LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386

Schneeweiß, Kukush:

Comparing the efficiency of structural and functional methods in measurement error models

Sonderforschungsbereich 386, Paper 491 (2006)

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

Projektpartner

Comparing the efficiency of structural and functional methods in measurement error models

Hans Schneeweiss and Alexander Kukush

Abstract

The paper is a survey of recent investigations by the authors and others into the relative efficiencies of structural and functional estimators of the regression parameters in a measurement error model. While 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, discards such knowledge and works 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 have to be estimated, this is no more 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 the authors and others that deal with the relative efficiencies of structural and functional estimation methods. The first ones take the regressor distribution into account, the latter ones do not rely on this distribution. We want to review these results focussing mainly on two estimators: the structural quasi-score (QS) estimator and the functional corrected score (CS) estimator. Both are consistent, but differ in their asymptotic covariance matrices (ACMs). The most important result will be that, generally speaking, QS is more efficient than CS, but that this property may become invalid when nuisance parameters describing the regressor distribution need to be estimated. The result has also to be qualified when the regressor distribution is misspecified because then QS becomes biased. If the measurement errors are small QS and CS are essentially equally efficient.

We briefly also mention other estimators. For a recent review on the broader field of measurement error models, see Schneeweiss and Augustin (2006). Books on measure and error models are Schneeweiss and Mittag (1986), Fuller (1987), Carroll *et al.* (1995), Cheng and Van Ness (1999), and Wansbeek and Meijer (2000).

Section 2 introduces the measurement error model and Section 3 the estimators we want to consider. Section 4 has some examples. In Section 5 we introduce the asymptotic covariance matrix and discuss a few "technical" assumptions needed to derive the asymptotic properties of the estimators. The main Section 6 reviews the various efficiency results, which are interpreted in the Conclusion.

2 The model

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

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

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

with

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

where β is the regression parameter vector and φ a scalar dispersion parameter such that $\theta = (\beta^T,\varphi)^T$ and a, c, and η are known functions, cf. Kukush and Schneeweiss (2005). This class comprises the class of generalized linear models, where $\eta = \eta(\xi^T\beta)$. The conditional mean and conditional variance of y given ξ are, respectively,

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

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

The conditional mean function m^* is the regression function to be estimated. Although we started from model (1) and derived the mean and variance functions (2) and (3), we could have also started 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 ξ by a measurement error variable δ, which is independent of ξ and y:

$$x = \xi + \delta$$

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

The parameter θ has to be estimated with the help of observable data $(x_i, y_i), i = 1, \ldots, n$. We assume that (y_i, ξ_i, δ_i) are i.i.d. variables.

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

3 Estimators

If the variable ξ were observable, one could estimate β (and also φ) by maximum likelihood. The corresponding likelihood-score function for β is given by

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

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

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

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

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

If the latent variable ξ is replaced with the surrogate x, we get an estimating function $\psi^*(y, x; \beta)$, which can be used to construct the so-called *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 ψ^* 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, cf. Kukush and Schneeweiss (2005).

In order to construct consistent estimators we typically need to be given some additional pieces of information. Here we assume that the measurement error covariance matrix Σ_δ is known. We distinguish between two types of estimators, functional and structural ones. The latter make use of the distribution of ξ, which therefore must be given, at least up to the unknown parameter vector γ. The former does not need the distribution of ξ and works even when ξ is not random (functional variant).

Among the functional methods, we consider the *corrected score* (CS) estimator, cf. Nakamura (1990), Stefanski (1989). (Another one is SIMEX, which however is not always consistent, cf. Cook and Stefanski (1994)). We want to construct an unbiased estimating function for β in the variables y and x. To this purpose, we need to find functions g_1 and g_2 of x and β such that

$$\mathbb{E}\left[g_1(x; \beta)|\xi\right] = \varphi v^{*-1}\, m^*_\beta = \eta_\beta \tag{6}$$

$$\mathbb{E}\left[g_2(x; \beta)|\xi\right] = \varphi m^*\, v^{*-1}\, m^*_\beta = c'(\eta)\, \eta_\beta. \tag{7}$$

Then

$$\psi_C(y, x; \beta) = y g_1(x; \beta) - g_2(x; \beta) \tag{8}$$

is the so-called corrected score function. Because of $\mathbb{E}(\psi_C|y, \xi) = \varphi\psi^*$ and $\mathbb{E}\psi^* = 0$, ψ_C is unbiased and can therefore be used to construct a consistent

estimator $\hat{\beta}_C$ of β as the solution to $\sum_{i=1}^{n} \psi_C(y_i, x_i; \hat{\beta}_C) = 0$. The functions g_1 and g_2 do not always exist. Stefanski (1989) gives conditions for their existence and shows how to find them in case they exist.

Among the structural methods we consider the *quasi score* (QS) and the *maximum likelihood* (ML) estimators (The *regression calibration* (RC) estimator, cf. Caroll et al. (1995), is not consistent in general, although it often reduces the bias considerably).
For QS, we construct the (obviously unbiased) quasi-score function very similar to (5) but with

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

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

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

Again we dropped the parameter φ taking it to be known. In addition, we disregarded in the notation the dependence of m and v on the nuisance parameter γ describing the regressor distribution. Indeed, in order to compute m and v we need the conditional distribution of ξ given x, and this depends on the distribution of ξ with its parameter γ. If $\xi \sim N(\mu_\xi, \Sigma_\xi)$, then $\xi|x \sim N(\mu(x), T)$ with

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

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

cf. Shklyar and Schneeweiss (2005) and, for an extension to a mixture of normals, Schneeweiss and Cheng (2003). Very often the first component of the vector x is the dummy variable 1 and the first component of δ is 0. Then the first row and column of Σ_δ and also of Σ_ξ are 0. In this case the inverse of $\Sigma_\xi + \Sigma_\delta$ is to be understood as 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:

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

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

Maximum likelihood (ML) is based on the joint density of x and y, which is given (again omitting φ and γ) 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 a conditional expectation of the model density $f(y|\xi)$ given y and x:

$$q(y, x; \beta) = \mathbb{E}\left[f(y|\xi; \beta)|y, x\right] k(x),$$

where k is the density of X and may be omitted if γ is known or has been estimated in advance. Thus in contrast to QS, which relies on the conditional expectations only of the error-free mean and variance functions, m^* and v^*, ML relies on the conditional expectation of 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 ξ) 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 efficient ML method.

4 Examples

A few examples will illustrate the concepts introduced in the previous sections, cf. also Carroll *et al.* (1995).

4.1 Polynomial model

For scalar ξ the polynomial model is given by the equation, cf. Cheng and Schneeweiss (2002),

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

where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ and ε is independent of ξ. 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

$$\mathbb{E}\left[t_r(x)|\xi\right] = \xi^r.$$

If turns out that, under normal δ, $t_r(x)$ is a polynomial of degree r given by, cf. Schneeweiss (2005),

$$\begin{aligned} t_r(x) &= \sum_{j=0}^{r} \left(\begin{array}{c} r \\ j \end{array} \right) \mu_j^+ x^{r-j} \\ \mu_j^+ &:= \begin{cases} 0 & \text{if} \quad j \quad \text{is odd} \\ (j-1)!!(-1)^{\frac{j}{2}} \sigma_\delta^j & \text{if} \quad j \quad \text{is even,} \end{cases} \end{aligned}$$

where $(j-1)!!$ is short for $1 \cdot 2 \cdot 3 \cdots (j-1)$ and $(-1)!! = 1$. Alternatively, $t_r(x)$ can be computed recursively by the recursion formula, cf. Stefanski (1989),

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

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 non-normal δ, cf. Cheng and Schneeweiss (1998). The CS function can now be set up as follows

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

where $t(x) = (t_0(x), \ldots, t_k(x))^\top$ and $T(x)$ is a $(k+1) \times (k+1)$ matrix with $T_{rs}(x) = t_{r+s}(x),\ r, s = 0, \ldots, k$. 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 $\mu_r(x) := \mathbb{E}(\xi^r|x)$. For normal ξ, we find that, cf. Schneeweiss (2005),

$$\mu_r(x) = \sum_{j=0}^{r} \left(\begin{array}{c} r \\ j \end{array} \right) \mu_j^* \mu_1(x)^{r-j}$$

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

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 $\tilde{\mu}(x) = (\mu_0(x), \ldots, \mu_k(x))^\top$ and $M(x)$ is a $(k+1)\times(k+1)$ matrix with $M_{rs}(x) = \mu_{r+s}(x),\ r, s = 0, \ldots, k$. 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, cf. Shklyar and Schneeweiss (2005),

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

so that here

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

For the QS function ψ_Q, we use the mean and variance functions, cf. Shklyar and Schneeweiss (2005):

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

with $\mu(x)$ and T from (10) and (11).

4.3 Loglinear Gamma model

Let $y \sim G(\mu, \nu)$, i.e.,

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

with $\mu = \exp(\beta_0 + \beta_1 \xi)$, ξ scalar. (In the special case $\nu = 1$, we have the loglinear exponential model). Here $\eta = -\frac{1}{\mu}$ and $\varphi = \frac{1}{\nu}$. 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, cf. Kukush et al. (2005a),

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

For QS, we need to compute m and v:

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

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

Here $\eta = \log(\frac{\pi}{1-\pi}) = \beta_0 + \beta_1\xi$, $\varphi = 1$, 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 Stefanski (1989), $g_2(x;\beta)$ does not exist in general. However, if (β_0,β_1,ξ) can be restricted such that $\beta_0 + \beta_1\xi > 0$ (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 g_2 is then given by

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

For QS, we need to construct

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

There is no closed form expression for $m(x;\beta)$. The expected value has to be computed by numerical integration, 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 $(1+e^{-\eta})^{-1}$ is closely approximated by $\Phi(\eta/c)$, where Φ is the standard normal distribution function and $c = 1.70$, cf. Johnson and Kotz (1970, Chapter 22). Thus assume that $\pi = \Phi\{\frac{1}{c}(\beta_0 + \beta_1\xi)\}$. 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 rather general assumptions, the CS and QS estimators of β exist uniquely (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

$$\Sigma = A^{-1}BA^{-\top} \tag{12}$$

$$A = -\mathbb{E}\frac{\partial\psi}{\partial\beta^\top} \tag{13}$$

$$B = \mathbb{E}\psi\psi^\top, \tag{14}$$

where ψ is either ψ_C or ψ_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 Σ_Q simplifies to $\Sigma_Q = B_Q^{-1}$. If φ has to be estimated along with β, the formula for Σ does not change, but if nuisance parameters γ are present and have been pre-estimated, the ACM of $\hat{\beta}_Q$ has to be supplemented by an additional term, i.e.,

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

where Σ_γ is the ACM of $\hat{\gamma}$ and $A_\gamma = -\mathbb{E}\frac{\partial\psi_Q}{\partial\gamma^\top}$.

We briefly discuss some of the assumptions underlying these results, cf. Kukush and Schneeweiss (2005). A typical assumption often made in nonlinear models is the requirement that β is an interior point of a given compact set. Furthermore the functions $c(\eta)$ and $\eta(\xi, \beta)$ should be sufficiently smooth and its derivatives should be exponentially bounded with respect to ξ. 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 guaranteeing the identifiability of β is the following strengthening of the unbiasedness property of ψ^*. Not only do we require that $\mathbb{E}\psi^*(y, \xi; \beta) = 0$, but even more that β is the unique solution, $b = \beta$, to $\mathbb{E}\psi^*(y, \xi; b) = 0$, where the expectation is taken with respect to the true value of β. A similar assumption is made with regard to ψ_Q. Finally, it is required that the matrix $\mathbb{E}m_\beta m_\beta^\top$ is positive definite. In the linear model this is equivalent to the familiar assumption that $\mathbb{E}xx^\top$ 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. It turns out that

$$\Sigma_C \geq \Sigma_Q \tag{16}$$

in the sense of the Loewner order for symmetric matrices. This can be shown by noting that ψ_C and ψ_Q are both linear in y and that ψ_Q is optimal within the class of linear-in-y unbiased estimating functions. Indeed, this follows from a general criterion of Heyde (1997) which, if applied to the present case, states that $\Sigma_C \geq \Sigma_Q$ if, and only if, $(\mathbb{E}\psi_{C\beta})^{-1}\mathbb{E}\psi_C\psi_Q^\top$ does not depend on β, and this independence can be verified. However, one can say more. One can construct a *simple score* (SS) estimator through a so-called simple score function given by

$$\psi_S(y, x; \beta) = [y - m(x; \beta)]g_2(y; \beta)$$

and one can show, cf. Kukush *et al.* (2005a), that for the corresponding ACMs

$$\Sigma_C \geq \Sigma_S \geq \Sigma_Q. \tag{17}$$

In special cases, one can also give conditions under which one or both of the $\geq$ signs can be replaced with $=, \neq$, or $>$ signs, cf. Kukush *et al.* (2005a).

Just to cite one result: if the components of $mg_1 - g_2$ are linearly independent as function of x, then $\Sigma_C > \Sigma_S$, cf. Kukush and Schneeweiss (2006). 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 $\beta_{(0)} = (\beta_1, \ldots, \beta_k)$, cf. Shklyar *et al.* (2005). In the Poisson model (see section 4.2), $\Sigma_C = \Sigma_Q$ if $\Sigma_\delta \beta = 0$, and $\Sigma_C > \Sigma_Q$ if $\Sigma_\delta \beta \neq 0$, cf. Shklyar and Schneeweiss (2005).

These results hold true if the nuisance parameters γ are known. If, however, they have to be estimated in advance, Σ_Q is given by (15), and (16) or (17) need not to be true any more. In the linear model, CS and QS coincide if the nuisance parameters μ_ξ and σ_ξ^2 have been pre-estimated, and so $\Sigma_C = \Sigma_Q$ in this case. In the quadratic model, $\det\ \Sigma_C < \det\ \Sigma_S$ for sufficiently large σ_δ^2, although we still find that $\operatorname{diag} \Sigma_C \geq \operatorname{diag} \Sigma_S$, cf. Schneeweiss (2005). For the Poisson model, Shkylar (2006) shows that $\Sigma_C \geq \Sigma_Q$ even if μ_ξ and σ_ξ^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 $\hat{\beta}_Q$ due to the estimation of γ, see (15), is the same for both estimators. Recently, it has been shown that QS can be modified so that, in general, $\Sigma_C \geq \Sigma_Q$ when $x \sim N(\mu_x, \sigma_x^2)$ even if the nuisance parameters μ_x and σ_x^2 are unknown and have to be estimated, but they must be estimated together with β, not in advance, Kukush *et al.* (2006).

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, cf. Kukush *et al.* (2005b).

Kukush and Schneeweiss (2005) have a number of results regarding the relative efficiencies of estimators when the measurement errors are small. They consider only the scalar case (δ one-dimensional), although their results can be extended to the vector case. They prove that

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

if $\sigma_\delta^2 \to 0$ and all other parameters are kept fixed, regardless of whether nuisance parameters are present or not. They show this by expanding Σ_C and Σ_Q in term of powers of σ_δ^2. In general, one cannot go beyond the forth power of σ_δ: e.g., for the Poisson model, the terms of order σ_δ^4 differ in the expansions of Σ_C and Σ_Q, cf. Shkylar and Schneeweiss (2005). For the naive estimator, the difference of Σ_N and Σ_Q (or Σ_C) is of the order σ_δ^2, not σ_δ^4, cf. Kukush and Schneweiss (2005). Similarly, the differences $\Sigma_Q - \Sigma^*$ and

$\Sigma_C - \Sigma^*$ are of order σ_δ^2, where Σ^* is the ACM of the ML estimator $\hat{\beta}^*$ in the error free model.

In another approach, Kukush and Schneeweiss (2005) show that different results emerge if along with σ_δ^2 also φ tends to zero such that there ratio remains constant. In this case Σ_C and Σ_Q differ at the order of σ_δ^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 σ_δ^2 and φ, $\Sigma_N - \Sigma_Q$ 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 σ_δ^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) rely on the knowledge of the distribution of ξ. If this distribution is misspecified, QS is (asymptotically) biased. Schneeweiss and Cheng (2006) investigate this bias by studying a distribution $h(\xi)$ which is a mixture of normals. When the true distribution is a mixture of two normals with equal variances but different means, whereas the assumed distribution is just a normal distribution, and when the difference ϑ of the two means tends to zero, the bias of $\hat{\beta}_Q$ is of the order ϑ^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 focused our review on two quite popular estimation methods for measurement error models with known measurement error variance (or covariance matrix): the functional CS and the structural QS method.

If the regressor distribution is known, 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, 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 β. Under this modification,QS is more efficient than CS. But there are other reasons why CS might be preferred to QS. First of all, for small measurement errors the efficiency difference between CS and

QS becomes almost negligible in the precise sense that the difference of the ACMs is of the order σ_δ^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. Lastly, at least for the polynomial model, one can modify the CS estimator so that it becomes more efficient in small samples, cf. Cheng *et al.* (2000).

In summary, one cannot give a clear-cut advice on which estimator to use. But the present paper at least gives some hints to a well-founded choice.

Bibliography

- Buzas, J.S., Stefanski, L.A. (1996), A note on corrected-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.

- Cheng, C.-L., Schneeweiss, H. (1998), Polynomial regression with errors in the variables, *Journal of the Royal Statistical Society* **B 60**, 189–199.

- Cheng, C.-L., Schneeweiss, H. (2002), On the polynomial measurement error model. In *Total Least Squares and Errors-in-Variables Modeling* (S. van Huffel, P. Lemmerling, eds.), 131-143. Kluwer, Dordrecht.

- Cheng, C.-L., Schneeweiss, H., Thamerus, M. (2000), A small sample estimator for a polynomial regression with errors in the variables *Journal of the Royal Statistical Society* **B 62**, 699–709.

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

- Cook, J. and Stefanski, L.A. (1995), A simulation extrapolation method for parametric measurement error models, *Journal of the American Statistical Association* **89**, 1314–1328.

- 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.

- Fuller, W.A. (1987), *Measurement Error Models*. Wiley, New York.
- Heyde, C.C. (1997), *Quasi-Likelihood And Its Application*. Springer, New York, 1997.
- 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. (2006), *Asymptotic optimality of the quasi-score estimator in a class of liner score estimators*. Discussion Paper 477/SFB 386, University of Munich.
- Kukush, A., Schneeweiss, H., Shklyar, S. (2005a), *Quasi Score is more efficient than Corrected Score in a general nonlinear measurement error model*. Discussion Paper 451/SFB 386, University of Munich.
- Kukush, A., Schneeweiss, H., Wolf, R. (2005b), 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.
- Kukush, A., Malenko, A. Schneeweiss, H., *Asymptotic optimality of the quasi-score-like estimator in a mean-variance model*. Unpublished manuscript.
- Nakamura, T. (1990), Corrected score functions for errors-in-variables models: Methodology and application to generalized linear models, *Biometrika* **77**, 127-137.
- Schneeweiß, H. (2005), *The polynomial and the Poisson measurement error models: some further results on quasi score and corrected score estimation*. Discussion Paper 446/SFB 386, University of Munich.
- Schneeweiss, H., Augustin, T. (2006), Some recent advances in measurement error models and methods, *Allgemeines Statistisches Archiv* **90**, 183-197.
- 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, University of Munich.

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

- Schneeweiss, H., Mittag, H.J. (1986), *Lineare Modelle mit fehlerbehafteten Daten*. Physica-Verlag, Heidelberg.

- Shklyar, S. (2006), A comparison of estimates in the Poisson regression model with measurement errors, to appear in *Bulletin of the University of Kyiv, Seria Physics and Mathematics* **N3**.

- 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, University of Munich, published online in *Metrika*.

- 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.

- Wansbeek, T., Meijer, E. (2000) *Measurement Error and Latent Variables in Econometrics*, Elsevier, Amsterdam.