Kukush, Schneeweiss:

A Comparison of Asymptotic Covariance Matrices of Adjusted Least Squares and Structural Least Squares in Error Ridden Polynomial Regression

Sonderforschungsbereich 386, Paper 218 (2000)

Online unter: http://epub.ub.uni-muenchen.de/

A Comparison of Asymptotic Covariance Matrices of Adjusted Least Squares and Structural Least Squares in Error Ridden Polynomial Regression

Discussion Paper 218

Alexander Kukush, SFB 386
Institut für Statistik, Universität München

Hans Schneeweiss
Institut für Statistik, Universität München

Abstract

A polynomial structural errors-in-variables model with normal underlying distributions is investigated. An asymptotic covariance matrix of the SLS estimator is computed, including the correcting terms which appear because in the score function the sample mean and the sample variance are plugged in. The ALS estimator is also considered, which does not need any asumption on the regressor distribution. The asymptotic covariance matrices of the two estimators are compared in border cases of small and of large errors. In both situations it turns out that under the normality assumption SLS is strictly more efficient than ALS.

Keywords: Polynomial regression; structural errors-in-variables model; asymptotic covariance matrix; efficiency.

1 Introduction

We deal with the structural case of a polynomial regression with measurement errors. For the normal case, Thamerus (1998) developed a structural least squares (SLS) estimator. The method is based on a quasi-likelihood score function. We consider a certain modification of the method. We follow the recommendation of Carroll et al. (1995), p. 271, and mention that a similar procedure for a quadratic regression was used by Kuha and Temple (1999). We show the consistency of the proposed quasi-likelihood method using the consistency criterion of Aitchison and Silvey (1958), which was adopted in Heyde (1997), p. 183.

Once the consistency is established, to prove asymptotic normality is an easy exercise in calculus. However a particular feature of the SLS procedure is that we plug in the sample mean and the mean variance instead of the unknown parameters of the regressor distribution. This causes the additional terms in the asymptotic covariance matrix of the estimator. These terms are written down below, and we remark that they were neglected in Kuha and Temple (1999).

Cheng and Schneeweiss (1998) elaborated the consistent adjusted least squares (ALS) estimator for a polynomial errors-in-variables model, which corresponds to a functional case and does not need any asumption on the regressor distribution. The SLS estimator, ALS estimator and its small sample modification MALS were compared by Schneeweiss and Nittner (2000). Via simulations it was shown in particular that SLS is more efficient than ALS or MALS whenever the normality assumption holds true.

We compare the asymptotic covariance matrices of the estimators and give theoretical proof of such efficiency conclusion. The structure of the matrices is quite complicated, and we were able to handle only the border cases of small and large errors. As criterions of relative efficiency we used (for small errors) the trace of the normalized asymptotic covariance matrix and (for both cases) its determinant. The trace characterizes the second moment of the normalized estimator, while the determinant is related to the volume of the asymptotic confidence ellipsoid. To compare the determinants, we established an interesting generalization of the Cauchy-Schwarz inequality, which seems to be new.

In the next section the ALS estimator is presented. In section 3 we introduce the SLS estimator and prove its consistency. In section 4 we derive the asymptotic covariance matrix of the SLS estimator. In the next two sections the asymptotic covariance matrices of the estimators are compared for small and for large errors, respectively. The auxiliary matrix inequalities and a general expression for the asymptotic covariance matrix are given in the Appendix.

2 The ALS estimator

Throughout this paper we consider a polynomial structural errors-in-variables model

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

$$x_i = \xi_i + \delta_i, \quad i = 1, \ldots, n. \tag{2}$$

We assume that $\xi_i \sim$ i. i. d. $N(\mu_\xi, \sigma_\xi^2)$ and the errors (ϵ_i, δ_i) to be i.i.d. Gaussian, independent of the ξ_i's, with variances σ_ϵ^2 and σ_δ^2 and covariance $\sigma_{\epsilon\delta} = 0$. Then it is possible to construct polynomials $t_r(x)$ of degree r, such that $\mathrm{E}t_r(\xi + \delta_1) = \xi^r$, for any non-random ξ. Let H_i be a $(k+1) \times (k+1)$ matrix with elements $(H_i)_{rs} = t_{r+s}(x_i)$, $r, s = 0, \ldots, k$ and h_i be a $(k+1) \times 1$ vector with elements $(h_i)_r = y_i t_r(x_i)$, $r = 0, \ldots, k$. Then the (unmodified) ALS estimator $\hat{\beta}_{\mathrm{ALS}}$ of $\beta = (\beta_0, \ldots, \beta_k)'$ is given as a measurable solution of

$$\overline{H} \hat{\beta}_{\mathrm{ALS}} = \overline{h}, \tag{3}$$

where the bar denotes averages, e. g. , $\overline{H} = \frac{1}{n} \sum_{i=1}^{n} H_i$.

Note that the ALS estimator can be defined by (3) also in a *functional* polynomial errors-in-variables model, i. e. , the model (1), (2) with non-random latent variables ξ_i. In that model the estimator is consistent and asymptotically normal, see Cheng and Schneeweiss (1998).

Of course, the ALS estimator preserves these asymptotic properties in a structural model (1), (2), with random ξ_i's. In the structural case,

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

in distribution, with the asymptotic covariance matrix

$$\Sigma_{\mathrm{ALS}} = (\mathrm{E}H)^{-1} \cdot \mathrm{E}(H\beta - h)(H\beta - h)' \cdot (\mathrm{E}H)^{-1}. \tag{4}$$

Here H and h have the same distribution as H_1 and h_1.

Remark 1. In Cheng et al. (2000) a small sample modification of the ALS procedure, called MALS, was introduced. But it was shown there that ALS and MALS estimators are asymptotically equivalent, i. e., $p\lim_{n\to\infty} \sqrt{n}(\hat{\beta}_{\mathrm{MALS}} - \hat{\beta}_{\mathrm{ALS}}) = 0$. Therefore these estimators have identical asymptotic covariance matrices, and in the present asymptotic comparision we do not consider the MALS estimator.

3 The SLS estimator and its consistency

3.1 The estimating equations

In the structural model (1), (2) we suppose that σ_δ^2 is known, while σ_ϵ^2, μ_x and σ_x^2 are unknown, and $\beta = (\beta_o, \ldots, \beta_k)'$ is the parameter of interest. We find a new mean-variance model in the observable variable x by taking conditional expectations given x:

$$\begin{aligned} \mathrm{E}(y|x) &= \sum_{j=0}^{k} \beta_j \mu_j(x) \stackrel{\mathrm{df}}{=} m(x, \beta), & (5) \\ \mathrm{V}(y|x) &= \sigma_\epsilon^2 + \sum_{j,l=1}^{k} \beta_j \beta_l \{\mu_{j+l}(x) - \mu_j(x)\mu_l(x)\} \stackrel{\mathrm{df}}{=} v(x, \beta), & (6) \end{aligned}$$

where $\mu_r(x) = \mathrm{E}(\xi^r|x)$. Now, $\mathcal{L}(\xi|x) = N(\mu_1(x), \tau^2)$ with

$$\begin{aligned} \mu_1(x) &= \mu_x + (1 - \sigma_\delta^2/\sigma_x^2)(x - \mu_x), & (7) \\ \tau^2 &= \sigma_\delta^2(1 - \sigma_\delta^2/\sigma_x^2). & (8) \end{aligned}$$

Here $\mu_x = \mathrm{E}x_1 = \mu_\xi$. The conditional moments $\mu_r(x)$ are given by

$$\mu_r(x) = \sum_{j=0}^{r} \binom{r}{j} \cdot \mu_j^* \cdot \mu_1(x)^{r-j}, \tag{9}$$

where $\mu_0^* = 1$, and for $j = 1, 2, \ldots$

$$\mu_j^* = \mathrm{E}[\{\xi - \mu_1(x)\}^j|x] = \begin{cases} 0 & \text{, if } j \text{ is odd} \\ 1 \cdot 3 \ldots (j-1)\tau^j & \text{, if } j \text{ is even.} \end{cases} \tag{10}$$

The nuisance parameters μ_x and σ_x^2 can be estimated by

$$\hat{\mu}_x = \overline{x}, \quad \hat{\sigma}_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2. \tag{11}$$

If these are substitued for μ_x and σ_x^2 in (7) to (10), estimates of $\mu_1(x)$ and τ^2 and finally of $\mu_r(x)$ arise. Replacing the $\mu_r(x)$ in (5) and (6) by their estimates $\hat{\mu}_r(x)$ and substituting the observable values x_i for the variable x, we finally get a mean variance model for the observable data with mean and variance functions

$$\begin{aligned} \hat{\mathrm{E}}(y|x = x_i) &= \sum_{j=0}^{k} \beta_j \hat{\mu}_j(x_i) \stackrel{\mathrm{df}}{=} \hat{m}(x_i, \beta), & (12) \\ \hat{\mathrm{V}}(y|x = x_i) &= \sigma_\epsilon^2 + \sum_{j,l=1}^{k} \beta_j \beta_l \{\hat{\mu}_{j+l}(x) - \hat{\mu}_j(x)\hat{\mu}_l(x)\} \\ &\stackrel{\mathrm{df}}{=} \hat{v}(x_i, \beta, \sigma_\epsilon^2). & (13) \end{aligned}$$

Quasilikelihood estimates $\hat{\beta}_{\mathrm{SLS}}$ and $\hat{\sigma}_\epsilon^2$ for β and σ_ϵ^2 are measurable solutions of the conditionally asymptotically unbiased estimating equations

$$\frac{1}{n} \sum_{i=1}^{n} \frac{y_i - \hat{m}(x_i, \beta)}{\hat{v}(x_i, \beta, \sigma_\epsilon^2)} \cdot \hat{\mu}(x_i) = 0 \tag{14}$$

$$\frac{1}{n} \sum_{i=1}^{n} \{ \frac{[y_i - \hat{m}(x_i, \beta)]^2}{\hat{v}(x_i, \beta, \sigma_\epsilon^2)} - \frac{n-k-1}{n} \} \cdot \frac{1}{\hat{v}(x_i, \beta, \sigma_\epsilon^2)} = 0, \tag{15}$$

where $\hat{\mu}(x) = (\hat{\mu}_0(x), \ldots, \hat{\mu}_k(x))'$. We follow here the recommendations of Carroll et al. (1995), p. 271, and construct the equations similarly to Kuha and Temple (1999). Now, $\hat{\beta}_{\mathrm{SLS}}$ is by definition the SLS estimator in the model (1), (2).

Remark 2. In Thamerus (1998) and Schneeweiss and Nittner (2000) instead of (15) it was proposed to update the estimate for σ_ϵ^2 using the residuals of the previous step. We do prefere (15) because we are able to write down the asymptotic covariance matrix for the solution of (14) and (15).

3.2 The algorithm

Denote the true values of parameters β and σ_ϵ^2 in the model (1), (2) by β^0 and $\sigma_{\epsilon 0}^2$. Introduce the following further assumptions for the model:

(i)	$\beta^0 \in G_\beta$	, where G_β is a given bounded open set in $I\!R^{k+1}$.
(ii)	$\sigma_{\epsilon 0}^2 \in (a_y, b_y)$	, where $0 < a_y < b_y < \infty$ and a_y, b_y are given.

We look for a solution of (14), (15) in the domain $G_\beta \times (a_y, b_y)$. The reason of such restrictions is that in (14) and (15) both score functions tend to 0, as $||\beta|| \to \infty$ or $\sigma_\epsilon^2 \to \infty$; for $\sigma_\epsilon^2 = 0$ we have a singularity in the score functions. In the latter case the expected values of the score functions might not exist, therefore, whenever there should be a solution with $\sigma_\epsilon^2 \to 0$, consistency is not guaranteed. From another point of view, the restrictions $\beta \in G_\beta$, $\sigma_\epsilon^2 \in (a_y, b_y)$ provide computational stability of the numerical procedure.

The following iterative algorithm to solve (14), (15) can be proposed, which is similar to the one from Kuha and Temple (1999).

1. Given estimates $\beta^{(j)} \in G_\beta$ from the j-th round of the algorithm, solve (15) for $\sigma_\epsilon^2 \in (a_y, b_y)$, treating $\beta^{(j)}$ as known. It is possible to use the Newton-Raphson algorithm for this purpose. Denote the solution by $\sigma_\epsilon^{2(j)}$.

2. Solve (14) for $\beta \in G_\beta$, using $\sigma_\epsilon^{2(j)}$ and $\beta^{(j)}$ for computing $\hat{v}(x_i, \beta, \sigma_\epsilon^2)$. The updated estimate $\beta^{(j+1)} \in G_\beta$ is given by a weighted least squares estimate from regressing y_i on $\hat{\mu}(x_i)$, with weights

$$w_i = [\hat{v}(x_i, \beta^{(j)}, \sigma_\epsilon^{2(j)})]^{-1}.$$

   The unweighted least squares estimate $\hat{\beta}^* \in G_\beta$, constructed from (12) alone, can be used as an initial value $\beta^{(0)}$ for β.

3.3 Existence and consistency

Lemma 1. Under the conditions (i) and (ii) from Section 3.2 the system (14), (15) has a solution $\hat{\beta}$, $\hat{\sigma}_\epsilon^2$ a. s. for all $n \geq n_0(\omega)$.

Proof. Denote by μ_{x0} and σ_{x0}^2 the true values of the parameters $\mu_x = \mu_\xi$ and σ_x^2, respectively. We define the score function G_n corresponding to the equations (14), (15):

$$G_n(\beta, \sigma_\epsilon^2; \mu_x, \sigma_x^2) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \frac{y_i - \mu'(x_i)\beta}{v(x_i,\beta,\sigma_\epsilon^2)} \cdot \mu(x_i) \\ \\ \frac{1}{n}\sum_{i=1}^n \left\{ \frac{[y_i - \mu'(x_i)\beta]^2}{v(x_i,\beta,\sigma_\epsilon^2)} - \frac{n-k-1}{n} \right\} \cdot \frac{1}{v(x_i,\beta,\sigma_\epsilon^2)} \end{pmatrix}. \qquad (16)$$

Here $\mu(x) = (\mu_0(x), \dots, \mu_k(x))'$, with $\mu_r(x)$ given in (7) to (10) but with parameter values μ_x and σ_x^2 which may be different from the true values μ_{x0} and σ_{x0}^2. The estimating equations (14), (15) can then be written as

$$G_n(\beta, \sigma_\epsilon^2; \hat{\mu_x}, \hat{\sigma_x^2}) = 0, \quad \beta \in G_\beta, \quad \sigma_\epsilon^2 \in (a_y, b_y). \tag{17}$$

Fix the finite intervals (μ_{x1}, μ_{x2}) and (a_x, b_x) containing μ_{x0} and σ_{x0}^2, respectively, with $a_x > 0$. Denote by E_0 the expectation under the condition that in the model (1), (2) $\beta = \beta^0$, $\sigma_\epsilon^2 = \sigma_{\epsilon 0}^2$, $\mu_x = \mu_{x0}$ and $\sigma_x^2 = \sigma_{x0}^2$, and let P_0 be the corresponding probability measure. Now, we list certain properties of the functions (16).

(a) P_0 – almost surely $G_n(\beta, \sigma_\epsilon^2; \mu_x, \sigma_x^2) \to G_\infty(\beta, \sigma_\epsilon^2; \mu_x, \sigma_x^2)$ uniformly in $\Theta = G_\beta \times (a_y, b_y) \times (\mu_{x1}, \mu_{x2}) \times (a_x, b_x)$, with
$G_\infty(\beta, \sigma_\epsilon^2; \mu_x, \sigma_x^2) = \lim_{n\to\infty} E_0 G_n(\beta, \sigma_\epsilon^2; \mu_x, \sigma_x^2)$.
It is easy to verify this property because we are in an i. i. d. case. Indeed, for any fixed argument $z \in \Theta$, $G_n(z) \to G_\infty(z)$, a. s. due to the LLN. Moreover, the functions G_n are equicontinuous in $z \in \Theta$, a. s. For instance, for the first component G_n^1 in (16), we have by the LLN a. s.

$$\begin{aligned} \sup_{z\in\Theta} \left\| \frac{\partial G_n^1}{\partial z'} \right\| &\le \frac{1}{n} \sum_{i=1}^{n} \sup_{z\in\Theta} \left\| \frac{\partial s_1(y_i, x_i; z)}{\partial z'} \right\| \\ &\to \mathrm{E} \sup_{z\in\Theta} \left\| \frac{\partial s_1(y_1, x_1; z)}{\partial z'} \right\|, \end{aligned} \tag{18}$$

with $s_1(y_i, x_i; z) = \frac{y_i - \mu'(x_i)\beta}{v(x_i, \beta, \sigma_\epsilon^2)} \cdot \mu(x_i)$.
Then $\sup_{n\ge1} \sup_{z\in\Theta} \|\frac{\partial G_n^1}{\partial z'}\| < \infty$ a.s., and therefore the functions $G_n^1(z)$ are equicontinuous on Θ, a. s. Similar considerations can be employed for $G_n^2(z)$. It is important also that $G_\infty(z)$ is continuous in $z \in \Theta$ a.s., see (b) below.

(b) $G_\infty(\beta, \sigma_\epsilon^2; \mu_{x0}, \sigma_{x0}^2) = (G_\infty^1, G_\infty^2)'$, with

$$G_\infty^1 = -\mathrm{E}_0 \left(\frac{\mu(x)\mu'(x)}{v(x, \beta, \sigma_\epsilon^2)} \right) \Delta\beta, \tag{19}$$

$$\begin{aligned} G_\infty^2 &= -\mathrm{E}_0 [v^{-2}(x, \beta, \sigma_\epsilon^2) \cdot (\sigma_\epsilon^2 - \sigma_{\epsilon 0}^2 \\ &+ \sum_{j,l=1}^{k} \Delta(\beta_j \beta_l) \{\mu_{j+l}(x) - \mu_j(x)\mu_l(x)\} - (\mu'(x)\Delta\beta)^2)] \end{aligned} \tag{20}$$

Here in (19), (20) $\mu(x)$ is given via true values μ_{x0} and σ_{x0}^2, and $\Delta\beta = \beta - \beta^0$, $\Delta(\beta_j\beta_l) = \beta_j\beta_l - \beta_j^0\beta_l^0$.

(c) The matrix

$$S_0 = \left. \frac{\partial G_\infty(\theta; \mu_{x0}, \sigma_{x0}^2)}{\partial \theta'} \right|_{\theta=\theta_0} \tag{21}$$

is nonsingular, where $\theta = (\beta', \sigma_\epsilon^2)'$ and $\theta_0 = ({\beta^0}', \sigma_{\epsilon 0}^2)'$.
This property follows from the relations

$$\begin{aligned} \frac{\partial G_\infty^1(\theta_0)}{\partial \beta'} &= -\mathrm{E}_0 \frac{\mu(x)\mu'(x)}{v(x, \beta^0, \sigma_{\epsilon 0}^2)}, \quad \frac{\partial G_\infty^1(\theta_0)}{\partial \sigma_\epsilon^2} = 0, \\ \frac{\partial G_\infty^2(\theta_0)}{\partial \sigma_\epsilon^2} &= -\mathrm{E}_0 \frac{1}{v^2(x, \beta^0, \sigma_{\epsilon 0}^2)}. \end{aligned}$$

Now, S_0 is nonsingular because $\frac{\partial G_\infty^1(\theta_0)}{\partial \beta'}$ is negative definite, $\frac{\partial G_\infty^1(\theta_0)}{\partial \sigma_\epsilon^2} = 0$, and $\frac{\partial G_\infty^2(\theta_0)}{\partial \sigma_\epsilon^2} < 0$.

Now, we apply Theorem 12.1 from Heyde (1997) to the sequence (17) of estimating functions. Set

$$H_n(\theta) = -S_0^{-1} \cdot G_n(\theta; \hat{\mu}_x, \hat{\sigma}_x^2), \quad \theta \in G_\theta = G_\beta \times (a_y, b_y).$$

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

$$q_\delta = \lim_{n\to\infty} \sup(\sup_{||\theta-\theta_0||=\delta} (\theta - \theta_0)' H_n(\theta)) < 0. \tag{22}$$

Due to property a) and because of the consistency of the estimators $\hat{\mu}_x$, $\hat{\sigma}_x^2$ we have

$$q_\delta = \sup_{||\theta-\theta_0||=\delta} (\theta - \theta_0)' \cdot (-S_0^{-1} \cdot G_\infty(\theta; \mu_{x0}, \sigma_{x0}^2)). \tag{23}$$

Now, $G_\infty(\theta_0; \mu_{x0}, \sigma_{x0}^2) = 0$ which is easily seen from (19), (20), and from the definition (21) of S_0 we get the expansion

$$(\theta - \theta_0)' \cdot (-S_0^{-1} \cdot G_\infty(\theta; \mu_{x0}, \sigma_{x0}^2)) = -||\theta - \theta_0||^2 + o(||\theta - \theta_0||^2),$$

as $\theta \to \theta_0$. From (23) we obtain that for all small $\delta > 0$, the inequality (22) holds. And by the above mentioned theorem from Heyde (1997) the equation $H_n(\theta) = 0$ has a solution, for all $n \geq n_0(w)$. This proves Lemma 1.

Now we can give a more rigorous definition of the SLS estimator. For those (small) n for which (14), (15) has no solutions we set $\hat{\beta}_{\text{SLS}} = \beta_f$, $\hat{\sigma}_\epsilon^2 = \sigma_{\epsilon f}^2$, where $\beta_f \in G_\beta$ and $\sigma_{\epsilon f}^2 \in (a_y, b_y)$ are arbitrary but fixed values. If n is such that (14), (15) has many solutions we choose one of them for every ω in such a way that $\hat{\beta}_{\text{SLS}}(\omega)$ and $\hat{\sigma}_\epsilon^2(\omega)$ are measurable. This is possible due to, e. g. , Pfanzagl (1969).

Theorem 1. Under the conditions (i), (ii) the estimators $\hat{\beta}_{\text{SLS}}$ and $\hat{\sigma}_\epsilon^2$ are strongly consistent, i.e. , P_0 – a. s. $\hat{\beta}_{\text{SLS}} \to \beta^0$ and $\hat{\sigma}_\epsilon^2 \to \sigma_{\epsilon 0}^2$, as $n \to \infty$.

Proof. By Lemma 1 the estimators are well defined by the estimating equations (17). Owing to property a) in the proof of Lemma 1 and because of the strong consistency of $\hat{\mu}_x$ and $\hat{\sigma}_x^2$, there is a set Ω_0 of probability 1 where $G_n(\beta; \sigma_\epsilon^2; \hat{\mu}_x, \hat{\sigma}_x^2) \to G_\infty(\beta; \sigma_\epsilon^2; \mu_{x0}, \sigma_{x0}^2)$ uniformly in $\beta \in cl(G_\beta)$, $\sigma_\epsilon^2 \in [a_y, b_y]$. Here $cl(G)$ is the closure of a set G. Fix $\omega \in \Omega_0$. The sequence $(\hat{\beta}_n(\omega))), \hat{\sigma}_{\epsilon n}^2(\omega)$ is in a bounded domain $G_\beta \times (a_y, b_y)$. Consider an arbitrary convergent subsequence $(\hat{\beta}_{n(k)}(\omega),\ \hat{\sigma}_{\epsilon n(k)}^2(\omega)) \to (\beta_*, \sigma_{\epsilon *}^2) \in I\!R^{k+1} \times (0, +\infty)$. Then $G_\infty(\beta_*, \sigma_{\epsilon *}^2; \mu_{x0}, \sigma_{x0}^2) = 0$ and hence $\beta_* = \beta^0$, $\sigma_{\epsilon *}^2 = \sigma_{\epsilon 0}^2$ because obviously $(\beta^0, \sigma_{\epsilon 0}^2)$ is the unique solution of $G_\infty(\beta; \sigma_\epsilon^2; \mu_{x0}, \sigma_{x0}^2) = 0$, see (19) – (20). This implies the convergence of the whole sequence $(\hat{\beta}_n(\omega)), \hat{\sigma}_{\epsilon n}^2(\omega))$ to the true value $(\beta^0, \sigma_{\epsilon 0}^2)$. Theorem 1 is proved.

4 The asymptotic covariance matrix of the SLS estimator

We apply Lemma 2 from the Appendix to the estimating equations (14), (15). The estimated parameter is $\theta = (\beta', \sigma_\epsilon^2)'$. As a nuisance parameter we shall consider $\gamma = (\mu_x, 1/\sigma_x^2)' = (\gamma_1, \gamma_2)'$. Denote $\gamma_0 = (\mu_{x0}, \frac{1}{\sigma_{x0}^2})$,

$$\Phi_{11} = \mathrm{E}_0 \frac{\mu(x_i, \gamma_0)\mu'(x_i, \gamma_0)}{v(x_i, \beta^0, \sigma_{\epsilon 0}^2)}$$

and

$$F_p = \mathrm{E}_0 \left(v^{-1}(x_i, \beta^0, \sigma^2_{\epsilon 0} \cdot \mu \frac{\partial \mu'}{\partial \gamma_p} \right) \beta^0; \quad p = 1, 2,$$

where $\mu = \mu(x_i, \gamma_0)$.

Theorem 2. Under the conditions (i), (ii) from Section 3.2, $\sqrt{n}(\hat{\beta}_{SLS} - \beta_0) \xrightarrow{d} N(0, \Sigma_{SLS})$, with

$$\Sigma_{\mathrm{SLS}} = \Phi_{11}^{-1} + \Phi_{11}^{-1} \left(\sigma^2_{x0} F_1 F_1' + \frac{2}{\sigma^4_{x0}} F_2 F_2' \right) \Phi_{11}^{-1}.$$

Proof. The compound score function (16) can be written as

$$S_n(\theta, \gamma) = G_n(\theta; \mu_x, \sigma_x^2) = \frac{1}{n} \sum_{i=1}^{n} s(x_i, y_i) + r_n, \tag{24}$$

with

$$s(x, y) = \begin{pmatrix} \frac{y - \mu'(x)\beta}{v(x, \beta, \sigma_\epsilon^2)} \cdot \mu(x) \\ \\ \frac{[y - \mu'(x)\beta]^2}{v^2(x, \beta, \sigma_\epsilon^2)} - \frac{1}{v(x, \beta, \sigma_\epsilon^2)} \end{pmatrix} \tag{25}$$

and r_n a $(r+2)$-vector with first $r+1$ components equal to zero and last component equal to

$$\tau_n = \frac{k+1}{n^2} \sum_{i=1}^{n} \frac{1}{v(x_i, \beta, \sigma_\epsilon^2)} = \frac{O_P(1)}{n\sqrt{n}}, \tag{26}$$

for each θ and γ. The estimators $\hat{\beta}_{\mathrm{SLS}}$ and $\hat{\sigma}_\epsilon^2$ satisfy the equation

$$S_n(\hat{\beta}_{\mathrm{SLS}}, \hat{\sigma}_\epsilon^2; \hat{\mu}_x, \hat{\sigma}_x^{-2}) = 0.$$

Let us check the conditions of Lemma 2. By Theorem 1, $\hat{\theta} = \left(\hat{\beta}'_{\mathrm{SLS}}, \hat{\sigma}_\epsilon^2 \right)'$ is consistent. The random field $S_n(\theta, \gamma)$, $\theta \in G_\beta \times (a_y, b_y)$, $\gamma \in G_\gamma = (\mu_{x1}, \mu_{x2}) \times (\frac{1}{b_x}, \frac{1}{a_x})$, has C^1-smooth paths, a. s. Consider condition c) of Lemma 2. Set $z_i = x_i - \mu_x$. We have under P_{μ_x, σ_x^2}:

$$\begin{aligned} \hat{\mu}_x &= \frac{1}{n} \sum_{i=1}^{n} x_i, \\ \hat{\sigma}_x^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (z_i - \overline{z})^2 = \frac{1}{n} \sum_{i=1}^{n} z_i^2 - \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} z_i z_j \\ &= \frac{1}{n} \sum_{i=1}^{n} z_i^2 + \frac{O_P(1)}{n}, \end{aligned} \tag{27}$$

because

$$\mathrm{E} \left| \frac{1}{n} \sum_{1 \le i < j \le n} z_i z_j \right|^2 = \frac{1}{n^2} \sum_{1 \le i < j \le n} \mathrm{E}(z_i^2 z_j^2) \le const.$$

Then

$$\hat{\sigma}_x^{-2} - \sigma_x^{-2} = -\frac{\hat{\sigma}_x^2 - \sigma_x^2}{\sigma_x^2 \hat{\sigma}_x^2} = -\frac{1}{\sigma_x^4} (\bar{z^2} - \sigma_x^2) + \frac{O_P(1)}{n}. \tag{28}$$

From (24), (26) – (28) we get

$$\begin{pmatrix} \sqrt{n}S_n(\theta_0,\gamma_0) \\ \\ \sqrt{n}(\hat{\gamma}_n-\gamma_0) \end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} s(x_i,y_i,\gamma_0) \\ \\ x_i-\mu_{x0} \\ \\ -\frac{z_i^2-\sigma_{x0}^2}{\sigma_{x0}^4} \end{pmatrix} + \frac{O_P(1)}{\sqrt{n}}. \tag{29}$$

Note that $\mathrm{E}_0 s(x_i,y_i;\gamma_0) = \mathrm{E}_0(\mathrm{E}_0(s(x_i,y_i)|x_i)) = 0$. From (29) we have by the CLT for i. i. d. random vectors:

$$(\sqrt{n}S_n'(\theta_0,\gamma_0), \sqrt{n}(\hat{\gamma}_n-\gamma_0))' \xrightarrow{d} N(0,\Sigma),$$

with

$$\Sigma = cov\begin{pmatrix} s(x_i,y_i) \\ \\ x_i-\mu_{x0} \\ \\ -\frac{(x_i-\mu_{x0})^2-\sigma_{x0}^2}{\sigma_{x0}^4} \end{pmatrix},$$

$$\Sigma = \mathrm{diag}(\mathrm{E}_0\frac{\mu(x_i,\gamma_0)\cdot\mu'(x_i,\gamma_0)}{v(x_i,\theta_0,\gamma_0)}, 2\mathrm{E}_0 v^{-2}(x_i,\theta_0,\gamma_0), \sigma_{x0}^2, 2\sigma_{x0}^{-4}). \tag{30}$$

Now, pass to condition d) of Lemma 2. By the LLN

$$V_1 = \mathrm{E}_0\frac{\partial s(x_i,y_i)}{\partial(\beta',\sigma_\epsilon^2)}, \quad \text{with} \quad \gamma=\gamma_0.$$

Let $s(x,y) = (s_1(x,y)', s_2(x,y))'$. Then

$$\begin{aligned} \mathrm{E}_0\frac{\partial s_1}{\partial\beta'} &= -\mathrm{E}_0\frac{\mu(x_i,\gamma_0)\cdot\mu'(x_i,\gamma_0)}{v} = -\Phi_{11}, \\ \mathrm{E}_0\frac{\partial s_2}{\partial\beta'} &= -\mathrm{E}_0(\frac{1}{v^2}\frac{\partial v}{\partial\beta'}) = -\Phi_{12}, \\ \mathrm{E}_0\frac{\partial s_1}{\partial\sigma_\epsilon^2} &= 0, \quad E_0\frac{\partial s_2}{\partial\sigma_\epsilon^2} = -\mathrm{E}v^{-2} = -\varphi_{22}. \end{aligned}$$

Here $v = v(x_i,\theta_0,\gamma_0)$. Now, Φ_{11} is positive definite, and $\varphi_{22} > 0$. Therefore V_1 is nonsingular, and

$$V_1^{-1} = -\begin{pmatrix} \Phi_{11}^{-1} & 0 \\ \\ \frac{1}{\varphi_{22}}\Phi_{12}\Phi_{11}^{-1} & \frac{1}{\varphi_{22}} \end{pmatrix} = -\Phi_1^{-1}. \tag{31}$$

Pass to the condition e) of Lemma 2. We have

$$V_2 = \mathrm{E}_0\frac{\partial s(x_i,y_i)}{\partial\gamma'}, \quad \text{with} \quad \gamma=\gamma_0.$$

In particular

$$\mathrm{E}_0\frac{\partial s_1(x_i,y_i)}{\partial\gamma'} = -\mathrm{E}_0(\frac{1}{v}\mu\beta^{0'}\frac{\partial\mu}{\partial\gamma'}) = -\Phi_{21}, \tag{32}$$

$$\mathrm{E}_0\frac{\partial s_2(x_i,y_i)}{\partial\gamma'} = -\mathrm{E}_0(\frac{1}{v^2}\frac{\partial v}{\partial\gamma'}) = -\Phi_{22}. \tag{33}$$

Therefore

$$V_2 = -\Phi_2, \quad \Phi_2 = \begin{pmatrix} \Phi_{21} \\ \\ \Phi_{22} \end{pmatrix}.$$

Here in (32), (33) $\mu = \mu(x_i, \gamma_0)$. At last, condition f) of Lemma 2 holds because for θ and γ in the ϵ-neighbourhood of θ_0 and γ_0,

$$\left|\left|\frac{\partial S_n(\theta,\gamma)}{\partial(\theta',\gamma')} - \frac{\partial S_n(\theta_0,\gamma_0)}{\partial(\theta',\gamma')}\right|\right| \leq$$
$$\leq \sup_{(||\theta-\theta_0||\leq\epsilon, ||\gamma-\gamma_0||\leq\epsilon)} \left|\left|\frac{\partial^2 S_n(\theta,\gamma)}{\partial(\theta,\gamma)\partial(\theta',\gamma')}\right|\right| \epsilon \cdot const,$$

and

$$E_0 \sup_{(||\theta-\theta_0||\leq\epsilon, ||\gamma-\gamma_0||\leq\epsilon)} \left|\left|\frac{\partial^2 S(x_i, y_i, \theta,\gamma)}{\partial(\theta,\gamma)\partial(\theta',\gamma')}\right|\right| < \infty,$$

for sufficiently small ϵ. All the conditions of Lemma 2 hold, and

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_{\text{SLS}} - \beta_0 \\ \hat{\sigma}^2_\epsilon - \sigma^2_{\epsilon 0} \end{pmatrix} \xrightarrow{d} N(0, \Sigma_\theta),$$

$$\Sigma_\theta = \Phi_1^{-1}(I_{k+2}, -\Phi_2)\Sigma \begin{pmatrix} I_{k+2} \\ -\Phi_2' \end{pmatrix} \Phi_1'^{-1}.$$

Introduce the $(k+1)\times(k+2)$ selection matrix $P_\beta = (I_{k+1}, 0)$. According to (31) we have for the asymptotic covariance matrix Σ_{SLS} for $\hat{\beta}_{\text{SLS}}$:

$$\Sigma_{\text{SLS}} = P_\beta \Sigma_\theta P_\beta' = (\Phi_{11}^{-1}, 0)(I_{k+2}, -\Phi_2)\Sigma \begin{pmatrix} I_{k+2} \\ -\Phi_2' \end{pmatrix} \begin{pmatrix} \Phi_{11}^{-1} \\ 0 \end{pmatrix}.$$

Denote

$$\Phi_{21} = (F_1, F_2)$$

with

$$F_p = \mathrm{E}_0\left(v^{-1}\mu\frac{\partial\mu'}{\partial\gamma_p}\right)\beta; \quad p = 1, 2. \tag{34}$$

Using the block-diagonal structure (30) of Σ, we have finally

$$\Sigma_{\text{SLS}} = \Phi_{11}^{-1} + \Phi_{11}^{-1}\left(\sigma^2_{x0}F_1F_1' + \frac{2}{\sigma^4_{x0}}F_2F_2'\right)\Phi_{11}^{-1}. \tag{35}$$

Theorem 2 is proved.

Note that

$$\Phi_{11}^{-1} = \left(\mathrm{E}_0\frac{\mu(x_i,\gamma_0)\mu'(x_i,\gamma_0)}{v(x_i,\theta_0,\gamma_0)}\right)^{-1} \tag{36}$$

is exactly the asymptotic covariance matrix of the quasilikelihood and variance function estimate of β, under the condition that the nuisance parameters γ_0 are known, see Carroll et al. (1995), p. 272. The correcting summands in (35) appear because we plug in the consistent estimator $\hat{\gamma}_n$ instead of γ_0.

5 Comparison of the estimators for small errors

Before we compare the asymptotic covariance matrices (4) and (35) we show that for the linear regression the ALS and SLS approaches almost coincide.

5.1 Case of linear regression

In this subsection we consider a model (1), (2), with $k = 1$. It is a linear structural errors-in-variables model.

Denote

$$S_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2, \quad S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}).$$

From (3) it is easy to obtain that the ALS estimators are given by

$$\hat{\beta}_{1,\mathrm{ALS}} = \frac{S_{xy}}{S_{xx} - \sigma_\delta^2}, \tag{37}$$

and $\hat{\beta}_{0,\mathrm{ALS}}$ is found from the equation

$$\overline{y} = \hat{\beta}_{0,\mathrm{ALS}} + \hat{\beta}_{1,\mathrm{ALS}} \cdot \overline{x}. \tag{38}$$

Now, according to (12) in the linear case we have

$$\begin{aligned}\hat{\mathrm{E}}(y|x = x_i) &= \left(\beta_0 + \beta_1 \cdot \frac{\sigma_\delta^2}{\hat{\sigma}_x^2} \cdot \hat{\mu}_x\right) + \beta_1\left(1 - \frac{\sigma_\delta^2}{\hat{\sigma}_x^2}\right) x_i \\ &= \tilde{\beta}_0 + \tilde{\beta}_1 x_i,\end{aligned}$$

where

$$\tilde{\beta}_0 = \beta_0 + \beta_1 \frac{\sigma_\delta^2}{\hat{\sigma}_x^2}\hat{\mu}_x, \quad \tilde{\beta}_1 = \beta_1\left(1 - \frac{\sigma_\delta^2}{\hat{\sigma}_x^2}\right). \tag{39}$$

Note that according to (13) $\hat{V}(y|x = x_i)$ does not depend upon x_i. Solving (14) with respect to $\tilde{\beta}_0$, $\tilde{\beta}_1$ we get

$$\tilde{\beta}_{1,\mathrm{SLS}} = \frac{S_{xy}}{S_{xx}},$$

and from (39) and (11) it follows that

$$\hat{\beta}_{1,\mathrm{SLS}} = \frac{S_{xy}}{S_{xx}} \cdot \frac{\hat{\sigma}_x^2}{\hat{\sigma}_x^2 - \sigma_\delta^2} = \frac{S_{xy}}{S_{xx} - \frac{n-1}{n}\sigma_\delta^2}. \tag{40}$$

Similarly to (38), $\hat{\beta}_{1,\mathrm{SLS}}$ is found from the equation

$$\overline{y} = \hat{\beta}_{0,\mathrm{SLS}} + \hat{\beta}_{1,\mathrm{SLS}} \cdot \overline{x}. \tag{41}$$

Comparing (37) and (40) we obtain $\sqrt{n}(\hat{\beta}_{1,\mathrm{SLS}} - \hat{\beta}_{1,\mathrm{ALS}}) \to 0$, a. s., and from (38), (41) we have also $\sqrt{n}(\hat{\beta}_{0,\mathrm{SLS}} - \hat{\beta}_{0,\mathrm{ALS}}) \to 0$, a. s. Therefore in the linear case the asymptotic covariance matrices of the two estimators coincide.

Note that if the estimate S_{xx} instead of $\hat{\sigma}_x^2$ is used in the SLS method, then in the linear case the SLS and ALS estimators coincide.

5.2 Case of small errors in the general model

Hereafter we consider the nonlinear regression (1), (2), i. e. , $k \geq 2$. Moreover we deal with a series of models (1), (2) and suppose that the parameters

$$\beta_0, \ldots, \beta_k; \mu_\xi, \sigma_\xi^2$$

do not change, while the variances σ_ϵ^2 and σ_δ^2 may change. Consider the following asumptions.

$$\begin{array}{ll} \text{(iii)} & \sigma_\delta \to 0 \\ \text{(iv)} & \chi = \dfrac{\sigma_\delta}{\sigma_\epsilon} \leq const. \end{array}$$

So both cases are possible:

$$\begin{array}{lll} \text{a)} & \sigma_\epsilon = const, & \text{then } \chi \to 0; \\ \text{b)} & \sigma_\epsilon \to 0, & \text{then } \sigma_\delta = O(\sigma_\epsilon). \end{array}$$

5.2.1 Asymptotics for the ALS method

We analyse the asymptotic covariance matrix (4) under the assumptions (iii), (iv). Denote $\rho = (1, \xi, \ldots, \xi^k)'$, and let hereafter (ϵ, ξ, δ) have the same distribution as $(\epsilon_1, \xi_1, \delta_1)$, $x = \xi + \delta$, $y = \sum_{j=0}^k \beta_j \xi^j + \epsilon$. Then by the definition of the matrices H_i and polynomials $t_r(x)$, we have

$$\mathrm{E}H_1 = \mathrm{E}(\mathrm{E}(H_1|\xi_1)) = \mathrm{E}(\xi^{r+s}, r, s = 0, \ldots, k) = \mathrm{E}(\rho\rho'). \tag{42}$$

Now, $H_1\beta - h_1$ has the same distribution as a vector with j-th component

$$\sum_{i=1}^k \beta_i (t_{j+i}(x) - \xi^i t_j(x)) - \epsilon t_j(x), \quad j = 0, 1, \ldots, k. \tag{43}$$

We expand (43) in the space $L_2(\Omega)$ of square-integrable random variables. The leading term of (43) is

$$x^j \left(\sum_{i=1}^k \beta_i ((\xi + \delta)^i - \xi^i) - \epsilon \right), \tag{44}$$

and the leading term of (44) is

$$\xi^j \left(\sum_{i=1}^k \beta_i \cdot i\xi^{i-1}\delta - \epsilon \right) = \sigma_\epsilon \cdot \xi^j \left(\sum_{i=1}^k i\beta_i \xi^{i-1} \chi n_\delta - n_\epsilon \right), \tag{45}$$

where $n_\delta = \frac{\delta}{\sigma_\delta}$ and $n_\epsilon = \frac{\epsilon}{\sigma_\epsilon}$ are mutually independent and independent of ξ standard normal random variables. Actually, the difference between (43) and (45) divided by σ_ϵ tends to 0 in mean square. (If $\chi = \frac{\sigma_\delta}{\sigma_\epsilon} \to 0$ then the summands containing χ also vanish but it is convenient to preserve these summands to be able to consider the cases of vanishing and nonvanishing χ simultaneously). Introduce the vector $z = (z_0, \ldots, z_k)'$,

$$z_j = \xi^j \left(\sum_{i=1}^k i\beta_i \xi^{i-1} \chi n_\delta - n_\epsilon \right), \quad j = 0, \ldots, k.$$

Then according to the leading terms (45) we have

$$\sigma_\epsilon^{-2} \mathrm{E}(H_1\beta - h_1)(H_1\beta - h_1)' = \mathrm{E}zz' + o(1).$$

But

$$\mathrm{E} z_j z_s = \mathrm{E}\xi^j \cdot \xi^s \cdot \left(1 + \chi^2 \left| \sum_{i=1}^{k} i\beta_i \xi^{i-1} \right|^2 \right).$$

Therefore

$$\sigma_\epsilon^{-2}\mathrm{E}(H_1\beta - h_1)(H_1\beta - h_1)' = \mathrm{E}[\rho\rho' v_0(\xi,\beta)] + o(1), \tag{46}$$

with

$$v_0(\xi,\beta) = 1 + \chi^2 \left| \sum_{i=1}^{k} i\beta_i \xi^{i-1} \right|^2. \tag{47}$$

From (4), (42) and (46) we obtain finally that under (iii) and (iv),

$$\sigma_\epsilon^{-2} \cdot \Sigma_{ALS} = \quad (\mathrm{E}\rho\rho')^{-1} \cdot \mathrm{E}[\rho\rho' v_0(\xi,\beta)] \cdot (\mathrm{E}\rho\rho')^{-1} + o(1), \tag{48}$$

as $\sigma_\delta \to 0$, $\frac{\sigma_\delta}{\sigma_\epsilon} \le const.$

5.2.2 Asymptotics for the SLS method

We analyze the asymptotic covariance matrix (35) under the same assumptions (iii) and (iv). Start with the matrix

$$\Phi_{11} = \mathrm{E}\left(\frac{\mu(x)\mu'(x)}{v(x,\beta)} \right). \tag{49}$$

Contrary to the section 4, hereafter we suppress the index 0 in the denotation of the true values of the parameters, and $\mu(x)$ is defined in subsection 3.3. Note that now $\frac{\tau^2}{\sigma_\delta^2} \to 1$, and $\tau^2 \to 0$. The leading terms of $\mu_{j+l}(x) - \mu_j(x)\mu_l(x)$ with respect to τ (when $j \ge 2, l \ge 2$) are:

$$\begin{aligned}&\mu_1^{l+j} + \tbinom{l+j}{2}\tau^2\mu_1^{l+j-2} - (\mu_1^j + \mu_1^{j-2}\tbinom{j}{2}\tau^2) \cdot (\mu_1^l + \mu_1^{l-2}\tbinom{l}{2}\tau^2) = \\ &\tau^2\mu_1^{l+j-2}(\tbinom{j+l}{2} - \tbinom{j}{2} - \tbinom{l}{2}) + o(\tau^2).\end{aligned}$$

Thus the leading term is $jl\tau^2\mu_1^{l+j-2}(x)$, and therefore

$$[\mu_{j+l}(x) - \mu_j(x)\mu_l(x)] - jl\sigma_\delta^2\xi^{l+j-2} = o_p(\sigma_\delta^2)\ , \tag{50}$$

as (iii) holds. The expansion (50) holds also if $j = 1$ or $l = 1$. Indeed:

a) for $j = l = 1,\ \mu_2(x) - \mu_1^2(x) = \tau^2 \sim \sigma_\delta^2 = jl\sigma_\delta^2\xi^{l+j-2}$,

b) for $j = 1,\ l \ge 2$ the leading terms are

$$\begin{aligned}&\mu_1^{l+1} + \mu_1^{l-1}\tau^2\tbinom{l+1}{2} - \mu_1(\mu_1^l + \mu_1^{l-2}\tau^2\tbinom{l}{2}) = \\ &l\mu_1^{l-1}\tau^2 + o(\tau^2) = jl\sigma_\delta^2\xi^{l+j-2} + o_p(\sigma_\delta^2).\end{aligned}$$

It follows that

$$\begin{aligned}&v(x,\beta) = \sigma_\epsilon^2 + \sigma_\delta^2 \textstyle\sum_{j,l=1}^{k} jl\beta_j\beta_l\xi^{l+j-2} + o_p(\sigma_\delta^2) = \\ &\sigma_\epsilon^2(v_o(\xi,\beta) + o_p(1)),\end{aligned}$$

with v_o given in (47). By Lebesgue's theorem of majorized convergence we have

$$\sigma_\epsilon^2 \cdot \Phi_{11} = \mathrm{E}_o\left(\frac{\mu(x)\cdot\mu'(x)}{v_o(\xi,\beta)} \right) + o(1) = \mathrm{E}\frac{\rho\cdot\rho'}{v_o(\xi,\beta)} + o(1). \tag{51}$$

The first summand in (35) has the order σ_ϵ^2. We shall show now that the other summands in (35) have smaller order. We use (34) and have

$$\frac{\partial \mu_1(x)}{\partial \mu_x} = \frac{\sigma_\delta^2}{\sigma_x^2}, \ \frac{\partial \mu_r(x)}{\partial \mu_x} = \sum_{j=0}^{r} \binom{r}{j} \cdot \mu_j^*(r-j)\mu_1^{r-j-1}\frac{\sigma_\delta^2}{\sigma_x^2}. \tag{52}$$

The leading term of the summand $\Phi_{11}^{-1}(\sigma_x^2 F_1 F_1')\Phi_{11}^{-1}$ has the order σ_δ^4, which is negligible with respect to σ_ϵ^2. Now, for $\gamma_2 = \sigma_x^{-2}$ we have

$$\frac{\partial \mu_1(x)}{\partial(\sigma_x^{-2})} = -(x-\mu_x)\sigma_\delta^2, \ \frac{\partial \tau^2}{\partial(\sigma_x^{-2})} = -\sigma_\delta^4. \tag{53}$$

Therefore the leading term of $\frac{\partial \mu_r(x)}{\partial(\sigma_x^{-2})}$ has the order σ_δ^2, and again the leading term of the summand $\Phi_{11}^{-1}(\frac{2}{\sigma_x^4}F_2 F_2')\Phi_{11}^{-1}$ has the order σ_δ^4. From (35) and (51) we obtain finally:

$$\sigma_\epsilon^{-2}\Sigma_{SLS} = \left(\mathrm{E}\frac{\rho\rho'}{v_0(\xi,\beta)}\right)^{-1} + o(1), \tag{54}$$

as $\sigma_\delta \to 0, \frac{\sigma_\delta}{\sigma_\epsilon} \leq const.$

5.2.3 Comparison of Σ_{SLS} and Σ_{ALS}

Now, we compare the expansions (54) and (48). We distinguish two cases regarding $\chi = \frac{\sigma_\delta}{\sigma_\epsilon}$: when $\chi \to 0$ and when χ is separated from 0.

Theorem 2. Let the conditions (i) to (iii) hold. Assume also

(v) $\chi = \frac{\sigma_\delta}{\sigma_\epsilon} \to 0$.

Then

$$\lim_{(\sigma_\delta \to 0, \frac{\sigma_\delta}{\sigma_\epsilon} \to 0)} \sigma_\epsilon^{-2}\Sigma_{SLS} = \lim_{(\sigma_\delta \to 0, \frac{\sigma_\delta}{\sigma_\epsilon} \to 0)} \sigma_\epsilon^{-2}\Sigma_{ALS} \ ,$$

and the limit is equal to a positive definite matrix $(\mathrm{E}\rho\rho')^{-1}$.

Proof. Under (v), $v_0(\xi,\beta) \to 1$ a.s., see (47). The statement follows now from (48) and (54).

Theorem 2 is applicable in the case $\sigma_\epsilon = const$, and $\sigma_\delta \to 0$. Simulation studies made by Schneeweiss and Nittner (2000), p. 5, corroborate this statement. The corresponding case in the simulations is $\sigma_\delta^2 = 0.01$ and normally distributed design points.

Denote

$$\mathrm{S} = \mathrm{E}\rho\rho'. \tag{55}$$

Theorem 3. Let the conditions (i) to (iii) hold.

a) Under (iv),

$$\liminf_{(\sigma_\delta \to 0, \sigma_\delta = O(\sigma_\epsilon))} \sigma_\epsilon^{-2}[tr(\Sigma_{ALS}\cdot \mathrm{S}) - tr(\Sigma_{SLS}\cdot \mathrm{S})] \geq 0, \tag{56}$$

and

$$\liminf_{(\sigma_\delta \to 0, \sigma_\delta = O(\sigma_\epsilon))} [det(\sigma_\epsilon^{-2}\Sigma_{ALS}) - det(\sigma_\epsilon^{-2}\Sigma_{SLS})] \geq 0. \tag{57}$$

b) Suppose that

(vi) $k \geq 2$ and $\sum_{j=2}^{k} \beta_j^2 \neq 0$,

(vii) for positive constants c_1 and c_2,

$$c_1 \leq \chi = \frac{\sigma_\delta}{\sigma_\epsilon} \leq c_2.$$

Then (56) and (57) become strict inequalities.

c) In the particular case of b) when $\chi \to \chi_0, 0 < \chi_0 < \infty$, the following inequalities hold:

$$\lim_{(\sigma_\delta \to 0, \frac{\sigma_\delta}{\sigma_\epsilon} \to \chi_0)} \sigma_\epsilon^{-2} \cdot tr(\Sigma_{SLS} \cdot \mathrm{S}) < \lim_{(\sigma_\delta \to 0, \frac{\sigma_\delta}{\sigma_\epsilon} \to \chi_0)} \sigma_\epsilon^{-2} \cdot tr(\Sigma_{ALS} \cdot \mathrm{S}), \tag{58}$$

$$\lim_{(\sigma_\delta \to 0, \frac{\sigma_\delta}{\sigma_\epsilon} \to \chi_0)} det(\sigma_\epsilon^{-2} \cdot \Sigma_{SLS}) < \lim_{(\sigma_\delta \to 0, \frac{\sigma_\delta}{\sigma_\epsilon} \to \chi_0)} det(\sigma_\epsilon^{-2} \cdot \Sigma_{ALS}). \tag{59}$$

Proof. Part a) follows immediately from the expansions (48), (54) and Lemma 4 from the Appendix with $\alpha = -1$, see also Remark 3 and Corollary from Lemma 6. Now, the relations (58), (59) follow from Lemma 4, see also Corollary from Lemma 6. It is important in this case that under (vi) the limit function

$$\lim_{\chi \to \chi_0} v_0(\xi, \beta) = 1 + \chi_0^2 |\sum_{i=1}^{k} i\beta_i \xi^{i-1}|^2$$

has a distribution without atoms, and this provides the strict inequalities (58) and (59).

Now, prove part b). For certain $\chi_0 \in [c_1, c_2]$ we have

$$\liminf_{(\sigma_\delta \to 0, c_1 \leq \frac{\sigma_\delta}{\sigma_\epsilon} \leq c_2)} \sigma_\epsilon^{-2} [tr(\Sigma_{ALS} \cdot \mathrm{S}) - tr(\Sigma_{SLS} \cdot \mathrm{S})] =$$
$$tr(\mathrm{S}^{-\frac{1}{2}} \cdot \mathrm{E}[\rho\rho' v_0(\xi, \beta; \chi_0)] \cdot \mathrm{S}^{-\frac{1}{2}}) - tr(\mathrm{S}^{\frac{1}{2}} (\mathrm{E} \frac{\rho\rho'}{v_0(\xi, \beta; \chi_0)})^{-1} \mathrm{S}^{\frac{1}{2}}) > 0,$$

according to (58). The consideration of the determinants relies similarly on (59) and Lemma 6. Theorem 3 is proved.

Relation (56) shows that the second strong moment of the limit distribution for the normalized estimator $\sqrt{n} \cdot \mathrm{S}^{\frac{1}{2}} (\hat{\beta}_{SLS} - \beta)$ is less then the respective one for the normalized estimator $\sqrt{n} \cdot \mathrm{S}^{\frac{1}{2}} (\hat{\beta}_{ALS} - \beta)$. In this sense $\hat{\beta}_{SLS}$ is more concentrated near β than $\hat{\beta}_{ALS}$, for large n.

And relation (57) shows that the volume of the asymptotic confidence ellipsoid for the SLS estimator is less than the respective volume for the ALS estimator.

Theorem 3 means that for small errors, when both errors in the model (1), (2) have the same order and under normality assumptions, the SLS estimator is asymptotically more efficient than the other one (with respect to both the trace and the determinant criteria).

5.3 Case of large errors

In the present subsection we deal with a series of models (1),(2) and suppose that the parameters $\beta_0, \ldots, \beta_k; \mu_\xi$ do not change, while all the variances in the model may change. Consider the following assumptions.

(viii) $\sigma_\xi \to \infty$.

(ix) $\sigma_\delta \geq const > 0$, and $\sigma_\delta = o(\sigma_\xi)$.

(x) For positive constants c_1 and c_2,

$$c_1 \leq \chi = \frac{\sigma_\epsilon}{\sigma_\xi^{k-1} \cdot \sigma_\delta} \leq c_2.$$

Remark that χ in (x) has another meaning than in (iv). Under (viii) to (x), σ_ξ and σ_ϵ tend to infinity, while the measurement errors do not vanish, and σ_ξ increases quicker than σ_δ.

5.3.1 Asymptotics for the ALS method

We analyze the asymptotic covariance matrix (4) under the assumptions (viii) to (x). We use the same denotations $\rho, \epsilon, \xi, \delta, x, y$ as in subsection 5.2.3. Denote also

$$\varphi = (1, N, \ldots, N^k)', \tag{60}$$

where N is a standard normal variable. From (42) we obtain

$$det(\mathrm{EH}_1) = \prod_{i=1}^k (\sigma_\xi^2)^i det(\mathrm{E}\varphi\varphi'). \tag{61}$$

Now, we consider $\mathrm{E}(\mathrm{H}_1\beta - h_1)(\mathrm{H}_1\beta - h_1)'$. The random vector $\mathrm{H}_1\beta - h_1$ has the same distribution as a vector with $j-th$ component given in (43). The leading term in $L_2(\Omega)$ of that expression is:

$$\begin{aligned} \beta_k(t_{j+k}(x) - \xi^k t_j(x)) - \epsilon t_j(x) \quad &= x^j \beta_k((\xi+\delta)^k - \xi^k) - \epsilon x^j + rest_1 \\ &= \xi^j (k\beta_k \xi^{k-1}\delta - \epsilon) + rest_2. \end{aligned} \tag{62}$$

Here $rest_1$ and $rest_2$ have smaller order under (viii) to (x) than the other summands, and we used here that $t_r(x) = x^r + a_{r-2}x^{r-2} + \ldots$, see Cheng and Schneeweiss (1998), Lemma 1. By condition (x), ϵ has the same order as $\xi^{k-1}\delta$. Therefore the leading term in (62) is

$$v_j = N^j \sigma_\xi^{j+k-1} \cdot \sigma_\delta (k\beta_k N^{k-1} n_\delta - \chi n_\epsilon).$$

Here $N = \frac{\xi - \mu_\xi}{\sigma_\xi}$, $n_\delta = \frac{\delta}{\sigma_\delta}$ and $n_\epsilon = \frac{\epsilon}{\sigma_\epsilon}$ are independent standardized random variables. Note that

$$\mathrm{E}(v_j v_s | N) = N^j N^s \sigma_\xi^{j+k-1} \sigma_\xi^{s+k-1} \sigma_\delta^2 (k^2 \beta_k^2 N^{2k-2} + \chi^2).$$

Therefore for $v = (v_0, \ldots, v_k)'$

$$\begin{aligned} &det(\mathrm{E}(H_1\beta - h_1)(H_1\beta - h_1)') \sim det\,\mathrm{E}vv' = \\ &(\sigma_\delta^2)^{k+1} \cdot \prod_{j=0}^k (\sigma_\xi^2)^{j+k-1} \cdot det\,\mathrm{E}[\varphi\varphi' v_1(N,\beta)], \end{aligned} \tag{63}$$

where

$$v_1(N, \beta) = \chi^2 + k^2 \beta_k^2 N^{2k-2}. \tag{64}$$

From (61), (63) and (4) we get finally that under (viii) to (x),

$$det\Sigma_{ALS} \sim (\sigma_\delta^2)^{k+1} \cdot \prod_{i=0}^k (\sigma_\xi^2)^{i-1} \frac{det\mathrm{E}[\varphi\varphi' v_1(N,\beta)]}{det^2(\mathrm{E}\varphi\varphi')}. \tag{65}$$

5.3.2 Asymptotics for the SLS method

Under the conditions (viii) to (x) we analyze

$$\Sigma_{SLS} = \Phi_{11}^{-\frac{1}{2}}[I_{k+1} + \Phi_{11}^{-\frac{1}{2}}(\sigma_x^2 F_1 F_1' + \frac{2}{\sigma_x^4} F_2 F_2')\Phi_{11}^{-\frac{1}{2}}]\Phi_{11}^{-\frac{1}{2}}, \tag{66}$$

see (35). Expand firstly $det\Phi_{11}$. Consider the conditional variance $v(x,\beta)$ given in (6). We have

$$v(x,\beta) = \sigma_\epsilon^2 + \beta_k^2(\mu_{2k}(x) - \mu_k^2(x)) + rest_3, \tag{67}$$

where $rest_3$ has smaller order. Write down the leading term of $\mu_{2k} - \mu_k^2$:

$$\begin{aligned} \mu_{2k} - \mu_k^2 &= \mu_1^{2k} + \mu_1^{2k-2}\tbinom{2k}{2}\tau^2 - (\mu_1^k + \mu_1^{k-2}\tbinom{k}{2}\tau^2)^2 + rest_4 = \\ &= \mu_1^{2k-2}\tau^2[\tbinom{2k}{2} - 2\tbinom{k}{2}] + rest_5, \end{aligned}$$

$$\mu_{2k}(x) - \mu_k^2(x) = \mu_1^{2k-2}(x)\tau^2 k^2 + rest_5.$$

Here the terms in $rest_3$, $rest_4$ and $rest_5$ have smaller order than $\mu_1(x)^{2k-2} \cdot \tau^2$. Note that by (x) σ_ϵ^2 is of the same order as $\mu_1(x)^{2k-2} \cdot \tau^2$. Now,

$$v(x,\beta) = \sigma_\xi^{2k-2}\sigma_\delta^2(\chi^2 - k^2\beta_k^2 N^{2k-2} + rest_6), \tag{68}$$

where $N = \frac{\xi - \mu_\xi}{\sigma_\xi} \sim N(0,1)$, and χ is fiven in (x). Our matrix

$$\Phi_{11} = \mathrm{E}(\frac{\mu(x)}{\sqrt{v(x,\beta)}} \cdot \frac{\mu'(x)}{\sqrt{v(x,\beta)}})$$

is a Gramm matrix of random values $z_i = \frac{\mu_i(x)}{\sqrt{v(x,\beta)}}$, $i = 0, \ldots, k$, in the space $L_2(\Omega)$, and $\mu_i(x)$ has a leading term $x^i = \xi^i + rest_7 = \sigma_\xi^i N^i + rest_8$. From (68) we get

$$z_i = \frac{\sigma_\xi^i}{\sigma_\xi^{k-1}\sigma_\delta}(\frac{N^i}{\sqrt{v_1(N,\beta)}} + rest_9), \tag{69}$$

with $||rest_9||_{L_2} \to 0$, where $v_1(N,\beta) = x^2 + k^2\beta_k N^{2k-2}$. Therefore for $z = (z_0, \ldots z_k)'$

$$\begin{aligned} det\Phi_{11} &\sim det\mathrm{E}(zz') \\ &\sim \frac{1}{(\sigma_\delta^2)^{k+1}\prod_{i=0}^k(\sigma_\xi^2)^{i-1}} \cdot det\mathrm{E}[\frac{\varphi\varphi'}{v_1(N,\beta)}]. \end{aligned} \tag{70}$$

We show now that under (viii) to (x),

$$\Phi_{11}^{-\frac{1}{2}}(\sigma_x^2 F_1 F_1' + \frac{2}{\sigma_x^4} F_2 F_2')\Phi_{11}^{-\frac{1}{2}} \to 0. \tag{71}$$

We use (34). Denote

$$W = \Phi_{11}^{-\frac{1}{2}}(\frac{\mu}{\sqrt{v}}), \text{with } \mu = (\mu_0(x), \ldots, \mu_k(x))', v = v(x,\beta),$$

and

$$\psi_p = \frac{\partial\mu'}{\partial\gamma_p}\beta, p = 1, 2.$$

Then the rs-element of the matrix (71) equals to:

$$a_{rs} = \sigma_x^2 \mathrm{E}_0(\frac{\psi_1 w_r}{\sqrt{v}})\mathrm{E}_0(\frac{\psi_1 w_s}{\sqrt{v}}) + \frac{2}{\sigma_x^4}\mathrm{E}_0(\frac{\psi_2 w_r}{\sqrt{v}})\mathrm{E}_0(\frac{\psi_2 w_s}{\sqrt{v}}). \tag{72}$$

Now, $v \geq \sigma_\epsilon^2$ and $\mathrm{E}w_r^2 = 1$. Therefore, by the Cauchy-Schwarz inequality,

$$|a_{rs}| \leq \frac{\sigma_x^2}{\sigma_\epsilon^2}\mathrm{E}_0\psi_1^2 + \frac{2}{\sigma_x^4\sigma_\epsilon^2}\mathrm{E}_0\psi_2^2. \tag{73}$$

We use firstly (52) to bound $\mathrm{E}_0\psi_1^2$. The leading term of ψ_1 is $\beta_k\mu_1^{k-1}\frac{\sigma_\delta^2}{\sigma_x^2}$, and the expectation of the leading term of ψ_1^2 has the order $(\sigma_x^2)^{k-1}\frac{\sigma_\delta^4}{\sigma_x^4}$. Therefore, using (x),

$$\begin{aligned} \tfrac{\sigma_x^2}{\sigma_\epsilon^2}\mathrm{E}_0\psi_1^2 &\leq const_1\tfrac{(\sigma_x^2)^{k-2}\sigma_\delta^4}{\sigma_\epsilon^2} \\ &\leq const_2\tfrac{(\sigma_x^2)^{k-2}\sigma_\delta^4}{(\sigma_x^2)^{k-1}\sigma_\delta^2} = const_2 \cdot \tfrac{\sigma_\delta^2}{\sigma_x^2} \to 0. \end{aligned} \tag{74}$$

Now, we use (53) to bound $\mathrm{E}_0\psi_2^2$. The leading term of ψ_2 is $\beta_k\mu_1^{k-1}\frac{\partial\mu_1}{\partial(\sigma_x^{-2})} = -\beta_k\mu_1^{k-1}(x-\mu_x)\sigma_\delta^2$, and the expectation of the leading term of ψ_2^2 has the order $(\sigma_x^2)^k\sigma_\delta^4$. Therefore

$$\frac{2}{\sigma_x^4\sigma_\epsilon^2}\mathrm{E}_0\psi_2^2 \leq const_3\frac{(\sigma_x^2)^{k-2}\sigma_\delta^4}{\sigma_\epsilon^2} \to 0, \tag{75}$$

see (74). Relations (73) to (75) show that $a_{rs} \to 0$. Thus (71) holds. By (66) and (70) we get finally

$$det\Sigma_{SLS} \sim (det\Phi_{11})^{-1} \sim \frac{(\sigma_\delta^2)^{k+1}\prod_{i=0}^k(\sigma_\xi^2)^{i-1}}{det\mathrm{E}[\frac{\varphi\varphi'}{v_1(N,\beta)}]}. \tag{76}$$

5.3.3 Comparison of Σ_{SLS} and Σ_{ALS}

Now, we compare the expansions (76) and (65).

Theorem 4. Let the conditions (i) to (iii), and (viii) to (x) hold. Then:

$$\text{a)} \liminf_{(\sigma_\xi\to\infty,\sigma_\delta=o(\sigma_\xi),c_1\leq\chi\leq c_2)} \frac{det\Sigma_{ALS}}{det\Sigma_{SLS}} \geq 1; \tag{77}$$

b) if additionally $k \geq 2$ and $\beta_k \neq 0$ then strict inequality in (77) holds;

c) in the particular case of b) when $\chi = \frac{\sigma_\epsilon}{\sigma_\xi^{k-1}\sigma_\delta} \to \chi_0$ the following inequality holds:

$$\lim_{(\sigma_\xi\to\infty,\sigma_\delta=o(\sigma_\xi),\chi\to\chi_0\in(0,+\infty))} \frac{det\Sigma_{ALS}}{det\Sigma_{SLS}} > 1.$$

Proof. Due to (76) and (65), under (i) to (iii) and (viii) to (x) we have:

$$\frac{det\Sigma_{ALS}}{det\Sigma_{SLS}} \sim \frac{det\mathrm{E}[\frac{\varphi\varphi'}{v_1(N,\beta)}]\cdot det\mathrm{E}[\varphi\varphi' v_1(N,\beta)]}{det^2(\mathrm{E}\varphi\varphi')}.$$

By the Corollary of Lemma 6 from the Appendix, the last ratio is greater or equal to 1, which proves (77). Moreover for any $\chi_0 \in [c_1, c_2]$

$$\begin{aligned} &\liminf_{(\sigma_\xi\to\infty,\sigma_\delta=o(\sigma_\xi),c_1\leq\chi\leq c_2)} \tfrac{det\Sigma_{ALS}}{det\Sigma_{SLS}} \geq \\ &\tfrac{det\mathrm{E}[\frac{\varphi\varphi'}{v_1(N,\beta;x_0)}]\det\mathrm{E}[\varphi\varphi' v_1(N,\beta;x_0)]}{det^2(\mathrm{E}\varphi\varphi')}, \end{aligned} \tag{78}$$

and by the Corollary of Lemma 6 the last ratio is greater than 1 in case b), because then $v_1(N,\beta)$ has a nonatomic distribution, see (64). This proves b). In case c) we observe simply that the underlying limit exists and equals the right hand side of (78). Theorem 4 is proved.

6 Conclusion

We derived an asymptotic covariance matrix of a structural least squares estimator in a structural polynomial errors-in-variables model. In doing this, we used a general method, which may be helpful in situations when the estimates of the parameters of interest depend on nuisance parameters which have to be estimated aforehand.
Under normal assumptions, we compared the asymptotic covariance matrices of the SLS and ALS estimators in border cases of small and large errors.
If the error in the response variable is fixed and the measurement error decreases, the covariance matrices coincide asymptotically. The simulations of Schneeweiss and Nittner (2000) corroborate this theoretical finding.
If both errors tend to 0 then the SLS method is strictly more efficient with respect to two criteria. The first one involves the trace of the normalized covariance matrix, and the second one is based on the determinants. Actually, the first criterion compares the asymptotic MSE of the normalized estimators, while the second one compares the volumes of the asymptotic confidence ellipsoids.
If the error in the response variable is large, and the variance of the latent variable is also large, while the measurement error is fixed, the SLS method is strictly more efficient as well, with respect to the second criterion.
Simulations made by Schneeweiss and Nittner (2000) show that under normal assumptions and nonvanishing measurement errors, SLS is more efficient than ALS. It is an open problem to prove this theoretically.

Acknowledgement. We thank T. Augustin for useful discussions.

7 Appendix

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

Consider a sequence of random fields $\mathcal{S}_n(\theta,\gamma), n = 1, 2, \ldots$ with values in $\mathbb{R}^d, \theta \in G_\theta$ and $\gamma \in G_\gamma$, where G_θ and G_γ are open sets in $\mathbb{R}^d$ and $\mathbb{R}^k$, respectively. We suppose that $\mathcal{S}_n(\theta,\gamma)$ are score functions constructed by an observed sample.
Let $\theta_0 \in G_\theta$ and $\gamma_0 \in G_\gamma$ be the true values of the parameters. Suppose that a consistent estimator $\hat{\gamma}_n$ of γ_0 is given. We define an estimator $\hat{\theta}_n$ of θ_0 as measurable solution of the equation

$$\mathcal{S}_n(\theta,\hat{\gamma}_n) = 0, \quad \theta \in G_\theta.$$

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

Lemma 2. Let the following conditions hold.

a) $\hat{\theta}_n$ is consistent, i.e. $\hat{\theta}_n \to \theta_0$ in probability $P_{\theta_0\gamma_0}$.

b) $\mathcal{S}_n(\theta,\gamma) \in C^1(G_\theta \times G_\gamma)$, a.s.

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

d) $\frac{\partial \mathcal{S}_n(\theta_0,\gamma_0)}{\partial \theta'} \to V_1$ in $P_{\theta_0\gamma_0}$, where V_1 is a non-random nonsingular matrix.

e) $\frac{\partial \mathcal{S}_n(\theta_0,\gamma_0)}{\partial \gamma'} \to V_2$ in $P_{\theta_0\gamma_0}$, where V_2 is a non-random matrix.

f) For each $\delta > 0$,

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} P_{\theta_0 \gamma_0} = 0,$$

where

$$A \stackrel{df}{=} \{\omega : \sup_{(||\theta - \theta_0|| \leq \epsilon, ||\gamma - \gamma_0|| \leq \epsilon)} ||\frac{\partial \mathcal{S}_n(\theta, \gamma)}{\partial (\theta, \gamma)'} - \frac{\partial \mathcal{S}_n(\theta_0, \gamma_0)}{\partial (\theta, \gamma)'}|| \geq \delta\}.$$

Then $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Sigma)$, with

$$\Sigma_\theta = V_1^{-1}(I_d, V_2)\Sigma(I_d, V_2)' V_1'^{-1}.$$

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

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

Here $(\bar{\theta}_i, \bar{\gamma}_i)$ are intermediate points on the line connecting (θ_o, γ_o) and $(\hat{\theta}_n, \hat{\gamma}_n)$. Then

$$\sqrt{n} S_n(\theta_0, \gamma_o) + \frac{\partial S_n(\theta_0, \gamma_0)}{\partial \theta'}\sqrt{n}(\hat{\theta}_n - \theta_o) + \frac{\partial S_n(\theta_0, \gamma_0)}{\partial \gamma'}\sqrt{n}(\hat{\gamma}_n - \gamma_0) + R_n = 0.$$

Here

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

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

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

Then for $\delta_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$ we obtain

$$\begin{aligned} &\left(\tfrac{\partial S_n(\theta_0, \gamma_0)}{\partial \theta'} + \Lambda_n\right)\delta_n = -\sqrt{n} S_n(\theta_o, \gamma_0) - \left(\tfrac{\partial S_n(\theta_0, \gamma_0)}{\partial \gamma'} + M_n\right) \times \\ &\sqrt{n}(\hat{\gamma}_n - \gamma_0). \end{aligned} \tag{79}$$

Now, $\Lambda_n \to 0$ on probability. Indeed

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

and

$$\begin{aligned} &\limsup_{n \to \infty} P_{\theta_0\gamma_0}(||\Lambda_n|| \geq \delta) \leq \\ &\limsup_{n \to \infty} P_{\theta_0\gamma_0}(\sup_{||\bar{\theta}_i - \theta_o|| \leq \epsilon, ||\bar{\gamma}_i - \gamma_0|| \leq \epsilon)} ||\Lambda_n|| \geq \delta) \to 0, \epsilon \to 0, \end{aligned}$$

by condition f). Thus $\Lambda_n \to 0$ in probability. Similarly $M_n \to 0$ in probability. Then (79) implies the desired convergence of $L(\delta_n)$. Indeed, using c), d), and e) we get

$$\delta_n \xrightarrow{d} V_1^{-1}(I_d, V_2) \cdot N(0, \Sigma),$$

which implies the statement of Lemma 2.

7.2 Matrix inequalities

Lemma 3. Suppose that x,y are r.v. with $x \geq 0$ and $y > 0$ a.s., $\mathrm{E}x = 1$, $\mathrm{E}(xy) < \infty$, and for each $d > 0, P(y \neq d, x \neq 0) > 0$. Then for each $\alpha \in (-\infty, 1)$ and $\beta \in (1, +\infty)$,

$$\{\mathrm{E}(xy^\alpha)\}^{\frac{1}{\alpha}} < \mathrm{E}(xy) < \{\mathrm{E}(xy^\beta)\}^{\frac{1}{\beta}}, \tag{80}$$

where by definition

$$\{\mathrm{E}(xy^\alpha)\}^{\frac{1}{\alpha}}\Big|_{\alpha=0} := e^{\mathrm{E}(x \ln y)}.$$

Proof. Let $F(x, y)$ be the joint d.f. of x and y. Then for each $c \neq 0$

$$\mathrm{E}(xy^c) = \int_{I\!R} y^c dG(y), \tag{81}$$

with

$$G(y) = \int_{I\!R} x dF(x, y), \quad y \in I\!R.$$

The funcion $G(y)$ is a probability d.f. as

$$G(+\infty) = \mathrm{E}x = 1.$$

Denote the integral on the right hand side of (81) by $\mathrm{E}_x y^c$. Note that $G(y)$ is not concentrated in a single point $y = d$ because $P(y \neq d, x \neq 0) > 0$. For $\alpha \neq 0$ the inequality (80) can be written in the form

$$(\mathrm{E}_x y^\alpha)^{\frac{1}{\alpha}} < \mathrm{E}_x y < (\mathrm{E}_x y^\beta)^{\frac{1}{\beta}}$$

and follows from Jensens's inequality, because $\mathrm{E}_x y = \mathrm{E}(xy) < \infty$. For $\alpha = 0$ we have to show that

$$e^{\mathrm{E}_x \ln y} < \mathrm{E}_x y, \tag{82}$$

where

$$\mathrm{E}_x \ln y := \int_{I\!R} \ln y dG(y).$$

But (82) also follows from Jensen's inequality. This completes the proof.

Lemma 4. Let v be a positive r.v. with distribution which has no atoms, i.e. for each $c > 0, P(v(\omega) \neq c) = 1$. Let w be a random (column) vector in $I\!R^m$, with $\mathrm{E}ww' = I_m$. Then for each $\alpha \in (-\infty, 0) \cup (0, 1)$:

$$tr\{\mathrm{E}(ww'v^\alpha)\}^{\frac{1}{\alpha}} < tr\mathrm{E}(ww'v), \tag{83}$$

if

$$\mathrm{E}(||w||^2 v^\alpha) < \infty, \quad \mathrm{E}(||w||^2 v) < \infty.$$

Proof. Let (λ, φ) be a pair of eigenvalue and normalized eigenvector of $\mathrm{E}(ww'v^\alpha)$. Then

$$\begin{aligned} \varphi'\{\mathrm{E}(ww'v^\alpha)\}^{\frac{1}{\alpha}}\varphi &= \lambda^{\frac{1}{\alpha}} = \{\varphi'\mathrm{E}(ww'v^\alpha)\varphi\}^{\frac{1}{\alpha}} \\ &= \{\mathrm{E}(\varphi'w)^2 v^\alpha\}^{\frac{1}{\alpha}} = \{\mathrm{E}(xv^\alpha)\}^{\frac{1}{\alpha}}, \end{aligned}$$

where $x = (\varphi'w)^2$, and $\mathrm{E}x = \varphi'(\mathrm{E}ww')\varphi = \varphi'\varphi = 1$. The distribution of v, has no atoms, therefore for each $d > 0, P(v \neq d, \varphi'w \neq 0) > 0$. From Lemma 3 we have as $\alpha < 1$

$$\{\mathrm{E}(xv^\alpha)\}^{\frac{1}{\alpha}} < \mathrm{E}(xv) = \varphi'\mathrm{E}(ww'v)\varphi.$$

Thus

$$\varphi'\{\mathrm{E}(ww'v^\alpha)\}^{\frac{1}{\alpha}}\varphi < \varphi'\mathrm{E}(ww'v)\varphi.$$

Now, summing up φ belonging to an eigenbasis of $\mathrm{E}(ww'v^\alpha)$, we obtain (83) and the Lemma is proved.

Remark 3. If the distribution of v has atoms then (83) holds with nonstrict inequality.

Remark 4. For $\beta > 1$ we will have similarly

$$tr\mathrm{E}(ww'v) < tr\{\mathrm{E}(ww'v^\beta)\}^{\frac{1}{\beta}}.$$

It is possible to extend (83) also to $\alpha = 0$. In this case

$$\{\mathrm{E}(ww'v^\alpha)\}^{\frac{1}{\alpha}}\big|_{\alpha=0} := e^{\mathrm{E}(ww'\ln v)}$$

and to prove (83) with $\alpha = 0$ we need additionally that $v > 1$.

Consider an application of Lemma 4.

Lemma 5. Let v and w satisfy the condition of Lemma 4, $\alpha \in I\!R$, and $v > 1$, and $\mathrm{E}(||w||^2 v^\alpha) < \infty, \mathrm{E}(||w||^2 v^\alpha \ln^2 v) < \infty$. Let

$$\begin{aligned} A &:= \mathrm{E}(ww'v^\alpha), \\ B &:= \mathrm{E}(ww'v^\alpha \ln v), \\ C &:= \mathrm{E}(ww'v^\alpha \ln^2 v). \end{aligned}$$

Then

$$tr(A^{-1}BA^{-1}B) < tr(A^{-1}C). \tag{84}$$

Proof. We set $\tilde{w} := (A^{-\frac{1}{2}}w)v^{\frac{\alpha}{2}}$ and $\tilde{v} := \ln^2 v$. Then $\mathrm{E}\tilde{w}\tilde{w}' = I_m, \tilde{v} > 0$. We have

$$\begin{aligned} tr(A^{-1}BA^{-1}B) &= tr(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})^2 = tr\{\mathrm{E}(\tilde{w}\tilde{w}'\tilde{v}^{\frac{1}{2}})\}^2, \\ tr(A^{-1}C) &= tr(A^{-\frac{1}{2}}CA^{-\frac{1}{2}}) = tr\mathrm{E}(\tilde{w}\tilde{w}'\tilde{v}). \end{aligned}$$

Now, (84) follows from (83), with the exponent $\frac{1}{2} < 1$.

Lemma 6. Let v and w satisfy the conditions of Lemma 4, and $v > 1$, and $\mathrm{E}(||w||^2 v) < \infty, \mathrm{E}(||w||^2 v^{-1}) < \infty$. Then

$$det\mathrm{E}(\frac{ww'}{v})det\mathrm{E}(ww'v) > 1. \tag{85}$$

Proof. Consider a matrix valued function

$$A = A(\alpha) := \mathrm{E}(ww'v^\alpha), -1 \le \alpha \le 1.$$

$A(\alpha)$ is positive definite for each $\alpha \in [-1, 1]$ and continuous in α. The inequality (85) is equivalent to

$$\frac{\ln detA(1) + \ln detA(-1)}{2} > \ln detA(0). \tag{86}$$

We prove that the continuous function

$$h(\alpha) := \ln detA(\alpha),\ \alpha \in [-1, 1],$$

is strictly convex. For $\alpha \in (-1, 1)$ we have $h'(\alpha) = tr(A^{-1}\frac{dA}{d\alpha})$, where $\frac{dA}{d\alpha} = \mathrm{E}(ww'v^\alpha \ln v)$. The last matrix is positive definite because $\ln v > 0$. Then

$$h''(\alpha) = -tr(A^{-1}\frac{dA}{d\alpha}A^{-1}\frac{dA}{d\alpha}) + tr(A^{-1}\frac{d^2 A}{d\alpha^2}),$$

where $\frac{d^2A}{d\alpha^2} = \mathrm{E}(ww'v^\alpha \ln^2 v)$. Now $h''(\alpha) > 0, -1 < \alpha < 1$, by Lemma 5. Thus $h(\alpha)$ is strictly convex. Therefore

$$\frac{h(1) + h(-1)}{2} > h(0).$$

We showed (86), and the Lemma is proved.

Corollary. Suppose that all the conditions of Lemma 6 hold, but $w = (w_1, \ldots, w_m)'$ is an arbitrary (not necessarily orthonormal) system of linearly independent in $L_2(\Omega)$ r.v. Then

$$det^2\mathrm{E}(ww') < det\mathrm{E}(\frac{ww'}{v})det\mathrm{E}(ww'v) \tag{87}$$

Proof. Denote $S := \mathrm{E}ww'$, S is positive definite. Let $z := S^{-\frac{1}{2}}w$. Then (87) is equivalent to

$$1 < det\mathrm{E}(\frac{zz'}{v})det\mathrm{E}(zz'v)$$

Now, $\mathrm{E}zz' = I_m$, and the inequality follows from Lemma 6. The Corollary is proved.

References

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

[2] CARROLL, R.J., RUPPERT, D. and STEFANSKI, L. A. (1995). *Measurement Error in Nonlinear Models.* Chapman and Hall, London.

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

[4] CHENG, C.-L., SCHNEEWEISS, H. and THAMERUS, M. (2000). *A small sample estimator for a polynomial regression with errors in the variables.* J.R. Statist. Soc. B **62**, 699–709.

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

[6] KUHA, J. T. and TEMPLE, J. (1999). *Covariate measurement error in quadratic regression.* Discussion Paper 1999-W2, Nuffield College Oxford.

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

[8] SCHNEEWEISS, H., and NITTNER, T. (2000). *Estimating a polynomial regression with measurement errors in the structural and in the functional case – a comparison.* Discussion Paper 197, Sonderforschungsbereich 386, University of Munich.

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