

# Quasi Score is more efficient than Corrected Score in a polynomial measurement error model

Sergiy Shklyar · Hans Schneeweiss ·  
Alexander Kukush

Received: 7 October 2005  
© Springer-Verlag 2006

**Abstract** We consider a polynomial regression model, where the covariate is measured with Gaussian errors. The measurement error variance is supposed to be known. The covariate is normally distributed with known mean and variance. Quasi score (QS) and corrected score (CS) are two consistent estimation methods, where the first makes use of the distribution of the covariate (structural method), while the latter does not (functional method). It may therefore be surmised that the former method is (asymptotically) more efficient than the latter one. This can, indeed, be proved for the regression parameters. We do this by introducing a third, so-called simple score (SS), estimator, the efficiency of which turns out to be intermediate between QS and CS. When one includes structural and functional estimators for the variance of the error in the equation, SS is still more efficient than CS. When the mean and variance of the covariate are not known and have to be estimated as well, one can still maintain that QS is more efficient than SS for the regression parameters.

**Keywords** Quasi score · Corrected score · Polynomial model · Measurement errors · Efficiency · Structural methods · Functional methods

---

S. Shklyar · A. Kukush  
Kiev National Taras Shevchenko University,  
Vladimirskaia st. 60, 01033, Kiev, Ukraine  
e-mail: shklyar@univ.kiev.ua

H. Schneeweiss (✉)  
University of Munich, Akademiestr. 1,  
80799 Munich, Germany  
e-mail: schneew@stat.uni-muenchen.de

A. Kukush  
e-mail: alexander\_kukush@univ.kiev.ua

## 1 Introduction

We consider a polynomial regression model:

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

with additive measurement errors  $\delta$  in the latent covariate  $\xi$ :

$$x = \xi + \delta,$$

$x$  being the observable surrogate for  $\xi$ . We assume  $\delta$  to be independent of  $\xi$  (classical measurement error model). We also assume

$$\mathbb{E}(\epsilon|\xi, \delta) = 0, \quad \mathbb{E}(\epsilon^2|\xi, \delta) = \sigma_\epsilon^2.$$

In addition, we often make use of the assumption

$$(\mathbf{N}) \quad \delta \sim N(0, \sigma_\delta^2), \quad \xi \sim N(\mu_\xi, \sigma_\xi^2).$$

For a review of the polynomial measurement error model see Cheng and Schneeweiss (2002), see also Cheng and Schneeweiss (1998). The quadratic model has been applied in practice (Fuller 1987; Kuha and Temple 2003).

The problem is to estimate the parameter vector  $\beta = (\beta_0, \beta_1, \dots, \beta_k)^\top$  consistently given an i.i.d. sample of observations  $(x_i, y_i), i = 1, \dots, n$ . It is well-known that replacing  $\xi$  with  $x$  and applying least square leads to an inconsistent estimator (the *naive* estimator). For constructing consistent estimators we rely on additional pieces of information. Here we assume that the distributions of  $\xi$  and  $\delta$  are given. Under **(N)** this means that  $\sigma_\delta^2$ ,  $\mu_\xi$ , and  $\sigma_\xi^2$  are known. This is a simplifying assumption, which in practice is met only approximately.  $\sigma_\delta^2$  can often be estimated by replicated measurements or with the help of validation data;  $\mu_\xi$  and  $\sigma_\xi^2$  can be estimated from the data  $x_i$  alone:

$$\hat{\mu}_\xi = \bar{x}, \quad \hat{\sigma}_\xi^2 = s_x^2 - \sigma_\delta^2.$$

There are two major estimation methods that can be applied to this model, apart from maximum likelihood (ML), which, however, is rather complex. Both construct unbiased (vector-valued) estimating functions  $\psi(y, x; b)$ , or shorter  $\psi(b)$ , such that  $\mathbb{E}[\psi(b)] = 0$  if, and only if,  $b = \beta$  and compute the estimator  $\hat{\beta}$  as the solution to the equation

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

Under some regularity conditions, the unbiasedness of the estimating function guarantees the consistency of the estimator.

The *quasi score* (QS) estimator uses the distribution of  $\xi$  in constructing  $\psi$  and is therefore a so-called structural method (Carroll et al. 1995). In Kukush et al. (2002) and in Kukush and Schneeweiss (2005), it is called “structural quasi score (SQS)”. The *corrected score* (CS) estimator does not use the distribution of  $\xi$ . It is a so-called functional method (Chan and Mak 1985; Stefanski 1989; Cheng and Schneeweiss 1998); in the latter, CS is called “adjusted least squares (ALS)”.

CS has the advantage that it is insensitive with respect to possible misspecifications of the regressor distribution, whereas QS is in danger of a bias (Schneeweiss and Cheng 2006). On the other hand, QS might be more efficient, as it uses the information inherent in the regressor distribution, which CS does not use. However, this is by no means obvious, as QS is not ML. Nevertheless, we can prove that, indeed, QS is more efficient than CS in the sense that its asymptotic covariance matrix (ACM)  $\Sigma_Q$  is smaller, in the Loewner sense, than the ACM  $\Sigma_C$ .

This result is proved with the help of an intermediate estimator, the so-called *simple score* (SS) estimator, which is a simplified version of the QS estimator. SS can be easier compared both to QS and to CS and it is also much easier to compute than QS. The inequality relation  $\Sigma_Q \leq \Sigma_C$  is proved without assumption (N). We need (N) only to prove the corresponding strict order relation.

Shklyar and Schneeweiss (2005) have proved a similar result for the log-linear Poisson model, but with different arguments. They were able to compute the ACMs of CS and SS explicitly and they could compare them directly. Here we do not have explicit formulas for the ACMs and their comparison has to be accomplished in an indirect way, see section 4.

We will also extend our investigations to include the estimation of  $\sigma_\epsilon^2$  and will show that the efficiency comparison, at least for CS and SS, also holds with respect to this parameter.

Our results partially depend on the assumption that the nuisance parameters of the distribution of  $\xi$  are known. If they have to be estimated along with  $\beta$  and  $\sigma_\epsilon^2$ , the superiority of QS over CS is not clear, although it can still be observed in many cases, see the simulation study in Wolf (2004). Nevertheless, one can at least show that QS is more efficient than SS even in the presence of nuisance parameters.

In section 2, we introduce the unbiased estimating functions for CS, QS, and SS. Section 3 has some remarks on the naive estimator. Section 4 has our main results: the ACMs of the CS, QS, and SS estimators and their comparison to each other. In sections 5 and 6 we extend our investigation to the estimation of  $\sigma_\epsilon^2$ . Section 7 studies the effect of the estimation of the nuisance parameters

of the distribution of  $\xi$ . In section 8 we study the multiple polynomial model. Section 9 has some concluding remarks. Proofs are transferred to an appendix.

## 2 Unbiased estimating (score) functions for $\beta$

We write the polynomial model in matrix form

$$y = \zeta^\top \beta + \epsilon$$

with  $\zeta^\top = (1, \xi, \dots, \xi^k)$ . We assume the distribution of  $\delta$  and  $\xi$  to be known.

We define a vector  $\mu = \mu(x)$  as the conditional mean

$$\mu = \mathbb{E}(\zeta|x).$$

Under **(N)** the  $r$ th component of  $\mu$  is a polynomial in  $x$  of degree  $r - 1$ , see Kukush et al. (2005). A corresponding matrix  $M = M(x)$  is defined by

$$M = \mathbb{E}(\zeta\zeta^\top|x).$$

We also define a vector  $t = t(x)$  such that

$$\mathbb{E}(t|\xi) = \zeta.$$

$t$  is uniquely defined a.s. The  $r$ th component is a polynomial in  $x$  of degree  $r - 1$  (Cheng and Schneeweiss 1998). Finally we define the  $(k + 1) \times (k + 1)$  matrix  $T = T(x)$  such that

$$\mathbb{E}(T|\xi) = \zeta\zeta^\top.$$

$\mu$ ,  $t$ , and  $T$  can be constructed from the data  $x$ . Indeed, if  $z$  is defined as  $z = (1, x, \dots, x^k)^\top$ , then  $t$  and  $\mu$  are linear transforms of  $z$ , the latter under **(N)**:

$$\begin{aligned} t &= K_1 z \\ \mu &= K_2 z, \end{aligned}$$

where  $K_1$  and  $K_2$  are lower triangular matrices with positive elements on the diagonal (Kukush et al. 2005).

The estimating (score) functions for the CS, SS, and QS estimators of  $\beta$  are now given, respectively, by (for CS and QS see Kukush et al. (2005), where, however, CS and QS are called ALS and SLS, respectively)

$$\begin{aligned} \psi_{\beta C}(b) &= ty - Tb \\ \psi_{\beta S}(b) &= ty - t\mu^\top b \\ \psi_{\beta Q}(b) &= v(b)^{-1}(\mu y - \mu\mu^\top b), \end{aligned}$$

where  $v(\beta) = v(x, \beta, \sigma_\epsilon^2)$  is the conditional variance  $\mathbb{V}(y|x)$ :

$$v(\beta) = \sigma_\epsilon^2 + \beta^\top (M - \mu\mu^\top)\beta,$$

(Thamerus 1998). Here we implicitly assume that  $\sigma_\epsilon^2$  is given, but see Section 5. In order to be sure that the QS estimator is consistent, we also need the “technical” assumption that  $\beta$  is restricted to an inner point of a given compact set.

These estimating functions imply the following CS and SS estimators, respectively,

$$\begin{aligned} \hat{\beta}_C &= \overline{T}^{-1} \overline{ty} \\ \hat{\beta}_S &= \overline{t\mu}^\top{}^{-1} \overline{ty}, \end{aligned}$$

where the bar denotes averaging over the sample data  $(x_i, y_i), i = 1, \dots, n$ . As under  $(\mathbf{N})$   $t = K_1 K_2^{-1} \mu$ ,  $\hat{\beta}_S$  can also be written as

$$\hat{\beta}_S = \overline{\mu\mu}^\top{}^{-1} \overline{\mu y},$$

which corresponds to the estimating function  $\psi_{\beta S}^*(b) = \mu y - \mu\mu^\top b$ , equivalent to  $\psi_{\beta S}(b)$ .  $\hat{\beta}_S$  is thus seen to be the LS estimator in the (homoscedastic) linear regression

$$y = \mu^\top \beta + u.$$

For the QS estimator the same linear regression can be used but with a heteroscedastic error term  $u$  with a variance equal to  $v(\beta)$ . An iteratively reweighted least squares method may be applied in order to compute the QS estimate of  $\beta$  (Kukush et al. 2001; Kukush and Schneeweiss 2005).

In the following, when we use estimating functions without the subscript  $C, S$ , or  $Q$ , we refer to any of the three types of estimating functions. Expectations are always meant with respect to the true model parameters.

The following lemma shows that the estimating functions are indeed unbiased estimating functions.

**Lemma 1** *For the CS, SS, and QS estimating functions, for the latter two under  $(\mathbf{N})$ , we have*

$$\mathbb{E}[\psi_\beta(b)] = 0 \quad \text{if, and only if, } b = \beta.$$

### 3 Naive estimation

The naive estimator  $\hat{\beta}_N$  is based on the estimating function

$$\psi_{\beta N}(b) = zy - z z^\top b.$$

$\hat{\beta}_N$  is inconsistent with  $\lim \hat{\beta}_N = \beta^* \neq \beta$ . To find  $\beta^*$  one has to solve the equation

$$\mathbb{E}[\psi_{\beta N}(\beta^*)] = 0.$$

But, under **(N)**,

$$\begin{aligned}\mathbb{E}[\psi_{\beta N}(\beta^*)] &= K_1^{-1} \mathbb{E} \left( t y - t \mu^\top K_2^{-\top} \beta^* \right) \\ &= K_1^{-1} \mathbb{E} \left[ \psi_{\beta S}(K_2^{-\top} \beta^*) \right].\end{aligned}$$

Thus

$$\beta^* = K_2^\top \beta.$$

A consistent estimator is constructed from  $\hat{\beta}_N$  by multiplying  $\hat{\beta}_N$  with  $K_2^{-\top}$ . But as  $\psi_{\beta N}(b)$  is equivalent to  $t y - t \mu^\top K_2^{-\top} b$ , which is  $\psi_{\beta S}(K_2^{-\top} b)$ , this estimator is just  $\hat{\beta}_S$ :

$$\hat{\beta}_S = K_2^{-\top} \hat{\beta}_N.$$

Under **(N)**, the SS estimator is thus seen to be the bias-corrected naive estimator.

#### 4 Comparing ACMs of $\hat{\beta}$

Estimators of  $\beta$  that are based on an unbiased estimating function  $\psi$  are asymptotically normal under general regularity assumptions:

$$\sqrt{n}(\hat{\beta} - \beta) \sim N(0, \Sigma_\beta),$$

where  $\Sigma_\beta$  is given by the sandwich formula

$$\Sigma_\beta = A_\beta^{-1} B_\beta A_\beta^{-\top}$$

with

$$\begin{aligned}A_\beta &= -\mathbb{E} \left( \frac{\partial \psi_\beta}{\partial \beta^\top} \right) \\ B_\beta &= \mathbb{E}[\psi_\beta \psi_\beta^\top] \\ \psi_\beta &= \psi_\beta(\beta).\end{aligned}$$

We compare the two factors  $A_\beta$  and  $B_\beta$  for QS and SS. We define  $\beta_{(0)}$  to be the  $k$ -dimensional vector that is derived from  $\beta$  by deleting its first element  $\beta_0$ , i.e.:  $\beta = (\beta_0, \beta_{(0)}^\top)^\top$ .

**Lemma 2**

$$A_{\beta C} = A_{\beta S} = \mathbb{E}(\zeta \zeta^\top).$$

**Lemma 3**

$$B_{\beta C} \geq B_{\beta S}$$

with equality if  $\beta_{(0)} = 0$ . Under  $(\mathbf{N})$ , if  $\beta_{(0)} \neq 0$ ,

$$B_{\beta C} > B_{\beta S}.$$

Lemmas 2 and 3 imply the following theorem.

**Theorem 1**

$$\Sigma_{\beta C} \geq \Sigma_{\beta S}$$

with equality if  $\beta_{(0)} = 0$ . Under  $(\mathbf{N})$ , if  $\beta_{(0)} \neq 0$ ,

$$\Sigma_{\beta C} > \Sigma_{\beta S}.$$

As to QS, we can rely on a general theorem of Heyde (1997, Theorem 2.3). According to this theorem  $\psi_{\beta Q}$  is optimal in the class of all linear-in- $y$  estimating functions  $\psi$  with  $\mathbb{E}(\psi|x) = 0$ . As  $\psi_{\beta S}$  belongs to this class, we can state the following theorem

**Theorem 2**

$$\Sigma_{\beta S} \geq \Sigma_{\beta Q}.$$

Under  $(\mathbf{N})$ ,  $\Sigma_{\beta S} = \Sigma_{\beta Q}$  if  $\beta_{(0)} = 0$ ; if  $\beta_{(0)} \neq 0$ ,  $\Sigma_{\beta S} > \Sigma_{\beta Q}$  if, and only if,  $r > \frac{k}{2} + 1$ , where  $r = \max\{j : \beta_j \neq 0\}$ .

As a corollary we have that in a quadratic model ( $k = 2$ ) the matrix  $\Sigma_{\beta S} - \Sigma_{\beta Q}$  is always singular, while in a polynomial model with ( $k > 2$ ) and with  $\beta_k \neq 0$  the matrix  $\Sigma_{\beta S} - \Sigma_{\beta Q}$  is nonsingular. It may also be noted that in a linear model  $v$  is a constant and thus  $\hat{\beta}_S$  and  $\hat{\beta}_Q$  coincide and consequently  $\Sigma_{\beta S} = \Sigma_{\beta Q}$  in this case even if  $\beta_{(0)} \neq 0$ , see Note 4 in the appendix.

The above argument cannot be extended to  $\Sigma_{\beta C}$  because  $\mathbb{E}(\psi_{\beta C}|x) \neq 0$ . Nevertheless, Theorems 1 and 2 imply

$$\Sigma_{\beta C} \geq \Sigma_{\beta Q},$$

where, under  $(\mathbf{N})$ , we have equality if  $\beta_{(0)} = 0$  and strict inequality if  $\beta_{(0)} \neq 0$ .

*Remark* By applying a criterion of Heyde (1997, Theorem 2.1), one can prove the inequality  $\Sigma_{\beta C} \geq \Sigma_{\beta Q}$  without the need to introduce the intermediate SS estimator. However, Heyde's criterion per se does not offer any way to investigate the cases of equality and strict inequality. This is best done with the help of the SS estimator. In addition, the SS estimator has its own merits: it is much simpler to compute than the QS estimator and it is more robust than the latter in the sense that one does not need to know the exact form of the variance function  $v$ , which in practice may deviate from the form that we used in our model. Thus Theorems 1 and 2 may be of interest for their own sake and not just as intermediate results for the main result  $\Sigma_{\beta C} \geq \Sigma_{\beta Q}$ .

## 5 Unbiased estimating functions for $\sigma_\epsilon^2$

We can also estimate  $\sigma_\epsilon^2$  by using the following estimating functions for CS and SS, respectively, together with the corresponding estimating functions for  $\beta$ :

$$\begin{aligned}\psi_{\epsilon C}(s^2, b) &= y^2 - b^\top T b - s^2 \\ \psi_{\epsilon S}(s^2, b) &= y^2 - b^\top M b - s^2.\end{aligned}$$

The last estimating function can also be used in connection with the QS estimator of  $\beta$  (i.e.,  $\psi_{\epsilon S} = \psi_{\epsilon Q}$ ). However, in this case,  $\psi_{\beta Q}$  and  $\psi_{\epsilon Q}$  have to be used together in order to estimate  $\beta$  and  $\sigma_\epsilon^2$  simultaneously:

$$\begin{aligned}\psi_{\beta Q}(b, s^2) &= v(b, s^2)^{-1}(\mu y - \mu \mu^\top b) \\ \psi_{\epsilon Q}(s^2, b) &= y^2 - b^\top M b - s^2.\end{aligned}$$

where

$$v(b, s^2) = s^2 + b^\top (M - \mu \mu^\top) b.$$

This does not change the unbiasedness property of  $\psi_{\beta Q}$ , as one can see from the proof of Lemma 1. We shall see in section 6 that the ACM of  $\hat{\beta}_Q$  does not change either.

By contrast, the CS and SS estimation of  $\beta$  and  $\sigma_\epsilon^2$  can be accomplished in two steps: first  $\beta$  is estimated using  $\psi_{\beta C}$  or  $\psi_{\beta S}$ , respectively, and then  $\sigma_\epsilon^2$  is estimated using  $\psi_{\epsilon C}$  or  $\psi_{\epsilon S}$ , respectively, and the estimate of  $\beta$  from the first step. For CS and SS, the estimation of  $\beta$  does not depend on the estimation of  $\sigma_\epsilon^2$ , and consequently their ACMs do not change whatever estimation method for  $\sigma_\epsilon^2$  is used.

The following lemma, together with Lemma 1, shows that  $\psi_\epsilon$  is, indeed, unbiased for all three estimation methods.

### Lemma 4

$$\mathbb{E}[\psi_\epsilon(s^2, \beta)] = 0 \quad \text{if, and only if, } s^2 = \sigma_\epsilon^2.$$



### 6 Comparing asymptotic variances of $\hat{\sigma}_\epsilon^2$

Let  $\alpha = (\beta^\top, \sigma_\epsilon^2)^\top$  and  $a = (b^\top, s^2)^\top$ . The estimating function for  $\alpha$  is just the compound of the estimating functions for  $\beta$  and  $\sigma_\epsilon^2$ :

$$\psi_\alpha(a) = \left( \psi_\beta^\top(b, s^2), \psi_\epsilon(s^2, b) \right)^\top.$$

The ACM of  $\hat{\alpha}$  is given by the sandwich formula

$$\Sigma_\alpha = A_\alpha^{-1} B_\alpha A_\alpha^{-\top}$$

with

$$\begin{aligned} A_\alpha &= -\mathbb{E} \left( \frac{\partial \psi_\alpha}{\partial \alpha^\top} \right) \\ B_\alpha &= \mathbb{E} \left( \psi_\alpha \psi_\alpha^\top \right) \\ \psi_\alpha &= \psi_\alpha(\alpha). \end{aligned}$$

We investigate the structure of  $A_\alpha$ . For CS and SS,  $\psi_\beta$  does not depend on  $\sigma_\epsilon^2$ . Therefore  $\partial \psi_\beta / \partial \sigma_\epsilon^2 = 0$ . For QS,  $\psi_{\beta Q}$  does depend on  $\sigma_\epsilon^2$ . Nevertheless,  $\mathbb{E}(\partial \psi_{\beta Q} / \partial \sigma_\epsilon^2) = 0$  because  $\mathbb{E}[(\mu y - \mu \mu^\top \beta) | x] = 0$ . Finally  $-\partial \psi_\epsilon / \partial \sigma_\epsilon^2 = 1$ . Therefore, for all three estimators,  $A_\alpha$  has the form

$$A_\alpha = \begin{pmatrix} A_\beta & 0 \\ A_{\epsilon\beta} & 1 \end{pmatrix},$$

where  $A_{\epsilon\beta} := -\mathbb{E} \left( \frac{\partial \psi_\epsilon}{\partial \beta^\top} \right)$  and  $\psi_\epsilon = \psi_\epsilon(\sigma_\epsilon^2, \beta)$ . Its inverse is

$$A_\alpha^{-1} = \begin{pmatrix} A_\beta^{-1} & 0 \\ -A_{\epsilon\beta} A_\beta^{-1} & 1 \end{pmatrix}.$$

Similarly  $B_\alpha$  has the form

$$B_\alpha = \begin{pmatrix} B_\beta & B_{\beta\epsilon} \\ B_{\epsilon\beta} & B_\epsilon \end{pmatrix},$$

with  $B_{\beta\epsilon} = \mathbb{E}(\psi_\beta \psi_\epsilon) = B_{\epsilon\beta}^\top$ ,  $B_\epsilon = \mathbb{E}(\psi_\epsilon^2)$ . The ACM of  $\hat{\alpha}$  can be partitioned in a similar way:

$$\Sigma_\alpha = \begin{pmatrix} \Sigma_\beta & \Sigma_{\beta\epsilon} \\ \Sigma_{\epsilon\beta} & \Sigma_\epsilon \end{pmatrix},$$

where  $\Sigma_\epsilon$  is the asymptotic variance of  $\hat{\sigma}_\epsilon^2$ .

An easy computation shows that  $\Sigma_\beta = A_\beta^{-1} B_\beta A_\beta^{-\top}$  just as in section 4. Thus  $\hat{\beta}$  (in particular  $\hat{\beta}_Q$ ) has the same ACM regardless of whether  $\sigma_\epsilon^2$  is known or has to be estimated. By the same arguments this result holds also true if any other estimating function for  $\sigma_\epsilon^2$  is employed. One can even show that  $\Sigma_\beta$  does not change if a consistent estimator of  $\sigma_\epsilon^2$  is used, which does not necessarily stem from an estimating function.

We can compare  $\Sigma_{\alpha C}$  to  $\Sigma_{\alpha S}$  and thus extend Theorem 1.

### Theorem 3

$$\Sigma_{\alpha C} \geq \Sigma_{\alpha S}$$

with equality if  $\beta_{(0)} = 0$ . Conversely, under **(N)**,  $\Sigma_{\alpha C} = \Sigma_{\alpha S}$  implies  $\beta_{(0)} = 0$ .

It follows that, in particular,  $\Sigma_{\epsilon C} \geq \Sigma_{\epsilon S}$ . Under **(N)**, we can also show that  $\Sigma_{\epsilon C} > \Sigma_{\epsilon S}$  if  $\beta_{(0)} \neq 0$ . We cannot maintain that  $\Sigma_{\alpha C} > \Sigma_{\alpha S}$  if  $\beta_{(0)} \neq 0$ . Indeed, there are counterexamples in the case of a linear model ( $k = 1$ ), see appendix after the proof of Theorem 3.

## 7 Nuisance parameters

Structural methods, like SS and QS, make use of the conditional distribution of  $\xi$  given  $x$ . Up to now this distribution was assumed to be known. We now relax this assumption and instead assume that this distribution is known only up to a vector  $\gamma$  of nuisance parameters. Under **(N)** and if  $\sigma_\delta^2$  is known,  $\gamma$  is a parameter vector characterizing the normal distribution of  $x$ , e.g.,  $\gamma = (\mu_x, \sigma_x^2)^\top$ . In this case,  $\gamma$  can be consistently estimated from the data  $x_i$  alone. We assume this to be so in general.

The structural estimators of  $\beta$  (and also of  $\sigma_\epsilon^2$ ) are then constructed in the same way as before except that  $\gamma$  is replaced with its preestimated value  $\hat{\gamma}$ . That, however, will affect the asymptotic covariance matrix of  $\hat{\beta}$ , which will be given by the same sandwich formula as before plus an additional term that stems from the estimation of  $\gamma$ . It turns out that, under **(N)**, this additional term is the same for SS and QS (and for other structural methods as well). Therefore the efficiency order of QS over SS that we found before is preserved in the presence of nuisance parameters.

We start from a more general model that comprises the polynomial model as a special case. The model is given by conditional mean and variance functions of  $y$  given  $x$ .

Let

$$m = m(x, \beta, \gamma) = \mathbb{E}(y|x)$$

$$v = v(x, \beta, \varphi, \gamma) = \mathbb{V}(y|x) = \mathbb{E}[(y - m)^2|x].$$

Here  $\varphi$  is a dispersion parameter. In the polynomial model,  $\varphi = \sigma_\epsilon^2$ . We assume that  $\gamma$  can be estimated via an unbiased estimating function that involves only  $x$ . We thus have the following set of estimating functions for  $\theta := (\beta^\top, \varphi, \gamma^\top)^\top$  (for simplicity we take them as functions of the true parameters):

$$\begin{aligned} \psi_\beta &= (y - m)e \\ \psi_\varphi &= y^2 - m^2 - v \\ \psi_\gamma &= g(x, \gamma). \end{aligned}$$

Here  $e = e(x, \beta, \varphi, \gamma)$  is an arbitrary function that defines the estimation method used. For SS,  $e = \partial m / \partial \beta$ , and for QS,  $e = v^{-1} \partial m / \partial \beta$ . The function  $g$  is such that  $\mathbb{E}g = 0$ . Then,  $\psi_\theta := (\psi_\beta^\top, \psi_\varphi, \psi_\gamma^\top)^\top$  is an unbiased estimating function:  $\mathbb{E}(\psi_\theta) = 0$ .

We remark that instead of the function  $\psi_\varphi$  introduced above we may also use the related function  $\psi_\varphi^* = (y - m)^2 - v$  without having to change the result of this section.

Let  $\alpha = (\beta^\top, \varphi)^\top$  and  $\psi_\alpha = (\psi_\beta^\top, \psi_\varphi)^\top$ , so that  $\psi_\theta = (\psi_\alpha^\top, \psi_\gamma^\top)^\top$ . Then  $A_\theta := -\mathbb{E}(\partial \psi_\theta / \partial \theta^\top)$  can be partitioned as follows

$$A_\theta = -\mathbb{E} \begin{pmatrix} \frac{\partial \psi_\alpha}{\partial \alpha^\top} & \frac{\partial \psi_\alpha}{\partial \gamma^\top} \\ \frac{\partial \psi_\gamma}{\partial \alpha^\top} & \frac{\partial \psi_\gamma}{\partial \gamma^\top} \end{pmatrix} =: \begin{pmatrix} A_\alpha & A_{\alpha\gamma} \\ 0 & A_\gamma \end{pmatrix}.$$

Similarly  $B_\theta := \mathbb{E}(\psi_\theta \psi_\theta^\top)$  is partitioned as

$$B_\theta = \begin{pmatrix} B_\alpha & 0 \\ 0 & B_\gamma \end{pmatrix},$$

because  $\mathbb{E}(\psi_\alpha \psi_\gamma^\top | x) = \mathbb{E}(\psi_\alpha | x) \psi_\gamma^\top = 0$ .

Now

$$A_\theta^{-1} = \begin{pmatrix} A_\alpha^{-1} & -A_\alpha^{-1} A_{\alpha\gamma} A_\gamma^{-1} \\ 0 & A_\gamma^{-1} \end{pmatrix},$$

and

$$\Sigma_\theta = A_\theta^{-1} B_\theta A_\theta^{-\top} = \begin{pmatrix} \Sigma_\alpha & \Sigma_{\alpha\gamma} \\ \Sigma_{\gamma\alpha} & \Sigma_\gamma \end{pmatrix},$$

with

$$\begin{aligned} \Sigma_\alpha &= A_\alpha^{-1} (B_\alpha + A_{\alpha\gamma} \Sigma_\gamma A_{\alpha\gamma}^\top) A_\alpha^{-\top} \\ \Sigma_{\alpha\gamma} &= -A_\alpha^{-1} A_{\alpha\gamma} \Sigma_\gamma = \Sigma_{\gamma\alpha}^\top \\ \Sigma_\gamma &= A_\gamma^{-1} B_\gamma A_\gamma^{-\top}. \end{aligned}$$

A further decomposition of  $\psi_\alpha$  into  $\psi_\beta$  and  $\psi_\varphi$  yields

$$\begin{aligned} A_\alpha &= \begin{pmatrix} A_\beta & 0 \\ A_{\varphi\beta} & A_\varphi \end{pmatrix} \\ B_\alpha &= \begin{pmatrix} B_\beta & B_{\beta\varphi} \\ B_{\varphi\beta} & B_\varphi \end{pmatrix}. \end{aligned}$$

We also have

$$A_{\alpha\gamma} = \begin{pmatrix} A_{\beta\gamma} \\ A_{\varphi\gamma} \end{pmatrix}.$$

We then can compute the two terms of  $\Sigma_\alpha$  as

$$A_\alpha^{-1} B_\alpha A_\alpha^{-\top} = \begin{pmatrix} A_\beta^{-1} B_\beta A_\beta^{-\top} & * \\ * & * \end{pmatrix}$$

$$A_\alpha^{-1} A_{\alpha\gamma} \Sigma_\gamma A_{\alpha\gamma}^\top A_\alpha^{-\top} = \begin{pmatrix} A_\beta^{-1} A_{\beta\gamma} \Sigma_\gamma A_{\beta\gamma}^\top A_\beta^{-\top} & * \\ * & * \end{pmatrix},$$

where  $*$  denotes submatrices of no further interest. We finally get  $\Sigma_\beta$  as the upper left corner of  $\Sigma_\alpha$ :

$$\Sigma_\beta = A_\beta^{-1} B_\beta A_\beta^{-\top} + A_\beta^{-1} A_{\beta\gamma} \Sigma_\gamma A_{\beta\gamma}^\top A_\beta^{-\top},$$

where the first term is the ACM of  $\hat{\beta}$  if  $\gamma$  were known and the second term is the correction due to the estimation of  $\gamma$ .

Returning to our original polynomial model, we see that

$$m = \mu^\top \beta$$

$$v = \sigma_\epsilon^2 + \beta^\top (M - \mu \mu^\top) \beta.$$

We want to show that the correction for estimation of  $\gamma$  in the formula of  $\Sigma_\beta$  is independent of  $e$  (i.e., independent of the estimation method used). For this we evaluate, under  $(\mathbf{N})$ ,

$$A_\beta = -\mathbb{E} \left( \frac{\partial \psi_\beta}{\partial \beta^\top} \right) = \mathbb{E} \left( e \frac{\partial m}{\partial \beta^\top} \right) = \mathbb{E} \left( e \mu^\top \right)$$

$$A_{\beta\gamma} = -\mathbb{E} \left( \frac{\partial \psi_\beta}{\partial \gamma^\top} \right) = \mathbb{E} \left( e \frac{\partial m}{\partial \gamma^\top} \right) = \mathbb{E} \left( e \beta^\top \frac{\partial \mu}{\partial \gamma^\top} \right).$$

Now with  $\mu = K_2 z$ , we find

$$\begin{aligned} \beta^\top \frac{\partial \mu}{\partial \gamma^\top} &= z^\top \frac{\partial K_2^\top \beta}{\partial \gamma^\top} \\ &= \mu^\top K_2^{-\top} \frac{\partial K_2^\top \beta}{\partial \gamma^\top} \\ &=: \mu^\top G, \end{aligned}$$

where  $G$  is a constant matrix that does not depend on  $e$ . Thus

$$A_{\beta\gamma} = A_\beta G$$

and

$$\Sigma_\beta = A_\beta^{-1} B_\beta A_\beta^{-\top} + G \Sigma_\gamma G^\top,$$

which proves that the correction term due to estimation of  $\gamma$  is, indeed, the same for all estimation methods  $e$ . In particular, it is the same for SS and QS.

This shows that Theorem 2 is still valid, even in the presence of nuisance parameters.

### 8 Multiple polynomials

The previous results can all be extended to the case of a multiple polynomial model.

We have to replace the random scalar variables  $\xi, x$ , and  $\delta$  with random vector variables  $\xi = (\xi_1, \dots, \xi_q)^\top$ ,  $x = (x_1, \dots, x_q)^\top$ , and  $\delta = (\delta_1, \dots, \delta_q)^\top$ , respectively, such that  $x = \xi + \delta$ , as before. The covariance matrix  $\Sigma_\xi$  of  $\xi$  is supposed to be nonsingular.

Consider the power term  $\xi_1^{k_1} \dots \xi_q^{k_q}$  with  $k_i$  being nonnegative integers. Let  $k_1 + \dots + k_q$  be the degree of this power term. A (multiple) polynomial in  $\xi$  of degree  $k$  is a linear combination of all such power terms up to the maximum degree  $k$ . Let  $\zeta$  be the vector consisting of all power terms up to the degree  $k$  arranged by increasing degree, where all terms with the same degree are arranged in an arbitrary way. Then

$$y = \zeta^\top \beta + \epsilon$$

is a (multiple) polynomial model with the unknown parameter vector  $\beta$ .

The vectors  $t$  and  $\mu$  and the matrices  $T$  and  $M$  are defined in the same way as in section 2. It turns out that a typical element of  $t$  is of the form

$$t_{k_1 \dots k_q} = x_1^{k_1} \dots x_q^{k_q} + p_{k_1 \dots k_q}(x),$$

where  $p_{k_1 \dots k_q}(x)$  is a polynomial in  $x$  of lower degree than  $k_1 + \dots + k_q$ . Let  $z$  be the vector defined just as  $\zeta$  but with the  $\xi_i$  replaced by  $x_i$ . Then

$$t = K_1 z$$

with  $K_1$  being a triangular matrix with ones on the diagonal.

Under  $(\mathbf{N})$ , the conditional distribution of  $\xi$  given  $x$  is  $N(\tilde{\mu}, \tilde{\Sigma})$  with

$$\tilde{\mu} = \mathbb{E}(\xi|x) = (\tilde{\mu}_1, \dots, \tilde{\mu}_q)^\top = Ax + b,$$

where  $A = \Sigma_\xi \Sigma_x^{-1}$  is nonsingular and  $b$  is a vector of constants. It follows that a typical element of  $\mu$  is of the form

$$\mu_{k_1 \dots k_q} = \mathbb{E}[(\xi_1 - \tilde{\mu}_1 + \tilde{\mu}_1)^{k_1} \dots (\xi_q - \tilde{\mu}_q + \tilde{\mu}_q)^{k_q} | x] = \tilde{\mu}_1^{k_1} \dots \tilde{\mu}_q^{k_q} + q_{k_1 \dots k_q}(\tilde{\mu}),$$

where  $q_{k_1 \dots k_q}(\tilde{\mu})$  is a polynomial in  $\tilde{\mu}$  of lower degree than  $k_1 + \dots + k_q$ . Moreover,  $\tilde{\mu}_1^{k_1} \dots \tilde{\mu}_q^{k_q}$  is a polynomial in  $x$  of degree  $k_1 + \dots + k_q$ , and so is  $\mu_{k_1 \dots k_q}$ . Thus

$$\mu = K_2 z.$$

Conversely, as  $x = A^{-1}\tilde{\mu} + c$  (with  $c = -A^{-1}b$ ), any term  $x_1^{l_1} \dots x_q^{l_q}$  is a polynomial in  $\tilde{\mu}$  of degree  $l_1 + \dots + l_q$  and so  $z$  is a linear transform of  $\mu$ . Thus  $K_2$  is nonsingular.

The arguments and results of the preceding section can now be taken over to the multiple case.

### 9 Conclusion

We proved that the (structural) QS estimator of the regression parameter  $\beta$  of a polynomial measurement error model is more efficient than the (functional) CS estimator under the assumption that  $\xi \sim N(\mu_\xi, \sigma_\xi^2)$  and  $\delta \sim N(0, \sigma_\delta^2)$  with the parameters  $\mu_\xi, \sigma_\xi^2$ , and  $\sigma_\delta^2$  known. The proof was accomplished by introducing an intermediate structural estimator, the SS estimator, which is simpler than QS, but is strongly related to QS so that a comparison of their ACMs can be carried out. It is also related to CS, albeit in a different way, so that the ACMs of SS and CS can be compared as well.

Apart from its role as a tool to prove the superiority of QS over CS, the SS estimator has its own merits as an estimator that can be very simply computed just using ordinary least squares, albeit not in the original polynomial model but in a related multiple linear model.

We extended this result to the estimation of  $\sigma_\epsilon^2$  in so far as we showed that the SS estimator of  $\sigma_\epsilon^2$  is more efficient than the CS estimator.

It should be noted that the result on the efficiency comparison between CS and QS holds true under the assumption that the nuisance parameters  $\mu_\xi$  and  $\sigma_\xi^2$  are known. It is an open question whether this result still holds under unknown nuisance parameters to be estimated by the data  $x_i$ . We can, however, show that QS is more efficient than SS regardless of whether the nuisance parameters are known or unknown.

It should also be noted that despite its inferiority with regard to its relative efficiency, the CS estimator has the big advantage not to depend on the knowledge of the distribution of  $\xi$ . It is therefore robust with regard to any misspecification of this distribution. This is different for SS and QS, which do depend on the correct knowledge of the  $\xi$ -distribution and which show an asymptotic bias if that distribution has been misspecified.

Finally, let us note that for small  $\sigma_\delta^2$  the difference  $\Sigma_C - \Sigma_Q$  is of order  $\sigma_\delta^4$  (Kukush et al. 2005; Kukush and Schneeweiss 2005).

We conclude with a technical remark. The main results concerning the (weak) inequalities ( $\leq$ ) of covariance matrices as stated in Theorems 1 to 3 do not rely on the normality assumption (**N**). They even hold true under much more general assumptions:  $\xi$  and  $\delta$  need not be independent as long as we can assume that the *joint* distribution of  $\xi$  and  $\delta$  is known. Thus these results remain valid also under the Berkson variant of the measurement error model where instead of  $\delta$  being independent of  $\xi$ ,  $\delta$  is independent of  $x$ .

### Appendix: Proofs

**Lemma 1** Start with CS. We have

$$\begin{aligned} \mathbb{E}[\psi_{\beta C}(b)] &= \mathbb{E}[(t\zeta^\top \beta - Tb)] + \mathbb{E}(t\epsilon) \\ &= \mathbb{E}\mathbb{E}(t\zeta^\top \beta - Tb|\xi) + \mathbb{E}[t\mathbb{E}(\epsilon|x)] \\ &= \mathbb{E}[\zeta \zeta^\top (\beta - b)], \end{aligned}$$

which is zero iff  $b = \beta$  because  $\mathbb{E}\zeta \zeta^\top$  is p.d.

For SS, we have

$$\begin{aligned}\mathbb{E}[\psi_{\beta S}(b)] &= \mathbb{E}\left[t\mathbb{E}(\zeta^\top|x)\beta - t\mu^\top b\right] + \mathbb{E}(t\epsilon) \\ &= \mathbb{E}[t\mu^\top(\beta - b)],\end{aligned}$$

which is zero iff  $b = \beta$  because (under **(N)**)

$$t\mu^\top = K_1 z z^\top K_2^\top$$

and  $\mathbb{E}(z z^\top)$  is p. d. The proof for QS proceeds in a similar way as for SS.

**Note 1** It can be seen from the proof that not only is  $\mathbb{E}[\psi_{\beta S}(\beta)] = 0$  but even

$$\mathbb{E}[\psi_{\beta S}(\beta)|x] = 0.$$

The same is true for  $\psi_{\beta Q}$ , but not for  $\psi_{\beta C}$ .

**Note 2** Assumption **(N)** was only needed to assume the nonsingularity of  $\mathbb{E}(t\mu^\top)$ . If it can be assumed that  $\mathbb{E}(t\mu^\top)$  is nonsingular, assumption **(N)** is not needed.

**Lemma 2** First note that

$$\begin{aligned}\frac{\partial \psi_{\beta C}}{\partial \beta^\top} &= -T \\ \frac{\partial \psi_{\beta S}}{\partial \beta^\top} &= -t\mu^\top.\end{aligned}$$

Now

$$\begin{aligned}\mathbb{E}(T) &= \mathbb{E}[\mathbb{E}(T|\xi)] = \mathbb{E}\zeta\zeta^\top \\ \mathbb{E}(t\mu^\top) &= \mathbb{E}[t\mathbb{E}(\zeta^\top|x)] = \mathbb{E}[\mathbb{E}(t\zeta^\top|x)] \\ &= \mathbb{E}(t\zeta^\top) = \mathbb{E}[\mathbb{E}(t\zeta^\top|\xi)] \\ &= \mathbb{E}[\mathbb{E}(t|\xi)\zeta^\top] = \mathbb{E}\zeta\zeta^\top.\end{aligned}$$

Thus  $A_{\beta C} = A_{\beta S}$ .

**Lemma 3** As  $\mathbb{E}(\psi_{\beta S}|x) = 0$  (compare Note 1) and  $\psi_{\beta C} - \psi_{\beta S}$  is a function of  $x$  alone, therefore

$$\mathbb{E}\left[\psi_{\beta S}(\psi_{\beta C} - \psi_{\beta S})^\top\right] = \mathbb{E}\left[\mathbb{E}(\psi_{\beta S}|x)(\psi_{\beta C} - \psi_{\beta S})^\top\right] = 0.$$

It follows that

$$\mathbb{E}\left(\psi_{\beta C}\psi_{\beta C}^\top\right) = \mathbb{E}\left(\psi_{\beta S}\psi_{\beta S}^\top\right) + \mathbb{E}\left[(\psi_{\beta C} - \psi_{\beta S})(\psi_{\beta C} - \psi_{\beta S})^\top\right] \geq \mathbb{E}\left(\psi_{\beta S}\psi_{\beta S}^\top\right)$$



or

$$B_{\beta C} \geq B_{\beta S},$$

which proves the first part of the lemma.

By the definition of the estimating functions  $\psi_\beta$ , the difference of  $B_{\beta C}$  and  $B_{\beta S}$  is

$$B_{\beta C} - B_{\beta S} = \mathbb{E} \left[ \left( T - t\mu^\top \right) \beta \beta^\top \left( T - t\mu^\top \right)^\top \right].$$

Consider the  $(i, j)$ th element,  $i, j = 0, \dots, k$ , of  $T - t\mu^\top$ , which is

$$T_{ij} - t_i \mu_j = t_{i+j} - t_i \mu_j.$$

For  $j = 0$  this is 0 because  $\mu_0 = 1$ . Thus the first column of  $T - t\mu^\top$  is zero. Consequently,  $\beta_{(0)} = 0$  implies  $(T - t\mu^\top) \beta = 0$  and thus  $B_{\beta C} = B_{\beta S}$ .

To prove the last part of the lemma, we have to show [under  $(\mathbf{N})$ ] that, for any  $(k + 1)$ -dimensional vector  $a \neq 0$ ,  $a^\top (B_{\beta C} - B_{\beta S}) a > 0$  if  $\beta_{(0)} \neq 0$ . This is equivalent to

$$\mathbb{P} \left( a^\top \left( T - t\mu^\top \right) \beta \neq 0 \right) > 0$$

if  $\beta_{(0)} \neq 0$ . But, under  $(\mathbf{N})$ ,  $a^\top (T - t\mu^\top) \beta$  is a polynomial in  $x$ . So it suffices to show that the degree of this polynomial is greater than zero if  $\beta_{(0)} \neq 0$ .

Now, under  $(\mathbf{N})$ ,  $\mu_j$  is a polynomial of degree  $j$  in  $\mu_1$  with highest coefficient 1, and since

$$\mu_1 = \mu_x \frac{\sigma_\delta^2}{\sigma_x^2} + \rho x, \quad \rho := 1 - \frac{\sigma_\delta^2}{\sigma_x^2},$$

(Kukush et al. 2005),  $\mu_j$  is a polynomial in  $x$  of degree  $j$  with highest coefficient  $\rho^j$ . Moreover, as  $t_i$  is a polynomial in  $x$  of degree  $i$  with highest coefficient 1, we have that, for  $j \neq 0$ ,  $T_{ij} - t_i \mu_j$  is a polynomial in  $x$  of degree  $i + j$  with highest coefficient  $1 - \rho^j \neq 0$ .

Now, let  $\beta_{(0)} \neq 0$ . Then there exists an index  $r$ ,  $1 \leq r \leq k$ , such that  $\beta_r \neq 0$  but  $\beta_j = 0$  for  $j > r$ . It follows that  $(T - t\mu^\top) \beta$  is a  $(k + 1)$ -dimensional vector the elements of which are polynomials in  $x$  of degree  $r, r + 1, \dots, r + k$ , respectively. Consequently, for any  $a \neq 0$ ,  $a^\top (T - t\mu^\top) \beta$  is a polynomial of degree at least  $r$  and  $r > 0$ .

**Theorem 2** The first part of the theorem is a consequence of general quasi score theory (Heyde 1997). If we assume  $(\mathbf{N})$ , which we shall do for the rest of the proof, we can prove the first part of the theorem more directly by other means.

Using the estimating functions  $\psi_{\beta S}^*$  and  $\psi_{\beta Q}$ , one can show that

$$\begin{aligned} \Sigma_{\beta S} &= (\mathbb{E}[\mu\mu^\top])^{-1}\mathbb{E}(v\mu\mu^\top)(\mathbb{E}[\mu\mu^\top])^{-1} \\ \Sigma_{\beta Q} &= (\mathbb{E}[v^{-1}\mu\mu^\top])^{-1}. \end{aligned}$$

The first part of the theorem then follows from the matrix inequality

$$\mathbb{E}(vw w^\top) \geq (\mathbb{E}[v^{-1}w w^\top])^{-1},$$

where  $w$  is a random vector with  $\mathbb{E}(w w^\top) = I$ , see Appendix in Shklyar and Schneeweiss (2005).

If  $\beta_{(0)} = 0$ , then  $v = \sigma_\epsilon^2$  because  $M_{00} - \mu_0^2 = 1 - 1 = 0$ . Hence  $\Sigma_{\beta S} = \Sigma_{\beta Q}$ .

Now, suppose  $\beta_{(0)} \neq 0$ . Let  $r$  be the degree of the true polynomial,  $r \geq 1$ . Then  $v(\beta)$  is a polynomial in  $\mu_1$  of degree  $2r - 2$  because, by the definition of  $\mu_r$  (Kukush et al. 2005),  $\mu_r = \mu_1^r + p_{r-2}(\mu_1)$ , where  $p_{r-2}(\mu_1)$  is a polynomial in  $\mu_1$  of degree at most  $r - 2$ . According to the Appendix in Shklyar and Schneeweiss (2005), the matrix  $\Sigma_S - \Sigma_Q$  is singular iff there exist  $f_1, f_2 \in \mathbb{R}^{k+1} \setminus \{0\}$ , such that

$$f_1^\top \mu = v(\beta) f_2^\top \mu \quad \text{a.s.} \tag{*}$$

The left-hand side of (\*) is a polynomial in  $\mu_1$  of degree at most  $k$ , the right-hand side is of a degree in the interval  $[2r - 2, 2r - 2 + k]$ . Therefore (\*) implies  $k \geq 2r - 2$ .

Conversely, if  $k \geq 2r - 2$ , then  $v(\beta) = f_3^\top \mu$  for some  $f_3$ . Then (\*) holds true with  $f_1 = f_3$  and  $f_2 = (1, 0, \dots, 0)^\top$ .

**Lemma 4** First consider CS. We have

$$\begin{aligned} \mathbb{E}[\psi_{\epsilon C}(s^2, \beta)] &= \mathbb{E}(\beta^\top \zeta \zeta^\top \beta) + \mathbb{E}[\mathbb{E}(2\epsilon \zeta^\top \beta | \xi)] + \mathbb{E}(\epsilon^2) - \mathbb{E}(\beta^\top T \beta) - s^2 \\ &= \beta^\top \mathbb{E}[\mathbb{E}(T | \xi)] \beta + \sigma_\epsilon^2 - \beta^\top \mathbb{E}(T) \beta - s^2 \\ &= \sigma_\epsilon^2 - s^2, \end{aligned}$$

which is 0 iff  $s^2 = \sigma_\epsilon^2$ .

Next consider SS or QS. Here we have

$$\begin{aligned} \mathbb{E} \left[ \psi_{\epsilon S}(s^2, \beta) \right] &= \mathbb{E} \left[ \mathbb{E}(\beta^\top \zeta \zeta^\top \beta + 2\epsilon \zeta^\top \beta + \epsilon^2 | x) \right] - \mathbb{E}(\beta^\top M \beta) - s^2 \\ &= \sigma_\epsilon^2 - s^2 \end{aligned}$$

which again is 0 iff  $s^2 = \sigma_\epsilon^2$ .

**Note 3** Note that not only do we have  $\mathbb{E}[\psi_{\epsilon S}(\sigma_\epsilon^2, \beta)] = 0$  but even

$$\mathbb{E} \left[ \psi_{\epsilon S}(\sigma_\epsilon^2, \beta) | x \right] = 0.$$

This can be seen from

$$\begin{aligned} \mathbb{E}(\epsilon^2|x) &= \mathbb{E} \left[ \mathbb{E}(\epsilon^2|x, \xi)|x \right] = \sigma_\epsilon^2 \\ \mathbb{E}(\epsilon \zeta|x) &= \mathbb{E} [\mathbb{E}(\epsilon \zeta|x, \xi)|x] = \mathbb{E} [\zeta \mathbb{E}(\epsilon|x, \xi)|x] = 0. \end{aligned}$$

**Theorem 3** We first show that  $\mathbb{E} [\psi_{\alpha S}(\psi_{\alpha C} - \psi_{\alpha S})^\top] = 0$ . Indeed, this matrix is equal to

$$\mathbb{E} \begin{pmatrix} \psi_{\beta S}(\psi_{\beta C} - \psi_{\beta S})^\top & \psi_{\beta S}(\psi_{\epsilon C} - \psi_{\epsilon S}) \\ \psi_{\epsilon S}(\psi_{\beta C} - \psi_{\beta S})^\top & \psi_{\epsilon S}(\psi_{\epsilon C} - \psi_{\epsilon S}) \end{pmatrix}.$$

All four parts of this matrix vanish. For the upper left part this has been shown in the proof of Lemma 3. Because  $\mathbb{E}(\psi_{\epsilon S}|x) = 0$  (compare Note 3), one can show in a similar way that the two lower parts also vanish. The upper right part is zero because  $\mathbb{E}(\psi_{\beta S}|x) = 0$ . It now follows immediately, just as in the proof of Lemma 3, that

$$\mathbb{E} \left( \psi_{\alpha C} \psi_{\alpha C}^\top \right) \geq \mathbb{E} \left( \psi_{\alpha S} \psi_{\alpha S}^\top \right)$$

or in other words,

$$B_{\alpha C} \geq B_{\alpha S}.$$

We have equality iff  $\psi_{\alpha C} = \psi_{\alpha S}$  a.s. If  $\beta_{(0)} = 0$ , then  $\psi_{\beta C} - \psi_{\beta S} = (T - t\mu^\top)\beta = 0$  by the proof of Lemma 3. But also  $\psi_{\epsilon C} = \psi_{\epsilon S}$  because  $\psi_{\epsilon C} - \psi_{\epsilon S} = \beta^\top (M - T) \beta$ , which is zero for  $\beta_{(0)} = 0$ , as  $M_{00} - T_{00} = \mu_0 - t_0 = 0$ . Thus  $\beta_{(0)} = 0$  implies  $\psi_{\alpha C} = \psi_{\alpha S}$ . Conversely, under **(N)**,  $\psi_{\alpha C} = \psi_{\alpha S}$  implies  $\psi_{\beta C} = \psi_{\beta S}$  and this in turn implies  $\beta_{(0)} = 0$  by Lemma 3.

We know already from Lemma 2 that  $A_{\beta C} = A_{\beta S}$ .

Similarly

$$\begin{aligned} A_{\epsilon \beta C} &= 2\beta^\top \mathbb{E}(T) = 2\beta^\top \mathbb{E}[\mathbb{E}(T|\xi)] \\ &= 2\beta^\top \mathbb{E}(\zeta \zeta^\top) = 2\beta^\top \mathbb{E}(\zeta \zeta^\top|x) \\ &= 2\beta^\top \mathbb{E}(M) = A_{\epsilon \beta S}. \end{aligned}$$

Thus

$$A_{\alpha C} = A_{\alpha S},$$

and the theorem is proved.

In addition, consider  $B_\epsilon$  under **(N)**. If  $\beta_{(0)} \neq 0$ ,  $\psi_{\epsilon C} - \psi_{\epsilon S} = \beta^\top (M - T) \beta$  is a polynomial of degree  $2r$  with  $r$  as in the proof of Lemma 3, and consequently  $\mathbb{P}(\psi_{\epsilon C} - \psi_{\epsilon S} \neq 0) > 0$ , which implies  $B_{\epsilon C} > B_{\epsilon S}$ , and, together with  $A_{\alpha C} = A_{\alpha S}$ , this implies  $\Sigma_{\epsilon C} > \Sigma_{\epsilon S}$ .

In order to see that  $\beta_{(0)} \neq 0$  does not necessarily imply  $\Sigma_{\alpha C} > \Sigma_{\alpha S}$ , consider a linear model ( $k = 1$ ) under  $(\mathbf{N})$ . We can show that for any  $\beta$ , there is always a vector  $a \neq 0$  such that  $a^\top (\psi_{\alpha C} - \psi_{\alpha S}) \equiv 0$  and thus  $a^\top (\Sigma_{\alpha C} - \Sigma_{\alpha S}) a = 0$ .

Indeed, with  $\mu_0 = t_0 = 1$ ,

$$\psi_{\alpha C} - \psi_{\alpha S} = \beta_1 \begin{pmatrix} \mu_1 - t_1 \\ \mu_1 t_1 - t_2 \\ 2\beta_0(\mu_1 - t_1) + \beta_1(\mu_2 - t_2) \end{pmatrix}.$$

Let  $\mu_x = 0$ . Then  $\mu_1 = \rho x$ , see proof of Lemma 3, and  $\mu_2 = \mu_1^2 + \sigma_\delta^2 \rho$ ,  $t_1 = x$ ,  $t_2 = x^2 - \sigma_\delta^2$  (Kukush et al. 2005). Hence

$$\psi_{\alpha C} - \psi_{\alpha S} = \beta_1 \begin{pmatrix} (\rho - 1)x \\ (\rho - 1)x^2 + \sigma_\delta^2 \\ 2\beta_0(\rho - 1)x + \beta_1(\rho + 1)((\rho - 1)x^2 + \sigma_\delta^2) \end{pmatrix}.$$

Let  $a = (-2\beta_0, -\beta_1(\rho + 1), 1)^\top$ , then  $a^\top (\psi_{\alpha C} - \psi_{\alpha S}) \equiv 0$ . (In a similar way we can construct a counterexample for  $k = 2$  if  $\beta_2 = 0$ ).

**Note 4** It might be noted that in the linear model, under  $(\mathbf{N})$ ,  $v = \sigma_\epsilon^2 + \beta_1^2(\mu_2 - \mu_1^2) = \sigma_\epsilon^2 + \beta_1^2 \sigma_\delta^2 \rho = \text{const}$  and hence that  $\psi_{\alpha S}$  is equivalent to  $\psi_{\alpha Q}$ , which implies  $\Sigma_{\alpha S} = \Sigma_{\alpha Q}$ .

## References

- Carroll RJ, Ruppert D, Stefanski LA (1995) Measurement error in nonlinear models. Chapman and Hall, London
- Chan LK, Mak TK (1985) On the polynomial functional relationship. *J Royal Stat Soc B* 47:510–518
- Cheng C-L, Schneeweiss H (1998) Polynomial regression with errors in the variables. *J Royal Stat Soc B* 60:189–199
- Cheng C-L, Schneeweiss H (2002) On the polynomial measurement error model. In: van Huffel S, Lemmerling P (eds) Total least squares and errors-in-variables modeling. Kluwer, Dordrecht, pp 131–143
- Fuller WA (1987) Measurement error models. Wiley, New York
- Heyde CC (1997) Quasi-likelihood and its application. Springer, Berlin Heidelberg New York
- Kuha JT, Temple J (2003) Covariate measurement error in quadratic regression. *Int Stat Rev* 71:131–150
- Kukush A, Schneeweiss H (2005) Comparing different estimators in a nonlinear measurement error model. *I. Math Methods Stat* 14:53–79
- Kukush A, Schneeweiss H, Wolf R (2001) Comparison of three estimators in a polynomial regression with measurement errors. Discussion Paper 233, SFB 386, Universität München
- Kukush A, Schneeweiss H, Wolf R (2002) Comparing different estimators in a nonlinear measurement error model. Discussion Paper 244, SFB 386, Universität München
- Kukush A, Schneeweiss H, Wolf R (2005) Relative efficiency of three estimators in a polynomial regression with measurement errors. *J Stat Plan Inference* 127:179–203
- Schneeweiss H, Cheng C-L (2006) Bias of the structural quasi-score estimator of a measurement error model under misspecification of the regressor distribution. *J Multivar Anal* 97:455–473
- Shklyar S, Schneeweiss H (2005) A comparison of asymptotic covariance matrices of three consistent estimators in the Poisson regression model with measurement errors. *J Multivar Anal* 94:250–270

- Stefanski LA (1989) Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Commun Stat Part A Theory Methods* 18:4335–4358
- Thamerus M (1998) Different nonlinear regression models with incorrectly observed covariate. In: Galata R, Küchenhoff H (eds) *Econometrics in theory and practice*, Festschrift for Hans Schneeweiß. Physica, Heidelberg, New York
- Wolf R (2004) *Vergleich von funktionalen und strukturellen Messfehlerverfahren*. Logos Verlag, Berlin