Theory of Stochastic Processes
Vol. 21 (37), no. 2, 2016, pp. 96–105

YA. V. TSAREGORODTSEV

ASYMPTOTIC NORMALITY OF ELEMENT-WISE WEIGHTED TOTAL LEAST SQUARES ESTIMATOR IN A MULTIVARIATE ERRORS-IN-VARIABLES MODEL

A multivariable measurement error model $AX \approx B$ is considered. Here A and B are input and output matrices of measurements and X is a rectangular matrix of fixed size to be estimated. The errors in $[A, B]$ are row-wise independent, but within each row the errors may be correlated. Some of the columns are observed without errors and the error covariance matrices may differ from row to row. The total covariance structure of the errors is known up to a scalar factor. The fully weighted total least squares estimator of X is studied. We give conditions for asymptotic normality of the estimator, as the number of rows in A is increasing. We provide that the covariance structure of the limiting Gaussian random matrix is nonsingular.

1. Introduction

We deal with an overdetermined set of linear equations $AX \approx B$, which is common in linear parameter estimation problems [12]. If both the data matrix A and observation matrix B are contaminated with errors, and all the errors are uncorrelated and have equal variances, the total least squares (TLS) technique is appropriate for solving this set [4], [12]. Under mild conditions, the TLS estimator of X is consistent and asymptotically normal, as the number of rows in A is increasing [3], [7].

In this paper we consider heteroscedastic errors. The errors in $[A, B]$ are row-wise independent, but within each row the errors may be correlated. Some of the columns are observed without errors, and the error covariance matrices may differ from row to row. The total error covariance structure is assumed known up to a scalar factor. For this model, the element-wise weighted total least squares (EW-TLS) estimator is introduced and its consistency is proven in [6]. Concerning the computation of the estimator see [10], [5]. The EW-TLS estimator $\hat{X}$ is applied, e.g., in geodesy [9].

Our goal is to extend the asymptotic normality result of [7] to the EW-TLS estimator. We work under the conditions of Theorem 2, [6] about the consistency of $\hat{X}$. We use the objective function of the estimator, see formula (22) in [6], and the rules of matrix calculus [2].

The paper is organized as follows. In section 2, we describe the model, introduce main assumptions, refer to the consistency result for $\hat{X}$ and present the objective function and the matrix estimating function. In Section 3, we state the asymptotic normality result and provide a nonsingular covariance structure for a limiting random matrix. In Section 4, we derive consistent estimators for nuisance parameters of the model in order to estimate consistently the asymptotic covariance structure of $\hat{X}$, and Section 5 concludes. The proofs are given in Appendix.

Throughout the paper all vectors are column ones, 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,

---

2000 *Mathematics Subject Classification.* 62E20; 62F12; 62J05; 62H12; 65F20.

Key words and phrases. Asymptotic normality, element-wise weighted total least squares estimator, heteroscedastic errors, multivariate errors-in-variables model.

notation $X_m = O_p(1)$ means that the sequence $\{||X_m||\}$ is stochastically bounded, and $X_m = o_p(1)$ means that $||X_m|| \xrightarrow{\mathrm{P}} 0$. I_p denotes the identity matrix of size p.

2. Observation model and consistency of the estimator

2.1. **The EW-TLS promblem.** We deal with the model $AX \approx B$. Here $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times d}$ are matrices of observations, and the matrix $X \in \mathbb{R}^{n\times d}$ is to be estimated. 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 a true output matrix, and $\tilde{A}$, $\tilde{B}$ are error matrices. X_0 is the true value of the matrix parameter.

It is useful to rewrite the model (2.1) and (2.2) as a classical errors-in-variables (EIV) model [1]. Denote $a_i^\top$, $a_{0i}^\top$, $\tilde{a}_i^\top$, $b_i^\top$, $b_{0i}^\top$, $\tilde{b}_i^\top$, $i = 1,\dots,m$, the rows of A, A_0, $\tilde{A}$, B, B_0 and $\tilde{B}$, respectively. Then the model above is equivalent to the EIV model

$$a_i = a_{0i} + \tilde{a}_i, \quad b_i = b_{0i} + \tilde{b}_i, b_{0i} = X_0^\top a_{0i}, \quad i = 1,\dots,m. \tag{2.3}$$

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

Rewrite the model (2.1) and (2.2) in an implicit way. Introduce matrices

$$C = [A, B], \quad C_0 = [A_0, B_0], \quad \tilde{C} = [\tilde{A}, \tilde{B}], \quad Z_0 = \begin{bmatrix} X_0 \\ -\mathrm{I}_d \end{bmatrix}. \tag{2.4}$$

Then (2.1), (2.2) is equivalent to the next relations:

$$C = C_0 + \tilde{C}, \quad C_0 Z_0 = 0.$$

Let $\tilde{C} = (\tilde{c}_{ij}, i = 1,\dots,m, \ \ j = 1,\dots,n+d)$. Following [6] we state global assumptions of the paper, conditions (i) to (iv).

(i). Vectors $\tilde{c}_i := (\tilde{c}_{i1},\dots,\tilde{c}_{i,n+d})^\top$, $i = 1,2,\dots$, are independent with zero mean and finite second moments.

Let $\sigma_{ij}^2 = \mathsf{E}\,\tilde{c}_{ij}^2$, $i = 1,2,\dots$, $j = 1,\dots,n+d$. We allow that some of σ_{ij}^2 are vanishing.

(ii). For a fixed $J \subset \{1,2,\dots,n+d\}$, every $j \notin J$ and every $i = 1,2,\dots$ satisfy $\sigma_{ij}^2 = 0$. Moreover

$$\mathbf{cov}(\tilde{c}_{ij}, j \in J) = \sigma^2 \Sigma_i, \quad i = 1,2,\dots,$$

with unknown positive factor of proportionality σ^2 and known matrices Σ_i.

(iii). There exists $\varkappa > 0$ such that for every $i = 1,2,\dots$, it holds $\lambda_{min}(\Sigma_i) \ge \varkappa^2$.

For the matrix $Z_0 = (z_{0,jk})$ given in (2.4) and the set J from condition (ii), denote

$$Z_{0J} = (z_{0,jk}, j \in J, k = 1,\dots,d).$$

(iv).

$$\operatorname{rank}(Z_{0J}) = d.$$

The EW-TLS problem consists in finding the value $\hat{X}$ of the unknown matrix X and values of disturbances $\Delta\hat{A}$, $\Delta\hat{B}$ minimizing the weighted sum of squared corrections:

$$\min_{(X\in\mathbb{R}^{n\times d}, \Delta A, \Delta B)} \sum_{i=1}^{m} ||\Sigma_i^{-1/2} \Delta c_i^J||^2 \tag{2.5}$$

subject to constrains

$$(A - \Delta A)X = B - \Delta B, \quad \Delta c_i^J = 0, \quad i = 1, \ldots, m, \quad j \notin J.$$

Here $C = [A, B] = (c_{ij})$, $\Delta C = [\Delta A, \Delta B] = (\Delta c_{ij})$ and the column vectors

$$\Delta c_i^J := (\Delta c_{ij}, j \in J) \in \mathbb{R}^{|J|}.$$

2.2. **EW-TLS estimator and its consistency.** For a random realization, it can happen that the problem (2.5) has no solution. Assume conditions (i) – (iv).

Definition 1. The EW-TLS estimator $\hat{X} = \hat{X}_{EW-TLS}$ of X_0 in the model (2.1), (2.2) is a Borel measurable mapping of the data matrix C into $\mathbb{R}^{n\times d} \cup \{\infty\}$, which solves the problem (2.5) under the additional constraint

$$rank(Z_J) = d \tag{2.6}$$

$\left(\text{here } Z = \begin{bmatrix} X \\ -\mathrm{I}_d \end{bmatrix} = (z_{jk}),\ Z_J := (z_{jk}, j \in J, k = 1, \ldots, d)\right)$, if there exists a solution, and $\hat{X} = \infty$ otherwise.

The EW-TLS estimator always exists due to [11]. We need more conditions to provide the consistency of $\hat{X}$.

(v). There exists $r \geq 2$ with $r > d\left(|J| - \dfrac{d+1}{2}\right)$ such that

$$\sup_{(i\geq 1, j\in J)} \mathsf{E}\,|\tilde{c}_{ij}|^{2r} < \infty.$$

(vi). $\dfrac{\lambda_{\min}(A_0^\top A_0)}{\sqrt{m}} \to \infty$, as $m \to \infty$.

(vii). $\dfrac{\lambda_{\min}^2(A_0^\top A_0)}{\lambda_{\max}(A_0^\top A_0)} \to \infty$, as $m \to \infty$.

The next result on weak consistency is stated in Theorem 2, [6].

Theorem 2. *Assume conditions (i) to (vii). Then the EW-TLS estimator $\hat{X}$ is finite with probability tending to one, and $\hat{X}$ tends to X_0 in probability, as $m \to \infty$.*

Notice that under a bit stronger assumptions on eigenvalues of $A_0^\top A_0$, the estimator $\hat{X}$ is strongly consistent, see Theorem 3, [6].

2.3. **The estimating function.** Remember that error vectors $\tilde{c}_i$ enter condition (i) and the matrix $Z = Z(X)$ is introduced in Definition 1. Let

$$S_i := \frac{1}{\sigma^2}\,\mathbf{cov}(\tilde{c}_i), i = 1, 2, \ldots$$

Denote also

$$q(c, S; X) = c^\top Z(Z^\top S Z)^{-1} Z^\top c, \tag{2.7}$$

where $c = \begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^{(n+d)\times 1}$, $S \in \mathbb{R}^{(n+d)\times(n+d)}$, and

$$Q(X) = \sum_{i=1}^{m} q(c_i, S_i; X), \quad X \in \mathbb{R}^{n\times d}, \quad \mathrm{rank}(Z_J) = d. \tag{2.8}$$

Notice that due to (iv) $|J| \geq d$, and under constraint (2.6) Z_J is of full rank. Then, under conditions (i) – (iii) the matrix $Z^\top S_i Z$ is nonsingular, $i = 1, 2, \ldots$

The EW-TLS estimator is known to minimize the objective function (2.7), see Theorem 1, [6].

Lemma 3. *Assume conditions (i) to (iv). The EW-TLS estimator $\hat{X}$ is finite if, and only if, there exists an unconditional minimum of the function* (2.8)*, and then $\hat{X}$ is a minimum point of this function.*

Introduce an estimating function related to the loss function (2.7):

$$s(a,b,S;X) = \tilde{s}\cdot(Z^\top SZ)^{-1}, \tag{2.9}$$

$$\tilde{s} = \tilde{s}(a,b,S;X) := ac^\top Z - [S_a, S_{ab}]Z(Z^\top SZ)^{-1}Z^\top cc^\top Z. \tag{2.10}$$

Here

$$c = \begin{bmatrix} a \\ b \end{bmatrix}, \quad a \in \mathbb{R}^{n\times 1}; \quad S = \begin{bmatrix} S_a & S_{ab} \\ S_{ba} & S_b \end{bmatrix}, \quad S_a \in \mathbb{R}^{n\times n}. \tag{2.11}$$

Corollary 4. *Assume conditions (i) – (vii). Then the next two statements hold true.*

(a) *With probability tending to one $\hat{X}$ is a solution to the equation*

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

(b) *The function* (2.9) *is an unbiased estimating function, i.e., for each $i \geq 1$,*

$$\mathsf{E}_{X_0}\, s(a_i, b_i, S_i; X_0) = 0.$$

For fixed a, b, S, the function (2.9) maps X into $\mathbb{R}^{n\times d}$. The derivative s'_X is a linear operator in this space.

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

$$\mathsf{E}_{X_0}[s'_X(a_i, b_i, S_i; X_0)\cdot H] = a_{0i}a_{0i}^\top H(Z_0 S_i Z_0)^{-1}.$$

3. Asymptotic normality of the estimator

Introduce further assumptions.

(viii). For some $\delta > 0$, $\sup\limits_{(i\geq 1, j\in J)} \mathsf{E}\,|\tilde{c}_{ij}|^{4+2\delta} < \infty$.

(ix). For δ from the condition (viii),

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

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

Notice that condition (x) implies assumptions (vi), (vii).

(xi). For matrices from condition the (ii), $\Sigma_i \to \Sigma_\infty$, as $m \to \infty$, where Σ_∞ is certain matrix.

Notice that conditions (xi), (iii) imply that Σ_∞ is nonsingular.

(xii). If $p, q, r \in J$ (they are not necessarily distinct) and $i \geq 1$, then

$$\mathsf{E}\,\tilde{c}_{ip}\tilde{c}_{iq}\tilde{c}_{ir} = 0.$$

(xiii). If $p, q, r, u \in J$ (they are not necessarily distinct), then $\dfrac{1}{m}\sum\limits_{i=1}^{m} \mathsf{E}\,\tilde{c}_{ip}\tilde{c}_{iq}\tilde{c}_{ir}\tilde{c}_{in}$ converges to a finite limit $\mu_4(p,q,r,u)$, as m tends to infinity.

Introduce a random element in the space of couples of matrices:

$$W_i = (a_{0i}\hat{c}_i^\top, \tilde{c}_i\tilde{c}_i^\top - \sigma^2 S_i). \tag{3.1}$$

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

Lemma 6. *Assume conditions (i), (ii) and (viii) – (xiii). Then*

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m} W_i \xrightarrow{\mathrm{d}} \Gamma = (\Gamma_1, \Gamma_2), \quad \text{as } m \to \infty, \tag{3.2}$$

where Γ *is a Gaussian centered random element with independent matrix components* Γ_1 *and* Γ_2.

Now, we state the asymptotic normality of the EW-TLS estimator.

Theorem 7. *Assume conditions (i) – (v) and (viii) – (xiii). Then*

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

$$\Gamma(X) := \Gamma_1 Z + P_a \Gamma_2 Z - [S_a^\infty, S_{ab}^\infty] Z (Z^\top S_\infty Z)^{-1}(Z^\top \Gamma_2 Z), \tag{3.4}$$

where V_A *enters condition (x),* P_a *is the projector with* $P_a \begin{bmatrix} a \\ b \end{bmatrix} = a$, Γ_1 *and* Γ_2 *enter relation* (3.2)*, and*

$$S_\infty = \begin{bmatrix} S_a^\infty & S_{ab}^\infty \\ S_{ba}^\infty & S_b^\infty \end{bmatrix} = \lim_{i\to\infty} S_i, \quad Z = \begin{bmatrix} X \\ -\mathrm{I}_d \end{bmatrix}. \tag{3.5}$$

Moreover the limiting random matrix $X_\infty := V_A^{-1}\Gamma(X_0)$ *has a nonsingular covariance structure, i.e., for each nonzero vector* $u \in \mathbb{R}^{d\times 1}$, $\mathbf{cov}(X_\infty u)$ *is a nonsingular matrix.*

4. Construction of confidence region for a linear functional of X_0

4.1. **Estimation of nuisance parameters.** Theorem 7 can be applied, e.g., to construct a confidence region for a linear functional of X_0. For this purpose one has to estimate consistently a covariance structure of the limiting random matrix $V_A^{-1}\Gamma(X_0)$. Such a structure, besides of X_0, depends on nuisance parameters. Some of them can be estimated consistently.

Hereafter bar means average for rows $i = 1, \dots, m$, e.g.,

$$\overline{ab^\top} = m^{-1} \cdot \sum_{i=1}^{m} a_i b_i^\top, \quad \bar{S} = m^{-1} \sum_{i=1}^{m} S_i.$$

Lemma 8. *Assume conditions of Theorem 7. Define*

$$\hat{Z} = \begin{pmatrix} \hat{X} \\ -\mathrm{I}_d \end{pmatrix}, \quad \hat{\sigma}^2 = \frac{1}{d}\mathrm{tr}\left[(\hat{Z}^\top \overline{cc^\top} \hat{Z})(\hat{Z}^\top \bar{S} \hat{Z})^{-1}\right], \tag{4.1}$$

$$\hat{V}_A = \overline{aa^\top} - \hat{\sigma}^2 \bar{S}. \tag{4.2}$$

Then, as $m \to \infty$,

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

4.2. **Estimation of the asymptotic covariance structure of** X_0**.** Let $u \in \mathbb{R}^{d\times 1}$, $u \neq 0$. Theorem 7 implies the convergence

$$\sqrt{m}(\hat{X}u - X_0 u) \xrightarrow{\mathrm{d}} N(0, S_u), \quad \text{as } m \to \infty, \tag{4.3}$$

with nonsingular matrix $S_u = \mathbf{cov}(V_A^{-1}\Gamma(X_0)u)$.

We start with the case of normal errors $\tilde{c}_i$, $i = 1, 2, \dots$ Then condition (xii) holds true, and Theorem 7 is applicable. The asymptotic covariance matrix S_u is a continuous function $S_u = S_u(X_0, V_A, \sigma^2, S_\infty)$ of unknown parameters (here the limiting covariance matrix S_∞ could be unknown, though for a given m, matrices $S_1, \dots, S_m$ are assumed known). Due to Theorem 2 and Lemma 8 the matrix

$$\hat{S}_u := S_u(\hat{X}, \hat{V}_A, \hat{\sigma}^2, \bar{S}) \tag{4.4}$$

is a consistent estimator of S_u.

Now, we do not assume the normality of the errors. Then the exact formula for S_u does not allow to estimate it consistently, because the formula involves higher moments of errors which are difficult to estimate consistently. Instead, we use Corollary 4 to construct the so-called sandwich estimator [1] for S_u. Denote

$$\hat{s}_i = \tilde{s}(a_i, b_i, S_i; \hat{X}), \quad i = 1, \ldots, m, \tag{4.5}$$

with $\tilde{s}$ introduced in (2.10)

Lemma 9. *Assume conditions of Theorem 7. For $u \in \mathbb{R}^{d\times 1}$, $u \neq 0$, define*

$$\hat{S}_u = \hat{V}_A^{-1} \cdot \frac{1}{m} \sum_{i=1}^{m} \hat{s}_i u u^\top \hat{s}_i^\top, \tag{4.6}$$

with $\hat{V}_A$ given in (4.2), (4.1). *Then $\hat{S}_u \xrightarrow{\mathrm{P}} S_u$, as $m \to \infty$.*

Remark. In the case of normal errors, the estimator (4.4) is asymptotically more efficient than the estimator (4.6), cf. the discussion in [1], p. 369.

Given a consistent estimator $\hat{S}_u$ of S_u, we have from (4.3) that

$$\sqrt{m}(\hat{S}_u)^{-1/2}(\hat{X}u - X_0 u) \xrightarrow{\mathrm{d}} N(0, \mathrm{I}_n), \quad \text{as } m \to \infty. \tag{4.7}$$

Based on (4.7), one can construct in a standard way an asymptotic confidence ellipsoid for $X_0 u$. Similarly a confidence ellipsoid can be constructed for any finite set of linear combinations of X_0 entries.

5. Conclusion

We proved the asymptotic normality of the EW-TLS estimator in a multivariate errors-in-variables model $AX \approx B$ with heteroscedastic errors. We assumed the convergence (xi) of the second error moments, vanishing third moments (xiii), and the convergence of averaged fourth moments (xiii). The condition (xii) ensured that the asymptotic covariance structure of $\hat{X}$ is nonsingular. This condition holds true in two cases: (a) all the error vectors $\tilde{c}_i$ are symmetrically distributed, or (b) for each i, random variables $\tilde{c}_{ip}$, $p \in J$, are independent and have vanishing coefficient of asymmetry.

The obtained asymptotic normality result made it possible to construct a confidence ellipsoid for a linear functional of X_0. Another plausible application is goodness-of-fit test in the model $AX \approx B$ with heteroscedastic errors (see [7] for such a test in the model with homoscedastic errors).

The author is grateful to Prof. A. Kukush for the problem statement and fruitful discussions.

Appendix

Proof of Corollary 4. (a) The space $\mathbb{R}^{n\times d}$ is endowed with natural inner product $< A, B >= \mathrm{tr}(AB^\top)$. The matrix derivative q'_X of the functional (2.7) is a linear functional on $\mathbb{R}^{n\times d}$, and based on the inner product, this functional can be identified with certain matrix from $\mathbb{R}^{n\times d}$.

Remember that $Z = Z(X)$ is introduced in Definition 1. Using the rules of matrix calculus [2], we have for $H \in \mathbb{R}^{n\times d}$:

$$< q'_X, H >= c^\top \begin{bmatrix} H \\ 0 \end{bmatrix} (Z^\top S Z)^{-1} Z^\top c + c^\top (Z^\top S Z)^{-1} \cdot [H^\top, 0]c-$$
$$-c^\top Z(Z^\top S Z)^{-1} \left([H^\top, 0] S Z + Z^\top S \begin{bmatrix} H \\ 0 \end{bmatrix} \right) (Z^\top S Z)^{-1} Z^\top c.$$

Remember relations (2.11). Collecting similar terms, we obtain:

$$\frac{1}{2} < q'_x, H >= a^\top H(Z^\top SZ)^{-1}Z^\top c-$$
$$-c^\top Z(Z^\top SZ)^{-1}Z^\top \begin{bmatrix} S_a \\ S_{ba} \end{bmatrix} H(Z^\top SZ)^{-1}Z^\top c,$$

and

$$\frac{1}{2} < q'_x, H >= \mathrm{tr}[ac^\top Z(Z^\top SZ)^{-1}H^\top]-$$
$$\mathrm{tr}\left[[S_a, S_{ab}]Z(Z^\top SZ)^{-1}Z^\top cc^\top Z(Z^\top SZ)^{-1}H^\top\right].$$

Using the inner product in $\mathbb{R}^{n\times d}$ we obtain

$$\frac{1}{2}q'_x = \tilde{s}(X)(Z^\top SZ)^{-1},$$

with $\tilde{s}(X) = \tilde{s}(a,b,S;X)$ given in (2.10). Now, Theorem 2 and Lemma 3 imply the statement of Corollary 4(a).

(b) We set

$$a = a_0 + \tilde{a}, \quad b = b_0 + \tilde{b}, \quad b_0 = X^\top a_0, \quad c = c_0 + \tilde{c} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \begin{bmatrix} \tilde{a} \\ \tilde{b} \end{bmatrix}, \tag{A.1}$$

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

$$\mathbf{cov}(\tilde{c}) = \sigma^2 S = \sigma^2 \begin{bmatrix} S_a & S_{ab} \\ S_{ba} & S_b \end{bmatrix}, \quad \mathsf{E}\,\tilde{c} = 0.$$

Then

$$\mathsf{E}_X\, ac^\top Z = a_0 c_0^\top \begin{bmatrix} X \\ -\mathrm{I}_d \end{bmatrix} + \mathsf{E}\,\tilde{a}\tilde{c}^\top Z = \sigma^2[S_a, S_{ab}]Z,$$
$$\mathsf{E}_X\, cc^\top Z = c_0 c_0^\top \begin{bmatrix} X \\ -\mathrm{I}_d \end{bmatrix} + \mathsf{E}\,\tilde{c}\tilde{c}^\top Z = \sigma^2 SZ.$$

Therefore, see (2.9),

$$\mathsf{E}_X\, \tilde{s}(a,b,S;X) = \sigma^2[S_a, S_{ab}]Z - \sigma^2[S_a, S_{ab}]Z(Z^\top SZ)^{-1}(Z^\top SZ) = 0$$

The statement (b) of Corollary 4 is proven.

Proof of Lemma 5. The derivative s'_X of the function (2.9) with respect to X is a linear operator in $\mathbb{R}^{n\times d}$. Denote $f = f(Z) = Z(Z^\top SZ)^{-1}$. For $H \in \mathbb{R}^{n\times d}$, it holds:

$$\tilde{s}'_X H = aa^\top H - [S_a, S_{ab}](f'_X H)(Z^\top cc^\top Z) - [S_a, S_{ab}]f \cdot \left([H^\top, 0]cc^\top Z + Z^\top cc^\top \begin{bmatrix} H \\ 0 \end{bmatrix}\right).$$

We set (A.1), use relations

$$\mathsf{E}\,aa^\top = a_0 a_0^\top + \sigma^2 S_a, \quad \mathsf{E}_X(cc^\top Z) = \sigma^2 SZ$$

and get:

$$\begin{aligned} \mathsf{E}_X(\tilde{s}'_X H) = (a_0 a_0^\top + \sigma^2 S_a)H - \sigma^2[S_a, S_{ab}](f'_X H)(Z^\top SZ)- \\ -\sigma^2[S_a, S_{ab}]f \cdot \left([H^\top, 0]SZ + Z^\top S \begin{bmatrix} H \\ 0 \end{bmatrix}\right). \end{aligned} \tag{A.2}$$

Next,

$$f'_X H = \begin{bmatrix} H \\ 0 \end{bmatrix} (Z^\top SZ)^{-1} - Z(Z^\top SZ)^{-1}\left([H^\top, 0]SZ + Z^\top S \begin{bmatrix} H \\ 0 \end{bmatrix}\right)(Z^\top SZ)^{-1}. \tag{A.3}$$

Combining (A.2) and (A.3) we see that on the right-hand side of (A.2) summands containing $H^\top$ are cancelled out. We get finally

$$\mathsf{E}_X(\tilde{s}'_X H) = a_0 a_0^\top H,$$

which implies the statement, because by Corollary 4(b) it holds $\mathsf{E}_X\, \tilde{s}(a, b, S; X) = 0$.

Proof of Lemma 6. The proof is similar to the proof of Lemmas 6 and 7 from [7] and based on Lyapunov's Central Limit Theorem. We just notice that due to condition (xii) the matrix components of W_i, namely $a_{0i}\tilde{c}_i^\top$ and $\tilde{c}_i\tilde{c}_i^\top - \sigma^2 S_i$, are uncorrelated, and this implies the independence of matrix components Γ_1 and Γ_2 in (3.2).

Proof of Theorem 7. We follow the line of [7], see there the proof of Theorem 8(a). By Corollary 4(a), it holds with probability tending to 1:

$$\sum_{i=1}^{m} s(a_i, b_i, S_i; \hat{X}) = 0. \tag{A.4}$$

Denote

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

Using Taylor's formula around X_0 (see [2], Theorem 5.6.2), we obtain from (A.4) that

$$\begin{gathered}\left(\frac{1}{m}U_m\right)\hat{\Delta} = -\frac{1}{\sqrt{m}}y_m + \mathrm{rest}_1,\\ ||\mathrm{rest}_1|| \le ||\hat{\Delta}|| \cdot ||\hat{X} - X_0|| \cdot O_p(1).\end{gathered} \tag{A.5}$$

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

$$\frac{1}{m}\sum_{i=1}^{m} \sup_{(||X - X_0|| \le \varepsilon_0)} ||s''_X(a_i, b_i, S_i; X)||, \tag{A.6}$$

with positive ε_0 chosen such that $\operatorname{rank}(Z_J) = d$, for all X with $||X - X_0|| \le \varepsilon_0$; the choice is possible due to condition (iv), and expression (A.6) is indeed $O_p(1)$ (i.e., stochastically bounded), because s''_X is quadratic in c_i and the averaged second moments of c_i are assumed bounded. Thus, the relation (A.5) holds true due to the consistency of $\hat{X}$ stated in Theorem 2.

We have $||\mathrm{rest}_1|| \le ||\hat{\Delta}|| \cdot o_p(1)$. Now, by Lemma 5 and condition (x) and (xi) it holds

$$\frac{1}{m}U_m H = V_A H (Z_0^\top S_\infty Z_0)^{-1} + o_p(1), \quad H \in \mathbb{R}^{n\times d},$$

and we derive from (A.5) the relation

$$V_A\hat{\Delta}(Z_0^\top S_\infty Z_0)^{-1} = -\frac{1}{\sqrt{m}}y_m + \mathrm{rest}_2, \quad ||\mathrm{rest}_2|| \le ||\hat{\Delta}|| \cdot o_p(1). \tag{A.7}$$

The summands in y_m have zero expectation by Corollary 4(b). Remember that $c_{0i}Z_0 = 0$ and the projector P_a is introduced in Theorem 7. Then, see (2.9),

$$\begin{aligned}\tilde{s}(a_i, b_i, S_i; X_0) &= (a_{0i} + \tilde{a}_i)\tilde{c}_i^\top Z_0 - [S_{ai}, S_{bi}]Z_0(Z_0^\top S_i Z_0)^{-1}(Z_0^\top \tilde{c}_i\tilde{c}_i^\top Z_0),\\ \tilde{s}(a_i, b_i, S_i; X_0) &= W_{i1}Z_0 + P_a W_{i2} Z_0 - [S_{ai}, S_{bi}]Z_0(Z_0^\top S_i Z_0)^{-1}(Z_0^\top W_{i2} Z_0).\end{aligned}$$

Here W_{ij} are components of (3.1). By Lemma 6 it holds, see (3.4) and condition (xi):

$$\frac{1}{\sqrt{m}}y_m \xrightarrow{\mathrm{d}} \Gamma(X_0)(Z_0^\top S_\infty Z_0)^{-1}, \quad \text{as } m \to \infty. \tag{A.8}$$

Now, relations (A.7), (A.8) and nonsingularity of V_A imply $\hat{\Delta} = O_p(1)$ and by Slutsky's lemma

$$V_A\hat{\Delta}(Z_0^\top S_\infty Z_0)^{-1} \xrightarrow{\mathrm{d}} \Gamma(X_0)(Z_0^\top S_\infty Z_0)^{-1}, \quad \text{as } m \to \infty.$$

This implies the desired convergence (3.3) – (3.5).

Let $u \in \mathbb{R}^{d\times 1}$, $u \neq 0$. By Lemma 6 the components Γ_1 and Γ_2 are independent. We have

$$\mathbf{cov}(\Gamma(X_0)u) \geq \mathbf{cov}(\Gamma_1 Z_0 u) = \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^m \mathsf{E}(a_{0i}\tilde{c}_i^\top Z_0 u u^\top Z_0^\top \tilde{c}_i a_{0i}^\top) = \\ = \sigma^2 V_A [u^\top (Z_0^\top S_\infty Z_0) u],$$

and the latter matrix is positive definite, because V_A and $Z_0^\top S_\infty Z_0$ are positive definite under the conditions of Theorem 7. Therefore, $\mathbf{cov}(X_\infty u)$ is a positive definite matrix as well.

Proof of Lemma 8. We have

$$\mathsf{E}\, c_i c_i^\top = c_{0i}c_{0i}^\top + \sigma^2 S_i,\;\; Z_0^\top(\mathsf{E}\, c_i c_i^\top) Z_0 = \sigma^2 Z_0^\top S_i Z_0,$$

$$\sigma^2 = \frac{1}{d}\operatorname{tr}\left[(Z_0^\top \overline{c_i c_i^\top} Z_0)(Z_0^\top \bar{S} Z_0)^{-1}\right] + o_p(1). \tag{A.9}$$

Relation (A.9) and the convergence $\hat{Z} \xrightarrow{\mathrm{P}} Z_0$ imply the desired convergence $\hat{\sigma}^2 \xrightarrow{\mathrm{P}} \sigma^2$, as $m \to \infty$.

Next,

$$\hat{V}_A = \mathsf{E}\,\overline{aa^\top} + o_p(1) - \hat{\sigma}^2\bar{S} = \overline{a_0 a_0^\top} + (\sigma^2 - \hat{\sigma}^2)\bar{S} + o_p(1),$$

$$\hat{V}_A \xrightarrow{\mathrm{P}} \lim_{m\to\infty} \overline{a_0 a_0^\top} = V_A.$$

Proof of Lemma 9. Denote $\tilde{s}_i = \tilde{s}(a_i, b_i, S_i; X_0)$, $i = 1, 2, \ldots$ Then expansion (A.7) implies that

$$S_u = V_A^{-1} \cdot \frac{1}{m}\sum_{i=1}^m \tilde{s}_i u_i u_i^\top \tilde{s}_i^\top + o_p(1),$$

and by Lemma 8

$$S_u = \hat{V}_A^{-1} \cdot \frac{1}{m}\sum_{i=1}^m \tilde{s}_i u_i u_i^\top \tilde{s}_i^\top + o_p(1),$$

$$\hat{S}_u - S_u = \hat{V}_A^{-1} \cdot \frac{1}{m}\sum_{i=1}^m (\hat{s}_i u_i u_i^\top \hat{s}_i^\top - \tilde{s}_i u_i u_i^\top \tilde{s}_i^\top) + o_p(1).$$

Then $\hat{S}_u - S_u \xrightarrow{\mathrm{P}} 0$, as $m \to \infty$, because $\hat{Z} \xrightarrow{\mathrm{P}} Z_0$ and $\overline{cc^\top} = O_p(1)$ (see formulas (2.9), (2.10) and (4.5)). Lemma 9 is proven.

References

1. R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu, *Measurement Error in Nonlinear Models: A Modern Perspective.* 2nd ed. Boca Raton, Chapman and Hall/CRC, 2006.
2. H. Cartan, *Differential Calculus.* Hermann/Houghton Mifflin Co., Paris/Boston, MA. Translated from French, 1971.
3. L. J. Gleser, *Estimation in a multivariate "errors in variables" regression model: large sample results*, Ann. Stat. **9** (1981), no. 1, 24–44.
4. G. H. Golub and C. F. Van Loan, *An analysis of the total least squares problem*, SIAM J. Numer. Anal. **17** (1980), no. 6, 883–893.
5. S. Jazaerti, A. R. Amiri-Simkooei, and M. A. Sharifi, *Iterative algorithm for weighted total least squares adjustment*, Survey Review **46** (2014), no. 334, 19–27.
6. A. Kukush and S. Van Huffel, *Consistency of elementwise-weighted total least squares estimator in a multivariate errors-in-variables model $AX = B$*, Metrika **59** (2004), no. 1, 75–97.

7. A. Kukush and Ya. Tsaregorodtsev, *Asymptotic normality of total least squares estimator in a multivariable errors-in-variables model $AX = B$*, Modern Stochastics: Theory and Applications **3** (2016), no. 1, 47–57.
8. A. Kukush and Ya. Tsaregorodtsev, *Goodness-of-fit test in a multivariate errors-in-variables model $AX = B$*, Modern Stochastics: Theory and Applications **3** (2016), no. 4, 287–302.
9. V. Mahboud, *On weighted total least squares for geodetic transformation*, J. of Geodesy **86** (2012), no. 5, 359–367.
10. I. Markovsky, M. L. Rastello, A. Premoli, A. Kukush, and S. Van Huffel, *The element-wise weighted total least-squares problem*, Comput. Statist. Data Anal. **50** (2006), no. 1, 181–209.
11. G. Pfanzagl, *On the measurability and consistency of minimum contrast estimates*, Metrika **14** (1969), no. 1, 249–273.
12. S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, Frontiers in Applied Mathematics, vol. 9, SIAM, Philadelphia, PA, 1991.

Department of Mathematical Analysis, Faculty of Mechanics and Mathematics, Taras Shevchenko National University of Kyiv, Building 4-e, Akademika Glushkova Avenue, Kyiv, Ukraine, 03127

E-mail address: 777Tsar777@mail.ru