Volume 14, No. 1 (2005), pp. 53–79 Allerton Press, Inc.

MATHEMATICAL METHODS OF STATISTICS

COMPARING DIFFERENT ESTIMATORS IN A NONLINEAR MEASUREMENT ERROR MODEL. I

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

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

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

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 of the regression parameter vector. 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 are proved. The corrected score (CS) estimator is based on an error-corrected likelihood score function. It turns out that for small error variances the SQS and CS estimators are approximately equally efficient.

In Part II of the paper, we shall, among other things, investigate the naive estimator, which is obtained by applying the maximum likelihood (ML) estimator ignoring the measurement error, and the true ML estimator. We shall also study two examples: the polynomial and the Poisson regression models.

Key words: exponential family, GLM, structural errors-in-variables model, asymptotic covariance matrix, efficiency, small measurement error variance.

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

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), Wansbeek and Meijer(2000), 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 possibly 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. This model is very much like a Generalized Linear Model (GLM), except that in the latter the covariates first are combined linearly and then mapped to the canonical parameter via a link function, whereas here the canonical parameter is directly linked to the covariates via an arbitrary function.

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 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 Gleser (1990), Carroll and Ruppert (1988), Heyde (1997), Carroll *et al.* (1995), and Thamerus (1998) for special cases. The estimator is a special GMM (Generalized Method of Moments) or M-estimator adapted to a measurement error model, see, e.g., Huber (1981), Schervish (1995). However, here we have to include the effect of nuisance parameters, see below. 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 order σ_u^4. It is due to these nuisance parameters that the derivation of the asymptotic properties of the SQS estimator needs special efforts that go beyond what can be found in the literature.

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, the CS procedure does not require knowledge of the distribution of the error-prone covariate. The CS estimator is asymptotically normal with an asymptotic covariance matrix, which can be evaluated, Gimenez and Bolfarine (1997).

We want to compare the relative asymptotic efficiency of these two 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 (2003), Schneeweiss and Nittner (2001). But to the best of our knowledge, theoretical comparative studies have not been carried out so far except for the special cases of the polynomial and the Poisson regression models, Kukush and Schneeweiss (2000), Kukush *et al.* (2004), (2005), Shklyar and Schneeweiss(2002). In the linear model, SQS and CS coincide, so there is nothing to be shown in this case.

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.

When only σ_u^2 tends to zero and φ stays fixed, it turns out, surprisingly enough, that the difference of the asymptotic covariance matrices of the SQS and CS estimators is of order σ_u^4. This is our main result. A similar result holds true for the ML estimator, see Part II.

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. Equality of the covariance matrices of the SQS and CS estimators up to the order of σ_u^4 is established. Section 6 contains some concluding remarks. The proofs are presented in Section 7. In the Appendix, we present a general expression for the covariance matrix of a parameter of interest when nuisance parameters are present.

2. The Model

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

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

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

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

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 sufficiently smooth (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

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

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

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

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 from the observations (Y_i, W_i, Z_i), $i = 1, \ldots, n$.

3. The Structural Quasi-Score (SQS) Estimator

Let us introduce the conditional mean and variance of Y,

$$m(W, Z, \beta) := \mathrm{E}\,(Y \mid W, Z), \tag{5}$$

$$v(W, Z, \beta, \varphi) := \mathrm{Var}(Y \mid W, Z). \tag{6}$$

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

$$m(W, Z, \beta) = \mathrm{E}\,[C'\{\xi(X, Z, \beta)\} \mid W, Z], \tag{7}$$

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

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

$$\mu(W) = W - \frac{\sigma_u^2}{\sigma_w^2}(W - \mu_w), \tag{9}$$

$$\tau^2 = \sigma_u^2 - \frac{\sigma_u^4}{\sigma_w^2}. \tag{10}$$

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

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

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

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

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

4. Asymptotic Properties of $\hat{\beta}_{SQS}$

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) $C(\xi) \in C^4(\mathbb{R})$, and for all x, z, β,

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

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

Note that one could just as well have taken a bound of the form $\text{const} \cdot \exp[B(|X| + \|Z\|)]$ 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 $A > 0$

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

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 $\xi(\cdot, Z, \cdot) \in C^5(\mathbb{R} \times U(\Theta_\beta))$.

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

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

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

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

$$\Phi := \mathrm{E}\left(v^{-1}\frac{\partial m}{\partial \beta}\frac{\partial m}{\partial \beta^t}\right)\Bigg|_{\theta=\theta_0}. \tag{13}$$

4.2. Consistency and uniqueness of $\hat{\boldsymbol{\beta}}_{\boldsymbol{SQS}}$. 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 $U_1, U_2, \dots$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$. A sequence of statements $A_n(U_n)$, $n = 1, 2, \dots$, is said to hold *eventually* if

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

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.*

Consistency of M-estimators under general conditions has also been proved by Huber (1981, Chapter 6, Theorem 2.2) and the existence of a solution has been shown by Schervish (1995, Theorem 7.72) but in both cases without the presence of nuisance parameters.

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

$$S_n(\beta, \alpha, \varphi) := \frac{1}{n} \sum_{i=1}^{n} \{Y_i - \hat{m}(W_i, Z_i, \beta)\} \hat{v}^{-1}(W_i, Z_i, \alpha, \varphi) \frac{\partial \hat{m}(W_i, Z_i, \beta)}{\partial \beta}. \tag{14}$$

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 equations (11), (12) to the form

$$S_n(\beta, \beta, \varphi) = 0, \tag{15}$$

$$\varphi = P \circ h_n(\beta). \tag{16}$$

The following algorithm to solve (15), (16) is a modification of the iteratively reweighted least squares procedure, see Carroll and Ruppert (1988), see also Small and Wang (1993).

1) Given an estimate $\beta^{(j)} \in \Theta_\beta$ from the j-th round of the algorithm, find $\varphi^{(j)}$ from (16) 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* (viii)*, 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

$$F_p := \mathrm{E}\left(v^{-1}\frac{\partial m}{\partial \beta}\frac{\partial m}{\partial \gamma_p}\right)\bigg|_{\theta=\theta_0, \gamma=\gamma_0}. \tag{17}$$

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

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

with

$$\Sigma_{SQS} = \Phi^{-1} + \Phi^{-1}\left(\sigma_{w0}^2 F_1 F_1^t + \frac{2}{\sigma_{w0}^4} F_2 F_2^t\right)\Phi^{-1}, \tag{18}$$

where Φ *is given in* (13)*.*

Remark 4.1. If μ_w and σ_w^2 are known, then $\Sigma_{SQS} = \Phi^{-1}$, see Carroll *et al.* (1995), see also Schervish (1995, Theorem 7.75). The additional terms in (18) 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:

$$S_0 := \mathrm{E}\,(C''(\xi)\xi_\beta \xi_\beta^t)\Big|_{\beta=\beta_0}, \qquad \xi = \xi(X, Z, \beta), \tag{19}$$

$$S := \mathrm{E}\,(C''(\xi)\xi_\beta \xi_\beta^t)\big|_{\beta=\beta_0}, \qquad \xi = \xi(W, Z, \beta). \tag{20}$$

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 (19) 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 sufficiently small 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, $P(a'\xi_\beta \neq 0) > 0$ for all $a \in \mathbb{R}^k$, $a \neq 0$. Note that when $\xi(X, Z, \beta) = g(\beta_0 + \beta_1 X + \beta_z' Z)$ with some function g, then (ix) implies $\sigma_x^2 > 0$.

Theorem 4.4. *Assume* (i) *to* (ix) *with* (vii) *for sufficiently small* σ_u^2. *Let* $\varphi = \varphi_0$ *be fixed and* $\sigma_u^2 \to 0$. *Then*

$$\Sigma_{SQS} = \Phi^{-1} + O(\sigma_u^4), \tag{21}$$

and furthermore

$$\begin{aligned} \Sigma_{SQS} = \varphi S^{-1} + \frac{1}{2}\sigma_u^2 \varphi S^{-1} \mathrm{E}\Big\{ & 2\varphi^{-1} C''^2 \xi_x^2 \xi_\beta \xi_\beta^t + C''[(\xi_\beta \xi_\beta^t)_{xx} + 2\xi_{x\beta}\xi_{x\beta}^t] \\ & + C'''[2(\xi_x \xi_\beta \xi_\beta^t)_x - \xi_{xx}\xi_\beta \xi_\beta^t] + C^{(4)}\xi_x^2 \xi_\beta \xi_\beta^t \Big\} S^{-1} + O(\sigma_u^4), \end{aligned} \tag{22}$$

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

$$\psi(y, x, z, \beta) = y\xi_\beta - C'(\xi)\xi_\beta, \tag{23}$$

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), $i = 1, \ldots, n$, 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

$$\mathrm{E}\,[\{C'(\xi_0) - C'(\xi)\}\xi_\beta] = 0, \quad \beta \in \Theta_\beta, \tag{24}$$

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 (24) has the unique solution $\beta = \beta_0$.

This is an identifiability 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

$$\mathrm{E}\{\psi_c(Y, W, Z, \beta) \mid Y, X, Z\} = \psi(Y, X, Z, \beta). \tag{25}$$

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

$$\mathrm{E}\{f_{ic}(W, Z, \beta) \mid X, Z\} = f_i(X, Z, \beta), \qquad i = 1, 2. \tag{26}$$

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 (26) defined for all β in a 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 $(w, z, \beta) \in \mathbb{R} \times \mathrm{E}_z \times \Theta_\beta$ as long as $\sigma_u^2 \le \sigma_0^2$ for some fixed $\sigma_0^2 > 0$:

$$D_\beta^j f_{ic} = D_\beta^j f_i - \frac{1}{2}\sigma_u^2 D_\beta^j (f_i)_{xx} + \sigma_u^4 \cdot R, \tag{27}$$

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

$$\|R\| \le \mathrm{const}(e^{A|w|} + e^{A\|z\|}) \tag{28}$$

with const > 0 and fixed $A > 0$. Here and in the sequel R denotes an unspecified remainder term.

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

$$\psi_c(y, w, z, \beta) := y f_{1c}(w, z, \beta) - f_{2c}(w, z, \beta) \tag{29}$$

satisfies (25). Condition (xii) with $j = 0$ gives an approximation of f_{1c}, f_{2c} (and therefore of ψ_c) for sufficiently small σ_u^2. This approximation is based on the representation of the solution f_{ic} in the form

$$f_{ic} = f_i - \frac{1}{2}\sigma_u^2 (f_i)_{xx} + \sigma_u^4 \sum_{k=2}^{\infty} \frac{(-\sigma_u^2)^{k-2}}{2^k k!} \frac{\partial^{2k} f_i}{\partial x^{2k}}, \tag{30}$$

which was shown in Stefanski (1989), p. 4344, in a regular case. Therefore for $j = 0$ in (27)

$$R = \sum_{k=2}^{\infty} \frac{(-\sigma_u^2)^{k-2}}{2^k k!} \frac{\partial^{2k} f_i}{\partial x^{2k}}, \tag{31}$$

and for this expression, with $\sigma_u^2 \le \sigma_0^2$, we require the bound (28). To justify (27) and (28) for $j = 1$, one can assume that (26) is differentiable with respect to β, i.e.,

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

and use a representation like (30) for $(f_{ic})_\beta$. We will see that assumptions (xi) and (xii) hold for the polynomial and Poisson regression models.

Now, for the corrected score function ψ_c given in (29), we define the corrected score (CS) estimator $\hat\beta_{CS}$ as a measurable solution to

$$\frac{1}{n}\sum_{i=1}^{n}\psi_c(Y_i, W_i, Z_i, \beta) = 0, \quad \beta \in \Theta_\beta. \tag{32}$$

Methods to solve equations like (32) are found, e.g., in Small and Wang (1993). Note that (24) is the limit estimating equation for $\hat\beta_{CS}$ (32) 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

$$A_c := -\mathrm{E}\,\frac{\partial\psi_c}{\partial\beta^t}\Big|_{\beta=\beta_0}, \qquad B_c := \mathrm{E}\,\psi_c\psi_c^t\big|_{\beta=\beta_0}, \tag{33}$$

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 (25), (23), 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))

$$A_c = \mathrm{E}\,(C''\xi_\beta\xi_\beta^t)\big|_{\beta=\beta_0} = S_0, \tag{34}$$

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, see also Schervish (1995, Theorem 7.75) and Giminez and Bolfarine (1997),

$$\Sigma_{CS} = A_c^{-1}B_cA_c^{-1}. \tag{35}$$

The following statement is the central result of the paper.

Theorem 5.1. *Assume* (i) *to* (vi), (vii) *for sufficiently small* 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. 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 σ_u^2 both estimators are approximately equally efficient, more precisely: the difference of their asymptotic covariance matrices is of order σ_u^4.

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.

7. 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".

7.1. Proof of Theorem 4.1(a). To simplify the following arguments we slightly modify the estimating equation (12) by replacing the term $\frac{1}{n-k}$ with $1/n$. The right-hand side of the modified equation (12) differs from the original one by a term of order $1/n^2$ 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

$$G(Y,W,Z;\theta;\mu_w,\sigma_w^2) = \begin{pmatrix} (Y-m)v^{-1}\frac{\partial m}{\partial \beta} \\ (Y-m)^2 - A_1 - \varphi A_2 \end{pmatrix}, \tag{36}$$

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), $i = 1, \dots, n$, the estimating function is defined by

$$G_n(\theta;\mu_w,\sigma_w^2) = \frac{1}{n}\sum_{i=1}^n G(Y_i,W_i,Z_i;\theta;\mu_w,\sigma_w^2). \tag{37}$$

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

$$G_n(\theta; \hat{\mu}_w, \hat{\sigma}_w^2) = 0, \qquad \theta \in \Theta := \Theta_\beta \times [a_1, b_1]. \tag{38}$$

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 (37).

1°. 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 (36), we have by the strong LLN a.s.

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

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 $\sup_{n \ge 1} \sup_{\eta \in \Theta_\eta} \| \frac{\partial G_n^1}{\partial \eta^t} \| < \infty$ 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 property 3° 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 $\forall \omega \in \Omega_{0i}$: $G_n(\eta_i) \to G_\infty(\eta_i)$. Then take $\Omega_0 = \cap_{i=1}^\infty \Omega_{0i}$. For any $\eta \in \Theta_\eta$ and $\omega \in \Omega_0$ we then have $|G_\infty(\eta) - G_n(\eta)| \le |G_\infty(\eta) - G_\infty(\eta_i)| + |G_\infty(\eta_i) - G_n(\eta_i)| + |G_n(\eta_i) - G_n(\eta)|$, which becomes arbitrarily small if η_i is 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 $\forall n > N$, $\forall \eta \in \Theta_\eta$: $|G_n(\eta) - G_\infty(\eta)| < \varepsilon$.

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

This property follows from property 1° and from the fact that $\hat{\mu}_w$ and $\hat{\sigma}_w^2$ are strongly consistent.

3°. $G_\infty(\eta) = (G_\infty^1(\eta)^t, G_\infty^2(\eta))^t$, with

$$G_\infty^1(\eta) = \mathrm{E}\left[(m_0 - m) v^{-1} \frac{\partial m}{\partial \beta} \right], \tag{39}$$

$$G_\infty^2(\eta) = \mathrm{E}\,[(m_0 - m)^2 + A_{10} + \varphi_0 A_{20} - A_1 - \varphi A_2], \tag{40}$$

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.

4°. The matrix

$$I_0 := \frac{\partial G_\infty(\eta)}{\partial \theta^t}\Big|_{\eta=\eta_0} \tag{41}$$

is non-singular.

This property follows from the relations

$$\frac{\partial G^1_\infty(\eta)}{\partial \beta^t}\Big|_{\eta=\eta_0} = -\mathrm{E}\left(v^{-1}\frac{\partial m}{\partial \beta}\frac{\partial m}{\partial \beta^t}\right)\Big|_{\eta=\eta_0},$$
$$\frac{\partial G^1_\infty(\eta)}{\partial \varphi} = 0, \qquad \frac{\partial G^2_\infty(\eta)}{\partial \varphi}\Big|_{\eta=\eta_0} = -\mathrm{E}\, A_2\Big|_{\eta=\eta_0},$$

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 (37) of estimating functions, see also Aitchison and Silvey (1958). Set

$$H_n(\theta) := -I_0^{-1} G_n(\theta; \hat\mu_w, \hat\sigma_w^2), \qquad \theta \in \Theta. \tag{42}$$

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

$$q_\delta := \limsup_{n\to\infty}\{\sup_{\|\theta-\theta_0\|=\delta}(\theta-\theta_0)^t H_n(\theta)\} < 0. \tag{43}$$

Indeed, due to property 2° we have

$$q_\delta = \sup_{\|\theta-\theta_0\|=\delta}(\theta-\theta_0)^t\{-I_0^{-1}G_\infty(\theta; \mu_{w0}, \sigma_{w0}^2)\}. \tag{44}$$

Now, $G_\infty(\theta_0; \mu_{w0}, \sigma_{w0}^2) = 0$, which is easily seen from (39), (40). Therefore, using the definition (41) 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 (44) we obtain that, for all small $\delta > 0$, inequality (43) 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 12.1. 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).

7.2. Proof of Theorem 4.1(b). Owing to property 2° 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_0 is the unique solution to $G_\infty(\theta;\mu_{w0},\sigma_{w0}^2)=0$, see (39), (40), 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.

7.3. Proof of Theorem 4.1(c). We apply an approach due to Foutz (1977) based on the Inverse Function Theorem. Due to property 2° from the proof of part (a), the functions H_n of (42) converge a.s. uniformly in Θ to $H_\infty(\theta)=-I_0^{-1}G_\infty\big(\theta;\mu_{w0},\sigma_{w0}^2\big)$, 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 (41), $\frac{\partial H_\infty(\theta)}{\partial\theta^t}\big|_{\theta=\theta_0}=-I_{k+1}$, where I_{k+1} is the $(k+1)\times(k+1)$ identity matrix. Now, we fix a sequence $\hat{\theta}_n$ of SQS estimators such that

$$H_n(\hat{\theta}_n)=0 \qquad \text{eventually.} \tag{45}$$

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

$$\hat{\theta}_n\to\theta_0 \qquad \text{a.s.} \tag{46}$$

Applying the arguments from Foutz (1977) we obtain the following result: if $\{\tilde{\theta}_n\}$ is another sequence satisfying (45), (46), then $\tilde{\theta}_n=\hat{\theta}_n$ eventually. This proves part (c). □

Remark 12.2. The approach of Foutz gives not only uniqueness, but also existence of the estimator with properties (45), (46). Therefore, in Subsection 7.1, 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.

7.4 Proof of Theorem 4.2(a). In a similar way as in Subsection 7.1 it can be shown that the functions S_n of (14) converge a.s. uniformly over $\Theta_\beta\times\Theta_\beta\times[a_1,b_1]$ 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 7.1, that eventually there exists a solution $\hat{\beta}_n(\alpha,\varphi)$ to

$$S_n(\beta,\alpha,\varphi)=0, \qquad \beta\in\Theta_\beta, \tag{47}$$

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 7.2, that the solution to (47) has the property that uniformly in $(\alpha,\varphi)\in\Theta$, as $n\to\infty$,

$$\hat{\beta}_n(\alpha,\varphi)\to\beta_0 \qquad \text{a.s.} \tag{48}$$

Moreover, applying again the arguments of Foutz (1977) as in Subsection 7.3, this solution is seen to be unique eventually.

7.5. Proof of Theorem 4.2(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 (47). 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 (16) 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 (47). Let $\beta_1\ne\beta_2$ and $l:=F(\beta_1)-F(\beta_2)\ne 0$. Then

$$l^t\frac{\partial S^*}{\partial\beta^t}l+l^t\frac{\partial S^*}{\partial\alpha^t}(\beta_1-\beta_2)+l^t\frac{\partial S^*}{\partial\varphi}\big(P\circ h(\beta_1)-P\circ h(\beta_2)\big)=0, \tag{49}$$

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(49) we get with $\vartheta=(\alpha,\varphi)$,

$$\begin{aligned}\|l\|\le\|\beta_1-\beta_2\|&\Big\{\inf_{\beta\in C,\vartheta\in\Theta}\lambda_{\min}\Big(-\frac{\partial S(\beta,\vartheta)}{\partial\beta^t}\Big)\Big\}^{-1}\\&\times\Big\{\sup_{\beta\in C,\,\vartheta\in\Theta}\Big\|\frac{\partial S(\beta,\vartheta)}{\partial\alpha^t}\Big\|+\sup_{\beta\in C,\,\vartheta\in\Theta}\Big\|\frac{\partial S(\beta,\vartheta)}{\partial\varphi}\Big\|\sup_{\beta\in\Theta_\beta}\Big\|\frac{\partial h}{\partial\beta^t}\Big\|\Big\}.\end{aligned} \tag{50}$$

Here $\lambda_{\min}(\cdot)$ is the minimal eigenvalue of a symmetric matrix. Let us consider the various terms of (50). First, due to uniform convergence (48), $C\subset B(\beta_0,\delta)$ for $n\ge n_0(\delta,\omega)$, 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 negative definite, see Subsection 7.4. Therefore, for large n and sufficiently small δ, $\partial S(\beta, \vartheta)/\partial\beta^t$ is negative definite on $B(\beta_0, \delta)$ and hence

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

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 (50) 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, (50) now implies:

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

where $\lambda_n \to 0$ with $n \to \infty$. 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 (15), (16). But $\hat\beta_{SQS}, \hat\varphi_{SQS}$ also satisfy (15), (16) 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 (15), (16) is unique eventually, therefore $\beta_F = \hat\beta_{SQS}$ and $\varphi_F = \hat\varphi_{SQS}$ eventually. This completes the proof. □

7.6. Proof of Theorem 4.3. We reparametrize the score function G of (36) and the estimating function G_n of (37) 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 (37), $G_n(\theta, \gamma)$ instead of $G_n(\theta, \mu_w, \sigma_w^2)$ and similarly for G in (36). The estimators $\hat\beta_{SQS}$ and $\hat\varphi_{SQS}$ satisfy the equation

$$G_n(\hat\beta_{SQS}, \hat\varphi_{SQS}; \hat\mu_w, \hat\sigma_w^{-2}) = 0. \tag{51}$$

We want to derive the asymptotic covariance matrix of $\hat\beta_{SQS}$ with the help of the Lemma in the Appendix. Let us check the conditions of the Lemma. By Theorem 4.1, $\hat\theta = (\hat\beta_{SQS}^t, \hat\varphi_{SQS})^t$ is consistent. The random field $G_n(\theta, \gamma)$, $\theta \in \Theta$, $\gamma \in \Theta_\gamma := [\mu_{w_1}, \mu_{w_2}] \times \big[\frac{1}{b_2}, \frac{1}{a_2}\big]$, has C^1-smooth paths a.s. So conditions (a) and (b) are satisfied.

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

$$\hat\mu_w = \frac{1}{n}\sum_{i=1}^n W_i = \overline{W}, \tag{52}$$

$$\hat\sigma_w^2 = \frac{1}{n-1}\sum_{i=1}^n (H_i - \overline{H})^2 = \overline{H^2} + O_p\Big(\frac{1}{n}\Big), \tag{53}$$

and

$$\hat\sigma_w^{-2} - \sigma_w^{-2} = -\frac{\hat\sigma_w^2 - \sigma_w^2}{\sigma_w^2\hat\sigma_w^2} = -\frac{1}{\sigma_w^4}(\overline{H^2} - \sigma_w^2) + O_p\Big(\frac{1}{n}\Big). \tag{54}$$

From (37), (52), and (54) we get

$$\text{(55)} \qquad \begin{pmatrix} \sqrt{n}G_n(\theta_0,\gamma_0) \\ \sqrt{n}(\hat\gamma_n-\gamma_0) \end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} G(Y_i,W_i,Z_i;\theta_0,\gamma_0) \\ W_i-\mu_{w0} \\ -\frac{H_i^2-\sigma_{w0}^2}{\sigma_{w0}^4} \end{pmatrix} + O_p\Big(\frac{1}{\sqrt{n}}\Big).$$

Note that

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

see (36), (5), (6), and (8). From (55) we have by the CLT for i.i.d. random vectors:

$$\left(\sqrt{n}G_n^t(\theta_0,\gamma_0), \sqrt{n}(\hat\gamma_n-\gamma_0)^t\right)^t \xrightarrow{d} N(0,\Sigma)$$

with

$$\text{(56)} \qquad \Sigma = \mathrm{cov}\begin{pmatrix} G(Y,W,Z;\theta_0,\gamma_0) \\ W-\mu_{w0} \\ -\frac{(W-\mu_{w0})^2-\sigma_{w0}^2}{\sigma_{w0}^4} \end{pmatrix}.$$

Obviously

$$\text{(57)} \qquad \Sigma = \mathrm{diag}(\Sigma_{11}, \sigma_{w0}^2, 2\sigma_{w0}^{-4}),$$

where

$$\text{(58)} \qquad \Sigma_{11} = \begin{pmatrix} \Phi & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix},$$

the matrix Φ comes from (13), σ_{12} and $\sigma_{21}=\sigma_{12}^t$ are unspecified vectors, and σ_{22} is a scalar.

Now pass to condition (d) of the Lemma. 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 (36): $G=(G_1^t,G_2)^t$. Then, with derivatives taken at the true parameter values,

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

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

$$\text{(59)} \qquad \mathrm{V}_1^{-1} = -\begin{pmatrix} \Phi^{-1} & 0 \\ -\frac{1}{\varphi_{22}}\Phi_{12}\Phi^{-1} & \frac{1}{\varphi_{22}} \end{pmatrix}.$$

We now turn to condition (e) of the Lemma. We have

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

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

$$\mathrm{E}\,\frac{\partial G_1}{\partial \gamma^t} = -\mathrm{E}\left(v^{-1}\frac{\partial m}{\partial \beta}\frac{\partial m}{\partial \gamma^t}\right) =: \mathrm{V}_{21}; \tag{60}$$

introduce also

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

Then

$$\mathrm{V}_2 = \begin{pmatrix} \mathrm{V}_{21} \\ \mathrm{V}_{22} \end{pmatrix}. \tag{61}$$

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

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

and, because of the exponential bounds conditions,

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

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

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

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

Introduce the $k\times(k+1)$ selection matrix $P_\beta := (I_k, 0)$. According to (59) 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 (61), (57), and (58):

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

Now by (60) $\mathrm{V}_{21} = -(F_1,F_2)$, where the F_p, $p=1,2$, are given in (17). We thus finally have

$$\Sigma_{SQS} = \Phi^{-1} + \Phi^{-1}\big(\sigma_{w0}^2 F_1F_1^t + 2\sigma_{w0}^{-4}F_2F_2^t\big)\Phi^{-1}. \quad \square$$

7.7. 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 $X \mid W \sim N\{\mu(w), \tau^2\}$:

$$m(w, z, \beta) = \mathrm{E}\, C'[\xi\{\mu(w) + \tau\zeta, z, \beta\}] \tag{62}$$

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

$$|f^{(i)}(w)| \le \mathrm{const} \cdot e^{A|w|}, \qquad w \in \mathbb{R}, \quad i = 0, 1, \ldots, 4. \tag{63}$$

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

$$\begin{aligned}
\mathrm{E}\, f\{\mu(w) + \tau\zeta\} &= \mathrm{E}\, f\Big(w + \tau\zeta - \frac{\sigma_u^2}{\sigma_w^2}(w - \mu_w)\Big) \\
&= \mathrm{E}\Big[\sum_{i=0}^{3} \frac{f^{(i)}(w)}{i!}\Big(\tau\zeta - \frac{\sigma_u^2}{\sigma_w^2}(w - \mu_w)\Big)^i + r_3\Big] \\
&= f(w) - f'(w)\frac{\sigma_u^2}{\sigma_w^2}(w - \mu_w) + \frac{f''(w)}{2}\tau^2 + O(\sigma_u^4) + \mathrm{E}\, r_3.
\end{aligned} \tag{64}$$

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

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

Now, from (63) and (10) we get a bound for r_3 when $\sigma_u^2 \le 1$:

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

with some $A > 0$, and hence

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

In (64) the term $O(\sigma_u^4)$ can be bounded in a similar way. Thus, again with the help of (10), we obtain

$$\mathrm{E}\, f\{\mu(w) + \tau\zeta\} = f(w) + \frac{1}{2}\sigma_u^2\Big(- 2f'(w)\frac{w - \mu_w}{\sigma_w^2} + f''(w)\Big) + \sigma_u^4 \cdot R, \tag{65}$$

$$|R| \le \mathrm{const} \cdot e^{A|w|}. \tag{66}$$

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

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

then for the expectation $\mathrm{E}f\{\mu(w)+\tau\zeta,z\}$ the expansion (65), with suppressed dependence on z, still holds, where the derivatives are taken with respect to w and

$$|R| \le \mathrm{const}(e^{A|w|}+e^{A\|z\|}). \tag{67}$$

Now, we specialize to the function $f=C'\{\xi(w,z,\beta)\}$ and derive an expansion for $m(W,Z,\beta)$ in (62). We use the bounds in (ii) and (vi) and obtain for $\sigma_u^2 \le 1$:

$$m(W,Z,\beta)=C'+\frac{1}{2}\sigma_u^2\Big(-2\frac{W-\mu_w}{\sigma_w^2}C''\xi_x+C''\xi_{xx}+C'''\xi_x^2\Big)+\sigma_u^4\cdot R, \tag{68}$$

$$|R| \le \mathrm{const}(e^{A|W|}+e^{A\|Z\|}). \tag{69}$$

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 (65) specialized, respectively, to $f=C'^2$ and to $f=C'$

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

where R satisfies (69). After simple transformations we get

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

where again R satisfies (69). 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 obtain the same expression as the right-hand side of (68), but with $C^{(i+1)}$ in place of $C^{(i)}$, $i=1,2,3$. Finally, we get from (8)

$$\begin{aligned} \varphi^{-1}v(W,Z,\beta,\varphi) = C''+\frac{\sigma_u^2}{2}\Big[&2\frac{C''^2\xi_x^2}{\varphi}+C'''\xi_{xx} \\ &+C^{(4)}\xi_x^2-2\frac{(W-\mu_w)}{\sigma_w^2}C'''\xi_x\Big]+\sigma_u^4\cdot R, \end{aligned} \tag{70}$$

where R satisfies (69).

(c) *Expansion of* $\frac{\partial m}{\partial\beta}(W,Z,\beta)$. We can differentiate both sides of (62) with respect to β and then use (65) with $f=C''\xi_\beta$. We obtain

$$\frac{\partial m}{\partial\beta}(W,Z,\beta)=C''\xi_\beta+\frac{1}{2}\sigma_u^2T_1+\sigma_u^4\cdot R, \tag{71}$$

$$\begin{aligned} T_1 := &-2\frac{(W-\mu_w)}{\sigma_w^2}(C''\xi_x)_\beta+C''\xi_{xx\beta} \\ &+C'''\big(\xi_{xx}\xi_\beta+2\xi_x\xi_{x\beta}\big)+C^{(4)}\xi_x^2\xi_\beta, \end{aligned} \tag{72}$$

and R satisfies (69).

(d) *Expansion of* $(v/\varphi)^{-1}$. We write (70) as

$$\varphi^{-1}v(W,Z,\beta) = C'' + \frac{\sigma_u^2}{2}T_2 + \sigma_u^4 \cdot R. \tag{73}$$

From condition (iv) and (8) we have

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

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

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

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

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

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

$$\varphi/v \le \text{const}\cdot e^{A|W|}e^{A\|Z\|} \le \text{const}(e^{2A|W|} + e^{2A\|Z\|}). \tag{74}$$

Now, according to (73) 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

$$\begin{aligned} &\left|\frac{\varphi}{v} - \frac{1}{C''}\Big(1 - \frac{\sigma_u^2}{2C''}T_2\Big)\right| \\ &\quad = \frac{\varphi}{v}\left|1 - \Big(C'' + \frac{\sigma_u^2}{2}T_2 + \sigma_u^4 \cdot R\Big)\frac{1}{C''}\Big(1 - \frac{\sigma_u^2}{2C''}T_2\Big)\right| \\ &\quad \le \frac{\varphi}{v}\left[\frac{\sigma_u^4 T_2^2}{4(C'')^2} + \frac{\sigma_u^4 \cdot |R| \cdot \big(1 + \frac{\sigma_u^2}{2C''}|T_2|\big)}{C''}\right]. \end{aligned} \tag{75}$$

Using (74) and (iv) and the fact that owing to (ii) and (vi) we have similar exponential bounds for T_2, we obtain from (75) the expansion

$$\frac{\varphi}{v} = \frac{1}{C''}\Big(1 - \frac{\sigma_u^2}{2C''}T_2\Big) + \sigma_u^4 \cdot R, \tag{76}$$

and R satisfies (69).

(e) *Proof of* (21). From (62) we have

$$\frac{\partial m(w,z,\beta)}{\partial \gamma_p} = \mathrm{E}\left[C''(\xi)\xi_x\Big(\frac{\partial \mu_\xi(w)}{\partial \gamma_p} + \frac{\partial \tau}{\partial \gamma_p}\zeta\Big)\right], \tag{77}$$

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 (77) we get that for $\sigma_u^2\le 1$

$$\left|\frac{\partial m(W,Z,\beta)}{\partial\gamma_p}\right|\le\sigma_u^2|R|, \tag{78}$$

where R satisfies (69) because of (ii) and (vi). Now, from (17), (71), (74), and (78) we have

$$F_p=O(\sigma_u^2),\qquad p=1,2,$$

and (21) follows from (18).

(f) *Proof of* (22). In the sequel R is a quantity which always satisfies (69). From (71) we have

$$\frac{\partial m}{\partial\beta}\frac{\partial m}{\partial\beta^t}=C''^2\xi_\beta\xi_\beta^t+\frac{1}{2}\sigma_u^2C''[T_1\xi_\beta^t]_S+\sigma_u^4\cdot R. \tag{79}$$

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

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

Now, (79) and (76) imply that

$$\frac{\varphi}{v}\frac{\partial m}{\partial\beta}\frac{\partial m}{\partial\beta^t}=C''\xi_\beta\xi_\beta^t+\frac{1}{2}\sigma_u^2\big([T_1\xi_\beta^t]_S-\xi_\beta\xi_\beta^tT_2\big)+\sigma_u^4\cdot R.$$

Taking expectations at $\beta=\beta_0$, we derive the following expression for Φ, as defined in (13), with S from (20):

$$\varphi\Phi=S+\frac{1}{2}\sigma_u^2\mathrm{E}\left([T_1\xi_\beta^t]_S-\xi_\beta\xi_\beta^tT_2\right)+O(\sigma_u^4),$$

and by inversion we get, according to (21),

$$\varphi^{-1}\Sigma_{SQS}=S^{-1}+\frac{1}{2}\sigma_u^2S^{-1}\mathrm{E}\left(\xi_\beta\xi_\beta^tT_2-[T_1\xi_\beta^t]_S\right)S^{-1}+O(\sigma_u^4). \tag{80}$$

To simplify (80), we use integration by part in the form

$$\mathrm{E}\left[\frac{W-\mu_w}{\sigma_w^2}h(W,Z,\beta)\right]=\mathrm{E}\,h_x(W,Z,\beta), \tag{81}$$

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

$$|h(W,Z,\beta)| + |h_x(W,Z,\beta)| \le \mathrm{const}(e^{A|W|} + e^{A\|Z\|}). \tag{82}$$

We have by (81) (with T_2 being the term in brackets of (70))

$$\begin{aligned}\mathrm{E}\,(\xi_\beta\xi_\beta^t T_2) = \mathrm{E}\,\big\{&2\varphi^{-1}C''^2\xi_x^2\xi_\beta\xi_\beta^t \\ &+ C'''\big(-2(\xi_x\xi_\beta\xi_\beta^t)_x + \xi_{xx}\xi_\beta\xi_\beta^t\big) - C^{(4)}\xi_x^2\xi_\beta\xi_\beta^t\big\}.\end{aligned} \tag{83}$$

Next, with T_1 from (72),

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

and applying (81) again we get

$$\begin{aligned}\mathrm{E}\,[T_1\xi_\beta^t]_S = \mathrm{E}\,\big\{&C''(-4\xi_{x\beta}\xi_{x\beta}^t - [\xi_{xx\beta}\xi_\beta^t]_S) \\ &+ C'''[-4(\xi_x\xi_\beta\xi_\beta^t)_x + 2\xi_{xx}\xi_\beta\xi_\beta^t] - 2C^{(4)}\xi_x^2\xi_\beta\xi_\beta^t\big\}.\end{aligned} \tag{84}$$

From (80), (83), and (84) and letting $\beta = \beta_0$ we finally obtain (22). □

7.8. 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 (27) to (29) we get

$$\begin{aligned}\psi_c(y,w,z,\beta) = {}&y\Big(\xi_\beta - \frac{1}{2}\sigma_u^2\xi_{xx\beta}\Big) - C'\xi_\beta \\ &+ \frac{1}{2}\sigma_u^2\big(C'\xi_{xx\beta} + C''\xi_{xx}\xi_\beta + 2C''\xi_x\xi_{x\beta} + C'''\xi_x^2\xi_\beta\big) + \sigma_u^4\cdot R,\end{aligned} \tag{85}$$

where

$$\|R\| \le \mathrm{const}(|y|+1)(e^{A|w|} + e^{A\|z\|}) \le \mathrm{const}\cdot 2(y^2 + e^{2A|w|} + e^{2A\|z\|}). \tag{86}$$

As to (86), note that, for $a>0$, $|y|e^a \le \frac{1}{2}(y^2 + e^{2a})$, and $e^a < e^{2a}$. Now, from (85) and (86) we have

$$\begin{aligned}\psi_c(Y,W,Z,\beta) = {}&(Y-m)\Big(\xi_\beta - \frac{1}{2}\sigma_u^2\xi_{xx\beta}\Big) \\ &+ \frac{1}{2}\sigma_u^2\Big[2C''\Big(\xi_x\xi_{x\beta} + \xi_{xx}\xi_\beta - \frac{W-\mu_w}{\sigma_w^2}\xi_x\xi_\beta\Big) + 2C'''\xi_x^2\xi_\beta\Big] + \sigma_u^4\cdot R,\end{aligned} \tag{87}$$

where R satisfies (86).

(b) *Expansion of* B_c. Substituting (87) in (33), we get at the true parameter values $\beta = \beta_0$, $\varphi = \varphi_0$

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

Using expansion (70) and integration by parts, (81) and (82), we get

$$(88)\quad \varphi^{-1} B_c = \mathrm{E}\Big\{C''\xi_\beta\xi_\beta^t - \frac{1}{2}\sigma_u^2\big[C''[\xi_{xx\beta}\xi_\beta^t]_S + C'''\big(2(\xi_x\xi_\beta\xi_\beta^t)_x - \xi_{xx}\xi_\beta\xi_\beta^t\big) + C^{(4)}\xi_x^2\xi_\beta\xi_\beta^t - 2\varphi^{-1}C''^2\xi_x^2\xi_\beta\xi_\beta^t\big]\Big\} + O(\sigma_u^4).$$

(c) *Expansion of* A_c. Remember that $A_c = S_0$, see (34). 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

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

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

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

$$(89)\quad A_c = \mathrm{E}\Big\{C''\xi_\beta\xi_\beta^t - \frac{1}{2}\sigma_u^2\big[C''(\xi_\beta\xi_\beta^t)_{xx} + C'''(2(\xi_x\xi_\beta\xi_\beta^t)_x - \xi_{xx}\xi_\beta\xi_\beta^t) + C^{(4)}\xi_x^2\xi_\beta\xi_\beta^t\big]\Big\} + O(\sigma_u^4).$$

(d) *Expansion for* (35). We represent (88) and (89) in the form

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

where S is given in (20). Note that S and ΔA are symmetric matrices. From (35) we now obtain

$$(90)\quad \varphi^{-1}\Sigma_{CS} = S^{-1} + \frac{1}{2}\sigma_u^2 S^{-1}(2\Delta A - \Delta B)S^{-1} + O(\sigma_u^4).$$

A simple calculation now shows that the right-hand sides of (90) and (22) are identical. This completes the proof of Theorem 5.1. $\square$

8. Appendix: Asymptotic Normality of an Estimator in the Presence of Nuisance Parameters

Consider a sequence of random fields $G_n(\theta, \gamma)$, $n = 1, 2, \ldots$, with values in $\mathbb{R}^d$, $\theta \in \Theta_\theta$ and $\gamma \in \Theta_\gamma$, where Θ_θ and Θ_γ are open sets in $\mathbb{R}^d$ and $\mathbb{R}^k$, 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

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

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. *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* $\delta > 0$,

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

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 ith component of the column vector G_n and let $B(\theta_0, r_1)$ and $B(\gamma_0, r_2)$ be open balls in $\mathbb{R}^d$ and $\mathbb{R}^k$ 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

$$R_n = \Lambda_n \sqrt{n}(\hat{\theta}_n - \theta_0) + M_n \sqrt{n}(\hat{\gamma}_n - \gamma_0),$$

$$\Lambda_n^{ij} = \frac{\partial G_n^i(\bar{\theta}_i, \bar{\gamma}_i)}{\partial \theta_j} - \frac{\partial G_n^i(\theta_0, \gamma_0)}{\partial \theta_j}, \qquad i, j = 1, 2, \ldots, d,$$

$$M_n^{ij} = \frac{\partial G_n^i(\bar{\theta}_i, \bar{\gamma}_i)}{\partial \gamma_j} - \frac{\partial G_n^i(\theta_0, \gamma_0)}{\partial \gamma_j}, \qquad i = 1, \ldots, d, \quad j = 1, \ldots, k.$$

Consequently we obtain

$$
(91) \qquad \left(\frac{\partial G_n(\theta_0,\gamma_0)}{\partial \theta^t} + \Lambda_n\right)\sqrt{n}(\hat{\theta}_n - \theta_0)
= -\sqrt{n}G_n(\theta_0,\gamma_0) - \left(\frac{\partial G_n(\theta_0,\gamma_0)}{\partial \gamma^t} + M_n\right)\sqrt{n}(\hat{\gamma}_n - \gamma_0).
$$

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

$$
P\big(\|\Lambda_n\| \geq \delta\big) \leq P\big(\|\hat{\theta}_n - \theta_0\| > \varepsilon \quad \text{or} \quad \|\hat{\gamma}_n - \gamma_0\| > \varepsilon\big) + P\Big(\sup_{\|\bar{\theta}_i - \theta_0\| \leq \varepsilon,\, \|\bar{\gamma}_i - \gamma_0\| \leq \varepsilon,\, i=1,\dots,d)} \|\Lambda_n\| \geq \delta\Big),
$$

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

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

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 (91) implies the desired convergence of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. Indeed, using (c), (d), and (e) we get

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

which implies the statement of the lemma. □

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

References

[1] J. Aitchison and S. D. Silvey (1958), *Maximum likelihood estimation of parameters subject to restraint*, Ann. Math. Statist. **29**, 813–828.

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

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

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

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

[6] R. V. Foutz (1977), *On the unique consistent solution to likelihood equations*, J. Amer. Statist. Assoc., **72**, 147–148.

[7] W. A. Fuller (1987), *Measurment Error Models*, Wiley, New York.

[8] P. Gimenez and H. Bolfarine (1997), *Corrected score functions in classical error-in-variables and incidental parameter models*, Austral. J. Statist., **39**, 325–344.

[9] L. J. Gleser (1990), *Improvement of the naive estimation in nonlinear errors-in-variables regression*, Statistical Analysis of Measurement Error Models and Application (P. J. Brown and W. A. Fuller eds.), Contemporary Mathematics, Vol. 112, 99–114.

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

[11] P. Huber (1981), *Robust Statistics*, Wiley, New York.
[12] J. T. Kuha and J. Temple (2003), *Covariate measurement error in quadratic regression*, Internat. Statist. Review, **71**, 131–150.
[13] A. Kukush and H. Schneeweiss (2000), *A Comparison of Asymptotic Covariance Matrices of Adjusted Least Squares and Structural Least Squares in Error Ridden Polynomial Regression*, Discussion Paper 218, Sonderforschungsbereich 386, Univ. of Munich.
[14] A. Kukush, H. Schneeweiss, and R. Wolf (2004), *Three Estimators in the Poisson Regression Model with Measurement Errors*, Statist. Papers, **45**, 351–368.
[15] A. Kukush, H. Schneeweiss, and R. Wolf (2005), *Relative efficiency of three estimators in a polynomial regression with measurement errors*, J. Statist. Plann. Inference, **127**, 179–203.
[16] P. McCullagh and J. A. Nelder (1989), *Generalized Linear Models*, 2nd ed., Chapman and Hall, London.
[17] T. Nakamura (1990), *Corrected score function for errors-in-variables models*, Biometrika **77**, 127–137.
[18] J. Pfanzagl (1969), *On the measurability and consistency of minimum contrast estimates*, Metrika, **14**, 249–273.
[19] M. Schervish (1995), *Theory of Statistics*, Springer, New York.
[20] H. Schneeweiss and H. J. Mittag (1986), *Lineare Modelle mit fehlerbehafteten Daten*, Physica-Verlag, Heidelberg.
[21] H. Schneeweiss and T. Nittner (2001), *Estimating a polynomial regression with measurement errors in the structural and in the functional case – a comparison.* In: *Data Analysis from Statistical Foundations. A Festschrift in Honour of the 75th Birthday of D. A. S. Fraser*, M. Sadeh (ed.), Nova Science, New York, pp. 195–205.
[22] S. Shklyar and H. Schneeweiss (2002), *A Comparison of Asymptotic Covariance Matrices of Three Consistent Estimators in the Poisson Regression Model with Measurement Errors*, Discussion Paper 283, Sonderforschungsbereich 386, Univ. of Munich; to appear in J. Mulivariate Anal.
[23] C. J. Small and J. Wang (1993), *Numerical Methods of Nonlinear Equations*, Clarendon Press, Oxford.
[24] L. A. Stefanski (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.
[25] M. Thamerus (1998), *Different nonlinear regression models with incorrectly observed covariates.* In: *Econometrics in Theory and Practice, Festschrift for Hans Schneeweiss*, R. Galata and H. Küchenhoff (eds.), Physica, Heidelberg–New York.
[26] T. Wansbeek and E. Meijer (2000), *Measurement Error and Latent Variables in Econometrics*, Elsevier, Amsterdam.

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