANCE MATRIX (ACM)

Under general assumptions, the CS and QS estimators of  $\beta$  exist and are unique (at least for large enough  $n$  and with probability going to 1). For CS and QS,  $\hat{\beta}$  is consistent

and  $\sqrt{n}(\hat{\beta} - \beta)$  is asymptotically normal with an asymptotic covariance matrix (ACM), which is given by a sandwich formula of the form

$$(12) \quad \Sigma = A^{-1}BA^{-\top},$$

$$(13) \quad A = -\mathbf{E} \frac{\partial \psi}{\partial \beta^\top},$$

$$(14) \quad B = \mathbf{E} \psi \psi^\top,$$

where  $\psi$  is either  $\psi_C$  or  $\psi_Q$  (or some other estimating function) depending on the estimator considered. For QS, the matrices  $A$  and  $B$  are equal, and the sandwich formula for  $\Sigma_Q$  simplifies to  $\Sigma_Q = B_Q^{-1}$ . If  $\varphi$  has to be estimated along with  $\beta$ , the formula for  $\Sigma$  does not change, but if nuisance parameters  $\gamma$  are present and have been pre-estimated, the ACM of  $\hat{\beta}_Q$  has to be supplemented by an additional term, i.e.,

$$(15) \quad \Sigma_Q = B_Q^{-1} + B_Q^{-1}A_\gamma \Sigma_\gamma A_\gamma^{-\top} B_Q^{-1},$$

where  $\Sigma_\gamma$  is the ACM of  $\hat{\gamma}$  and  $A_\gamma = -\mathbf{E} \partial \psi_Q / \partial \gamma^\top$ .

Usually there are no closed-form expressions for the expectations in (13) and (14), but for the Poisson model, a closed-form expression for the ACM of CS exists (see Shklyar and Schneeweiss, 2005).

We briefly discuss some of the assumptions underlying these results [13]. A typical assumption often made in nonlinear models is the requirement that  $\beta$  is an interior point of a given compact set. Furthermore, the functions  $c(\eta)$  and  $\eta(\xi; \beta)$  should be sufficiently smooth and their derivatives should be exponentially bounded with respect to  $\xi$ . This guarantees the existence of the conditional expectations introduced above and the interchangeability of taking expectations and going to limits in the parameter space. An important condition, which guarantees the identifiability of  $\beta$ , is the following strengthening of the unbiasedness property of  $\psi^*$ . Not only do we require that  $\mathbf{E} \psi^*(y, \xi; \beta) = 0$ , but even more that  $\beta$  is the unique solution,  $b = \beta$ , to  $\mathbf{E} \psi^*(y, \xi; b) = 0$ , where the expectation is taken with respect to the true value of  $\beta$ . A similar assumption is made with regard to  $\psi_Q$ . Finally, it is required that the matrix  $\mathbf{E} m_\beta m_\beta^\top$  is positive definite. In the linear model, this is equivalent to the familiar assumption that  $\mathbf{E} x x^\top$  is positive definite. With these and some more assumptions the results described in the next section can be proved.

## 6. EFFICIENCY COMPARISON

We compare the relative efficiencies of  $\hat{\beta}_C$  and  $\hat{\beta}_Q$  by comparing their ACMs. In case there are no nuisance parameters, it turns out that

$$(16) \quad \Sigma_C \geq \Sigma_Q$$

in the sense of the Loewner order for symmetric matrices. This can be shown by noting that  $\psi_C$  and  $\psi_Q$  are both linear in  $y$  and that  $\psi_Q$  is optimal within the class of linear-in- $y$  unbiased estimating functions. Indeed, this follows from a general criterion of [10], which, if applied to the present case, states that  $\Sigma_C \geq \Sigma_Q$  if, and only if,  $(\mathbf{E} \psi_{C\beta})^{-1} \mathbf{E} \psi_C \psi_Q^\top$  does not depend on  $\beta$ , and this independence can be verified. However, one can refine inequality (16). One can construct a *simple score* (SS) estimator through a so-called simple score function given by

$$\psi_S(y, x; \beta) = [y - m(x; \beta)]g_1(y; \beta),$$

with no involvement of the variance function  $v(x, \beta)$ , and can show [16] that for the corresponding ACMs,

$$(17) \quad \Sigma_C \geq \Sigma_S \geq \Sigma_Q.$$



To show that  $\Sigma_S \geq \Sigma_Q$ , one compares

$$\Sigma_S = (\mathbf{E} g_1 m_\beta^\top)^{-1} \mathbf{E} v g_1 g_1^\top (\mathbf{E} g_1 m_\beta^\top)^{-\top}$$

to

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

and uses a matrix variant of the Cauchy–Schwarz inequality. To prove the other inequality  $\Sigma_C \geq \Sigma_S$ , one notes that  $A_C = A_S$  and

$$B_C - B_S = \mathbf{E}(m g_1 - g_2)(m g_1 - g_2)^\top \geq 0,$$

from which  $\Sigma_C \geq \Sigma_S$  follows immediately. In both cases, one can find conditions under which the  $\geq$  sign can be replaced with  $=$ ,  $\neq$ , or  $>$  signs [15]. Just to cite a few results: if the components of  $m g_1 - g_2$  are linearly independent as functions of  $x$ , then  $\Sigma_C > \Sigma_S$ . In the polynomial model (see section 4.1),  $\Sigma_C = \Sigma_Q$  if  $\beta_{(0)} = 0$  and  $\Sigma_C > \Sigma_Q$  if  $\beta_{(0)} \neq 0$ , where  $\beta_{(0)} = (\beta_1, \dots, \beta_k)$  [30]. In the Poisson model (see section 4.2),  $\Sigma_C = \Sigma_Q$  if  $\Sigma_\delta \beta = 0$ , and  $\Sigma_C > \Sigma_Q$  if  $\Sigma_\delta \beta \neq 0$  [29].

These results hold true if the nuisance parameters  $\gamma$  are known. If, however, they have to be estimated in advance,  $\Sigma_Q$  is given by (15), and (16) or (17) need not be true any more. In the linear model, CS and QS coincide if the nuisance parameters  $\mu_\xi$  and  $\sigma_\xi^2$  are pre-estimated, and so  $\Sigma_C = \Sigma_Q$  in this case. For the Poisson model, [28] shows that  $\Sigma_C \geq \Sigma_Q$  even if  $\mu_\xi$  and  $\sigma_\xi^2$  have to be estimated. In the polynomial as well as in the Poisson model, we still have  $\Sigma_S \geq \Sigma_Q$  because the additional term in the ACMs of  $\hat{\beta}_S$  and  $\hat{\beta}_Q$ , which is due to the estimation of  $\gamma$ , see (15), is the same for both the estimators. Recently, it has been shown that QS can be modified so that, in general,  $\Sigma_C \geq \Sigma_Q$  when  $x \sim N(\mu_x, \sigma_x^2)$ . This result is true even if the nuisance parameters  $\mu_x$  and  $\sigma_x^2$  are unknown and have to be estimated, but they must be estimated together with  $\beta$ , and not in advance, [18, 19]. A multivariate version of this approach has been applied to estimate a zero-inflated Poisson measurement error model [20].

Considering the naive estimator  $\hat{\beta}_N$ , which is (asymptotically) biased, one might expect that  $\Sigma_C \geq \Sigma_N$  because CS corrects for the bias and therefore loses efficiency relative to  $N$ . However, in the polynomial model, there are cases where  $\Sigma_C - \Sigma_N$  is indefinite [17].

In [13], we derive a number of results regarding the relative efficiencies of estimators when the measurement errors are small. We consider only the scalar case ( $\delta$  one-dimensional), although our results can be extended to the vector case. We prove that

$$\Sigma_C - \Sigma_Q = O(\sigma_\delta^4)$$

if  $\sigma_\delta^2 \rightarrow 0$  and all other parameters are kept fixed, regardless whether nuisance parameters are present or not. We show this by expanding  $\Sigma_C$  and  $\Sigma_Q$  in terms of powers of  $\sigma_\delta^2$ . In general, one cannot go beyond the fourth power of  $\sigma_\delta$ : e.g., for the Poisson model, the terms of order  $\sigma_\delta^4$  differ in the expansions of  $\Sigma_C$  and  $\Sigma_Q$  [29]. For the naive estimator, the difference of  $\Sigma_N$  and  $\Sigma_Q$  (or  $\Sigma_C$ ) is of the order  $\sigma_\delta^2$ , not  $\sigma_\delta^4$  [14]. Similarly, the differences  $\Sigma_Q - \Sigma^*$  and  $\Sigma_C - \Sigma^*$  are of order  $\sigma_\delta^2$ , where  $\Sigma^*$  is the ACM of the ML estimator  $\hat{\beta}^*$  in the error-free model.

In another approach, [14], we show that different results emerge if, along with  $\sigma_\delta^2$ ,  $\varphi$  also tends to zero such that their ratio remains constant. In this case,  $\Sigma_C$  and  $\Sigma_Q$  differ at the order of  $\sigma_\delta^2$  such that  $\Sigma_C - \Sigma_Q$  is positive definite at that order. At the same time,  $\Sigma_N - \Sigma_C = O(\sigma_\delta^4)$ . Thus for small  $\sigma_\delta^2$  and  $\varphi$ ,  $\Sigma_N - \Sigma_Q$  is positive definite, despite the fact that  $\hat{\beta}_N$  is biased and  $\hat{\beta}_Q$  is not.

In the efficiency comparison of CS and QS, it seems that QS comes out best, at least when  $\sigma_\delta^2$  is not too large, so that the influence of the nuisance parameters is not yet felt.

But one must keep in mind that QS (just as SS and ML) relies on the knowledge of the distribution of  $\xi$ . If this distribution is misspecified, QS is (asymptotically) biased. [26] investigates this bias by studying a distribution  $h(\xi)$  which is wrongly assumed to be a normal distribution, although it is, in fact, a mixture of normals. Suppose the true distribution is a mixture of two normals with equal variances but different means. When the difference  $\vartheta$  of the two means tends to zero, the bias of  $\hat{\beta}_Q$  is of the order  $\vartheta^2$  and therefore most often negligible. But when the two means differ a lot and one of the components of the mixture has a weight  $p$  that tends to zero, the bias is of the order  $p$  and therefore not negligible.

## 7. CONCLUSION

We have focused our review on two popular estimation methods for measurement error models with known measurement error variance (or covariance matrix): the functional CS and the structural QS methods.

If the regressor distribution is known, then QS is more efficient than CS. This result is plausible, as QS uses more information than CS. But it is by no means self-evident, as QS is not ML. Indeed, when the regressor distribution has unknown (nuisance) parameters, which need to be estimated in advance, then this result is no more universally valid, although the superiority of QS can still be claimed in many cases. One can, however, modify the QS procedure so that the nuisance parameters are not pre-estimated, but are estimated jointly with the regression parameter  $\beta$ . Under this modification, QS is more efficient than CS.

But there are reasons why CS might be preferred to QS in some cases. First of all, for small measurement errors, the efficiency difference between CS and QS becomes almost negligible in the sense that the difference of the ACMs is of the order  $\sigma_\delta^4$ . Furthermore, the QS estimator will typically be biased if the regressor distribution, on which it relies, has been misspecified. In some cases this bias may be negligible, but in other cases it is relevant. In those other cases one can try to rectify the misspecification by specifying a mixture of normals for the regressor distribution. This, however, introduces more parameters with the danger that QS will lose efficiency. It is an open question whether, despite of this loss of efficiency, QS is still more efficient than CS.

It should be mentioned that all the results presented in this paper refer to large sample properties. Typically CS does not perform well in small samples. But, at least for the polynomial model, one can modify the CS estimator so that it becomes more efficient in small samples [5].

Weighing the pros and cons of both methods, it seems that QS comes out better than CS in general unless one has virtually no idea about the underlying regressor distribution, in which case CS is to be preferred.

## ACKNOWLEDGMENT

Support by the Deutsche Forschungsgemeinschaft (German Research Foundation) is gratefully acknowledged. A. Kukush is supported also by the Swedish Institute grant SI-01424/2007.

## BIBLIOGRAPHY

1. J. S. Buzas and L. A. A. Stefanski, *Note on corrected-score estimation*, *Statistics and Probability Letters* **28** (1996), 1–8. MR1394413 (97d:62144)
2. R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu, *Measurement Error in Non-linear Models*, Second edition, Chapman and Hall, London, 2006. MR2243417 (2007e:62004)
3. C.-L. Cheng and H. Schneeweiss, *Polynomial regression with errors in the variables*, *Journal of the Royal Statistical Society* **B 60** (1998), 189–199. MR1625632

4. C.-L. Cheng and H. Schneeweiss, *On the polynomial measurement error model*, Total Least Squares and Errors-in-Variables Modeling (S. van Huffel and P. Lemmerling, eds.), Kluwer, Dordrecht, 2002, pp. 131–143. MR1952942
5. C.-L. Cheng, H. Schneeweiss, and M. Thamerus, *A small sample estimator for a polynomial regression with errors in the variables*, Journal of the Royal Statistical Society **62** (2000), 699–709. MR1796286
6. C.-L. Cheng and J. W. Van Ness, *Statistical Regression with Measurement Error*, Arnold, London, 1999. MR1719513 (2001k:62001)
7. J. Cook and L. A. Stefanski, *Simulation-extrapolation estimation in parametric measurement error models*, Journal of the American Statistical Association **89** (1994), 1314–1328.
8. E. A. Crouch and D. Spiegelman, *The evaluation of integrals of the form  $\int_{-\infty}^{\infty} f(t) \exp(-t^2) dt$ : application to logistic-normal models*, Journal of the American Statistical Association **85** (1990), 464–467. MR1141749 (92h:65032)
9. W. A. Fuller, *Measurement Error Models*, Wiley, New York, 1987. MR898653 (89a:62160)
10. C. C. Heyde, *Quasi-Likelihood and its Application*, Springer, New York, 1997. MR1461808 (99f:62003)
11. C. C. Heyde and R. Morton, *Multiple roots in general estimating equations*, Biometrika **85** (1998), 967–972.
12. N. L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Univariate Distributions*, vol. 2, Houghton Mifflin Co., Boston, MA, 1970. MR0270476 (42:5364)
13. A. Kukush and H. Schneeweiss, *Comparing different estimators in a nonlinear measurement error model, I*, Mathematical Methods of Statistics **14** (2005a), 53–79. MR2158071 (2006j:62068a)
14. A. Kukush and H. Schneeweiss, *Comparing different estimators in a nonlinear measurement error model, II*, Mathematical Methods of Statistics **14** (2005b), 203–223. MR2160395 (2006j:62068b)
15. A. Kukush and H. Schneeweiss, *Asymptotic optimality of the quasi-score estimator in a class of linear score estimators*, Discussion Paper, [www.stat.uni-muenchen.de/sfb386](http://www.stat.uni-muenchen.de/sfb386), vol. 477, University of Munich, 2006.
16. A. Kukush, H. Schneeweiss, and S. Shklyar, *Quasi Score is more efficient than Corrected Score in a general nonlinear measurement error model*, Discussion Paper, [www.stat.uni-muenchen.de/sfb386](http://www.stat.uni-muenchen.de/sfb386), vol. 451, University of Munich, 2005a.
17. A. Kukush, H. Schneeweiss, and R. Wolf, *Relative efficiency of three estimators in a polynomial regression with measurement errors*, Journal of Statistical Planning and Inference **127** (2005b), 179–203. MR2103032 (2005h:62087)
18. A. Kukush, A. Malenko, and H. Schneeweiss, *Optimality of the quasi-score estimator in a mean-variance model with applications to measurement error models*, Discussion Paper, [www.stat.uni-muenchen.de/sfb386](http://www.stat.uni-muenchen.de/sfb386), vol. 494, University of Munich, 2006.
19. A. Kukush, A. Malenko, and H. Schneeweiss, *Comparing the efficiency of estimates in concrete errors-in-variables models under unknown nuisance parameters*, Theory of Stochastic Processes **13(29)** (2007a), 69–81. MR2482252 (2010b:62256)
20. A. Kukush, A. Malenko, H. Schneeweiss, and Shalabh, *Optimality of Quasi-Score in the multivariate mean-variance model with application to the zero-inflated Poisson model with measurement errors*, Discussion Paper, [www.stat.uni-muenchen.de/sfb386](http://www.stat.uni-muenchen.de/sfb386), vol. 498, University of Munich, 2007b.
21. J. Monahan and L. A. Stefanski, *Normal scale mixture approximations to  $F^*(z)$  and computation of the logistic-normal integral*, Handbook of the Logistic Distribution (N. Balakrishnan, ed.), Marcel Dekker, New York, 1992, pp. 529–540.
22. T. Nakamura, *Corrected score functions for errors-in-variables models: Methodology and application to generalized linear models*, Biometrika **77** (1990), 127–137. MR1049414 (91e:62179)
23. H. Schneeweiss, *Quasi score and corrected score estimation in the polynomial measurement error model*, Recent Advances in Linear Models and Related Areas (Shalabh and C. Heumann, eds.), Physica, Heidelberg, 2008, pp. 45–58. MR2523524
24. H. Schneeweiss and T. Augustin, *Some recent advances in measurement error models and methods*, Allgemeines Statistisches Archiv **90** (2006), 183–197. MR2255581
25. H. Schneeweiss and C.-L. Cheng, *Bias of the quasi-score estimator of a measurement error model under misspecification of the regressor distribution*, Discussion Paper, [www.stat.uni-muenchen.de/sfb386](http://www.stat.uni-muenchen.de/sfb386), vol. 340, University of Munich, 2003.
26. H. Schneeweiss and C.-L. Cheng, *Bias of the structural quasi-score estimator of a measurement error model under misspecification of the regressor distribution*, Journal of Multivariate Analysis **97** (2006), 455–473. MR2234032

27. H. Schneeweiss and H. J. Mittag, *Lineare Modelle mit fehlerbehafteten Daten*, Physica-Verlag, Heidelberg, 1986. MR914793 (89e:62084)
28. S. Shklyar, *A comparison of estimates in the Poisson regression model with measurement errors*, Bulletin of the University of Kyiv, Seria Physics and Mathematics **3** (2006), 60–67.
29. S. Shklyar and H. Schneeweiss, *A comparison of asymptotic covariance matrices of three consistent estimators in the Poisson regression model with measurement errors*, Journal of Multivariate Analysis **94** (2005), 250–270. MR2167914
30. S. Shklyar, H. Schneeweiss, and A. Kukush, *Quasi Score is more efficient than Corrected Score in a polynomial measurement error model*, Metrika **65** (2007), 275–295. MR2299552 (2008h:62071)
31. L. A. Stefanski, *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** (1989), 4335–4358. MR1046712 (91e:62064)
32. L. A. Stefanski and R. J. Carroll, *Conditional scores and optimal scores in generalized linear measurement error models*, Biometrika **74** (1987), 703–716. MR919838 (89h:62048)
33. T. Wansbeek and E. Meijer, *Measurement Error and Latent Variables in Econometrics*, Elsevier, Amsterdam, 2000. MR1804397 (2002i:62184)

DEPARTMENT OF STATISTICS, UNIVERSITY OF MUNICH, AKADEMIESTRASSE 1, MUNICH 80799, GERMANY

*E-mail address:* `schneew@stat.uni-muenchen.de`

KYIV NATIONAL TARAS SHEVCHENKO UNIVERSITY, VOLODYMYRSKA STREET, 64, KYIV 01033, UKRAINE

*E-mail address:* `alexander_kukush@univ.kiev.ua`

Received 10/10/2008

Originally published in English