Volume 14, No. 2 (2005), pp. 203–223 Allerton Press, Inc.

MATHEMATICAL METHODS OF STATISTICS

COMPARING DIFFERENT ESTIMATORS IN A NONLINEAR MEASUREMENT ERROR MODEL. II

A. Kukush[1] and H. Schneeweiss[2]

[1]Department of Mechanics and Mathematics, Kiev National Taras Shevchenko University
Volodymyrska str. 60, 01033 Kiev, Ukraine
E-mail: alexander_kukush@univ.kiev.ua

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

As in Part I of the paper, we deal with a nonlinear structural errors-in-variables model, where the response variable has a density belonging to an exponential family and the error prone covariate follows a Gaussian distribution. The error variance is assumed to be known. In addition to the two consistent estimators SQS and CS studied in Part I, we now investigate the naive estimator and the ML estimator. For small measurement error variances, ML, SQS, and CS are equally efficient up to the order of the measurement error variance. The naive estimator differs in its asymptotic covariance matrix from SQS and CS. We also study the case when not only the measurement error variance but also the dispersion parameter of the model goes to zero and obtain some results for this case. The polynomial model and the Poisson regression model are explored in more detail.

Key words: exponential family, structural errors-in-variables model, efficiency, naive estimator, ML, polynomial regression, Poisson regression, small measurement error variance, small dispersion parameter.

2000 Mathematics Subject Classification: 62J02, 62H12, 62F12, 62J12.

1. Introduction

As in Part I of this paper (Kukush and Schneeweiss (2005)), we study the model with response variable Y having the density

$$f(y \mid \xi) = \exp\left\{\frac{y\xi - C(\xi)}{\varphi} + c(y, \varphi)\right\},$$

where $\xi = \xi(X, Z, \beta)$ is a function of the vector variables X and Z, with X being normally distributed and measured with errors:

$$W = X + U,$$

where $U \sim N(0, \sigma_u^2)$. We want to estimate the parameter vector β within the set Θ_β (see Part I) from the data (Y_i, W_i, Z_i), $i = 1, \ldots, n$, when σ_u^2 is known.

For this model, various regularity conditions have been laid down in Part I. They are assumed to hold also for this paper. If necessary, we refer to them with the numbers (i), (ii) etc. as they occur in Part I. We also refer to the equations of Part I with their numbers in Part I but with an added KS, e.g., KS(23). In a similar way we refer to sections, theorems, and remarks of Part I.

In Part I, we studied two consistent estimators: Structural Quasi Score (SQS) and Corrected Score (CS). We compared their asymptotic covariance matrices for small σ_u^2. Here we want to do the same for two more estimators: the Naive (N) and the Maximum Likelihood (ML) estimators.

The naive estimator simply ignores the measurement errors and estimates the parameters by conventional methods (ML or Quasi ML) as if there were no errors. Although the naive estimator is biased, it is still worthwhile to study its asymptotic covariance matrix, as in some situations it may turn out that the estimator is the best in terms of the MSE. We can determine the asymptotic bias of the naive estimator.

The ML estimator is most efficient within the class of consistent estimators. However, it turns out that its efficiency for small σ_u^2 is almost the same as that of SQS and CS. More precisely, the asymptotic covariance matrix of ML differs from those of SQS and CS only by terms of order $O(\sigma_u^4)$. This result supplements the main result of Part I.

We also study the case where not only σ_u^2 but also the dispersion parameter φ of the model becomes small. In this case, SQS turns out to be more efficient than CS and the latter becomes almost as efficient as N. (Another borderline case, for large errors, has been studied for the polynomial model in Kukush and Schneeweiss (2000)).

These findings, including the results of Part I, are specialized to two particular cases: the polynomial regression and the Poisson regression, where more detailed results can be presented. Naturally, these results also hold true for the linear model. However, in the linear model SQS, CS, and ML coincide anyway, so that our results become trivial in this case.

It should be mentioned that there are other estimation methods apart from those studied in this paper, e.g., Regression Calibration and SIMEX, see Caroll *et al.* (1995), but they are not consistent, although they typically reduce the bias.

In the next section, the naive estimator is introduced. Its asymptotic bias is determined and its asymptotic covariance matrix is compared with that of SQS (and implicitly also of CS), both for small σ_u^2. Section 3 deals with two special cases: the polynomial model (and more generally a linear-in-β regression model) and the Poisson model. In Section 4 the case when both σ_u^2 and φ go to zero simultaneously is investigated. Section 5 contains some numerical results illustrating the general theory. Section 6 introduces ML as a method to which the simpler methods SQS and

CS can be compared. Some concluding remarks are found in Section 7. Section 8 contains the proofs. A matrix inequality lemma is given in the Appendix, which might be of some interest also outside its particular application in the present paper.

2. The Naive (N) Estimator

Starting from the likelihood score function of β for the error free model:

$$\psi(y, x, z, \beta) = y\xi_\beta - C'(\xi)\xi_\beta, \tag{1}$$

where ξ_β is short for $\partial\xi/\partial\beta$ and ξ and ξ_β are taken at the point (x, z, β), the naive estimator $\hat{\beta}_N$ of β is defined as a measurable solution to

$$\frac{1}{n}\sum_{i=1}^{n}\psi(Y_i, W_i, Z_i, \beta) = 0, \qquad \beta \in \Theta_\beta. \tag{2}$$

(Methods for solving nonlinear equations can be found, e.g., in Small and Wang (1993).) Thus in the likelihood score function, the unobservable regressor X_i is replaced with the observed surrogate covariate W_i. The limit equation is given by

$$\mathrm{E}\left[m(W, Z, \beta_0)\xi_\beta - C'(\xi)\xi_\beta\right] = 0, \qquad \beta \in \Theta_\beta, \tag{3}$$

where

$$m(W, Z, \beta_0) = \mathrm{E}\,(Y \mid W, Z),$$

β_0 being the true parameter value, and ξ and ξ_β are now taken at the point (W, Z, β). We shall investigate the properties of the naive estimator under the following restriction on the function $\xi(x, z, \beta)$:

(xiii) ξ is linear in β.

For instance, canonical generalized linear models, with $\xi = \beta_0 + \beta_x x + \beta_z^t z$ are linear in β. The same is true for the polynomial model. In fact, for most common models $\xi(x, z, \cdot)$ is a linear function.

Theorem 2.1. *Assume* (i) *to* (iii), (v), (vi), (ix), (x), *and* (xiii). *Then for sufficiently small* σ_u^2:

(a) *the equation* (3) *has a unique solution* $\beta_* = \beta_*(\sigma_u^2)$,

(b) *eventually, the equation* (2) *has a solution* $\hat{\beta}_N$,

(c) $\hat{\beta}_N \to \beta_*$ *a.s. as* $n \to \infty$,

(d) $\beta_* = \beta_0 + \frac{1}{2}\sigma_u^2\Delta\beta_* + O(\sigma_u^4)$ *as* $\sigma_u^2 \to 0$, *with*

$$\Delta\beta_* := -S^{-1}\mathrm{E}\left[(C''\xi_x\xi_\beta)_x + C''\xi_x\xi_{x\beta}\right], \tag{4}$$

where

$$S := \mathrm{E}\,(C''\xi_\beta\xi_\beta^t), \tag{5}$$

$C'' = C''(\xi)$, *and* ξ *and its derivatives are taken at the point* (W, Z, β_0).

Here a property is said to hold true "eventually", if, with probability 1, it holds true for sufficiently large n; for an exact definition see Definition KS 4.1. The asymptotic covariance matrix of $\hat{\beta}_N$ has a sandwich structure, see, e.g., Schervish (1995). Introduce the following two symmetric matrices:

$$A_* := -\mathrm{E}\,\psi_{\beta^t}(\beta_*), \qquad B_* := \mathrm{E}\,\psi(\beta_*)\psi^t(\beta_*), \tag{6}$$

where $\psi(\beta_*)$ is short for $\psi(Y, W, Z, \beta_*)$. The symmetry of A_* follows directly from the definition (1) of ψ. According to the theory of estimating equations, under the conditions of Theorem 2.1 for sufficiently small σ_u^2,

$$\sqrt{n}(\hat{\beta}_N - \beta_*) \xrightarrow{d} N(0, \Sigma_N),$$

where

$$\Sigma_N = A_*^{-1} B_* A_*^{-1}. \tag{7}$$

This asymptotic covariance matrix can be compared to the asymptotic covariance matrices of $\hat{\beta}_{SQS}$ or $\hat{\beta}_{CS}$ for small σ_u^2:

Theorem 2.2. *Under the assumptions* (i) *to* (iii), (v), (vi), *and* (ix) *to* (xiii), *let* $\sigma_u^2 \to 0$ *and* φ *be fixed, then*

$$\begin{aligned}(\Sigma_{SQS} - \Sigma_N) = \frac{1}{2}\sigma_u^2 \varphi S^{-1} \mathrm{E}\Big\{ & C''[(\xi_\beta \xi_\beta^t)_{xx} + 2\xi_{x\beta}\xi_{x\beta}^t] \\ & + 2C'''[2(\xi_x \xi_\beta \xi_\beta^t)_x - \xi_{xx}\xi_\beta\xi_\beta^t + (\xi_\beta^t \Delta\beta_*)\xi_\beta\xi_\beta^t] \\ & + 2C^{(4)}\xi_x^2 \xi_\beta \xi_\beta^t \Big\} S^{-1} + O(\sigma_u^4),\end{aligned} \tag{8}$$

where $C^{(i)} = C^{(i)}(\xi)$, ξ *and its derivatives are taken at the point* (W, Z, β_0), *and* $\Delta\beta_*$ *is given in* (4) *and* S *in* (5).

In general, we cannot say which of the two estimators has the greater asymptotic covariance matrix. We can say more, however, in special cases.

3. Special Cases

3.1. Linear-in-$\boldsymbol{\beta}$ regression model. Consider the following structural errors-in-variables model

$$Y_i = \sum_{j=0}^{k} \beta_j h_j(X_i, Z_i) + \varepsilon_i, \tag{9}$$

$$W_i = X_i + U_i, \qquad i = 1, \dots, n. \tag{10}$$

We assume that h_j, $j = 0, \dots, k$, are known measurable functions, $X_i \sim$ i.i.d. $N(\mu_x, \sigma_x^2)$, Z_i are i.i.d. random vectors with values in a Euclidean space E_Z, and the errors (ε_i, U_i) are i.i.d. Gaussian, independent of the X_i's and Z_i's, with zero expectations, variances σ_ε^2 and σ_u^2, and covariance $\sigma_{\varepsilon u} = 0$. This is a particular case of our model with $\xi = \sum_{j=0}^k \beta_j h_j(X, Z)$, $C(\xi) = \xi^2/2$, $\varphi = \sigma_\varepsilon^2$. Let $\beta^0 = (\beta_{00}, \beta_{10}, \dots, \beta_{k0})^t$ and $\varphi_0 = \sigma_{\varepsilon 0}^2$ be the true values of $\beta = (\beta_0, \dots, \beta_k)^t$ and σ_ε^2

and let $\beta^0 \in \Theta_\beta$, where Θ_β is a convex compact set in $\mathbb{R}^{k+1}$. We assume that $h_j(\cdot, z) \in C^2(\mathbb{R})$ and

$$|h_j(x,z)| + |h_{jx}(x,z)| \le \mathrm{const}(e^{A|x|} + e^{A\|z\|})$$

with const > 0 and a fixed $A > 0$, $j = 0, \dots, k$. In this case the matrix (5) is calculated as

$$S = \mathrm{E}\, h(W,Z)h^t(W,Z), \tag{11}$$

with $h := (h_0, \dots, h_k)^t$. We require the corresponding matrix

$$S_0 := \mathrm{E}\, h(X,Z)h^t(X,Z)$$

to be nonsingular and assume (i) and (iii). Then Theorem KS 5.1 and Theorem 2.2 are applicable, and as $\sigma_u^2 \to 0$ we obtain

$$\begin{aligned} \Sigma_{SQS} &= \Sigma_{CS} + O(\sigma_u^4), \\ \Sigma_{CS} - \Sigma_N &= \frac{1}{2}\sigma_u^2\sigma_\varepsilon^2 S^{-1}\mathrm{E}\left\{(hh^t)_{xx} + 2h_x h_x^t\right\} S^{-1} + O(\sigma_u^4), \end{aligned} \tag{12}$$

where h and its derivatives are taken at the point (W,Z).

We cannot say that the leading term in (12) is always semidefinite. In fact, we have a counterexample in Subsection 3.2. Nevertheless, in some important cases, at least the variances of the components of $S\beta$ are smaller for N than for CS or SQS.

Let a_{jj} be the difference of corresponding diagonal elements of the asymptotic covariance matrices of the estimators $S\hat\beta_{CS}$ and $S\hat\beta_N$:

$$a_{jj} := [S(\Sigma_{CS} - \Sigma_N)S]_{jj}, \qquad j = 0, \dots, k. \tag{13}$$

From (12) we conclude that, as $\sigma_u^2 \to 0$,

$$a_{jj} = \frac{1}{2}\sigma_u^2\sigma_\varepsilon^2 \mathrm{E}\left[\frac{\partial^2(h_j^2)}{\partial x^2} + 2\left(\frac{\partial h_j}{\partial x}\right)^2\right] + O(\sigma_u^4) \tag{14}$$

with the derivative of h_j taken at (W,Z). The following statement is an easy consequence of (14).

Proposition 3.1. *Suppose that for fixed j the function $h_j^2(\cdot, Z)$ is convex a.s. and $P\{\frac{\partial h_j}{\partial x}(X,Z) \ne 0\} > 0$. Then $a_{jj} > 0$ for sufficiently small σ_u^2.*

The proposition gives a sufficient condition for the asymptotic variance of $(S\hat\beta_N)_j$ to be smaller than the asymptotic variance of $(S\hat\beta_{CS})_j$ for small measurement errors.

Analyzing the leading term of expansion (14), it is interesting to give a geometric interpretation of the inequality $\partial^2(h_j^2)/\partial x^2 + 2(\partial h_j/\partial x)^2 \ge 0$. An easy calculation leads to the following statement.

Proposition 3.2. *Let $g \in C^2(\mathbb{R})$. The inequality $(g^2)_{xx} + 2(g_x)^2 \geq 0$ holds for all $x \in \mathbb{R}$ iff the function $|g|^3$ is convex on $\mathbb{R}$. Moreover, if $|g|^3$ is strictly convex on some interval (α, β), then*

$$\mathrm{E}\left[\frac{\partial^2(g^2)}{\partial x^2}(X) + 2\left(\frac{\partial g}{\partial x}(X)\right)^2\right] > 0,$$

where $X \sim N(\mu_x, \sigma_x^2)$.

Thus if h_j does not depend on Z and satisfies the conditions of Proposition 3.2, then $a_{jj} > 0$, for sufficiently small σ_u^2.

3.2. Polynomial regression. The polynomial regression is a particular case of the model (9), (10), with

$$h_j = h_j(X) = X^j, \qquad j = 0, 1, \ldots, k. \tag{15}$$

This model has found extensive treatment in the literature, see Cheng and Schneeweiss (2002). We mention that, in the polynomial case, $\hat{\beta}_{CS}$ is called the adjusted least squares estimator ($\hat{\beta}_{ALS}$), see Cheng and Schneeweiss (1998), $\hat{\beta}_{SQS}$ is called the structural least squares estimator ($\hat{\beta}_{SLS}$), see Kukush *et al.* (2005), and $\hat{\beta}_N$ is the ordinary least squares estimator ($\hat{\beta}_{OLS}$). From (14), we get the following result.

Proposition 3.3. *In the polynomial model* (9), (10), (15) *we have for a_{jj} of* (13) *as $\sigma_u^2 \to 0$:*

$$a_{00} = O(\sigma_u^4),$$
$$a_{jj} = j(3j-1)\sigma_u^2\sigma_\varepsilon^2 \mathrm{E}\,(W^{2j-2}) + O(\sigma_u^4), \quad j = 1, \ldots, k.$$

Thus $a_{jj} > 0$, $j = 1, \ldots, k$, for sufficiently small σ_u^2, in accordance with Proposition 3.1 or Proposition 3.2. This statement about the diagonal elements of the matrix $S(\Sigma_{CS} - \Sigma_N)S$ can be supplemented by a statement about the whole matrix. For $k \geq 2$ we consider the leading term of $S(\Sigma_{CS} - \Sigma_N)S$, see (12):

$$M := \frac{1}{2}\sigma_u^2\sigma_\varepsilon^2 \mathrm{E}\,\{(hh^t)_{xx} + 2h_x h_x^t\}.$$

Due to Proposition 3.3, $M_{00} = 0$ and $M_{jj} > 0$, $j = 1, \ldots, k$. But the matrix M is not necessarily positive semidefinite. E.g., if $W \sim N(0, 1)$, then

$$\begin{pmatrix} M_{00} & M_{01} & M_{02} \\ M_{10} & M_{11} & M_{12} \\ M_{20} & M_{21} & M_{22} \end{pmatrix} = \sigma_u^2\sigma_\varepsilon^2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 10 \end{pmatrix},$$

and the determinant of this submatrix is negative; so in this case M is not positive semidefinite. Therefore, for the polynomial model it is not true that $\Sigma_{CS} \geq \Sigma_N$ in the matrix sense.

We mention that, for the polynomial model, (12) holds also true when h is evaluated at the point (X, Z) instead of (W, Z) and S is replaced with S_0. Similarly, in Proposition 3.3, $\mathrm{E}(W^{2j-2})$ may be replaced with $\mathrm{E}(X^{2j-2})$. A corresponding remark applies to the matrix M, see Kukush *et al.* (2005).

3.3. Poisson regression. The Poisson regression is one of the better known and often applied Generalized Linear Models, see, e.g., Cameron and Trivedi (1998) and Winkelmann (1997).

Suppose that $Y \mid \xi$ has a Poisson distribution with parameter $\lambda = e^{\xi}$, and $\xi = \beta_0 + \beta_1 X$, where $\beta = (\beta_0, \beta_1)^t$ is the parameter of interest. Again $X \sim N(\mu_x, \sigma_x^2)$, and $W = X + U$ is the surrogate covariate, where U is independent of X and Y, and $U \sim N(0, \sigma_u^2)$ with σ_u^2 known. We observe independent realizations $(Y_i, W_i), i = 1, \dots, n$. This is a particular case of the model of Section 2, with $C(\xi) = e^{\xi}$ and known $\varphi = 1$. A measure m dominating the distribution of Y is the counting measure.

We suppose that Θ_β is a convex compact set in $\mathbb{R}^2$, and the true value of β is an interior point of Θ_β. The statements of Theorems KS 4.3, KS 4.4, KS 5.1, 2.1, and 2.2 hold true for the Poisson regression. For further details see Thamerus (1998), Kukush *et al.* (2004), and Shklyar and Schneeweiss (2005).

Now, let β be the true value of the parameter of interest, and let

$$g := \mu_w + \sigma_w^2 \beta_1. \tag{16}$$

Theorem 3.1. *In the Poisson model the following statements hold as* $\sigma_u^2 \to 0$:

(a) $\hat{\beta}_N \to \beta_*$ *a.s. as* $n \to \infty$, *and*

$$\beta_* = \beta + \frac{1}{2}\sigma_u^2 \Delta\beta_* + O(\sigma_u^4), \tag{17}$$

$$\Delta\beta_* = \frac{\beta_1}{\sigma_w^2}(\mu_w + g, -2)^t. \tag{18}$$

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

(c)

$$\Sigma_{CS} - \Sigma_N = \frac{2\sigma_u^2}{\sigma_w^4} \exp\left\{ -(\beta_0 + \beta_1\mu_w + \frac{1}{2}\beta_1^2\sigma_w^2) \right\} \begin{pmatrix} g^2 & -g \\ -g & 1 \end{pmatrix} + O(\sigma_u^4). \tag{19}$$

In contrast to the expansion (12) in the polynomial case, we see that the term of order σ_u^2 in the expansion (19) is a positive semidefinite matrix. Moreover, we conclude from (19) that for sufficiently small σ_u^2, the asymptotic variance of $\hat{\beta}_{1,N}$ is smaller than the asymptotic variance of $\hat{\beta}_{1,CS}$, and if $g \neq 0$, the same holds true for the estimators of β_0.

4. Asymptotics when Both Errors are Small

Hereafter we consider again the general model presented in Section 1. Again we deal with a series of such models, but now we suppose that only the parameters β, μ_x, and σ_x^2 stay fixed, while the dispersion parameter φ and the variance σ_u^2 tend to zero simultaneously. As to the relation between φ and σ_u^2, we consider two cases:

(xiv) $\chi^2 := \varphi/\sigma_u^2 = \text{const}$, and

(xv) $C_1 \le \varphi/\sigma_u^2 \le C_2$ for some positive constants C_1 and C_2.

Let

$$v_0(W, Z, \beta_0) := \chi^2 + C''\xi_x^2, \tag{20}$$

where $C'' = C''(\xi)$, and ξ and ξ_x are taken at the point (W, Z, β_0). Below ξ and ξ_β are also taken at this point.

Theorem 4.1. *Assume* (xiv) *and let* $\sigma_u^2 \to 0$.

(a) *Under the conditions of Theorem* KS 4.4,

$$\Sigma_{SQS} = \sigma_u^2 \left[\mathrm{E}\,(C''\xi_\beta\xi_\beta^t v_0^{-1})\right]^{-1} + O(\sigma_u^4).$$

(b) *Under the conditions of Theorem* KS 5.1,

$$\Sigma_{CS} = \sigma_u^2 S^{-1}\mathrm{E}\,(C''\xi_\beta\xi_\beta^t v_0)S^{-1} + O(\sigma_u^4).$$

(c) *Under the conditions of Theorem* 2.2,

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

Remark 4.1. In Theorem 4.1, one can replace S with S_0 and also the variable W with X in the argument of ξ and its derivatives without changing the statement of the theorem. See also Remark KS 4.3 and the remark at the end of Subsection 3.2.

The next result compares Σ_{SQS} with Σ_{CS} for both errors small.

Theorem 4.2. *Let the conditions of Theorem* KS 5.1 *hold. Assume additionally that the distribution of* $C''\xi_x^2$ *as a function of* (W, Z, β_0) *has no atoms. Then*:

(a) *under* (xiv), *the difference*

$$\lim_{\sigma_u^2\to 0}(\sigma_u^{-2}\Sigma_{CS}) - \lim_{\sigma_u^2\to 0}(\sigma_u^{-2}\Sigma_{SQS})$$

is positive semidefinite and is not equal to zero.

(b) *under* (xv),

$$\liminf_{\sigma_u^2\to 0}[\sigma_u^{-2}\lambda_{\max}(\Sigma_{CS} - \Sigma_{SQS})] > 0 \tag{21}$$

and

$$\liminf_{\sigma_u^2\to 0}[\sigma_u^{-2}\lambda_{\min}(\Sigma_{CS} - \Sigma_{SQS})] \geq 0. \tag{22}$$

Theorem 4.2 states that, for small errors and under normality assumptions, when φ and σ_u^2 are of the same order, the SQS estimator is asymptotically more efficient than the CS estimator.

Remark 4.2. In the polynomial model of Subsection 3.2 with degree $k \geq 2$, and $\beta_2 \neq 0$, the distribution of $C''\xi_x^2$ has no atoms, and the results of Theorems 4.1 and 4.2 hold true. Thus in a polynomial model of degree $k \geq 2$ in case of small σ_ε^2 and σ_u^2, the SQS estimator is asymptotically more efficient than the CS estimator.

5. Numerical Results

In order to see how close the small-σ_u approximations come to the true asymptotic covariance matrices (ACM), we numerically computed these covariance matrices for two special models: a quadratic regression model and a Poisson regression model, see Subsections 3.2 and 3.3. In the quadratic model we studied both limiting cases: (a) $\sigma_u^2 \to 0$, $\varphi = \sigma_\varepsilon^2$ fixed, and (b) $\sigma_u^2 \to 0$, $\chi^2 = \sigma_\varepsilon^2/\sigma_u^2$ fixed. In the Poisson model only case (a) was investigated; case (b) does not exist, as $\varphi = 1$. For the quadratic model, we took $\beta = (0, 1, -0.5)'$, $X \sim N(0, 1)$, and for case (a), $\sigma_u^2 = 0.05$ and $\sigma_\varepsilon^2 = 20$, while for case (b), $\sigma_u^2 = 0.01$ and $\sigma_\varepsilon^2 = 0.002$, so that $\chi^2 = 0.2$. For the Poisson model, we took $\beta = (-1, 0.5)'$, $X \sim N(2, 1)$, and $\sigma_u^2 = 0.05$. We also performed some simulation studies not shown here in order to check the theoretical results, see Kukush *et al.* (2002) and Wolf (2004).

We computed the ACM's of the SQS, CS, and N estimators according to KS(18), KS(35), and (7), respectively, and their small-σ_u approximations according to KS(22), (32), and Theorem 4.1. However, to save space, we only present the asymptotic variances. They are shown in Tables 1–3. For the N estimator we also computed the asymptotic bias according to Theorem 2.1.

The bias of N turns out to be large enough to justify the employment of consistent methods like SQS or CS, especially if the sample size should be large. When (in the polynomial model) only σ_u^2 is small, the ACM's of CS and SQS come very close to each other and also quite close to their approximation, while the ACM of N is quite a bit smaller. When both σ_u^2 and σ_ϵ^2 are small, the ACM's of N and CS almost coincide and are well approximated, while the ACM of SQS now is clearly a good deal smaller. These numerical results corroborate the general theory.

Figure 1 shows how the ACM's of the SQS, CS, and N estimates of the parameter β_1 of the Poisson model approach the ACM of the error free-model when $\sigma_u^2 \to 0$. Clearly the ACM curves of SQS and CS approach this value at the same angle, whereas the ACM curve of N has a different angle.

A few words about the computations (for more detail see Schneeweiss (2005)). In the polynomial model, the bias of N is $\beta^* - \beta$, where β^* is the solution to the equation (3), which turns out to be a linear system of equations for β^* with higher moments and linear combinations of moments of W as its coefficients, see Kukush *et al.* (2005). Thus the bias can be computed with the help of higher moments of W only. The same is true for the exact and approximate ACM's of N and CS. It should be noted, though, that in Kukush *et al.* (2005) moments of X rather than of W were used, which, however, can be easily transformed into each other. Only the exact ACM of SQS, and its approximation when both errors are small, cannot be reduced to simple moments. In these cases the expectation of a rational function in W has to be computed by numerical integration.

In the Poisson model, Shklyar and Schneeweiss (2005) derived an explicit formula for the ACM of CS, which can be used to compute it numerically. Kukush *et al.* (2004) have formulas for the bias of N and ACM's of N, CS, and SQS. These

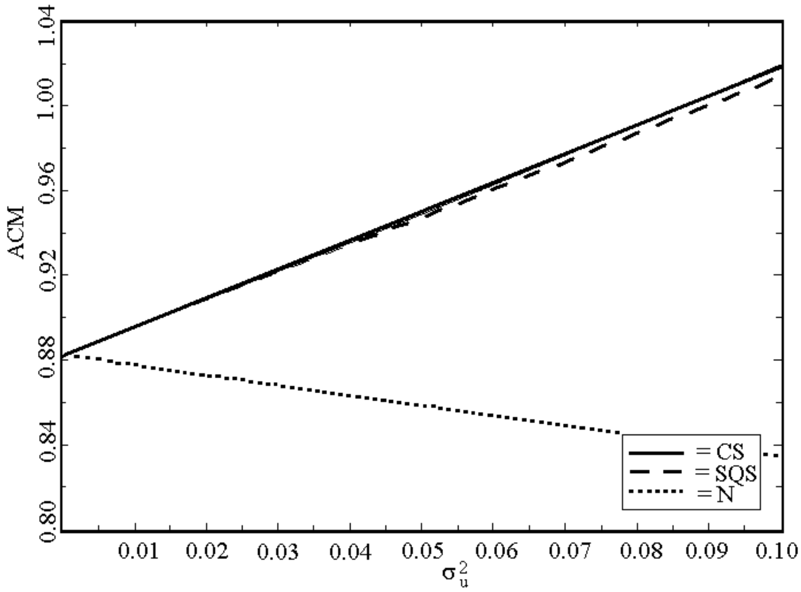

Fig. 1. ACM's of SQS, CS, and N for the parameter β_1 of a Poisson model as functions of σ_u^2

formulas employ the expectations of expressions of the form $W^r e^{aW}$. Again only the computation of the ACM of SQS needs numerical integration, viz., in computing $\Phi := \mathrm{E}\,(v^{-1}\frac{\partial m}{\partial \beta}\frac{\partial m}{\partial \beta^t})$, see KS(18).

In both models, the ACM of SQS is computed with the inclusion of those parts of the ACM that stem from the estimation of the nuisance parameters μ_w and σ_w^2, see the second term on the r.h.s. of KS(18). Their contribution to the ACM is, however, almost negligible in the present parameter setting. It turns out that in the polynomial and Poisson models the computation of F_p (see KS(17)) can be reduced to that of Φ, so that numerical integration needs to be performed only once in each model.

Table 1. Bias and exact and approximate asymptotic variances of different estimators in a polynomial model, case (a): $\beta = (0, 1, -0.5)^t$, $\mu_x = 0$, $\sigma_x^2 = 1$, $\sigma_\epsilon^2 = 20$, $\sigma_u^2 = 0.05$

	N	N	N	CS	SQS	CS/SQS
	Bias	exact	appr.	exact	exact	appr.
β_0	−0.024	30.14	30.15	31.20	31.18	31.11
β_1	−0.048	19.22	19.25	21.22	21.20	21.06
β_2	0.046	9.20	9.21	11.20	11.18	10.94

TABLE 2. Bias and exact and approximate asymptotic variances of different estimators in a polynomial model, case (b): $\beta = (0, 1, -0.5)^t$, $\mu_x = 0$, $\sigma_x^2 = 1$, $\sigma_\epsilon^2 = 0.002$, $\sigma_u^2 = 0.01$.

	N	N	CS	N/CS	QS	QS
	Bias	exact	exact	appr.	exact	appr.
β_0	−0.0050	0.033	0.032	0.033	0.017	0.016
β_1	−0.0099	0.041	0.041	0.042	0.027	0.027
β_2	0.0099	0.030	0.031	0.031	0.015	0.014

TABLE 3. Bias and exact and approximate asymptotic variances of different estimators in a Poisson model: $\beta = (-1, 0.5)^t$, $\mu_x = 2$, $\sigma_x^2 = 1$, $\sigma_u^2 = 0.05$.

	N	N	N	CS	SQS	CS/SQS
	Bias	exact	appr.	exact	exact	appr.
β_0	−0.024	6.23	6.24	6.79	6.78	6.74
β_1	−0.054	0.86	0.86	0.95	0.95	0.94

6. Maximum Likelihood

Up to now we only considered estimation methods which were not Maximum Likelihood (ML), and this for good reason: ML is much more difficult to apply than CS or SQS because it implies the integration over the latent variable X. Nevertheless we can compare the asymptotic covariance matrix Σ_{ML} of $\widehat{\beta}_{ML}$ with those of $\widehat{\beta}_{CS}$ and $\widehat{\beta}_{SQS}$, in particular, for small σ_u^2. For simplicity, let us drop the variable Z in the sequel.

The ML method is based on the joint density of (Y, W) given by

$$\rho(y, w; \beta) = \frac{1}{2\pi\sigma_x} \int \exp\left\{ \frac{y\xi^* - C(\xi^*)}{\varphi} + c(y, \varphi) - \frac{(w - \mu_x - \sigma_u u)^2}{2\sigma_x^2} - \frac{u^2}{2} \right\} du,$$

where $\xi^* := \xi(w - \sigma_u u, \beta)$. Σ_{ML} is the inverse of the information matrix I_β, which is given by

$$I_\beta = \mathrm{E}\, \frac{\rho_\beta(Y, W; \beta)\ \rho_\beta^t(Y, W; \beta)}{\rho^2(Y, W; \beta)}.$$

Although in general $\Sigma_{ML} \leq \Sigma_{SQS}$ and $\Sigma_{ML} \leq \Sigma_{CS}$, the covariance matrices tend to become equal up to the order of σ_u^2 when σ_u^2 becomes small:

Theorem 6 *Under general regularity conditions*

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

Remark 6.1. For brevity of exposition, we do not present the exact conditions for the theorem to hold true. However, these are similar to the conditions given in the previous theorems. We only provide a sketch of a proof in Subsection 8.9, which does not use any regularity conditions but only exploits the algebra that leads to the result of the theorem. For more details, see Kukush and Schneeweiss (2004b).

Remark 6.2. The theorem implies that, although ML is generally more efficient than SQS or CS, the efficiency of the latter methods is only slightly worse than that of ML as long as the measurement error variance is not too large. As SQS and CS are methods much simpler to apply than ML, this is an assuring message.

7. Conclusion

In Part I of this paper, we studied the relative asymptotic efficiency of the SQS and the CS estimators of the parameters of a nonlinear regression model with Gaussian measurement errors in one of the covariates. For small measurement error variance σ_u^2, their asymptotic covariance matrices (ACM's) turned out to be equal up to the order of σ_u^2. In the present paper (Part II), we prove that the same holds true for the ML estimator. Thus for small σ_u^2, all three estimators are almost equally efficient. In this case, it is not necessary to apply the complicated ML procedure. The simpler SQS or CS estimators are almost just as good.

We also included in our investigation the so-called naive estimator, which is the ML estimator of the model computed without regard to measurement errors. The naive estimator is inconsistent, but its ACM may still be of interest, in particular, if it turns out to be smaller, in a sense, than those of the consistent estimators. For the polynomial and Poisson regression models this can, in fact, be shown when σ_u^2 becomes small, albeit in the polynomial model only in a limited sense, because the difference of the ACM's for N and SQS is indefinite in the polynomial case.

A different picture is seen if together with σ_u^2 also the dispersion parameter φ of the exponential family goes to zero in constant proportion to σ_u^2. In this case, the SQS estimator is, in a sense, more efficient for small σ_u^2 than the CS estimator, and the naive estimator has the same ACM (up to the order of σ_u^2) as the CS estimator and is thus seen to be less efficient than the SQS estimator.

The efficiency study of the present paper is confined to the case of small measurement errors. It is an open problem to evaluate the relative efficiencies of the various estimators for arbitrary error variances. For the Poisson and polynomial models, however, it can be shown that SQS is more efficient than CS for whatever σ_u^2, see Shklyar and Kukush (2002), Shklyar and Schneeweiss (2005), Shklyar *et al.* (2005).

8. Proofs

8.1. Proof of Theorem 2.1, part (a). Consider the limit estimating function in (3)

$$Q(\beta, \sigma_u^2) := \mathrm{E}\,[m(W, Z, \beta_0)\xi_\beta - C'(\xi)\xi_\beta], \qquad \beta \in \Theta_\beta, \quad \sigma_u^2 \geq 0.$$

If $\sigma_u^2 = 0$, then $Q(\beta_0, 0) = 0$ because $\sigma_u^2 = 0$ implies $W = X$ and $m(X, Z, \beta_0) = E(Y \mid X, Z) = C'(\xi)$. We also have

$$\left.\frac{\partial Q(\beta, \sigma_u^2)}{\partial \beta^t}\right|_{\beta=\beta_0,\, \sigma_u^2=0} = -S_0, \tag{23}$$

where $S_0 = \mathrm{E}\,(C''(\xi)\xi_\beta\xi_\beta^t)$ with $\xi = \xi(X, Z, \beta)$. The matrix S_0 is non-singular due to (ix). Therefore, by the implicit function theorem, a unique solution β to the equation $Q(\beta, \sigma_u^2) = 0$ exists in a neighborhood of β_0 if $\sigma_u^2 \leq \sigma_0^2$ with some $\sigma_0^2 > 0$. Obviously $\beta_* = \beta_*(\sigma_u^2)$.

But this solution is unique not only in a neighborhood of β_0 but on the whole convex set Θ_β. Indeed, suppose we had two solutions β_1 and β_2 such that $Q(\beta_1) = Q(\beta_2) = 0$, where we suppressed the dependence on σ_u^2. Let $\beta(t) = t\beta_1 + (1-t)\beta_2$, $0 \leq t \leq 1$. Then $\beta(t) \in \Theta_\beta$. Let $q(t) = (\beta_1 - \beta_2)^t Q[\beta(t)]$. Then $q(0) = q(1) = 0$, and there exists a $\bar{t}$, $0 \leq \bar{t} \leq 1$, such that $\frac{dq}{dt}(\bar{t}) = 0$. But

$$\frac{dq}{dt} = (\beta_1 - \beta_2)^t \frac{\partial Q}{\partial \beta^t}(\beta_1 - \beta_2) < 0$$

unless $\beta_1 = \beta_2$ because, due to the linearity of $\xi(\beta)$, see (xiii), $\xi_{\beta\beta^t} = 0$ and thus

$$\frac{\partial Q(\beta, \sigma_u^2)}{\partial \beta^t} = -\mathrm{E}\,[C''(\xi)\xi_\beta\xi_\beta^t], \tag{24}$$

which is negative definite for all $\beta \in \Theta_\beta$ by assumption (ix).

8.2. Proof of Theorem 2.1, parts (b) and (c). We want to apply Theorem 12.1 from Heyde (1997). Consider

$$p_\delta := \lim_{n\to\infty} \sup_{\|\beta-\beta_0\|=\delta} (\beta - \beta_0)^t \frac{1}{n} \sum_{i=1}^{n} \psi(Y_i, W_i, Z_i, \beta).$$

As Q is the limit of $\frac{1}{n}\sum \psi$, see Subsection 8.1,

$$p_\delta = \sup_{\|\beta-\beta_0\|=\delta} (\beta - \beta_0)^t Q(\beta, \sigma_u^2).$$

We have to show that $p_\delta < 0$ for some small $\delta > 0$. Let C_0 be the compact set $C_0 := \{\beta \mid \|\beta - \beta_0\| = \delta\}$. Since $Q(\beta, \sigma_u^2)$ tends to $Q(\beta, 0)$ uniformly on C_0 as $\sigma_u^2 \to 0$, we have $Q(\beta, \sigma_u^2) = Q(\beta, 0) + R_1$ with $\sup_{C_0} \|R_1\| < \delta^2$ for $\sigma_u^2 < \sigma_0^2$ with some sufficiently small $\sigma_0^2 = \sigma_0^2(\delta)$. Now, since $Q(\beta_0, 0) = 0$, see Subsection 8.1, we have by (23)

$$Q(\beta, 0) = -S_0(\beta - \beta_0) + R_2$$

with $\sup_{C_0} \|R_2\| = O(\delta^2)$. Therefore,

$$p_\delta = \sup_{C_0} \big[-(\beta - \beta_0)^t S_0 (\beta - \beta_0) \big] + O(\delta^3) \qquad \text{as} \quad \delta \to 0.$$

By assumption (ix), S_0 is positive definite. Hence, $p_\delta < 0$ for sufficiently small δ and $\sigma_u^2 < \sigma_0^2$, and by Theorem 12.1 from Heyde (1997), for $\sigma_u^2 < \sigma_0^2$, the equation (2) eventually has a solution $\hat{\beta}_N$.

By arguments similar to those in the proof of Theorem KS 4.1, part (b), we can now prove that, for $\sigma_u^2 < \sigma_0^2$, $\hat{\beta}_N$ converges a.s. to the unique solution $\beta_*(\sigma_u^2)$ of the limit estimating equation (3).

8.3. Proof of Theorem 2.1, part (d). Let β_* be the (unique) solution to equation (3). Substituting KS(68) in (3) we obtain

$$\text{(25)}\quad \mathrm{E}\left\{\left[C' + \frac{1}{2}\sigma_u^2\left(-2\frac{W-\mu_w}{\sigma_w^2}C''\xi_x + C''\xi_{xx} + C'''\xi_x^2\right)\right]\xi_\beta - C'(\xi_*)\xi_\beta\right\} + O(\sigma_u^4) = 0.$$

Here ξ, ξ_x, ξ_{xx} are taken at the point (W, Z, β_0) and $C^{(i)} = C^{(i)}(\xi)$, $i = 1, 2, 3$, while $\xi_* = \xi(Z, W, \beta_*)$. Note that ξ_β is independent of β by assumption (xiii). We expand $C'(\xi_*)$ at $\beta = \beta_0$ with $\Delta\beta := \beta_* - \beta_0$, taking account of the linearity of $\xi(\beta)$:

$$\text{(26)}\qquad C'(\xi_*) = C' + C''\xi_\beta^t\Delta\beta + R\cdot\|\Delta\beta\|^2,$$

where, due to (ii) and (vi),

$$|R| \le \mathrm{const}(e^{A|W|} + e^{A\|Z\|}).$$

We substitute (26) in (25) and obtain

$$\text{(27)}\quad \mathrm{E}\left\{\frac{1}{2}\sigma_u^2\left[-2\frac{W-\mu_w}{\sigma_w^2}C''\xi_x + C''\xi_{xx} + C'''\xi_x^2\right]\xi_\beta - C''\xi_\beta\xi_\beta^t\Delta\beta\right\} + O(\sigma_u^4) + O(1)\|\Delta\beta\|^2 = 0.$$

Now, due to the implicit function theorem (see Subsecton 8.1), $\Delta\beta = \Delta\beta(\sigma_u^2)$ and $\Delta\beta \in C^1([0, \sigma_0^2])$ for sufficiently small σ_0^2. Also $\Delta\beta(0) = 0$ and hence $\Delta\beta(\sigma_u^2) = O(\sigma_u^2)$ and $\|\Delta\beta(\sigma_u^2)\|^2 = O(\sigma_u^4)$. Therefore we get from (27) with S from (5)

$$\Delta\beta = \frac{1}{2}\sigma_u^2 S^{-1}\mathrm{E}\left[-2\frac{W-\mu_w}{\sigma_w^2}C''\xi_x + C''\xi_{xx} + C'''\xi_x^2\right]\xi_\beta + O(\sigma_u^4).$$

Now, integration by parts according to KS(81) yields (4) and proves the statement. □

8.4. Proof of Theorem 2.2. First, we expand B_*, see (6). From (1) we have

$$\text{(28)}\qquad \psi(Y, W, Z, \beta_*) = (Y - m)\xi_\beta + (m - C'_*)\xi_\beta,$$

where $m = m(W, Z, \beta_0)$ and $C'_* := C'\{\xi(W, Z, \beta_*)\}$; note that ξ_β is independent of β. Therefore,

$$\text{(29)}\qquad B_* = \mathrm{E}\,[v\xi_\beta\xi_\beta^t + (m - C'_*)^2\xi_\beta\xi_\beta^t)].$$

The difference $m - C'_* = (m - C'_0) + (C'_0 - C'_*)$, where $m - C'_0 = m - C'\{\xi(W, Z, \beta_0)\}$, is of order σ_u^2, see KS(68), and note that $\|\beta_* - \beta_0\| = O(\sigma_u^2)$ by Theorem 2.1 (d). Therefore

$$B_* = \mathrm{E}\,(v\xi_\beta\xi_\beta^t) + O(\sigma_u^4).$$

Using KS(70) and again KS(81), we get with S from (5)

$$\varphi^{-1}B_* = S + \frac{1}{2}\sigma_u^2 \mathrm{E}\Big\{2\varphi^{-1}C''^2\xi_x^2\xi_\beta\xi_\beta^t + (C'''\xi_{xx} - C^{(4)}\xi_x^2)\xi_\beta\xi_\beta^t - 2C'''(\xi_x\xi_\beta\xi_\beta^t)_x\Big\} + O(\sigma_u^4) =: S + \frac{1}{2}\sigma_u^2 \Delta B_* + O(\sigma_u^4), \tag{30}$$

where ξ and its derivatives are taken at (W, Z, β_0).

Next we consider A_*. From (6) we have, using again Theorem 2.1 (d),

$$\begin{aligned} A_* &= \mathrm{E}\, C_*''\xi_\beta\xi_\beta^t = \mathrm{E}\left\{(C'' + \frac{1}{2}\sigma_u^2 C'''\xi_\beta^t\Delta\beta_*)\xi_\beta\xi_\beta^t\right\} + O(\sigma_u^4) \\ &= S + \frac{1}{2}\sigma_u^2\mathrm{E}\left\{C'''\xi_\beta^t\Delta\beta_*\xi_\beta\xi_\beta^t\right\} + O(\sigma_u^4). \end{aligned} \tag{31}$$

Then by (7), (30), and (31)

$$\varphi^{-1}\Sigma_N = S^{-1} + \frac{\sigma_u^2}{2}S^{-1}\big[-2\mathrm{E}\left\{C'''\xi_\beta^t\Delta\beta_*\xi_\beta\xi_\beta^t\right\} + \Delta B_*\big]S^{-1} + O(\sigma_u^4). \tag{32}$$

Now, (32) and KS(22) imply (8). □

8.5. Proof of Theorem 3.1. We apply Theorem 2.1 and evaluate (4). We have

$$\xi = \beta_0 + \beta_1 W, \qquad \xi_x = \beta_1, \qquad \xi_\beta = (1, W)^t, \qquad \xi_{x\beta} = (0,1)^t,$$

$$S = \mathrm{E}\, e^\xi \begin{pmatrix} 1 & W \\ W & W^2 \end{pmatrix} = d\begin{pmatrix} 1 & g \\ g & g^2 + \sigma_w^2 \end{pmatrix},$$

where g is given in (16), and

$$d := \exp\Big(\beta_0 + \beta_1\mu_w + \frac{1}{2}\beta_1^2\sigma_w^2\Big).$$

Here we used the following identities (except for the last one, which we shall need below):

$$\mathrm{E}\, e^\xi = d, \quad \mathrm{E}\, We^\xi = gd, \quad \mathrm{E}\, W^2e^\xi = (g^2 + \sigma_w^2)d, \quad \mathrm{E}\, W^3e^\xi = (3g\sigma_w^2 + g^3)d.$$

Next,

$$S^{-1} = \sigma_w^{-2}d^{-1}\begin{pmatrix} g^2 + \sigma_w^2 & -g \\ -g & 1 \end{pmatrix}. \tag{33}$$

By (4) and because $C = e^\xi$,

$$\begin{aligned} \Delta\beta_* &= -S^{-1}\mathrm{E}\,[C'''\xi_x^2\xi_\beta + 2C''\xi_x\xi_{x\beta}] = -S^{-1}\mathrm{E}\, e^\xi[\beta_1^2(1,W)^t + 2\beta_1(0,1)^t] \\ &= -S^{-1}d(\beta_1^2, \beta_1^2 g + 2\beta_1)^t = \frac{\beta_1}{\sigma_w^2}(\mu_w + g, -2)^t, \end{aligned}$$

and part (a) is proved. Part (b) follows from Theorem KS 5.1. In order to show (19), we have to evaluate the right-hand side of (8). The expectation in (8) equals

$$\begin{aligned}
&\mathrm{E}\left\{C''4\begin{pmatrix}0&0\\0&1\end{pmatrix}+4C'''\beta_1\frac{\partial}{\partial W}\begin{pmatrix}1&W\\W&W^2\end{pmatrix}+2C^{(4)}\beta_1^2\begin{pmatrix}1&W\\W&W^2\end{pmatrix}\right.\\
&\qquad\left.+2C'''[(1,W)\Delta\beta_*]\begin{pmatrix}1&W\\W&W^2\end{pmatrix}\right\}\\
&=\mathrm{E}\left\{e^{\xi}\left[4\begin{pmatrix}0&0\\0&1\end{pmatrix}+4\beta_1\begin{pmatrix}0&1\\1&2W\end{pmatrix}+2\beta_1^2\begin{pmatrix}1&W\\W&W^2\end{pmatrix}\right.\right.\\
&\qquad\left.\left.+\frac{2\beta_1}{\sigma_w^2}(\mu_w+g-2W)\begin{pmatrix}1&W\\W&W^2\end{pmatrix}\right]\right\}\\
&=4d\left[\begin{pmatrix}0&0\\0&1\end{pmatrix}+\beta_1\begin{pmatrix}0&1\\1&2g\end{pmatrix}+\frac{\beta_1 g}{\sigma_w^2}\begin{pmatrix}1&g\\g&g^2+\sigma_w^2\end{pmatrix}\right.\\
&\qquad\left.-\frac{\beta_1}{\sigma_w^2}\begin{pmatrix}g&g^2+\sigma_w^2\\g^2+\sigma_w^2&3g\sigma_w^2+g^3\end{pmatrix}\right]=4d\begin{pmatrix}0&0\\0&1\end{pmatrix}.
\end{aligned}$$

Next, by (33),

$$S^{-1}4d\begin{pmatrix}0&0\\0&1\end{pmatrix}S^{-1}=\frac{4d^{-1}}{\sigma_w^4}\begin{pmatrix}g^2&-g\\-g&1\end{pmatrix}. \tag{34}$$

From (34) and (8) we get (19). □

8.4. Proof of Theorem 4.1, part (a). We first consider

$$\Phi=\mathrm{E}\left(v^{-1}\frac{\partial m}{\partial\beta}\frac{\partial m}{\partial\beta^t}\right)$$

from condition (viii). From KS(70) we have

$$\varphi^{-1}v=C''+\chi^{-2}C''^2\xi_x^2+\sigma_u^2\cdot R, \tag{35}$$

where, as usual, here and in the sequel, R satisfies KS(69). Similarly, by KS(79),

$$\frac{\partial m}{\partial\beta}\frac{\partial m}{\partial\beta^t}=C''^2\xi_\beta\xi_\beta^t+\sigma_u^2\cdot R. \tag{36}$$

Now (35) and (36) imply

$$\frac{\varphi}{v}\frac{\partial m}{\partial\beta}\frac{\partial m}{\partial\beta^t}=\frac{\chi^2C''\xi_\beta\xi_\beta^t}{\chi^2+C''\xi_x^2}+\sigma_u^2\cdot R=\chi^2\frac{C''\xi_\beta\xi_\beta^t}{v_0}+\sigma_u^2\cdot R$$

with v_0 from (20). Hence

$$\varphi\Phi=\chi^2\mathrm{E}\,(C''\xi_\beta\xi_\beta^t v_0^{-1})+O(\sigma_u^2). \tag{37}$$

We now turn to F_p, $p = 1, 2$, defined in KS(17). By KS(78)

$$\left|\frac{\partial m}{\partial \gamma_p}\right| \le \varphi \cdot |R|, \qquad p = 1, 2,$$

where R satisfies KS(69). By KS(74) it follows that $\left|\frac{1}{v}\frac{\partial m}{\partial \gamma_p}\right|$ also satisfies KS(69). Therefore

$$F_p = O(1). \tag{38}$$

Now applying (38) to KS(18), we get

$$\varphi^{-1}\Sigma_{SQS} = (\varphi\Phi)^{-1} + O(\varphi), \tag{39}$$

and the statement follows by substituting (37) in (39) and taking account of (xiv).

8.7. Proof of Theorem 4.1, parts (b) and (c). From KS(88) we have

$$\varphi^{-1}B_c = \mathrm{E}\, C''\xi_\beta\xi_\beta^t(1 + \chi^{-2}C''\xi_x^2) + O(\sigma_u^2). \tag{40}$$

From KS(89) we get

$$A_c = S + O(\sigma_u^2). \tag{41}$$

Therefore

$$\varphi^{-1}\Sigma_{CS} = A_c^{-t}(\varphi^{-1}B_c)A_c^{-1} = S^{-1}\mathrm{E}\, C''\xi_\beta\xi_\beta^t(1 + \chi^{-2}C''\xi_x^2)S^{-1} + O(\sigma_u^2). \tag{42}$$

Multiplying by φ and taking account of (xiv) and (20), we get the desired result.

For Σ_N we have the same expansion (42) because the expansions (40), (41) are valid for A_* and B_*, too, see (30), (31). □

8.8. Proof of Theorem 4.2. We apply the expressions for Σ_{SQS} and Σ_{CS} from Theorem 4.1 (a) and (b). Note that by our assumptions the distribution of $v_0 = \chi^2 + C''\xi_x^2$ as a function of (W, Z, β_0) has no atoms. Note also that according to Remark KS 4.3 S can be assumed nonsingular for sufficiently small σ_u^2. Let $w = S^{-1/2}\sqrt{C''}\xi_\beta$. Then $\mathrm{E}ww' = I$, and part (a) of Theorem 4.2 follows from the Lemma in the Appendix.

To prove (21), consider the difference

$$D(\sigma_u^2, \chi^2) := \sigma_u^{-2}\lambda_{\max}(\Sigma_{CS} - \Sigma_{SQS})$$

as a function of σ_u^2 and χ^2. From Theorem 4.1 and by the continuity of $\lambda_{\max}(\cdot)$ we have

$$D(\sigma_u^2, \chi^2) = \lambda_{\max}\left\{\mathrm{E}\,(S^{-1}C''\xi_\beta\xi_\beta^t v_0 S^{-1}) - [\mathrm{E}\,(C''\xi_\beta\xi_\beta^t v_0^{-1})]^{-1}\right\} + O(\sigma_u^2).$$

Denote the first term on the right-hand side by $D(0, \chi^2)$, i.e.,

$$D(\sigma_u^2, \chi^2) = D(0, \chi^2) + O(\sigma_u^2).$$

Note that $D(0,\chi^2)$ does not depend on σ_u^2. It follows that

$$\liminf_{(\sigma_u^2\to 0, C_1\le\chi^2\le C_2)} D(\sigma_u^2,\chi^2) = \inf_{C_1\le\chi^2\le C_2} D(0,\chi^2).$$

But $D(0,\chi^2) > 0$ for all χ^2 because the difference in the argument of $\lambda_{\max}$ is positive semidefinite and $\neq 0$ by the Lemma of the Appendix, as proved in part (a). As χ^2 is restricted to a closed interval and $D(0,\chi^2)$ is continuous in χ^2, we finally have

$$\inf_{C_1\le\chi^2\le C_2} D(0,\chi^2) > 0,$$

wich implies (21). Inequality (22) is proved in a similar way. □

8.9. Sketch of a proof of Theorem 6. First it can be shown by arguments similar to those used in KS 7.7 to prove KS(21) that the estimation of the nuisance parameters including φ affects the asymptotic covariance matrix of $\widehat{\beta}_{ML}$ only in terms of order σ_u^4. Therefore, we may proceed in the following as if μ_x, σ_x^2, and φ were known.

The first step is to expand the model density ρ in powers of σ_u^2. Expanding ξ^* and $C(\xi^*)$, we find that

$$\rho = \frac{e^A}{\sqrt{2\pi}\sigma_x}\left[1 + (D+B^2)\frac{\sigma_u^2}{2}\right] + O(\sigma_u^4), \tag{43}$$

where

$$B := -\varphi^{-1}(y-C')\xi_x + V, \tag{44}$$

$$D := \varphi^{-1}[(y-C')\xi_{xx} - C''\xi_x^2] - \frac{1}{\sigma_x^2}, \tag{45}$$

$$V := \frac{w-\mu_x}{\sigma_x^2}, \tag{46}$$

and A is an expression that will cancel out later on. From (43) we derive an expansion of the integrand of the information matrix:

$$\frac{\rho_\beta\rho_\beta^t}{\rho^2} = \varphi^{-2}(y-C')^2\xi_\beta\xi_\beta^t + \frac{\sigma_u^2}{2}G + O(\sigma_u^4) \tag{47}$$

with

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

where the subscript S denotes the symmetrization operator, see the definition after KS(79). The matrix G can be further evaluated with the help of (44) and (45):

$$\begin{aligned}\varphi^2 G = {} & 2(y-C')\big[C''(2\xi_x\xi_\beta\xi_\beta^t V - \xi_{xx}\xi_\beta\xi_\beta^t - \xi_x(\xi_{x\beta}\xi_\beta^t)_S) - C'''\xi_x^2\xi_\beta\xi_\beta^t\big] \\ & + (y-C')^2\big[-2(\xi_{x\beta}\xi_\beta^t)_S V + (\xi_{xx\beta}\xi_\beta^t)_S - 4\varphi^{-1}C''\xi_x^2\xi_\beta\xi_\beta^t\big] \\ & + 2\varphi^{-1}(y-C')^3\xi_x(\xi_{x\beta}\xi_\beta^t)_S.\end{aligned} \tag{48}$$

As can be seen from (48), G consists of terms of the form $(y - C')^k h$, $k = 1, 2, 3$, where h is a function of w. The same is true for the other term in the integrand of the information matrix, see (47). We therefore investigate the expectation of these terms with (Y, W) in place of (y, w) 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 find with some algebra for $k = 1, 2, 3$:

$$\mathrm{E}\,(Y - C')h = O(\sigma_u^2), \tag{49}$$

$$\mathrm{E}\,(Y - C')^2 h = \varphi \mathrm{E}\,C'' h + \frac{\sigma_u^2}{2}\mathrm{E}\,[2C''^2 \xi_x^2 h - \varphi C'''(2\xi_x h_x + \xi_{xx} h) - \varphi C^{(4)} \xi_x^2 h] + O(\sigma_u^4), \tag{50}$$

$$\mathrm{E}\,(Y - C')^3 h = \varphi^2 \mathrm{E}\,C''' h + O(\sigma_u^2). \tag{51}$$

The most difficult to derive of these three equations is (50). We first express C' and h as functions of X rather than W in order to use KS(2) for the conditional mean and variance of Y given X. Then we transform the resulting functions of X back into functions of W using KS(9) and KS(10) and the integration by part formula KS(81).

With the help of (49) to (51) we can now expand $\mathrm{E}\,G$ and finally the information matrix:

$$\begin{aligned} I_\beta = \varphi^{-1}\Big\{ &\mathrm{E}\,C''\xi_\beta\xi_\beta^t - \frac{\sigma_u^2}{2}\mathrm{E}\,\big[2\varphi^{-1}C''^2\xi_x^2\xi_\beta\xi_\beta^t + C''\big((\xi_{xx\beta}\xi_\beta^t)_S + 4\xi_{x\beta}\xi_{x\beta}^t\big) \\ &+ C'''\big(2\xi_x(\xi_\beta\xi_\beta^t)_x + \xi_{xx}\xi_\beta\xi_\beta^t\big) + C^{(4)}\xi_x^2\xi_\beta\xi_\beta^t\big]\Big\} + O(\sigma_u^4) \\ =:{}& \varphi^{-1}(S - \frac{\sigma_u^2}{2}Q) + O(\sigma_u^4). \end{aligned}$$

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

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

and this is seen to be equal to KS(22). □

9. Appendix

9.1. A matrix inequality.

Lemma. *Let v be a positive random variable and w a random (column) vector with $E(ww^t) = I$. Assume $\mathrm{E}\,(v^{-1}w^t w) < \infty$ and $\mathrm{E}\,(v w^t w) < \infty$, then*

$$\Delta := \mathrm{E}\,(v w w^t) - [\mathrm{E}\,(v^{-1} w w^t)]^{-1}$$

is positive semidefinite. If in addition the distribution of v has no atoms, then the difference Δ is not equal to zero.

Proof. Let

$$q = [\mathrm{E}\,(v^{-1} w w^t)]^{-1}\frac{w}{\sqrt{v}} - \sqrt{v}w.$$

Then $\mathrm{E} q q^t = \Delta$, and Δ is positive semidefinite, see also Shklyar and Schneeweiss (2005).

In Kukush and Schneeweiss (2004a) it is proved that, under the assumptions of the Lemma,

$$\operatorname{tr} \mathrm{E}\,(vww^t) > \operatorname{tr}[\mathrm{E}\,(v^{-1}ww^t)]^{-1}.$$

This implies $\Delta \neq 0$. $\square$

Acknowledgement. We thank T. Augustin and R. Wolf for useful discussions. We also thank an anonymous referee for some very helpful recommendations, in particular, for the suggestion to include an investigation of the properties of the ML estimator in the paper. Support from Deutsche Forschungsgemeinschaft is gratefully acknowledged.

References

[1] A. C. Cameron and P. K. Trivedi (1998), *Regression Analysis of Count Data*, Cambridge University Press, London.

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

[3] C.-L. Cheng and H. Schneeweiss (1998), *Polynomial regression with errors in the variables*, J. Roy. Statist. Soc. B, **60**, 189–199.

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

[5] C. C. Heyde (1997), *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*, Springer, New York.

[6] A. Kukush and H. Schneeweiss (2000), *A Comparison of Asymptotic Covariance Matrices of Adjusted Least Squares and Structural Least Squares in Error Ridden Polynomial Regression*, Discussion Paper 218, Sonderforschungsbereich 386, Univ. Munich.

[7] A. Kukush and H. Schneeweiss (2004a), *A note on a matrix inequality for generalized means*, Linear Algebra and its Applications, **388**, 289–294.

[8] A. Kukush and H. Schneeweiss (2004b), *Relative Efficiency of Maximum Likelihood and Other Estimators in a Nonlinear Regression Model with Small Measurement Errors*, Discussion Paper 396, Sonderforschungsbereich 386, Univ. Munich.

[9] A. Kukush and H. Schneeweiss (2005), *Comparing the efficiencies of different estimators in a nonlinear measurement error model, I*, Math. Methods Statist., **14**, 53–79.

[10] A. Kukush, H. Schneeweiss, and R. Wolf (2002), *Comparing Different Estimators in a Nonlinear Measurement Error Model*, Discussion Paper 244, Sonderforschungsbereich 386, Univ. Munich.

[11] A. Kukush, H. Schneeweiss, and R. Wolf (2004), *Three estimators in the Poisson regression model with measurement errors*, Statist. Papers, **45**, 351–368.

[12] A. Kukush, H. Schneeweiss, and R. Wolf (2005), *Relative efficiency of three estimators in a polynomial regression with measurement errors*, J. Statist. Plann. Inference, **127**, 179–203.

[13] M. Schervish (1995), *Theory of Statistics*, Springer, New York.

[14] H. Schneeweiss (2005), *The Polynomial and the Poisson Measurement Error Models — Some Further Results on Quasi Score and Corrected Score Estimation*, Discussion Paper 446, Sonderforschungsbereich 386, Univ. Munich.

[15] S. Shklyar and A. Kukush (2002), *Comparison of Three Estimators in Poisson Errors-in-Variables Models with One Covariate*, Discussion Paper 293, Sonderforschungsbereich 386, Univ. Munich.

[16] S. Shklyar and H. Schneeweiss (2005), *A comparison of asymptotic covariance matrices of three consistent estimators in the Poisson regression model with measurement errors*, J. Mulivariate Anal., **94**, 250–270.

[17] S. Shklyar and H. Schneeweiss, and A. Kukush (2005), *Quasi Score is more Efficient than Corrected Score in a Polynomial Measurement Error Model*, Discussion Paper 445, Sonderforschungsbereich 386, Univ. Munich.

[18] C. J. Small and J. Wang (1993), *Numerical Methods of Nonlinear Equations*, Clarendon Press, Oxford.

[19] M. Thamerus (1998), *Different nonlinear regression models with incorrectly observed covariates.* In: *Econometrics in Theory and Practice*, R. Galata and H. Küchenhoff (eds.), Festschrift for Hans Schneeweiss, Physica, Heidelberg–New York.

[20] R. Winkelmann (1997), *Econometric Analysis of Count Data*, 2nd ed., Springer, Berlin.

[21] R. Wolf (2004), *Vergleich von funktionalen und strukturellen Messfehlerverfahren*, Logos, Berlin.

[*Received October* 2002; *revised December* 2004]