Theory of Stochastic Processes
Vol. 14 (30), no. 1, 2008, pp. 117–125

UDC 519.21

MARIA POLEKHA

ESTIMATION IN AN IMPLICIT MULTIVARIATE MEASUREMENT ERROR MODEL WITH CLUSTERING IN THE REGRESSOR

An implicit linear multivariate model $DZ \approx 0$ is considered, where the data matrix D is observed with errors, and Z is a parameter matrix. The error matrix is partitioned into two uncorrelated blocks, and the total covariance structure in each block is supposed to be known up to a corresponding scalar factor. Moreover, the row data are clustered into two groups. Based on the method of corrected objective function, the strongly consistent estimators of scalar factors and the kernel of the matrix D are constructed, as the numbers of rows in the clusters tend to infinity.

1. Introduction

We deal with an implicit multivariate model $\bar{D}Z = 0,\ D = \bar{D} + \tilde{D}$, where Z is the parameter matrix, D is the data matrix, $\bar{D}$ is the unobserved matrix, and $\tilde{D}$ is the error matrix. The model is important for the system identification. In fact, our model can be obtained from an explicit model $AX \approx B$, where the matrices A and B are observed with errors, and X is a matrix parameter. The approximate equality $AX \approx B$ means that, for certain unobserved matrices $\bar{A}$ and $\bar{B}$, the equality $\bar{A}X = \bar{B}$ holds, and $A = \bar{A} + \tilde{A}$, $B = \bar{B} + \tilde{B}$, where $\tilde{A}$ and $\tilde{B}$ are error matrices. For this model, there are some results concerning the estimation of the parameter. In [1, 2], the covariance structure of $[A\ B]$ was known up to a constant, and the consistency of the total least squares estimator was proven. In [4], the consistent estimator was constructed for the situation when the covariance structure of $[A\ B]$ is unknown. In [3,8,9], the model $AX \approx B$ was studied when the covariance structure of A is known up to one scaling factor, while the covariance structure of B is known up to another scaling factor. For identifiability reasons, it was assumed in [3,8,9] that the data consist of two independent clusters, and, in [4], the number of clusters depended of the size of X. The idea of the use of clusters belongs to A. Wald [6]. He used it for a linear scalar model and proposed a slope estimator based on the first empirical moments. In [4], similar moments were used, while, in [3,8,9], the second empirical moments were exploited.

We modify the results of [3]. But our model is more general, although the matrix D can be considered as a compound matrix $[A\ B]$, here the input and output signals are not separated.

The paper is organized as follows. Section 2 describes the model and introduces main assumptions. In Section 3, we use the method of corrected objective function to derive an estimator of the kernel of Z in case of known scalar factors λ_1^0 and λ_2^0. A preliminary attempt to derive an objective function for λ_1^0 and λ_2^0, when they are unknown, is made in Section 4. In Section 5, we introduce a model with two clusters and the final objective function for the scalar factors and state the consistency result. In Section 6, the consistent

---

2000 *AMS Mathematics Subject Classification.* Primary 65F20, 93E12, 62H30, 62J05, 62H12, 62F12, 65P99.

Key words and phrases. Linear measurement error model, corrected objective function, clusters, consistent estimator.

estimator for the kernel of Z is proposed, and Section 7 concludes. The proofs of the results are given in Appendix.

In the paper, we use a standard notation: $\|A\|$ is the Frobenius norm of a matrix A; the symbol **tr** denotes the trace of a matrix; I_p denotes a unit matrix of size p; and the symbol **E** denotes the expectation of a random matrix. All the vectors in the paper are column vectors. For a symmetric matrix C, $\mu_1(C) \le \ldots \le \mu_p(C)$ are p smallest eigenvalues of C.

2. The model and assumptions

Consider the model of observations

$$DZ \approx 0, \tag{1}$$

where $Z \in \mathbb{R}^{(n+p)\times p}$ is the unknown matrix parameter, and the data matrix $D \in \mathbb{R}^{m\times(n+p)}$ is observed, $D = \bar{D} + \tilde{D}$. Here, $\bar{D}$ is the unknown nonrandom matrix, and $\tilde{D}$ is the matrix of random errors. For the true matrix $\bar{D}$, we have

$$\bar{D}Z = 0.$$

We want to estimate the matrix parameter Z with fixed n and p and increasing m.

We assume that dim Ker$\bar{D} = p$, and the kernel is not changed with increase in m. Let $Z = [z_1, \ \ldots \ , \ z_p]$ and rank$Z = p$. Then

$$\bar{D}[z_1, \ \ldots \ , \ z_p] = 0, \ \bar{D}z_i = 0, \ i = 1, ..., p,$$

and

$$\mathrm{span}(z_1, \ \ldots \ , \ z_p) = \mathrm{Ker}\bar{D} =: V_p.$$

Thus, we have to construct the estimator $\hat{V}_p$ of the kernel Ker$\bar{D}$.

Let $\tilde{D}^\top = [\tilde{d}_1, \ \cdots \ , \ \tilde{d}_m]$. By concerning the error vectors $\tilde{d}_i$, we make some assumptions:

(i) $\mathbf{E}\tilde{d}_i = 0$, for all i.

(ii) There exists $\delta > 0$ such that $\sup_{i\ge 1} \mathbf{E}\|\tilde{d}_i\|^{4+\delta} < \infty$.

(iii) The sequence of random vectors $\{\tilde{d}_i, \ i \ge 1\}$ is independent.

(iv) There exists n_1, $1 \le n_1 \le n + p - 1$, such that $\tilde{D} = [\tilde{D}_1 \ \tilde{D}_2]$, $\tilde{D}_1 \in \mathbb{R}^{m\times n_1}$, and $\mathbf{E}\tilde{D}_1^\top \tilde{D}_2 = 0$.

This means that the error matrix $\tilde{D}$ can be partitioned into two uncorrelated blocks.

(v) $\mathbf{E}\tilde{D}_k^\top \tilde{D}_k = \lambda_k^0 W_k$, where W_k are the known positive definite matrices, and λ_k^0 are the unknown positive scalars, $k = 1, 2$.

(vi) $\frac{1}{m}\|\bar{D}\|^2 \le const$, $m \ge 1$.

3. The estimator under known scalar factors

Suppose that λ_1^0 and λ_2^0 are known. We will use the corrected objective function method [5]. The least squares objective function is

$$Q_{\mathrm{ls}}(\bar{D}; Z) := \|\bar{D}Z\|^2, \qquad Z \in \mathbb{R}^{(n+p)\times p},$$

or

$$Q_{\mathrm{ls}}(\bar{D}; Z) = \mathbf{tr}(Z^\top \Psi_{\mathrm{ls}}(D)Z),$$

where

$$\Psi_{\mathrm{ls}}(\bar{D}) := \bar{D}^\top \bar{D}.$$

By the corrected objective function method, we have to construct a matrix $Q_{\mathrm{c}}(D; Z)$ such that

$$\mathbf{E}[Q_{\mathrm{c}}(D; Z)] = Q_{\mathrm{ls}}(\bar{D}; Z), \ \text{for all } Z.$$

It is possible under the known λ_j^0 and W_j, $j = 1, 2$, defined in conditions (iv) - (v). Let

$$\Psi_c(D) = D^\top D - \begin{bmatrix} \lambda_1^0 W_1 & 0 \\ 0 & \lambda_2^0 W_2 \end{bmatrix}.$$

Then

$$Q_c(D; Z) = \mathbf{tr}(Z^\top \Psi_c(D) Z).$$

We minimize this objective function under the condition $Z^\top Z = I_p$. Consider the ordered eigenvalues of $\Psi_c(D)$, $\mu_1 \le \mu_2 \le \ldots \le \mu_{n+p}$, and the corresponding orthonormal eigenbasis $\{\varphi_i, i = 1, 2, \ldots, n+p\}$. The function $Q_c(D; Z)$ is minimized at $Z_0 = [\varphi_1, ..., \varphi_p]$. Therefore, we obtain the estimator $\hat{V}_p = \mathrm{span}(\varphi_1, ..., \varphi_p)$. It is a linear span of p eigenvectors which correspond to p smallest eigenvalues. The next lemma is similar to the corresponding lemma in [3].

Lemma 1. *Assume (i) to (vi). Then*

$$\left\| \frac{1}{m} \Psi_c(D) - \frac{1}{m} \Psi_{ls}(\bar{D}) \right\| \to 0, \quad \textit{as} \quad m \to \infty, \quad \textit{a.s.}$$

4. The estimator under unknown scalar factors

Let λ_j^0, $j = 1, 2$, be the unknown scalar factors. Then, for any $\lambda_1, \lambda_2 \ge 0$, we define

$$\Psi_c(D; \lambda_1, \lambda_2) = \Psi_c(\lambda_1, \lambda_2) := D^\top D - \begin{bmatrix} \lambda_1 W_1 & 0 \\ 0 & \lambda_2 W_2 \end{bmatrix},$$

and

$$\Psi_{ls}(\bar{D}; \lambda_1, \lambda_2) = \Psi_{ls}(\lambda_1, \lambda_2) := \bar{D}^\top \bar{D} - \begin{bmatrix} (\lambda_1 - \lambda_1^0) W_1 & 0 \\ 0 & (\lambda_2 - \lambda_2^0) W_2 \end{bmatrix}.$$

By Lemma 1,

$$\left\| \frac{1}{m} \Psi_c(\lambda_1, \lambda_2) - \frac{1}{m} \Psi_{ls}(\lambda_1, \lambda_2) \right\| \to 0, \quad \text{as} \quad m \to \infty, \quad \text{a.s.}$$

We will study the properties of the approximating matrix $\frac{1}{m} \Psi_{ls}(\lambda_1, \lambda_2)$. Assume the following condition.

(vii) $\liminf_{m \to \infty} \mu_{p+1}(\bar{D}^\top \bar{D}/m) > 0$.

For large m, we have the approximate equality

$$\frac{1}{m} \Psi_c(\lambda_1, \lambda_2) \approx \frac{1}{m} \Psi_{ls}(\lambda_1, \lambda_2),$$

and, for the approximate matrix $\Psi_{ls}(\lambda_1^0, \lambda_2^0)$, we have

$$\mu_i\big(\Psi_{ls}(\lambda_1^0, \lambda_2^0)\big) = 0, \quad \text{for all } i = 1, \ldots, p,$$

and $\mu_{p+1}\big(\Psi_{ls}(\lambda_1^0, \lambda_2^0)\big) > 0$. Note that

$$V_p(\lambda_1^0, \lambda_2^0) = \mathrm{span}(z_1, \ldots, z_p) = \mathrm{Ker}(\Psi_{ls}(\lambda_1^0, \lambda_2^0)).$$

Therefore, in order to estimate λ_1^0, λ_2^0, it seems natural to use the objective function

$$Q(\lambda_1, \lambda_2) := \sum_{i=1}^{p} \mu_i^2 \Big(\frac{1}{m} \Psi_c(\lambda_1, \lambda_2) \Big).$$

Unfortunately, its minimization does not yield a consistent estimator of $\lambda_1^0,\ \lambda_2^0$, since, for the approximating matrix-valued function $\Psi_{\rm ls}(\lambda_1,\lambda_2)/m$, there could exist other values λ_1^* and λ_2^* which are separated from λ_1^0 and λ_2^0, and such that

$$\mu_i\big(\Psi_{\rm ls}(\lambda_1^*,\lambda_2^*)\big)=0,\quad \text{for all } 1\le i\le p.$$

Therefore, we cannot estimate the scalars λ_1^0 and λ_2^0 by this way.

5. Model with two clusters

Consider two copies of model (1):

$$D(k)Z\approx 0,\ \ k=1,2, \tag{2}$$

where $Z\in\mathbb{R}^{(n+p)\times p}$ is the unknown parameter matrix to be estimated with rank$Z=p$, and $D(k)\in\mathbb{R}^{m_k\times(n+p)}$ are observed,

$$D(k)=\bar D(k)+\tilde D(k),$$

$$\bar D(k)Z=0,\ \ k=1,2.$$

Here, $\bar D(k)$ are the unknown nonrandom matrices, and $\tilde D(k)$ are the error matrices, $\tilde D^\top(k)=[\tilde d_1(k)\ \cdots\ \tilde d_{m_k}(k)]$, $k=1,2$. We assume that dim Ker$\bar D(k)=p$, $k=1,2$. We want to estimate $V_p:=\mathrm{Ker}\bar D(1)=\mathrm{Ker}\bar D(2)$ for increasing m_1 and m_2 under fixed n and p. For each $k=1,2$ we assume the following.

(a) $\mathbf{E}\tilde d_i(k)=0,\ i\ge 1.$

(b) There exists $\delta>0$ such that $\sup_{i\ge1}\mathbf{E}\|\tilde d_i(k)\|^{4+\delta}<\infty$.

(c) $\{\,\tilde d_i(k),\ i\ge 1\,\}$ are two mutually independent sequences of random vectors.

(d) There exists n_1, $1\le n_1\le n+p-1$, such that $\tilde D(k)=[\tilde D_1(k)\ \tilde D_2(k)]$, $\tilde D_1(k)\in\mathbb{R}^{m_k\times n_1}$ and $\mathbf{E}\tilde D_1^\top(k)\tilde D_2(k)=0$.

(e) $\mathbf{E}\tilde D_j^\top(k)\tilde D_j(k)=\lambda_j^0W_j(k)$, $j=1,2$, where $W_j(k)$ are the known positive definite matrices, and λ_i^0 are the unknown positive scalars.

(f) $\frac{1}{m_k}\|\bar D(k)\|^2\le const$, $m_k\ge 1$.

(g) $\liminf_{m_k\to\infty}\mu_{p+1}(\bar D^\top(k)\bar D(k)/m_k)>0$.

(h) $\liminf_{m_1,m_2\to\infty}\sigma_{p+1}(\bar D^\top(1)\bar D(1)/m_1-\bar D^\top(2)\bar D(2)/m_2)>0$, where $\sigma_{p+1}(C)$ is the $(p+1)$-th singular value of the symmetric matrix C.

[We mention that, for the matrix C in brackets in (h), $\sigma_1(C)=\ldots=\sigma_p(C)=0$].

Let $Z=\begin{bmatrix}Z_1\\ Z_2\end{bmatrix}$, $Z_1\in\mathbb{R}^{n_1\times p}$, $Z_2\in\mathbb{R}^{n_2\times p}$, $n_1+n_2=n+p$. Then $\tilde D(k)Z=\tilde D_1(k)Z_1+\tilde D_2(k)Z_2$.

(i) $\liminf_{m_1\to\infty}\mathbf{tr}(Z_j^T\frac{W_j(1)}{m_1}Z_j)>0$, $j=1,2$, for certain fixed $Z=[z_1\ldots z_p]$ with span$(z_1,\ldots,z_p)=V_p$.

(j) $\frac{W_j(1)}{m_1}-\frac{W_j(2)}{m_2}\to 0$, as $m_1,m_2\to\infty$, $j=1,2$.

Let, for $\lambda:=(\lambda_1,\lambda_2)\in[0,\infty)\times[0,\infty)$,

$$\Psi_{\rm c}^{(k)}(\lambda):=D^\top(k)D(k)-\begin{bmatrix}\lambda_1W_1(k) & 0\\ 0 & \lambda_2W_2(k)\end{bmatrix},$$

and let $\mu_{1k}(\lambda)\le\mu_{2k}(\lambda)\le\ldots\le\mu_{pk}(\lambda)$ be p smallest eigenvalues of the matrix $\Psi_{\rm c}^{(k)}(\lambda)$ with the corresponding orthonormal eigenvectors $f_{1k}(\lambda),f_{2k}(\lambda),\ldots,f_{pk}(\lambda)$, $k=1,2$. Note that if the given eigenvalues are multiple, then the eigenvectors are not uniquely defined. In this case, we define them in such a way that they are Borel measurable vector functions of $D(k)$ and λ.

Let $V_{pk}(\lambda) = \mathrm{span}(f_{1k}(\lambda), f_{2k}(\lambda), \ldots, f_{pk}(\lambda))$. Then an objective function for estimating λ is

$$Q_c(\lambda) := \sum_{k=1}^{2}\sum_{i=1}^{p} \mu_{ik}^2(\lambda) + c\|\sin\Theta(\lambda)\|^2,$$

where c is a fixed positive constant, $\Theta(\lambda)$ is a diagonal matrix of canonical angles between the subspaces $V_{p_1}(\lambda)$ and $V_{p_2}(\lambda)$, and $\sin\Theta(\lambda)$ is defined as the diagonal matrix with the sines of these angles as diagonal elements. Recall the next definition of canonical angles.

Definition 1. [7]. *Let $\Re(X)$ and $\Re(Y)$ be two subspaces in $\mathbb{R}^n$ with the same dimension. The given subspaces are spanned by columns of matrices X and Y, respectively. Let the columns of the matrix $X_\perp$ form an orthogonal basis for $(\Re(X))^\perp$ which is an orthogonal complement to $\Re(X)$. Then nonzero singular values of the matrix $X_\perp^T Y$ are the sines of nonzero canonical angles between the subspaces $\Re(X)$ and $\Re(Y)$.*

Let $\{\varepsilon_t, t \geq 1\}$ be a fixed positive sequence, and $\varepsilon_t \to 0$ as $t \to \infty$. Then the estimator $\hat\lambda = \hat\lambda(t)$ is defined as any measurable solution to the inequality

$$Q_c(\hat\lambda) \leq \inf_{\lambda_1,\lambda_2 \geq 0} Q_c(\lambda) + \varepsilon_t, \tag{3}$$

where $t := \min(m_1, m_2)$.

Theorem 1. *Suppose that conditions (a) to (j) are satisfied, then*

$$\hat\lambda \to \lambda^0 := (\lambda_1^0, \lambda_2^0), \ \textit{as } m_1, m_2 \to \infty, \textit{a.s.}$$

6. Estimator of the kernel

Under the conditions of Theorem 1, one can obtain the estimator $\hat\lambda$ for the unknown scalars λ^0 and then construct two estimators $V_{p_1}(\hat\lambda)$ and $V_{p_2}(\hat\lambda)$ for a subspace V_p. But we want to construct a single subspace estimator using both clusters. It is reasonable to consider a compound data matrix $D_c := \begin{bmatrix} D(1) \\ D(2) \end{bmatrix}$ and $W_{cj} := W_j(1) + W_j(2),\ j = 1, 2$. Then $D_c = \bar D_c + \tilde D_c$ and

$$H := E\tilde D_c^T \tilde D_c = \begin{bmatrix} \lambda_1^0 W_{c1} & 0 \\ 0 & \lambda_2^0 W_{c2} \end{bmatrix}.$$

Define the matrix

$$\hat H = D_c^T D_c - \begin{bmatrix} \hat\lambda_1 W_{c1} & 0 \\ 0 & \hat\lambda_2 W_{c2} \end{bmatrix}.$$

Let $V_p(\hat H)$ denote the subspace spanned by the first p eigenvectors of $\hat H$ corresponding to the smallest eigenvalues, then $\hat V_p = V_p(\hat H) = \mathrm{span}(\hat z_1, \hat z_2, \ldots, \hat z_p)$.

Definition 2. *Let V_1 and V_2 be two subspaces in $\mathbb{R}^n$ with the same dimension p, $\sin\Theta$ be a diagonal matrix of the sines of canonical angles between the subspaces V_1 and V_2. A measurement of the proximity of the subspace V_1 to the subspace V_2 is a norm $\|\sin\Theta\| = \sqrt{\sum_{i\geq 1} \sin^2\Theta_{ii}}$.*

Theorem 2. *Under the conditions of Theorem 1, $\hat V_p \to V_p$ holds, i.e. $\|\sin\hat\Theta\| \to 0$, a.s., as $m_1, m_2 \to \infty$, where $\hat\Theta$ is the matrix of canonical angles between the subspaces $\hat V_p$ and V_p.*

7. Conclusions

We considered a multivariate errors-in-variables model $DZ \approx 0$ for the case where the total error covariance structure of the data matrix is known up to two scalar factors. The main assumption was that we observe two independent copies of the model. In practice, this means that the data can be partitioned in two separated groups, i.e. clusters. Based on the objective function method and using the second empirical moments, we have constructed the consistent estimators of the scalar factors and of the kernel of the data matrix.

Acknowledgement

The author thanks Prof. A. Kukush for posing the problem and fruitful discussions.

Appendix

Proof of Theorem 1.

1. Behavior of $Q_{\rm c}(\lambda^0)$

We have

$$Q_{\rm c}(\lambda^0) = \sum_{k=1}^{2}\sum_{i=1}^{p}\mu_{ik}^2(\lambda^0) + c\|\sin\Theta(\lambda^0)\|^2.$$

By Lemma 1, $\alpha_k := \left\|\frac{1}{m_k}\Psi_{\rm c}^{(k)}(\lambda^0) - \frac{1}{m_k}\bar{D}^\top(k)\bar{D}(k)\right\| \to 0$, as $m_k \to \infty$, a.s., for $k = 1, 2$. We have $\mu_i\big(\bar{D}^\top(k)\bar{D}(k)/m_k\big) = 0$, $1 \le i \le p$. Let

$$\delta_N := \inf_{m_k \ge N}\mu_{p+1}\left(\frac{1}{m_k}\bar{D}^\top(k)\bar{D}(k)\right).$$

Then, due to assumption (g), $\lim_{N\to\infty}\delta_N > 0$. Thus,

$$\lim_{m_1,m_2\to\infty}\sum_{k=1}^{2}\sum_{i=1}^{p}\mu_{ik}^2(\lambda^0) = 0,$$

and $\|\sin\Theta_k(\lambda^0)\| \le \frac{\alpha_k}{\delta_N}$, by Wedin's theorem [7]. Then $\|\sin\Theta_k(\lambda^0)\| \to 0$, as $m_k \to \infty$, a.s., $k = 1, 2$. Here, $\Theta_k(\lambda^0)$ is the diagonal matrix of canonical angles between the subspaces $V_p(\Psi_{\rm c}^{(k)}(\lambda^0)/m_k)$ and $V_p(\bar{D}^\top(k)\bar{D}(k)/m_k)$, and V_p denotes the span of p eigenvectors corresponding to p smallest eigenvalues.

Now,

$$V_p(\bar{D}^\top(1)\bar{D}(1)/m_1) = V_p(\bar{D}^\top(2)\bar{D}(2)/m_2) = \operatorname{span}(z_1, \dots, z_p).$$

Hence, $\|\sin\Theta(\lambda^0)\| \to 0$, as $m_1, m_2 \to \infty$, a.s. Thus,

$$Q_{\rm c}(\lambda^0) \to 0 \quad \text{as} \quad m_1, m_2 \to \infty, \quad \text{a.s.}$$

By inequality (3), we have

$$Q_{\rm c}(\hat{\lambda}) \to 0 \quad \text{as} \quad m_1, m_2 \to \infty, \quad \text{a.s.} \tag{4}$$

2. $\hat{\lambda}$ is eventually bounded

Now we want to construct such a nonrandom $L > 0$ that, eventually,

$$\|\hat{\lambda}\| \le L. \tag{5}$$

("Eventually" means: for all $n \geq n_0(\omega)$, a.s.) From (4) for any $\varepsilon_0 > 0$, we have

$$\sum_{k=1}^{2}\sum_{i=1}^{p}\mu_{ik}^2(\hat\lambda) \leq \varepsilon_0, \quad \text{eventually.} \tag{6}$$

By Lemma 1,

$$\left|\frac{1}{m_1}\Psi_{\mathrm{c}}^{(1)}(\hat\lambda) - \frac{1}{m_1}\Psi_{\mathrm{ls}}^{(1)}(\hat\lambda)\right| \to 0, \quad \text{as} \quad m_1, m_2 \to \infty, \quad \text{a.s.,} \tag{7}$$

where

$$\Psi_{\mathrm{ls}}^{(1)}(\lambda) := \bar D^\top(1)\bar D(1) - \begin{bmatrix}(\lambda_1-\lambda_1^0)W_1(1) & 0\\ 0 & (\lambda_2-\lambda_2^0)W_2(1)\end{bmatrix},$$

$\lambda := (\lambda_1, \lambda_2) \in [0,\infty)\times[0,\infty)$. Let Z satisfy condition (i). Then

$$\begin{aligned}&\mathbf{tr}\left(Z^\top(\Psi_{\mathrm{ls}}^{(1)}(\hat\lambda)/m_1)Z\right)\\ &\quad= -\mathbf{tr}\left((\hat\lambda_1-\lambda_1^0)Z_1^\top\big(W_1(1)/m_1\big)Z_1 + (\hat\lambda_2-\lambda_2^0)Z_2^\top\big(W_2(1)/m_1\big)Z_2\right).\end{aligned}$$

Suppose that $\hat\lambda_1 - \lambda_1^0 > L_0$, where $L_0 > 0$. Then

$$\mathbf{tr}\left(Z^\top(\Psi_{\mathrm{ls}}^{(1)}(\hat\lambda)/m_1 Z\right) \leq -L_0\,\mathbf{tr}\big(Z_1^\top\big(W_1(1)/m_1\big)Z_1\big) + \mathrm{const}\cdot\lambda_2^0.$$

But, due to (i) and (j),

$$\liminf_{m_1,m_2\to\infty}\ \mathbf{tr}\big(Z_1^\top(W_1(1)/m_1)Z_1\big) > 0.$$

Thus, for L_0 large enough and all $m_1 \geq m_{1_0}$, we have

$$\mathbf{tr}\left(Z^\top(\Psi_{\mathrm{ls}}^{(1)}(\hat\lambda)/m_1)Z\right) \leq -1.$$

So $\mu_{p+1}\big(\Psi_{\mathrm{ls}}^{(1)}(\hat\lambda)/m_1\big)$ is negative and separated from 0. Then we have from (7) that $\mu_{p+1}\big(\Psi_{\mathrm{c}}^{(1)}(\hat\lambda)/m_1\big)$ is also negative and separated from 0, starting from some random number. But this contradicts (6). Therefore, for a large enough nonrandom L_0, $\hat\lambda_1-\lambda_1^0 \leq L_0$ holds eventually. A similar inequality can be shown for $\hat\lambda_2 - \lambda_2^0$, based on condition (i) for $j = 2$. Thus, (5) holds eventually.

3. Consistency of the estimator

Let us have any fixed set Ω_0, $\Pr(\Omega_0) = 1$, such that, for all $\omega \in \Omega_0$ and for $m_1 \geq m_{10}(\omega)$, $m_2 \geq m_{20}(\omega)$, $\|\hat\lambda(\omega)\| \leq L$ holds. Fix $\omega \in \Omega_0$ and consider a bounded sequence

$$\{\,\hat\lambda(\omega; m_1, m_2):\ m_1 \geq m_{10}(\omega), m_2 \geq m_{20}(\omega)\}. \tag{8}$$

We will show that sequence (8) tends to λ^0, as $m_1, m_2 \to \infty$. Let

$$\left\{\,\hat\lambda\big(\omega; m_1(q), m_2(q)\big),\ q \geq 1\right\}$$

be any convergent subsequence, i.e. for a certain $\lambda^\infty \in R^2$

$$\lim_{q\to\infty}\hat\lambda\big(\omega; m_1(q), m_2(q)\big) = \lambda^\infty.$$

We want to prove the convergence of (8) and show that $\lambda^\infty = \lambda^0$. Let

$$M^{(k)}(m_k) := \mathrm{diag}(\mu_{1k}, \mu_{2k}, \dots, \mu_{pk})$$

and

$$Y^{(k)}(m_k) := [f_1^{(k)}(m_k) \quad \dots \quad f_p^{(k)}(m_k)]$$

be a matrix, the columns of which are the first p eigenvalues of the matrix, $\Psi_{\mathrm{c}}^{(k)}(\hat{\lambda})/m_k$. These columns form an orthogonal basis for $V_{pk}(\hat{\lambda})$. According to (4), $M^{(k)}(m_k) \to 0$, as $m_k \to \infty$. Moreover, $\sin\Theta(\hat{\lambda}) \to 0$, as $q \to \infty$. Here, $\hat{\lambda} = \hat{\lambda}(m_1(q), m_2(q))$. Therefore, we can assume

$$\left\| Y^{(1)}(m_1) - Y^{(2)}(m_2) \right\| \to 0, \quad \text{as} \quad m_1, m_2 \to \infty,$$

where $m_k = m_k(q),\ k = 1, 2$. Then

$$\frac{1}{m_k}\Psi_{\mathrm{c}}^{(k)}(\hat{\lambda})Y^{(k)}(m_k) = M^{(k)}(m_k)Y^{(k)}(m_k). \tag{9}$$

We suppose that

$$Y^{(k)}\big(m_k(q)\big) \to Y_\infty^{(k)},\ q \to \infty,\ k = 1, 2, \tag{10}$$

and $Y_\infty^{(k)} = [y_{\infty_1}^{(k)} \dots y_{\infty_p}^{(k)}]$. (Otherwise we should consider an embedded subsequence of numbers $m_k(q'))$. The matrix $\Theta(\hat{\lambda})$ is a matrix of canonical angles between

$$\operatorname{span}(f_1^{(1)}(m_1)\ \dots\ f_p^{(1)}(m_1)) \text{ and } \operatorname{span}(f_1^{(2)}(m_2)\ \dots\ f_p^{(2)}(m_2)).$$

Then, due to (10), we have $\sin\Theta_\infty = 0$, where Θ_∞ is the matrix of canonical angles between $\operatorname{span}(y_{\infty_1}^{(1)} \dots y_{\infty_p}^{(1)})$ and $\operatorname{span}(y_{\infty_1}^{(2)} \dots y_{\infty_p}^{(2)})$. Thus, these spans coincide. Since $Y^{(k)}(m_k)(q)$ are matrices of orthonormal columns, the matrices $Y_\infty^{(1)}$ and $Y_\infty^{(2)}$ have the same property. Then, from the coincidence of the linear spans of the columns of the matrices, we have $Y_\infty^{(1)} = Y_\infty^{(2)}U$, where U is the $p \times p$ orthogonal matrix corresponding to the choose of another basis in the linear span of columns of the matrix $Y_\infty^{(1)}$. Due to (9),

$$\sup_{\|\lambda\| \le L} \left\| \frac{1}{m_k}\Psi_{\mathrm{c}}^{(k)}(\lambda) - \frac{1}{m_k}\Psi_{\mathrm{ls}}^{(k)}(\lambda) \right\| \to 0, \quad \text{as} \quad q \to \infty.$$

Thus

$$\frac{1}{m_k(q)}\Psi_{\mathrm{ls}}^{(k)}(\lambda^\infty)Y_\infty^{(k)} \to 0, \quad \text{as} \quad q \to \infty, \quad k = 1, 2.$$

Let $Y_\infty = Y_\infty^{(2)}$. Multiplying it by U for $k = 1$, we obtain

$$\frac{1}{m_k(q)}\Psi_{\mathrm{ls}}^{(k)}(\lambda^\infty)Y_\infty \to 0,$$

but this also holds for $k = 2$.

Next,

$$\left(\frac{1}{m_1(q)}\Psi_{\mathrm{ls}}^{(1)}(\lambda^\infty) - \frac{1}{m_2(q)}\Psi_{\mathrm{ls}}^{(2)}(\lambda^\infty)\right) Y_\infty \to 0, \quad \text{as} \quad q \to \infty.$$

Due to (j), we have

$$\left(\frac{1}{m_1(q)}\bar{D}^\top(1)\bar{D}(1) - \frac{1}{m_2(q)}\bar{D}^\top(2)\bar{D}(2)\right) Y_\infty \to 0, \quad \text{as} \quad q \to \infty.$$

On the other hand, assumption (h) yields

$Y_\infty = [y_{\infty_1} \dots y_{\infty_p}]$ and $\operatorname{span}(y_{\infty_1} \dots y_{\infty_p}) = \operatorname{span}(z_1, \dots, z_p)$. Due to (9),

$$\frac{1}{m_k(q)}\Psi_{\mathrm{ls}}^{(k)}(\lambda^\infty)Z \to 0, \quad \text{as} \quad q \to \infty, \quad k = 1, 2.$$

Therefore,

$$\begin{bmatrix} (\lambda_1^\infty - \lambda_1^0)W_1(1)/m_1(q) & 0 \\ 0 & (\lambda_2^\infty - \lambda_2^0)W_2(1)/m_2(q) \end{bmatrix} Z \to 0,$$

$$(\lambda_j^\infty - \lambda_j^0)\mathbf{tr}\big(Z_j^\top W_j(1) Z_j/m_j(q)\big) \to 0, \quad \text{as} \quad t\to\infty, \quad j=1,2.$$

Conditions (i) and (j) imply that $\lambda_j^\infty = \lambda_j^0$, $j=1,2$. Hence, any convergent subsequence of the bounded sequence (8) converges to λ^0; therefore, the sequence (8) itself converges to λ^0. The convergence holds for all $\omega\in\Omega_0$, $\Pr(\Omega_0)=1$. Thus, $\hat\lambda\to\lambda^0$, as $m_1, m_2\to\infty$, a.s. □

Proof of Theorem 2.

By Theorem 1,

$$\left\|\frac{1}{m_1+m_2}(\hat H - \bar D_c^\top \bar D_c)\right\| \to 0, \quad \text{as} \quad m_1,m_2\to\infty, \quad \text{a.s.}$$

Due to condition (g),

$$\mu_1\left(\frac{1}{m_1+m_2}\bar D_c^\top \bar D_c\right) = \ldots = \mu_p\left(\frac{1}{m_1+m_2}\bar D_c^\top \bar D_c\right) = 0,$$

and, by Lemma 1, $\mu_{p+1}\left(\bar D_c^\top \bar D_c/m\right) > 0$ for large m_1, m_2. Moreover, the kernel equals

$$V_p\left(\frac{1}{m_1+m_2}\bar D_c^\top \bar D_c\right) = \operatorname{span}(z_1,\ldots,z_p).$$

By Wedin's theorem [7], this implies that $\hat\Theta\to 0$, as $m_1,m_2\to\infty$., where $\hat\Theta$ is the diagonal matrix of the canonical angles between $L_p(\hat H)$ and $\operatorname{span}(z_1,\ldots,z_p)$. □

Bibliography

1. A.Kukush, I.Markovsky, and S.Van Huffel, *Consistency of the structured total least squares estimator in a multivariate errors-in-variables model*, Journal of Statistical Planning and Inference **133** (2005), no. 2, 315-358.
2. A.Kukush and S.Van Huffel, *Consistency of element-wise weighted total least squares estimator in multivariate errors-in-variables model $AX = B$*, Metrika **59** (2004), no. 1, 75-97.
3. A.Kukush, I.Markovsky, and S.Van Huffel, *Estimation in a linear multivariate measurement error model with clustering in the regressor*, Internal Report 05-170. ESAT-SISTA. K.U.Leuven (Leuven, Belgium) (2005).
4. A.Kukush and M.Polekha, *Consistent estimator in multivariate errors-in-variables model under unknown error covariance structure*, Ukrainian Mathematical Journal **59** (2007), no. 8, 1026-1033.
5. A.Kukush and S.Zwanzing, *About the adaptive minimum contrast estimator in a model with nonlinear functional relations*, Ukrainian Mathematical Journal **53** (2001), 1145-1452.
6. A.Wald, *The fitting of straight lines if both variables are subject to error*, Ann. Math. Stat. **11** (1940), 284-300.
7. G.Stewart and J.Sun, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.
8. I.Markovsy, A.Kukush, and S.Van Huffel, *On errors-in-variables estimation with unknown noise variance ratio*, 14th IFAC Symposium on System Identification. Newcastle. Australia (2006).
9. I.Markovsy, A.Kukush, and S.Van Huffel, *Estimation in a linear multivariate measurement error model with a change point in data*, Computational Statistics and Data Analysis **52** (2007), no. 2, 1167-1182.

Department of Mathematical Analysis, Taras Shevchenko Kyiv National University, 6, Academician Glushkov Ave., Kyiv 03022, Ukraine

E-mail: poleha@bigmir.net