Modern Stochastics: Theory and Applications 0 (0000) 0–0
DOI:

Asymptotic normality of total least squares estimator in a multivariate errors-in-variables model $AX = B$

A. Kukush*, **Ya. Tsaregorodtsev**

Taras Shevchenko National University of Kyiv, Kuiv, Ukraine

alexander_kukush@univ.kiev.ua (A. Kukush), 777Tsar777@mail.ru (Ya. Tsaregorodtsev)

Abstract A multivariate functional measurement error model $AX \approx B$ is considered. The errors in $[A, B]$ are uncorrelated, row-wise independent, and have equal (unknown) variances. The total least squares estimator of X is studied, which in the case of normal errors coincides with the maximum likelihood one. We give conditions for asymptotic normality of the estimator, when the number of rows in A is increasing. Under mild assumptions, the covariance structure of the limit Gaussian random matrix is nonsingular. For normal errors, the results can be used to construct the asymptotic confidence interval for a linear functional of X.

Keywords Asymptotic normality, multivariate errors-in-variables model, total least squares
2010 MSC 15A52, 65F20, 62E20, 62S05, 62F12, 62H12

1 Introduction

We deal with overdetermined system of linear equations $AX \approx B$, which is common in linear parameter estimation problem [9]. If the data matrix A and observation matrix B are contaminated with errors, and all the errors are uncorrelated and have equal variances, then the total least squares (TLS)

*Corresponding author.

Preprint submitted to VTeX / Modern Stochastics: Theory and Applications
<June 20, 2016>

technique is appropriate for solving this system [9]. In Kukush and Van Huffel [5] the statistical consistency of the TLS estimator $\hat{X}_{tls}$ was shown, as the number m of rows in A grows, provided the errors in $[A, B]$ are row-wise i.i.d. with zero mean and covariance matrix proportional to a unit matrix; the covariance matrix was assumed known up to a factor of proportionality; the true input matrix A_0 was supposed to be nonrandom. Actually in [5] a more general, element-wise weighted TLS estimator was studied, where the errors in $[A, B]$ were row-wise independent, but within each row the entries could be observed without errors, and addition the error covariance matrix could differ from row to row. In [6] an iterative numerical procedure was developed to compute the elementwise-weighted TLS estimator, and the rate of convergence of the procedure was established.

In a univariate case where B and X are column vectors, the asymptotic normality of $\hat{X}_{tls}$ was shown by Gallo [4], as m grows. In [7] that result was extended to mixing error sequences. Both [4] and [7] utilized an explicit form of the TLS solution.

In the present paper we extend the Gallo's asymptotic normality result to a multivariate case, where A, X, and B are matrices.

Now, a closed form solution is unavailable, and we work instead with the cost function. More precisely we deal with the estimating function, which is a matrix derivative of the cost function. In fact we show that under mild conditions, the normalized estimator converges in distribution to a Gaussian random matrix with nonsingular covariance structure. For normal errors, the latter structure can be estimated consistently based on the observed matrix $[A, B]$. The results can be used to construct the asymptotic confidence ellipsoid for a vector Xu, where u is a column vector of corresponding dimension.

The paper is organized as follows. In Section 2, we describe the model, refer to the consistency result for the estimator, and present the objective function and corresponding matrix estimating function. In Section 3, we state the asymptotic normality of $\hat{X}_{tls}$ and provide a nonsingular covariance structure for a limit random matrix. The latter structure depends continuously on some nuisance parameters of the model, and we derive consistent estimators for those parameters. Section 4 concludes. The proofs are given in Appendix. There we work with the estimating function and derive an expansion for the normalized estimator using Taylor's formula. The expansion holds true with probability tending to 1.

Throughout the paper all vectors are column ones, $\mathbf{E}$ stands for expectation and acts as an operator on the total product, $\mathbf{cov}(x)$ denotes the covariance matrix of a random vector x, and for a sequence of random matrices $\{X_m, m \geq 1\}$ of the same size, notation $X_m = O_p(1)$ means that the sequence $\{||X_m||\}$ is stochastically bounded, and $X_m = o_p(1)$ means that $||X_m|| \xrightarrow{\text{P}} 0$. I_p denotes a unit matrix of size p.

2 Model, objective and estimating

2.1 The TLS problem

Consider the model $AX \approx B$. Here $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times d}$ are observations, and $X \in \mathbb{R}^{n\times d}$ is a parameter of interest. Assume that

$$A = A_0 + \tilde{A}, \quad B = B_0 + \tilde{B}, \tag{2.1}$$

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

$$A_0 X_0 = B_0. \tag{2.2}$$

Here A_0 is nonrandom true input matrix, B_0 is true output matrix, and $\tilde{A}$, $\tilde{B}$ are error matrices. The matrix X_0 is the true value of the parameter.

One can rewrite the model (2.1) and (2.2) as a classical functional errors-in-variables (EIV) model, with vector regressor and vector response [3]. Denote $a_i^\top$, $a_{0i}^\top$, $\tilde{a}_i^\top$, $b_i^\top$, $b_{0i}^\top$ and $\tilde{b}_i^\top$ the rows of A, A_0, $\tilde{A}$, B, B_0 and $\tilde{B}$, respectively, $i = 1, \ldots, m$. Then the model above is equivalent to the following EIV model:

$$a_i = a_{0i} + \tilde{a}_i, \quad b_i = b_{0i} + \tilde{b}_i, \quad b_{oi} = X_0^\top a_{0i}, i = 1, \quad \ldots, m.$$

Based on observations a_i, b_i, $i = 1, \ldots, m$, one has to estimate X_0. The vectors a_{0i} are nonrandom and unknown, and the vectors $\tilde{a}_i$, $\tilde{b}_i$ are random errors.

State a global assumption of the paper.

(i). Vectors $\tilde{z}_i$ with $\tilde{z}_i^\top = [\tilde{a}_i^\top, \tilde{b}_i^\top]$, $i = 1, 2, \ldots$, are i.i.d., with zero mean and variance-covariance matrix

$$S_{\tilde{z}} := \mathbf{cov}(\tilde{z}_1) = \sigma^2 \, \mathrm{I}_{n+d}, \tag{2.3}$$

where a factor of proportionality σ^2 is positive and unknown.

The TLS problem consists in finding values of disturbances $\Delta\hat{A}$, $\Delta\hat{B}$ minimizing the sum of squared corrections

$$\min_{(X\in\mathbb{R}^{n\times d},\, \Delta A,\, \Delta B)} (||\Delta A||_F^2 + ||\Delta B||_F^2) \tag{2.4}$$

subject to the constraints:

$$(A - \Delta A)X = B - \Delta B. \tag{2.5}$$

Here in (2.4), for a matrix $C = (c_{ij})$, $||C||_F$ denotes the Frobenius norm, $||C||_F^2 = \sum_{i,j} c_{ij}^2$. Later on we will use the operator norm as well, $||C|| = \sup_{x\neq 0} \dfrac{||Cx||}{||x||}$.

2.2 TLS estimator and its consistency

It can happen that for some random realization, the problem (2.4), (2.5) has no solution. In that case put $\hat{X}_{tls} = \infty$. Now, we give a formal definition of the TLS estimator.

Definition 1. The TLS estimator $\hat{X}_{tls}$ of X_0 in the model (2.1), (2.2) is a measurable mapping of the underlying probability space into $\mathbb{R}^{n\times d} \cup \{\infty\}$, which solves the problem (2.4), (2.5) if there exists a solution, and $\hat{X}_{tls} = \infty$ otherwise.

We need the following conditions for the consistency of $\hat{X}_{tls}$.

(ii). $\mathbf{E}\,||\tilde{z}_1||^4 < \infty$, where $\tilde{z}_1$ enters condition (i).

(iii). $\dfrac{1}{m} A_0^\top A_0 \to V_A$, as $m \to \infty$, where V_A is nonsingular matrix.

The next consistency result is contained in Theorem 4a), [5].

Theorem 2. *Assume condition (i) to (iii). Then $\hat{X}_{tls}$ is finite with probability tending to one, and $\hat{X}_{tls}$ tends to X_0 in probability, as $m \to \infty$.*

2.3 The objective and estimating functions

Denote

$$q(a, b; X) = (a^\top X - b^\top)(\mathrm{I}_d + X^\top X)^{-1}(X^\top a - b), \tag{2.6}$$

$$Q(X) = \sum_{i=1}^{m} q(a_i, b_i; X), \quad X \in \mathbb{R}^{n\times d}. \tag{2.7}$$

The TLS estimator is known to minimize the objective function (2.7), see [8] or formula (24) in [5].

Lemma 3. *The TLS estimator $\hat{X}_{tls}$ is finite iff there exists an unconstrained minimum of the function (2.7), and then $\hat{X}_{tls}$ is a minimum point of that function.*

Introduce an estimating function related to the loss function (2.6):

$$s(a, b; X) := a(a^\top X - b^\top) - X(\mathrm{I}_d + X^\top X)^{-1}(X^\top a - b)(a^\top X - b^\top). \tag{2.8}$$

Corollary 4. *(a) Under conditions (i) to (iii), with probability tending to one $\hat{X}_{tls}$ is a solution to the equation*

$$\sum_{i=1}^{m} s(a_i, b_i; X) = 0, \quad X \in \mathbb{R}^{n\times d}.$$

(b) Under assumption (i), the function s(a,b;X) is unbiased estimating function, i.e., for each $i \geq 1$, $\mathbf{E}_{X_0}\, s(a_i, b_i; X_0) = 0$.

Expression (2.8) as a function of X is a mapping in $\mathbb{R}^{n\times d}$. Its derivative s'_X is a linear operator in this space.

Lemma 5. *Under condition (i), for each $H \in \mathbb{R}^{n\times d}$ and $i \geq 1$ it holds*

$$\mathbf{E}_{X_0}[s'_X(a_i, b_i; X_0) \cdot H] = a_{0i} a_{0i}^\top H. \tag{2.9}$$

Therefore, we can identify $\mathbf{E}_{X_0}\, s'_X(a_i, b_i; X_0)$ with the matrix $a_{0i} a_{0i}^\top$.

3 Main results

Introduce further assumptions to state the asymptotic normality of $\hat{X}_{tls}$. We need a bit higher moments compared with conditions (ii) and (iii) in order to use Lyapunov CLT. Remember that $\tilde{z}_i$ enters condition (i).

(iv). For some $\delta > 0$, $\mathbf{E}\,||\tilde{z}_1||^{4+2\delta} < \infty$.

(v). For δ from condition (iv),

$$\frac{1}{m^{1+\delta/2}} \sum_{i=1}^{m} ||a_{0i}||^{2+\delta} \to 0, \quad \text{as } m \to \infty.$$

(vi). $\displaystyle\frac{1}{m} \sum_{i=1}^{m} a_{0i} \to \mu_a$, as $m \to \infty$, where $\mu_a \in \mathbb{R}^{n\times 1}$.

(vii). Distribution of $\tilde{z}_1$ is symmetric around the origin.

Introduce a random element in the space of systems consisting of 5 matrices:

$$W_i = (a_{0i}\tilde{a}_i^\top, a_{0i}\tilde{b}_i^\top, \tilde{a}_i\tilde{a}_i^\top - \sigma^2\,\mathrm{I}_n, \tilde{a}_i\tilde{b}_i^\top, \tilde{b}_i\tilde{b}_i^\top - \sigma^2\,\mathrm{I}_d). \tag{3.1}$$

Hereafter $\xrightarrow{\mathrm{d}}$ stands for the convergence in distribution.

Lemma 6. *Assume conditions (i) and (iii) – (vi). Then*

$$\frac{1}{\sqrt{m}} \sum_{i=1}^{m} W_i \xrightarrow{\mathrm{d}} \Gamma = (\Gamma_1, \dots, \Gamma_5), \text{ as } m \to \infty, \tag{3.2}$$

where Γ is a Gaussian centered random element with matrix components.

Lemma 7. *In assumptions of Lemma 6, replace condition (vi) with condition (vii). Then the convergence (3.2) still holds true, with independent components $\Gamma_1, \dots, \Gamma_5$.*

Now, we state the asymptotic normality of $\hat{X}_{tls}$.

Theorem 8. *(a) Assume conditions (i) and (iii) – (vi). Then*

$$\sqrt{m}(\hat{X}_{tls} - X_0) \xrightarrow{\mathrm{d}} V_A^{-1}\Gamma(X_0), \quad \text{as } m \to \infty, \tag{3.3}$$

$$\Gamma(X) := \Gamma_1 X - \Gamma_2 + \Gamma_3 X - \Gamma_4 - X(\mathrm{I}_d + X^\top X)^{-1}(X^\top\Gamma_3 X - X^\top\Gamma_4 - \Gamma_4^\top X + \Gamma_5), \tag{3.4}$$

where V_A enters condition (iii), and Γ_i's enter relation (3.2).

(b) In the assumption of part (a), replace condition (vi) with condition (vii). Then the convergence (3.3) still holds true, and moreover the limit random matrix $X_\infty := V_A^{-1}\Gamma(X_0)$ has nonsingular covariance structure, i.e., for each nonzero vector $u \in \mathbb{R}^{d\times 1}$, $\mathbf{cov}(X_\infty u)$ is nonsingular matrix.

Remark 9. Conditions of Theorem 8(a) are similar to Gallo's conditions [4] for the asymptotic normality in the univariate case, see also, [9], pp. 240-243. Compared with Theorems 2.3 and 2.4 [7] stated for univariate case with mixing errors, we need not the requirement for entries of the true input A_0 to be totally bounded.

In [7], Section 2, one can find a discussion of importance of the asymptotic normality result for $\hat{X}_{tls}$. It is claimed there that the formula for the asymptotic covariance structure of $\hat{X}_{tls}$ is computationally useless, but in case where the limit distribution is nonsingular, one can use the block-bootstrap techniques when constructing confidence intervals and testing hypotheses.

But in the case of normal errors $\tilde{z}_i$'s, one can apply Theorem 8(b) to construct the asymptotic confidence ellipsoid, say, for $X_0 u$, $u \in \mathbb{R}^{d\times 1}$, $u \neq 0$. Indeed relations (3.1)–(3.4) show that the nonsingular matrix

$$S_u := \mathbf{cov}(\mathbf{V}_A^{-1}\Gamma(X_0)u)$$

is continuous function $S_u = S_u(X_0, \mathbf{V}_A, \sigma^2)$ of unknown parameters X_0, $\mathbf{V}_A$, and σ^2. (It is important here that now the components Γ_j of Γ are independent, and the covariance structure of each Γ_j depends on σ^2 and $\mathbf{V}_A$, not on some other limit characteristics of A_0, see Lemma 6.) Once we possess consistent estimators $\hat{\mathbf{V}}_A$ and $\hat{\sigma}^2$ of $\mathbf{V}_A$ and σ^2, the matrix $\hat{S}_u := S_u(\hat{X}_{tls}, \hat{\mathbf{V}}_A, \hat{\sigma}^2)$ is consistent estimator for the covariance matrix S_u.

Hereafter bar means averaging for rows $i = 1, \ldots, m$, e.g., $\overline{ab^\top} = \frac{1}{m}\sum_{i=1}^{m} a_i b_i^\top$.

Lemma 10. *Assume conditions of Theorem 2. Define*

$$\hat{\sigma}^2 = \frac{1}{d}\operatorname{tr}\left[(\overline{bb^\top} - 2\hat{X}_{tls}^\top \overline{ab^\top} + \hat{X}_{tls}^\top \overline{aa^\top}\hat{X}_{tls})(\mathrm{I}_d + \hat{X}_{tls}^\top \hat{X}_{tls})^{-1}\right], \tag{3.5}$$

$$\hat{V}_A = \overline{aa^\top} - \hat{\sigma}^2\, \mathrm{I}_n\,.$$

Then

$$\hat{\sigma}^2 \xrightarrow{\mathrm{P}} \sigma^2, \quad \hat{V}_A \xrightarrow{\mathrm{P}} V_A. \tag{3.6}$$

Remark 11. The estimator (3.5) is a multivariate analogue of the maximum likelihood estimator (1.53), [2], in the functional scalar EIV model.

Finally, for the case $\tilde{z}_1 \sim N(0, \sigma^2\, \mathrm{I}_{n+d})$, based on Lemma 10 and relations

$$\sqrt{m}(\hat{X}_{tls}u - X_0 u) \xrightarrow{\mathrm{d}} N(0, S_u), \quad S_u > 0, \quad \hat{S}_u \xrightarrow{\mathrm{P}} S_u,$$

in a standard way one can construct the asymptotic confidence ellipsoid for the vector $X_0 u$.

Remark 12. In a similar way a confidence ellipsoid can be constructed for any finite set of linear combinations of X_0 entries with fixed known coefficients.

4 Conclusion

We extended the result of Gallo [4] and proved the asymptotic normality of the TLS estimator in a multivariate model $AX \approx B$. The normalized estimator converges in distributin to a random matrix with quite complicated covariance structure. If the error distribution is symmetric around the origin, the latter covariance structure is nonsingular. For the case of normal errors, this makes it possible to construct the asymptotic confidence region for a vector $X_0 u$, $u \in \mathbb{R}^{d\times 1}$, where X_0 is the true value of X.

In future papers, we will extend the result for the element-wise weighted TLS estimator [5] in the model $AX \approx B$, where some columns of $[A, B]$ matrix may be observed without errors and in addition the error covariance matrix may differ from row to row.

Appendix

Proof of Corollary 4

(a) For any n, d, the space $\mathbb{R}^{n\times d}$ is endowed with natural inner product $< A, B >= \mathrm{tr}(AB^\top)$ and the Frobenius norm. The matrix derivative q'_X of the functional (2.6) is a linear functional on $\mathbb{R}^{n\times d}$, which can be identified with certain matrix from $\mathbb{R}^{n\times d}$ based on the inner product.

Using the rules of matrix calculus [1], we have for $H \in \mathbb{R}^{n\times d}$:

$$< q'_X, H >= a^\top H(\mathrm{I}_d + X^\top X)^{-1}(X^\top a - b)-$$
$$-(a^\top X - b^\top)(\mathrm{I}_d + X^\top X)^{-1}(H^\top X + X^\top H)(\mathrm{I}_d + X^\top X)^{-1}(X^\top a - b)+$$
$$+(a^\top X - b^\top)(\mathrm{I}_d + X^\top X)^{-1} H^\top a.$$

Collecting similar terms we obtain:

$$\frac{1}{2} < q'_X, H >= (a^\top X - b^\top)(\mathrm{I}_d + X^\top X)^{-1} H^\top a-$$
$$-(a^\top X - b^\top)(\mathrm{I}_d + X^\top X)^{-1} H^\top X(\mathrm{I}_d + X^\top X)^{-1}(X^\top a - b),$$

and

$$\frac{1}{2} < q'_X, H >= \mathrm{tr}\left[a(a^\top X - b^\top)(\mathrm{I}_d + X^\top X)^{-1} H^\top\right] -$$
$$-\mathrm{tr}\left[X(\mathrm{I}_d + X^\top X)^{-1}(X^\top a - b)(a^\top X - b^\top)(\mathrm{I}_d + X^\top X)^{-1} H^\top\right].$$

Using the inner product in $\mathbb{R}^{n\times d}$ we get: $\frac{1}{2} q'_X = s(x)(\mathrm{I}_d + X^\top X)^{-1}$, where $s(x)$ is the left-hand side of (2.8). In view of Theorem 2 and Lemma 3 this implies the statement of Corollary 4(a).

(b) Now, we set

$$a = a_0 + \tilde{a}, \quad b = b_0 + \tilde{b}, \quad b_0 = X^\top a_0, \tag{4.1}$$

where a_0 is a nonrandom vector and like in (2.3),

$$\mathbf{cov}\left(\begin{bmatrix} \tilde{a} \\ \tilde{b} \end{bmatrix}\right) = \sigma^2\,\mathrm{I}_{n+d}, \quad \mathbf{E}\begin{bmatrix} \tilde{a} \\ \tilde{b} \end{bmatrix} = 0. \tag{4.2}$$

Then

$$\mathbf{E}_X\, a(a^\top X - b^\top) = \mathbf{E}\, a(\tilde{a}^\top X - \tilde{b}^\top) = \sigma^2 X, \tag{4.3}$$

$$\mathbf{E}_X (X^\top a - b)(a^\top X - b^\top) = \mathbf{E}(X^\top \tilde{a} - \tilde{b})(\tilde{a}^\top X - \tilde{b}^\top) = \sigma^2(\mathrm{I}_d + X^\top X). \tag{4.4}$$

Therefore, see (2.8),

$$\mathbf{E}_X\, s(a, b; X) = \sigma^2 X - \sigma^2 X(\mathrm{I}_d + X^\top X)^{-1}(\mathrm{I}_d + X^\top X) = 0.$$

This implies the statement of Corollary 4(b).

Proof of Lemma 5

The derivative s'_X of the function (2.8) is a linear operator in $\mathbb{R}^{n\times d}$. For $H \in \mathbb{R}^{n\times d}$, we have:

$$\begin{aligned} s'_X H = {} & aa^\top H - H(\mathrm{I}_d + X^\top X)^{-1}(X^\top a - b)(a^\top X - b^\top) + \\ & + X(\mathrm{I}_d + X^\top X)^{-1}(H^\top X + X^\top H)(\mathrm{I}_d + X^\top X)^{-1}(X^\top a - b) \times \\ & \times (a^\top X - b^\top) - X(\mathrm{I}_d + X^\top X)^{-1}(H^\top a(a^\top X - b^\top) + (X^\top a - b)a^\top H). \end{aligned} \tag{4.5}$$

As above we set (4.1), (4.2) and use relations (4.3), (4.4), and relation $\mathbf{E}\, aa^\top = a_0 a_0^\top + \sigma^2\,\mathrm{I}_n$. We obtain:

$$\begin{aligned} \mathbf{E}_X\, s'_X H = {} & (a_0 a_0^\top + \sigma^2\,\mathrm{I}_n)H - \sigma^2 H + \sigma^2 X(\mathrm{I}_d + X^\top X)^{-1}(H^\top X + X^\top H) - \\ & - \sigma^2 X(\mathrm{I}_d + X^\top X)^{-1}(H^\top H + X^\top H) = a_0 a_0^\top H. \end{aligned}$$

This implies (2.9).

Proof of Lemma 6

The random elements W_i, $i \ge 1$, in (3.1) are independent and centered. We want to apply Lyapunov CLT for the left-hand side of (3.2).

(a) All the second moments of $m^{-\frac{1}{2}} \sum_{i=1}^{m} W_i$ converge to finite limits. E. g., for 1st component we have:

$$\frac{1}{m}\sum_{i=1}^{m} \mathbf{E}(< a_{0i}\tilde{a}_i^\top, H_1 >)^2 = \frac{1}{m}\sum_{i=1}^{m} \mathbf{E}(\mathrm{tr}\ a_{0i}\tilde{a}_1^\top H_1^\top)^2,$$

and this has finite limit due to assumption (iii). Here $H_1 \in \mathbb{R}^{n\times n}$ and we use the inner product introduced in the proof of Corollary 4.

For 5th component,

$$\frac{1}{m}\sum_{i=1}^{m}\mathbf{E}(< \tilde{b}_i\tilde{b}_i^{\mathsf{T}} - \sigma^2\,\mathrm{I}_d, H_2 >)^2 = \mathbf{E}[\mathrm{tr}((\tilde{b}_1\tilde{b}_1^{\mathsf{T}} - \sigma^2\,\mathrm{I}_d)H_2)]^2 < \infty,$$

because 4th moments of $\tilde{b}_i$ are finite. Here $H_2 \in \mathbb{R}^{d\times d}$.

For mixed moments of 1st and 5th components, we have:

$$\begin{aligned} &\frac{1}{m}\sum_{i=1}^{m}\mathbf{E} < a_{0i}\tilde{a}_i^{\mathsf{T}}, H_1 > \cdot < \tilde{b}_i\tilde{b}_i^{\mathsf{T}} - \sigma^2\,\mathrm{I}_d, H_2 >= \\ &= \mathbf{E} < \left(\frac{1}{m}\sum_{i=1}^{m} a_{0i}\right)\tilde{a}_1^{\mathsf{T}}, H_1 > \cdot < \tilde{b}_1\tilde{b}_1^{\mathsf{T}} - \sigma^2\,\mathrm{I}_d, H_2 >, \end{aligned} \tag{4.6}$$

and this due to condition (vi) converges towards

$$\mathbf{E} < \mu_a\tilde{a}_1^{\mathsf{T}}, H_1 > \cdot < \tilde{b}_1\tilde{b}_1^{\mathsf{T}} - \sigma^2\,\mathrm{I}_d, H_2 > .$$

Other second moments can be considered in a similar way.

(b) The Lyapunov's condition holds for each component of (3.1). Let δ be a quantity from assumptions (iv), (v). Then

$$\frac{1}{m^{1+\delta/2}}\sum_{i=1}^{m}\mathbf{E}\,||a_{0i}\tilde{a}_i^{\mathsf{T}}||^{2+\delta} \le \frac{\mathbf{E}\,||\tilde{a}_1||^{2+\delta}}{m^{1+\delta/2}}\sum_{i=1}^{m}||a_{0i}||^{2+\delta} \to 0,$$

as $m \to \infty$, by condition (v). For 5th component,

$$\begin{aligned} &\frac{1}{m^{1+\delta/2}}\sum_{i=1}^{m}\mathbf{E}\,||\tilde{b}_i\tilde{b}_i^{\mathsf{T}} - \sigma^2\,\mathrm{I}_d\,||^{2+\delta} = \frac{1}{m^{\delta/2}}\,\mathbf{E}\,||\tilde{b}_1\tilde{b}_1^{\mathsf{T}} - \sigma^2\,\mathrm{I}_d\,||^{2+\delta} \le \\ &\le \frac{\mathrm{const}}{m^{\delta/2}}\,\mathbf{E}\,||\tilde{b}_1||^{4+2\delta} \to 0, \text{ as } m \to \infty. \end{aligned}$$

The latter expectation is finite by condition (iv).

The Lyapunov's condition for other components is considered similarly.

(c) Parts (a) and (b) of the present proof imply (3.2) by Lyapunov CLT.

Proof of Lemma 7

Under conditions (vii) and (i), all the five components of W_i, which is given in (3.1), are uncorrelated (e.g. cross-correlation like (4.6) equals zero, and condition (vi) is not needed). As in proof of Lemma 6, the convergence (3.2) still holds true. The component $\Gamma_1, \ldots, \Gamma_5$ of Γ are independent, because the components of W_i are uncorrelated.

Proof of Theorem 8(a)

Our reasoning are typical for Generalized Estimating Equations theory, with specific feature that a matrix rather than vector parameter is estimated.

By Corollary 4(a), with probability tending to 1 it holds:

$$\sum_{i=1}^{m} s(a_i, b_i; \hat{X}_{tls}) = 0. \tag{4.7}$$

Now, we use Taylor's formula around X_0, with the remainder in the Lagrange form, see [1], Theorem 5.6.2. Denote

$$\hat{\Delta} = \sqrt{m}(\hat{X}_{tls} - X_0), \quad y_m = \sum_{i=1}^{m} s(a_i, b_i; X_0), \quad U_m = \sum_{i=1}^{m} s'_X(a_i, b_i; X_0).$$

Then (4.7) implies the relation:

$$\begin{gathered}\left(\frac{1}{m}U_m\right)\hat{\Delta} = -\frac{1}{\sqrt{m}}y_m + rest_1, \\ ||rest_1|| \le ||\hat{\Delta}|| \cdot ||\hat{X}_{tls} - X_0|| \cdot O_p(1).\end{gathered} \tag{4.8}$$

Here $O_p(1)$ is a factor of the form

$$\frac{1}{m}\sum_{i=1}^{m} \sup_{(||X|| \le ||X_0||+1)} ||s''_x(a_i, b_i; X)||. \tag{4.9}$$

Relation (4.8) holds with probability tending to 1, because due to Theorem 2, $\hat{X}_{tls} \xrightarrow{\text{P}} X_0$; expression (4.9) is indeed $O_p(1)$, because the derivative s''_x is quadratic in a_i, b_i, cf. (4.5), and the averaged 2nd moments of $[a_i^\top, b_i^\top]$ are assumed bounded.

Now, $||rest_1|| \le ||\hat{\Delta}|| \cdot o_p(1)$. Next, by Lemma 5 and condition (iii),

$$\frac{1}{m}U_m = \frac{1}{m}\mathbf{E}\,U_m + o_p(1) = V_A + o_p(1).$$

Therefore, (4.8) implies that

$$V_A\hat{\Delta} = -\frac{1}{\sqrt{m}}y_m + rest_2, \tag{4.10}$$

$$||rest_2|| \le ||\hat{\Delta}|| \cdot o_p(1). \tag{4.11}$$

Now, we find the limit in distribution of $y_m/\sqrt{m}$. The summands in y_m have zero expectation due to Corollary 4(b). Moreover, see (2.8),

$$s(a_i, b_i; X_0) = (a_{0i} + \tilde{a}_i)(\tilde{a}_i^\top X_0 - \tilde{b}_i^\top) - X_0(\mathrm{I}_d + X_0^\top X_0)^{-1}(X_0^\top \tilde{a}_i - \tilde{b}_i)(\tilde{a}_i^\top X_0 - \tilde{b}_i^\top),$$

$$\begin{aligned}s(a_i, b_i; X_0) = W_{i1}X_0 - W_{i2} + W_{i3}X_0 - W_{i4} - X_0(\mathrm{I}_d + X_0^\top X_0)^{-1} \times \\ \times (X_0^\top W_{i3} X_0 - X_0^\top W_{i4} - W_{i4}^\top X_0 + W_{i5}).\end{aligned}$$

Here W_{ij} are components of (3.1). By Lemma 6 we have, see (3.4):

$$\frac{1}{\sqrt{m}} y_m \xrightarrow{\mathrm{d}} \Gamma(X_0), \quad \text{as } m \to \infty. \tag{4.12}$$

Finally, relations (4.10), (4.11), (4.12) and nonsingularity of V_A imply that $\hat{\Delta} = O_p(1)$, and by Slutsky's lemma we get

$$V_A \hat{\Delta} \xrightarrow{\mathrm{d}} \Gamma(X_0), \quad \text{as } m \to \infty. \tag{4.13}$$

By condition (iii) the matrix V_A is nonsingular. Thus, the desired relation (3.3) follows from (4.13).

Proof of Theorem 8(b)

The convergence (3.3) is grounded as above, but using Lemma 7 instead of Lemma 6. It is enough to show that $\mathbf{cov}(\Gamma(X_0)u)$ is nonsingular, for $u \in \mathbb{R}^{d\times 1}$, $u \neq 0$.

Now, the components $\Gamma_1, \ldots, \Gamma_5$ are independent. Then, see (3.4),

$$\mathbf{cov}(\Gamma(X_0)u) \geq \mathbf{cov}(\Gamma_2 u) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} \mathbf{E}(u^\top \tilde{b}_i a_{0i}^\top a_{0i} \tilde{b}_i^\top u) =$$
$$= \mathrm{trV_A} \cdot \mathbf{E}\, ||\tilde{\mathrm{b}}_1^\top \mathrm{u}||^2 = \sigma^2 \mathrm{trV_A} \cdot ||\mathrm{u}||^2 > 0.$$

Proof of Lemma 10

We have by condition (i):

$$\mathbf{E}\, a_i a_i^\top = a_{0i} a_{0i}^\top + \sigma^2\, \mathrm{I}_n, \quad \mathbf{E}\, a_i b_i^\top = a_{i0} a_{i0}^\top X_0,$$
$$\mathbf{E}\, b_i b_i^\top = X_0^\top a_{0i} a_{0i}^\top X_0 + \sigma^2\, \mathrm{I}_d,$$

$$\mathbf{E}\, b_i b_i^\top - 2X_0^\top\, \mathbf{E}\, a_i b_i^\top + X_0^\top (\mathbf{E}\, a_i a_i^\top) X_0 = \sigma^2 (\mathrm{I}_d + X_0^\top X_0). \tag{4.14}$$

Equality (4.14) implies the first relation in (3.6), because $\hat{X}_{tls} \xrightarrow{\mathrm{P}} X_0$ and $\overline{aa^\top} - \mathbf{E}\, \overline{aa^\top} \xrightarrow{\mathrm{P}} 0$, $\overline{ab^\top} - \mathbf{E}\, \overline{ab^\top} \xrightarrow{\mathrm{P}} 0$, $\overline{bb^\top} - \mathbf{E}\, \overline{bb^\top} \xrightarrow{\mathrm{P}} 0$,

Finally,
$\hat{V}_A = \mathbf{E}\, \overline{aa^\top} + o_p(1) - \hat{\sigma}^2\, \mathrm{I}_n = \overline{a_0 a_0^\top} + (\sigma^2 - \hat{\sigma}^2)\, \mathrm{I}_n + o_p(1)$,
$\hat{V}_A \xrightarrow{\mathrm{P}} \lim_{m\to\infty} \overline{a_0 a_0^\top} = V_A$.

References

[1] Cartan, H.: Differential Calculus. Hermann, Paris; Houghton Mifflin Co., Boston, Mass. (1971). Translated from the French. MR0344032

[2] Cheng, C.-L., Van Ness, J.W.: Statistical Regression with Measurement Error. Kendall's Library of Statistics, vol. 6. Arnold, London; co-published by Oxford University Press, New York (1999). MR1719513

[3] Fuller, W.A.: Measurement Error Models. John Wiley & Sons, Inc., New York (1987). doi:10.1002/9780470316665. MR898653

[4] Gallo, P.P.: Properties of estimators errors-in-variables models. PhD thesis, The University of North Carolina at Chapel Hill, NC (1982). MR2632121

[5] Kukush, A., Van Huffel, S.: Consistency of elementwise-weighted total least squares estimator in a multivariate errors-in-variables model $AX = B$. Metrika **59**(1), 75–97 (2004). doi:10.1007/s001840300272. MR2043433

[6] Markovsky, I., Rastello, M.L., Premoli, A., Kukush, A., Van Huffel, S.: The element-wise weighted total least-squares problem. Comput. Statist. Data Anal. **50**(1), 181–209 (2006). doi:10.1016/j.csda.2004.07.014. MR2196229

[7] Pešta, M.: Asymptotics for weakly dependent errors-in-variables. Kybernetika (Prague) **49**(5), 692–704 (2013). MR3182634

[8] Sprent, P.: A generalized least-squares approach to linear functional relationships. J. Roy. Statist. Soc. Ser. B **28**, 278–297 (1966). MR0230432

[9] Van Huffel, S., Vandewalle, J.: The Total Least Squares Problem. Frontiers in Applied Mathematics, vol. 9. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1991). doi:10.1137/1.9781611971002. MR1118607