Kukush, Schneeweiss, Wolf:

Comparing Different Estimators in a Nonlinear Measurement Error Model

Sonderforschungsbereich 386, Paper 244 (2002)

Comparing Different Estimators in a Nonlinear Measurement Error Model

Discussion Paper 244, 2002

Alexander Kukush, Department of Mathematics, Taras Shevchenko University, 01601 Kiev, Ukraine, alexander_kukush@univ.kiev.ua

Hans Schneeweiss, Institute of Statistics, University of Munich, Akademiestr. 1, 80799 Munich, Germany, schneew@stat.uni-muenchen.de

Roland Wolf, Institute of Statistics, University of Munich, 80799 Munich, Akademiestr. 1, Germany, wolf@stat.uni-muenchen.de

Abstract

A nonlinear structural errors-in-variables model is investigated, where the response variable has a density belonging to an exponential family and the error-prone covariate follows a Gaussian distribution. Assuming the error variance to be known, we consider two consistent estimators in addition to the naive estimator. We compare their relative efficiencies by means of their asymptotic covariance matrices for small error variances. The structural quasi score (SQS) estimator is based on a quasi score function, which is constructed from a conditional mean-variance model. Consistency and asymptotic normality of this estimator is proved. The corrected score (CS) estimator is based on an error-corrected likelihood score function. For small error variances the SQS and CS estimators are approximately equally efficient. The polynomial model and the Poisson regression model are explored in greater detail.

Key Words: Exponential family, structural errors-in-variables model, asymptotic covariance matrix, efficiency, polynomial regression, Poisson regression, small measurement error variance

1 Introduction

Measurement error (or errors-in-variables) models have been extensively studied over the last decades, see, e.g., the monographs of Schneeweiss and

Mittag (1986), Fuller (1987), Cheng and Van Ness (1999), and Carroll et al (1995). In particular, the last book deals almost exclusively with nonlinear regression models.

In this paper, we study a rather general nonlinear model, where the response variable has a density belonging to an exponential family, the canonical parameter of which depends on covariates in a nonlinear way with an unknown parameter vector β to be estimated. We also have a dispersion parameter φ, which may or may not be known. One of the covariates is unobservable; it can only be observed with a Gaussian latent measurement error u.

It is well known that, in such a situation, ignoring the measurement error in the estimation procedure gives rise to an estimator - the so-called *naive* estimator - which will typically be inconsistent. However, consistent estimators of β are available, in particular if the measurement error variance σ_u^2 is known, and this will be assumed in the present paper.

We will consider two consistent estimators. Assuming a Gaussian distribution for the error-prone latent covariate, we can construct a *structural quasi score* (SQS) estimator of β. It is based on a conditional mean-variance model, conditioned on the observed covariate, which can be derived from the original model, see Carroll and Ruppert (1988), Heyde (1997), Carroll et al (1995), and Thamerus (1998) for special cases. We show that the SQS estimator is consistent, asymptotically normal, and eventually (i.e., for large enough sample size) unique. We also prove the convergence of an iteratively reweighted least squares algorithm. The asymptotic covariance matrix of the SQS estimator turns out to include special terms which stem from the necessity of estimating the mean and variance of the error-prone covariate distribution as nuisance parameters. These additional terms can, however, be neglected for small σ_u^2, more precisely: they are of the order σ_u^4.

The other consistent estimator of β considered in this paper is a *corrected score* (CS) estimator, which is based on solving a corrected score estimating equation, see Stefanski (1989), Nakamura (1990), Buonaccorsi (1996), Carroll et al (1995). In contrast to the SQS estimator, knowledge of the distribution of the error-prone covariate is not required for the CS procedure. The CS estimator is asymptotically normal with an asymptotic covariance matrix, wich can be evaluated.

The same can be done for the naive estimator, although this estimator is asymptotically biased. We can, however, determine its asymptotic bias.

We want to compare the relative asymptotic efficiency of these three estimators in terms of their asymptotic covariance matrices. Such comparisons have been carried out with the help of Monte Carlo simulations, e.g., Kuha and Temple (1999), Schneeweiss and Nittner (2001). But to the best of our knowledge, theoretical comparative studies have not been carried out before except for the special cases of the polynomial and the Poisson regression models, Kukush and Schneeweiss (2000), Kukush et al (2001a), (2001b).

It seems that the asymptotic covariance matrices are hard to compare in general. We can only do so in certain border line cases where either σ_u^2 or both σ_u^2 and φ are small. (Another borderline case, for large errors, has been dealt with for the polynomial regression in Kukush and Schneeweiss (2000).)

When only σ_u^2 tends to zero and φ stays fixed, it turns out, surprisingly enough, that SQS and CS coincide in their asymptotic properties up to the order of σ_u^2. This is our main result.

On the other hand, when φ is proportional to σ_u^2 and both tend to zero, SQS is more efficient than CS and the latter becomes almost as efficient as the naive estimator.

These findings are specialized to two particular cases: the polynomial regression and the Poisson regression, where more detailed results can be presented.

In the next two sections the model is presented, and the SQS estimator is defined. In Section 4 its consistency, uniqueness and asymptotic normality are shown. The convergence of an iterative algorithm for constructing the estimator is also established. An expansion of the asymptotic covariance matrix for $\sigma_u^2 \to 0$ is given. In Section 5 we introduce the CS estimator and derive an expansion of its asymptotic covariance matrix. The equivalence of the covariance matrices of the SQS and CS estimators up to the order of σ_u^2 is established. In Section 6 we compute the asymptotic bias of the naive estimator and expand its asymptotic covariance matrix. Section 7 deals with two special models: polynomial and Poisson regression. In Section 8 we study the case when both σ_u^2 and φ are small. Section 9 has some simulation results. Section 10 contains some concluding remarks. Proofs are presented in Section 11. In the appendix, we first present a general expression for the covariance matrix of a parameter of interest when nuisance parameters are present. Secondly, we give an auxiliary matrix inequality, which might be of some interest also outside its particular application in the present paper.

2 The model

Throughout this paper we suppose that a response scalar random variable Y has a density $f(y|\xi)$ with respect to a σ-finite measure m on the Borel σ-field in $\mathbb{R}$ given by

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

This relation describes a density belonging to an exponential family with canonical parameter ξ. φ is a dispersion parameter, $\varphi > 0$, $c(y,\varphi)$ is measurable, the function $C(\cdot)$ is smooth enough (see Section 4.1), and $C''(\xi) > 0$, for all ξ. Then the mean and the variance of Y given ξ are, respectively,

$$\mathrm{E}(Y|\xi) = C'(\xi), \qquad \mathrm{V}(Y|\xi) = \varphi \cdot C''(\xi), \tag{2}$$

see, e.g., McCullagh and Nelder (1989). In general, we suppose that φ is unknown, but is known to belong to a fixed interval $[a_1, b_1], a_1 > 0, b_1 < \infty$. In some special cases, e.g., in the Poisson model, φ may be known.

We assume that

$$\xi = \xi(X, Z, \beta), \tag{3}$$

where X is an unobservable random scalar explanatory variable (covariate), Z is an observable random vector of further explanatory variables and β is a nonrandom vector of regression parameters; $\xi(\cdot,\cdot,\cdot)$ is smooth enough (see Section 4.1).

For each $i = 1, 2, \ldots, n$, let the triple (Y_i, X_i, Z_i) have a distribution given by

$$p_X(x)dx \cdot m_Z(dz) \cdot f\{y|\xi(x,z,\beta)\} \cdot m(dy),$$

where $p_X(\cdot)$ is the density of X, and m_Z is the distribution of Z. We assume in particular that $X_i \sim N(\mu_x, \sigma_x^2)$, with unknown μ_x, σ_x^2. Suppose also that the triples (Y_i, X_i, Z_i), $i = 1, 2, \ldots$ are i.i.d. The true predictors X_i are related to the observed surrogate covariates W_i through

$$W_i = X_i + U_i, \quad i = 1, \ldots, n,$$

where the U_i are i.i.d., independent of (Y_i, X_i, Z_i), and

$$U_i \sim N(0, \sigma_u^2). \tag{4}$$

We assume σ_u^2 to be known. Finally, $\beta \in \Theta_\beta$, where Θ_β is the closure of a given convex open bounded subset of $\mathbb{R}^k$.

The parameter β is to be estimated by the observations (Y_i, W_i, Z_i), $i = 1, \dots, n$.

3 The structural quasi score (SQS) estimator

Let us introduce the conditional mean and variance of Y

$$\begin{aligned} m(W, Z, \beta) &:= \mathrm{E}(Y|W, Z), & (5)\\ v(W, Z, \beta, \varphi) &:= \mathrm{V}(Y|W, Z). & (6) \end{aligned}$$

Using (2), we have, compare, e.g., Carroll et al (1995), Section 7.8, and Thamerus (1998):

$$\begin{aligned} m(W, Z, \beta) &= \mathrm{E}[C'\{\xi(X, Z, \beta)\}|W, Z], & (7)\\ v(W, Z, \beta, \varphi) &= \mathrm{V}[C'\{\xi(X, Z, \beta)\}|W, Z] + \varphi \mathrm{E}[C''\{\xi(X, Z, \beta)\}|W, Z] \\ &=: A_1(W, Z, \beta) + \varphi A_2(W, Z, \beta). & (8) \end{aligned}$$

The conditional distribution of X given W is $N\{\mu(W), \tau^2\}$, with

$$\begin{aligned} \mu(W) &= W - \frac{\sigma_u^2}{\sigma_w^2}(W - \mu_w), & (9)\\ \tau^2 &= \sigma_u^2 - \frac{\sigma_u^4}{\sigma_w^2}. & (10) \end{aligned}$$

Here $\mu_w := \mathrm{E}W = \mu_x$, $\sigma_w^2 := Var(W) = \sigma_x^2 + \sigma_u^2$.

As the conditional distrubution of X given W depends on the parameters μ_w and σ_w^2, therefore $m(W, Z, \beta)$ as well as $v(W, Z, \beta, \varphi)$, $A_1(W, Z, \beta)$, and $A_2(W, Z, \beta)$ all depend on these unknown nuisance parameters. Let $\hat{\mu}_w$ and $\hat{\sigma}_w^2$ be the sample mean and sample variance of W, respectively. Replacing μ_w and σ_w^2 with their estimates $\hat{\mu}_w$ and $\hat{\sigma}_w^2$, we obtain $\hat{m}$, $\hat{A}_1$, $\hat{A}_2$, and $\hat{v}$ from m, A_1, A_2, and v, respectively.

Now, we define the SQS estimators $\hat{\beta}_{SQS}$ and $\hat{\varphi}_{SQS}$ as measurable solutions to the conditionally asymptotically unbiased estimating equations

$$\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{m}(W_i, Z_i, \beta)\}\hat{v}^{-1}(W_i, Z_i, \beta, \varphi)\frac{\partial \hat{m}(W_i, Z_i, \beta)}{\partial \beta} = 0 \qquad (11)$$

$$\begin{aligned}\varphi &= \left[\frac{1}{n}\sum_{i=1}^{n}\hat{A}_2(W_i, Z_i, \beta)\right]^{-1} \\ &\times \left\{\frac{1}{n-k}\sum_{i=1}^{n}[Y_i - \hat{m}(W_i, Z_i, \beta)]^2 - \frac{1}{n}\sum_{i=1}^{n}\hat{A}_1(W_i, Z_i, \beta)\right\}, \qquad (12)\end{aligned}$$

$\beta \in \Theta_\beta$, $\varphi \in [a_1, b_1]$.

The idea behind (12) can be explained als follows. For the true values $\beta = \beta_0$, $\varphi = \varphi_0$, the right hand side of (12) converges a.s. to

$$\begin{aligned}& [\mathrm{E}A_2(W, Z, \beta_0)]^{-1}\{\mathrm{E}[Y - m(W, Z, \beta_0)]^2 - \mathrm{E}A_1(W, Z, \beta_0)\} \\ = \; & [\mathrm{E}A_2(W, Z, \beta_0)]^{-1}\{\mathrm{E}v(W, Z, \beta_0, \varphi_0) - \mathrm{E}A_1(W, Z, \beta_0)\} = \varphi_0,\end{aligned}$$

see (8).

4 Asymptotic properties of $\hat{\beta}_{\mathbf{SQS}}$

4.1 Further assumptions

Consider the model described in Section 2. Hereafter β_0 and φ_0 denote the true values of β and φ, and the expectation E is always taken with respect to the true parameter values. We introduce some further assumptions

(i) β_0 is an interior point of Θ_β, and $\varphi_0 \in (a_1, b_1)$.

(ii) $C(\xi) \in C^4(I\!R)$, and for all x, z, β,

$$|C^{(i)}\{\xi(x, z, \beta)\}| \le const(e^{A|x|} + e^{A\|z\|}), \quad i = 1, \dots, 4,$$

with some fixed $A > 0$ and $const > 0$.

Note that one could just as well have taken a bound of the form $const \cdot \exp[B(|X| + \|Z\|)]$ with some $B > 0$. Both types of bounds are equivalent. Adding or multiplying two such bounds yields a bound of the same type albeit with different constants.

(iii) $\mathrm{E}e^{A\|Z\|} < \infty$, for each $A > 0$.

Under (ii) and (iii) the conditional mean and variance of Y given W and Z are well defined and satisfy (7), (8).

(iv) For each x, z, β, with some $const > 0$ and $A > 0$

$$C''\{\xi(x,z,\beta)\} \geq const \cdot e^{-A|x|-A\|z\|},$$

We need (iv) to bound v^{-1}.

(v) $\xi(x,z,\beta)$ is defined for β in a neighborhood $U(\Theta_\beta)$ of Θ_β, and $\xi(\cdot, Z, \cdot) \in C^5(I\!R \times U(\Theta_\beta))$.

(vi) $\|D_x^i D_\beta^j \xi\| \leq const(e^{A|x|} + e^{A\|z\|})$,
for $0 \leq i \leq 4, j = 0, 1$.

(vii) For each $\varphi \in [a_1, b_1]$, the equation

$$\mathrm{E}[(m_0 - m)v^{-1}\frac{\partial m}{\partial \beta}] = 0$$

has the unique solution $\beta = \beta_0$, where $m_0 := m(W,Z,\beta_0)$, $m = m(W,Z,\beta)$, and $v = v(W,Z,\beta,\varphi)$.

Due to (vii), the limit equations for the system (11), (12) have the unique solution $\beta = \beta_0$, $\varphi = \varphi_0$. We introduce the compound parameter $\theta := (\beta^t, \varphi)^t$ and let $\theta_0 := (\beta_0^t, \varphi_0)^t$.

(viii) The matrix

$$\mathrm{E}\left(\frac{\partial m}{\partial \beta}\frac{\partial m}{\partial \beta^t}\right)\Bigg|_{\theta=\theta_0}$$

is positive definite.

It then follows that the following matrix Φ is also positive definite, where

$$\Phi := \mathrm{E}(v^{-1}\frac{\partial m}{\partial \beta}\frac{\partial m}{\partial \beta^t})\Bigg|_{\theta=\theta_0}.$$

4.2 Consistency and uniqueness of $\hat{\beta}_{\mathbf{SQS}}$

When we study various asymptotic properties of the estimators we shall often use the expression "eventually" to indicate that a certain property holds true for large enough n. The following definition makes this precise.

Definition 4.1. Let $U_1, U_2, \ldots$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. A sequence of statements $A_n(U_n)$, $n = 1, 2, \ldots$ is said to hold true *eventually* if

$$\exists\,\Omega_0, P(\Omega_0) = 1, \forall\omega \in \Omega_0 \quad \exists N(\omega) \quad \forall n \geq N(\omega) : A_n\{U_n(\omega)\} \text{ holds true.}$$

We can now state the following theorem about existence, uniqueness, and consistency of the SQS estimator.

Theorem 4.1. Assume (i) to (viii). Then:

a) eventually, the estimating equations (11), (12) have a solution $\hat{\beta}_{SQS} \in \Theta_\beta$ and $\hat{\varphi}_{SQS} \in [a_1, b_1]$,

b) as $n \to \infty$, $\hat{\beta}_{SQS} \to \beta_0$ and $\hat{\varphi}_{SQS} \to \varphi_0$ a.s.,

c) eventually, the solution of (11), (12) is unique.

4.3 The algorithm

We look for an algorithm to solve the system (11), (12) in the domain $\Theta_\beta \times [a_1, b_1]$. The fact that this domain is compact will be useful in the proof of asymptotic properties of the estimators. In addition, the restriction of the estimators to a compact domain may provide computational stability of the numerical procedure.

Denote by $h_n(\beta)$ the function on the right hand side of (12), and let

$$S_n(\beta, \alpha, \varphi) := \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{m}(W_i, Z_i, \beta)\}\hat{v}^{-1}(W_i, Z_i, \alpha, \varphi)\frac{\partial \hat{m}(W_i, Z_i, \beta)}{\partial \beta}. \quad (13)$$

Introduce also the projector P onto the interval $[a_1, b_1]$ with $P(u) = a_1$ if $u \leq a_1$, $P(u) = u$ if $a_1 \leq u \leq b_1$, and $P(u) = b_1$ if $u \geq b_1$. Now, we modify the equations (11), (12) to the form

$$S_n(\beta, \beta, \varphi) = 0, \quad (14)$$

$$\varphi = P \circ h_n(\beta). \quad (15)$$

The following algorithm to solve (14), (15) is a modification of the iteratively reweighted least squares procedure, see Carroll and Ruppert (1988).

1. Given an estimate $\beta^{(j)} \in \Theta_\beta$ from the j-th round of the algorithm, find $\varphi^{(j)}$ from (15), treating $\beta^{(j)}$ as known.

2. Solve the equation $S_n(\beta, \beta^{(j)}, \varphi^{(j)}) = 0$ for $\beta \in \Theta_\beta$, using $\hat{v}(W_i, Z_i, \alpha, \varphi)$ with $\alpha = \beta^{(j)}$ and $\varphi = \varphi^{(j)}$. The updated estimate $\beta^{(j+1)} \in \Theta_\beta$ is given by a weighted least squares estimate from regressing Y_i on $\hat{m}(W_i, Z_i, \beta)$, with weights

$$w_i^{(j)} := [\hat{v}(W_i, Z_i, \beta^{(j)}, \varphi^{(j)})]^{-1}.$$

   The corresponding unweighted least squares estimate $\hat{\beta}^* \in \Theta_\beta$ can be used as an initial value $\beta^{(0)}$ for β.

To show the convergence of the iterative procedure we have to strengthen assumption (vii).

(vii)' For each $\alpha \in \Theta_\beta$ and $\varphi \in [a_1, b_1]$, the equation

$$\mathrm{E}\left\{(m_0 - m)[v(W, Z, \alpha, \varphi)]^{-1} \frac{\partial m}{\partial \beta}\right\} = 0,$$

where m_0 and m are as in (vii), has the unique solution $\beta = \beta_0$.

Thus in contrast to (vii), we fix α in the function v.

Theorem 4.2. Assume (i) to (vi), and also (vii)'. Then:

a) eventually, the equation $S_n(\beta, \alpha, \varphi) = 0$ has a unique solution $\hat{\beta}_n(\alpha, \varphi)$ for arbitrary $\alpha \in \theta_\beta$ and $\varphi \in [a_1, b_1]$,

b) eventually, $\beta^{(j)} \to \hat{\beta}_{SQS}$ and $\varphi^{(j)} \to \hat{\varphi}_{SQS}$, as $j \to \infty$.

4.4 Asymptotic normality

According to (10) and (9), the conditional mean $m(W, Z, \beta)$ involves the nuisance parameters μ_w and σ_w^2. In view of (9), it is convenient to use instead $\gamma := (\mu_w, \sigma_w^{-2})^t =: (\gamma_1, \gamma_2)^t$ as a nuisance parameter. Let $\gamma_0 = (\mu_{w0}, \sigma_{w0}^{-2})^t$ be the true value of γ. For $p = 1, 2$ let

$$F_p := \mathrm{E}\left(v^{-1} \frac{\partial m}{\partial \beta} \frac{\partial m}{\partial \gamma_p}\right)\Bigg|_{\theta=\theta_0, \gamma=\gamma_0}. \tag{16}$$

Theorem 4.3. Assume (i) to (viii). Then

$$\sqrt{n}(\hat{\beta}_{SQS} - \beta_0) \xrightarrow{d} N(0, \Sigma_{SQS}),$$

with

$$\Sigma_{SQS} = \Phi^{-1} + \Phi^{-1}(\sigma_{w0}^2 F_1 F_1^t + \frac{2}{\sigma_{w0}^4} F_2 F_2^t)\Phi^{-1}, \qquad (17)$$

where Φ is given in condition (viii).

Remark 4.1. If μ_w and σ_w^2 are known, then $\Sigma_{SQS} = \Phi^{-1}$, see Carroll et al (1995). The additional terms in (17) appear because the sample estimators of γ are plugged in.

Remark 4.2. Theorem 4.3 and the formula for the asymptotic covariance matrix can be extended to the case, where the measurement error variance σ_u^2 is unknown, but some validation data (i.e., some additional observations of the latent variable X) are used to estimate it.

4.5 Expansion of Σ_{SQS}

We want to find approximating expressions for Σ_{SQS} for fixed φ and $\sigma_u^2 \to 0$. Hereafter ξ_β denotes the column vector $\frac{\partial \xi}{\partial \beta}$, and similarly $\xi_x := \frac{\partial \xi}{\partial x}$, $\xi_{xx} := \frac{\partial^2 \xi}{\partial x^2}$, etc. We shall need the following matrices:

$$S_0 \quad := \quad \mathrm{E}(C''(\xi)\xi_\beta \xi_\beta^t)\Big|_{\beta=\beta_0}, \quad \xi = \xi(X, Z, \beta), \qquad (18)$$

$$S \quad := \quad \mathrm{E}(C''(\xi)\xi_\beta \xi_\beta^t)\big|_{\beta=\beta_0}, \quad \xi = \xi(W, Z, \beta). \qquad (19)$$

We introduce a new assumption, related to (viii).

(ix) The matrix $\mathrm{E}(\xi_\beta \xi_\beta^t)\big|_{\beta=\beta_0}$ is positive definite, where $\xi_\beta = \xi_\beta(X, Z, \beta)$.

Remark 4.3. Assumptions (iv) and (ix) imply that the matrix S_0 of (18) is positive definite. Under assumptions (ii) to (vi), S tends to S_0, as $\sigma_u^2 \to 0$, and therefore under the additional assumption (ix), S is also positive definite for small enough σ_u^2.

Remark 4.4. Assumption (ix) is equivalent to the statement that the components of ξ_β at $\beta = \beta_0$ are linearly independent with positive probability, more precisely: For all $a \in R^k, a \neq 0 : P(a'\xi_\beta \neq 0) > 0$. Note that when $\xi(X, Z, \beta) = g(\beta_0 + \beta_1 X + \beta_z' Z)$ with some function g, then (ix) implies $\sigma_x^2 > 0$.

Therorem 4.4 Assume (i) to (vi), and (vii) for small enough σ_u^2, and (ix). Let φ_0 be fixed and $\sigma_u^2 \to 0$. Then

$$\Sigma_{SQS} = \Phi^{-1} + O(\sigma_u^4), \tag{20}$$

and furthermore

$$\begin{aligned} \Sigma_{SQS} &= \varphi S^{-1} \\ &+ \frac{1}{2}\sigma_u^2 \varphi S^{-1} \mathrm{E}\{2\varphi^{-1} C''^2 \xi_x^2 \xi_\beta \xi_\beta^t + C''[(\xi_\beta \xi_\beta^t)_{xx} + 2\xi_{x\beta}\xi_{x\beta}^t] \\ &+ C'''[2(\xi_x \xi_\beta \xi_\beta^t)_x - \xi_{xx}\xi_\beta \xi_\beta^t] + C^{(4)} \xi_x^2 \xi_\beta \xi_\beta^t\} S^{-1} \\ &+ O(\sigma_u^4), \end{aligned} \tag{21}$$

where $C^{(i)} = C^{(i)}(\xi)$ and ξ and the derivatives of ξ are taken at the point (W, Z, β_0).

5 The corrected score (CS) estimator

We start with the likelihood score function of β for the original model (1). Due to the structure of the exponential family (1), the likelihood score function is given by

$$\psi(y, x, z, \beta) = y\xi_\beta - C'(\xi)\xi_\beta, \tag{22}$$

where ξ and ξ_β are taken at the point (x, z, β). The estimating equation for the maximum likelihood estimator based on the observations $(X_i, Z_i, Y_i), i = 1, \dots, n$, is then given by $n^{-1}\sum_{i=1}^n \psi(Y_i, X_i, Z_i, \beta) = 0$, and its limit as $n \to \infty$ is

$$\mathrm{E}[\{C'(\xi_0) - C'(\xi)\}\xi_\beta] = 0,\ \beta \in \Theta_\beta, \tag{23}$$

where $\xi_0 := \xi(X, Z, \beta_0)$, and ξ and ξ_β are taken at the point (X, Z, β). We need the following assumption.

(x) The equation (23) has the unique solution $\beta = \beta_0$.

This is an identifiabily condition for the error-free model (1).

According to the approach of Carroll et al (1995), Chapter 6, see also Nakamura (1990), we want to introduce a corrected score function $\psi_c(y, w, z, \beta)$ such that

$$\mathrm{E}\{\psi_c(Y, W, Z, \beta)|Y, X, Z\} = \psi(Y, X, Z, \beta). \tag{24}$$

To this purpose, consider the functions $f_1(x,z,\beta) = \xi_\beta(x,z,\beta)$ and $f_2(x,z,\beta) = C'\{\xi(x,z,\beta)\}\xi_\beta(x,z,\beta)$. We are looking for new functions $f_{ic}(w,z,\beta)$ such that

$$\mathrm{E}\{f_{ic}(W,Z,\beta)|X,Z\} = f_i(X,Z,\beta),\ i=1,2. \tag{25}$$

We demand that:

(xi) For $f_1 = \xi_\beta$ and $f_2 = C'(\xi)\xi_\beta$ there exist solutions f_{1c}, f_{2c} of equations (25) defined for all β in the neighborhood $U(\Theta_\beta)$ of Θ_β, and $f_{ic}(w,z,\cdot) \in C^1(U(\Theta_\beta)), i=1,2$; see condition (v).

(xii) For $f_1 = \xi_\beta$ and $f_2 = C'(\xi)\xi_\beta$ and for the respective solutions f_{1c}, f_{2c} satisfying (xi) the following expansion holds at any point $(w,z,\beta) \in I\!R \times \mathrm{E}_z \times \Theta_\beta$ as long as $\sigma_u^2 \le \sigma_0^2$ for some fixed $\sigma_0^2 > 0$.

$$D_\beta^j f_{ic} = D_\beta^j f_i - \frac{1}{2}\sigma_u^2 D_\beta^j (f_i)_{xx} + \sigma_u^4 \cdot rest, \tag{26}$$

$i=1,2,\ j=0,1$, and

$$\|rest\| \le const(e^{A|w|} + e^{A\|z\|}) \tag{27}$$

with $const > 0$ and fixed $A > 0$.

We comment on the new assumptions. Due to (xi), the function

$$\psi_c(y,w,z,\beta) := y f_{1c}(w,z,\beta) - f_{2c}(w,z,\beta) \tag{28}$$

satisfies (24). The condition (xii) with $j=0$ gives an approximation of f_{1c}, f_{2c} (and therefore of ψ_c) for small enough σ_u^2. This approximation is based on the representation of the solution f_{ic} in the form

$$f_{ic} = f_i - \frac{1}{2}\sigma_u^2 (f_i)_{xx} + \sigma_u^4 \sum_{k=2}^{\infty} \frac{(-\sigma_u^2)^{k-2}}{2^k k!} \frac{\partial^{2k} f_i}{\partial x^{2k}}, \tag{29}$$

which was shown in Stefanski (1989), p.4344, in a regular case. Therefore for $j=0$ in (26)

$$rest = \sum_{k=2}^{\infty} \frac{(-\sigma_u^2)^{k-2}}{2^k k!} \frac{\partial^{2k} f_i}{\partial x^{2k}}, \tag{30}$$

and for this expression, with $\sigma_u^2 \leq \sigma_0^2$, we require the bound (27). To justify (26) and (27) for $j = 1$, one can assume that (25) is differentiable with respect to β, i.e.,

$$\mathrm{E}\{(f_{ic})_\beta(W, Z, \beta)|X, Z\} = (f_i)_\beta(X, Z, \beta),$$

and use a representation like (29) for $(f_{ic})_\beta$. We will see that assumptions (xi) and (xii) hold for the polynomial regression and for the Poisson regression model.

Now, for the corrected score function ψ_c given in (28), we define the corrected score (CS) estimator $\hat{\beta}_{CS}$ as a measurable solution to

$$\frac{1}{n}\sum_{i=1}^{n} \psi_c(Y_i, W_i, Z_i, \beta) = 0, \quad \beta \in \Theta_\beta. \tag{31}$$

Note that (23) is the limit estimating equation for $\hat{\beta}_{CS}$ (31) as $n \to \infty$.

In Carroll et al (1995) the asymptotic properties of $\hat{\beta}_{CS}$ are studied. Under (x)-(xii), $\hat{\beta}_{CS}$ is strictly consistent, i.e., $\hat{\beta}_{CS} \to \beta_0$ a.s. Introduce

$$A_c := -\mathrm{E}\frac{\partial \psi_c}{\partial \beta^t}\Big|_{\beta=\beta_0}, \; B_c := \mathrm{E}\psi_c \psi_c^t\Big|_{\beta=\beta_0}, \tag{32}$$

where ψ_c and $\psi_{c\beta^t}$ are taken at the point (Y, W, Z, β). The matrix A_c is symmetric and positive definite. Indeed, by (24), (22) and (2) we have

$$\mathrm{E}\psi_c(Y, W, Z, \beta) = \mathrm{E}\psi(Y, X, Z, \beta) = \mathrm{E}[\{C'(\xi_0) - C'(\xi)\}\xi_\beta],$$

where $\xi_0 := \xi(X, Z, \beta_0)$, $\xi = \xi(X, Z, \beta)$, and $\xi_\beta = \frac{\partial}{\partial \beta}\xi(X, Z, \beta)$. Differentiating with respect to β^t and setting $\beta = \beta_0$, we get (bearing in mind assumptions (ii) and (vi))

$$A_c = \mathrm{E}(C'' \xi_\beta \xi_\beta^t)\big|_{\beta=\beta_0} = S_0, \tag{33}$$

which is positive definite under assumption (ix), see Remark 4.3. Now

$$\sqrt{n}(\hat{\beta}_{CS} - \beta_0) \xrightarrow{d} N(0, \Sigma_{CS}),$$

where Σ_{CS} is given by the sandwich formula

$$\Sigma_{CS} = A_c^{-1} B_c A_c^{-1}. \tag{34}$$

The next statement is the central result of the paper.

Theorem 5.1. Assume (i) to (vi), and (vii) for small enough σ_u^2, and (ix) to (xii). Let φ be fixed and $\sigma_u^2 \to 0$. Then

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

6 The naive (N) estimator

The naive estimator $\hat{\beta}_N$ of β is defined as a measurable solution to

$$\frac{1}{n}\sum_{i=1}^{n} \psi(Y_i, W_i, Z_i, \beta) = 0, \qquad \beta \in \Theta_\beta. \tag{35}$$

Thus the likelihood score function (22) is used, but in (35) the unobservable regressor X_i is replaced with the observed surrogate covariate W_i. The limit equation is given by

$$\mathrm{E}[m(W, Z, \beta_0)\xi_\beta - C'(\xi)\xi_\beta] = 0, \qquad \beta \in \Theta_\beta, \tag{36}$$

where ξ and ξ_β are taken at the point (W, Z, β). We shall now investigate the properties of the naive estimator under the following restriction of 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 6.1. Assume (i) to (iii), and (v), (vi), (ix), (x), (xiii). Then for small enough σ_u^2:

a) the equation (36) has a unique solution $\beta_* = \beta_*(\sigma_u^2)$,

b) eventually, the equation (35) 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

$$\Delta\beta_* := -S^{-1}\mathrm{E}[(C''\xi_x\xi_\beta)_x + C''\xi_x\xi_{x\beta}], \tag{37}$$

where S is given in (19), $C'' = C''(\xi)$, and ξ and its derivatives are taken at the point (W, Z, β_0).

The asymptotic covariance matrix of $\hat{\beta}_N$ has a sandwich structure, which is similar to (34). Introduce the following two symmetric matrices

$$A_* := -\mathrm{E}\psi_{\beta^t}(\beta_*), \qquad B_* := \mathrm{E}\psi(\beta_*)\psi^t(\beta_*), \tag{38}$$

where $\psi(\beta_*)$ is short for $\psi(Y, W, Z, \beta_*)$. The symmetry of A_* follows directly from the definition (22) of ψ. According to the theory of estimating equations, under the conditions of Theorem 6.1 for small enough σ_u^2

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

where

$$\Sigma_N = A_*^{-1} B_* A_*^{-1}. \tag{39}$$

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

$$\begin{aligned}(\Sigma_{SQS} - \Sigma_N) &= \frac{1}{2}\sigma_u^2 \varphi S^{-1} \mathrm{E}\{C''[(\xi_\beta \xi_\beta^t)_{xx} + 2\xi_{x\beta}\xi_{x\beta}^t] \\ &+ 2C'''[2(\xi_x \xi_\beta \xi_\beta^t)_x - \xi_{xx}\xi_\beta\xi_\beta^t + (\xi_\beta^t \Delta\beta_*)\xi_\beta\xi_\beta^t] \\ &+ 2C^{(4)}\xi_x^2 \xi_\beta \xi_\beta^t\} S^{-1} + O(\sigma_u^4), \end{aligned} \tag{40}$$

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 (37) and S in (19).

7 Special cases

7.1 Linear-in-β regression model

Consider the following structural errors-in-variables model

$$Y_i = \sum_{j=0}^{k} \beta_j h_j(X_i, Z_i) + \varepsilon_i \tag{41}$$

$$W_i = X_i + U_i, \qquad i = 1, \dots, n. \tag{42}$$

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 and variances σ_ε^2 and σ_u^2 and covariance $\sigma_{\varepsilon u} = 0$. This is a particular case of the model of Section 2, 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

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

with $const > 0$ and fixed $A > 0$, $j = 0, \dots, k$. In this case the matrices (18) and (19) are calculated as

$$S_0 = \mathrm{E}h(X,Z)h^t(X,Z), \qquad S = \mathrm{E}h(W,Z)h^t(W,Z), \tag{43}$$

with $h := (h_0, \dots, h_k)^t$. We require S_0 to be nonsingular and assume (i) and (iii). Then Theorem 5.1 and Theorem 6.2 are applicable, and as $\sigma_u^2 \to 0$ we obtain

$$\begin{aligned} \Sigma_{SQS} &= \Sigma_{CS} + O(\sigma_u^4), \\ \Sigma_{CS} - \Sigma_N &= \frac{1}{2}\sigma_u^2\sigma_\varepsilon^2 S^{-1}\mathrm{E}\left\{(hh^t)_{xx} + 2h_x h_x^t\right\} S^{-1} + O(\sigma_u^4), \end{aligned} \tag{44}$$

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

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

$$a_{jj} := [S(\Sigma_{CS} - \Sigma_N)S]_{jj}, \qquad j = 0, \dots, k. \tag{45}$$

From (44) we conclude that, as $\sigma_u^2 \to 0$,

$$a_{jj} = \frac{1}{2}\sigma_u^2\sigma_\varepsilon^2 \mathrm{E}\left[\frac{\partial^2(h_j^2)}{\partial x^2} + 2\left(\frac{\partial h_j}{\partial x}\right)^2\right] + O(\sigma_u^4) \tag{46}$$

with the derivative of h_j taken at (W, Z). The following statement is an easy consequence of (46).

Proposition 7.1. 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 for small enough σ_u^2 it holds that $a_{jj} > 0$.

This 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 (46), it is interesting to give a geometric interpretation of the inequality $\partial^2(h_j^2)/\partial x^2 + 2(\partial h_j/\partial x)^2 \geq 0$. An easy calculation leads to

Proposition 7.2. Let $g \in C^2(I\!R)$. The inequality $(g^2)_{xx} + 2(g_x)^2 \geq 0$ holds for all $x \in I\!R$ iff the function $|g|^3$ is convex on $I\!R$. Moreover, if $|g|^3$ is strictly convex on some interval (α, β) then

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

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

Thus if h_j does not depend on Z and satisfies the conditons of Proposition 7.2, then $a_{jj} > 0$, for small enough σ_u^2.

7.2 Polynomial regression

The polynomial regression is a particular case of the model (41), (42), with

$$h_j = h_j(X) = X^j, \qquad j = 0, 1, \ldots, k. \tag{47}$$

It 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 (2001a), and $\hat{\beta}_N$ is the ordinary least squares estimator ($\hat{\beta}_{OLS}$). From (46), we get the following result.

Proposition 7.3. In the polynomial model (41), (42), (47) we have for a_{jj} of (45) as $\sigma_u^2 \to 0$:

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

Thus for small enough σ_u^2, $a_{jj} > 0$, $j = 1, \ldots, k$, in accordance with Proposition 7.1 or Proposition 7.2. Now, for $k \geq 2$ we consider the leading term of the matrix $S(\Sigma_{CS} - \Sigma_N)S$, see (44):

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

Due to Proposition 7.3, $M_{00} = 0$ 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

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

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 a matrix sense.

We mention that, for the polynomial model, (44) 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 7.3, $\mathrm{E}(W^{2j-2})$ may be replaced with $\mathrm{E}(X^{2j-2})$. A corresponding remark applies to the matrix M, see Kukush et al (2001a).

7.3 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 $Y|\xi$ 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 $(Y_i, W_i), i = 1, \ldots, n$. This is a particular case of the model of Section 2, with $C(\xi) = e^\xi$ and known $\varphi = 1$. A measure m which dominates the distribution of Y is the counting measure.

We suppose that Θ_β is a convex compact set in $I\!R^2$, and the true value of β is an interior point of Θ_β. The statements of Theorem 4.3, Theorem 4.4, Theorem 5.1, Theorem 6.1, and Theorem 6.2 hold true for the Poisson regression. Note that $\hat{\beta}_{SQS}$ is defined by (11) only, (12) is not needed; for details see Thamerus (1998) and Kukush et al (2001b).

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

$$g := \mu_w + \sigma_w^2 \beta_1 . \tag{48}$$

Theorem 7.1. 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

$$\beta_* = \beta + \frac{1}{2}\sigma_u^2 \Delta\beta_* + O(\sigma_u^4), \tag{49}$$

$$\Delta\beta_* = \frac{\beta_1}{\sigma_w^2}(\mu_w + g, -2)^t . \tag{50}$$

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

c) $$\Sigma_{CS} - \Sigma_N = \frac{2\sigma_u^2}{\sigma_w^4} \exp\{-(\beta_0 + \beta_1\mu_w + \tfrac{1}{2}\beta_1^2\sigma_w^2)\} \begin{pmatrix} g^2 & -g \\ -g & 1 \end{pmatrix} + O(\sigma_u^4) \tag{51}$$

In contrast to the expansion (44) in the polynomial case, we see that the term of order σ_u^2 in the expansion (51) is a positive semidefinite matrix. Moreover, we conclude from (51) that for small enough σ_u^2, the asymptotic variance of $\hat{\beta}_{1,N}$ is less than the asymptotic variance of $\hat{\beta}_{1,CS}$, and if $g \neq 0$ the same holds true for the estimators of β_0.

We mention that (51) also holds true if everywhere W is replaced with X, see Kukush et al (2001b).

8 Asymptotics when both errors are small

Hereafter we consider again the general model presented in Section 2. 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) $\chi^2 := \varphi/\sigma_u^2 = const.$

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

Let

$$v_0(W, Z, \beta_0) := \chi^2 + C''\xi_x^2, \tag{52}$$

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 8.1. Assume (xiv) and let $\sigma_u^2 \to 0$.

a) Under the conditions of Theorem 4.4,

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

b) Under the conditions of Theorem 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 6.2,

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

Remark 8.1.: In Theorem 8.1, 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 4.3 and the remark at the end of 7.2. The next result compares Σ_{SQS} to Σ_{CS} for both errors small.

Theorem 8.2. Let the conditions of Theorem 5.1 hold. Assume additionally that $L(C''\xi_x^2)$ with arguments (W, Z, β_0) has no atoms. Then:

a) under (xiv), the difference

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

ist positive semidefinite and does not equal zero.

b) under (xv),

$$\liminf_{\sigma_u^2 \to 0} [\sigma_u^{-2} \lambda_{max}(\Sigma_{CS} - \Sigma_{SQS})] > 0 \tag{53}$$

and

$$\liminf_{\sigma_u^2 \to 0} [\sigma_u^{-2} \lambda_{min}(\Sigma_{CS} - \Sigma_{SQS})] \ge 0, \tag{54}$$

Theorem 8.2 states that, for small errors, when φ and σ_u^2 have the same order and under normality assumptions, the SQS estimator is asymptotically more efficient than the CS estimator. Note that part a) can be formulated in a similar way as part b) by writing (53) and (54) with "lim" in place of "lim inf".

Remark 8.2. In the polynomial model of Subsection 7.2. with degree $k \geq 2$, and $\beta_0 \neq 0$, $L(C''\xi_x^2)$ has no atoms, and both the results of Theorem 8.1 and Theorem 8.2 hold true. Thus in a polynomial model of degree $k \geq 2$ in the case of small σ_ε^2 and σ_u^2, the SQS estimator is asymptotically more efficient than the CS estimator.

9 Simulation results

In order to check the theoretical results we performed some Monte Carlo simulation studies with two special models: a quadratic regression model and a Poisson regression model, see Sections 7.2 and 7.3. 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 $\sigma_u^2 = 0.05$.

In both models the sample size was taken to be $n = 800$ so that one could hope that the asymptotic theory would apply. The number of replications was $N = 1000$.

In the quadratic model the CS estimator was modified to an estimator (MCS) which has the same asymptotic properties as CS but is much stabler for small n, see Cheng et al (2000), where MCS was called MALS.

We computed bias and standard error of the three estimators N, CS (or rather MCS), and SQS from the 1000 replications and compared them to the theoretical approximation values as computed from (37), (21), and (40), respectively. The resulting values are presented in the following tables 1 to 3 under the heading "theory".

Table 1: *Standard deviations and differences of covariance matrices for a polynomial model in case (a):* $\beta = (0, 1, -0.5)', \mu_x = 0, \sigma_x^2 = 1, \sigma_\epsilon^2 = 20, \sigma_u^2 = 0.05$

$n = 800$	Standard Deviations				
	Naive theory	Naive simulation	MCS simulation	SQS simulation	CS or SQS theory
β_0	0.1941	0.1911	0.1948	0.1944	0.1973
β_1	0.1549	0.1555	0.1636	0.1633	0.1628
β_2	0.1069	0.1050	0.1161	0.1158	0.1181

$$\frac{n}{\sigma_u^2}(\Sigma_{MCS} - \Sigma_N) = \begin{pmatrix} 22.598 & 0.240 & -30.746 \\ 0.240 & 41.505 & -0.348 \\ -30.746 & -0.348 & 39.329 \end{pmatrix}$$

$$\frac{n}{\sigma_u^2}(\Sigma_{MCS} - \Sigma_{SQS}) = \begin{pmatrix} 2.281 & -0.638 & -1.691 \\ -0.638 & 1.496 & 0.871 \\ -1.691 & 0.871 & 1.106 \end{pmatrix}$$

Table 2: *Standard deviations and differences of covariance matrices for a polynomial model in case (b):* $\beta = (0, 1, -0.5)', \mu_x = 0, \sigma_x^2 = 1, \sigma_\epsilon^2 = 0.002, \sigma_u^2 = 0.010$

$n = 800$	Standard Deviations			
	Naive or CS theory	Naive simulation	MCS simulation	SQS simulation
β_0	0.0064	0.0063	0.0064	0.0048
β_1	0.0072	0.0072	0.0073	0.0059
β_2	0.0062	0.0062	0.0064	0.0044

$$\frac{n}{\sigma_u^2}(\Sigma_{MCS} - \Sigma_N) = \begin{pmatrix} 0.059 & 0.000 & -0.073 \\ 0.000 & 0.128 & -0.045 \\ -0.073 & -0.045 & 0.128 \end{pmatrix}$$

$$\frac{n}{\sigma_u^2}(\Sigma_{MCS} - \Sigma_{SQS}) = \begin{pmatrix} 1.464 & 0.827 & -1.569 \\ 0.827 & 1.481 & -0.875 \\ -1.569 & -0.875 & 1.686 \end{pmatrix}$$

Table 3: *Standard deviations and differences of covariance matrices for the Poisson model:* $\beta = (-1, 0.5)', \mu_x = 2, \sigma_x^2 = 1, \sigma_u^2 = 0.05$

$n = 800$	Standard Deviations				
	Naive theory	Naive simulation	CS simulation	SQS simulation	CS or SQS theory
β_0	0.0883	0.0888	0.0928	0.0929	0.0918
β_1	0.0328	0.0334	0.0351	0.0351	0.0343

$$\frac{n}{\sigma_u^2}(\Sigma_{CS} - \Sigma_N) = \begin{pmatrix} 11.6419 & -4.6875 \\ -4.6875 & 1.8869 \end{pmatrix}$$

$$\frac{n}{\sigma_u^2}(\Sigma_{CS} - \Sigma_{SQS}) = \begin{pmatrix} -0.3536 & -0.0161 \\ -0.0161 & 0.0201 \end{pmatrix}$$

It is seen that our theoretical results are nicely corroborated by our simulation study.

10 Conclusion

We studied the relative asymptotic efficiency of two consistent estimators of the parameters of a nonlinear regression model with Gaussian measurement errors in one of the covariates. The error variance σ_u^2 is supposed to be known, the response variable has a density belonging to an exponential family, and the error-ridden covariate has a Gaussian distribution. We are thus faced with the so-called structural variant of a measurement error model.

For this variant a structural quasi score (SQS) estimator can be constructed. The SQS estimator is consistent if the assumption of a Gaussian covariate holds true, otherwise it is biased. On the other hand, the corrected score (CS) estimator does not depend on any distributional assumptions for the covariates and is consistent whatever this distribution looks like. It thus belongs to the so-called functional variant of the model and is more robust than the SQS estimator with regard to the shape of the covariate distribution.

However, if the normality assumption does, in fact, hold true, the SQS estimator, which utilizes this extra information, might be thought to be more efficient than the CS estimator, which does not use this information.

It turns out, however, that for small error variances this is not true. On the contrary, both estimators are approximately equally efficent for small σ_u^2, more precisely: their asymptotic covariance matrices coincide up to the order of σ_u^2.

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.

In deriving the asymptotic covariance matrix of the SQS estimator, we took account of the fact, that before setting up the estimating equations for the regression parameters, the parameters of the normal covariate distribution (i.e., mean and variance) have to be estimated. This fact implies that additional terms have to be incorporated into the formula for Σ_{SQS}, which would not appear if the parameters of the covariate distribution were known. In deriving these additional terms we used a general approach, which might be helpful also in other situations where the estimates of the parameters of interest depend on nuisance parameters to be estimated aforehand.

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 asymptotic covariance matrix 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 the Poisson regression models this can, in fact, be shown when σ_u^2 becomes small. When both φ and σ_u^2 become small, the naive estimator is as efficient (up to the order of σ_u^2) as the CS estimator, and the SQS estimator is more efficient, in a sense. These results are corroborated by the results of a small simulation study which was carried out for the polynomial and the Poisson regression model. More simulation results can be found in Kukush et al (2001a), (2001b).

Simulations performed by Schneeweiss and Nittner (2001) for a polynomial model show that, for large σ_u^2, SQS seems to be more efficient than CS. It is an open problem to prove this theoretically. For the Poisson model, however, it can be shown that SQS is more efficient than CS for whatever σ_u^2, Shklyar and Kukush (2002).

11 Proofs

All random variables appearing in the model of Section 2 are defined on a common probability space $(\Omega, \mathcal{F}, P_0)$ where P_0 is the law under the true parameter values $\beta_0, \varphi_0, \mu_{w0}^2, \sigma_{w0}^2$. The operator "E" always denotes expectation under P_0 and "a.s." is an abbreviation for "P_0 - almost surely".

11.1 Proof of Theorem 4.1, part a)

To simplify the following arguments we slightly modify the estimating equation (12) by replacing the term $\frac{1}{n-k}$ with $\frac{1}{n}$. The right hand side of the modified equation (12) differs from the original one by a term of order $\frac{1}{n}$ a.s. The estimator resulting from the modified equations (11), (12) therefore has the same asymptotic properties as the one coming from the original equations (11), (12).

The score function corresponding to the (modified) equations (11), (12) is then

$$G(Y, W, Z; \theta; \mu_w, \sigma_w^2) = \begin{pmatrix} (Y-m)v^{-1}\frac{\partial m}{\partial \beta} \\ (Y-m)^2 - A_1 - \varphi A_2 \end{pmatrix}, \tag{55}$$

where m, v, A_1, and A_2 are functions of W and Z as well as of the parameters $\theta = (\beta^t, \varphi)^t, \mu_w, \sigma_w^2$, which may differ from the true values $\theta_0, \mu_{w0}, \sigma_{w0}^2$. We shall sometimes use the abbreviation $\eta := (\theta^t, \mu_w, \sigma_w^2)^t$ with η_0 being its true value. Given an i.i.d sample (Y_i, W_i, Z_i), $i = 1, \ldots, n$, the estimating function is defined by

$$G_n(\theta; \mu_w, \sigma_w^2) = \frac{1}{n}\sum_{i=1}^{n} G(Y_i, W_i, Z_i; \theta; \mu_w, \sigma_w^2). \tag{56}$$

The (modified) estimating equations (11), (12) can then be written as

$$G_n(\theta; \hat{\mu}_w, \hat{\sigma}_w^2) = 0, \qquad \theta \in \Theta := \Theta_\beta \times [a_1, b_1]. \tag{57}$$

Fix finite intervals (μ_{w1}, μ_{w2}) and $(a_2, b_2), a_2 > 0$, that contain μ_{w0} and σ_{w0}^2, respectively. Now, we list some properties of the functions (56).

(a) Almost surely $G_n(\theta; \mu_w, \sigma_w^2) \to G_\infty(\theta; \mu_w, \sigma_w^2)$ uniformly in $\Theta_\eta := \Theta \times [\mu_{w1}, \mu_{w2}] \times [a_2, b_2]$, with
$G_\infty(\theta; \mu_w, \sigma_w^2) = \mathrm{E}G_n(\theta; \mu_w, \sigma_w^2) = \mathrm{E}G(Y, Z, W; \theta; \mu_w, \sigma_w^2)$.

This property is based on three facts. First, for any fixed argument $\eta \in \Theta_\eta$, $G_n(\eta) \to G_\infty(\eta)$ a.s. due to the strong LLN.

Second, the functions $G_n(\eta)$ are equicontinuous in $\eta \in \Theta_\eta$ a.s. For instance, for the first component $G_n^1(\eta)$, see (55), we have by the strong LLN a.s.

$$\sup_{\eta\in\Theta_\eta} \left\| \frac{\partial G_n^1}{\partial \eta^t} \right\| \leq \frac{1}{n} \sum_{i=1}^n \sup_{\eta\in\Theta_\eta} \left\| \frac{\partial G^1(Y_i, W_i, Z_i; \eta)}{\partial \eta^t} \right\|$$
$$\to \mathrm{E} \sup_{\eta\in\Theta_\eta} \left\| \frac{\partial G^1(Y, W, Z; \eta)}{\partial \eta^t} \right\| < \infty,$$

where $G^1(Y_i, W_i, Z_i; \eta) := (Y_i - m_i) v_i^{-1} \frac{\partial m_i}{\partial \beta}$, $m_i = m(W_i, Z_i, \beta)$, and $v_i = v(W_i, Z_i, \theta)$. (To derive this result we used the exponential bounds of the conditions of Theorem 4.1) It follows that $\sup_{n\geq 1} \sup_{\eta\in\Theta_\eta} \| \frac{\partial G_n^1}{\partial \eta^t} \| < \infty$ a.s., and therefore the functions $G_n^1(\eta)$ are equicontinuous on Θ_η a.s. Similar arguments can be applied to $G_n^2(\eta)$.

Third, $G_\infty(\eta)$ is continuous in $\eta \in \theta_\eta$ see (b) below.

These three facts imply the existence of $\Omega_0 \subset \Omega$ with $P(\Omega_0) = 1$ such that $\forall \omega \in \Omega_0, \forall \eta \in \Theta_\eta : G_n(\eta) \to G_\infty(\eta)$. Indeed, let Θ_η^* be a countable dense subset of Θ_η and for any $\eta_i \in \Theta_\eta^*$ let $\Omega_{0i} \subset \Omega$ with $P(\Omega_{0i}) = 1$ be such that $\forall \omega \in \Omega_{0i} : G_n(\eta_i) \to G_\infty(\eta)$. Then take $\Omega_0 = \cap_{i=1}^\infty \Omega_{0i}$. For any $\eta \in \Theta_\eta$ and $\omega \in \Omega_0$ we then have $|G_\infty(\eta) - G_n(\eta)| \leq |G_\infty(\eta) - G_\infty(\eta_i)| + |G_\infty(\eta_i) - G_n(\eta_i)| + |G_n(\eta_i) - G_n(\eta)|$, which becomes arbitrary small if η_i us chosen close to η and n is sufficiently large.

Again using these three facts and the further fact that Θ_η is compact, we can now prove that $G_n(\eta) \to G_\infty(\eta)$ uniformly on Θ_η for all $\omega \in \Omega_0$. Indeed for any $\eta_0 \in \Theta_\eta$ and all η in a δ-neighborhood of η_0 the difference $|G_n(\eta) - G_\infty(\eta)|$ can be made less than any $\varepsilon > 0$ if $\delta = \delta(\varepsilon, \eta_0)$ is chosen sufficiently small and $n > N(\varepsilon, \eta_0)$. As a finite set of such δ-neighborhoods covers Θ_η, an $N = N(\varepsilon)$ can be chosen such that $\forall n > N, \forall \eta \in \Theta_\eta : |G_n(\eta) - G_\infty(\eta)| < \varepsilon$.

(b) Almost surely $G_n(\theta; \hat{\mu}_w, \hat{\sigma}_w^2) \to G_\infty(\theta; \mu_{w,0}\sigma_{w0}^2)$ uniformly in Θ.

This property follows from property (a) and from the fact that $\hat{\mu}_w$ and $\hat{\sigma}_w^2$ are strongly consistent.

(c) $G_\infty(\eta) = (G^1_\infty(\eta)^t, G^2_\infty(\eta))^t$, with

$$G^1_\infty(\eta) \;=\; \mathrm{E}\left[(m_0 - m)v^{-1}\frac{\partial m}{\partial \beta}\right], \tag{58}$$

$$G^2_\infty(\eta) \;=\; \mathrm{E}[(m_0 - m)^2 + A_{10} + \varphi_0 A_{20} - A_1 - \varphi A_2], \tag{59}$$

where $m_0 = m(W, Z, \beta_0), A_{10} = A_1(W, Z, \beta_0), A_{20} = A_2(W, Z, \beta_0)$ with $\mu_w = \mu_{w0}$ and $\sigma_w^2 = \sigma_{w0}^2$, and m, v, A_1, A_2 are as before.

(d) The matrix

$$I_0 := \left.\frac{\partial G_\infty(\eta)}{\partial \theta^t}\right|_{\eta=\eta_0} \tag{60}$$

is non-singular.

This property follows from the relations

$$\left.\frac{\partial G^1_\infty(\eta)}{\partial \beta^t}\right|_{\eta=\eta_0} = \left.-\mathrm{E}\left(v^{-1}\frac{\partial m}{\partial \beta}\frac{\partial m}{\partial \beta^t}\right)\right|_{\eta=\eta_0},$$

$$\frac{\partial G^1_\infty(\eta)}{\partial \varphi} = 0,$$

$$\left.\frac{\partial G^2_\infty(\eta)}{\partial \varphi}\right|_{\eta=\eta_0} = \left.-\mathrm{E}A_2\right|_{\eta=\eta_0},$$

where the matrix of the first relation is negative definite due to (viii), and $-\mathrm{E}A_2 < 0$ because $C''(\xi) > 0$, see (8); therefore $\det I_0 \neq 0$.

Next, we want to apply Theorem 12.1 from Heyde (1997) to the sequence (56) of estimating functions, see also Aitchison and Silvey (1958). Set

$$H_n(\theta) := -I_0^{-1} G_n(\theta; \hat{\mu}_w, \hat{\sigma}_w^2), \qquad \theta \in \Theta. \tag{61}$$

The functions $H_n(\theta)$ are continuous in θ. We have to show that for all small $\delta > 0$ a.e. on the set Ω

$$q_\delta := \limsup_{n\to\infty}\{\sup_{\|\theta-\theta_0\|=\delta} (\theta - \theta_0)^t H_n(\theta)\} < 0. \tag{62}$$

Indeed, due to property (b) we have

$$q_\delta = \sup_{\|\theta-\theta_0\|=\delta} (\theta-\theta_0)^t\{-I_0^{-1}G_\infty(\theta;\mu_{w0},\sigma_{w0}^2)\}. \tag{63}$$

Now, $G_\infty(\theta_0;\mu_{w0},\sigma_{w0}^2)=0$, which is easily seen from (58), (59). Therefore, using the definition (60) of I_0, we get the expansion

$$(\theta-\theta_0)^t\{-I_0^{-1}G_\infty(\theta;\mu_{w0},\sigma_{w0}^2)\} = -\|\theta-\theta_0\|^2 + o(\|\theta-\theta_0\|^2)$$

as $\theta \to \theta_0$. From (63) we obtain that, for all small $\delta>0$, the inequality (62) holds. Thus, by the above mentioned theorem from Heyde (1997), the equation $H_n(\theta)=0$ has a solution eventually. This proves statement a) of Theorem 4.1.

Remark 11.1. Now we can give a more rigorous definition of the SQS estimator. For those (small) n for which (11), (12) has no solution we set $\hat\beta_{SQS}=\beta_f$, $\hat\varphi_{SQS}=\varphi_f$, where $\beta_f \in \Theta_\beta$ and $\varphi_f \in [a_1,b_1]$ are arbitrary but fixed values. If n is such that (11), (12) has many solutions we choose one of them for every $\omega\in\Omega$ in such a way that $\hat\beta_{SQS}(\omega)$ and $\hat\varphi_{SQS}(\omega)$ are measurable. This is possible due to, e.g., Pfanzagl (1969).

11.2 Proof of Theorem 4.1, part b)

Owing to property (b) in the proof of part a), there is a set Ω_0 of probability 1 where $G_n(\theta;\hat\mu_w,\hat\sigma_w^2) \to G_\infty(\theta;\mu_{w0},\sigma_{w0}^2)$ uniformly in Θ. Fix $\omega\in\Omega_0$. The sequence $\hat\theta_n(w)$ of SQS-estimators lies in the compact set Θ. Consider an arbitrary convergent subsequence $\hat\theta_{n(k)}(\omega)\to\theta_*$. The sequence $G_n(\hat\theta_{n(k)};\hat\mu_w,\hat\sigma_w^2)$ converges to $G_\infty(\theta_*,\mu_{w0},\sigma_{w0}^2)$, which is zero because $G_n(\hat\theta_{n(k)};\hat\mu_w,\hat\sigma_w^2)=0$ eventually. Hence $\theta_*=\theta_0$ because obviously θ_o is the unique solution to $G_\infty(\theta;\mu_{w0},\sigma_{w0}^2)=0$, see (58), (59) and assumption (vii). This implies the convergence of the whole sequence $\hat\theta_n(\omega)$ to the true value θ_0, and strong consistency is proved.

11.3 Proof of Theorem 4.1, part c)

We apply an approach due to Foutz (1977) based on the Inverse Function Theorem. Due to property (b) from the proof of part a), the functions H_n of (61) converge a.s. uniformly in Θ to $H_\infty(\theta)=-I_0^{-1}G_\infty(\theta;\mu_{w0},\sigma_{w0}^2)$, and $H_\infty(\theta_0)=0$. Moreover, by similar arguments, $\frac{\partial H_n(\theta)}{\partial\theta^t}$ converges a.s.

uniformly in Θ to $\frac{\partial H_\infty(\theta)}{\partial \theta^t}$, and, because of (60), $\left.\frac{\partial H_\infty(\theta)}{\partial \theta^t}\right|_{\theta=\theta_0} = -I_{k+1}$, where I_{k+1} is the $(k+1) \times (k+1)$ unit matrix. Now, we fix a sequence $\hat{\theta}_n$ of SQS estimators such that

$$H_n(\hat{\theta}_n) = 0 \qquad \text{eventually.} \tag{64}$$

Such a sequence exists according to part a), and according to part b)

$$\hat{\theta}_n \to \theta_0 \qquad \text{a.s.} \tag{65}$$

Applying the arguments from Foutz (1977) we obtain the following result: if $\{\tilde{\theta}_n\}$ is another sequence satisfying (64), (65), then $\tilde{\theta}_n = \hat{\theta}_n$ eventually. This proves part c).

Remark 11.2. The approach of Foutz gives not only uniqueness, but also existence of the estimator with properties (64), (65). Therefore, in Subsection 11.1, we could have referred to Foutz instead of to Heyde. But our version of the proof shows that the stonger the convergence properties of $H_n(\theta)$ are, the better are the properties of the resulting estimators: just uniform convergence of $H_n(\theta)$ implies only existence, while additionally uniform convergence of the derivatives $\frac{\partial H_n(\theta)}{\partial \theta^t}$ implies uniqueness of the solution.

11.4 Proof of Theorem 4.2, part a)

In a similar way as in Subsection 11.1 it can be shown that the functions S_n of (13) converge a.s. uniformly in $\Theta_\beta \times \Theta_\beta \times [a_1, b_1]$ to

$$S_\infty(\beta, \alpha, \varphi) := \mathrm{E}\left[v^{-1}(W, Z, \alpha, \varphi)\{m(W, Z, \beta_0) - m(W, Z, \beta)\}\frac{\partial m(W, Z, \beta)}{\partial \beta}\right].$$

Moreover, $\frac{\partial S_n}{\partial \beta^t} \to \frac{\partial S_\infty}{\partial \beta^t}$ in the same sense, and

$$\left.\frac{\partial S_\infty(\beta, \alpha, \varphi)}{\partial \beta^t}\right|_{\beta=\beta_0} = -\mathrm{E}\left[v^{-1}(W, Z, \alpha, \varphi)\frac{\partial m(W, Z, \beta)}{\partial \beta}\frac{\partial m(W, Z, \beta)}{\partial \beta^t}\right]\Bigg|_{\beta=\beta_0},$$

which is negative definite due to (viii). We then find, as in Subsection 11.1, that eventually there exists a solution $\hat{\beta}_n(\alpha, \varphi)$ to

$$S_n(\beta, \alpha, \varphi) = 0, \qquad \beta \in \Theta_\beta, \tag{66}$$

for all $(\alpha,\varphi)\in\Theta$. But owing to (vii)' the limit equation $S_\infty(\beta,\alpha,\varphi)=0$ has the unique solution $\beta=\beta_0$. Therefore we see, like in Subsection 11.2, that the solution to (66) has the property that uniformly in $(\alpha,\varphi)\in\Theta$, as $n\to\infty$,

$$\hat\beta_n(\alpha,\varphi)\to\beta_0 \qquad \text{a.s.} \tag{67}$$

Moreover, applying again the arguments of Foutz (1977) as in Subsection 11.3, this solution is seen to be unique eventually.

11.5 Proof of Theorem 4.2, part b)

According to part a) of Theorem 4.2 there exists $\Omega_0\subset\Omega$ with $P(\Omega_0)=1$ such that, for all $\omega\in\Omega_0$ and $n\ge N(\omega)$, $\hat\beta_n(\alpha,\varphi)$ satisfies (66). Let us fix $\omega\in\Omega_0$ and $n\ge N(\omega)$, and let us drop the index n for notational ease. Denote the function $\hat\beta_n(\alpha,\beta)$ by $s(\alpha,\beta)$. Then

$$\beta^{(j+1)}=s(\beta^{(j)},\varphi^{(j)}),$$

and because of (15) we get

$$\beta^{(j+1)}=s\{\beta^{(j)},P\circ h(\beta^{(j)})\}=:F(\beta^{(j)}).$$

We show that, for large n, $F(\cdot)$ is a contraction on Θ_β, i.e., $F:\Theta_\beta\to\Theta_\beta$ and $\|F(\beta_1)-F(\beta_2)\|\le\lambda\|\beta_1-\beta_2\|$, where $\lambda<1$ and λ does not depend on β_1,β_2. We have the identity

$$S\{F(\beta),\beta,P\circ h(\beta)\}=0,$$

where $S=S_n$ from (66). Let $\beta_1\ne\beta_2$ and $l:=F(\beta_1)-F(\beta_2)\ne 0$. Then

$$l^t\frac{\partial S^*}{\partial\beta^t}l+l^t\frac{\partial S^*}{\partial\alpha^t}(\beta_1-\beta_2)+l^t\frac{\partial S^*}{\partial\varphi}(P\circ h(\beta_1)-P\circ h(\beta_2))=0, \tag{68}$$

where the asterisk indicates that the derivatives are taken at an intermediate point between the points $(F(\beta_i),\beta_i,P\circ h(\beta_i)),i=1,2$. Denote by C the convex hull of $F(\Theta_\beta)$. From (68) we get with $\vartheta=(\alpha,\varphi)$,

$$\begin{aligned}||l||\le||\beta_1-\beta_2||&\left\{\inf_{\beta\in C,\vartheta\in\Theta}\lambda_{\min}\left(\frac{\partial S(\beta,\vartheta)}{\partial\beta^t}\right)\right\}^{-1}\\ \times\left\{\sup_{\beta\in C,\vartheta\in\Theta}\left|\left|\frac{\partial S(\beta,\vartheta)}{\partial\alpha^t}\right|\right|\right.&+\left.\sup_{\beta\in C,\vartheta\in\Theta}\left|\left|\frac{\partial S(\beta,\vartheta)}{\partial\varphi}\right|\right|\sup_{\beta\in\Theta_\beta}\left|\left|\frac{\partial h}{\partial\beta^t}\right|\right|\right\}.\end{aligned} \tag{69}$$

Here $\lambda_{\min}(\cdot)$ is the minimal eigenvalue of a symmetric matrix. Let us consider the various terms of (69). First, due to uniform convergence (67), $C \subset B(\beta_0, \delta)$ for $n \geq n_0(\delta, \omega)$, where $B(\beta_0, \delta)$ is a ball with center β_0 and radius δ. Second, the derivatives of $S(\beta, \vartheta)$ converge uniformly to the derivatives of $S_\infty(\beta, \vartheta)$, and $\partial S_\infty(\beta_0, \vartheta)/\partial \beta^t$ is positive definite, see Subsection 11.4. Therefore, for large n and small enough δ, $\partial S(\beta, \vartheta)/\partial \beta^t$ is positive definite on $B(\beta_0, \delta)$ and hence

$$\inf_{\beta \in C, \vartheta \in \Theta} \lambda_{\min} \left(-\frac{\partial S(\beta, \vartheta)}{\partial \beta^t} \right)$$

is positive and bounded away from 0. Third, the derivatives of S_∞ at the point $(\beta_0, \alpha, \varphi)$ with respect to φ and α equal 0, and so $\sup ||\partial S/\partial \alpha^t||$ and $\sup ||\partial S/\partial \varphi||$ in (69) tend to zero with $n \to \infty$ and $\delta \to 0$. Finally $\sup_{\beta \in \Theta_\beta} ||\frac{\partial h}{\partial \beta^t}|| = O(1)$. Taking all these results together, (69) now implies:

$$||F(\beta_1) - F(\beta_2)|| \leq \lambda_n \cdot ||\beta_1 - \beta_2||,$$

where $\lambda_n \to 0, n \to \infty$. Thus, for large n, F is a contraction. Then by Banach's contraction theorem, for sufficiently large n, $\beta^{(j)}$ converges to a fixed point β_F of the mapping F as $j \to \infty$, and $\varphi^{(j)} = P \circ h(\beta^{(j)}) \to \varphi_F = P \circ h(\beta_F)$ as $j \to \infty$. The limit values β_F, φ_F satisfy (14), (15). But $\hat{\beta}_{SQS}, \hat{\varphi}_{SQS}$ also satisfy (14), (15) eventually. As $\hat{\varphi}_{SQS} \to \varphi_0$ a.s. and $\varphi_0 \in (a_1, b_1)$, therefore $\hat{\varphi}_{SQS} \in (a_1, b_1)$ eventually, and then $P \circ h(\hat{\beta}_{SQS}) = h(\hat{\beta}_{SQS})$. Since the solution to (14), (15) is unique eventually, therefore $\beta_F = \hat{\beta}_{SQS}$ and $\varphi_F = \hat{\varphi}_{SQS}$ eventually. This completes the proof.

11.6 Proof of Theorem 4.3

We reparameterize the score function G of (55) and the estimating function G_n of (56) by replacing the nuisance parameters (μ_w, σ_w^2) in the arguments with $\gamma = (\gamma_1, \gamma_2)^t$ without changing the notation of G or G_n, i.e., we write, e.g., in (56), $G_n(\theta, \gamma)$ instead of $G_n(\theta, \mu_w, \sigma_w^2)$ and similarly for G in (55). The estimators $\hat{\beta}_{SQS}$ and $\hat{\varphi}_{SQS}$ satisfy the equation

$$G_n(\hat{\beta}_{SQS}, \hat{\varphi}_{SQS}; \hat{\mu}_w, \hat{\sigma}_w^{-2}) = 0. \tag{70}$$

We want to derive the asymptotic covariance matrix of $\hat{\beta}_{SQS}$ with the help of Lemma A1 of the Appendix. Let us check the conditions of Lemma

A1. By Theorem 4.1, $\hat{\theta} = (\hat{\beta}^t_{SQS}, \hat{\varphi}_{SQS})^t$ is consistent. The random field $G_n(\theta,\gamma)$, $\theta \in \Theta$, $\gamma \in \Theta_\gamma := [\mu_{w_1}, \mu_{w_2}] \times \left[\frac{1}{b_2}, \frac{1}{a_2}\right]$, has C^1-smooth paths a.s. So conditions a) and b) are satisfied.

Consider condition c) of Lemma A1. Set $H_i := W_i - \mu_w$. We have:

$$\hat{\mu}_w = \frac{1}{n}\sum_{i=1}^n W_i = \overline{W}, \tag{71}$$

$$\hat{\sigma}^2_w = \frac{1}{n-1}\sum_{i=1}^n (H_i - \overline{H})^2 = \overline{H^2} + O_p\left(\frac{1}{n}\right), \tag{72}$$

and

$$\hat{\sigma}^{-2}_w - \sigma^{-2}_w = -\frac{\hat{\sigma}^2_w - \sigma^2_w}{\sigma^2_w \hat{\sigma}^2_w} = -\frac{1}{\sigma^4_w}(\overline{H^2} - \sigma^2_w) + O_p\left(\frac{1}{n}\right). \tag{73}$$

From (56), (71) and (73) we get

$$\begin{pmatrix} \sqrt{n} G_n(\theta_0,\gamma_0) \\ \sqrt{n}(\hat{\gamma}_n - \gamma_0) \end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} G(Y_i, W_i, Z_i; \theta_0, \gamma_0) \\ W_i - \mu_{w0} \\ -\frac{H_i^2 - \sigma^2_{w0}}{\sigma^4_{w0}} \end{pmatrix} + O_p\left(\frac{1}{\sqrt{n}}\right). \tag{74}$$

Note that

$$\mathrm{E}G(Y,W,Z;\theta_0,\gamma_0) = \mathrm{E}[\mathrm{E}\{G(Y,W,Z;\theta_0,\gamma_0)|W,Z\}] = 0,$$

see (55), (5), (6) and (8). From (74) we have by the CLT for i.i.d. random vectors:

$$(\sqrt{n} G^t_n(\theta_0,\gamma_0), \sqrt{n}(\hat{\gamma}_n - \gamma_0)^t)^t \xrightarrow{d} N(0,\Sigma)$$

with

$$\Sigma = \mathrm{cov}\begin{pmatrix} G(Y,W,Z;\theta_0,\gamma_0) \\ W - \mu_{w0} \\ -\frac{(W-\mu_{w0})^2 - \sigma^2_{w0}}{\sigma^4_{w0}} \end{pmatrix}. \tag{75}$$

Obviously

$$\Sigma = \mathrm{diag}(\Sigma_{11}, \sigma^2_{w0}, 2\sigma^{-4}_{w0}), \tag{76}$$

where

$$\Sigma_{11} = \begin{pmatrix} \Phi & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}, \tag{77}$$

Φ comes from (viii), σ_{12} and $\sigma_{21} = \sigma_{12}^t$ are unspecified vectors, and σ_{22} is a scalar.

Now pass to condition d) of Lemma A1. By the LLN

$$\mathrm{V}_1 = \mathrm{E}\frac{\partial G(Y, W, Z; \beta_0, \varphi_0, \gamma_0)}{\partial(\beta^t, \varphi)}.$$

Let us decompose G in accordance with (55): $G = (G_1^t, G_2)^t$. Then, with derivatives taken at the true parameter values,

$$\mathrm{E}\frac{\partial G_1}{\partial \beta^t} = -\Phi, \qquad \mathrm{E}\frac{\partial G_1}{\partial \varphi} = 0,$$
$$\mathrm{E}\frac{\partial G_2}{\partial \varphi} = -\mathrm{E}A_2 =: -\varphi_{22}.$$

Introduce also $\Phi_{12} := -\mathrm{E}\frac{\partial G_2}{\partial \beta^t}$. By condition (viii) Φ is positive definite, and by (8) $\varphi_{22} = \mathrm{E}C'' > 0$. Therefore V_1 is non-singular, and

$$\mathrm{V}_1^{-1} = -\begin{pmatrix} \Phi^{-1} & 0 \\ -\frac{1}{\varphi_{22}}\Phi_{12}\Phi^{-1} & \frac{1}{\varphi_{22}} \end{pmatrix}. \tag{78}$$

Let us now pass to condition e) of Lemma A1. We have

$$\mathrm{V}_2 = \mathrm{E}\frac{\partial G(Y, W, Z; \beta_0, \varphi_0, \gamma_0)}{\partial \gamma^t}.$$

In particular, again taking derivatives at the true parameter values,

$$\mathrm{E}\frac{\partial G_1}{\partial \gamma^t} = -\mathrm{E}\left(v^{-1}\frac{\partial m}{\partial \beta}\frac{\partial m}{\partial \gamma^t}\right) =: \mathrm{V}_{21}, \tag{79}$$

introduce also

$$\mathrm{V}_{22} := \mathrm{E}\frac{\partial G_2}{\partial \gamma^t}.$$

Then

$$\mathrm{V}_2 = \begin{pmatrix} \mathrm{V}_{21} \\ \mathrm{V}_{22} \end{pmatrix}. \tag{80}$$

Finally, condition f) of Lemma A1 can be shown to hold. For θ and γ in the ε-neighborhood of θ_0 and γ_0,

$$\left\|\frac{\partial G_n(\theta,\gamma)}{\partial(\theta^t,\gamma^t)} - \frac{\partial G_n(\theta_0,\gamma_0)}{\partial(\theta^t,\gamma^t)}\right\| \leq \sup_{\|\theta-\theta_0\|\leq\varepsilon, \|\gamma-\gamma_0\|\leq\varepsilon} \left\|\frac{\partial^2 G_n(\theta,\gamma)}{\partial(\theta,\gamma)\partial(\theta^t,\gamma^t)}\right\|\sqrt{2}\varepsilon,$$

and, because of the exponential bounds conditions,

$$\mathrm{E} \sup_{\|\theta-\theta_0\|\leq\varepsilon, \|\gamma-\gamma_0\|\leq\varepsilon} \|\frac{\partial^2 G(Y,W,Z,\theta,\gamma)}{\partial(\theta,\gamma)\partial(\theta^t,\gamma^t)}\| < \infty,$$

for sufficiently small ε. Now condition f) follows by applying Chebyshev's inequality in the form: $P(|X| \geq \delta) \leq \frac{\mathrm{E}|X|}{\delta}$.

So all the conditions of Lemma A1 are satisfied, and hence

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_{SQS} - \beta_0 \\ \hat{\varphi}_{SQS} - \varphi_0 \end{pmatrix} \xrightarrow{d} N(0, \Sigma_\theta),$$
$$\Sigma_\theta = \mathrm{V}_1^{-1}(I_{k+1}, \mathrm{V}_2)\Sigma(I_{k+1}, \mathrm{V}_2)^t \mathrm{V}_1^{-t}$$

Introduce the $k \times (k+1)$ selection matrix $P_\beta := (I_k, 0)$. According to (78) we have for the asymptotic covariance matrix Σ_{SQS} of $\hat{\beta}_{SQS}$:

$$\Sigma_{SQS} = P_\beta \Sigma_\theta P_\beta^t = (\Phi^{-1}, 0)(I_{k+1}, \mathrm{V}_2)\Sigma(I_{k+1}, \mathrm{V}_2)^t(\Phi^{-1}, 0)^t,$$

where 0 is $k \times 1$, and with (80), (76) and (77):

$$\begin{aligned} \Sigma_{SQS} &= (\Phi^{-1}, 0; \Phi^{-1}\mathrm{V}_{21})\mathrm{diag}(\Sigma_{11}, \sigma_{w0}^2, 2\sigma_{w0}^{-4})(\Phi^{-1}, 0; \Phi^{-1}\mathrm{V}_{21})^t \\ &= \Phi^{-1} + \Phi^{-1}\mathrm{V}_{21}\mathrm{diag}(\sigma_{w0}^2, 2\sigma_{w0}^{-4})\mathrm{V}_{21}^t\Phi^{-1}. \end{aligned}$$

Now by (79) $\mathrm{V}_{21} = -(F_1, F_2)$, where the $F_p, p = 1, 2$, are given in (16). We thus finally have

$$\Sigma_{SQS} = \Phi^{-1} + \Phi^{-1}(\sigma_{w0}^2 F_1 F_1^t + 2\sigma_{w0}^{-4} F_2 F_2^t)\Phi^{-1}.$$

11.7 Proof of Theorem 4.4

We divide the proof into several steps.
(a) *Expansion of the conditional mean.* We want to derive an expansion of $m(W, Z, \beta)$ in (7) in terms of powers of σ_u^2. We use the following representation of $m(w, z, \beta)$ with non-random w and z, remembering that $X|W \sim N\{\mu(w), \tau^2\}$:

$$m(w, z, \beta) = \mathrm{E}C'[\xi\{\mu(w) + \tau\zeta, z, \beta\}] \tag{81}$$

with $\zeta \sim N(0, 1)$. We start, however, with a slightly more general situation. Let $f \in C^{(4)}(I\!R)$, and for some $A > 0$

$$|f^{(i)}(w)| \le const \cdot e^{A|w|}, \quad w \in I\!R, \quad i = 0, 1, \ldots, 4. \tag{82}$$

Let $\zeta \sim N(0, 1)$ and consider the following expansion as $\sigma_u^2 \to 0$, see (9), (10):

$$\begin{aligned} \mathrm{E}f\{\mu(w) + \tau\zeta\} &= \mathrm{E}f\left(w + \tau\zeta - \frac{\sigma_u^2}{\sigma_w^2}(w - \mu_w)\right) = \\ &= \mathrm{E}\left[\sum_{i=0}^{3} \frac{f^{(i)}(w)}{i!}\left(\tau\zeta - \frac{\sigma_u^2}{\sigma_w^2}(w - \mu_w)\right)^i + r_3\right] = \\ &= f(w) - f'(w)\frac{\sigma_u^2}{\sigma_w^2}(w - \mu_w) + \frac{f''(w)}{2}\tau^2 + O(\sigma_u^4) + \mathrm{E}r_3. \end{aligned} \tag{83}$$

Here, with some $\lambda \in [0, 1]$,

$$r_3 = \frac{1}{4!}f^{(4)}\left\{w + \lambda\left(\tau\zeta - \frac{\sigma_u^2}{\sigma_w^2}(w - \mu_w)\right)\right\}\left(\tau\zeta - \frac{\sigma_u^2}{\sigma_w^2}(w - \mu_w)\right)^4.$$

Now, from (82) and (10) we get a bound for r_3 when $\sigma_u^2 \le 1$:

$$|r_3| \le const \cdot \sigma_u^4 e^{A|w|} e^{A|\zeta|}$$

with some $A > 0$, and hence

$$|\mathrm{E}r_3| \le const \cdot \sigma_u^4 e^{A|w|}.$$

In (83) the term $O(\sigma_u^4)$ can be bounded with a similar bound. Thus, again with the help of (10), we obtain

$$\mathrm{E}f\{\mu(w) + \tau\zeta\} = f(w) \quad + \quad \frac{1}{2}\sigma_u^2\left(-2f'(w)\frac{w - \mu_w}{\sigma_w^2} + f''(w)\right) \tag{84}$$

$$
\begin{aligned}
&+ \quad \sigma_u^4 \cdot rest, \\
|rest| \quad &\leq \quad const \cdot e^{A|w|}. \qquad (85)
\end{aligned}
$$

If $f = f(w,z), w \in I\!R, z \in \mathrm{E}_z, f(\cdot,z) \in C^4(I\!R)$, and

$$
|D_w^i f(w,z)| \leq const(e^{A|w|} + e^{A\|z\|}), \quad 0 \leq i \leq 4,
$$

then for the expectation $\mathrm{E}f\{\mu(w) + \tau\zeta, z\}$ the expansion (84), with suppressed dependence on z, still holds, where the derivatives are taken with respect to w and

$$
|rest| \leq const(e^{A|w|} + e^{A\|z\|}). \qquad (86)
$$

Now, we specialize to the function $f = C'\{\xi(w,z,\beta)\}$ and derive an expansion for $m(W,Z,\beta)$ in (81). We use the bounds in (ii) and (vi) and obtain for $\sigma_u^2 \leq 1$:

$$
\begin{aligned}
m(W,Z,\beta) = C' \quad &+ \quad \frac{1}{2}\sigma_u^2 \left(-2\frac{W-\mu_w}{\sigma_w^2}C''\xi_x + C''\xi_{xx} + C'''\xi_x^2\right) \qquad (87) \\
&+ \quad \sigma_u^4 \cdot rest, \\
|rest| \quad &\leq \quad const(e^{A|W|} + e^{A\|Z\|}). \qquad (88)
\end{aligned}
$$

The function ξ and its derivatives are taken at (W,Z,β).

(b) *Expansion of conditional variance.* First we expand A_1 defined in (8). We have with (84) specialized, respectively, to $f = C'^2$ and to $f = C'$

$$
\begin{aligned}
&A_1(W,Z,\beta) = \mathrm{E}\left[C'^2\{\xi(X,Z,\beta)\}|W,Z\right] - \{\mathrm{E}[C'\{\xi(X,Z,\beta)\}|W,Z]\}^2 \\
&= C'^2 + \frac{\sigma_u^2}{2}\left[-4\frac{(W-\mu_w)}{\sigma_w^2}C'C''\xi_x + 2(C''^2\xi_x^2 + C'C'''\xi_x^2 + C'C''\xi_{xx})\right] \\
&\quad - \left\{C' + \frac{\sigma_u^2}{2}\left[-2\frac{(W-\mu_w)}{\sigma_w^2}C''\xi_x + C''\xi_{xx} + C'''\xi_x^2\right]\right\}^2 + \sigma_u^4 \cdot rest,
\end{aligned}
$$

where *rest* satisfies (88). After simple transformations we get

$$
A_1(W,Z,\beta) = \frac{\sigma_u^2}{2}2C''^2\xi_x^2 + \sigma_u^4 \cdot rest.
$$

Here and in the sequel the function ξ and its derivatives are taken at (W,Z,β). Next, A_2, as defined in (8), is expanded in the same way. We receive the same expression as the right hand side of (87), but with $C^{(i+1)}$ in place of $C^{(i)}, i = 1, 2, 3$. Finally we get from (8)

$$\begin{aligned}\varphi^{-1} v(W, Z, \beta, \varphi) &= C'' \\ &+ \frac{\sigma_u^2}{2}\left[2\frac{C''^2 \xi_x^2}{\varphi} + C''' \xi_{xx} + C^{(4)} \xi_x^2 - 2\frac{(W-\mu_w)}{\sigma_w^2} C''' \xi_x\right] \\ &+ \sigma_u^4 \cdot rest, \end{aligned} \qquad (89)$$

where $rest$ satisfies (88).

(c) *Expansion of* $\frac{\partial m}{\partial \beta}(W, Z, \beta)$. We can differentiate both sides of (81) with respect to β and then use (84) with $f = C'' \xi_\beta$. We obtain

$$\frac{\partial m}{\partial \beta}(W, Z, \beta) = C'' \xi_\beta + \frac{1}{2}\sigma_u^2 T_1 + \sigma_u^4 \cdot rest, \qquad (90)$$

$$\begin{aligned} T_1 &:= -2\frac{(W-\mu_w)}{\sigma_w^2}(C'' \xi_x)_\beta + C'' \xi_{xx\beta} \\ &+ C'''(\xi_{xx}\xi_\beta + 2\xi_x \xi_{x\beta}) + C^{(4)} \xi_x^2 \xi_\beta, \end{aligned} \qquad (91)$$

and $rest$ satisfies (88).

(d) *Expansion of* $(v/\varphi)^{-1}$. We write (89) as

$$\varphi^{-1} v(W, Z, \beta) = C'' + \frac{\sigma_u^2}{2} T_2 + \sigma_u^4 \cdot rest. \qquad (92)$$

From condition (iv) and (8) we have

$$\frac{\varphi}{v} \leq \frac{1}{A_2}.$$

But by condition (iv) and the independence of Z and X

$$A_2 \geq const\, e^{-A\|Z\|} \mathrm{E}(e^{-A|X|}|W).$$

Now write $X = \mu(W) + \tau\zeta$, where $\zeta \sim N(0,1)$ and ζ is independent of W, see (9) and (10). Then

$$\begin{aligned} \mathrm{E}(e^{-A|X|}|W) &\geq e^{-A|\mu(W)|} \mathrm{E} e^{-A\tau|\zeta|} \\ &= const\, e^{-A\left|\frac{\sigma_x^2}{\sigma_w^2} W + \frac{\sigma_u^2}{\sigma_w^2}\mu_w\right|} \\ &\geq const\, e^{-A|W|}, \end{aligned}$$

where "$const$" depends on μ_w and σ_w^2. Thus

$$\varphi/v \le const \cdot e^{A|W|} e^{A\|Z\|} \le const(e^{2A|W|} + e^{2A\|Z\|}). \tag{93}$$

Now, according to (92) the leading terms of the expansion of φ/v will have the form $\frac{1}{C''}(1 - \frac{\sigma_u^2}{2C''}T_2)$. Therefore consider the difference

$$\begin{aligned} & \left| \frac{\varphi}{v} - \frac{1}{C''}\left(1 - \frac{\sigma_u^2}{2C''}T_2\right) \right| \\ = \; & \frac{\varphi}{v}\left| 1 - \left(C'' + \frac{\sigma_u^2}{2}T_2 + \sigma_u^4 \cdot rest\right)\frac{1}{C''}\left(1 - \frac{\sigma_u^2}{2C''}T_2\right) \right| \\ \le \; & \frac{\varphi}{v}\left[\frac{\sigma_u^4 T_2^2}{4(C'')^2} + \frac{\sigma_u^4 \cdot |rest| \cdot \left(1 + \frac{\sigma_u^2}{2C''}|T_2|\right)}{C''} \right]. \end{aligned} \tag{94}$$

Using (93) and (iv) and the fact that owing to (ii) and (vi) we have similar exponential bounds for T_2, we obtain from (94) the expansion

$$\varphi/v = \frac{1}{C''}\left(1 - \frac{\sigma_u^2}{2C''}T_2\right) + \sigma_u^4 \cdot rest, \tag{95}$$

and $rest$ satisfies (88).

(e) *Proof of (20).* From (81) we have

$$\frac{\partial m(w, z, \beta)}{\partial \gamma_p} = \mathrm{E}\left[C''(\xi)\xi_x\left(\frac{\partial \mu_\xi(w)}{\partial \gamma_p} + \frac{\partial \tau}{\partial \gamma_p}\zeta\right)\right], \tag{96}$$

where ξ and ξ_x are functions of $\mu(w) + \tau\zeta$, z, β. If $p = 1$, then $\gamma_1 = \mu_w$, and by (9) and (10)

$$\frac{\partial \mu(w)}{\partial \gamma_1} = \frac{\sigma_u^2}{\sigma_w^2}, \qquad \frac{\partial \tau}{\partial \gamma_1} = 0,$$

and if $p = 2$, then $\gamma_2 = \sigma_w^{-2}$, and

$$\frac{\partial \mu(w)}{\partial \gamma_2} = -\sigma_u^2(w - \mu_w), \qquad \frac{\partial \tau}{\partial \gamma_2} = \frac{1}{2\tau}\frac{\partial \tau^2}{\partial \gamma_2} = -\frac{\sigma_u^3}{2(1 - \sigma_u^2/\sigma_w^2)^{1/2}}.$$

Therefore from (96) we get that for $\sigma_u^2 \le 1$

$$\left|\frac{\partial m(W,Z,\beta)}{\partial \gamma_p}\right| \le \sigma_u^2 |rest|, \tag{97}$$

where $rest$ satisfies (88) because of (ii) and (vi). Now, from (16), (90) (93) and (97) we have

$$F_p = O(\sigma_u^2),$$

and (20) follows from (17).

(f) *Proof of (21).* In the sequel $rest$ is a quantity which always satisfies (88). From (90) we have

$$\frac{\partial m}{\partial \beta}\frac{\partial m}{\partial \beta^t} = C''^2 \xi_\beta \xi_\beta^t + \frac{1}{2}\sigma_u^2 C''[T_1 \xi_\beta^t]_S + \sigma_u^4 \cdot rest. \tag{98}$$

Hereafter $[\cdot]_S$ means symmetrization, i.e., for a square matrix M,

$$[M]_S := M + M^t.$$

Now, (98) and (95) imply that

$$\frac{\varphi}{v}\frac{\partial m}{\partial \beta}\frac{\partial m}{\partial \beta^t} = C''\xi_\beta \xi_\beta^t + \frac{1}{2}\sigma_u^2([T_1\xi_\beta^t]_S - \xi_\beta\xi_\beta^t T_2) + \sigma_u^4 \cdot rest.$$

Taking expectations at $\beta = \beta_0$, we derive the following expression for Φ, as defined right after (viii), with S from (19):

$$\varphi\Phi = S + \frac{1}{2}\sigma_u^2 \mathrm{E}([T_1\xi_\beta^t]_S - \xi_\beta\xi_\beta^t T_2) + O(\sigma_u^4),$$

and by inversion we get, according to (20),

$$\varphi^{-1}\Sigma_{SQS} = S^{-1} + \frac{1}{2}\sigma_u^2 S^{-1}\mathrm{E}(\xi_\beta\xi_\beta^t T_2 - [T_1\xi_\beta^t]_S)S^{-1} + O(\sigma_u^4). \tag{99}$$

To simplify (99), we use integration by part in the form

$$\mathrm{E}\left[\frac{W-\mu_w}{\sigma_w^2} h(W,Z,\beta)\right] = \mathrm{E}h_x(W,Z,\beta), \tag{100}$$

where $W \sim N(\mu_w, \sigma_w^2), h : I\!R \times \mathrm{E}_Z \times \Theta_\beta \to I\!R$ is measurable, $h(\cdot, Z, \beta) \in C^1(I\!R)$, and

$$|h(W, Z, \beta)| + |h_x(W, Z, \beta)| \le const(e^{A|W|} + e^{A\|Z\|}). \tag{101}$$

We have by (100) (with T_2 being the term in brackets of (89))

$$\begin{aligned} \mathrm{E}(\xi_\beta \xi_\beta^t T_2) &= \mathrm{E}\{2\varphi^{-1} C''^2 \xi_x^2 \xi_\beta \xi_\beta^t \\ &+ C'''(-2(\xi_x \xi_\beta \xi_\beta^t)_x + \xi_{xx} \xi_\beta \xi_\beta^t) - C^{(4)} \xi_x^2 \xi_\beta \xi_\beta^t\}. \end{aligned} \tag{102}$$

Next, with T_1 from (91),

$$\begin{aligned} [T_1 \xi_\beta^t]_S &= -\frac{2(W - \mu_w)}{\sigma_w^2}(2C''' \xi_x \xi_\beta \xi_\beta^t + C''[\xi_{x\beta} \xi_\beta^t]_S) + C''[\xi_{xx\beta} \xi_\beta^t]_S \\ &+ C'''(2\xi_{xx} \xi_\beta \xi_\beta^t + 2\xi_x [\xi_{x\beta} \xi_\beta^t]_S) + 2C^{(4)} \xi_x^2 \xi_\beta \xi_\beta^t, \end{aligned}$$

and applying (100) again we get

$$\begin{aligned} \mathrm{E}[T_1 \xi_\beta^t]_S &= \mathrm{E}\{C''(-4\xi_{x\beta} \xi_{x\beta}^t - [\xi_{xx\beta} \xi_\beta^t]_S) \\ &+ C'''[-4(\xi_x \xi_\beta \xi_\beta^t)_x + 2\xi_{xx} \xi_\beta \xi_\beta^t] - 2C^{(4)}(\xi_x^2 \xi_\beta \xi_\beta^t\}. \end{aligned} \tag{103}$$

From (99), (102) and (103) and letting $\beta = \beta_0$ we finally obtain (21).

11.8 Proof of Theorem 5.1

We divide the proof into several steps. In the sequel ξ and its derivatives are considered as functions of W, Z, β.

(a) *Expansion of* ψ_c. From (26) to (28) we get

$$\begin{aligned} \psi_c(y, w, z, \beta) &= y\left(\xi_\beta - \frac{1}{2}\sigma_u^2 \xi_{xx\beta}\right) - C' \xi_\beta \\ &+ \frac{1}{2}\sigma_u^2 (C' \xi_{xx\beta} + C'' \xi_{xx} \xi_\beta + 2C'' \xi_x \xi_{x\beta} + C''' \xi_x^2 \xi_\beta) \\ &+ \sigma_u^4 \cdot rest, \end{aligned} \tag{104}$$

where

$$\begin{aligned} \|rest\| &\le const(|y| + 1)(e^{A|w|} + e^{A\|z\|}) \\ &\le const \cdot 2(y^2 + e^{2A|w|} + e^{2A\|z\|}). \end{aligned} \tag{105}$$

As to (105), note that, for $a > 0$, $|y|e^a \leq \frac{1}{2}(y^2 + e^{2a})$ and $e^a < e^{2a}$. Now, from (104) and (87) we have

$$\begin{aligned}\psi_c(Y, W, Z, \beta) &= (Y - m)\left(\xi_\beta - \frac{1}{2}\sigma_u^2 \xi_{xx\beta}\right) \qquad (106)\\ &+ \frac{1}{2}\sigma_u^2 \left[2C''\left(\xi_x\xi_{x\beta} + \xi_{xx}\xi_\beta - \frac{W - \mu_w}{\sigma_w^2}\xi_x\xi_\beta\right) + 2C'''\xi_x^2\xi_\beta\right]\\ &+ \sigma_u^4 \cdot rest,\end{aligned}$$

where $rest$ satisfies (105).

(b) *Expansion of* B_c. Substituting (106) in (32), we get at the true parameter values $\beta = \beta_0, \varphi = \varphi_0$

$$B_c = \mathrm{E}v(\xi_\beta\xi_\beta^t - \frac{1}{2}\sigma_u^2[\xi_{xx\beta}\xi_\beta^t]_S) + O(\sigma_u^4).$$

Using expansion (89) and integraton by part, (100) and (101), we get

$$\begin{aligned}\varphi^{-1}B_c &= \mathrm{E}\{C''\xi_\beta\xi_\beta^t - \frac{1}{2}\sigma_u^2[C''[\xi_{xx\beta}\xi_\beta^t]_S\\ &+ C'''(2(\xi_x\xi_\beta\xi_\beta^t)_x - \xi_{xx}\xi_\beta\xi_\beta^t) \qquad (107)\\ &+ C^{(4)}\xi_x^2\xi_\beta\xi_\beta^t \ - 2\varphi^{-1}C''^2\xi_x^2\xi_\beta\xi_\beta^t]\} + O(\sigma_u^4).\end{aligned}$$

(c) *Expansion of* A_c. Remember that $A_c = S_0$, see (33). Define the function $F(\cdot) = C''(\xi)\xi_\beta\xi_\beta^t$, $\xi = \xi(\cdot, Z, \beta)$, so that $A_c = \mathrm{E}F(X)$. As $W = X + U$, we have the expansion

$$F(W) = F(X) + F'(X)U + \frac{1}{2}F''(X)U^2 + O(U^3).$$

We can replace $F''(X)$ by $F''(W)$ because $F''(W) = F''(X) + O(U^2)$. Therefore

$$\mathrm{E}F(W) = \mathrm{E}F(X) - \frac{1}{2}\sigma_u^2\mathrm{E}F''(W) + O(\sigma_u^4).$$

Now

$$\begin{aligned}F''(W) &= C''(\xi_\beta\xi_\beta^t)_{xx} + C'''\{2(\xi_x\xi_\beta\xi_{\beta^t})_x - \xi_{xx}\xi_\beta\xi_\beta^t\}\\ &+ C^{(4)}\xi_x^2\xi_\beta\xi_\beta^t + O(\sigma_u^4).\end{aligned}$$

As always $C^{(i)} = C^{(i)}(\xi)$, and ξ and its derivatives are taken at (W, Z, β). It follows that

$$\begin{aligned} A_c &= \mathrm{E}F(W) - \frac{1}{2}\sigma_u^2 \mathrm{E}F''(W) + O(\sigma_u^2), \\ A_c &= \mathrm{E}\{C''\xi_\beta\xi_\beta^t - \frac{1}{2}\sigma_u^2[C''(\xi_\beta\xi_\beta^t)_{xx} \qquad (108) \\ &+ C'''(2(\xi_x\xi_\beta\xi_\beta^t)_x - \xi_{xx}\xi_\beta\xi_\beta^t) + C^{(4)}\xi_x^2\xi_\beta\xi_\beta^t]\} + O(\sigma_u^4). \end{aligned}$$

(d) *Expansion for (34).* We represent (107) and (108) in the form

$$\varphi^{-1}B_c = S - \frac{1}{2}\sigma_u^2\Delta B + O(\sigma_u^4),$$
$$A_c = S - \frac{1}{2}\sigma_u^2\Delta A + O(\sigma_u^4),$$

where S is given in (19). Note that S and ΔA are symmetric matrices. From (34) we now obtain

$$\varphi^{-1}\Sigma_{CS} = S^{-1} + \frac{1}{2}\sigma_u^2 S^{-1}(2\Delta A - \Delta B)S^{-1} + O(\sigma_u^4). \qquad (109)$$

A simple calculation now shows that the right hand sides of (109) and (21) are identical. This completes the proof of Theorem 5.1.

11.9 Proof of Theorem 6.1, part a)

Consider the limit estimating function in (36)

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

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

$$\left.\frac{\partial Q(\beta, \sigma_u^2)}{\partial \beta^t}\right|_{\beta=\beta_0, \sigma_u^2=0} = -S_0, \qquad (110)$$

where S_0 is given in (18). 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 neighbohood 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 $Q(\beta_1) = Q(\beta_2)$, where we suppressed the dependence on σ_u^2. Let $\beta(t) = t\beta_1 + (1-t)\beta_2, 0 \le t \le 1$. 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 $\bar{t}, 0 \le \bar{t} \le 1$, 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

$$\frac{\partial Q(\beta, \sigma_u^2)}{\partial \beta^t} = -\mathrm{E}[C''(\xi)\xi_\beta \xi_\beta^t], \tag{111}$$

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

11.10 Proof of Theorem 6.1, parts b) and c)

We want to apply Theorem 12.1 from Heyde (1997). Consider

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

As Q is the limit of $\frac{1}{n}\sum \psi$, see Subsection 11.9,

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

We have to show that $p_\delta < 0$ for some small $\delta > 0$. Let C_0 be the compact set $C_0 := \{\beta| \quad \|\beta - \beta_0\| = \delta\}$. Because $Q(\beta, \sigma_u^2)$ tends to $Q(\beta, 0)$ uniformly on C_0 as $\sigma_u^2 \to 0$, therefore $Q(\beta, \sigma_u^2) = Q(\beta, 0) + R_1$ 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, because $Q(\beta_0, 0) = 0$, see Subsection 11.9, we have with (110)

$$Q(\beta, 0) = \left.\frac{\partial Q}{\partial \beta^t}\right|_{\beta=\beta_0, \sigma_u^2=0} (\beta - \beta_0) + R_2 = -S_0(\beta - \beta_0) + R_2$$

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

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

By assumption (ix) S_0 is positive definite. Hence, $p_\delta < 0$ for small enough 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 (35) eventually has a solution $\hat{\beta}_N$.

By arguments similar to those of the proof of Theorem 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 (36).

11.11 Proof of Theorem 6.1, part d)

Let β_* be the (unique) solution to equation (36). Substituting (87) in (36) we obtain

$$\mathrm{E}\left\{\left[C' + \frac{1}{2}\sigma_u^2\left(-2\frac{W-\mu_w}{\sigma_w^2}C''\xi_x + C''\xi_{xx} + C'''\xi_x^2\right)\right]\xi_\beta - C'(\xi_*)\xi_\beta\right\}$$
$$+O(\sigma_u^4) = 0. \tag{112}$$

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 $\xi(\beta)$:

$$C'(\xi_*) = C' + C''\xi_\beta^t\Delta\beta + rest \cdot \|\Delta\beta\|^2, \tag{113}$$

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

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

We substitute (113) in (112) and obtain

$$\mathrm{E}\left\{\frac{1}{2}\sigma_u^2\left[-2\frac{W-\mu_w}{\sigma_w^2}C''\xi_x + C''\xi_{xx} + C'''\xi_x^2\right]\xi_\beta - C''\xi_\beta\xi_\beta^t\Delta\beta\right\}$$
$$+O(\sigma_u^4) + O(1)\|\Delta\beta\|^2 = 0. \tag{114}$$

Now, due to the implicit function theorem (see Subsecton 11.9), $\Delta\beta = \Delta\beta(\sigma_u^2)$ and $\Delta\beta \in C^1([0,\sigma_o^2])$ for small enough 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 (114) with S from (19)

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

Now, integration by part according to (100) yields (37) and proves the statement.

11.12 Proof of Theorem 6.2

We expand firstly B_*, see (38). From (22) we have

$$\psi(Y,W,Z,\beta_*) = (Y-m)\xi_\beta + (m-C_*)\xi_{\beta x} \tag{115}$$

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

$$B_* = \mathrm{E}(v\xi_\beta\xi_\beta^t + (m-C_*')^2\xi_\beta\xi_\beta^t). \tag{116}$$

The difference $m-C_*' = (m-C_0')+(C_0'-C_*')$, where $C_0' = C'\{\xi(W,Z,\beta_0)\}$, is of the order σ_u^2, see (87) and note the fact that according to Theorem 6.1, part d) $\|\beta_*-\beta_0\| = O(\sigma_u^2)$. Therefore

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

Using (89) and again (100), we get with S from (19)

$$\begin{aligned}
\varphi^{-1}B_* &= S + \frac{1}{2}\sigma_u^2\mathrm{E}\{2\varphi^{-1}C''^2\xi_x^2\xi_\beta\xi_\beta^t + (C'''\xi_{xx} - C^{(4)}\xi_x^2)\xi_\beta\xi_\beta^t \\
&- 2C'''(\xi_x\xi_\beta\xi_\beta^t)_x\} + O(\sigma_u^4) \\
&= : S + \frac{1}{2}\sigma_u^2\Delta B_* + O(\sigma_u^4),
\end{aligned} \tag{117}$$

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

Next we consider A_*. From (38) we have, using again Theorem 6.1, part d),

$$\begin{aligned}
A_* = \mathrm{E}C_*''\xi_\beta\xi_\beta^t &= \mathrm{E}\{(C'' + \frac{1}{2}\sigma_u^2C'''\xi_\beta^t\Delta\beta_*)\xi_\beta\xi_\beta^t\} + O(\sigma_u^4) \\
&= S + \frac{1}{2}\sigma_u^2\mathrm{E}\{C'''(\xi_\beta^t\Delta\beta_*)\xi_\beta\xi_\beta^t\} + O(\sigma_u^4).
\end{aligned} \tag{118}$$

Then by (39), (117), and (118)

$$\varphi^{-1}\Sigma_N = S^{-1} + \frac{\sigma_u^2}{2}S^{-1}[-2\mathrm{E}\{C'''(\xi_\beta^t\Delta\beta_*)\xi_\beta\xi_\beta^t\} + \Delta B_*]S^{-1} + O(\sigma_u^4) \tag{119}$$

Now, (119), (21), and Theorem 5.1 imply (40).

11.13 Proof of Theorem 7.1

We apply Theorem 6.1 and evaluate (37). We have

$$\xi = \beta_0 + \beta_1 W, \quad \xi_x = \beta_1, \quad \xi_\beta = (1, W)^t, \quad \xi_{x\beta} = (0,1)^t,$$
$$S = \mathrm{E}e^\xi \begin{pmatrix} 1 & W \\ W & W^2 \end{pmatrix} = d \begin{pmatrix} 1 & g \\ g & g^2 + \sigma_W^2 \end{pmatrix},$$

where g is given in (48), and

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

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

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

Next,

$$S^{-1} = \sigma_w^{-2} d^{-1} \begin{pmatrix} g^2 + \sigma_W^2 & -g \\ -g & 1 \end{pmatrix}. \tag{120}$$

By (37)

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

and part a) is proved. Part b) follows from Theorem 5.1. In order to show (51), we have to evaluate the right hand side of (40). The expectation in (40) equals

$$\mathrm{E}\Big\{ C''4\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + 4C'''\beta_1 \frac{\partial}{\partial W}\begin{pmatrix} 1 & W \\ W & W^2 \end{pmatrix} + 2C^{(4)}\beta_1^2 \begin{pmatrix} 1 & W \\ W & W^2 \end{pmatrix}$$
$$+ 2C'''[(1,W)\Delta\beta_*]\begin{pmatrix} 1 & W \\ W & W^2 \end{pmatrix}\Big\}$$

$$= \mathrm{E}\left\{e^\xi \left[4\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + 4\beta_1\begin{pmatrix} 0 & 1 \\ 1 & 2W \end{pmatrix} + 2\beta_1^2\begin{pmatrix} 1 & W \\ W & W^2 \end{pmatrix}\right.\right.$$
$$\left.\left. + \frac{2\beta_1}{\sigma_w^2}(\mu_w + g - 2W)\begin{pmatrix} 1 & W \\ W & W^2 \end{pmatrix}\right]\right\}$$

$$= 4d \left[\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \beta_1 \begin{pmatrix} 0 & 1 \\ 1 & 2g \end{pmatrix} \right. + \frac{\beta_1 g}{\sigma_w^2} \begin{pmatrix} 1 & g \\ g & g^2 + \sigma_w^2 \end{pmatrix}$$

$$\left. - \frac{\beta_1}{\sigma_w^2} \begin{pmatrix} g & g^2 + \sigma_w^2 \\ g^2 + \sigma_w^2 & 3g\sigma_w^2 + g^3 \end{pmatrix} \right]$$

$$= 4d \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Next, with (120),

$$S^{-1} 4d \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} S^{-1} = \frac{4d^{-1}}{\sigma_w^4} \begin{pmatrix} g^2 & -g \\ -g & 1 \end{pmatrix} \tag{121}$$

From (121) and (40) we get (51).

11.14 Proof of Theorem 8.1, part a)

We first consider Φ from condition (viii). From (89) we have

$$\varphi^{-1} v = C'' + \chi^{-2} C''^2 \xi_x^2 + \sigma_u^2 \cdot rest, \tag{122}$$

where, as usual, here and in the sequel, $rest$ satisfies (88). Similarly, by (98),

$$\frac{\partial m}{\partial \beta} \frac{\partial m}{\partial \beta^t} = C''^2 \xi_\beta \xi_\beta^t + \sigma_u^2 \cdot rest. \tag{123}$$

(122) and (123) imply

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

with v_0 from (52). Hence

$$\varphi \Phi = \chi^2 \mathrm{E}(C'' \xi_\beta \xi_\beta^t v_0^{-1}) + O(\sigma_u^2). \tag{124}$$

We now turn to $F_p, p = 1, 2$, defined in (16). By (97)

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

where *rest* satisfies (88). With (93) it follows that $\left|\frac{1}{v}\frac{\partial m}{\partial \gamma_p}\right|$ also satisfies (88). Therefore

$$F_p = O(1). \tag{125}$$

Now applying (125) to (17), we get

$$\varphi^{-1}\Sigma_{SQS} = (\varphi\Phi)^{-1} + O(\varphi), \tag{126}$$

and the statement follows by substituting (124) in (126), and taking account of (xiv).

11.15 Proof of Theorem 8.1, parts b) and c)

From (107) we have

$$\varphi^{-1}B_c = \mathrm{E}C''\xi_\beta\xi_\beta^t(1+\chi^{-2}C''\xi_x^2) + O(\sigma_u^2). \tag{127}$$

From (108) we get

$$A_c = S + O(\sigma_u^2). \tag{128}$$

Therefore

$$\varphi^{-1}\Sigma_{CS} = A_c^{-t}(\varphi^{-1}B_c)A_c^{-1} = S^{-1}\mathrm{E}C''\xi_\beta\xi_\beta^t(1+\chi^{-2}C''\xi_x^2)S^{-1} + O(\sigma_u^2) \tag{129}$$

Multiplying by φ and taking account of (xiv) and (52), we get the desired result.
For Σ_N, we have the same expansion (129) because the expansions (127), (128) are valid for A_* and B_* too, see (117), (118).

11.16 Proof of Theorem 8.2

We apply the expressions for Σ_{SQS} and Σ_{CS} from Theorem 8.1 a) and b). Note that by our assumptions the distribution of $v_o = \chi^2 + C''\xi_x^2$ as a function of (W, Z, β_0) has no atoms. Note also that according to Remark 4.3 S can be assumed as nonsingular for small enough σ_u^2. Let $w = S^{-1/2}\sqrt{C''}\xi_\beta$. Then $\mathrm{E}ww' = I$, and part a) of Theorem 8.2 follows from Lemma A2 (see Appendix).

To prove (53), 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 8.1 and by the continuity of $\lambda_{max}(\cdot)$ we have

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

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

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

$D(0, \chi^2)$ 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 λ_{max} is positive definite by Lemma A2, 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 (53). (54) is proved in a similar way.

12 Appendix

12.1 Asymptotic normality of an estimator in the presence of nuisance parameters

Consider a sequence of random fields $G_n(\theta,\gamma), n = 1,2,\dots$ with values in $I\!R^d, \theta \in \Theta_\theta$ and $\gamma \in \Theta_\gamma$, where Θ_θ and Θ_γ are open sets in $I\!R^d$ and $I\!R^k$, respectively. We may think of $G_n(\theta,\gamma)$ as score functions constructed from an observed sample.
Let $\theta_0 \in \Theta_\theta$ and $\gamma_0 \in \Theta_\gamma$ be the true values of the parameters. Suppose that a consistent estimator $\hat{\gamma}_n$ of γ_0 is given. We define an estimator $\hat{\theta}_n$ of θ_0 as a measurable solution to the equation

$$G_n(\theta, \hat{\gamma}_n) = 0, \quad \theta \in \Theta_\theta.$$

More precisely, we suppose that the equality $G_n(\hat{\theta}_n, \hat{\gamma}_n) = 0$ holds with probability tending to 1 as $n \to \infty$.

Lemma A1. Let the following conditions hold.

a) $\hat{\theta}_n$ is consistent, i.e., $\hat{\theta}_n \to \theta_0$ in probability.

b) $G_n(\theta,\gamma) \in C^1(\Theta_\theta \times \Theta_\gamma)$, a.s.

c) $\binom{\sqrt{n}G_n(\theta_0,\gamma_0)}{\sqrt{n}(\hat{\gamma}_n-\gamma_0)} \xrightarrow{d} N(0,\Sigma)$, where Σ is a positive semidefinite matrix.

d) $\frac{\partial G_n(\theta_0,\gamma_0)}{\partial \theta^t} \to V_1$ in probability, where V_1 is a non-random nonsingular matrix.

e) $\frac{\partial G_n(\theta_0,\gamma_0)}{\partial \gamma^t} \to V_2$ in probability, where V_2 is a non-random matrix.

f) For any $\delta > 0$,

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} \sup P\left\{ \sup_{\substack{||\theta-\theta_0||\le\varepsilon,\\ ||\gamma-\gamma_0||\le\varepsilon}} ||\frac{\partial G_n(\theta,\gamma)}{\partial(\theta^t,\gamma^t)} - \frac{\partial G_n(\theta_0,\gamma_0)}{\partial(\theta^t,\gamma^t)}|| \ge \delta \right\} = 0.$$

Then $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0,\Sigma_\theta)$, with $\Sigma_\theta = V_1^{-1}(I_d,V_2)\Sigma(I_d,V_2)^t V_1^{-t}$.

Proof: Let G_n^i be the i-th component of the column vector G_n and let $B(\theta_0, r_1)$ and $B(\gamma_0, r_2)$ be open balls in $I\!R^d$ and $I\!R^k$ with centers at θ_0 and γ_0, respectively. Consistency of $\hat{\theta}_n$ and $\hat{\gamma}_n$ implies that $\hat{\theta}_n \in B(\theta_0, r_1) \subset \Theta_\theta$ and $\hat{\gamma}_n \in B(\gamma_0, r_2) \subset \Theta_\gamma$ with large probability, i.e., with probability tending to 1 as $n \to \infty$. As $G_n(\hat{\theta}_n, \hat{\gamma}_n) = 0$, we obtain that with large probability

$$G_n^i(\theta_0, \gamma_0) + \frac{\partial G_n^i(\bar{\theta}_i, \bar{\gamma}_i)}{\partial \theta^t}(\hat{\theta}_n - \theta_0) + \frac{\partial G_n^i(\bar{\theta}_i, \bar{\gamma}_i)}{\partial \gamma^t}(\hat{\gamma}_n - \gamma_0) = 0,$$

where $(\bar{\theta}_i, \bar{\gamma}_i)$ are intermediate points on the line connecting (θ_0, γ_0) and $(\hat{\theta}_n, \hat{\gamma}_n)$. It follows that

$$\sqrt{n}G_n(\theta_0, \gamma_0) + \frac{\partial G_n(\theta_0, \gamma_0)}{\partial \theta^t}\sqrt{n}(\hat{\theta}_n - \theta_0) + \frac{\partial G_n(\theta_0, \gamma_0)}{\partial \gamma^t}\sqrt{n}(\hat{\gamma}_n - \gamma_0) + R_n = 0$$

where

$$\begin{aligned} R_n &= \Lambda_n\sqrt{n}(\hat{\theta}_n - \theta_0) + M_n\sqrt{n}(\hat{\gamma}_n - \gamma_0), \\ \Lambda_n^{ij} &= \frac{\partial G_n^i(\bar{\theta}_i, \bar{\gamma}_i)}{\partial \theta_j} - \frac{\partial G_n^i(\theta_0, \gamma_0)}{\partial \theta_j}, \quad i, j = 1, 2, \ldots, d. \\ M_n^{ij} &= \frac{\partial G_n^i(\bar{\theta}_i, \bar{\gamma}_i)}{\partial \gamma_j} - \frac{\partial G_n^i(\theta_0, \gamma_0)}{\partial \gamma_j}, \quad i = 1, \ldots, d, j = 1, \ldots, k. \end{aligned}$$

Consequently we obtain

$$\begin{aligned} &\left(\frac{\partial G_n(\theta_0, \gamma_0)}{\partial \theta^t} + \Lambda_n\right)\sqrt{n}(\hat{\theta}_n - \theta_0) \\ &= -\sqrt{n}G_n(\theta_0, \gamma_0) - \left(\frac{\partial G_n(\theta_0, \gamma_0)}{\partial \gamma^t} + M_n\right)\sqrt{n}(\hat{\gamma}_n - \gamma_0). \end{aligned} \tag{130}$$

Now, $\Lambda_n \to 0$ in probability. Indeed, for any $\varepsilon > 0$ and $\delta > 0$,

$$\begin{aligned} P(||\Lambda_n|| \geq \delta) &\leq P(||\hat{\theta}_n - \theta_0|| > \varepsilon \text{ or } ||\hat{\gamma}_n - \gamma_0|| > \varepsilon) \\ &+ P(\sup_{\substack{||\bar{\theta}_i - \theta_0|| \leq \varepsilon, ||\bar{\gamma}_i - \gamma_0|| \leq \varepsilon, \\ i = 1, \ldots, d)}} ||\Lambda_n|| \geq \delta, \end{aligned}$$

and, due to consistency of $\hat{\theta}_n$ and $\hat{\gamma}_n$,

$$\limsup_{n\to\infty} P(||\Lambda_n|| \geq \delta) \leq \limsup_{n\to\infty} P(\sup_{\substack{||\bar{\theta}_i - \theta_0|| \leq \varepsilon, \\ ||\bar{\gamma}_i - \gamma_0|| \leq \varepsilon}} ||\Lambda_n|| \geq \delta).$$

But because of condition f) the last expression tends to 0 as $\varepsilon \to 0$. Thus $\Lambda_n \to 0$ in probability. Similarly $M_n \to 0$ in probability. Then (130) implies the desired convergence of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. Indeed, using c), d), and e) we get

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \stackrel{d}{\to} V_1^{-1}(I_d, V_2) \cdot N(0, \Sigma),$$

which implies the statement of Lemma A1.

12.2 A matrix inequality

Lemma A2. Let v be a positive random variable and w a random (column) vector with $\mathrm{E}(ww^t) = I$. Assume $\mathrm{E}(v^{-1}w^t w) < \infty$ and $\mathrm{E}(vw^t w) < \infty$, then

$$\Delta := \mathrm{E}(vww^t) - [\mathrm{E}(v^{-1}ww^t)]^{-1}.$$

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

Proof: Let

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

Then $\mathrm{E}qq^t = \Delta$, which is positive semidefinite.

In Kukush et al (2000), Lemma 4, it was proved that, under the assumptions of the Lemma,

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

This implies $\Delta \neq 0$.

Acknowledgement: We thank T. Augustin for useful discussions. Support from Deutsche Forschungsgemeinschaft is gratefully acknowledged.

13 References

Aitchison, J. and Silvey, S.D. (1958), *Maximum likelihood estimation of parameters subject to retraint*, Ann. Math. Statist. 29, 813-828.

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

Carroll, R.J. and Ruppert, D. (1988), *Transformation and Weighting in Regression*, Chapman and Hall, London.

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

Cheng, C.-L. and Schneeweiss, H. (1998), *Polynomial regression with errors in the variables*, J.R. Statist. Soc. B60, 189-199.

Cheng, C.-L. and Schneeweiss, H. (2002), *On the Polynomial Measurement Error Model*. In S. van Huffel and P. Lemmerling (Eds.): Total Least Squares and Errors-in-Variables Modeling. Kluwer, Dordrecht, 131-143.

Cheng, C.-L., Schneeweiss, H. and Thamerus, M. (2000), *A small sample estimator for a polynomial regression with errors in the variables*, J.R. Statist. Soc. B 62, 699-709.

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

Foutz, R.V. (1997), *On the unique consistent solution to the likelihood equations*, JASA, 72, 357, 147-148.

Fuller, W.A. (1987), *Measurment Error Models*, New York: Wiley

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

Kuha, J.T. and Temple, J. (1999), *Covariate measurement error in quadratic regression* Discussion Paper 1999-W2, Nuffield College Oxford.

Kukush, A. and Schneeweiss, H.(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, University of Munich.

Kukush, A., Schneeweiss, H. and Wolf, R. (2001a), *Comparison of three Estimators in a Polynomial Regression with Measurement Errors*, Discussion Paper 233, Sonderforschungsbereich 386, University of Munich

Kukush, A., Schneeweiss, H. and Wolf, R. (2001b), *Three Estimators in the Poisson Regression Model with Measurement Errors*, Discussion Paper 243, Sonderforschungsbereich 386, University Munich

McCullagh and Nelder (1989), *Generalized Linear Models*, 2nd edition. Chapman & Hall London.

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

Pfanzagl, J. (1969), *On the measurability and consistency of minimum contrast estimates*, Metrika 14, 249-273.

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

Schneeweiss, H. and Nittner, T. (2001), *Estimating a polynomial regression with measurement errors in the structural and in the functional case - a comparison.* In M. Sadeh (Ed.): Data Analysis from Statistical Foundations, A Festschrift in Honour of the 75th Birthday of D.A.S. Fraser, Nova Science, New York, 195-205.

Shklyar, S. and Kukush, A. (2002), *Comparison of three estimators in the Poisson errors-in-variables model with one covariate*, unpublished paper, Department of Mathematics, Taras Shevchenko University, Kiev, Ukraine.

Stefanski, L.A. (1989), *Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models*, Commun. Statist. - Theory Meth, 18 (12), 4335-4358.

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

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