Relative Efficiency of Maximum Likelihood and Other Estimators in a Nonlinear Regression Model with Small Measurement Errors

Alexander Kukush
Department of Mathematics,
Taras Shevchenko University
01601 Kiew, Ukraine

Hans Schneeweiss
Department of Statistics,
University of Munich
Akademiestr. 1, 80799 München, Germany

Discussion Paper 396, 2004

Abstract

We compare the asymptotic covariance matrix of the ML estimator in a nonlinear measurement error model to the asymptotic covariance matrices of the CS and SQS estimators studied in Kukush et al (2002). For small measurement error variances they are equal up to the order of the measurement error variance and thus nearly equally efficient.

Keywords: Measurement Errors, Maximum Likelihood, Efficiency, Small Error Variance

1 Introduction

Kukush, Schneeweiss, and Wolf (2002), thereafter KSW, study the relative efficiency of three estimators in a nonlinear model of the exponential family

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

where $\xi = \xi(x,\beta)$ is a known function of the covariate X at the point x with an unknown parameter vector β. The variable X is measured with errors: $W = X + \sigma_u U, \quad U \sim N(0,1), \sigma_u U$ being the measurement error.

The naive estimator, which simply substitutes W for X in the original model and then uses Maximum Likelihood, is inconsistent.

Two, quite different, consistent estimators can be constructed when σ_u is known. The Corrected Score (CS) estimator starts from the likelihood score function $\psi(y,x;\beta,\varphi)$ of the model and corrects it by constructing a corrected score function $\psi_c(y,w;\beta,\varphi)$ such that

$$E\ \psi_c(Y,W;\beta,\varphi|Y,X) = \psi(Y,X;\beta,\varphi).$$

The estimator $\widehat{\beta}_{CS}$ (together with $\widehat{\varphi}_{CS}$) is then the solution to the equation

$$\sum_{i=1}^{n} \psi_c(y_i,w_i;\widehat{\beta}_{CS},\widehat{\varphi}_{CS}) = 0,$$

where $(y_i,w_i),\ i = 1,\ldots,n,$ is a sample of observations.

The Structural Quasi Score (SQS) estimator uses the distribution of X, which here is assumed to be a Gaussian distribution. The SQS procedure is based on the conditional mean and variance functions of Y given W:

$$E(Y|W) = m(W;\beta,\varphi)$$

$$V(Y|W) = v(W;\beta,\varphi),$$

from which a quasi score function for β and φ, is constructed, where the β-part is of the form

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

The estimator is then the solution to the equation

$$\sum_{i=1}^{n} \psi^*(y_i,w_i;\widehat{\beta}_{SQS},\widehat{\varphi}_{SQS}) = 0$$

together with a second equation for $\widehat{\varphi}$.

Both consistent estimators are asymptotically normally distributed with asymptotic covariance matrices Σ_{CS} and Σ_{SQS}, respectively. It is not known whether $\Sigma_{CS} - \Sigma_{SQS}$ is positive (semi)definite in general, i.e., whether $\widehat{\beta}_{SQS}$ is more efficient than $\widehat{\beta}_{CS}$, although this may be expected and has been proved in the special case of the Poisson model, see Shklyar and Schneeweiss (2002). However, KSW were able to show that for small measurement error variances σ_u^2, both estimators are nearly equally efficient. More precisely: their covariance matrices differ only by a term of order σ_u^4 as $\sigma_u^2 \to 0$:

$$\Sigma_{CS} - \Sigma_{SQS} = O(\sigma_u^4).$$

In the earlier paper, the ML estimator, being far more complicated than the CS and SQS estimators, was completely ignored and was not investigated. The purpose of the present paper is to fill this gap and to study the efficiency of ML relative to CS and SQS.

Of course, on general grounds, ML is more efficient than CS and SQS. However, it turns out that for small σ_u^2 the ML estimator is approximately as efficient as the two other estimators:

$$\Sigma_{ML} = \Sigma_{SQS} + O(\sigma_u^4) = \Sigma_{CS} + O(\sigma_u^4).$$

The proof of this proposition rests on heavy algebra. Here we concentrate on the algebra only and leave aside all questions of a rigorous justification of the various algebraic manipulations. Generally speaking, the functions C and ξ should be smooth and should be regular in the sense that all expectations that arise in the course of the arguments exist and that differentiation and forming of expectations are interchangeable. In KSW, exact conditions for this to hold were given for the SQS and CS estimators. The corresponding conditions for the ML estimator should be quite similar. But we do not go into these details.

In the next section, we state the model and derive its likelihood. Section 3 gives an expansion of the model density in terms of powers of the measurement error variance σ_u^2, Section 4 does the same for the integrand of the information matrix, and Sections 5 and 6 evaluate the various terms of this expansion, where it can be seen how the normality assumptions of the model are utilized. Section 7 has the main result; it presents an asymptotic expression for small σ_u^2 of the covariance matrix of the ML estimator of β and thereby proves the approximate efficiency of SQS (and CS) for small σ_u^2. Section 8 extends this result to the case of unknown nuisance parameters, which have to be estimated along with the parameter vector β of interest. Section 9 has some concluding remarks.

2 The model and its likelihood

We start from model (1) which specifies the conditional density (with respect to a fixed σ-finite measure on the Borel σ-field of $\mathbb{R}$) of a response variable Y given the covariate X. The density belongs to an exponential family with canonical parameter ξ, which depends on X and the parameter vector β to be estimated. For simplicity, the dispersion parameter φ is assumed to be known. In Section 8 we will analyse the model without assuming φ known.

The variable X is assumed to be normally distributed:

$$X \sim N(\mu, \sigma^2)$$

with parameters μ and σ^2, which, for simplicity, are assumed to be known. (They can easily be estimated from the data w_i if σ_u^2 is known.) In the end (Section 8) we will drop this assumption. X is unobservable (latent). Instead we observe the variable W, which is X together with an additive measurement error:

$$W = X + \sigma_u U, \tag{2}$$

where $U \sim N(0,1)$ and U is independent of (X, Y). σ_u is assumed to be known.

The joint density of the observable variables (Y, W) equals

$$\rho(y, w; \beta) = \frac{1}{2\pi\sigma} \int exp\{\frac{y\xi^* - C(\xi^*)}{\varphi} + c(y, \varphi) - \frac{(w - \mu - \sigma_u u)^2}{2\sigma^2} - \frac{u^2}{2}\}\ du, \tag{3}$$

where here $\xi^* := \xi(w - \sigma_u u, \beta)$. This is the likelihood function of β given one observation (y, w). The ML estimator of β is found by maximizing

$$\sum_{i=1}^{n} log\ \rho(y_i, w_i; \beta) \tag{4}$$

with respect to β, where $(y_i, w_i),\ \ i = 1, \ldots, n\ ,$ is an $i.i.d.$ sample. Because of the integral in (3) the maximization of (4) may be prohibitively difficult.

Therefore other estimation techniques have been proposed, which are simpler to carry out. Among these the Structural Quasi Score (SQS) method is most prominent, see KSW. It only uses the conditional mean and variance of Y given W and not the complete likelihood.

Nevertheless, we can still evaluate the asymptotic covariance matrix of the ML estimator, at least for small σ_u^2, and compare it to the asymptotic covariance of the SQS estimator. This is the purpose of the present paper.

The asymptotic covariance matrix Σ_{ML} of the ML estimator of β is the inverse of the information matrix I_β, and this is given by

$$I_\beta = E \, \frac{\rho_\beta(Y, W; \beta) \; \rho^t_\beta(Y, W; \beta)}{\rho^2(Y, W; \beta)}, \tag{5}$$

where $\rho_\beta := \frac{\partial \rho}{\partial \beta}$, and t is the transposition sign.

Note on Notation: In the sequel we will generally omit the arguments in the various functions. Thus ρ stands for $\rho(Y, W; \beta)$, and C stands for $C(\xi)$. As to the function ξ, we let $\xi := \xi(W, \beta)$, i.e. we replace X by W in the function ξ. In some places, e.g. in (1), we take ξ as a function of X rather than of W. We then write $\overline{\xi} := \xi(X, \beta)$ and $\overline{C} = C(\overline{\xi})$. The derivatives of C with respect to ξ are denoted by primes. Partial derivatives of ξ and functions of ξ with respect to the first or second argument of ξ are denoted by the corresponding subscripts, x or β, respectively. Note that ξ_β is column a vector.

As usual, a variable like x will be written in capitals to denote the random variable X or in small letters to denote a realization of X. The expectation sign E is understood to operate on the whole term following the sign, so that brackets will not be necessary, terms being separated by + or − signs.

Remember that in an exponential family

$$\overline{C}' = E(Y|X), \quad \varphi\overline{C}'' = V(Y|X), \quad \varphi^2\overline{C}''' = E[(Y - \overline{C}')^3|X] \tag{6}$$

3 Expansion of the model density ρ for small σ_u^2

The sign $\approx$ always denotes equality up to terms of order σ_u^2. We have

$$\xi^* = \xi(w - \sigma_u u, \beta) \approx \xi - \xi_x \sigma_u u + \frac{1}{2}\xi_{xx}\sigma_u^2 u^2 \tag{7}$$

and

$$\begin{aligned} C(\xi^*) &\approx C + C'(-\xi_x\sigma_u u + \frac{1}{2}\xi_{xx}\sigma_u^2 u^2) + \frac{C''}{2}(\xi_x \sigma_u u)^2 \\ &= C - C'\xi_x\sigma_u u + (C'\xi_{xx} + C''\xi_x^2)\frac{\sigma_u^2}{2}u^2. \end{aligned} \tag{8}$$

Denote the exponent in (3) by M. Then using (7) and (8), we have

$$\begin{aligned} M &\approx y\varphi^{-1}(\xi - \xi_x\sigma_u u + \frac{1}{2}\xi_{xx}\sigma_u^2 u^2) + c(y,\varphi) \\ &\quad -\varphi^{-1}[C - C'\xi_x\sigma_u u + (C'\xi_{xx} + C''\xi_x^2)\frac{\sigma_u^2}{2}u^2] \\ &\quad -\frac{(w-\mu)^2}{2\sigma^2} + \frac{(w-\mu)\sigma_u u}{\sigma^2} - \frac{\sigma_u^2 u^2}{2\sigma^2} - \frac{u^2}{2} \\ &:= A + B\sigma_u u + D\frac{\sigma_u^2}{2}u^2 - \frac{u^2}{2} \end{aligned}$$

with

$$A := \varphi^{-1}(y\xi - C) + c(y,\varphi) - \frac{(w-\mu)^2}{2\sigma^2} \tag{9}$$

$$B := -\varphi^{-1}(y - C')\xi_x + V \tag{10}$$

$$D := \varphi^{-1}[(y - C')\xi_{xx} - C''\xi_x^2] - \frac{1}{\sigma^2}, \tag{11}$$

where

$$V := \frac{w-\mu}{\sigma^2}, \tag{12}$$

and therefore

$$\begin{aligned} e^M &= e^{A-\frac{u^2}{2}} exp(B\sigma_u u + D\tfrac{\sigma_u^2}{2}u^2) \\ &\approx e^{A-\frac{u^2}{2}}[1 + B\sigma_u u + (D+B^2)\tfrac{\sigma_u^2}{2}u^2]. \end{aligned}$$

Finally,

$$\begin{aligned} \rho &= \frac{1}{2\pi\sigma}\int e^M du \\ &\approx \frac{e^A}{\sqrt{2\pi}\sigma}\frac{1}{\sqrt{2\pi}}\int e^{-\frac{u^2}{2}}[1 + B\sigma_u u + (D+B^2)\frac{\sigma_u^2}{2}u^2]du \\ &= \frac{e^A}{\sqrt{2\pi}\sigma}[1 + (D+B^2)\frac{\sigma_u^2}{2}]. \end{aligned} \tag{13}$$

4 Expansion of the integrand of the information matrix

First note, see (9), that

$$A_\beta = \varphi^{-1}(y - C')\xi_\beta.$$

Therefore, by (13),

$$\rho_\beta \approx \frac{e^A}{\sqrt{2\pi}\sigma}[\varphi^{-1}(y-C')\xi_\beta\{1 + (D+B^2)\frac{\sigma_u^2}{2}\} + (D_\beta + 2BB_\beta)\frac{\sigma_u^2}{2}]$$

and

$$\begin{aligned} \frac{\rho_\beta}{\rho} &\approx \frac{\varphi^{-1}(y-C')\xi_\beta + [\varphi^{-1}(y-C')(D+B^2)\xi_\beta + D_\beta + 2BB_\beta]\dfrac{\sigma_u^2}{2}}{1 + (D+B^2)\frac{\sigma_u^2}{2}} \\ &\approx \varphi^{-1}(y-C')\xi_\beta + (D_\beta + 2BB_\beta)\tfrac{\sigma_u^2}{2} \end{aligned}$$

Finally, the integrand of the information matrix is

$$\frac{\rho_\beta \rho_\beta^t}{\rho^2} \approx \varphi^{-2}(y - C')^2 \xi_\beta \xi_\beta^t + \frac{\sigma_u^2}{2} G \tag{14}$$

with

$$G := \varphi^{-1}(y - C')[(D_\beta + 2BB_\beta)\xi_\beta^t]_s, \tag{15}$$

where the subscript S denotes the symmetrization operator:

$$A_s := A + A^t$$

5 Evaluation of G

We first evaluate the derivatives of B and D from (10) and (11):

$\varphi B_\beta = -(y - C')\xi_{x\beta} + C''\xi_x\xi_\beta$

$\varphi D_\beta = (y - C')\xi_{xx\beta} - C''(\xi_{xx}\xi_\beta + 2\xi_x\xi_{x\beta}) - C'''\xi_x^2\xi_\beta$

Consequently,

$$\begin{aligned}\varphi(D_\beta + 2BB_\beta) &= C''(2\xi_x\xi_\beta V - \xi_{xx}\xi_\beta - 2\xi_x\xi_{x\beta}) - C'''\xi_x^2\xi_\beta \\ &\quad +(y - C')(\xi_{xx\beta} - 2\varphi^{-1}C''\xi_x^2\xi_\beta - 2V\xi_{x\beta}) \\ &\quad +2\varphi^{-1}(y - C')^2\xi_x\xi_{x\beta}.\end{aligned} \tag{16}$$

Substitution (16) in (15) we get

$$\begin{aligned}\varphi^2 G &= 2(y - C')[C''(2\xi_x\xi_\beta\xi_\beta^t V - \xi_{xx}\xi_\beta\xi_\beta^t - \xi_x(\xi_{x\beta}\xi_\beta^t)_s) - C'''\xi_x^2\xi_\beta\xi_\beta^t] \\ &\quad +(y - C')^2[-2(\xi_{x\beta}\xi_\beta^t)_s V + (\xi_{xx\beta}\xi_\beta^t)_s - 4\varphi^{-1}C''\xi_x^2\xi_\beta\xi_\beta^t] \\ &\quad +2\varphi^{-1}(y - C')^3\xi_x(\xi_{x\beta}\xi_\beta^t)_s.\end{aligned} \tag{17}$$

6 Expansion of $E(Y - C')^k h$

As can be seen from (17), G consists of terms of the form $(Y - C')^k h, \quad k = 1, 2, 3$, where h is a function of W. The same is true for the other term in the integrand of the information matrix, see (14). We therefore investigate the expectation of these terms for alternative values of k and expand them in terms of powers of σ_u^2. However, it is only for $k = 2$ that we need an expansion up to the order of σ_u^2. For $k = 1$ and 3 we only need to know the first term in the expansion.

We first expand ξ:

$$\xi \approx \overline{\xi} + \overline{\xi}_x \sigma_u U + \overline{\xi}_{xx} \frac{\sigma_u^2}{2} U^2,$$

where $\overline{\xi} := \xi(X, \beta)$ etc.; see the note on notation at the end of Section 2. It follows that

$$C' \approx \overline{C}' + \overline{C}''(\overline{\xi}_x \sigma_u U + \overline{\xi}_{xx} \frac{\sigma_u^2}{2} U^2) + \overline{C}''' \overline{\xi}_x^2 \frac{\sigma_u^2}{2} U^2. \tag{18}$$

Now, with $k = 1$,

$$\begin{aligned} E(Y - C')h \approx{} & E[Y - \overline{C}' - \overline{C}''(\overline{\xi}_x \sigma_u U + \overline{\xi}_{xx} \tfrac{\sigma_u^2}{2} U^2) - \overline{C}''' \overline{\xi}_x^2 \tfrac{\sigma_u^2}{2} U^2] \\ & * [\overline{h} + \overline{h}_x \sigma_u U + \overline{h}_{xx} \tfrac{\sigma_u^2}{2} U^2]. \end{aligned}$$

Due to (6), $E(Y - \overline{C}'|X) = 0$, and as U is independent of X and Y and $U \sim N(0, 1)$, we thus have

$$E(Y - C')h = O(\sigma_u^2). \tag{19}$$

In a similar way we can now treat the case $k = 2$. Using (18) and (6), we first get

$$E(Y - C')^2 h \approx E\varphi\overline{C}''\overline{h} + \frac{\sigma_u^2}{2} E(\varphi\overline{C}''\overline{h}_{xx} + 2\overline{C}''^2\overline{\xi}_x^2\overline{h}). \tag{20}$$

We need to express this result as a function of the observable W rather than the latent X. For this remember that because of (2) and the joint normality of X and U we have

$$X = W - \frac{\sigma_u^2}{\sigma^2 + \sigma_u^2}(W - \mu) + \tau N,$$

$$\tau^2 = \frac{\sigma_u^2\sigma^2}{\sigma^2 + \sigma_u^2} \quad ,$$

where N is a standard Gaussian variable independent of W, see KSW. Up to the order of σ_u^2 we have

$$\frac{\sigma_u^2}{\sigma^2 + \sigma_u^2} \approx \frac{\sigma_u^2}{\sigma^2}$$

and consequently, with V from (12),

$$X \approx W - \sigma_u^2 V + \tau N.$$

For any function g (with some regularity properties) we therefore have, because of $\tau^2 \approx \sigma_u^2$,

$$Eg(X) \approx E[g(W) + g'(W)(\tau N - \sigma_u^2 V) + \frac{1}{2}g''(W)\sigma_u^2 N^2]$$

or, for short, with $\overline{g} := g(X)$ and $g := g(W)$,

$$E\overline{g} \approx Eg + E(-2g'V + g'')\frac{\sigma_u^2}{2}. \tag{21}$$

This can be simplified by using the following lemma, which is proved in the appendix, see also KSW, equation (100).

Lemma

For any function $f(W)$ (with some regularity properties) and for V from (12) we have

$$Ef(W)V = (1 + \frac{\sigma_u^2}{\sigma^2})Ef'(W).$$

Applying this lemma to (21), we get

$$E\overline{g} \approx Eg - \frac{\sigma_u^2}{2}Eg''. \tag{22}$$

In this way we can transform expectations of functions of X into expectations of functions of W.

We can apply this result to (20) and obtain

$$E(Y - C')^2 h \approx E\varphi C'' h + \frac{\sigma_u^2}{2}E[-\varphi(C''h)_{xx} + \varphi C'' h_{xx} + 2C''^2\xi_x^2 h].$$

With

$$(C''h)_{xx} = C'' h_{xx} + C'''(2\xi_x h_x + \xi_{xx} h) + C^{(4)}\xi_x^2 h$$

we finally have

$$\begin{aligned} E(Y - C')^2 h \;\; &\approx \;\; \varphi E C'' h \\ &+ \frac{\sigma_u^2}{2}E[2C''^2\xi_x^2 h - \varphi C'''(2\xi_x h_x + \xi_{xx} h) - \varphi C^{(4)}\xi_x^2 h]. \end{aligned} \tag{23}$$

Now for the case $k = 3$. Here we need only the first term of the expansion. Using (18) an (6), we have

$$E(Y - C')^3 h = \varphi^2 E\overline{C}'''\overline{h} + O(\sigma_u^2).$$

This can again be transformed into an expression with observables by using (22):

$$E(Y - C')^3 h = \varphi^2 E C''' h + O(\sigma_u^2). \tag{24}$$

7 Expansion of I_β and Σ_{ML}

Using (19), (23), and (24), we can now expand $\frac{\sigma_u^2}{2} EG$ from (17) in terms of powers of σ_u^2:

$$\begin{aligned}\frac{\sigma_u^2}{2} EG \;\;\approx\;\; & \varphi^{-1}\frac{\sigma_u^2}{2} E[C''\{-2(\xi_{x\beta}\xi_\beta^t)_s V + (\xi_{xx\beta}\xi_\beta^t)_s - 4\varphi^{-1}C''\xi_x^2\xi_\beta\xi_\beta^t\} \\ & +2C'''\xi_x(\xi_{x\beta}\xi_\beta^t)_s].\end{aligned}$$

With the help of the lemma this can be simplified to

$$\frac{\sigma_u^2}{2} EG \approx -\frac{\sigma_u^2}{2}\varphi^{-1} E[C''\{(\xi_{xx\beta}\xi_\beta^t)_s + 4\xi_{x\beta}\xi_{x\beta}^t\} + 4\varphi^{-1}C''^2\xi_x^2\xi_\beta\xi_\beta^t]. \tag{25}$$

In a similar way we also expand the expectation of the first term on the right-hand side of (14):

$$\begin{aligned}\varphi^{-2}E(Y - C')^2\xi_\beta\xi_\beta^t \;\;\approx\;\; & \varphi^{-1}\{EC''\xi_\beta\xi_\beta^t + \frac{\sigma_u^2}{2}E[2\varphi^{-1}C''^2\xi_x^2\xi_\beta\xi_\beta^t \\ & -C'''\{2\xi_x(\xi_\beta\xi_\beta^t)_x + \xi_{xx}\xi_\beta\xi_\beta^t\} - C^{(4)}\xi_x^2\xi_\beta\xi_\beta^t]\}.\end{aligned} \tag{26}$$

Finally we find an expansion for the information matrix I_β, defined in (5). According to (14), this is just the sum of (25) and (26).

$$\begin{aligned} I_\beta \approx \quad & \varphi^{-1}\{EC''\xi_\beta\xi_\beta^t \\ & -\tfrac{\sigma_u^2}{2}E[2\varphi^{-1}C''^2\xi_x^2\xi_\beta\xi_\beta^t + C''((\xi_{xx\beta}\xi_\beta^t)_s + 4\xi_{x\beta}\xi_{x\beta}^t) \\ & \qquad +C'''(2\xi_x(\xi_\beta\xi_\beta^t)_x + \xi_{xx}\xi_\beta\xi_\beta^t) \\ & \qquad +C^{(4)}\xi_x^2\xi_\beta\xi_\beta^t]\} \\ =: \quad & \varphi^{-1}(S - \tfrac{\sigma_u^2}{2}Q). \end{aligned}$$

The asymptotic covariance matrix of $\widehat{\beta}_{ML}$ is then found to be

$$\Sigma_{ML} = \varphi(S^{-1} + \frac{\sigma_u^2}{2}S^{-1}QS^{-1}) + O(\sigma_u^4).$$

Noting that

$$(\xi_{xx\beta}\xi_\beta^t)_s + 4\xi_{x\beta}\xi_{x\beta}^t = (\xi_\beta\xi_\beta^t)_{xx} + 2\xi_{x\beta}\xi_{x\beta}^t$$

and

$$2\xi_x(\xi_\beta\xi_\beta^t)_x + \xi_{xx}\xi_\beta\xi_\beta^t = 2(\xi_x\xi_\beta\xi_\beta^t)_x - \xi_{xx}\xi_\beta\xi_\beta^t,$$

we see that Σ_{ML} equals the expression for Σ_{SQS} in (21) of KSW up to the order of σ_u^2, i.e.:

$$\Sigma_{ML} = \Sigma_{SQS} + O(\sigma_u^4).$$

8 Nuisance parameters

Up to now we assumed the nuisance parameters φ, μ, and σ^2 to be known. We shall now drop this assumption.

Let $\nu := (\varphi, \mu, \sigma^2)^t$ be the vector of the nuisance parameters and $\theta := (\beta^t, \nu^t)^t$ the vector of all unknown parameters of the model. Suppose θ is estimated by Maximum Likelihood. The covariance matrix of $\widehat{\theta}$ is then given as the inverse of the information matrix of θ, which can be decomposed in the following way

$$I_\theta = \begin{pmatrix} I_\beta & F \\ F^t & I_\nu \end{pmatrix},$$

where I_β is as in Section 7, I_ν is the information matrix of ν:

$$I_\nu = E\rho^{-2}\rho_\nu\rho_\nu^t,$$

and

$$F = E\rho^{-2}\rho_\beta\rho_\nu^t.$$

By arguments as in the previous sections but without the need to compute all the terms, it is seen that F can be expanded as

$$F \approx F_0 + F_1\frac{\sigma_u^2}{2}.$$

In the error-free model (i.e., $\sigma_u^2 = 0$), the ML estimators of β, φ, μ, and σ^2 are asymptotically independent. Thus, in this case, $F = 0$, and therefore, in the error-ridden model, $F_0 = 0$ and thus $F = O(\sigma_u^2)$. In addition, in the error-free model, I_β and I_ν are positive definite and so, in the error-ridden model, I_β and I_ν are also positive definite, at least for small σ_u^2. It then follows that

$$I_\theta^{-1} = \begin{pmatrix} I_\beta^{-1} & -I_\beta^{-1} F I_\nu^{-1} \\ -I_\nu^{-1} F^t I_\beta^{-1} & I_\nu^{-1} \end{pmatrix} + O(\sigma_u^4).$$

Considering only the upper left corner of this matrix, we see that the covariance matrix of $\widehat{\beta}_{ML}$ is now given by

$$\Sigma_{ML} = I_\beta^{-1} + O(\sigma_u^4).$$

But according to our main result of the previous section,

$$I_\beta^{-1} = \Sigma_{SQS} + O(\sigma_u^4).$$

Therefore we have again, even in the presence of nuisance parameters,

$$\Sigma_{ML} = \Sigma_{SQS} + O(\sigma_u^4).$$

9 Conclusion

In a measurement error model Maximum Likelihood (ML) methods are often extremely difficult to apply. For this reason other estimation techniques that are easier to apply and yet lead to consistent estimators have been proposed. Among these Structural Quasi Score (SQS) and Corrected Score (CS) methods are most prominent. These methods are, however, less efficient than ML. Fortunately, it can be shown that for small error variances these three methods are almost equally efficient. This result is a further justification for the use of the simpler methods in practical applications.

ML and SQS, as far as these methods have been developed here, both rely on the assumption that the latent covariate is normally distributed. Deviations from normality lead to biased estimators, see Cheng and Schneeweiss (2003). One can, however, modify the ML and SQS methods so that they take into account a (finite) mixture of normal distributions instead of just one normal distribution, for SQS see Cheng and Schneeweiss (2003). This possibility renders these methods rather flexible in so far as any (continuous) distribution can be approximated by a mixture of normals.

The near equal efficiency result was proved for a scalar covariate X (but for a vector valued parameter β). One should be able to prove the same result for a vector of covariates each measured with measurement errors.

Appendix

Proof of the Lemma in Section 6:

First note that $W \sim N(\mu, \sigma_w^2)$ with $\sigma_w^2 = \sigma^2 + \sigma_u^2$. Therefore, with V from (12),

$$\begin{aligned} Ef(W)V &= Ef(W)\tfrac{W-\mu}{\sigma_w^2}\tfrac{\sigma_w^2}{\sigma^2} \\ &= \tfrac{\sigma_w^2}{\sigma^2}\tfrac{1}{\sqrt{2\pi}\sigma_w}\int f(w)\tfrac{w-\mu}{\sigma_w^2}e^{-\frac{(w-\mu)^2}{2\sigma_w^2}}dw, \end{aligned}$$

which by partial integration is equal to

$$\begin{aligned} Ef(W)V &= \frac{\sigma_w^2}{\sigma^2}\frac{1}{\sqrt{2\pi}\sigma_w}\int f'(w)e^{-\frac{(w-\mu)^2}{2\sigma_w^2}}dw \\ &= \frac{\sigma_w^2}{\sigma^2}Ef'(W) \\ &= (1+\frac{\sigma_u^2}{\sigma^2})Ef'(W). \end{aligned}$$

References

[1] Kukush, A., Schneeweiss, H., Wolf, R. (2002): Comparing different estimators in a nonlinear measurement error model. *Discussion Paper 244, SFB 386*, University of Munich.

[2] Shklyar, S., Schneeweiss, H. (2002): A comparison of asymptotic covariance matrices of three consistent estimators in the Poisson regression model with measurement error. *Discussion Paper 283, SFB 386*, University of Munich. To appear in Journ. Multivariate Analysis.

[3] Cheng, C.L., Schneeweis, H. (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.