Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT

Journal of Statistical Planning and Inference 127 (2005) 179–203

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Relative efficiency of three estimators in a polynomial regression with measurement errors

Alexander Kukush[a], Hans Schneeweiss[b,*], Roland Wolf[b]

[a]Department of Mathematics, Kiev Taras Shevchenko University, 01601 Kiev, Ukraine
[b]Department of Statistics, University of Munich, Akademiestr. 1, 80799 Munich, Germany

Received 28 August 2002; accepted 6 September 2003

Abstract

In a polynomial regression with measurement errors in the covariate, the latter being supposed to be normally distributed, one has (at least) three ways to estimate the unknown regression parameters: one can apply ordinary least squares (OLS) to the model without regard to the measurement error or one can correct for the measurement error, either by correcting the estimating equation (ALS) or by correcting the mean and variance functions of the dependent variable, which is done by conditioning on the observable, error ridden, counter part of the covariate (SLS). While OLS is biased, the other two estimators are consistent. Their asymptotic covariance matrices and thus their relative efficiencies can be compared to each other, in particular for the case of a small measurement error variance. In this case, it appears that ALS and SLS become almost equally efficient, even when they differ noticeably from OLS.

MSC: 62F12; 62J99

Keywords: Adjusted least squares; Measurement errors; Ordinary least squares; Polynomial regression; Structural least squares; Small error variance

1. Introduction

The polynomial structural regression model with measurement errors in the covariate and normally distributed variables is given by the equations

$$y_i = \sum_{j=0}^{k} \beta_j \xi_i^j + \varepsilon_i, \tag{1}$$

* Corresponding author. Fax: +49-89-2180-5044.
E-mail addresses: alexander_kukush@univ.kiev.ua (A. Kukush), schneew@stat.uni-muenchen.de (H. Schneeweiss), wolf@stat.uni-muenchen.de (R. Wolf).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2003.09.016

$$x_i = \xi_i + \delta_i, \tag{2}$$

where $(\xi_i, \varepsilon_i, \delta_i) \sim$ are i.i.d. random variables, ξ_i has a density and moments up to the order of $2k$, $\varepsilon_i \sim \mathrm{N}(0, \sigma_\varepsilon^2)$, $\delta_i \sim \mathrm{N}(0, \sigma_\delta^2)$ and ξ_i, ε_i, δ_i are mutually independent, $i = 1, \ldots, n$.

According to the usual interpretation of measurement error models, y is an observable variable depending on the regressor variable ξ, here via a polynomial regression function. The variable ξ, however, is not directly observable; it is latent. Instead a surrogate variable x is observed, which is related to ξ as described in (2), where δ is a normally distributed measurement error, independent of ξ, with expectation 0 and variance σ_δ^2. In addition to the error δ, there is also an "error in the equation", ε, which is also normally distributed with expectation 0 and variance σ_ε^2 and is independent of ξ and δ. Although δ is latent, it is assumed that σ_δ^2 is known. Below, see Section 4, we will additionally assume that ξ is normally distributed with mean μ and variance σ^2.

Two consistent estimators for the parameter vector $\beta = (\beta_0, \ldots, \beta_k)'$ have been considered in Kukush and Schneeweiss (2000): the *adjusted least squares* (*ALS*) and the *structural least squares* (*SLS*) estimators, see Cheng and Schneeweiss (1998) or Stefanski (1989) for ALS and Thamerus (1998) or Carroll et al. (1995) for SLS. Both assume σ_δ^2 to be known. The first one does not take the distribution of ξ into account, whereas the latter one heavily rests on the assumption that ξ is normally distributed.

If ξ is indeed normally distributed, one might expect SLS to be more efficient than ALS, since SLS uses the information about the distribution of ξ, while ALS does not. However, as neither of these estimators are ML estimators, this presumption is not evident at the outset and therefore needs further investigation. (Note that ML is very difficult to apply in this model.) Some simulation results in Schneeweiss and Nittner (2001) and also in Kuha and Temple (2003) seem to indicate that SLS has in fact smaller asymptotic standard errors than ALS. The theoretical investigations of Kukush and Schneeweiss (2000) come to the conclusion that at least in certain border line cases SLS is, in a sense, more efficient than ALS. In particular, they dealt with the situation where both error variances σ_ε^2 and σ_δ^2 were small.

Here we take an approach which is slightly different from the one in Kukush and Schneeweiss (2000), yet leads to a completely different result. We consider a model with small σ_δ^2 but comparatively large σ_ε^2. More precisely, we let σ_δ^2 go to zero keeping σ_ε^2 constant, cf. also Kukush et al. (2001). In doing so, we study how the asymptotic standard errors of the ALS and SLS estimators approach each other for small σ_δ^2.

Clearly, for small σ_δ^2 the asymptotic covariance matrices of the ALS and SLS estimators come close to each other. Indeed, for $\sigma_\delta^2 = 0$ (i.e., when measurement errors are not present) they simply coincide. The not so obvious question is, how fast do they approach each other when σ_δ^2 tends to zero. Is the difference of the covariance matrices for small σ_δ^2 proportional to σ_δ^2 or is it rather proportional to σ_δ^4? In the latter case, this would mean that small error variances lead to extremely (more than proportional) small differences in the covariance matrices, so that the two estimators could be deemed almost equally efficient. It turns out that this is just the case: the asymptotic covariance matrices of $\hat{\beta}_{\mathrm{ALS}}$ and $\hat{\beta}_{\mathrm{SLS}}$ are *equal up to the order of* σ_δ^2. This result, which is the main result of our paper, can be derived using Taylor series expansions of the formulas for the asymptotic covariance matrices of $\hat{\beta}_{\mathrm{ALS}}$ and $\hat{\beta}_{\mathrm{SLS}}$. However, as these formulas do not exist in an explicit form, the derivation of this result is not straightforward.

In a similar way, we also derive a small-σ_δ^2 approximation to the asymptotic bias and the asymptotic covariance matrix of the naive *ordinary least squares* (*OLS*) estimator of β. The OLS estimator is constructed by simply using the observed, error contaminated, variable x in place of the latent, error free, variable ξ and applying the method of least squares to the polynomial regression. The OLS estimator is inconsistent, but—after a suitable normalisation—the asymptotic variances of the estimators of the components of β are less than those of the ALS or SLS estimators for small σ_δ^2.

These theoretical results are corroborated by a simulation study.

In Section 2, we study the bias and the asymptotic covariance matrix of $\hat{\beta}_{\mathrm{OLS}}$ for small σ_δ^2. In Sections 3 and 4 we do the same for $\hat{\beta}_{\mathrm{ALS}}$ and $\hat{\beta}_{\mathrm{SLS}}$, respectively. Section 5 compares $\hat{\beta}_{\mathrm{OLS}}$ to $\hat{\beta}_{\mathrm{ALS}}$ or $\hat{\beta}_{\mathrm{SLS}}$. In Section 6 we briefly reexamine the case when both error variances are small and present some new results. Section 7 has some simulation results and an example, and Section 8 concludes with some further remarks. Proofs are relegated to the appendix. In the appendix we also prove a matrix inequality that we need in Section 6.

As we deal with asymptotic properties exclusively, we often omit the adjective “asymptotic” in the sequel.

2. The naive estimator $\hat{\beta}_{\mathrm{OLS}}$

2.1. Bias and covariance matrix

With $\zeta_i = (1, \xi_i, \ldots, \xi_i^k)'$ the model equation (1) can be written as

$$y_i = \zeta_i'\beta + \varepsilon_i. \tag{3}$$

The (naive) OLS estimator of β is found by replacing ζ_i in (3) with the corresponding vector $z_i = (1, x_i, \ldots, x_i^k)'$ of the observable variable x, and then applying ordinary least squares (OLS). It is given by

$$\hat{\beta}_{\mathrm{OLS}} = \left(\sum_1^n z_i z_i'\right)^{-1} \sum_1^n z_i y_i. \tag{4}$$

The estimating error, as derived from (3) and (4), is

$$\hat{\beta}_{\mathrm{OLS}} - \beta = -\left(\sum_1^n z_i z_i'\right)^{-1} \left\{\sum_1^n z_i(z_i - \zeta_i)'\beta - \sum_1^n z_i \varepsilon_i\right\}. \tag{5}$$

By the Strong Law of Large Numbers, the bias of $\hat{\beta}_{\mathrm{OLS}}$,

$$b = \lim_{n\to\infty} (\hat{\beta}_{\mathrm{OLS}} - \beta),$$

is found to be

$$b = -(Ezz')^{-1}Ez(z-\zeta)'\beta. \tag{6}$$

Here and in the following we do not use brackets for the expectation operator. The operator E is always understood to operate on the whole term to the right of E, terms being separated by $+$ or $-$. I.i.d. variables without the observation index i are understood to be the variables with index $i=1$, say.

In order to derive the asymptotic covariance matrix of $\hat\beta_{\mathrm{OLS}}$, we find the limiting distribution of $\sqrt{n}(\hat\beta_{\mathrm{OLS}}-\beta-b)$. Note that we use $\lim \hat\beta_{\mathrm{OLS}} = \beta+b$ as the center of the distribution, not β. As $\hat\beta_{\mathrm{OLS}}$ is biased, use of β as the center of the distribution would lead the expression $\sqrt{n}(\hat\beta_{\mathrm{OLS}}-\beta)$ to diverge as $n\to\infty$. We have the following result, the proof of which is found in Appendix A.

Proposition 2.1. *The OLS estimator* (4) *has an asymptotic bias given by* (6). *It has an asymptotic normal distribution*

$$\sqrt{n}(\hat\beta_{\mathrm{OLS}}-\beta-b) \to \mathrm{N}(0,\Sigma_{\mathrm{OLS}})$$

with a covariance matrix given by

$$\Sigma_{\mathrm{OLS}} = E(zz')^{-1}E\{(z-\zeta)'\beta + z'b\}^2 zz'(Ezz')^{-1} + \sigma_\varepsilon^2(Ezz')^{-1}. \tag{7}$$

2.2. Expansion of bias and covariance matrix

We want to find approximate expressions for b and Σ_{OLS} as σ_δ^2 tends to zero. We do not have explicit expressions for b and Σ_{OLS}. Still, by using the Taylor series expansion

$$z = \zeta + \frac{\mathrm{d}\zeta}{\mathrm{d}\xi}\delta + \frac{1}{2}\frac{\mathrm{d}^2\zeta}{\mathrm{d}\xi^2}\delta^2 + \mathrm{O}(\delta^3), \tag{8}$$

which relates the observable vector z to the corresponding latent vector ζ, we can find expansions of b and Σ_{OLS} up to the order of σ_δ^2.

Proposition 2.2. *The bias and the asymptotic covariance matrix of* $\hat\beta_{\mathrm{OLS}}$ *can be expressed in terms up to the order of* σ_δ^2 *for* $\sigma_\delta \to 0$ *as follows*:

$$b = -\sigma_\delta^2\Phi^{-1}E\left(\frac{\mathrm{d}\zeta}{\mathrm{d}\xi}\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} + \frac{1}{2}\zeta\frac{\mathrm{d}^2\zeta'}{\mathrm{d}\xi^2}\right)\beta + \mathrm{O}(\sigma_\delta^4), \tag{9}$$

$$\Sigma_{\mathrm{OLS}} = \sigma_\varepsilon^2\Phi^{-1} + \sigma_\delta^2\Phi^{-1}\left\{E\left(\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi}\beta\right)^2\zeta\zeta' - \frac{1}{2}\sigma_\varepsilon^2 E\frac{\mathrm{d}^2(\zeta\zeta')}{\mathrm{d}\xi^2}\right\}\Phi^{-1} + \mathrm{O}(\sigma_\delta^4), \tag{10}$$

where $\Phi = E\zeta\zeta'$.

Remark. In the special case of a linear regression ($k=1$) the well-known attenuation effect for β_1 can be seen:

$$\mathrm{bias}(\hat\beta_1) = -(\sigma_\delta^2/\sigma_\xi^2)\beta_1 + \mathrm{O}(\sigma_\delta^4).$$

Actually, in this case, the bias is exactly given by $\text{bias}(\hat{\beta}_1) = -(\sigma_\delta^2/\sigma_x^2)\beta_1$, see, e.g., Schneeweiss and Mittag (1986, p. 140).

3. The adjusted least squares estimator $\hat{\beta}_{\text{ALS}}$

3.1. Covariance matrix

The adjusted least squares (ALS) estimator of β is the (a.s. unique) solution of the unbiased estimating equation

$$\sum_1^n H_i \hat{\beta}_{\text{ALS}} = \sum_1^n h_i, \tag{11}$$

where $(H_i)_{rs} = t_{r+s}(x_i)$, $r, s = 0, \dots, k$, $(h_i)_r = y_i t_r(x_i)$, $r = 0, \dots, k$, with $t_r(x)$ being a polynomial in x of degree r such that, for $x = \xi + \delta$,

$$E\{t_r(x)|\xi\} = \xi^r. \tag{12}$$

The polynomials $t_r(x)$ satisfy the recursion formula

$$t_{r+1}(x) = x\, t_r(x) - \sigma_\delta^2\, r\, t_{r-1}(x) \tag{13}$$

with initial conditions $t_0(x) = t_{-1}(x) = 1$, cf. Cheng and Schneeweiss (1998). σ_δ^2 is assumed to be known.

According to the theory of unbiased estimating equations, see, e.g., Heyde (1997), the estimator $\hat{\beta}_{\text{ALS}}$ is consistent and asymptotically normally distributed with asymptotic covariance matrix

$$\Sigma_{\text{ALS}} = (EH)^{-1} E(H\beta - h)(H\beta - h)'(EH)^{-1}. \tag{14}$$

This can be given another form more suitable to the derivation of the small-σ_δ result of the next subsection.

Proposition 3.1. *The asymptotic covariance matrix of $\hat{\beta}_{\text{ALS}}$ is given by*

$$\Sigma_{\text{ALS}} = \Phi^{-1}\{\sigma_\varepsilon^2 Ett' + E(H - t\zeta')\beta\beta'(H - \zeta t')\}\Phi^{-1}, \tag{15}$$

where $t = (t_0(x), \dots, t_k(x))'$ and $\Phi = E\zeta\zeta'$.

3.2. Expansion of Σ_{ALS}

Similarly as in Section 2.2, we want to derive an approximate expression for the asymptotic covariance matrix Σ_{ALS}, which is valid for small σ_δ. The key to this derivation is again the Taylor series expansion (8) supplemented by another expansion of t about z:

$$t = z - \frac{1}{2}\sigma_\delta^2 \frac{\mathrm{d}^2 z}{\mathrm{d}x^2} + \mathrm{O}(\sigma_\delta^4), \tag{16}$$

which follows from the recursion formula (13). In (16), $\mathrm{O}(\sigma_\delta^4)$ is not only a function of σ_δ but also of the random variable $x=\xi+\delta$: the components of $\mathrm{O}(\sigma_\delta^4)$ are polynomials in ξ and δ. We have the following result

Proposition 3.2. *The asymptotic covariance matrix of $\hat{\beta}_{\mathrm{ALS}}$ can be expressed in terms up to the order of σ_δ^2 for $\sigma_\delta \to 0$ as follows*:

$$\Sigma_{\mathrm{ALS}} = \sigma_\varepsilon^2 \Phi^{-1} + \sigma_\delta^2 \Phi^{-1} \left\{ E\left(\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi}\beta\right)^2 \zeta\zeta' + \sigma_\varepsilon^2 E \frac{\mathrm{d}\zeta}{\mathrm{d}\xi}\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} \right\} \Phi^{-1} + \mathrm{O}(\sigma_\delta^4). \tag{17}$$

4. The structural least squares estimator $\hat{\boldsymbol{\beta}}_{\mathrm{SLS}}$

4.1. Covariance matrix

From now on we enrich our original model (1), (2) by assuming additionally that ξ is normally distributed:

$$\xi \sim \mathrm{N}(\mu_\xi, \sigma_\xi^2).$$

This assumption may seem rather restrictive. But it allows us to derive a new, so-called structural least squares (SLS), estimator, which is based on this assumption. One can actually relax this assumption by introducing a finite mixture of normals rather than a single normal distribution, rendering the approach rather flexible, see Küchenhoff and Carroll (1997). Here we prefer to work with the simple assumption of just one normal distribution, but the subsequent results would still hold in the more general setting.

We can reformulate (1) as a mean–variance model in the latent variable ξ:

$$E(y|\xi) = \zeta'\beta,$$

$$V(y|\xi) = \sigma_\varepsilon^2.$$

We find a new (conditional) mean–variance model in the observable variable x by taking conditional expectations given x, cf. Thamerus (1998):

$$E(y|x) = \mu(x)'\beta \stackrel{\mathrm{df}}{=} m(x,\beta), \tag{18}$$

$$V(y|x) = \sigma_\varepsilon^2 + \beta'\{M(x) - \mu(x)\mu(x)'\}\beta \stackrel{\mathrm{df}}{=} v(x,\beta,\sigma_\varepsilon^2). \tag{19}$$

Here $\mu(x) = (1, \mu_1(x), \ldots, \mu_k(x))'$ is a vector consisting of components

$$\mu_r(x) = E(\xi^r|x)$$

and $(M(x))_{r,s} = \mu_{r+s}(x),\ r,s = 0,\ldots,k$.

Note that the first row and column of $M - \mu\mu'$ are zero and that after deletion of the first row and column the remaining matrix is positive definite. Therefore $v > 0$ if $\sigma_\varepsilon^2 + \sum_{j=1}^k \beta_j^2 > 0$.

Now, because of (2) and because ξ and x are jointly normally distributed, the conditional distribution of ξ given x is $\mathrm{N}(\mu_1(x), \tau^2)$ with

$$\mu_1(x) = \mu_x + (1 - \sigma_\delta^2/\sigma_x^2)(x - \mu_x), \tag{20}$$

$$\tau^2 = \sigma_\delta^2(1 - \sigma_\delta^2/\sigma_x^2), \tag{21}$$

where $\mu_x = \mu_\xi$ and $\sigma_x^2 = \sigma_\xi^2 + \sigma_\delta^2$. We can therefore compute the conditional moments of ξ given x as

$$\mu_r(x) = \sum_{j=0}^{r} \binom{r}{j} \mu_j^* \mu_1(x)^{r-j}, \tag{22}$$

$$\mu_j^* = \begin{cases} 0 & \text{if } j \text{ is odd,} \\ 1 \cdot 3 \cdots (j-1)\tau^j & \text{if } j \text{ is even, } \mu_0^* = 1, \end{cases}$$

cf. Thamerus (1998). In these formulas the nuisance parameters μ_x and σ_x^2 have to be replaced with their estimates in order to derive a modified mean–variance model corresponding to (18), (19), in which $\hat{\beta}_{\mathrm{SLS}}$ and $\hat{\sigma}^2_{\varepsilon\mathrm{SLS}}$ are then found as the solution to the (asymptotically unbiased) estimating equations

$$\sum_1^n \frac{y_i - \hat{m}(x_i, \beta)}{\hat{v}(x_i, \beta, \sigma_\varepsilon^2)} \hat{\mu}(x_i) = 0,$$

$$\sigma_\varepsilon^2 = \frac{1}{n-k-1} \sum_1^n \{y_i - \hat{m}(x_i, \beta)\}^2 - \frac{1}{n} \sum_1^n \beta' \{\hat{M}(x_i) - \hat{\mu}(x_i)\hat{\mu}(x_i)'\}\beta,$$

cf. Kukush and Schneeweiss (2000); the second equation is different from (15) in Kukush and Schneeweiss (2000), but can just as well be used to estimate and update σ_ε^2. The "hat" serves to remind us of the replacement of μ_x and σ_x^2 with their estimates $\bar{x}$ and s_x^2.

The estimating equations are solved by an iterative procedure (iteratively reweighted least squares), which can be shown to converge eventually, i.e., for sufficiently large n, see Kukush et al. (2001). During the iterations and also as a final result $\hat{\sigma}_\varepsilon^2$ may become negative. The estimation method can be greatly improved by providing bounds for $\hat{\sigma}^2_{\varepsilon\mathrm{SLS}}$. If there are no prior bounds for σ_ε^2, one might choose bounds like

$$\frac{1}{n} \hat{\sigma}^2_{\varepsilon\mathrm{OLS}} < \hat{\sigma}^2_{\varepsilon\mathrm{SLS}} < n\hat{\sigma}^2_{\varepsilon\mathrm{OLS}}.$$

Whenever one of these bounds is exceeded, the estimate $\hat{\sigma}_\varepsilon^2$ is set equal to that bound. These bounds, which have been found on empirical grounds, have the advantage that they tend to disappear for large n, where they are of no importance any more, whereas for small n they are simple multiples or fractions of a plausible surrogate estimate of σ_ε^2, viz., $\hat{\sigma}^2_{\varepsilon\mathrm{OLS}}$, and thus provide genuine restrictions for $\hat{\sigma}^2_{\varepsilon\mathrm{SLS}}$. But, of course, other bounds are also possible. Whatever bounds are chosen, they have no influence on the results of the paper, which deals with asymptotic properties exclusively.

As to the asymptotic properties of $\hat{\beta}_{\mathrm{SLS}}$ we have the following result.

Proposition 4.1. *Suppose ξ is normally distributed. Under the "technical" assumption that β lies in a given bounded open set of $\mathbb{R}^{k+1}$ and σ_ε^2 lies in a given open interval of $\mathbb{R}^+$, the resulting SLS estimators of β and σ_ε^2 are consistent and asymptotically normally distributed. The asymptotic covariance matrix Σ_{SLS} of $\hat\beta_{\text{SLS}}$ is given by*

$$\Sigma_{\text{SLS}} = \left(E\,\frac{\mu(x)\mu(x)'}{v(x,\beta,\sigma_\varepsilon^2)}\right)^{-1} + \left(E\,\frac{\mu(x)\mu(x)'}{v(x,\beta,\sigma_\varepsilon^2)}\right)^{-1}\left(\sigma_x^2F_1F_1' + \frac{2}{\sigma_x^4}F_2F_2'\right)\left(E\,\frac{\mu(x)\mu(x)'}{v(x,\beta,\sigma_\varepsilon^2)}\right)^{-1}, \tag{23}$$

where

$$F_j = -E\frac{\mu(x)}{v(x,\beta,\sigma_\varepsilon^2)}\,\frac{\partial\mu(x)'}{\partial\gamma_j}\,\beta, \quad j=1,2$$

and $\gamma_1 = \mu_x$, $\gamma_2 = \sigma_x^{-2}$.

The last term in (23) is due to the fact that the "nuisance" parameters μ_x and σ_x^2 have to be estimated before the SLS estimating equations can even be set up. If these parameters were known, (23) would reduce to a much simpler formula with only the first term on the right-hand side, cf. Carroll et al. (1995, A.4.2). Fortunately, for our purpose, the second term is irrelevant, as it is of the order $\mathrm{O}(\sigma_\delta^4)$. We thus can simplify (23) for small σ_δ^2.

Proposition 4.2. *When $\sigma_\delta^2 \to 0$, the asymptotic covariance matrix of $\hat\beta_{\text{SLS}}$ is, up to order of σ_δ^2, given by*

$$\Sigma_{\text{SLS}} = \left(E\frac{\mu(x)\mu(x)'}{v(x,\beta,\sigma_\varepsilon^2)}\right)^{-1} + \mathrm{O}(\sigma_\delta^4). \tag{24}$$

4.2. Expansion of Σ_{SLS} and its comparison to Σ_{ALS}

The SLS estimator $\hat\beta_{\text{SLS}}$ was constructed in a completely different way than the ALS estimator. Whereas the latter does not make use of the distribution of ξ, the former relies heavily on the normality assumption for ξ. No wonder then that the asymptotic covariance matrices of these two estimators look quite different, see (14) or (15) and (23) or (24). Once the higher moments of x are known—and with the normality assumption they are indeed well-known—, Σ_{ALS} can easily be computed via (14) or (15). On the other hand, Σ_{SLS} does not lend itself to an easy computation. Even under the normality assumption the expectation in (23) or (24) cannot be found explicitly but must be computed by numerical integration. Despite these differences between ALS and SLS, it turns out that, for small σ_δ, the two asymptotic covariance matrices are essentially equal. They only differ by a term of order σ_δ^4. This is our main result.

Theorem. *The asymptotic covariance matrices of the two estimators $\hat\beta_{\text{ALS}}$ and $\hat\beta_{\text{SLS}}$ of the parameter vector β of the polynomial model (1), (2) with normally distributed*

ξ are equal up to the order of σ_δ^2, when $\sigma_\delta^2 \to 0$, i.e.

$$\Sigma_{\mathrm{ALS}} = \Sigma_{\mathrm{SLS}} + \mathrm{O}(\sigma_\delta^4).$$

In other words, if σ_δ^2 is sufficiently small, then ALS and SLS are almost equally efficient.

5. Comparison of Σ_{OLS} to Σ_{ALS} and Σ_{SLS}

We now want to compare OLS to ALS or to SLS, again for small σ_δ^2. As Σ_{ALS} and Σ_{SLS} are equal up to the order of σ_δ^2 we need only consider the difference of (17) and (10):

$$\Sigma_{\mathrm{ALS}} - \Sigma_{\mathrm{OLS}} = \sigma_\delta^2 \sigma_\varepsilon^2 \Phi^{-1} \left\{ E \frac{\mathrm{d}\zeta}{\mathrm{d}\xi} \frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} + \frac{1}{2} E \frac{\mathrm{d}^2(\zeta\zeta')}{\mathrm{d}\xi^2} \right\} \Phi^{-1} + \mathrm{O}(\sigma_\delta^4).$$

Obviously, Σ_{ALS} and Σ_{OLS} are not equal up to the order of σ_δ^2. One might expect the difference of Σ_{ALS} and Σ_{OLS} to be positive definite at least for small σ_δ^2. However, the matrix in braces—call it B—is not even positive semidefinite in general. Indeed for $k \geqslant 2$ and $\xi \sim \mathrm{N}(0,1)$, the submatrix of B consisting of the first three rows and columns is

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 10 \end{pmatrix},$$

which is indefinite, as it has a negative determinant. Thus, e.g., for the linear combination $\alpha = 4\beta_0 + 3\beta_2$ the OLS estimator has a larger asymptotic variance than the ALS estimator up to the order of σ_δ^2. More precisely,

$$\sigma^2(\hat{\alpha}_{\mathrm{OLS}}) - \sigma^2(\hat{\alpha}_{\mathrm{ALS}}) = 2\sigma_\delta^2\sigma_\varepsilon^2 + \mathrm{O}(\sigma_\delta^4).$$

However, the diagonal elements of B are obviously all positive except for the first one, which is zero. This means that the difference of the asymptotic variances of the ALS and OLS estimators of the jth element of $\Phi\beta$, $j = 1, \ldots, k$, is always positive for small enough σ_δ^2.

In this sense, OLS is superior to ALS or SLS, as far as the asymptotic variances go. It should be remembered, however, that OLS is a biased estimator.

6. Case of small σ_δ^2 and σ_ε^2

In this section we briefly return to a case dealt with in Kukush and Schneeweiss (2000), where σ_δ^2 and σ_ε^2 both tend to zero. The results of that earlier paper can now be derived very simply. We make, however, the simplifying assumption that σ_δ^2 and σ_ε^2 remain proportional to each other while going to zero, i.e.

$$\sigma_\varepsilon^2 = \lambda \sigma_\delta^2 \tag{25}$$

with some fixed constant $\lambda > 0$. We also derive a formula for the asymptotic covariance matrix of $\hat{\beta}_{\mathrm{OLS}}$, not considered in Kukush and Schneeweiss (2000). Indeed, Σ_{OLS} is found from (10) by simply omitting the last term in braces. With some additional algebra we get the following result.

Proposition 6.1. *Suppose* (25) *holds and* σ_δ^2 *tends to zero, then*

$$\Sigma_{\mathrm{OLS}} = \sigma_\delta^2 \Phi^{-1} E v_1 \zeta\zeta' \Phi^{-1} + \mathrm{O}(\sigma_\delta^4), \tag{26}$$

where

$$v_1 = v_1(\xi, \beta) = \lambda + \left(\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} \beta \right)^2. \tag{27}$$

Similarly, by omitting the last term in braces of (17) we can derive a corresponding expression for Σ_{ALS}. However, it turns out that this expression is the same as for Σ_{OLS}:

Proposition 6.2. *Under the conditions of Proposition* 6.1

$$\Sigma_{\mathrm{ALS}} = \Sigma_{\mathrm{OLS}} + \mathrm{O}(\sigma_\delta^4).$$

Finally Σ_{SLS} can also be expanded for $\sigma_\delta^2 \to 0$, but this time with some more refined arguments (see Appendix A):

Proposition 6.3. *Under the conditions of Proposition* 6.1

$$\Sigma_{\mathrm{SLS}} = \sigma_\delta^2 \left(E \frac{\zeta\zeta'}{v_1} \right)^{-1} + \mathrm{O}(\sigma_\delta^4). \tag{28}$$

Thus if σ_δ^2 and σ_ε^2 both go to zero, Σ_{ALS} (or equivalently Σ_{OLS}) and Σ_{SLS} differ at the order $\mathrm{O}(\sigma_\delta^2)$. It turns out that SLS is more efficient, in a sense, than ALS (and also than OLS) in this case:

Proposition 6.4. *Under the conditions of Proposition* 6.1 *and if* $\sum_{j=2}^k \beta_j^2 \neq 0$,

$$\lim_{\sigma_\delta^2 \to 0} \frac{1}{\sigma_\delta^2} \Sigma_{\mathrm{OLS}} = \lim_{\sigma_\delta^2 \to 0} \frac{1}{\sigma_\delta^2} \Sigma_{\mathrm{ALS}} \geqslant \lim_{\sigma_\delta^2 \to 0} \frac{1}{\sigma_\delta^2} \Sigma_{\mathrm{SLS}},$$

where the last two matrices are not equal.

This result is in stark contrast to the results of the theorem and of Section 5.

7. Some simulation results and an example

We simulated a quadratic model with $\beta = (0, 1, -0.5)'$ and a cubic model with $\beta = (0, 1, -0.5, 0.5)'$. The sample size was taken to be $n = 900$. We let $\varepsilon \sim \mathrm{N}(0, 1)$.

Table 1
Quadratic model: $\beta = (0, 1, -\frac{1}{2})'$, $\mu_\xi = 0$, $\sigma_\xi^2 = 1$, $\sigma_\varepsilon^2 = 20$, $\sigma_\delta^2 = 0.05$

	OLS		MALS simulation	MALS or SLS theory	SLS simulation
	Theory	Simulation			
Bias					
β_0	−0.0250	−0.0260	−0.0026	0	−0.0024
β_1	−0.0500	−0.0442	0.0033	0	0.0037
β_2	0.0500	0.0515	0.0054	0	0.0053
Standard deviation					
β_0	0.1830	0.1816	0.1851	0.1860	0.1848
β_1	0.1461	0.1483	0.1558	0.1535	0.1559
β_2	0.1008	0.1011	0.1119	0.1113	0.1116

In Schneeweiss and Nittner (2001) the error variances σ_δ^2 and σ_ε^2 were set equal to each other. Here, however, we studied two different scenarios. In the first one we took $\sigma_\varepsilon^2 = 20$ and $\sigma_\delta^2 = 0.05$, in accordance with Sections 2–5, where a small σ_δ^2 but not so small σ_ε^2 was required. The other scenario deals with the case of Section 6, when both error variances are small. We took $\sigma_\varepsilon^2 = 0.002$ and $\sigma_\delta^2 = 0.01$. The number of replications was $N = 1000$.

Note that in the first scenario σ_δ^2 is very small relative to the variance of ξ so that the results of Sections 2–5 come out quite clearly. On the other hand, it is still large enough so that the naive OLS method shows a noticeable bias, suggesting the use of more refined methods like ALS or SLS. For larger σ_δ^2 the bias would be even bigger but the approximations of our theory, though still applicable, would not be so clear cut any more. A similar remark can be made with respect to the second scenario.

We replaced ALS by a modified ALS method, called MALS. MALS has the same asymptotic properties as ALS, but is much stabler for small n, see Cheng et al. (2000).

We computed bias and standard error of the various estimators directly from the 1000 replications and compared them to the theoretical approximation values as computed from (9), (10), and (17) in the first scenario and (9) and (26) in the second scenario. As in these formulas the expected values consist of higher moments of ξ, these can be easily computed under the normality assumption for ξ. The results corroborate the theory, see Tables 1–4. We found similar results when the sample size was smaller, e.g. $n = 200$, not shown here, but see Kukush et al. (2001). Naturally, the simulation results are less stable for the cubic regression, but even there they agree sufficiently well with the theoretical results, Tables 2 and 4.

In the second scenario, Tables 3 and 4, we also computed the differences of the empirical covariance matrices of the three estimators. They were standardized by multiplying them with n/σ_δ^2. The results agree well with the theoretical statements in that the difference $\Sigma_{ALS} - \Sigma_{OLS}$ is very small, whereas $\Sigma_{ALS} - \Sigma_{SLS}$ is large.

When looking for a real data example to illustrate the results of the paper, one should keep in mind that, in contrast to the simulation study, the model

Table 2
Cubic model: $\beta=(0,1,-\frac{1}{2},\frac{1}{2})'$, $\mu_\xi=0$, $\sigma_\xi^2=1$, $\sigma_\varepsilon^2=20$, $\sigma_\delta^2=0.05$

	OLS		MALS simulation	MALS or SLS theory	SLS simulation
	Theory	Simulation			
Bias					
β_0	−0.0250	−0.0231	0.0017	0	0.0011
β_1	0.0250	0.0188	−0.0191	0	−0.0008
β_2	0.0500	0.0473	−0.0006	0	0.0002
β_3	−0.0750	−0.0678	0.0075	0	0.0003
Standard deviation					
β_0	0.1872	0.1915	0.1960	0.1902	0.1936
β_1	0.2510	0.2515	0.2780	0.2682	0.2715
β_2	0.1141	0.1167	0.1304	0.1235	0.1269
β_3	0.0706	0.0681	0.0820	0.0780	0.0782

Table 3
Quadratic model: $\beta=(0,1,-\frac{1}{2})'$, $\mu_\xi=0$, $\sigma_\xi^2=1$, $\sigma_\varepsilon^2=0.002$, $\sigma_\delta^2=0.010$

	OLS		MALS simulation	MALS or SLS theory	SLS simulation
	Theory	Simulation			
Bias					
β_0	−0.0050	−0.0051	−0.0002	0	0.0000
β_1	−0.0100	−0.0095	0.0003	0	0.0004
β_2	0.0100	0.0100	0.0001	0	−0.0001
Standard deviation					
β_0		0.0059	0.0061	0.0059	0.0043
β_1		0.0067	0.0068	0.0069	0.0056
β_2		0.0059	0.0059	0.0061	0.0042

Difference of the covariance matrices:

$$\frac{n}{\sigma_\delta^2}(\Sigma_{\mathrm{ALS}}-\Sigma_{\mathrm{OLS}})=\begin{pmatrix}0.091 & 0.056 & -0.112\\ 0.056 & 0.211 & -0.147\\ -0.112 & -0.147 & 0.210\end{pmatrix},$$

$$\frac{n}{\sigma_\delta^2}(\Sigma_{\mathrm{ALS}}-\Sigma_{\mathrm{SLS}})=\begin{pmatrix}1.535 & 1.013 & -1.670\\ 1.013 & 1.440 & -0.998\\ -1.670 & -0.998 & 1.814\end{pmatrix},$$

assumptions as required by the theory will be most often only approximately satisfied in practice.

Table 4
Cubic model: $\beta = (0, 1, -\frac{1}{2}, \frac{1}{2})'$, $\mu_\xi = 0$, $\sigma_\xi^2 = 1$, $\sigma_\varepsilon^2 = 0.002$, $\sigma_\delta^2 = 0.010$

	OLS		MALS simulation	MALS or SLS theory	SLS simulation
	Theory	Simulation			
Bias					
β_0	−0.0050	−0.0054	−0.0008	0	0.0002
β_1	0.0050	0.0049	−0.0017	0	−0.0012
β_2	0.0100	0.0103	0.0008	0	−0.0003
β_3	−0.0150	−0.0146	0.0004	0	0.0007
Standard deviation					
β_0		0.0174	0.0187	0.0178	0.0063
β_1		0.0384	0.0453	0.0398	0.0144
β_2		0.0227	0.0246	0.0235	0.0119
β_3		0.0163	0.0191	0.0169	0.0089

Difference of covariance matrices:

$$\frac{n}{\sigma_\delta^2}(\Sigma_{\text{ALS}} - \Sigma_{\text{OLS}}) = \begin{pmatrix} 1.322 & -0.288 & -2.041 & 0.157 \\ -0.288 & 9.246 & 0.361 & -3.650 \\ -2.041 & 0.361 & 3.349 & -0.340 \\ 0.157 & -3.650 & -0.340 & 1.744 \end{pmatrix},$$

$$\frac{n}{\sigma_\delta^2}(\Sigma_{\text{ALS}} - \Sigma_{\text{SLS}}) = \begin{pmatrix} 25.017 & -18.048 & -30.203 & 7.438 \\ -18.048 & 123.636 & 18.956 & -47.744 \\ -30.203 & 18.956 & 36.921 & -7.873 \\ 7.438 & -47.744 & -7.873 & 18.640 \end{pmatrix}$$

Cheng and Van Ness (1999, Example 6.3) gave an example of a quadratic model with measurement errors which we can use to illustrate our theorem. They modelled the dependency of yearly corn yield (y) on rainfall (x) in six Corn Belt States from 1890 to 1928 by the following quadratic equation.

$$y = \beta_0 + \beta_1\xi + \beta_2\xi^2 + \varepsilon.$$

Rainfall is measured with errors δ, however the error variance is not available. Here we assumed $\sigma_\delta^2 = 0.2$, which is relatively small with respect to the sample variance $\hat{\sigma}_x^2 = 5.1$. We tried also other values for σ_δ^2 with essentially the same results. The sample consisted of 38 data points (x_i, y_i), see Cheng and Van Ness (1999, Table 6.1). This is not much, and asymptotic theory may give poor results for a sample of this size.

Nevertheless, we got the following results for the parameter estimates where the values in parentheses are the estimated standard errors (Table 5). These were obtained by computing estimates of the covariance matrices of the three estimators and taking the square root of the diagonal elements. The covariance matrices were estimated by

Table 5
Corn yield and rainfall with $\sigma_\delta^2 = 0.2$

	OLS	SLS	ALS
β_0	−5.098	−7.465	−7.711
	(10.923)	(11.842)	(11.997)
β_1	6.015	6.431	6.454
	(1.857)	(2.023)	(2.043)
β_2	−0.230	−0.248	−0.248
	(0.078)	(0.085)	(0.086)

sandwich formulas of the form $\hat{\Sigma} = A^{-1}BA^{-1}$, where for the ALS, SLS, and OLS we took, respectively:

$$A_{\mathrm{ALS}} = \sum_{i=1}^{n} H_i, \quad B_{\mathrm{ALS}} = \sum_{i=1}^{n} (H_i\hat{\beta}_{\mathrm{ALS}} - h_i)(H_i\hat{\beta}_{\mathrm{ALS}} - h_i)',$$

$$A_{\mathrm{SLS}} = \sum_{i=1}^{n} \frac{\hat{\mu}(x_i)\hat{\mu}'(x_i)}{\hat{v}(x_i, \hat{\beta}_{\mathrm{SLS}}, \hat{\sigma}_\varepsilon^2)}, \quad B_{\mathrm{SLS}} = \sum_{i=1}^{n} \left(\frac{y_i - \hat{\mu}(x_i)\hat{\beta}_{\mathrm{SLS}}}{\hat{v}(x_i, \hat{\beta}_{\mathrm{SLS}}, \hat{\sigma}_\varepsilon^2)}\right)^2 \hat{\mu}(x_i)\hat{\mu}'(x_i),$$

$$A_{\mathrm{OLS}} = \sum_{i=1}^{n} z_i z_i', \quad B_{\mathrm{OLS}} = \sum_{i=1}^{n} (y_i - z_i'\hat{\beta}_{\mathrm{OLS}})^2 z_i z_i'.$$

First note that the SLS and ALS estimates both differ from the OLS estimates always in the same direction, indicating the presence of a bias for OLS. True, the differences are not significant, but this is due to the small sample size. Asymptotically, we can be sure of a bias for OLS.

It is seen that, for all three parameters, the standard errors of SLS and ALS come quite close to each other, while those of OLS are smallest. This corresponds nicely with the results of the theorem and Section 5.

8. Conclusion

We compared three estimators of the parameter vector β of a polynomial regression with measurement error: the naive ordinary least squares (OLS), the adjusted least squares (ALS), and the structural least squares (SLS) estimators. Although OLS is inconsistent, it is still worthwhile to compare it to the consistent ALS and SLS estimators because often the standard errors of OLS are so small as compared to those of ALS and SLS that in certain occasions OLS might be preferable despite its inconsistency. SLS relies on the knowledge of the regressor distribution (structural case), which here is taken to be a normal distribution. ALS is not based on any distributional assumption for the regressor (functional case) and is therefore more robust than SLS. Here, however, we do not study robustness properties (for this see Schneeweiss and Nittner (2001) and Schneeweiss et al. (2002)). Instead, we assume that SLS takes the normal regressor distribution correctly into account. In this case one might suspect that SLS is more efficient than ALS.

In order to study efficiency properties of the estimators we compute their asymptotic covariance matrices. Note that we do not deal with finite sample properties. This can be justified on the ground that in many practical applications, notably in epidemiology, large samples are quite common, so that asymptotic results do apply. Because the covariance matrices of the estimators are hard to compare in general, we restrict our investigations to borderline cases with small error variances. Border line cases are also interesting per se in that they give rise to particularly simple relations between the covariance matrices. We study two scenarios: in the first, the measurement error variance σ_δ^2 tends to zero, in the second, both error variances, σ_δ^2 and σ_ε^2, tend to zero, where σ_ε^2 is the variance of the error in the equation. For both cases approximate formulas of the covariance matrices are derived, which then can be compared to each other.

It should be clear that by studying limiting processes like $\sigma_\delta^2 \to 0$ we do not intend to study the limit model itself, which would just be an ordinary polynomial regression without measurement errors, rather our intention is to find out about the behavior of our estimators when σ_δ^2 is small. Of course, σ_δ^2 should not be too small. It should be large enough so that the measurement errors cannot be neglected and the naive OLS estimator would give wrong results. On the other hand, it should be small enough so that our approximations can be applied. Such situations are possible as testified in our simulation study and may also occur in practice. An example is studied in Section 7.

For small σ_δ^2 it turns out that, surprisingly, $\Sigma_{\mathrm{ALS}} = \Sigma_{\mathrm{SLS}}$ up to the order of σ_δ^2, whereas Σ_{OLS} differs clearly from Σ_{ALS} and Σ_{SLS}. We can, however, not say that $\Sigma_{\mathrm{OLS}} < \Sigma_{\mathrm{ALS}}$. In fact, there are linear combinations $\alpha = a'\beta$ for which the variance of $\hat\alpha_{\mathrm{OLS}}$ is larger than the variance of $\hat\alpha_{\mathrm{SLS}}$. On the other hand, for the components of $\Phi\beta$, where $\Phi = E\zeta\zeta'$, we can say that the OLS estimator has smaller variance than the ALS (or SLS) estimator if σ_δ^2 is small enough.

In the case where σ_ε^2 and σ_δ^2 are both small it turns out, again surprisingly, that $\Sigma_{\mathrm{OLS}} = \Sigma_{\mathrm{ALS}}$ up to the order of σ_δ^2, whereas now Σ_{SLS} differs from Σ_{ALS} and Σ_{OLS}, and $\hat\beta_{\mathrm{SLS}}$ turns out to be the best estimator for small σ_δ.

We not only have these qualitative results, we also have explicit quantitative formulas for the covariance matrices (see (10), (17), (26), (28)) and for the bias of OLS (see (9)) up to the order of σ_δ^2. These formulas can be used to compute approximately these covariance matrices. Simulations show that these approximations are fairly accurate for small error variances. It might be advantageous, though, to replace the matrix $\Phi = E\zeta\zeta'$ with Ezz' in these computations, but we did not try this modification.

Acknowledgements

We thank Chi-Lun Cheng from Academia Sinica, Taipei, and Thomas Augustin from the Department of Statistics, University of Munich, for valuable discussions. We also thank two anonymous referees for their suggestions to improve the presentation of the paper. Financial support from Deutsche Forschungsgemeinschaft is gratefully acknowledged.

Appendix A. Proofs

Proof of Proposition 2.1. Eq. (6) has already been proved. In order to prove (7), we introduce the abbreviation $S=(1/n)\sum_1^n z_i z_i'$ and consider, using (5),

$$\begin{aligned}\sqrt{n}(\hat\beta_{\text{OLS}}-\beta-b) &= S^{-1}\frac{1}{\sqrt{n}}\sum_1^n\{z_i\varepsilon_i - z_i(z_i-\zeta_i)'\beta - z_i z_i' b\}\\ &= S^{-1}\frac{1}{\sqrt{n}}\sum_1^n\{z_i\varepsilon_i - z_i(z_i-\zeta_i)'\beta - Ezz'b-(z_iz_i'-Ezz')b\}\\ &= S^{-1}\frac{1}{\sqrt{n}}\sum_1^n[z_i\varepsilon_i - \{z_i(z_i-\zeta_i)' - Ez(z-\zeta)'\}\beta\\ &\quad -(z_iz_i'-Ezz')b],\end{aligned}$$

where we used (6) in the last equation. Proposition 2.1 is then a consequence of the CLT with

$$\begin{aligned}\Sigma_{\text{OLS}} = {} & (Ezz')^{-1}[E\{(z-\zeta)'\beta\}^2 zz' - Ez(z-\zeta)'\beta\beta'E(z-\zeta)z'\\ & +2E\{(z-\zeta)'\beta\}(z'b)zz' - 2Ez(z-\zeta)'\beta b'Ezz'\\ & +E(z'b)^2zz' - Ezz'bb'Ezz'](Ezz')^{-1}\\ & +\sigma_\varepsilon^2(Ezz')^{-1}.\end{aligned}$$

With the help of (6) this can be simplified to (7). □

Proof of Proposition 2.2. We start from (8) and note that the components of the vector $\mathrm{O}(\delta^3)$ are actually polynomials in δ with coefficients being polynomials in ξ. This can also be said of all the following terms of the form $\mathrm{O}(\delta^m)$. As δ is assumed to be normally distributed and ξ has finite moments, $E[\mathrm{O}(\delta^m)]$ always exists. Now (8) implies

$$zz' = \zeta\zeta' + \delta\left(\zeta\frac{d\zeta'}{d\xi} + \frac{d\zeta}{d\xi}\zeta'\right) + \frac{\delta^2}{2}\frac{d^2(\zeta\zeta')}{d\xi^2} + \mathrm{O}(\delta^3), \tag{A.1}$$

where we used the identity

$$\frac{d^2(\zeta\zeta')}{d\xi^2} = 2\frac{d\zeta}{d\xi}\frac{d\zeta'}{d\xi} + \zeta\frac{d^2\zeta'}{d\xi^2} + \frac{d^2\zeta}{d\xi^2}\zeta'. \tag{A.2}$$

It follows that

$$Ezz' = \Phi + \frac{1}{2}\sigma_\delta^2 E\frac{d^2(\zeta\zeta')}{d\xi^2} + \mathrm{O}(\sigma_\delta^4) \tag{A.3}$$

because terms of the form $\xi^j\delta^3$ have expectation zero and therefore only terms of the form $\xi^j\delta^m$, $m \geqslant 4$, contribute to $E[\mathrm{O}(\delta^3)]$.

Similarly by (8),

$$z(z-\zeta)' = \delta\zeta \frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} + \delta^2\left(\frac{\mathrm{d}\zeta}{\mathrm{d}\xi}\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} + \frac{1}{2}\zeta\frac{\mathrm{d}^2\zeta'}{\mathrm{d}\xi^2}\right) + \mathrm{O}(\delta^3)$$

and thus

$$Ez(z-\zeta)' = \sigma_\delta^2 E\left(\frac{\mathrm{d}\zeta}{\mathrm{d}\xi}\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} + \frac{1}{2}\zeta\frac{\mathrm{d}^2\zeta'}{\mathrm{d}\xi^2}\right) + \mathrm{O}(\sigma_\delta^4). \tag{A.4}$$

From (6), (A.3) and (A.4) we finally get (9).

Turning to the covariance matrix (7), we first note that by dropping the term $z'b$ the right-hand side of (7) will change only by a term of order $\mathrm{O}(\sigma_\delta^4)$ because, due to (9), $b = \mathrm{O}(\sigma_\delta^2)$ and, due to (8), $z-\zeta$ is of the form $a\delta + \mathrm{O}(\delta^2)$. Now again by (8) and using (A.1),

$$E\{(z-\zeta)'\beta\}^2 zz' = \sigma_\delta^2 E\left(\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi}\beta\right)^2 \zeta\zeta' + \mathrm{O}(\sigma_\delta^4). \tag{A.5}$$

With the help of (A.3) and (A.5) we finally get the desired result (10). □

Note that Φ is p.d. because ξ has a density. We used the following general result, which holds true for any nonsingular matrix A and arbitrary square matrix B and for $a \to 0$, cf. e.g. Dhrymes (1984, Corollary 41):

$$(A + aB)^{-1} = A^{-1} - aA^{-1}BA^{-1} + \mathrm{O}(a^2).$$

Proof of Proposition 3.1. We obviously have $E(H|\xi) = \zeta\zeta'$ and thus

$$EH = E\zeta\zeta' = \Phi. \tag{A.6}$$

We further have $h = yt = t\zeta'\beta + \varepsilon t$, which implies

$$H\beta - h = (H - t\zeta')\beta - \varepsilon t,$$

and thus

$$E(H\beta - h)(H\beta - h)' = E(H - t\zeta')\beta\beta'(H - \zeta t') + \sigma_\varepsilon^2 Ett'.$$

Substituting this expression and (A.6) in (14), we get the desired result. □

Proof of Proposition 3.2. Eq. (16) together with (8) and a corresponding expansion for $\mathrm{d}^2z/\mathrm{d}x^2$ yield the expansion

$$t = \zeta + \frac{\mathrm{d}\zeta}{\mathrm{d}\xi}\delta + \frac{1}{2}\frac{\mathrm{d}^2\zeta}{\mathrm{d}\xi^2}(\delta^2 - \sigma_\delta^2) + R_1 \tag{A.7}$$

with $R_1 = \sigma_\delta^2 \mathrm{O}(\delta) + \mathrm{O}(\delta^3) + \mathrm{O}(\sigma_\delta^4)$, where $\mathrm{O}(\sigma_\delta^4)$ is as in (16) and the terms $\mathrm{O}(\delta)$ and $\mathrm{O}(\delta^3)$ are as before, see proof of Proposition 2.2. As a consequence, $ER_1 = \mathrm{O}(\sigma_\delta^4)$. All the following remainder terms R_j (up to $j = 12$) will be of the same form. Because the elements of H consist of the polynomials $t_r(x)$, we have an analogous expansion for H:

$$H = \zeta\zeta' + \frac{\mathrm{d}(\zeta\zeta')}{\mathrm{d}\xi}\,\delta + \frac{1}{2}\,\frac{\mathrm{d}^2(\zeta\zeta')}{\mathrm{d}\xi^2}(\delta^2 - \sigma_\delta^2) + R_2.$$

From (A.7) it follows that

$$t\zeta' = \zeta\zeta' + \frac{\mathrm{d}\zeta}{\mathrm{d}\xi}\,\zeta'\delta + \frac{1}{2}\,\frac{\mathrm{d}^2\zeta}{\mathrm{d}\xi^2}\,\zeta'(\delta^2 - \sigma_\delta^2) + R_3$$

and thus

$$H - t\zeta' = \zeta\,\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi}\,\delta + C(\delta^2 - \sigma_\delta^2) + R_4,$$

where C is a square matrix. It follows that

$$E(H - t\zeta')\beta\beta'(H - \zeta t') = \sigma_\delta^2 E\left(\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi}\,\beta\right)^2 \zeta\zeta' + \mathrm{O}(\sigma_\delta^4). \tag{A.8}$$

From (A.7) it also follows that

$$Ett' = E\zeta\zeta' + \sigma_\delta^2 E\,\frac{\mathrm{d}\zeta}{\mathrm{d}\xi}\,\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} + \mathrm{O}(\sigma_\delta^4). \tag{A.9}$$

Substituting (A.8) and (A.9) in (15), we finally get the desired result (17). □

Proof of Propositions 4.1 and 4.2. We only sketch the proof. A more rigorous proof is found in Kukush and Schneeweiss (2000).

Let $\gamma = (\mu_x, \sigma_x^{-2})'$ be the vector of the nuisance parameters and $\hat\gamma = (\bar{x}, s_x^{-2})'$ its estimate.

The expressions m, v and μ, defined by (18), (19), and (22), respectively, all depend on γ, although this has not been made explicit in the notation. They are the building blocks of the quasi score function on which the estimating equation of $\hat\beta_{\mathrm{SLS}}$ is based, cf. Section 4.1. We therefore write this quasi score function more explicitly in the form

$$s(\beta, \sigma_\varepsilon^2, \gamma) = \frac{y - m(x, \beta, \gamma)}{v(x, \beta, \sigma_\varepsilon^2, \gamma)}\,\mu(x, \gamma).$$

The estimating equation of Section 4.1 can than be written as

$$\bar{s}(\hat\beta, \hat\sigma_\varepsilon^2, \hat\gamma) = 0,$$

where the bar over s indicates the formation of an arithmetic mean from a sample of i.i.d. observations (x_i, y_i), $i = 1, \ldots, n$. The denotation "SLS" has been omitted.

We now expand $\bar{s}(\hat\beta, \hat\sigma_\varepsilon^2, \hat\gamma)$ into a Taylor series around the true parameter point $(\beta', \sigma_\varepsilon^2, \gamma')$ omitting terms of higher order:

$$\bar{s}(\hat\beta, \hat\sigma_\varepsilon^2, \hat\gamma) \approx \bar{s}(\beta, \sigma_\varepsilon^2, \gamma) + \frac{\partial \bar{s}}{\partial \beta'}(\hat\beta - \beta) + \frac{\partial \bar{s}}{\partial \sigma_\varepsilon^2}(\hat\sigma_\varepsilon^2 - \sigma_\varepsilon^2) + \frac{\partial \bar{s}}{\partial \gamma'}(\hat\gamma - \gamma),$$

where the derivatives are taken at $(\beta', \sigma_\varepsilon^2, \gamma')$.

For the various terms of this expansion we have

$$\sqrt{n}\bar{s} \to \mathrm{N}(0, \Sigma_s), \quad \Sigma_s = Ess' = Ev^{-1}\mu\mu',$$

$$\sqrt{n}(\hat{\gamma} - \gamma) \to \mathrm{N}(0, \Sigma_{\hat{\gamma}}), \quad \Sigma_{\hat{\gamma}} = \mathrm{diag}(\sigma_x^2, 2\sigma_x^{-4}),$$

$$\sqrt{n}(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2) \to \mathrm{N}(0, \Sigma_{\hat{\sigma}_\varepsilon^2}),$$

$$\frac{\partial \bar{s}}{\partial \beta'} \to E\,\frac{\partial s}{\partial \beta'} = -E\left(v^{-1}\mu\,\frac{\partial m}{\partial \beta'}\right) = -Ev^{-1}\mu\mu',$$

$$\frac{\partial \bar{s}}{\partial \gamma'} \to E\,\frac{\partial s}{\partial \gamma'} = -E\left(v^{-1}\mu\,\frac{\partial m}{\partial \gamma'}\right) = -Ev^{-1}\mu\beta'\,\frac{\partial \mu}{\partial \gamma'},$$

$$\frac{\partial \bar{s}}{\partial \sigma_\varepsilon^2} \to E\,\frac{\partial s}{\partial \sigma_\varepsilon^2} = 0,$$

$$Es(\hat{\gamma} - \gamma)' = EE\{(y - m)v^{-1}\mu(\hat{\gamma} - \gamma)' \,|\, x\} = 0.$$

It then follows that

$$\sqrt{n}(\hat{\beta} - \beta) \to \mathrm{N}(0, \Sigma_{\mathrm{SLS}}),$$

$$\Sigma_{\mathrm{SLS}} = \left(E\,\frac{\partial s}{\partial \beta'}\right)^{-1}\left(\Sigma_s + E\,\frac{\partial s}{\partial \gamma'}\,\Sigma_\gamma E\,\frac{\partial s'}{\partial \gamma}\right)\left(E\,\frac{\partial s'}{\partial \beta}\right)^{-1}$$

$$= \left(E\,\frac{\mu\mu'}{v}\right)^{-1} + \left(E\,\frac{\mu\mu'}{v}\right)^{-1}\left(\sigma_x^2 F_1 F_1' + \frac{2}{\sigma_x^4}\,F_2 F_2'\right)\left(E\,\frac{\mu\mu'}{v}\right)^{-1}$$

with

$$F_j = -Ev^{-1}\mu\,\frac{\partial \mu'}{\partial \gamma_j}\,\beta, \quad j = 1, 2,$$

which proves (23).

Moreover, from the definition of $\mu(x) = \mu(x, \gamma)$, cf. (20)–(22), we see that $\partial\mu'/\partial\gamma_j = \mathrm{O}(\sigma_\delta^2)$ and hence

$$F_j = \mathrm{O}(\sigma_\delta^2).$$

Therefore, the second term of Σ_{SLS} is of the order σ_δ^4. This proves (24).

It turns out that (24) is also valid if $\sigma_\varepsilon^2 = \lambda\sigma_\delta^2$, $\lambda > 0$ fixed, and again $\sigma_\delta^2 \to 0$. In this case, we can use (A.14), which implies $v = \mathrm{O}(\sigma_\delta^2)$ and hence $F_j = \mathrm{O}(1)$. But now $(E\mu\mu'/v)^{-1} = \mathrm{O}(\sigma_\delta^2)$; see (A.15). Therefore, the second term of Σ_{SLS} is again of order σ_δ^4 and (24) holds true. □

Proof of the Theorem. In order to derive an expansion of Σ_{SLS} analogous to (17) for small σ_δ we have to expand all the constituents of Σ_{SLS}. We start by expanding

(20)–(22), using $\mu_x = \mu_\xi$ and $\sigma_x^2 = \sigma_\xi^2 + \sigma_\delta^2$:

$$\tau^2 = \sigma_\delta^2 + \mathrm{O}(\sigma_\delta^4),$$

$$\mu_1(x) = \xi + \delta - \frac{\sigma_\delta^2}{\sigma_\xi^2}(\xi - \mu_\xi) + R_5,$$

$$\begin{aligned} \mu_r(x) &= \mu_1(x)^r + \binom{r}{2} \tau^2 \mu_1(x)^{r-2} + R_6 \\ &= \xi^r + r\xi^{r-1}\delta + \binom{r}{2} \xi^{r-2}\delta^2 - r\xi^{r-1} \frac{\sigma_\delta^2}{\sigma_\xi^2}(\xi - \mu_\xi) \\ &\quad + \binom{r}{2} \sigma_\delta^2 \xi^{r-2} + R_7, \end{aligned}$$

where R_5, R_6, R_7 (and the following R_j as well) are of the same form as, e.g., R_1 in (A.7). The latter equation, for $r = 0, \ldots, k$, can also be written in vector form:

$$\mu = \mu(x) = \zeta + \delta \frac{\mathrm{d}\zeta}{\mathrm{d}\xi} + \frac{\delta^2}{2} \frac{\mathrm{d}^2\zeta}{\mathrm{d}\xi^2} - \sigma_\delta^2 \frac{\xi - \mu_\xi}{\sigma_\xi^2} \frac{\mathrm{d}\zeta}{\mathrm{d}\xi} + \frac{1}{2} \sigma_\delta^2 \frac{\mathrm{d}^2\zeta}{\mathrm{d}\xi^2} + R_8. \tag{A.10}$$

From this we get

$$\begin{aligned} \mu\mu' &= \zeta\zeta' + \delta \frac{\mathrm{d}(\zeta\zeta')}{\mathrm{d}\xi} + \frac{\delta^2}{2} \frac{\mathrm{d}^2(\zeta\zeta')}{\mathrm{d}\xi^2} \\ &\quad + \sigma_\delta^2 \left\{ \frac{1}{2} \left(\zeta \frac{\mathrm{d}^2\zeta'}{\mathrm{d}\xi^2} + \frac{\mathrm{d}^2\zeta}{\mathrm{d}\xi^2} \zeta' \right) - \frac{\xi - \mu_\xi}{\sigma_\xi^2} \frac{\mathrm{d}(\zeta\zeta')}{\mathrm{d}\xi} \right\} + R_9, \end{aligned} \tag{A.11}$$

where we used again identity (A.2) and in addition the similar identity

$$\frac{\mathrm{d}(\zeta\zeta')}{\mathrm{d}\xi} = \zeta \frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} + \frac{\mathrm{d}\zeta}{\mathrm{d}\xi} \zeta'.$$

Now the elements of $M(x)$ are the same $\mu_r(x)$ as the elements of $\mu(x)$. Therefore we have an expansion for $M(x)$ analogous to that of $\mu(x)$ in (A.10):

$$\begin{aligned} M = M(x) &= \zeta\zeta' + \delta \frac{\mathrm{d}(\zeta\zeta')}{\mathrm{d}\xi} + \frac{\delta^2}{2} \frac{\mathrm{d}^2(\zeta\zeta')}{\mathrm{d}\xi^2} - \sigma_\delta^2 \frac{\xi - \mu_\xi}{\sigma_\xi^2} \frac{\mathrm{d}(\zeta\zeta')}{\mathrm{d}\xi} \\ &\quad + \frac{1}{2} \sigma_\delta^2 \frac{\mathrm{d}^2(\zeta\zeta')}{\mathrm{d}\xi^2} + R_{10}. \end{aligned}$$

So finally, again using (A.2), we get

$$M - \mu\mu' = \sigma_\delta^2 \frac{\mathrm{d}\zeta}{\mathrm{d}\xi} \frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} + R_{11}.$$

Now substituting this result in (19), we find

$$v = v(x, \beta, \sigma_\varepsilon^2) = \sigma_\varepsilon^2 + \sigma_\delta^2 \left(\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} \beta \right)^2 + R_{12} \tag{A.12}$$

and

$$\begin{aligned} \frac{\mu\mu'}{v} &= \frac{1}{\sigma_\varepsilon^2} \mu\mu' \left\{ 1 - \frac{\sigma_\delta^2}{\sigma_\varepsilon^2} \left(\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} \beta \right)^2 \right\} + R_{13} \\ &= \frac{1}{\sigma_\varepsilon^2} \left\{ \mu\mu' - \frac{\sigma_\delta^2}{\sigma_\varepsilon^2} \left(\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} \beta \right)^2 \zeta\zeta' \right\} + R_{14}, \end{aligned}$$

where the last equation follows from (A.11). The remainder terms R_{13} and R_{14} are more complicated than the R_j before, but they still have the property that $ER_j = \mathrm{O}(\sigma_\delta^4)$. Taking expectations and using (A.11) again, we derive

$$\begin{aligned} E \frac{\mu\mu'}{v} = \frac{1}{\sigma_\varepsilon^2} \Bigg[\Phi + \frac{\sigma_\delta^2}{2} \left\{ E \left(\frac{\mathrm{d}^2(\zeta\zeta')}{\mathrm{d}\xi^2} + \zeta \frac{\mathrm{d}^2\zeta'}{\mathrm{d}\xi^2} + \frac{\mathrm{d}^2\zeta}{\mathrm{d}\xi^2} \zeta' \right) - 2A \right\} \\ - \frac{\sigma_\delta^2}{\sigma_\varepsilon^2} E \left(\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} \beta \right)^2 \zeta\zeta' \Bigg] + \mathrm{O}(\sigma_\delta^4), \end{aligned} \tag{A.13}$$

where

$$A = E \frac{\xi - \mu_\xi}{\sigma_\xi^2} \frac{\mathrm{d}(\zeta\zeta')}{\mathrm{d}\xi},$$

which will now be further analyzed.

Let $\xi_* = (\xi - \mu_\xi)/\sigma_\xi$. Then

$$E\xi_*^{r+1} = rE\xi_*^{r-1}, \quad r = 1, 2, \ldots,$$

because $\xi_* \sim \mathrm{N}(0,1)$. By binomial expansion it follows that

$$E(\xi_* + a)^r \xi_* = rE(\xi_* + a)^{r-1}.$$

Applying this to $a = \mu_\xi/\sigma_\xi$, we get

$$E\xi^r \frac{\xi_*}{\sigma_\xi} = rE\xi^{r-1} = E \frac{\mathrm{d}\xi^r}{\mathrm{d}\xi}.$$

As the elements of $\mathrm{d}(\zeta\zeta')/\mathrm{d}\xi$ are of the form ξ^r, we can apply this result to the expression A and find

$$A = E \frac{\mathrm{d}(\zeta\zeta')}{\mathrm{d}\xi} \frac{\xi_*}{\sigma_\xi} = E \frac{\mathrm{d}^2(\zeta\zeta')}{\mathrm{d}\xi^2}.$$

Substituting this expression for A in (A.13) and again using (A.2) we get

$$E \frac{\mu\mu'}{v} = \frac{1}{\sigma_\varepsilon^2} \left[\Phi - \sigma_\delta^2 E \frac{\mathrm{d}\zeta}{\mathrm{d}\xi} \frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} - \frac{\sigma_\delta^2}{\sigma_\varepsilon^2} E \left(\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi} \beta \right)^2 \zeta\zeta' \right] + \mathrm{O}(\sigma_\delta^4).$$

Inverting this expression and taking account of (24), we see that Σ_{SLS} is exactly of the same form as (17) except, of course, that the O-terms may differ. This proves the theorem. □

Proof of Proposition 6.1. As in the various approximation formulas that were used in the derivation of (10) only δ or σ_δ^2 was involved, (10) remains essentially unchanged except that the last term in braces multiplied by σ_δ^2 is $\mathrm{O}(\sigma_\delta^4)$ and can therefore be dropped. We thus have

$$\Sigma_{\mathrm{OLS}} = \sigma_\delta^2 \Phi^{-1} \left\{ \lambda\Phi + E\left(\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi}\beta\right)^2 \zeta\zeta' \right\} \Phi^{-1} + \mathrm{O}(\sigma_\delta^4).$$

If we substitute the term v_1, as given in (27), we can write Σ_{OLS} as stated in the proposition. □

Proof of Proposition 6.2. Just as in the proof of Proposition 6.1 we see that (17) remains essentially unchanged except that the last term in braces multiplied by σ_δ^2 is $\mathrm{O}(\sigma_\delta^4)$ and can therefore be dropped. Consequently, by (26), Σ_{ALS} is the same as Σ_{OLS} up to $\mathrm{O}(\sigma_\delta^2)$. □

Proof of Proposition 6.3. In expansion (39) for $\mu\mu'$ we need only the first term

$$\mu\mu' = \zeta\zeta' + \mathrm{O}(\delta) + \mathrm{O}(\sigma_\delta^2).$$

Expansion (A.12) for v is taken over, but with σ_ε^2 replaced by $\lambda\sigma_\delta^2$, so that, using abbreviation (27), we have

$$v = \sigma_\delta^2 \left\{ \lambda + \left(\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi}\beta\right)^2 \right\} + R_{12} = \sigma_\delta^2 v_1 + R_{12} \tag{A.14}$$

with $ER_{12} = \mathrm{O}(\sigma_\delta^4)$. The last two equations imply

$$\frac{\mu\mu'}{v} = \frac{1}{\sigma_\delta^2}\frac{\zeta\zeta'}{v_1} + R$$

with $ER = \mathrm{O}(\sigma_\delta^2)$. Hence

$$E\frac{\mu\mu'}{v} = \frac{1}{\sigma_\delta^2}\left\{ E\frac{\zeta\zeta'}{v_1} + \mathrm{O}(\sigma_\delta^2) \right\}$$

and finally

$$\left(E\frac{\mu\mu'}{v}\right)^{-1} = \sigma_\delta^2 \left\{ \left(E\frac{\zeta\zeta'}{v_1}\right)^{-1} + \mathrm{O}(\sigma_\delta^2) \right\}$$

$$= \sigma_\delta^2 \left(E\frac{\zeta\zeta'}{v_1}\right)^{-1} + \mathrm{O}(\sigma_\delta^4). \tag{A.15}$$

Therefore by (24), which holds true also in this case (see the last paragraph in the proof of Propositions 4.1 and 4.2), (28) is an immediate consequence of (A.13). □

Proof of Proposition 6.4. The equality in Proposition 6.4 is just a reformulation of Proposition 6.2. The inequality follows immediately from (26) and (28) by applying Lemmas 1 and 2 of Appendix B with $w = \Phi^{-1/2}\zeta$ and $v = v_1$. The condition on the β_j's implies that v is not a constant.

Appendix B. A matrix inequality

Lemma B.1. *Let v be a positive random variable and w a random column (vector) in $\mathbb{R}^m$. Assume $E(\frac{1}{v}w'w) < \infty$ and $E(vw'w) < \infty$, then (in the Loewner order)*

$$E(vww') \geqslant \left[E\left(\frac{1}{v}ww'\right)\right]^{-1}. \tag{B.1}$$

Assume further that for all $x \in \mathbb{R}^m$ with $x \neq 0$ we have $P(x'w = 0) = 0$, then equality in (B.1) holds if, and only if, v is constant a.s.

Proof. First note that $E((1/v)ww')$ is p.d. and therefore invertible. Indeed, $x'E((1/v)ww')x \geqslant 0$ for any $x \in \mathbb{R}^m$, and $x'E((1/v)ww')x = 0$ implies $w'x = 0$ a.s., but then $E(x'ww'x) = x'x = 0$ and thus $x = 0$. Now let

$$q := \left[E\left(\frac{1}{v}ww'\right)\right]^{-1}\frac{w}{\sqrt{v}} - \sqrt{v}w.$$

Then

$$E(qq') = E(vww') - \left[E\left(\frac{1}{v}ww'\right)\right]^{-1},$$

which is p.s.d. This proves (B.1).

Equality of the two matrices is equivalent to $E(qq') = 0$ and thus $q = 0$ a.s., that is

$$\left[E\left(\frac{1}{v}ww'\right)\right]^{-1} w = vw \quad \text{a.s.} \tag{B.2}$$

Assume (B.2) to be true. Let $(Ev^{-1}ww')^{-1} = T'\Lambda T$ where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$ and T is orthogonal. Let $u = Tw$, then (B.2) reduces to

$$\Lambda u = vu \quad \text{a.s.}$$

In particular

$$\lambda_1 u_1 = vu_1 \quad \text{a.s.},$$

$$\lambda_2 u_2 = vu_2 \quad \text{a.s.},$$

where $u=(u_1,\ldots,u_m)$. But $P(u_1\neq 0$ and $u_2\neq 0)=1$ because $P(u_1=0$ or $u_2=0)=P(e_1'u=0$ or $e_2'u=0)\leqslant P(e_1'u=0)+P(e_2'u=0)=P(e_1'Tw=0)+P(e_2'Tw=0)=0$ by assumption, where e_i is the ith unit vector. Thus $\lambda_1=\lambda_2$. Similarly $\lambda_1=\lambda_2=\cdots\lambda_m=:\lambda$. It follows that

$$\lambda u = vu \quad \text{a.s.}$$

and thus $v=\lambda$ is a constant. Conversely, if v is a constant, then (B.1) holds with an equality sign. This proves the lemma. □

The next lemma shows that the condition formulated in Lemma A.1 is satisfied in the polynomial model.

Lemma A.2. *Let $\zeta=(1,\xi,\ldots,\xi^k)'$ and let ξ have a density. Then for all $x\in\mathbb{R}^k$ with $x\neq 0: P(x'\zeta=0)=0$.*

Proof. Suppose $P(x'\zeta=0)>0$. Then there exist infinitely many ξ_i such that

$$\sum_{j=0}^{k} x_j\xi_i^j=0, \quad i=1,2,\ldots\ .$$

This implies $x_0=\cdots=x_k=0$. □

References

Carroll, R.J., Ruppert, D., Stefanski, L.A., 1995. Measurement Error in Nonlinear Models. Chapman & Hall, London.

Cheng, C.-L., Schneeweiss, 1998. Polynomial regression with errors in the variables. J. Roy. Statist. Soc. B 60, 189–199.

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

Cheng, C.-L., Van Ness, J.W., 1999. Statistical Regression with Measurement Error. Arnold, London.

Dhrymes, P.J., 1984. Mathematics for Econometrics, 2nd Edition. Springer, New York.

Heyde, C.C., 1997. Quasi-likelihood and Application. Springer, New York.

Küchenhoff, H., Carroll, R.J., 1997. Segmented regression with errors in predictors: semi-parametric and parametric methods. Statist. Med. 16, 169–188.

Kuha, J.T., Temple, J., 2003. Covariate measurement error in quadratic regression. Internat. Statist. Rev. 71, 131–150.

Kukush, A., Schneeweiss, H., 2000. A comparison of asymptotic co-variance matrices of adjusted least squares and structural least squares in error ridden polynomial regression. Discussion Paper 218, Sonderforschungsbereich 386, University of Munich.

Kukush, A., Schneeweiss, H., Wolf, R., 2001. A comparison of three estimators in a polynomial regression with measurement errors. Discussion Paper 233, Sonderforschungsbereich 386, University of Munich.

Schneeweiss, H., Mittag, H.-J., 1986. Lineare Modelle mit Fehlerbehafteten Daten. Physica, Heidelberg.

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

Schneeweiss, H., Cheng, C.-L., Wolf, R., 2002. On the bias of structural estimation methods in a polynomial regression with measurement error when the distribution of the latent covariate is misspecified. In: Klein, I., Mittnik, S. (Eds.), Contributions to Modern Econometrics, From Data Analysis to Economic Policy, Kluwer, Boston, pp. 209–222.

Stefanski, L.A., 1989. Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. Comm. Statist. Part A—Theory Methods 18, 4335–4358.

Thamerus, M., 1998. Different nonlinear regression models with incorrectly observed covariates. In: Galata, R., Kuechenhoff, H. (Eds.), Econometrics in Theory and Practice, Festschrift for Hans Schneeweiss. Physica, Heidelberg, New York, pp. 31–44.