Metrika (2003) 57: 253–285

Metrika
© Springer-Verlag 2003

Consistent estimation in the bilinear multivariate errors-in-variables model

A. Kukush[1], **I. Markovsky**[1], **S. Van Huffel**[1]

ESAT, SCD-SISTA, K.U. Leuven, Kasteelpark 10, B-3001 Leuven-Heverlee, Belgium
(e-mail: Sabine.VanHuffel@esat.kuleuven.ac.be)

Abstract. A bilinear multivariate errors-in-variables model is considered. It corresponds to an overdetermined set of linear equations $AXB = C$, $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{p\times q}$, in which the data A, B, C are perturbed by errors. The total least squares estimator is inconsistent in this case.

An adjusted least squares estimator $\hat{X}$ is constructed, which converges to the true value X, as $m \to \infty$, $q \to \infty$. A small sample modification of the estimator is presented, which is more stable for small m and q and is asymptotically equivalent to the adjusted least squares estimator. The theoretical results are confirmed by a simulation study.

Key words: bilinear multivariate measurement error models, errors-in-variables models, adjusted least squares, consistency, asymptotic normality, small sample modification

1 Introduction

Many linear parameter estimation problems [VV91] can be reduced to solving an overdetermined set of linear equations

$$AX \approx B. \tag{1}$$

Whenever all measurements in both matrices A and B are affected by errors, the popular ordinary least squares estimator gives biased estimates. Measurement error models [Ful87], also called errors-in-variables models, should be considered in order to derive consistent estimators. If the errors are non-correlated and equally sized, the total least squares (TLS) method [VV91, GV80] provides a consistent estimate of the unknown parameter X. This method, better known as orthogonal regression in the statistical literature,

computes correction matrices ΔA and ΔB of minimal Frobenius norm in order to make the corrected set of equations

$$(A - \Delta A)X = B - \Delta B$$

compatible and has become very popular in engineering since the early eighties due to the existence of computationally attractive algorithms based on the singular value decomposition (SVD) [GV80, VV91].

Since then several generalizations of the TLS estimator have been presented. In particular, we mention the generalized TLS estimator, based on the generalized SVD, which provides consistent estimates of X provided the measurement errors in $[A\ B]$ are row-wise i.i.d. with zero mean and same covariance matrix, known up to a factor of proportionality.

In this paper we generalize the linear model in (1) to a bilinear model, represented as

$$AXB \approx C \tag{2}$$

This model can be considered as a special case of a polynomial model, namely a quadratic measurement error model [Ful87, CRS95].

It should be noted that the TLS principle can no longer be applied to model (2) in order to provide consistent estimates. Indeed, as mentioned in [Ful87], adding correction matrices ΔA, ΔB and ΔC of minimal Frobenius norm in order to make $(A - \Delta A)X(B - \Delta B) = C - \Delta C$ compatible results in biased estimates for the parameter X. In this paper an adjusted least squares (ALS) estimator [CRS95, CS98] of X is presented and shown to be consistent.

Next we give two examples where the bilinear measurement errors model (2) arises.

Example 1 (Total production cost model). Assume that r production inputs (materials, parts, labor, etc.) are combined to make n products. Let b_k be the price per unit of the k-th production input and x_{jk} be the number of units of the k-th production input, required to produce one unit of the j-th product. The production costs per unit of the j-th product is the j-th element of the vector

$$y = Xb.$$

Let a_j be a required quantity to be produced of the j-th product. The total quantity of the k-th production input needed is the k-th element of the vector

$$z = X^T a.$$

The total production cost c is $z^T b$, which gives a "single measurement" $AXB = C$ model

$$a^T Xb = c.$$

A situation in which we have multiple measurements could be: given is a set of (approximate) quantities to be produced of the n products, $a^1, \ldots, a^m$, (approximate) prices per unit of the production inputs, $b^1, \ldots, b^q$,

and (approximate) total costs c^{il}, corresponding to all combinations of the given quantities to be produced and prices. Then the model is

$$\underbrace{\begin{bmatrix} (a^1)^T \\ \vdots \\ (a^m)^T \end{bmatrix}}_{A} X \underbrace{[b^1 \dots b^q]}_{B} = \underbrace{\begin{bmatrix} c^{11} \cdots c^{1q} \\ \vdots \quad \vdots \\ c^{m1} \cdots c^{mq} \end{bmatrix}}_{C}.$$

Estimation of x_{jk} in the model $AXB = C$, can be interpreted as: estimate the number of units of the k-th production input, required to produce one unit of the j-th product.

Example 2 (Estimation of the fundamental matrix [MM98]). Two images are captured by a mobile camera and m matching pixels are located. Let

$$u_i = \begin{bmatrix} u_i(1) \\ u_i(2) \\ 1 \end{bmatrix} \quad \text{and} \quad v_i = \begin{bmatrix} v_i(1) \\ v_i(2) \\ 1 \end{bmatrix}, \quad i = 1, \dots, m$$

represent the homogeneous pixel coordinates in the first and second image respectively. The so called *epipolar constraint* relates the corresponding pixel coordinates by the model

$$v_i^T F u_i = 0, \quad i = 1, \dots, N, \tag{3}$$

where $F \in \mathbb{R}^{3\times 3}$, $\operatorname{rank}(F) = 2$ is the *fundamental matrix* which is identical for all pairs of corresponding vectors u_i, v_i, $1 \le i \le N$. Estimation of F from the given noisy data is called *structure from motion problem* and is a central problem in computer vision.

In [KMV01] we modified the adjusted least squares estimator derived in this paper for the model (3).

The notation we use is standard. For any matrix T, t_{ij} denotes the i, j-th element of T. The bold symbol $\mathbf{E}$ denotes mathematical expectation. It acts on the expression on the right up to an addition or subtraction sign. Conditional expectation of x, conditioned on C, is denoted by $\mathbf{E}[x|C]$. The notation $\operatorname{cov}(x)$ denotes the covariance matrix $\mathbf{E}x^T x - \mathbf{E}x^T \mathbf{E}x$ and $A^\dagger$ denotes the pseudo-inverse of A. In the formulas "const" denotes *any* constant value (for example, we can write $\text{const}^2 = \text{const}$).

The paper is organized as follows. Section 2 introduces the model and the global assumptions. In Section 3, an ALS estimator for the bilinear model (2) is derived. In Section 4, weak and strong consistency of the ALS estimator is stated. In Section 5, a bound on the rate of convergence is derived. In Section 6, asymptotic normality is shown. In Section 7, a small sample correction of the ALS estimator is derived. Section 8 gives numerical results and Section 9 concludes and discusses future work. Technical proofs are presented in the Appendix.

2 The model

We consider the model

$$AXB = C. \tag{4}$$

Here $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{p\times q}$, $C \in \mathbb{R}^{m\times q}$ are *observations* and $X \in \mathbb{R}^{n\times p}$ is a *parameter* of interest. We suppose that

$$A = A_0 + \tilde{A}, \quad B = B_0 + \tilde{B}, \quad C = C_0 + \tilde{C}, \tag{5}$$

and that there exists $X_0 \in \mathbb{R}^{n\times p}$ such that

$$A_0 X_0 B_0 = C_0.$$

Here A_0, B_0, and C_0 are *nominal* or *true values* and $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$ are *errors*. The matrix X_0 is the nominal or true value of the parameter. From the point of view of errors-in-variables models, $\tilde{C}$ represents the *equation error*, while $\tilde{A}$ and $\tilde{B}$ represent the *measurement errors*.

Looking for asymptotic results in the estimation of X in the model (4), we fix the dimensions of X, n and p, and let the number of measurements, m and q, increase. The measurements are represented by the rows of A, the columns of B and the elements of C. For the whole paper we denote

$$V_{\tilde{A}} \triangleq \mathbf{E}\tilde{A}^T\tilde{A}, \quad V_{\tilde{B}} \triangleq \mathbf{E}\tilde{B}\tilde{B}^T.$$

The matrices $V_{\tilde{A}}$ and $V_{\tilde{B}}$ are supposed to be known, while the variances of the entries of $\tilde{C}$ are unknown.

Specific notation is set in the course of exposition. The assumptions used in the paper are enumerated. Global assumption for the paper is assumption (i).

(i). The errors $\{\tilde{a}_{ij}, i \geq 1, 1 \leq j \leq n\}$, $\{\tilde{b}_{kl}, 1 \leq k \leq p, l \geq 1\}$ and $\{\tilde{c}_{il}, i \geq 1, l \geq 1\}$ form three independent arrays of r.v., which are centered and possess finite second order moments.

More assumptions are stated where necessary.

The model (4–5) is a bilinear regression measurement error model. In a scalar form it can be written as

$$\begin{aligned} c_{il} &= \sum_{j,k} \tilde{a}^0_{ij} x_{jk} \tilde{b}^0_{kl} + \tilde{c}_{il}, \quad 1 \leq i \leq m, 1 \leq l \leq q, \\ a_{ij} &= a^0_{ij} + \tilde{a}_{ij}, \quad 1 \leq i \leq m, 1 \leq j \leq n, \\ b_{kl} &= b^0_{kl} + \tilde{b}_{kl}, \quad 1 \leq k \leq p, 1 \leq l \leq q. \end{aligned} \tag{6}$$

Here the design points a^0_{ij} and b^0_{kl} are unobservable non-stochastic variables and the true value c^0_{il} is a nonlinear function of A_0 and B_0.

It is known that orthogonal regression is inconsistent for nonlinear measurement error models, see comments in [Ful87] and a mathematical proof of inconsistency in [KZ96]. The orthogonal regression estimator is a (weighted) TLS estimator [GV80, VV91], therefore due to nonlinearity of the model (4), the TLS estimator is inconsistent in this case.

The model (6) is a particular case of a quadratic model; it is bilinear with respect to the compound nuisance parameters $[A_0 \;\; B_0]$. For polynomial errors-in-variables models an ALS estimator is proposed in [CS98]. It is consistent. In [CRS95, Chapter 6], the method of corrected score functions is presented and in [Bar00] it is mentioned that an ALS estimator in a polynomial model is generated by the method of corrected score functions.

3 The score equation and an ALS estimator

We start with the LS objective function

$$q_{\mathrm{LS}}(X, A, B, C) \triangleq \|AXB - C\|_F^2. \tag{7}$$

In the space of matrices $\mathbb{R}^{n\times p}$, we introduce a scalar product

$$\langle T, S\rangle \triangleq \operatorname{tr}(TS^T), \quad T, S \in \mathbb{R}^{n\times p}.$$

The derivative $\partial q_{\mathrm{LS}}/\partial X$ is a linear functional on $\mathbb{R}^{n\times p}$. It acts on $H \in \mathbb{R}^{n\times p}$ according to the rule

$$\begin{aligned}\frac{1}{2}\frac{\partial q_{\mathrm{LS}}}{\partial X}(H) &= \operatorname{tr}((AXB - C)(AHB)^T) = \operatorname{tr}(A^T(AXB - C)B^T H^T)\\ &= \langle A^T(AXB - C)B^T, H\rangle.\end{aligned} \tag{8}$$

We can identify the derivative $\partial q_{\mathrm{LS}}/\partial X$ with a matrix, which represents it in the equality (8). Thus it is redefined as

$$\frac{1}{2}\frac{\partial q_{\mathrm{LS}}}{\partial X} = (A^TA)X(BB^T) - A^TCB^T.$$

In the absence of measurement errors, i.e. when $\tilde{A} = 0$ and $\tilde{B} = 0$, see (5), the LS estimator is obtained by minimizing (7) or (what is asymptotically equivalent) via the score equation $\frac{1}{2}\partial q_{\mathrm{LS}}/\partial X = 0$. Thus the score function for the LS method is

$$\psi_{\mathrm{LS}}(X, A, B, C) \triangleq (A^TA)X(BB^T) - A^TCB^T,$$

and the LS estimator is consistent in the absence of measurement errors.

Now, we are looking for a corrected score function ψ, such that

$$\mathbf{E}[\psi(X, A_0 + \tilde{A}, B_0 + \tilde{B}, C) \mid C] = \psi_{\mathrm{LS}}(X, A_0, B_0, C) \quad \text{for all } X, A_0, B_0, C.$$

We seek ψ in the form $\psi = \psi_{\mathrm{LS}} - \psi_1$. We have by assumption (i)

$$\begin{aligned}&\mathbf{E}[\psi_{\mathrm{LS}}(X, A_0 + \tilde{A}, B_0 + \tilde{B}, C) \mid C]\\ &\quad= \psi_{\mathrm{LS}}(X, A_0, B_0, C) + \mathbf{E}\tilde{A}^T\tilde{A}XB_0B_0^T + \mathbf{E}A_0^TA_0X\tilde{B}\tilde{B}^T + V_{\tilde{A}}XV_{\tilde{B}}\\ &\quad= \psi_{\mathrm{LS}} + \psi_{11}(B_0) + \psi_{12}(A_0) + V_{\tilde{A}}XV_{\tilde{B}},\end{aligned}$$

with

$$\psi_{11}(B_0) \triangleq V_{\tilde{A}} X B_0 B_0^T,$$

$$\psi_{12}(A_0) \triangleq A_0^T A_0 X V_{\tilde{B}}.$$

To find a proper correction term ψ_1 we consider

$$\begin{aligned} \mathbf{E}\psi_{11}(B_0 + \tilde{B}) &= V_{\tilde{A}} X B_0 B_0^T + V_{\tilde{A}} X V_{\tilde{B}}, \\ \mathbf{E}\psi_{12}(A_0 + \tilde{A}) &= A_0^T A_0 X V_{\tilde{B}} + V_{\tilde{A}} X V_{\tilde{B}}. \end{aligned} \tag{9}$$

Therefore

$$\psi_1(A, B) = \psi_{11}(B) + \psi_{12}(A) - V_{\tilde{A}} X V_{\tilde{B}}$$

and

$$\begin{aligned} &\psi(X, A, B, C) \\ &\quad = (A^T A) X (BB^T) - A^T C B^T - V_{\tilde{A}} X (BB^T) - (A^T A) X V_{\tilde{B}} + V_{\tilde{A}} X V_{\tilde{B}} \\ &\quad = (A^T A - V_{\tilde{A}}) X (BB^T - V_{\tilde{B}}) - A^T C B^T. \end{aligned}$$

The ALS estimator $\hat{X}$ is defined from the equation

$$\psi(X, A, B, C) = 0. \tag{10}$$

As an estimator we can take

$$\hat{X} \triangleq (A^T A - V_{\tilde{A}})^{\dagger} (A^T C B^T) (BB^T - V_{\tilde{B}})^{\dagger}. \tag{11}$$

If $A^T A - V_{\tilde{A}}$ and $BB^T - V_{\tilde{B}}$ are non-singular, then (11) satisfies (10). Later on we shall show that these matrices are non-singular with probability tending to one as the number of measurements (rows of A and columns of B) is tending to infinity. Observe that (11) reduces to the generalized TLS estimator [VV89, Gal82], in the case $B = I_p$, $\tilde{B} = 0$ under the assumption (i).

4 Weak and strong consistency

We introduce further assumptions.

(ii). The rows of $\tilde{A}$ are independent, i.e. $(\tilde{a}_{ij}, i \geq 1, 1 \leq j \leq n)$ are independent, the columns of $\tilde{B}$ are independent, i.e. $(\tilde{b}_{kl}, 1 \leq k \leq p, l \geq 1)$ are independent, and all elements of $\tilde{C}$ are independent, i.e. $(\tilde{c}_{il}, i \geq 1, l \geq 1)$ are independent.

(iii). $\mathbf{E}\tilde{a}_{ij}^4 \leq \text{const}$, $\mathbf{E}\tilde{b}_{kl}^4 \leq \text{const}$, and $\mathbf{E}\tilde{c}_{il}^2 \leq \text{const}$.

(iv). We denote

$$V_{A_0} \triangleq A_0^T A_0, \quad V_{B_0} \triangleq B_0 B_0^T$$

and assume that

$$\frac{\lambda_{\max}(V_{A_0})+m}{\lambda^2_{\min}(V_{A_0})} \to 0 \quad \text{as } m \to \infty, \quad \text{and} \quad \frac{\lambda_{\max}(V_{B_0})+q}{\lambda^2_{\min}(V_{B_0})} \to 0 \quad \text{as } q \to \infty.$$

The assumption (iv) corresponds to the condition of weak consistency, given in [Gal82] for the maximum likelihood estimator in the model (1).

Theorem 1 (Weak consistency). *Assume that assumptions (i) to (iv) hold. Then the estimator $\hat{X}$ given in (11) converges to X_0 in probability, as $m \to \infty$ and $q \to \infty$.*

Proof. Denote

$$U_A \triangleq A^TA - V_{\tilde{A}}, \quad U_B \triangleq BB^T - V_{\tilde{B}}. \tag{12}$$

By assumption (iv), V_{A_0} is non-singular for $m \geq m_0$ and V_{B_0} is non-singular for $q \geq q_0$ for some fixed m_0 and q_0. For $m \geq m_0$, $q \geq q_0$, we rewrite equation (10) in the form

$$V_{A_0}^{-1} U_A X U_B V_{B_0}^{-1} = V_{A_0}^{-1}(A^TA_0X_0B_0B^T + A^T\tilde{C}B^T)V_{B_0}^{-1}. \tag{13}$$

For consistency, it is enough to show that

$$V_{A_0}^{-1}U_A \xrightarrow{\text{P}} I_n \quad \text{and} \quad U_BV_{B_0}^{-1} \xrightarrow{\text{P}} I_p, \tag{14}$$

$$V_{A_0}^{-1}(A^TA_0) \xrightarrow{\text{P}} I_n \quad \text{and} \quad B_0B_0^TV_{B_0}^{-1} \xrightarrow{\text{P}} I_p, \tag{15}$$

$$V_{A_0}^{-1}A^T\tilde{C}B^TV_{B_0}^{-1} \xrightarrow{\text{P}} 0. \tag{16}$$

The proofs of (14)–(16) are given in the Appendix. □

The main probabilistic tool to prove the strong consistency is the following matrix analogue of the Rosenthal inequality, see [Ros70].

Lemma 1. *Let $\{\eta_i, i \geq 1\}$ be a sequence of independent r.v., $\mathbf{E}\eta_i = 0$, $i = 1, 2, \ldots$ Then for any real number $t \geq 2$, and for all $m \geq 1$*

$$\mathbf{E}\left|\sum_{i=1}^m \eta_i\right|^t \leq c(t)\max\left(\sum_{i=1}^m \mathbf{E}|\eta_i|^t, \left(\sum_{i=1}^m \mathbf{E}\eta_i^2\right)^{t/2}\right),$$

where $c(t)$ depends on t, but it does not depend on m.

We strengthen assumptions (iii) and (iv).

(v). For fixed real number $r \geq 2$, $\mathbf{E}|\tilde{a}_{ij}|^{2r} \leq \text{const}$, $\mathbf{E}|\tilde{b}_{kl}|^{2r} \leq \text{const}$, and $\mathbf{E}|\tilde{c}_{il}|^{2r} \leq \text{const}$.

(vi). For fixed $m_0 \geq 1$,

$$\sum_{m=m_0}^{\infty}\left(\frac{m^{r/2}}{\lambda^r_{\min}(V_{A_0})} + \frac{\lambda^r_{\max}(V_{A_0})}{\lambda^{2r}_{\min}(V_{A_0})}\right) < \infty,$$

and for fixed $q_0 \geq 1$

$$\sum_{q=q_0}^{\infty} \left(\frac{q^{r/2}}{\lambda_{\min}^{r}(V_{B_0})} + \frac{\lambda_{\max}^{r}(V_{B_0})}{\lambda_{\min}^{2r}(V_{B_0})} \right) < \infty,$$

where r is defined in assumption (v).

Theorem 2 (Strong consistency). *Assume that assumptions (i), (ii), (v) and (vi) hold. Then the estimator $\hat{X}$ given in (11) converges to X_0 a.s., as $m \to \infty$, $q \to \infty$.*

Proof. See Appendix.

5 Rate of convergence

Let the assumptions of Theorem 1 hold. With probability tending to 1 we have

$$(A^T A - V_{\tilde{A}})\hat{X}(BB^T - V_{\tilde{B}}) = A^T C B^T. \tag{17}$$

We set $\hat{X} \triangleq X_0 + \hat{\Delta}$ and consider $m \geq m_0$, $q \geq q_0$, for which V_{A_0} and V_{B_0} are non-singular. From (17) we have

$$\begin{aligned}(A^T A - V_{\tilde{A}})\hat{\Delta}(BB^T - V_{\tilde{B}}) = {} & A^T(A_0 X_0 B_0 + \tilde{C})B^T \\ & - (A^T A - V_{\tilde{A}})X_0(BB^T - V_{\tilde{B}}).\end{aligned} \tag{18}$$

Using the notations (12) we have

$$\begin{aligned}V_{A_0}^{-1} U_A \hat{\Delta} U_B V_{B_0}^{-1} = {} & V_{A_0}^{-1} A^T \tilde{C} B^T V_{B_0}^{-1} + V_{A_0}^{-1}(A^T A_0 X_0 B_0 B^T \\ & - (A^T A - V_{\tilde{A}})X_0(BB^T - V_{\tilde{B}}))V_{B_0}^{-1} \triangleq R_1 + R_2.\end{aligned} \tag{19}$$

By (14), the LHS of (19) equals

$$\text{LHS} = (I_n + \mathrm{o_p}(1))\hat{\Delta}(I_p + \mathrm{o_p}(1)) \tag{20}$$

Next, see Section 4,

$$\begin{aligned}R_1 &= V_{A_0}^{-1} A^T \tilde{C} B^T V_{B_0}^{-1} \\ &= \frac{\sqrt{\lambda_{\max}(V_{A_0}) + m}}{\lambda_{\min}(V_{A_0})} \cdot \frac{\sqrt{\lambda_{\max}(V_{B_0}) + q}}{\lambda_{\min}(V_{b_0})} \cdot \mathrm{O_p}(1).\end{aligned} \tag{21}$$

We decompose $R_2 = R_{21} - R_{22}$.

$$R_{21} = V_{A_0}^{-1}(A^T A_0 X_0 B_0 B^T - V_{A_0} X_0 V_{B_0}) V_{B_0}^{-1}$$

$$= \left(I_n + \frac{\sqrt{\lambda_{\max}(V_{A_0})}}{\lambda_{\min}(V_{A_0})} \mathrm{O_p}(1) \right) X_0 \left(I_p + \frac{\sqrt{\lambda_{\max}(V_{B_0})}}{\lambda_{\min}(V_{B_0})} \mathrm{O_p}(1) \right) - X_0, \quad (22)$$

$$\|R_{21}\|_F = \left(\frac{\sqrt{\lambda_{\max}(V_{A_0})}}{\lambda_{\min}(V_{A_0})} + \frac{\sqrt{\lambda_{\max}(V_{B_0})}}{\lambda_{\min}(V_{B_0})} \right) \mathrm{O_p}(1) \quad (23)$$

and

$$R_{22} = V_{A_0}^{-1}(A^T A - V_{\tilde{A}}) X_0 (BB^T - V_{\tilde{B}} - V_{A_0} X_0 V_{B_0}) V_{B_0}^{-1}$$

$$= \left(I_n + \frac{\sqrt{m} + \sqrt{\lambda_{\max}(V_{A_0})}}{\lambda_{\min}(V_{A_0})} \mathrm{O_p}(1) \right) X_0$$

$$\left(I_p + \frac{\sqrt{q} + \sqrt{\lambda_{\max}(V_{B_0})}}{\lambda_{\min}(V_{B_0})} \mathrm{O_p}(1) \right) - X_0, \quad (24)$$

$$\|R_{22}\|_F = \left(\frac{\sqrt{m} + \sqrt{\lambda_{\max}(V_{A_0})}}{\lambda_{\min}(V_{A_0})} + \frac{\sqrt{q} + \sqrt{\lambda_{\max}(V_{B_0})}}{\lambda_{\min}(V_{B_0})} \right) \mathrm{O_p}(1). \quad (25)$$

Therefore from (19), (20), and (21) to (25) we obtain

$$\|\hat{X} - X_0\|_F = (u_m + v_q) \mathrm{O_p}(1), \quad (26)$$

where

$$u_m \triangleq \frac{\sqrt{m} + \sqrt{\lambda_{\max}(V_{A_0})}}{\lambda_{\min}(V_{A_0})}, \quad v_q \triangleq \frac{\sqrt{q} + \sqrt{\lambda_{\max}(V_{B_0})}}{\lambda_{\min}(V_{B_0})}. \quad (27)$$

6 Asymptotic normality

6.1 Expansion for $\hat{\Delta}$

Now, we strengthen assumption (iv).

(vii). $\frac{1}{m} V_{A_0} \to V_{A\infty}$ as $m \to \infty$, and $\frac{1}{q} V_{B_0} \to V_{B\infty}$ as $q \to \infty$, where $V_{A\infty}$ and $V_{B\infty}$ are positive definite matrices.

Under (vii), $\frac{1}{m}\lambda_{\max}(V_{A_0}) \to \lambda_{\max}(A_{A\infty}) > 0$, and $\frac{1}{m}\lambda_{\min}(V_{A_0}) \to \lambda_{\min}(A_{A\infty}) > 0$, similarly for V_{B_0}, therefore (vii) implies (iv).

We shall assume (i) to (iii) and (vii), and in the process of establishing asymptotic normality, we shall set some more assumptions.

From (19), (20) and (21) we have now

$$(I_n + \mathrm{o_p}(1)) \hat{\Delta} (I_p + \mathrm{o_p}(1)) = \frac{1}{\sqrt{mq}} \mathrm{O_p}(1) + R_{21} - R_{22}. \quad (28)$$

$$R_{21} = (I_n + V_{A_0}^{-1}\tilde{A}^T A_0)X_0(I_p + B_0\tilde{B}^T V_{B_0}^{-1}) - X_0$$

$$= V_{A_0}^{-1}\tilde{A}^T A_0 X_0 + X_0 B_0 \tilde{B}^T V_{B_0}^{-1} + \frac{1}{\sqrt{mq}}\mathrm{O_p}(1),$$

see (22). Next, see (24),

$$R_{22} = (I_n + V_{A_0}^{-1}(\tilde{A}^T A_0 + A_0^T\tilde{A} + \tilde{A}^T\tilde{A} - V_{\tilde{A}}))X_0$$

$$+ (I_p + (\tilde{B}B_0^T + B_0\tilde{B}^T + \tilde{B}\tilde{B}^T - V_{\tilde{B}})V_{B_0}^{-1}) - X_0$$

$$= V_{A_0}^{-1}(\tilde{A}^T A_0 + A_0^T\tilde{A} + \tilde{A}^T\tilde{A} - V_{\tilde{A}})X_0$$

$$+ X_0(\tilde{B}B_0^T + B_0\tilde{B}^T + \tilde{B}\tilde{B}^T - V_{\tilde{B}})V_{B_0}^{-1} + \frac{1}{\sqrt{mq}}\mathrm{O_p}(1),$$

and

$$R_{21} - R_{22} = -V_{A_0}^{-1}L_1X_0 - X_0L_2V_{B_0}^{-1} + \frac{1}{\sqrt{mq}}\mathrm{O_p}(1). \tag{29}$$

Here

$$L_1 \triangleq A_0^T\tilde{A} + (\tilde{A}^T\tilde{A} - V_{\tilde{A}}), \quad L_2 \triangleq \tilde{B}B_0^T + (\tilde{B}\tilde{B}^T - V_{\tilde{B}}). \tag{30}$$

Thus

$$-(I_n + \mathrm{o_p}(1))\hat{\Delta}(I_p + \mathrm{o_p}(1))$$

$$= \left(\left(\frac{1}{m}V_{A_0}\right)^{-1}\frac{L_1}{\sqrt{m}}X_0\right)\frac{1}{\sqrt{m}} + \left(X_0\frac{L_2}{\sqrt{q}}\left(\frac{1}{q}V_{B_0}\right)^{-1}\right)\frac{1}{\sqrt{q}}$$

$$+ \frac{1}{\sqrt{mq}}\mathrm{O_p}(1). \tag{31}$$

By assumption (vii), $\frac{1}{m}V_{A_0}$ and $\frac{1}{q}V_{B_0}$ converge to the corresponding positive definite matrices $V_{A\infty}$ and $V_{B\infty}$. Random matrices L_1 and L_2 are independent by (i). We need an assumption which ensures the convergence in distribution of $\frac{1}{\sqrt{m}}L_1$ and $\frac{1}{\sqrt{q}}L_2$.

6.2 Behavior of $\frac{1}{\sqrt{m}}L_1$

We denote by $\tilde{a}_i^T$, a_i^{0T} and a_i^T, $i \geq 1$, the rows of $\tilde{A}$, A_0 and A respectively.

$$\tilde{A} = \begin{bmatrix} \tilde{a}_1^T \\ \vdots \\ \tilde{a}_m^T \end{bmatrix}, \quad A_0 = \begin{bmatrix} a_1^{0T} \\ \vdots \\ a_m^{0T} \end{bmatrix}, \quad \text{and} \quad A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix}.$$

We have

$$L_1 = \sum_{i=1}^{m} (a_i^0 \tilde{a}_i^T + \tilde{a}_i a_i^{0T} + \tilde{a}_i \tilde{a}_i^T - \mathbf{E}\tilde{a}_i \tilde{a}_i^T). \tag{32}$$

To apply the central limit theorem (CLT), we consider $l_1 \triangleq \operatorname{vec}(L_1)$. We have

$$\operatorname{vec}(a_i^0 \tilde{a}_i^T) = (I_n \otimes a_i^0)\tilde{a}_i, \tag{33}$$

and

$$\operatorname{vec}(\tilde{a}_i \tilde{a}_i^T) = \tilde{a}_i \otimes \tilde{a}_i. \tag{34}$$

Therefore, see (32) to (34),

$$l_1 = \sum_{i=1}^{m} (a_i^{0T} \otimes \tilde{a}_i + \tilde{a}_i \otimes a_i^0 + \tilde{a}_i \otimes \tilde{a}_i - \mathbf{E}\tilde{a}_i \otimes \tilde{a}_i). \tag{35}$$

We introduce the following assumptions in order to apply the CLT to $\frac{1}{\sqrt{m}} l_1$.

(viii). The rows $\{\tilde{a}_i^T,\ i \geq 1\}$ are identically distributed as random vector $\tilde{\mathbf{a}}$. By assumption (i), $\tilde{\mathbf{a}}$ is centered and has covariance matrix $V_{\tilde{\mathbf{a}}} = \frac{1}{m} V_{\tilde{A}}$.
In order to distinguish the vectors $\tilde{a}_i$ from the scalars $\tilde{\mathbf{a}}_i$, we will use the notation $\tilde{\mathbf{a}}(i)$ for the elements of $\tilde{\mathbf{a}}$.

(ix). For fixed $\delta > 0$, $\mathbf{E}|\tilde{\mathbf{a}}(j)|^{4+\delta} < \infty$, $j = 1, \ldots, n$.

(x). For each $\{j, k, l\} \subset \{1, \ldots, n\}$, $\mathbf{E}\tilde{\mathbf{a}}(j)\tilde{\mathbf{a}}(k)\tilde{\mathbf{a}}(l) = 0$.
Assumption (x) holds, e.g., when $\tilde{\mathbf{a}}$ has a symmetric distribution. It is possible to avoid (x), but then the asymptotic covariance matrix of $\hat{X}$ will be more complicated. For instance, (x) holds in the case of normal errors $\tilde{\mathbf{a}}(j)$.

(xi). For some $\tau > 0$, $\frac{1}{m}\sum_{i=1}^{m} \|a_i^0\|^{2+\tau} \leq \text{const}$, and $\frac{1}{m} \max_{1 \leq i \leq m} \|a_i^0\|^2 \to 0$.

Denote

$$U_A' \triangleq \lim_{m \to \infty} \frac{1}{m} (I_n \otimes a_i^0) V_{\tilde{\mathbf{a}}} (I_n \otimes a_i^0)^T = V_{\tilde{\mathbf{a}}} \otimes V_{A\infty},$$

and

$$\begin{aligned} U_A'' &\triangleq \operatorname{cov}(\tilde{\mathbf{a}} \otimes \tilde{\mathbf{a}} - \mathbf{E}\tilde{\mathbf{a}} \otimes \tilde{\mathbf{a}}) \\ &= \mathbf{E}((\tilde{\mathbf{a}} \otimes \tilde{\mathbf{a}} - \mathbf{E}\tilde{\mathbf{a}} \otimes \tilde{\mathbf{a}})(\tilde{\mathbf{a}} \otimes \tilde{\mathbf{a}} - \mathbf{E}\tilde{\mathbf{a}} \otimes \tilde{\mathbf{a}})^T) \\ &= \mathbf{E}((\tilde{\mathbf{a}}\tilde{\mathbf{a}}^T) \otimes (\tilde{\mathbf{a}}\tilde{\mathbf{a}}^T)) - \operatorname{vec}(V_{\tilde{\mathbf{a}}})(\operatorname{vec}(V_{\tilde{\mathbf{a}}}))^T. \end{aligned} \tag{36}$$

The elements of $\mathbf{E}((\tilde{\mathbf{a}}\tilde{\mathbf{a}}^T) \otimes (\tilde{\mathbf{a}}\tilde{\mathbf{a}}^T))$ are the fourth moments of $\tilde{\mathbf{a}}$, $\mathbf{E}\tilde{\mathbf{a}}(i)\tilde{\mathbf{a}}(j)\tilde{\mathbf{a}}(k)\tilde{\mathbf{a}}(l)$. We note that U_A' and U_A'' are positive semidefinite. The last assumption in this subsection is assumption (xii).

(xii). $U_A' + U_A''$ is positive definite.

Now, under new assumptions (viii) to (xi) we apply the CLT to $\frac{1}{\sqrt{m}} l_1$, see (35). We have

$$\operatorname{cov}\left(\frac{l_1}{\sqrt{m}}\right) = \frac{1}{m}\operatorname{cov}(l_1) = \frac{1}{m}\sum_{i=1}^{m}(I_n \otimes a_i^0) V_{\tilde{\mathbf{a}}} (I_n \otimes a_i^0)^T + U_A''$$

and $\operatorname{cov}(l_1/\sqrt{m}) \to U_A \triangleq U_A' + U_A''$, which is positive definite.

Next, we check the Lyapunov condition, with $v \triangleq \min(\tau, \delta/2)$, where τ and δ are defined in assumptions (ix) and (xi). We have, see (35),

$$\begin{aligned}
&\frac{1}{m^{1+v/2}}\sum_{i=1}^{m}(\mathbf{E}\|(I_n \otimes a_i^0)\tilde{\mathbf{a}}\|_F^{2+v} + \mathbf{E}\|\tilde{\mathbf{a}} \otimes \tilde{\mathbf{a}} - \mathbf{E}\tilde{\mathbf{a}} \otimes \tilde{\mathbf{a}}\|_F^{2+v}) \\
&\quad \le \frac{1}{m^{1+v/2}}\left(\sum_{i=1}^{m} \text{const} \cdot \|a_i^0\|^{2+v} \mathbf{E}\|\tilde{\mathbf{a}}\|^{2+v} + \text{const} \cdot m\right) \\
&\quad \le \text{const} \cdot \frac{1}{m^{v/2}} \to 0, \quad \text{as } m \to \infty.
\end{aligned}$$

We have also by assumptions (iii), (viii) and (xi), that the second moments of the summands in (35) are bounded. Therefore by the CLT

$$\frac{l_1}{\sqrt{m}} \xrightarrow{d} N(0, U_A).$$

Then, see (31),

$$\begin{aligned}
&\operatorname{vec}\left(\left(\frac{1}{m} V_{A_0}\right)^{-1} \frac{L_1}{\sqrt{m}} X_0\right) \\
&\quad = \left(X_0^T \otimes \left(\frac{1}{m} V_{A_0}\right)^{-1}\right) \frac{l_1}{\sqrt{m}} \xrightarrow{d} (X_0^T \otimes V_{A\infty}^{-1}) \cdot N(0, U_A) \\
&\quad = N(0, \Sigma_A), \quad \text{as } m \to \infty
\end{aligned}$$

with

$$\Sigma_A \triangleq (X_0 \otimes V_{A\infty}^{-1})^T U_A (X_0 \otimes V_{A\infty}^{-1}). \tag{37}$$

6.3 *Behavior of* $\frac{1}{\sqrt{q}} L_2$

We list similar assumptions for B_0 and $\tilde{\boldsymbol{B}}$.

(viii)′. The columns $\{\tilde{b}_l, l \ge 1\}$ are identically distributed as random vector $\tilde{\mathbf{b}}$. Here $\tilde{\boldsymbol{B}} = [\tilde{b}_1 \dots \tilde{b}_q]$. By assumption (i), $\tilde{\mathbf{b}}$ is centered, with covariance matrix $V_{\tilde{\mathbf{b}}} = \frac{1}{q} V_{\tilde{\boldsymbol{B}}}$.

(ix)′. For fixed $\delta > 0$, $\mathbf{E}|\tilde{\mathbf{b}}(k)|^{4+\delta} < \infty$, $k = 1, \ldots, p$.
(x)′. For each $\{j,k,l\} \subset \{1, \ldots, p\}$, $\mathbf{E}\tilde{\mathbf{b}}(j)\tilde{\mathbf{b}}(k)\tilde{\mathbf{b}}(l) = 0$.
(xi)′. For some $\tau > 0$, $\frac{1}{q}\sum_{l=1}^{q}\|b_i^0\|^{2+\tau} \le \text{const}$, and $\frac{1}{q}\max_{1\le l\le q}\|b_l^0\|^2 \to 0$.

Denote

$$U_B' \triangleq \lim_{q\to\infty}\frac{1}{q}\sum_{l=1}^{q}(b_l^0 \otimes I_p)V_{\tilde{\mathbf{b}}}(b_l^0 \otimes I_p)^T = V_{B\infty} \otimes V_{\tilde{\mathbf{b}}},$$

and

$$\begin{aligned} U_B'' &\triangleq \operatorname{cov}(\tilde{\mathbf{b}} \otimes \tilde{\mathbf{b}} - \mathbf{E}\tilde{\mathbf{b}} \otimes \tilde{\mathbf{b}}) \\ &= \mathbf{E}(\tilde{\mathbf{b}} \otimes \tilde{\mathbf{b}} - \mathbf{E}\tilde{\mathbf{b}} \otimes \tilde{\mathbf{b}})(\tilde{\mathbf{b}} \otimes \tilde{\mathbf{b}} - \mathbf{E}\tilde{\mathbf{b}} \otimes \tilde{\mathbf{b}})^T \\ &= \mathbf{E}((\tilde{\mathbf{b}}\tilde{\mathbf{b}}^T) \otimes (\tilde{\mathbf{b}}\tilde{\mathbf{b}}^T)) - \operatorname{vec}(V_{\tilde{\mathbf{b}}})(\operatorname{vec}(V_{\tilde{\mathbf{b}}}))^T. \end{aligned}$$

(xii)′. $U_B \triangleq U_B' + U_B''$ is positive definite.

Then similarly to the previous subsection, we have

$$l_2 \triangleq \operatorname{vec}(L_2) = \sum_{l=1}^{q}((b_l^0 \otimes I_p)\tilde{\mathbf{b}} + \tilde{\mathbf{b}} \otimes \tilde{\mathbf{b}} - \mathbf{E}\tilde{\mathbf{b}} \otimes \tilde{\mathbf{b}}).$$

$$\frac{l_2}{\sqrt{q}} \xrightarrow{d} N(0, U_B),$$

and

$$\begin{aligned} &\operatorname{vec}\left(X_0\frac{L_2}{\sqrt{q}}\left(\frac{1}{q}V_{B_0}\right)^{-1}\right) \\ &\quad= \left(\left(\frac{1}{q}V_{B_0}\right)^{-1} \otimes X_0\right)\frac{l_2}{\sqrt{q}} \xrightarrow{d} (V_{B\infty}^{-1} \otimes X_0)\cdot N(0, U_B) \\ &\quad= N(0, \Sigma_B), \end{aligned} \tag{38}$$

where $\Sigma_B \triangleq (V_{B\infty}^{-1} \otimes X_0)U_B(V_{B\infty}^{-1} \otimes X_0)^T$. From (31) we obtain

$$(I_n + \mathrm{o_p}(1))\hat{\Delta}(I_n + \mathrm{o_p}(1)) = \frac{1}{\sqrt{m}}\xi_m + \frac{1}{\sqrt{q}}\eta_q + \frac{\mathrm{O_p}(1)}{\sqrt{mq}} \quad \text{as } m \to \infty, q \to \infty,$$

where $\{\xi_m\}$ and $\{\eta_q\}$ are independent random matrices, and $\operatorname{vec}(\xi_m) \xrightarrow{d} N(0, \Sigma_A)$ as $m \to \infty$ and $\operatorname{vec}(\eta_q) \xrightarrow{d} N(0, \Sigma_B)$ as $q \to \infty$.

Let $m = m(r)$ and $q = q(r)$, with $\frac{r}{m} \to \lambda_A$, $\frac{r}{q} \to \lambda_B$, as $r \to \infty$, where $0 \le \lambda_A$, $\lambda_B < \infty$, $\lambda_A + \lambda_B > 0$. Then we have

$$(I_n + \mathrm{o_p}(1))(\sqrt{r}\hat{\Delta})(I_n + \mathrm{o_p}(1)) = \sqrt{\frac{r}{m}}\xi_m + \sqrt{\frac{r}{q}}\eta_q + \frac{\mathrm{O_p}(1)}{\sqrt{r}} \cdot \sqrt{\frac{r}{m} \cdot \frac{r}{q}}.$$

Therefore we proved the following asymptotic normality result.

Theorem 3. *Assume that assumptions (i) to (iii), (vii) to (xii), and (viii)′ to (xii)′ hold. Then for $m = m(r)$, $q = q(r)$, $\frac{r}{m(r)} \to \lambda_A$, $\frac{r}{q(r)} \to \lambda_B$, as $r \to \infty$ with $0 \le \lambda_A < \infty$, $0 \le \lambda_B < \infty$, $\lambda_A + \lambda_B > 0$, we have*

$$\sqrt{r} \cdot \mathrm{vec}(\hat{X} - X_0) \xrightarrow{d} N(0, \lambda_A \Sigma_A + \lambda_B \Sigma_B).$$

Now, we investigate the rank of the matrix $\lambda_A \Sigma_A + \lambda_B \Sigma_B$. We analyze (37). Suppose that X_0 is of full rank, i.e. $\mathrm{rank}(X_0) = \min(n, p)$. Then

$$\mathrm{rank}(X_0^T \otimes V_{A\infty}^{-1}) = \mathrm{rank}(X_0 \otimes V_{A\infty}^{-1}) = n \min(n, p).$$

U_A has full rank equal n^2, and

$$\mathrm{rank}(\Sigma_A) = n \min(n, p).$$

Similarly

$$\mathrm{rank}(\Sigma_B) = p \min(n, p).$$

Denote

$$\Sigma \triangleq \lambda_A \Sigma_A + \lambda_B \Sigma_B.$$

- If $\lambda_A > 0$, $\lambda_B > 0$ then $\mathrm{rank}(\Sigma) = \max(\mathrm{rank}(\Sigma_A), \mathrm{rank}(\Sigma_B)) = np$, i.e. Σ is a positive definite $np \times np$ matrix.
- If $\lambda_A = 0$ then $\mathrm{rank}(\Sigma) = \mathrm{rank}(\Sigma_B) = p \min(n, p)$, and Σ is positive definite when $n \le p$.
- If $\lambda_B = 0$ then $\mathrm{rank}(\Sigma) = \mathrm{rank}(\Sigma_A) = n \min(n, p)$, and Σ is positive definite when $n \ge p$.

In the case when Σ is positive definite, we have

$$\|\sqrt{r}\Sigma^{-1/2} \mathrm{vec}(\hat{\Delta})\|^2 \xrightarrow{d} \chi^2_{np} \quad \text{as } r \to \infty,$$

or

$$r \cdot (\mathrm{vec}(\hat{\Delta}))^T \Sigma^{-1} \mathrm{vec}(\hat{\Delta}) \xrightarrow{d} \chi^2_{np} \quad \text{as } r \to \infty. \tag{39}$$

6.4 Approximate expression

Next, we give an approximate expression for Σ, constructed via observations, which converges in probability to Σ.

$$\Sigma^{\mathrm{app}} = \lambda_A \Sigma_A^{\mathrm{app}} + \lambda_B \Sigma_B^{\mathrm{app}},$$

see (37),

$$\Sigma_A^{\text{app}} = \left(\hat{X} \otimes \left(\frac{1}{m} A^T A - V_{\tilde{A}}\right)^{-1}\right)^T U_A^{\text{app}} \left(\hat{X} \otimes \left(\frac{1}{m} A^T A - V_{\tilde{A}}\right)^{-1}\right),$$

$$U_A^{\text{app}} = U'_{\text{app},A} + U''_{\text{app},A},$$

$$U'_{\text{app},A} = V_{\tilde{\mathbf{a}}} \otimes \left(\frac{1}{m} A^T A - V_{\tilde{\mathbf{a}}}\right), \quad \text{and} \quad U''_{\text{app},A} = U''_A.$$

Now, see (38),

$$\Sigma_B^{\text{app}} = \left(\left(\frac{1}{q} BB^T - V_{\tilde{B}}\right)^{-1} \otimes \hat{X}\right) U_B^{\text{app}} \left(\left(\frac{1}{q} BB^T - V_{\tilde{B}}\right)^{-1} \otimes \hat{X}\right)^T,$$

$$U_B^{\text{app}} = U'_{\text{app},B} + U''_{\text{app},B}$$

$$U_{B1}^{\text{app}} = \left(\frac{1}{q} BB^T - V_{\tilde{\mathbf{b}}}\right) \otimes V_{\tilde{b}}, \quad \text{and} \quad U_{B2}^{\text{app}} = U''_{B2}.$$

The approximate asymptotic covariance matrix Σ^{app} can be used to construct an asymptotic confidence ellipsoid for $\text{vec}(X_0)$, based on the convergence

$$r \cdot (\text{vec}(\hat{\Delta}))^T (\Sigma^{\text{app}})^{-1} \text{vec}(\hat{\Delta}) \xrightarrow{d} \chi^2_{np}, \quad \text{as } r \to \infty, m \to \infty, q \to \infty.$$

7 Small sample correction

7.1 Construction

In [CST00] a small sample estimator for a polynomial regression with errors in the variables was constructed. We apply this approach for the model (4), (5). Our goal is to modify the ALS estimator (11) in such a way that it shows good results in small samples without loosing the asymptotic properties for large samples.

We construct a modification of the ALS estimator as follows. For arbitrary positive integers β and γ, $\beta \le \gamma$, denote

$$f_{\beta\gamma} \triangleq [\underbrace{1 \ldots 1}_{\beta} \underbrace{0 \ldots 0}_{\gamma-\beta}]^T \in \mathbb{R}^{\gamma \times 1}.$$

First we introduce two matrices of the same size, with $t \le q$

$$T_A \triangleq \left[\frac{1}{\sqrt{t}} Cf_{tq} \quad A\right]^T \left[\frac{1}{\sqrt{t}} Cf_{tq} \quad A\right], \quad \text{and} \quad W_A \triangleq \begin{bmatrix} 0 & 0 \\ 0 & V_{\tilde{A}} \end{bmatrix},$$

and let λ_A be the smallest positive root of

$$\det(T_A - \lambda W_A) = 0. \tag{40}$$

Our polynomial measurement error model is of degree $d = 2$. Following the advice of [CST00] we set $\alpha \triangleq d + 1 = 3$. We define

$$\mu_A = \begin{cases} \frac{m-\alpha}{m}, & \text{if } \lambda_A > 1 + \frac{1}{m}, \\ \frac{\lambda_A(m-\alpha)}{m+1}, & \text{if } \lambda_A \le 1 + \frac{1}{m}. \end{cases}$$

Similarly we introduce the other two matrices of equal size, $s \le m$,

$$T_B \triangleq \begin{bmatrix} \frac{1}{\sqrt{s}} C^T f_{sm} & B^T \end{bmatrix}^T \begin{bmatrix} \frac{1}{\sqrt{s}} C^T f_{sm} & B^T \end{bmatrix}, \quad \text{and} \quad W_B \triangleq \begin{bmatrix} 0 & 0 \\ 0 & V_{\tilde{B}} \end{bmatrix},$$

and let λ_B be the smallest positive root of

$$\det(T_B - \lambda W_B) = 0. \tag{41}$$

We set

$$\mu_B = \begin{cases} \frac{q-\alpha}{q}, & \text{if } \lambda_B > 1 + \frac{1}{q}, \\ \frac{\lambda_B(q-\alpha)}{q+1}, & \text{if } \lambda_B \le 1 + \frac{1}{q}. \end{cases}$$

The modified estimator is defined by

$$\hat{X}_{\text{M}} \triangleq (A^T A - \mu_A V_{\tilde{A}})^{\dagger} A^T C B^T (BB^T - \mu_B V_{\tilde{B}})^{\dagger}. \tag{42}$$

7.2 T_A and T_B are positive definite, a.s.

We need the next assumption.

(xiii). The following distributions have no atoms

$$\mathscr{L}(\tilde{c}_{il}),\ i \ge 1, l \ge 1, \quad \mathscr{L}(\tilde{a}_{ij}),\ i \ge 1, j \le n, \quad \mathscr{L}(\tilde{b}_{kl}),\ k \le p, l \ge 1.$$

We remind that the distribution of a r.v. ξ has no atoms iff $\mathbf{P}(\xi = a) = 0$, for all $a \in \mathbb{R}$.

Lemma 2. *Assume that assumptions (i) and (xiii) hold. Then for all $m \ge n + 1$, $q \ge 1$, $t \le q$, T_A is positive definite a.s., and for all $m \ge 1$, $q \ge p + 1$, $s \le m$, T_B is positive definite a.s.*

Proof. We shall give a proof for T_A only. The sum of independent r.v. with non-atomic distribution also has non-atomic distribution, hence the components of $\frac{1}{\sqrt{t}} C f_{tq}$ have non-atomic distribution. We suppose that $m \ge n + 1$, $q \ge 1$. Denote $\frac{1}{\sqrt{t}} C f_{tq} \triangleq [u_1 \dots u_m]^T$. It is sufficient to show that the vectors

$$h_1^T \triangleq \begin{bmatrix} u_1 \\ a_1 \end{bmatrix}, \dots, h_{n+1}^T \triangleq \begin{bmatrix} u_{n+1} \\ a_{n+1} \end{bmatrix} \tag{43}$$

are linearly independent, a.s. Note that $h_1, \dots, h_{n+1}$ are independent as random vectors. Using induction by $n \ge 1$ we prove the following statement.

Let $h_1, \dots, h_{n+1}$ given in (43) be independent random vectors, $a_i \in \mathbb{R}^{n \times 1}$, $u_i \in \mathbb{R}$, and $u_1, \dots, u_{n+1}$ have non-atomic distribution, and all the coordinates of $a_1, \dots, a_n$ have non-atomic distribution as well. Then

$$\det\begin{bmatrix} h_1 \\ \vdots \\ h_{n+1} \end{bmatrix} \neq 0, \quad \text{a.s.}$$

a). Indeed, for $n = 1$

$$\det\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = u_1 a_2 - u_2 a_1 \quad \text{and} \quad a_1 \neq 0, \quad \text{a.s.}$$

Then

$$\mathbf{P}(u_1 a_2 - u_2 a_1 = 0) = \mathbf{E}(\mathbf{E}[I(u_1 a_2 - u_2 a_1 = 0) \,|\, a_1, a_2]),$$

where $I(\cdot)$ denotes the indicator function of a random event. But for deterministic $a_1, a_2, a_1 \neq 0$, we have: $\mathscr{L}(u_1 a_2 - u_2 a_1)$ is non-atomic because it is a sum of two independent r.v. with non-atomic distribution (if $a_2 \neq 0$) or it is exactly $\mathscr{L}(-u_2 a_1)$, which is non-atomic. Then

$$\mathbf{E}[I(u_1 a_2 - u_2 a_1 = 0) \,|\, a_1, a_2] = 0 \quad \text{a.s.}$$

and $\mathbf{P}(u_1 a_2 - u_2 a_1 = 0) = 0$. We proved the statement for $n = 1$.

b). Suppose it holds for $n - 1 \geq 1$, and we prove it for n.

$$\det\begin{bmatrix} h_1 \\ \vdots \\ h_{n+1} \end{bmatrix} = \sum_{i=1}^{n+1} u_i A_i,$$

where $A_i = A_i(a_1, \ldots, a_{n+1})$ is the corresponding algebraic complement. Here

$$A_{n+1} = \pm\det\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \neq 0 \quad \text{a.s.}$$

by the assumption of induction. Then

$$\mathbf{P}\left(\sum_{i=1}^{n+1} u_i A_i = 0\right) = \mathbf{E}\left(\mathbf{E}\left[I\left(\sum_{i=1}^{n+1} u_i A_i = 0\right) \,|\, a_1, \ldots, a_{n+1}\right]\right) = 0,$$

because for deterministic $A_1, \ldots, A_{n+1}$, $A_{n+1} \neq 0$, we have $\mathscr{L}(\sum_{i=1}^{n+1} u_i A_i)$ is non-atomic. Thus we proved the statement for n.

This accomplishes the proof of the auxiliary statement.

Thus a.s. $\operatorname{rank}\left(\left[\frac{1}{\sqrt{t}} Cf_{tq} \;\; A\right]\right) = n + 1$, and T_A is positive definite a.s. Lemma 2 is proved. □

7.3 The matrices $A^TA - \mu_A V_{\tilde{A}}$ and $BB^T - \mu_B V_{\tilde{B}}$ are positive definite, a.s.

We suppose that $m \geq n+1$, $q \geq p+1$. We may and do assume that T_A and T_B are positive definite. If $V_{\tilde{A}} \neq 0$ and $V_{\tilde{B}} \neq 0$ then λ_A and λ_B exist, because (40) is equivalent to

$$\det\left(T_A^{-1/2} W_A T_A^{-1/2} - \frac{1}{\lambda} I_{n+1} \right) = 0,$$

and this equation has a positive solution λ, as $T_A^{-1/2} W_A T_A^{-1/2}$ is positive semi-definite and $T_A^{-1/2} W_A T_A^{-1/2} \neq 0$; the same for (41).

Now, we show that $A^TA - \mu_A V_{\tilde{A}}$ is positive definite. We have

$$\mu_A \leq \frac{m-\alpha}{m+1} \lambda_A$$

and

$$\begin{aligned} A^TA - \mu_A V_{\tilde{A}} &\geq A^TA - \frac{m-\alpha}{m+1} \lambda_A V_{\tilde{A}} \\ &= \frac{m-\alpha}{m+1} (A^TA - \lambda_A V_{\tilde{A}}) + \frac{\alpha+1}{m+1} A^TA \geq \frac{\alpha+1}{m+1} A^TA > 0. \end{aligned}$$

Similarly $B^TB - \mu_B V_{\tilde{B}}$ is also positive definite. Thus for $m \geq n+1$, $q \geq p+1$ we have a.s.

$$\hat{X}_{\text{M}} = (A^TA - \mu_A V_{\tilde{A}})^{-1} A^T C B^T (BB^T - \mu_B V_{\tilde{B}})^{-1}.$$

7.4 $\hat{X}_{\text{M}}$ has the same asymptotic properties as $\hat{X}$

First we show that $\mathbf{P}(\lambda_A > 1) \to 1$ as $m, q \to \infty$, and the same for λ_B. We need some new assumptions.

(xiv). $t = t_m = o(\sqrt[3]{m})$, $m \to \infty$, $t \leq q$ and $s = s_q = o(\sqrt[3]{q})$, $q \to \infty$, $s \leq m$ and t_m, s_q are nondecreasing sequences of numbers. E.g., it is possible to set $t = [m^{1/4}]$ and $s = [q^{1/4}]$, where $[\cdot]$ are Gaussian brackets, if $m/q \to \lambda$, with arbitrary $\lambda \in (0, \infty)$.

(xv). $\{\tilde{c}_{il}, i \geq 1, l \geq 1\}$ are identically distributed and $\mathbf{E}\tilde{c}_{11}^2 > 0$.

Lemma 3. *Assume (i) to (iii), (vii), (viii), (viii)′, (xiv) and (xv). Then for $m \to \infty$, $q \to \infty$, we have*

$$\mathbf{P}(\lambda_A > 1) \to 1 \quad \textit{and} \quad \mathbf{P}(\lambda_B > 1) \to \infty.$$

Proof. See Appendix.

Theorem 4. *Assume (i) to (iii), (vii), (viii), (viii)′, (xiv) and (xv). Then for $m = m(u)$, $q = q(u)$, $u/m(u) \to \lambda_A$, $u/q(u) \to \lambda_B$, as $u \to \infty$, with $0 \leq \lambda_A < \infty$, $0 \leq \lambda_B < \infty$, $\lambda_A + \lambda_B > 0$, we have*

$$p \min_{u \to \infty} \sqrt{u} (\hat{X}_{\text{M}} - \hat{X}) = 0.$$

Proof. We follow the line of [CST00]. From Lemma 3 and the definition of μ_A, see Subsection 7.1, we obtain that with probability tending to one,

$$\frac{m-\alpha}{m+1} \le \mu_A \le \frac{m-\alpha}{m},$$

and

$$0 \le (\mu_A - 1)\sqrt{m} \le \frac{\alpha}{\sqrt{m}}.$$

Therefore

$$p \lim_{\substack{m\to\infty \\ q\to\infty}} \sqrt{m}(\mu_A - 1) = 0. \tag{44}$$

Similarly

$$p \lim_{\substack{m\to\infty \\ q\to\infty}} \sqrt{q}(\mu_B - 1) = 0. \tag{45}$$

Now, we consider the difference between the estimators $\hat{X} - \hat{X}_{\text{M}}$. From (9) and (42) we have

$$(A^TA - V_{\tilde{A}})\hat{X}(BB^T - V_{\tilde{B}}) = (A^TA - \mu_A V_{\tilde{A}})\hat{X}_M(BB^T - \mu_B V_{\tilde{B}}).$$

We set $\mu_A = 1 + \frac{h_1}{\sqrt{m}}$, $\mu_B = 1 + \frac{h_2}{\sqrt{q}}$, $h_1 \xrightarrow{\text{P}} 0$, $h_2 \xrightarrow{\text{P}} 0$ and $\hat{X}_{\text{M}} \triangleq \hat{X} + \tilde{X}$.

We have

$$\begin{aligned}
&(A^TA - \mu_A V_{\tilde{A}} + (\mu_A - 1)V_{\tilde{A}})\hat{X}(BB^T - \mu_B V_{\tilde{B}} + (\mu_B - 1)V_{\tilde{B}}) \\
&\qquad = (A^TA - \mu_A V_{\tilde{A}})(\hat{X} + \tilde{X})(BB^T - \mu_B V_{\tilde{B}}), \\
&(A^TA - \mu_A V_{\tilde{A}})\hat{X}(BB^T - \mu_B V_{\tilde{B}}) \\
&\qquad = \frac{h_1}{\sqrt{m}} V_{\tilde{A}}\hat{X}(BB^T - V_{\tilde{B}}) + \frac{h_2}{\sqrt{q}}(A^TA - V_{\tilde{A}})\hat{X}V_{\tilde{B}} + \frac{h_1h_2}{\sqrt{mq}} V_{\tilde{A}}\hat{X}V_{\tilde{B}}.
\end{aligned}$$

But under the assumptions of Section 4 we have as $\sqrt{m}(\mu_A - 1) \xrightarrow{\text{P}} 0$, $\sqrt{q}(\mu_B - 1) \xrightarrow{\text{P}} 0$,

$$\begin{aligned}
(I_n + \text{o}_\text{p}(1))\tilde{X}(I_p + \text{o}_\text{p}(1)) &= \frac{h_1}{\sqrt{m}}\text{O}_\text{p}(1) + \frac{h_2}{\sqrt{q}}\text{O}_\text{p}(1) + \frac{h_1h_2}{\sqrt{mq}}\text{O}_\text{p}(1) \\
&= \frac{\text{o}_\text{p}(1)}{\sqrt{m}} + \frac{\text{o}_\text{p}(1)}{\sqrt{q}} + \frac{\text{o}_\text{p}(1)}{\sqrt{mq}}.
\end{aligned}$$

We assume that $m = m(u)$, $q = q(u)$, and we have as $u \to \infty$ that $m \to \infty$, $q \to \infty$ and

$$\frac{u}{m} \to \lambda_A, \quad \frac{u}{q} \to \lambda_B, \quad 0 \le \lambda_A < \infty, \quad 0 \le \lambda_B < \infty, \quad \lambda_A + \lambda_B > 0.$$

Then we have as $u \to \infty$

$$(I_n + \mathrm{o_p}(1))(\sqrt{u}\tilde{X})(I_p + \mathrm{o_p}(1))$$
$$= \sqrt{\frac{u}{m}}\mathrm{o_p}(1) + \sqrt{\frac{u}{q}}\mathrm{o_p}(1) + \sqrt{\frac{u}{m}}\sqrt{\frac{u}{q}}\,\frac{\mathrm{o_p}(1)}{\sqrt{u}}. \qquad \square$$

From Theorem 4 under the conditions of asymptotic normality of $\hat{X}$, see Subsections 6.2 and 6.3, we will have

$$\sqrt{u} \cdot \mathrm{vec}(\hat{X}_{\mathrm{M}} - \hat{X}) \xrightarrow{d} N(0, \lambda_A \Sigma_A + \lambda_B \Sigma_B).$$

If $\lambda_A \Sigma_A + \lambda_B \Sigma_B > 0$ then we can say that $\hat{X}_{\mathrm{M}}$ and $\hat{X}$ have the same asymptotic properties. It happens, see Subsection 6.3, if

– either $\lambda_A > 0$, $\lambda_B > 0$,
– or $\lambda_A = 0$ and $n \le p$,
– or $\lambda_B = 0$ and $n \ge p$.

Thus in the case $n = p$ we guarantee that $\hat{X}_{\mathrm{M}}$ and $\hat{X}$ are asymptotically equivalent only for the convergence $m \to \infty$, $q \to \infty$, $m/q \to \lambda$, $\lambda \in (0, \infty)$.

8 Examples

In this section we apply the ALS estimator to a hypothetical example.

We consider the model (4), (5) with $m = q$ and $n = p = 2$, i.e.

$$\underbrace{A}_{m\times 2}\,\underbrace{X}_{2\times 2}\,\underbrace{B}_{2\times m} = \underbrace{C}_{m\times m}.$$

The true data is

$$A_0 = \begin{bmatrix} I_2 \\ \vdots \\ I_2 \end{bmatrix}, \quad B_0 = [I_2 \ldots I_2], \quad \text{and} \quad C_0 = \begin{bmatrix} I_2 \cdots I_2 \\ \vdots \quad\quad \vdots \\ I_2 \cdots I_2 \end{bmatrix},$$

and the true value of the parameter is $X_0 = I_2$. The perturbations $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ are selected in three different ways.

1. *Equally sized errors*. All elements a_{ij}, b_{kl}, c_{il} are independent, centered, and normally distributed with common variance 0.01.
2. *Differently sized errors*. All elements a_{ij}, b_{kl}, c_{il} are independent, centered, and normally distributed. The elements in the first column of $\tilde{A}$ have variance 0.05 and the elements in the second column of $\tilde{A}$ have variance 0.01. The elements in the first row of $\tilde{B}$ have variance 0.01 and the elements in the second row of $\tilde{B}$ have variance 0.05. All elements of $\tilde{C}$ have variance 0.01;

3. *Correlated errors.* All elements a_{ij}, b_{kl}, c_{il} are independent and normally distributed. All rows of $\tilde{A}$ have covariance 0.01 $\left[\begin{smallmatrix}5 & 1\\ 1 & 1\end{smallmatrix}\right]$ and the elements are independent from row to row. All columns of $\tilde{B}$ have covariance 0.01 $\left[\begin{smallmatrix}1 & 1\\ 1 & 5\end{smallmatrix}\right]$ and the elements are independent from column to column.

The estimation is performed for increasing number of measurements m. As measure of the estimation quality, we use the empirical (relative) mean square error

$$e(m) = \frac{1}{N}\sum_{s=1}^{N} \frac{\|X_0 - \hat{X}^{(s)}\|_F^2}{\|X_0\|_F^2},$$

where $\hat{X}^{(s)}$ is the estimate computed for the s-th noise realization.

The ALS and the small sample modified ALS estimators are compared with the LS estimator

$$\hat{X}_{\mathrm{LS}} \triangleq (A^TA)^{\dagger}A^TCB^T(BB^T)^{\dagger},$$

and with the partial LS estimators,

$$\hat{X}_{\mathrm{PA}} \triangleq \text{TLS solution of } XB = (A^TA)^{\dagger}A^TC$$

and

$$\hat{X}_{\mathrm{PB}} \triangleq \text{TLS solution of } AX = CB^T(BB^T)^{\dagger}.$$

Figure 1 shows small sample size result for equally sized errors; the number of measurements m is between 10 and 20. On the left plot is the mean square error of estimation $e(m)$ for LS (dotted line), ALS (solid line), small sample modified ALS (dashed-dotted line) and partial LS (dashed lines) estimators. The plots are averaged for $N = 200$ noise realizations.

The right plot of Figure 1 illustrates application of the asymptotic normality results for confidence region computation, see Subsection 6.4. The

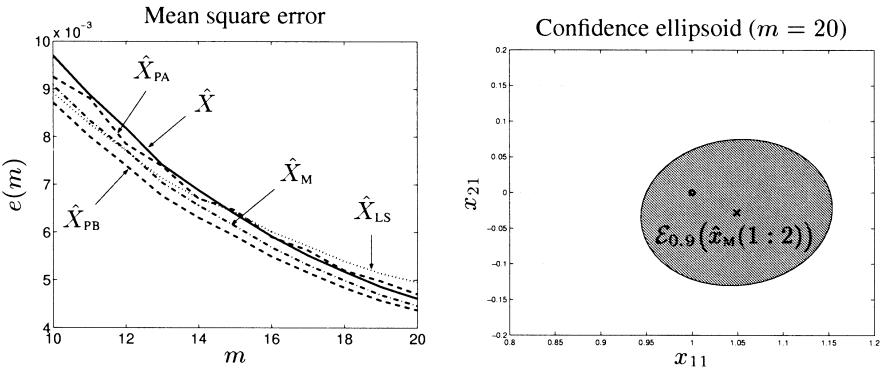

Fig. 1. Equally sized errors, small sample size results

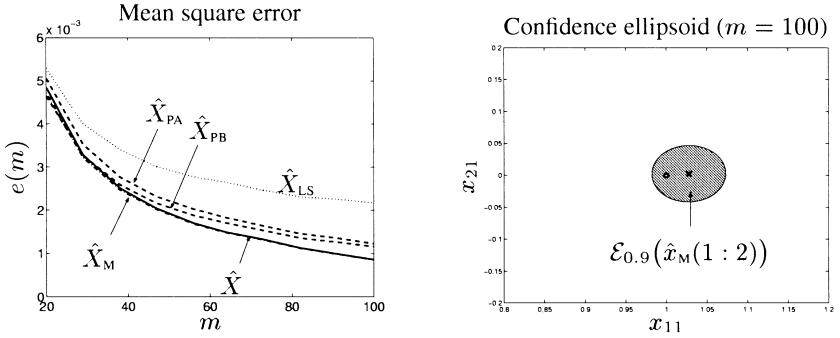

Fig. 2. Equally sized errors, result for $m \in \{20, \ldots, 100\}$

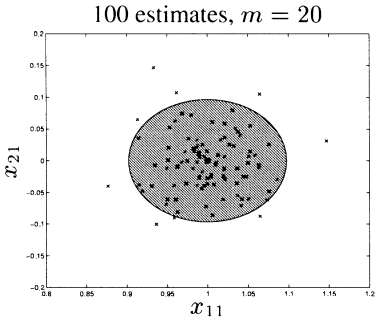

Table 1. Percentage of estimates inside

	inside
$\hat{x}_M(1:2) \in \mathscr{E}_{0.9}(\hat{x}_M(1:2))$	89%
$\hat{x}_M \in \mathscr{E}_{0.9}(\hat{x}_M)$	91%

Fig. 3. Clustering of the modified ALS estimates

confidence region of $\hat{x}_M \triangleq \mathrm{vec}(\hat{X}_M)$ with $1-\alpha$ confidence probability is the ellipsoid

$$\mathscr{E}^{\mathrm{app}}_{1-\alpha}(\hat{x}_M) = \left\{ x \,|\, (x-\hat{x}_M)(\Sigma^{\mathrm{app}})^{-1}(x-\hat{x}_M) \le \frac{1}{m}(\chi^2_l)_\alpha \right\},$$

where $(\chi^2_l)_\alpha$ is the α quantile of the χ^2_l distribution, i.e. $\mathbf{P}(\chi^2_l \ge (\chi^2_l)_\alpha) = \alpha$, and l is the number of degrees of freedom. For $\hat{x}_M$ in the example, $l = 4$.

In order to be able to visualize the results, we use the first two elements of $\hat{x}_M$, denoted by $\hat{x}_M(1:2)$. For $\hat{x}_M(1:2)$, $l = 2$ and the approximate asymptotic covariance matrix is the upper left, 2 by 2 submatrix of Σ^{app}.

The computed confidence region $\mathscr{E}^{\mathrm{app}}_{0.9}(\hat{x}_M)$ for $m = 20$ is shown as shaded area on the plot. The symbol "$\circ$" indicates the true value point $[1\ 0]^T$ and the symbol "$\times$" indicates the point estimate $\hat{x}_M(1:2)$.

Figure 2 shows analogous results for equally sized errors for larger sample size; m between 20 and 100.

Figure 3, shows how the estimates are clustered. Again "$\circ$" corresponds to the true value $[1\ 0]^T$. The "$\times$" symbols correspond to 100 estimates $\hat{x}_M(1:2)$. The shaded area is the ellipsoid,

$$\mathscr{E}_{1-\alpha}(\hat{x}_M) = \left\{ x \,|\, (x-\hat{x}_M)\Sigma^{-1}(x-\hat{x}_M) \le \frac{1}{m}(\chi^2_l)_\alpha \right\},$$

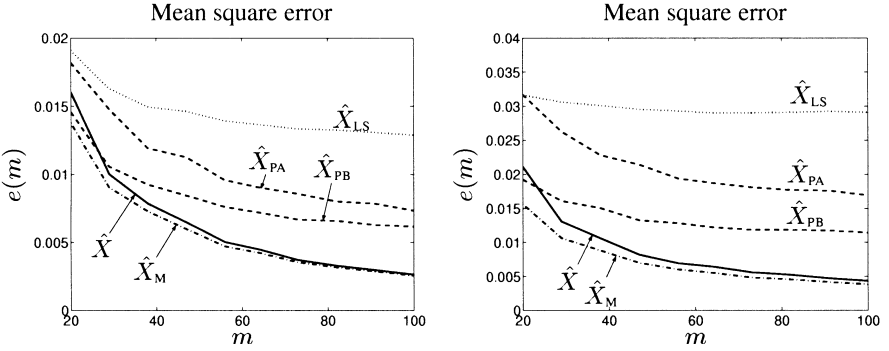

Fig. 4. Left: differently sized errors; right: correlated errors

described by the true asymptotic covariance matrix Σ, see (39), and centered at the true value $\hat{x}_{\mathrm{M}}(1:2)$.

Figure 4 show the mean square error of the compared estimators for differently sized uncorrelated errors (left plot) and for correlated errors (right plot).

9 Conclusion

We considered the multivariable model $AXB = C$. In the situation when the TLS estimator is inconsistent, we construct the ALS estimator, which is consistent. We gave the conditions of weak and strong consistency, and of asymptotic normality. It turns out that the asymptotic covariance matrix of the estimator does not depend upon the covariance structure of $\tilde{C}$. We introduced a small sample modification of the estimator, which has better properties for small samples and preserves the asymptotic properties of the estimator.

An open question is what are the optimality properties of the ALS estimator. In [KM00] for the model $AX = B$ in the scalar case it was shown that the ALS estimator is asymptotically efficient in the situation where $V_{\tilde{A}}$ is known exactly and $\mathbf{E}\tilde{b}_{kl}^2$ are known up to a constant factor. It would be interesting to check the following conjecture:

> In the model $AXB = C$ the ALS estimator is asymptotically efficient in the situation where $V_{\tilde{A}}$ and $V_{\tilde{B}}$ are known exactly and $\mathbf{E}\tilde{c}_{il}^2$ are known up to a constant factor.

10 Appendix

10.1 Proof of (14)

We have

$$V_{A_0}^{-1} U_A = I_n + V_{A_0}^{-1}(\tilde{A}^T\tilde{A} - V_{\tilde{A}}) + V_{A_0}^{-1}(\tilde{A}^T A_0 + A_0^T\tilde{A}). \tag{46}$$

Next

$$\mathbf{E}\|V_{A_0}^{-1}(\tilde{A}^T\tilde{A} - V_{\tilde{A}})\|_F^2 \le \|V_{A_0}^{-1}\|^2 \mathbf{E}\|\tilde{A}^T\tilde{A} - V_{\tilde{A}}\|_F^2 \tag{47}$$

and $\|V_{A_0}^{-1}\| = \lambda_{\min(V_{A_0})}^{-1}$. By (ii) and (iii), we have

$$\mathbf{E}\|\tilde{A}^T\tilde{A} - V_{\tilde{A}}\|_F^2 = \sum_{i,k=1}^{n} \operatorname{var}\left(\sum_{j=1}^{m} \tilde{a}_{ji}\tilde{a}_{jk}\right)$$

$$= \sum_{i,k=1}^{n}\sum_{j=1}^{m} \operatorname{var}(\tilde{a}_{ji}\tilde{a}_{jk}) \le \text{const} \cdot m,$$

thus by (iv)

$$\mathbf{E}\|V_{A_0}^{-1}(\tilde{A}^T\tilde{A} - V_{\tilde{A}})\|_F^2 \le \frac{\text{const} \cdot m}{\lambda_{\min}^2(V_{A_0})} \to 0 \quad \text{as } m \to \infty, \tag{48}$$

this proves that $V_{A_0}^{-1}(\tilde{A}^T\tilde{A} - V_{\tilde{A}}) \xrightarrow{\mathrm{P}} 0$.

Next

$$\mathbf{E}\|V_{A_0}^{-1}(\tilde{A}^T A_0 + A_0^T\tilde{A})\|_F^2 \le \frac{2}{\lambda_{\min}^2(V_{A_0})}\mathbf{E}\|A_0^T\tilde{A}\|_F^2 \tag{49}$$

and

$$\mathbf{E}\|A_0^T\tilde{A}\|_F^2 = \sum_{i,k=1}^{n} \mathbf{E}\left(\sum_{j=1}^{m} a_{ji}^0\tilde{a}_{jk}\right)^2 = \sum_{i,k=1}^{n}\sum_{j=1}^{m}(a_{ji}^0)^2 \operatorname{var}(\tilde{a}_{jk})$$

$$\le \text{const} \cdot \sum_{i=1}^{n}\sum_{j=1}^{m}(a_{ji}^0)^2 \le \text{const} \cdot \lambda_{\max}(V_{A_0})$$

and from (49) under assumption (iv) we have

$$V_{A_0}^{-1}(\tilde{A}^T A_0 + A_0^T\tilde{A}) \xrightarrow{\mathrm{P}} 0. \tag{50}$$

Now, (46), (48) and (50) imply the first relation in (14), and the second one in (14) holds similarly.

10.2 Proof of (15)

$$V_{A_0}^{-1}(A^T A_0) = I_n + V_{A_0}^{-1}\tilde{A}^T A_0,$$

and this converges in probability to I_n, see Subsection 10.1. The second condition in (15) is shown similarly.

10.3 Proof of (16)

We have

$$\|V_{A_0}^{-1}A^T\tilde{C}B^TV_{B_0}^{-1}\|_F \le \frac{1}{\lambda_{\min}(V_{A_0})\lambda_{\min}(V_{B_0})}\|A^T\tilde{C}B^T\|_F, \tag{51}$$

and

$$\begin{aligned}\mathbf{E}\|A^T\tilde{C}B^T\|_F^2 &= \sum_{j=1}^{n}\sum_{k=1}^{p}\mathbf{E}\left(\sum_{i=1}^{m}\sum_{l=1}^{q}a_{ij}\tilde{c}_{il}b_{kl}\right)^2 = \sum_{i,j,k,l}\mathbf{E}a_{ij}^2\operatorname{var}(\tilde{c}_{il})\mathbf{E}b_{kl}^2\\ &\le \text{const}\cdot\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{E}a_{ij}^2\right)\left(\sum_{k=1}^{p}\sum_{l=1}^{q}\mathbf{E}b_{kl}^2\right)\\ &\le \text{const}\cdot(\lambda_{\max}(V_{A_0})+m)(\lambda_{\max}(V_{B_0})+q).\end{aligned}$$

Then from (51) we have

$$\mathbf{E}\|V_{A_0}^{-1}A^T\tilde{C}B^TV_{B_0}^{-1}\|_F^2 \le \text{const}\cdot\frac{(\lambda_{\max}(V_{A_0})+m)}{\lambda_{\min}^2(V_{A_0})}\cdot\frac{(\lambda_{\max}(V_{B_0})+q)}{\lambda_{\min}^2(V_{B_0})},$$

and this tends to zero, as $m\to\infty$, $q\to\infty$ by assumption (iv).

10.4 Proof of Theorem 2

We have to show that in (14) to (16) the convergence is with probability one. After that the statement of Theorem 2 will follow from equation (13) for the estimator $\hat{X}$.

We use Lemma 1. First, see (46). Consider

$$\mathbf{E}\|\tilde{A}^T\tilde{A}-V_{\tilde{A}}\|_F^r = \mathbf{E}\left\|\sum_{i=1}^{m}(\tilde{a}_i^T\tilde{a}_i-\mathbf{E}\tilde{a}_i^T\tilde{a}_i)\right\|_F^r \le \text{const}\cdot m^{r/2},$$

because by (v) $\mathbf{E}\|\tilde{a}\|^{2r}\le\text{const}$. Then

$$\mathbf{E}\|V_{A_0}^{-1}(\tilde{A}^T\tilde{A}-V_{\tilde{A}})\|_F^r \le \text{const}\cdot\frac{m^{r/2}}{\lambda_{\min}^r(V_{A_0})},$$

$$\sum_{m=m_0}^{\infty}\frac{m^{r/2}}{\lambda_{\min}^r(V_{A_0})}<\infty$$

and by Borel-Cantelli lemma

$$V_{A_0}^{-1}(\tilde{A}^T\tilde{A}-V_{\tilde{A}})\to 0 \quad \text{a.s.} \quad \text{as } m\to\infty.$$

Next, consider

$$\|\tilde{A}^T A_0 + A_0^T \tilde{A}\|_F \le 2\|A_0^T \tilde{A}\|_F$$

and

$$\mathbf{E}\|A_0^T \tilde{A}\|_F^{2r} = \mathbf{E}\left\|\sum_{i=1}^m a_i^{0T} \tilde{a}_i\right\|_F^{2r} \le \text{const} \cdot \left(\sum_{i=1}^m \|a_i^0\|^2\right)^r \le \text{const} \cdot \lambda_{\max}^r(V_{A_0}).$$

Therefore,

$$\mathbf{E}\| V_{A_0}^{-1}(\tilde{A}^T A_0 + A_0^T \tilde{A})\|^{2r} \le \text{const} \cdot \frac{\lambda_{\max}^r(V_{A_0})}{\lambda_{\min}^{2r}(V_{A_0})},$$

$$\sum_{m=m_0}^{\infty} \frac{\lambda_{\max}^r(V_{A_0})}{\lambda_{\min}^{2r}(V_{A_0})} \le \infty,$$

and by Borel-Cantelli lemma

$$V_{A_0}^{-1}(\tilde{A}^T A_0 + A_0^T \tilde{A}) \to 0 \quad \text{a.s.,} \quad \text{as } m \to \infty.$$

Thus, see (46), $V_{A_0}^{-1} U_A \to I_n$ a.s., as $m \to \infty$. Similarly, $U_B V_{B_0}^{-1} \to I_p$ a.s., as $q \to \infty$.

Secondly, in (15) we also have convergence a.s., compare with the proof of Theorem 1.

Thirdly, consider

$$\mathbf{E}[\|A^T \tilde{C} B^T\|_F^{2r} \mid \tilde{A}, \tilde{B}] = \mathbf{E}\left[\left\|\sum_{i=1}^m \sum_{l=1}^q a_i^T \tilde{c}_{il} b_l^T\right\|_F^{2r} \mid \tilde{A}, \tilde{B}\right]$$

$$\le \text{const} \cdot \left(\sum_{i=1}^m \sum_{l=1}^q \mathbf{E}[\|a_i^T \tilde{c}_{il} b_l^T\|^{2r} \mid \tilde{A}, \tilde{B}] + \left(\sum_{i=1}^m \sum_{l=1}^q \mathbf{E}[\|a_i^T \tilde{c}_{il} b_l^T\|^2 \mid \tilde{A}, \tilde{B}]\right)^r\right).$$

Now,

$$\sum_{i=1}^m \sum_{l=1}^q \mathbf{E}[\|a_i^T \tilde{c}_{il} b_l^T\|^{2r} \mid \tilde{A}, \tilde{B}] \le \text{const} \cdot \sum_{i=1}^m \|a_i\|^{2r} \sum_{l=1}^q \|b_l\|^{2r},$$

and

$$\mathbf{E}\left(\sum_{i=1}^m \sum_{l=1}^q \mathbf{E}[\|a_i^T \tilde{c}_{il} b_l^T\|^{2r} \mid \tilde{A}, \tilde{B}]\right) \le \text{const} \cdot (\lambda_{\max}^r(V_{A_0}) + m)(\lambda_{\max}^r(V_{B_0}) + q).$$

Next,

$$\mathbf{E}\left(\sum_{i=1}^{m}\sum_{l=1}^{q}\mathbf{E}[\|a_i^T\tilde{c}_{il}b_l^T\|^2 \mid \tilde{A},\tilde{B}]\right)^r$$

$$\leq \text{const}\cdot\mathbf{E}\left(\sum_{i=1}^{m}\|a_i\|^2\right)^r\mathbf{E}\left(\sum_{l=1}^{q}\|b_l\|^2\right)^r$$

$$\leq \text{const}\cdot(\lambda_{\max}^r(V_{A_0})+m^r)(\lambda_{\max}^r(V_{B_0})+q^r)$$

Therefore

$$\mathbf{E}\|A^T\tilde{C}B^T\|_F^{2r} \leq \text{const}\cdot(\lambda_{\max}^r(V_{A_0})+m^r)(\lambda_{\max}^r(V_{B_0})+q^r),$$

and, see (46),

$$\mathbf{E}\|V_{A_0}^{-1}A^T\tilde{C}B^TV_{B_0}^{-1}\|_F^{2r} \leq \text{const}\cdot\frac{\lambda_{\max}^r(V_{A_0})+m^r}{\lambda_{\min}^{2r}(V_{A_0})}\,\frac{\lambda_{\max}^r(V_{B_0})+q^r}{\lambda_{\min}^{2r}(V_{B_0})},$$

$$\sum_{m=m_0}^{\infty}\sum_{q=q_0}^{\infty}\frac{\lambda_{\max}^r(V_{A_0})+m^r}{\lambda_{\min}^{2r}(V_{A_0})}\,\frac{\lambda_{\max}^r(V_{B_0})+q^r}{\lambda_{\min}^{2r}(V_{B_0})}$$

$$=\left(\sum_{m=m_0}^{\infty}\frac{\lambda_{\max}^r(V_{A_0})+m^r}{\lambda_{\min}^{2r}(V_{A_0})}\right)\left(\sum_{q=q_0}^{\infty}\frac{\lambda_{\max}^r(V_{B_0})+q^r}{\lambda_{\min}^{2r}(V_{B_0})}\right)<\infty.$$

Then for each $\varepsilon>0$,

$$\mathbf{P}(\|V_{A_0}^{-1}A^T\tilde{C}B^TV_{B_0}^{-1}\|>\varepsilon)\leq\frac{1}{\varepsilon^{2r}}\mathbf{E}\|V_{A_0}^{-1}A^T\tilde{C}B^TV_{B_0}^{-1}\|^{2r}$$

and by Borel-Cantelli lemma with probability one the event

$$\mathscr{D}_{mq}=\{\|V_{A_0}^{-1}A^T\tilde{C}B^TV_{B_0}^{-1}\|>\varepsilon\}$$

happens only for finite number of indices m and q. Then almost surely there exists $m_1=m_1(\omega)$ and $q_1=q_1(\omega)$, such that for all $m\geq m_1$, $q\geq q_1$, $\|V_{A_0}^{-1}A^T\tilde{C}B^TV_{B_0}^{-1}\|\leq\varepsilon$.

This means that $V_{A_0}^{-1}A^T\tilde{C}B^TV_{B_0}^{-1}\to 0$ a.s., as $m\to\infty$, $q\to\infty$.

We proved that in (14) to (16) the convergence is with probability one and Theorem 2 is proved. □

10.5 Proof of Lemma 3

We give the proof for λ_A only. It is sufficient to show that

$$\mathbf{P}(T_A>W_A)\to 1 \quad \text{as } m\to\infty,\ q\to\infty.$$

We have

$$T_A - W_A = \begin{bmatrix} \frac{1}{t} f_{tq}^T C^T C f_{tq} & \frac{1}{\sqrt{t}} f_{tq}^T C^T A \\ \frac{1}{\sqrt{t}} A^T C f_{tq} & A^T A - V_{\tilde{A}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{t}} C_0 f_{tq} & A_0 \end{bmatrix}^T \begin{bmatrix} \frac{1}{\sqrt{t}} C_0 f_{tq} & A_0 \end{bmatrix} + \begin{bmatrix} \frac{1}{t} f_{tq}^T \tilde{C}^T \tilde{C} f_{tq} & 0 \\ 0 & 0 \end{bmatrix}$$

$$+ \begin{bmatrix} \frac{1}{t} f_{tq}^T (C_0^T \tilde{C} + \tilde{C}^T C_0) f_{tq} & \frac{1}{\sqrt{t}} f_{tq}^T (C_0^T \tilde{A} + \tilde{C}^T A_0) \\ \frac{1}{\sqrt{t}} (\tilde{A}^T C_0 + A_0^T \tilde{C}) f_{tq} & A_0^T \tilde{A} + \tilde{A}^T A_0 + (\tilde{A}^T \tilde{A} - V_{\tilde{A}}) \end{bmatrix}$$

$$\triangleq H_1 + H_2 + H_3.$$

10.5.1 Behavior of H_1

We have $H_1 \geq 0$, and $H_1 = \begin{bmatrix} H_{11}' & H_{12}' \\ H_{21}' & H_{22}' \end{bmatrix}$,

$$\frac{1}{m} H_{22}' = \frac{1}{m} A_0^T A_0 \to V_{A\infty} > 0, \quad \text{see assumption (vii).}$$

10.5.2 Behavior of H_2

We have $H_2 = \begin{bmatrix} H_{11}'' & 0 \\ 0 & 0 \end{bmatrix}$,

$$\frac{1}{m} H_{11}'' = \frac{1}{mt} \sum_{i=1}^{m} \sum_{k=1}^{t} \tilde{c}_{ik} \sum_{l=1}^{t} \tilde{c}_{il} = \frac{1}{mt} \sum_{\substack{1 \leq i \leq m \\ 1 \leq k \leq t}} \tilde{c}_{ik}^2 + \frac{2}{mt} \sum_{\substack{1 \leq i \leq m \\ 1 \leq k < l \leq t}} \tilde{c}_{ik} \tilde{c}_{il} \triangleq S_A + S_B.$$

We have

$$\frac{1}{4} \mathbf{E} S_B^2 = \frac{1}{m^2 t^2} \sum_{\substack{1 \leq i \leq m \\ 1 \leq k < l \leq t}} \mathbf{E}(\tilde{c}_{ik}^2) \mathbf{E}(\tilde{c}_{il}^2) \leq \text{const} \cdot \frac{mt^2}{m^2 t^2} \to 0 \quad \text{as } m \to \infty,\ q \to \infty.$$

Next, by assumption (xv) and by the law of large numbers

$$S_A \to \mathbf{E} \tilde{c}_{11}^2 \quad \text{a.s.} \quad \text{as } m \to \infty,\ q \to \infty$$

Thus

$$H_2 \to \begin{bmatrix} \mathbf{E}\tilde{c}_{11}^2 & 0 \\ 0 & 0 \end{bmatrix} \text{ a.s.}$$

10.5.3 Behavior of H_3

We prove that $\frac{1}{m}H_3 \xrightarrow{\text{P}} 0$. We write $H_3 = \begin{bmatrix} H_{11}''' & H_{12}''' \\ H_{21}''' & H_{22}''' \end{bmatrix}$,

a).

$$\begin{aligned}
\frac{1}{m^2}\mathbf{E}\left(\frac{1}{t}f_{tq}^T C_0^T \tilde{C} f_{tq}\right)^2 &= \frac{1}{m^2}\mathbf{E}\left(\frac{1}{t}f_{tq}^T \tilde{C}^T C_0 f_{tq}\right)^2 \\
&= \frac{1}{m^2 t^2}\mathbf{E}\left(\sum_{i=1}^{m}\sum_{k=1}^{t} c_{ik}^0 \sum_{l=1}^{t}\tilde{c}_{il}\right)^2 \\
&= \frac{1}{m^2 t^2}\sum_{i=1}^{m}\sum_{l=1}^{t}\mathbf{E}\tilde{c}_{11}^2\left(\sum_{k=1}^{t} c_{ik}^0\right)^2 \\
&\le \text{const}\cdot\frac{1}{m^2}\sum_{i=1}^{m}\sum_{l=1}^{t}(c_{ik}^0)^2 \\
&\le \text{const}\cdot\frac{1}{m^2}\lambda_{\max}(V_{A_0})\lambda_{\max}\left(\sum_{i=1}^{t} b_i^0 b_i^{0T}\right) \\
&\le \text{const}\cdot\frac{t}{m} \to 0,
\end{aligned}$$

because $t = \text{o}(\sqrt[3]{m})$. Thus $H_{11}''' \xrightarrow{\text{P}} 0$.

b).

$$\begin{aligned}
\frac{1}{m^2}\mathbf{E}\left\|\frac{1}{\sqrt{t}}\tilde{A}^T C_0 f_{tq}\right\|^2 &= \frac{1}{m^2 t}\mathbf{E}\sum_{k=1}^{n}\left(\sum_{i=1}^{m}\tilde{a}_{ik}\sum_{l=1}^{t} c_{il}^0\right)^2 \\
&\le \text{const}\cdot\frac{1}{m^2 t}\sum_{k=1}^{n}\sum_{i=1}^{m}\left(\sum_{l=1}^{t} c_{il}^0\right)^2 \\
&\le \text{const}\cdot\frac{1}{m^2}\sum_{i=1}^{m}\sum_{l=1}^{t}(c_{il}^0)^2 \\
&\le \text{const}\cdot\frac{t}{m} \to 0.
\end{aligned}$$

c).

$$\frac{1}{m^2}\mathbf{E}\left\|\frac{1}{\sqrt{t}}A_0^T\tilde{C}f_{tq}\right\|^2 = \frac{1}{m^2 t}\mathbf{E}\sum_{k=1}^{n}\left(\sum_{i=1}^{m}a_{ik}^0\sum_{l=1}^{t}\tilde{c}_{il}\right)^2$$

$$= \frac{1}{m^2 t}\sum_{k=1}^{n}\sum_{l=1}^{t}\mathbf{E}\tilde{c}_{11}^2\left(\sum_{i=1}^{m}a_{ik}^0\right)^2$$

$$\le \text{const}\cdot\frac{1}{m^2}\sum_{i=1}^{m}\sum_{k=1}^{n}(a_{ik}^0)^2$$

$$\le \text{const}\cdot\frac{t}{m}\to 0.$$

From b) and c) we have $H_{12}''' \xrightarrow{\text{P}} 0$, $H_{21}''' \xrightarrow{\text{P}} 0$.

d). It was shown above that $H_{22}''' \xrightarrow{\text{P}} 0$, see section 10.1. Finally, $H_3 \xrightarrow{\text{P}} 0$.

10.5.4 End of proof

Summarizing Subsection 10.5.3, we have

$$\left\|\frac{1}{m}H_3\right\|_F = \sqrt{\frac{t}{m}}\cdot \text{O}_\text{p}(1).$$

We need to know the behavior of the other blocks of H_1.

$$\frac{1}{m}H_{11}' = \frac{1}{mt}\sum_{i=1}^{m}\left(\sum_{l=1}^{t}c_{il}^0\right)^2$$

$$\le \frac{1}{m}\sum_{i=1}^{m}\sum_{l=1}^{t}(c_{il}^0)^2$$

$$\le \text{const}\cdot t.$$

$$\frac{1}{m}H_{21}' = \frac{1}{m\sqrt{t}}A_0^T C_0 f_{tq} = \frac{1}{m}A_0^T A_0\cdot\frac{1}{\sqrt{t}}X_0 B_0 f_{tq},$$

$$\left\|\frac{1}{m}H_{21}'\right\|_F \le \text{const}\cdot\sqrt{t},$$

and

$$\left\|\frac{1}{m}H_{12}'\right\|_F = \left\|\frac{1}{m}H_{21}'\right\|_F \le \text{const}\cdot\sqrt{t}.$$

We consider the matrix

$$M_A \triangleq \frac{1}{m}(H_1 + H_2).$$

It is positive semidefinite and we look for $\varepsilon_0 > 0$ such that $M_A \geq \frac{\varepsilon_0}{t} I_{n+1}$ with probability tending to one. We have

$$M_A - \frac{\varepsilon_0}{t} I_{n+1} = \begin{bmatrix} \frac{1}{m} H'_{11} + \left(\frac{1}{m} H''_{11} - \frac{\varepsilon_0}{t}\right) & \frac{1}{m} H'_{12} \\ \frac{1}{m} H'_{21} & \frac{1}{m} H'_{22} - \frac{\varepsilon_0}{t} I_n \end{bmatrix}.$$

Now, for $m \geq m_0$

$$\frac{1}{m} H'_{22} - \frac{\varepsilon_0}{t} I_n \geq \left(\frac{1}{2}\lambda_{\min}(V_{A\infty}) - \frac{\varepsilon_0}{t}\right) I_n \geq \left(\frac{1}{2}\lambda_{\min}(V_{A\infty}) - \varepsilon_0\right) I_n$$

$$\geq \frac{1}{4}\lambda_{\min}(V_{A\infty}), \quad \text{if } \varepsilon_0 \leq \frac{1}{4}\lambda_{\min}(V_{A\infty}).$$

We apply Silvester's criterion to the matrix $M_A - \frac{\varepsilon_0}{t} I_{n+1}$. We have for

$$\varepsilon_0 \leq \frac{1}{4}\lambda_{\min}(V_{A\infty}), \tag{52}$$

$$\det\left(M_A - \frac{\varepsilon_0}{t} I_n\right) = \det\left(\frac{1}{m} H_1\right) + \left(\frac{1}{m} H''_{11}\right) \det\left(\frac{1}{m} H'_{22} - \frac{\varepsilon_0}{t} I_n\right) + \varepsilon_0 \cdot \mathrm{O}(1). \tag{53}$$

The last term comes from $\frac{1}{m} H'_{11} \det\left(\frac{1}{m} H'_{22} - \frac{\varepsilon_0}{t} I_n\right)$ and from the product of components $\frac{1}{m} H'_{21}$, $\frac{1}{m} H'_{12}$ and $\frac{\varepsilon_0}{t} I_n$. In both cases we have $t \frac{\varepsilon_0}{t} \cdot \mathrm{O}(1) = \varepsilon_0 \cdot \mathrm{O}(1)$.

Now, $\frac{1}{m} H_1 \geq 0$, therefore (53) implies

$$\det\left(M_A - \frac{\varepsilon_0}{t} I_{n+1}\right) \geq \frac{1}{m} H''_{11} \cdot \det\left(\frac{1}{m} H'_{22} - \frac{\varepsilon_0}{t} I_n\right) - \mathrm{const}_1 \cdot \varepsilon_0,$$

with some $\mathrm{const}_1 > 0$, and a.s. we have

$$\liminf_{\substack{m \to \infty \\ q \to \infty}} \det\left(M_A - \frac{\varepsilon_0}{t} I_{n+1}\right) \geq \mathbf{E}\tilde{c}_{11}^2 \cdot \det(V_{A\infty}) - \mathrm{const}_1 \cdot \varepsilon_0 > 0,$$

if

$$\varepsilon_0 < \frac{1}{\mathrm{const}_1} \mathbf{E}\tilde{c}_{11}^2 \det(V_{A\infty}). \tag{54}$$

Therefore if ε_0 satisfies (52) and (54) then

$$M_A - \frac{\varepsilon_0}{t} I_{n+1} > 0,$$

with probability tending to one, and

$$\lambda_{\min}(M_A) \geq \frac{\varepsilon_0}{t},$$

with probability tending to one. We have

$$\lambda_{\min}\left(\frac{1}{m}(T_A - W_A)\right) \geq \lambda_{\min}(M_A) - \left\|\frac{1}{m}H_3\right\| \geq \lambda_{\min}(M_A) - \sqrt{\frac{t}{m}} \cdot \mathrm{O_p}(1),$$

and with probability tending to one we have

$$\lambda_{\min}(T_A - W_A) \geq \frac{\varepsilon_0}{t} - \sqrt{\frac{t}{m}}\mathrm{O_p}(1).$$

But with probability tending to one

$$\frac{\varepsilon_0}{t} - \sqrt{\frac{t}{m}}\mathrm{O_p}(1) > 0 \Leftrightarrow \sqrt{\frac{m}{t^3}} \cdot \varepsilon_0 > \mathrm{O_p}(1),$$

and the last holds because by assumption $\frac{m}{t^3} \to \infty$. Thus $T_A > W_A$ with probability tending to one. Lemma 3 is proved. □

Acknowledgements. We thank two anonymous reviewers for their suggestions and corrections.

A. Kukush is supported by a postdoctoral research fellowship of the Belgian office for Scientific, Technical and Cultural Affairs, promoting Scientific and Technical Collaboration with Central and Eastern Europe.

S. Van Huffel is a full professor with the Katholieke Universiteit Leuven.

I. Markovsky is a research assistant with the Katholieke Universiteit Leuven.

This paper presents research results of the Belgian Programme on Interuniversity Poles of Attraction (IUAP V-22), initiated by the Belgian State, Prime Minister's Office – Federal Office for Scientific, Technical and Cultural Affairs of the Concerted Research Action (GOA) projects of the Flemish Government MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology), of the IDO/99/03 project (K.U. Leuven) "Predictive computer models for medical classification problems using patient data and expert knowledge", of the FWO projects G.0078.01, G.0200.00, and G0.0270.02.

The scientific responsibility is assumed by its authors.

References

[Bar00] Baran S (2000) A consistent estimator in general functional errors-in-variables models. Metrika 51(2):117–132

[CRS95] Carroll RJ, Ruppert D, Stefanski LA (1995) Measurement error in nonlinear models. Number 63 in Monographs on Statistics and Applied Probability. Chapman & Hall/CRC

[CS98] Cheng C, Schneeweiss H (1998) Polynomial regression with errors in the variables. J. R. Statistical Society B (60):189–199

[CST00] Cheng C, Schneeweiss H, Thamerus M (2000) A small sample estimator for a polynomial regression with errors in the variables. J. R. Statistical Society (62):699–709

[Ful87] Fuller WA (1987) Measurement error models. New York: Wiley

[Gal82] Gallo PP (1982) Consistency of regression estimates when some variables are subject to error. Comm. Statist. B-Theory Methods (11):973–893

[GV80] Golub GH, Van Loan CF (1980) An analysis of the total least squares problem. SIAM J. Numer. Anal. (17):883–893

[KM00] Kukush A, Maschke E-O (2000) The efficiency of adjusted least squares in the linear functional relationship. Discussion Paper 208, SFB 386, Ludwig-Maximilians-University of Munich

[KMV01] Kukush A, Markovsky I, Van Huffel S. Consistent fundamental matrix estimation in a quadratic measuremnt error model arising in motion analysis. Technical Report 01-64, Dept. EE, K.U. Leuven, 2001. Accepted for publication in Computational Statistics and Data Analysis, Special Issue of Matrix Computations and Statistics

[KZ96] Kukush A, Zwanzig S (1996) On inconsistency of the least squares estimator in non-linear functional error-in-variables models. Preprint N96-12, Institut für Mathematische Stochastik, Universität Hamburg

[MM98] Mühlich M, Mester R (1998) The role of total least squares in motion analysis. In Burkhardt H (ed) Proc. European Conference on Computer Vision (ECCV'98), Springer Lecture Notes on Computer Science, pages 305–321. Springer-Verlag, June

[Ros70] Rosenthal HP (1970) On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables. Isr. J. Math. (8):273–303

[VV89] Van Huffel S, Vandewalle J (1989) Analysis and properties of the generalized total least squares problem $AX \approx B$ when some or all columns in A are subject to error. SIAM J. Matrix Anal. 10(3):294–315

[VV91] Van Huffel S, Vandewalle J (1991) The total least squares problem: Computational aspects and analysis. SIAM, Philadelphia