Communications for Statistical Applications and Methods
2018, Vol. 25, No. 6, 577–590

https://doi.org/10.29220/CSAM.2018.25.6.577
Print ISSN 2287-7843 / Online ISSN 2383-4757

Application of covariance adjustment to seemingly unrelated multivariate regressions

Lichun Wang[1,a], Lawrence Pettit[b]

[a]Department of Mathematics, Beijing Jiaotong University, China;
[b]School of Mathematical Sciences, Queen Mary University of London, UK

Abstract

Employing the covariance adjustment technique, we show that in the system of two seemingly unrelated multivariate regressions the estimator of regression coefficients can be expressed as a matrix power series, and conclude that the matrix series only has a unique simpler form. In the case that the covariance matrix of the system is unknown, we define a two-stage estimator for the regression coefficients which is shown to be unique and unbiased. Numerical simulations are also presented to illustrate its superiority over the ordinary least square estimator. Also, as an example we apply our results to the seemingly unrelated growth curve models.

Keywords: seemingly unrelated regressions, matrix power series, two-stage estimator

1. Introduction

The system of seemingly unrelated regressions (SUR) has been investigated by many authors since the pioneering works of Zellner (1962, 1963), which can be used to model subtle interactions among individual statistical relationships. For more details, the readers are referred to Revankar (1974), Schmidt (1977), Wang (1989), Percy (1992), Liu (2000), Liu (2002). Among these, the cases of orthogonal regressors (Zeller, 1963) and triangular SUR models (Revankar, 1974) and an SUR with unequal numbers of observations (Schmidt, 1977) are more impressive. Some examples in the econometrics literature (Srivastava and Giles, 1987) suggest that the SUR model is appropriate and useful for a wide range of applications. Further, Velu and Richards (2008) focuses on some applications of reduced-rank model in the context of SUR. Alkhamisi (2010) proposes two SUR type estimators based on combining the SUR ridge regression and the restricted least squares methods as well as evaluates their performances by means of some designated criteria. Zhou *et al.* (2011) also employs seemingly unrelated nonparametric regression models to fit the multivariate panel data. Shukur and Zeebari (2012) considers median regression for SUR models with the same explanatory variables and obtains an interesting feature of the generalized least absolute deviations method. However, this paper will show some interesting facts about the SUR system by employing the covariance adjustment technique. We start from the system of seemingly unrelated multivariate regressions (Gupta and Kabe, 1998), namely

$$\begin{cases} Y_1 = X_1 B_1 + E_1, \\ Y_2 = X_2 B_2 + E_2, \end{cases} \tag{1.1}$$

[1] Corresponding author: Department of Mathematics, Beijing Jiaotong University, No.3 Shangyuancun Haidian District, Beijing 100044, China. E-mail: lchwang@bjtu.edu.cn

Published 30 November 2018 / journal homepage: http://csam.or.kr

where Y_i $(i = 1, 2)$ are $n \times q$ observation variables; X_i $(i = 1, 2)$ are $n \times p_i$ matrices with full column rank; B_i $(i = 1, 2)$ are $p_i \times q$ unknown regression coefficients; E_1 and E_2 are random error matrices and the row variable of (E_1, E_2) follow a common unspecified multivariate distribution with mean zero and covariance matrix V, where V is a 2×2 non-diagonal partitioned matrix and given by

$$V = \begin{pmatrix} V_1 & D \\ D^T & V_2 \end{pmatrix}, \tag{1.2}$$

where V_i is the variance-covariance matrix of the row variable of E_i $(i = 1, 2)$ and D denotes the covariance matrix between the row variable of E_1 and the corresponding row variable of E_2. The different rows of (E_1, E_2) are also assumed to be uncorrelated. The case of multivariate SUR is common in biological science. For instance, if the i^{th} row of Y_1 denotes the observation vector of the weight of the i^{th} rabbit at q different time points and the i^{th} row of Y_2 denotes the observation vector of the length of the i^{th} rabbit at the same q time points, and the observation values of different rabbits are uncorrelated, then the multivariate SUR (1.1) reasonably model the interactions among the observation vectors of weight and length of n rabbits.

If one neglects the correlation between Y_1 and Y_2, i.e., taking D as zero, then only by the first equation of the system (1.1), one would obtain the least square estimator (LSE) for $\mathrm{Vec}(B_1)$ as

$$\widehat{\mathrm{Vec}}(B_1) = \left(I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right) \mathrm{Vec}(Y_1), \tag{1.3}$$

and correspondingly, the LSE of the coefficients matrix B_1 is $\hat{B}_1 = (X_1^T X_1)^{-1} X_1^T Y_1$, where $\mathrm{Vec}(A)$ denotes the direct operator of matrix A, $\otimes$, and I_q are the Kronecker product operator and the identity matrix of q order, respectively.

However, if we denote $Y = (Y_1, Y_2)$, $B = (B_1, B_2)$, and $E = (E_1, E_2)$, then the system (1.1) can also be represented as:

$$\mathrm{Vec}(Y) = \begin{pmatrix} I_q \otimes X_1 & 0 \\ 0 & I_q \otimes X_2 \end{pmatrix} \mathrm{Vec}(B) + \mathrm{Vec}(E). \tag{1.4}$$

Hence, from (1.4), one can obtain the LSE of $\mathrm{Vec}(B)$, say $\overline{\mathrm{Vec}}(B)$, and accordingly another estimator for $\mathrm{Vec}(B_1)$, denoted by $\overline{\mathrm{Vec}}(B_1)$, can be proposed since $\overline{\mathrm{Vec}}(B)^T = (\overline{\mathrm{Vec}}(B_1)^T, \overline{\mathrm{Vec}}(B_2)^T)$. We think it makes sense that $\overline{\mathrm{Vec}}(B_1)$ and its corresponding two-stage estimator version $\overline{\mathrm{Vec}}(B_1)_{\text{2-stage}}$ (in case of unknown V) should outperform $\widehat{\mathrm{Vec}}(B_1)$ (1.3) since they take the another equation information on B_1 into account.

The covariance adjustment technique is usually employed to obtain an optimal unbiased estimator of a vector parameter θ via linearly combining an unbiased estimator of θ, say T_1, and an unbiased estimator of a zero vector, say T_2 (Rao, 1967; Baksalary, 1991).

Applying the covariance adjustment technique to the estimator $\widehat{\mathrm{Vec}}(B_1)$, which only uses the first equation information on $\mathrm{Vec}(B_1)$, we firstly use $(I_q \otimes N_2)\mathrm{Vec}(Y_2)$ to improve $\widehat{\mathrm{Vec}}(B_1)$ noting $E[(I_q \otimes N_2)\mathrm{Vec}(Y_2)] = (I_q \otimes N_2)(I_q \otimes X_2)\mathrm{Vec}(B_2) = 0$ and obtain $\widehat{\mathrm{Vec}}(B_1)^{(1)}$, secondly we again improve $\widehat{\mathrm{Vec}}(B_1)^{(1)}$ by $(I_q \otimes N_1)\mathrm{Vec}(Y_1)$ due to $E[(I_q \otimes N_1)\mathrm{Vec}(Y_1)] = 0$ and get $\widehat{\mathrm{Vec}}(B_1)^{(2)}$. Repeating this process, we obtain the following estimator sequence ($k \geq 1$) for $\mathrm{Vec}(B_1)$:

$$\begin{aligned}
\widehat{\mathrm{Vec}}(B_1)^{(2k-1)} &= \widehat{\mathrm{Vec}}(B_1)^{(2k-2)} - \mathrm{Cov}\left(\widehat{\mathrm{Vec}}(B_1)^{(2k-2)}, (I_q \otimes N_2)\mathrm{Vec}(Y_2)\right) \\
&\quad \times \left[\mathrm{Cov}\left((I_q \otimes N_2)\mathrm{Vec}(Y_2)\right)\right]^{-} (I_q \otimes N_2)\mathrm{Vec}(Y_2), \qquad (1.5) \\
\widehat{\mathrm{Vec}}(B_1)^{(2k)} &= \widehat{\mathrm{Vec}}(B_1)^{(2k-1)} - \mathrm{Cov}\left(\widehat{\mathrm{Vec}}(B_1)^{(2k-1)}, (I_q \otimes N_1)\mathrm{Vec}(Y_1)\right) \\
&\quad \times \left[\mathrm{Cov}\left((I_q \otimes N_1)\mathrm{Vec}(Y_1)\right)\right]^{-} (I_q \otimes N_1)\mathrm{Vec}(Y_1), \qquad (1.6)
\end{aligned}$$

where $\widehat{\mathrm{Vec}}(B_1)^{(0)} = \widehat{\mathrm{Vec}}(B_1)$, $N_i = I_n - X_i(X_i^T X_i)^{-1}X_i^T$ $(i = 1, 2)$, and A^- denotes any a generalized inverse matrix of A.

Note that $\mathrm{Cov}(\mathrm{Vec}(Y_i), \mathrm{Vec}(Y_i)) = V_i \otimes I_n$ $(i = 1, 2)$ and $\mathrm{Cov}(\mathrm{Vec}(Y_1), \mathrm{Vec}(Y_2)) = D \otimes I_n$. By some algebra computations, we obtain that for $k \geq 1$

$$\widehat{\mathrm{Vec}}(B_1)^{(2k-1)} = \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right] \sum_{i=0}^{k-1} \left(DV_2^{-1}D^T V_1^{-1} \otimes N_2N_1\right)^i \times \left[\mathrm{Vec}(Y_1) - \left(DV_2^{-1} \otimes N_2\right)\mathrm{Vec}(Y_2)\right], \qquad (1.7)$$

$$\begin{aligned}
\widehat{\mathrm{Vec}}(B_1)^{(2k)} &= \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right] \sum_{i=0}^{k} \left(DV_2^{-1}D^T V_1^{-1} \otimes N_2N_1\right)^i \mathrm{Vec}(Y_1) \\
&\quad - \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right] \sum_{i=0}^{k-1} \left(DV_2^{-1}D^T V_1^{-1} \otimes N_2N_1\right)^i \left(DV_2^{-1} \otimes N_2\right) \times \mathrm{Vec}(Y_2). \qquad (1.8)
\end{aligned}$$

Denote $V^{-1} = \begin{pmatrix} V^{11} & V^{12} \\ V^{21} & V^{22} \end{pmatrix}$ and $Q = (V^{11})^{-1}V^{12}(V^{22})^{-1}V^{21}$. By (1.2) and the inverse of partitioned matrix, we have

$$\begin{aligned}
Q &= \left(V^{11}\right)^{-1} V^{12} \left(V^{22}\right)^{-1} V^{21} \\
&= \left[V_1 - DV_2^{-1}D^T\right] \cdot \left[V_1 - DV_2^{-1}D^T\right]^{-1} DV_2^{-1} \cdot \left(V_2 - D^T V_1^{-1} D\right) \cdot V_2^{-1}D^T \left[V_1 - DV_2^{-1}D^T\right]^{-1} \\
&= DV_2^{-1}D^T \left[\left(V_1 - DV_2^{-1}D^T\right)^{-1} - V_1^{-1}DV_2^{-1}D^T \left(V_1 - DV_2^{-1}D^T\right)^{-1}\right] \\
&= DV_2^{-1}D^T V_1^{-1}, \qquad (1.9)
\end{aligned}$$

and $(V^{11})^{-1}V^{12} = -DV_2^{-1}$. Thus we have

$$\begin{aligned}
\overline{\mathrm{Vec}}(B_1) &= \widehat{\mathrm{Vec}}(B_1)^{(\infty)} = \lim_{k\to\infty} \widehat{\mathrm{Vec}}(B_1)^{(2k-1)} = \lim_{k\to\infty} \widehat{\mathrm{Vec}}(B_1)^{(2k)} \\
&= \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right] \sum_{i=0}^{\infty} (Q \otimes N_2N_1)^i \times \left[\mathrm{Vec}(Y_1) + \left(\left(V^{11}\right)^{-1} V^{12} \otimes N_2\right)\mathrm{Vec}(Y_2)\right] \\
&= \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right] \left\{I_{qn} - Q \sum_{i=0}^{\infty} (Q \otimes P_2P_1)^i P_2N_1\right\} \\
&\quad \times \left\{\mathrm{Vec}(Y_1) + \left[\left(V^{11}\right)^{-1} V^{12} \otimes N_2\right] \mathrm{Vec}(Y_2)\right\}, \qquad (1.10)
\end{aligned}$$

where $P_i = I_n - N_i = X_i(X_i^T X_i)^{-1}X_i^T$ and we use the facts that $(Q \otimes P_2P_1)^0 = I_q \otimes I_n = I_{qn}$, $(V^{11})^{-1}(Q^T)^k V^{11} = Q^k$ for $k \geq 0$ and $X_1^T(N_2N_1)^k = -X_1^T(P_2P_1)^{k-1}P_2N_1$ for $k \geq 1$.

Then, we integrate the above conclusions into the following theorem, which indicates the limit of the covariance adjustment sequence and the covariance of $\overline{\mathrm{Vec}}(B_1)$.

Theorem 1. *For the system (1.1), the limit of the covariance adjustment sequence of $Vec(B_1)$ equals to $\overline{Vec}(B_1)$, i.e., $\lim_{k\to\infty}\widehat{Vec}(B_1)^{(k)} = \overline{Vec}(B_1)$, and*

$$Cov\left(\overline{Vec}(B_1)\right) = V_1 \otimes \left(X_1^T X_1\right)^{-1} - \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right] \cdot G \cdot \left[I_q \otimes X_1\left(X_1^T X_1\right)^{-1}\right],$$

where $G = \sum_{i=0}^{\infty}[Q^i D V_2^{-1} D^T] \otimes [(P_1P_2P_1)^i - (P_1P_2P_1)^{i+1}]$.

Proof: The first conclusion follows from the above discussion. Denote $\overline{\mathrm{Vec}}(B_1) = M(Q)\{\mathrm{Vec}(Y_1) + [(V^{11})^{-1}V^{12} \otimes N_2]\mathrm{Vec}(Y_2)\}$ with $M(Q) = [I_q \otimes (X_1^T X_1)^{-1} X_1^T]\{I_{qn} - Q\sum_{i=0}^{\infty}(Q \otimes P_2P_1)^i P_2 N_1\}$, we have

$$\begin{aligned}
&\mathrm{Cov}\left(\overline{\mathrm{Vec}}(B_1)\right)\\
&= M(Q)\left[V_1 \otimes I_n + \left(V^{11}\right)^{-1} V^{12} V_2 V^{21}\left(V^{11}\right)^{-1} \otimes N_2 + D V^{21}\left(V^{11}\right)^{-1} \otimes N_2 + \left(V^{11}\right)^{-1} V^{12} D^T \otimes N_2\right] M^T(Q)\\
&= M(Q)[V_1 \otimes I_n - Q V_1 \otimes N_2] M^T(Q), \qquad (1.11)
\end{aligned}$$

where we use the following fact

$$\left(V^{11}\right)^{-1} V^{12} V_2 V^{21}\left(V^{11}\right)^{-1} + D V^{21}\left(V^{11}\right)^{-1} + \left(V^{11}\right)^{-1} V^{12} D^T = -D V_2^{-1} D^T. \qquad (1.12)$$

Together with the expression of $M(Q)$, we have

$$\begin{aligned}
&\mathrm{Cov}\left(\overline{\mathrm{Vec}}(B_1)\right)\\
&= \left\{I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T - \sum_{i=0}^{\infty}\left[Q^{i+1} \otimes \left(X_1^T X_1\right)^{-1} X_1^T (P_2P_1)^i P_2 N_1\right]\right\}\\
&\quad \times \left\{V_1 \otimes X_1\left(X_1^T X_1\right)^{-1} - \sum_{i=0}^{\infty}\left[V_1\left(Q^T\right)^{i+1} \otimes N_1 P_2 (P_1P_2)^i X_1\left(X_1^T X_1\right)^{-1}\right]\right\}\\
&\quad - \left\{I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T - \sum_{i=0}^{\infty}\left[Q^{i+1} \otimes \left(X_1^T X_1\right)^{-1} X_1^T (P_2P_1)^i P_2 N_1\right]\right\}\\
&\quad \times \left\{Q V_1 \otimes N_2 X_1\left(X_1^T X_1\right)^{-1} - \sum_{i=0}^{\infty}\left[Q V_1\left(Q^T\right)^{i+1} \otimes N_2 N_1 P_2 (P_1P_2)^i X_1\left(X_1^T X_1\right)^{-1}\right]\right\}\\
&= V_1 \otimes \left(X_1^T X_1\right)^{-1} - \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right][Q V_1 \otimes N_2]\left[I_q \otimes X_1\left(X_1^T X_1\right)^{-1}\right]\\
&\quad + \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right]\left[\sum_{i=0}^{\infty} Q^{i+2} V_1 \otimes (P_2P_1)^i P_2 N_1 N_2\right]\left[I_q \otimes X_1\left(X_1^T X_1\right)^{-1}\right]\\
&\quad + \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right]\left[\sum_{i=0}^{\infty} Q V_1\left(Q^T\right)^{i+1} \otimes N_2 N_1 P_2 (P_1P_2)^i\right]\left[I_q \otimes X_1\left(X_1^T X_1\right)^{-1}\right]\\
&\quad + \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right]\left[\sum_{i=0}^{\infty} Q^{i+1} \otimes (P_2P_1)^i P_2 N_1 \sum_{i=0}^{\infty} V_1\left(Q^T\right)^{i+1} \otimes N_1 P_2 (P_1P_2)^i\right]\left[I_q \otimes X_1\left(X_1^T X_1\right)^{-1}\right]\\
&\quad - \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right]\left[\sum_{i=0}^{\infty} Q^{i+1} \otimes (P_2P_1)^i P_2 N_1 \sum_{i=0}^{\infty} Q V_1\left(Q^T\right)^{i+1} \otimes N_2 N_1 P_2 (P_1P_2)^i\right]\left[I_q \otimes X_1\left(X_1^T X_1\right)^{-1}\right]
\end{aligned}$$

$$= V_1 \otimes \left(X_1^T X_1\right)^{-1} - \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right] [QV_1 \otimes N_2] \left[I_q \otimes X_1 \left(X_1^T X_1\right)^{-1}\right]$$
$$+ \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right] \left[\sum_{i=0}^{\infty} QV_1 \left(Q^T\right)^{i+1} \otimes P_2 (P_1 P_2)^{i+1} - \sum_{i=0}^{\infty} Q^{i+2} V_1 \otimes (P_2 P_1)^{i+1}\right]$$
$$\times \left[I_q \otimes X_1 \left(X_1^T X_1\right)^{-1}\right]. \tag{1.13}$$

Using $QV_1 = DV_2^{-1}D^T$, $X_1^T N_2 X_1 = X_1^T (I_n - P_1 P_2 P_1) X_1$, and $X_1^T P_1 = X_1^T$, we have

$$\mathrm{Cov}\left(\overline{\mathrm{Vec}}(B_1)\right)$$
$$= V_1 \otimes \left(X_1^T X_1\right)^{-1} - \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right] \left[DV_2 D^T \otimes (I_n - P_1 P_2 P_1)\right] \left[I_q \otimes X_1 \left(X_1^T X_1\right)^{-1}\right]$$
$$- \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right] \left\{\sum_{i=1}^{\infty} Q^i DV_2^{-1} D^T \otimes (P_1 P_2 P_1)^i - \sum_{i=1}^{\infty} DV_2^{-1} D^T \left(Q^T\right)^i \otimes (P_1 P_2 P_1)^{i+1}\right\}$$
$$\times \left[I_q \otimes X_1 \left(X_1^T X_1\right)^{-1}\right]$$
$$= V_1 \otimes \left(X_1^T X_1\right)^{-1} - \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right] \left\{\sum_{i=0}^{\infty} Q^i DV_2^{-1} D^T \otimes (P_1 P_2 P_1)^i \sum_{i=0}^{\infty} DV_2^{-1} D^T \left(Q^T\right)^i \otimes (P_1 P_2 P_1)^{i+1}\right\}$$
$$\times \left[I_q \otimes X_1 \left(X_1^T X_1\right)^{-1}\right]$$
$$= V_1 \otimes \left(X_1^T X_1\right)^{-1} - \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right] \left\{\sum_{i=0}^{\infty} \left[Q^i DV_2^{-1} D^T\right] \otimes \left[(P_1 P_2 P_1)^i - (P_1 P_2 P_1)^{i+1}\right]\right\}$$
$$\times \left[I_q \otimes X_1 \left(X_1^T X_1\right)^{-1}\right], \tag{1.14}$$

where the last step uses the facts that $(P_1 P_2 P_1)^0 = I_n$ and $Q^i DV_2^{-1} D^T = DV_2^{-1} D^T (Q^T)^i$ for $i \geq 0$.

The proof of Theorem 1 is finished. □

Note that $Q^0 DV_2^{-1} D^T = DV_2^{-1} D^T \geq 0$, $I_n - P_1 P_2 P_1 \geq 0$ and for $i \geq 1$

$$Q^i DV_2^{-1} D^T = DV_2^{-1} D^T \left(Q^T\right)^i$$
$$= \begin{cases} DV_2^{-1} D^T \left(V_1^{-1} DV_2^{-1} D^T\right)^{k-1} V_1^{-1} \left(DV_2^{-1} D^T V_1^{-1}\right)^{k-1} DV_2^{-1} D^T \geq 0, & i = 2k-1, \\ DV_2^{-1} D^T \left(V_1^{-1} DV_2^{-1} D^T\right)^{k-1} V_1^{-1} DV_2^{-1} D^T V_1^{-1} \left(DV_2^{-1} D^T V_1^{-1}\right)^{k-1} DV_2^{-1} D^T \geq 0, & i = 2k, \end{cases}$$
$$k = 1, 2, \ldots, \tag{1.15}$$

and $(P_1 P_2 P_1)^i - (P_1 P_2 P_1)^{i+1} \geq 0$. Hence

$$G = \sum_{i=0}^{\infty} \left[Q^i DV_2^{-1} D^T\right] \otimes \left[(P_1 P_2 P_1)^i - (P_1 P_2 P_1)^{i+1}\right] \geq 0. \tag{1.16}$$

Further, since $\mathrm{Cov}(\widehat{\mathrm{Vec}}(B_1)) = V_1 \otimes (X_1^T X_1)^{-1}$, we have

$$\mathrm{Cov}\left(\overline{\mathrm{Vec}}(B_1)\right) = \mathrm{Cov}\left(\widehat{\mathrm{Vec}}(B_1)\right) - \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right] \cdot G \cdot \left[I_q \otimes X_1 \left(X_1^T X_1\right)^{-1}\right]$$
$$\leq \mathrm{Cov}\left(\widehat{\mathrm{Vec}}(B_1)\right), \tag{1.17}$$

which means $\overline{\mathrm{Vec}}(B_1)$ is superior to $\widehat{\mathrm{Vec}}(B_1)$ in the sense of having less covariance. This result is exactly consistent with the fact that $\widehat{\mathrm{Vec}}(B_1)$ only uses the first regression information on $\mathrm{Vec}(B_1)$, whereas $\overline{\mathrm{Vec}}(B_1)$ combines the second regression equation with the first one via covariance adjustment.

2. The characteristics of matrix series

Note that for $i = 1, 2, \ldots,$

$$X_1^T (P_2P_1)^{i-1} P_2N_1 = 0 \Longleftrightarrow X_1^T (P_2P_1)^{i-1} P_2N_1N_2 = 0. \tag{2.1}$$

We only need to prove that the right equality implies the left equality. Note that $X_1^T (P_2P_1)^{i-1} P_2N_1N_2 = 0$ concludes $X_1^T (P_2P_1)^{i-1} N_2N_1N_2 = 0$, hence one has $X_1^T (P_2P_1)^{i-1} N_2N_1N_2(P_1P_2)^{i-1} X_1 = 0$, thus $X_1^T (P_2P_1)^{i-1} N_2N_1 = 0$, where we use $N_1^2 = N_1$. Further, replace N_2 by $I_n - P_2$ and note that $X_1^T N_1 = 0$ and $P_1N_1 = 0$, we have $X_1^T (P_2P_1)^{i-1} P_2N_1 = 0$.

Therefore, (2.1) implies that for $i = 1, 2, \ldots,$

$$\left(X_1^T X_1\right)^{-1} X_1^T (P_2P_1)^{i-1} P_2N_1 = 0 \Longleftrightarrow \left(X_1^T X_1\right)^{-1} X_1^T (P_2P_1)^{i-1} P_2N_1N_2 = 0, \tag{2.2}$$

which further shows that for $i = 1, 2, \ldots,$

$$Q^i \otimes \left(X_1^T X_1\right)^{-1} X_1^T (P_2P_1)^{i-1} P_2N_1 = 0 \Longleftrightarrow Q^i \left(V^{11}\right)^{-1} V^{12} \otimes \left(X_1^T X_1\right)^{-1} X_1^T (P_2P_1)^{i-1} P_2N_1N_2 = 0, \tag{2.3}$$

where we note that $Q = DV_2^{-1} D^T V_1^{-1}$ and $Q^i (V^{11})^{-1} V^{12} = -Q^i DV_2^{-1}$ and D is the covariance matrix of E_1 and E_2, and that both Q and $Q^i (V^{11})^{-1} V^{12}$ are invertible.

Set

$$\overline{\mathrm{Vec}}(B_1)_s = \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right] \mathrm{Vec}(Y_1) + \left[\left(V^{11}\right)^{-1} V^{12} \otimes \left(X_1^T X_1\right)^{-1} X_1^T N_2\right] \mathrm{Vec}(Y_2). \tag{2.4}$$

The following theorem shows that the matrix series (1.10) only have one degeneration form $\overline{\mathrm{Vec}}(B_1)_s$.

Theorem 2. *$\overline{Vec}(B_1)_s$ is the unique simpler form of $\widehat{Vec}(B_1)^{(\infty)}$.*

Proof: Note that for any a fixed i ($i \geq 1$) that: if $X_1^T (P_2P_1)^{i-1} P_2N_1 = 0$, then $X_1^T (P_2P_1)^{i} P_2N_1 = X_1^T (P_2P_1)^{i-1} P_2(I_n - N_1)P_2N_1 = 0$. Step by step, we come to

$$X_1^T (P_2P_1)^{k-1} P_2N_1 = 0, \quad k = i + 1, i + 2, \ldots. \tag{2.5}$$

Thus, we find

$$\begin{aligned} & Q^i \otimes \left(X_1^T X_1\right)^{-1} X_1^T (P_2P_1)^{i-1} P_2N_1 = 0, \quad \text{for any a fixed } i \, (i \geq 1) \\ \Longrightarrow\ & Q^k \otimes \left(X_1^T X_1\right)^{-1} X_1^T (P_2P_1)^{k-1} P_2N_1 = 0, \quad k = i + 1, i + 2, \ldots. \end{aligned} \tag{2.6}$$

On the other hand, if for any a fixed i ($i \geq 2$) one has $X_1^T(P_2P_1)^{i-1}P_2N_1 = 0$, then it is easy to see that

$$\begin{aligned}
& X_1^T(P_2P_1)^{i-1}(I_n - N_2)N_1 = 0 \\
\Longrightarrow\ & X_1^T(P_2P_1)^{i-1}N_2N_1 = 0 \\
\Longrightarrow\ & X_1^T(P_2P_1)^{i-2}P_2(I_n - N_1)N_2N_1 = 0 \\
\Longrightarrow\ & X_1^T(P_2P_1)^{i-2}P_2N_1N_2N_1 = 0 \\
\Longrightarrow\ & X_1^T(P_2P_1)^{i-2}P_2N_1N_2N_1P_2(P_1P_2)^{i-2}X_1 = 0 \\
\Longrightarrow\ & X_1^T(P_2P_1)^{i-2}P_2N_1N_2 = 0 \\
\Longrightarrow\ & X_1^T(P_2P_1)^{i-2}P_2N_1 = 0,
\end{aligned} \tag{2.7}$$

where the last step comes from the fact (2.1). Thus, step by step we conclude that

$$\begin{aligned}
& Q^i \otimes \left(X_1^TX_1\right)^{-1} X_1^T(P_2P_1)^{i-1}P_2N_1 = 0, \quad \text{for any a fixed } i\,(i \geq 2) \\
\Longrightarrow\ & Q^k \otimes \left(X_1^TX_1\right)^{-1} X_1^T(P_2P_1)^{k-1}P_2N_1 = 0, \quad k = 1, 2, \ldots, i-1.
\end{aligned} \tag{2.8}$$

Combining (2.6) with (2.8), we know that for any a fixed i ($i \geq 1$) if

$$Q^i \otimes \left(X_1^TX_1\right)^{-1} X_1^T(P_2P_1)^{i-1}P_2N_1 = 0, \tag{2.9}$$

then the infinite series

$$\sum_{i=1}^{\infty} \left[Q^i \otimes \left(X_1^TX_1\right)^{-1} X_1^T(P_2P_1)^{i-1}P_2N_1\right] = 0, \tag{2.10}$$

and by (2.3), concurrently we conclude that the infinite series

$$\sum_{i=1}^{\infty} \left[Q^i \left(V^{11}\right)^{-1} V^{12} \otimes \left(X_1^TX_1\right)^{-1} X_1^T(P_2P_1)^{i-1}P_2N_1N_2\right] = 0. \tag{2.11}$$

Hence, $\widehat{\mathrm{Vec}}(B_1)^{(\infty)}$ has unique simpler form $\overline{\mathrm{Vec}}(B_1)_s$ in the sense that if one term in (2.10) or (2.11) is zero, then both infinite sums turn into zero.

The proof of Theorem 2 is finished. □

3. The properties of two-stage estimator

If the covariance matrix V is unknown, then both $\widehat{\mathrm{Vec}}(B_1)^{(\infty)}$ and the simpler form $\overline{\mathrm{Vec}}(B_1)_s$ are not available to use. Set $\tilde{X} = (X_1, X_2)$, we estimate V by

$$\hat{V} = \frac{1}{n - R\left(\tilde{X}\right)} \begin{pmatrix} Y_1^T \\ Y_2^T \end{pmatrix} (I_n - P_{\tilde{X}})(Y_1, Y_2), \tag{3.1}$$

where $R(\tilde{X})$ is the rank of $\tilde{X}$ and $P_{\tilde{X}} = \tilde{X}(\tilde{X}^T\tilde{X})^{-}\tilde{X}^T$.

Following from $E(a^T Ab) = \text{trace}[A\text{Cov}(b,a)] + (Ea)^T A(Eb)$ and $(I_n - P_{\tilde{X}})X_i = 0 (i = 1,2)$, where a and b denote two random vectors, we have $E[Y_i^T(I_n - P_{\tilde{X}})Y_i] = V_i[n - R(\tilde{X})]$ $(i = 1,2)$ and $E[Y_1^T(I_n - P_{\tilde{X}})Y_2] = D[n - R(\tilde{X})]$, which show that

$$E\hat{V} = \begin{pmatrix} V_1 & D \\ D^T & V_2 \end{pmatrix} = V. \tag{3.2}$$

Substituting the estimator $\hat{V}$ for V in the expressions of $\widehat{\text{Vec}}(B_1)^{(\infty)}$ and $\overline{\text{Vec}}(B_1)_s$, we obtain the following two two-stage estimators

$$\widehat{\text{Vec}}(B_1)^{(\infty)}_{\text{2-stage}} = M\left(\hat{Q}\right)\left\{\text{Vec}(Y_1) + \left[-\hat{D}\hat{V}_2^{-1} \otimes N_2\right]\text{Vec}(Y_2)\right\} \tag{3.3}$$

with $M(\hat{Q}) = [I_q \otimes (X_1^T X_1)^{-1} X_1^T]\{I_{qn} - \hat{Q}\sum_{i=0}^{\infty}(\hat{Q} \otimes P_2 P_1)^i P_2 N_1\}$ and $\hat{Q} = \hat{D}\hat{V}_2^{-1}\hat{D}^T\hat{V}_1^{-1}$, and

$$\overline{\text{Vec}}(B_1)_{s,\text{2-stage}} = \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right]\text{Vec}(Y_1) + \left[-\hat{D}\hat{V}_2^{-1} \otimes \left(X_1^T X_1\right)^{-1} X_1^T N_2\right]\text{Vec}(Y_2). \tag{3.4}$$

Similar to Theorem 2, we know that $\overline{\text{Vec}}(B_1)_{s,\text{2-stage}}$ is the unique simpler form of $\widehat{\text{Vec}}(B_1)^{(\infty)}_{\text{2-stage}}$. Hence, we focus on the performances of $\overline{\text{Vec}}(B_1)_{s,\text{2-stage}}$.

The matrix-variate normal distribution is a commonly used distribution in the class of matrix elliptically symmetric distributions. It plays an important role in the investigation of multivariate regression models such as the growth curve model (GCM). In what follows, in order to establish the unbiasedness of $\overline{\text{Vec}}(B_1)_{s,\text{2-stage}}$, we first briefly present the definition of the matrix-variate normal distribution as well as two related properties and then make some assumptions on the distributions of random error matrices E_i $(i = 1,2)$.

Definition 1. *A random matrix Z with order $n\times q$ is said to follow a matrix-variate normal distribution if its probability function is of the form*

$$f(Z) = (2\pi)^{-\frac{nq}{2}}[\det(\Sigma)]^{-\frac{q}{2}}[\det(\Omega)]^{-\frac{n}{2}}\exp\left(-\frac{1}{2}trace\left\{\Omega^{-1}[Z-M]^T\Sigma^{-1}[Z-M]\right\}\right),$$

where M, $\Sigma > 0$, and $\Omega > 0$ are $n \times q$, $n \times n$, and $q \times q$ matrices, respectively, and $\det(A)$ *is the determinant of the square matrix A. In this case, it is usually denoted that $Z \sim N_{n,q}(M,\Sigma,\Omega)$.*

The following two lemmas point out that the relationship between the matrix-variate and vector-variate normal distributions and an affine transformation of a matrix-variate normal variable also follows a matrix-variate normal distribution. The readers are referred to the first chapter of Pan and Fang (2007) for more details.

Lemma 1. *Let Z be a $n \times q$ random matrix and $z = Vec(Z)$. Then $Z \sim N_{n,q}(M,\Sigma,\Omega)$ if $z \sim N_{nq}$ $(Vec(M), \Omega \otimes \Sigma)$.*

Lemma 2. *Suppose $Z \sim N_{n,q}(M,\Sigma,\Omega)$, and that C, $A_1 > 0$, and $A_2 > 0$ are given matrices with orders $n \times q$, $n \times n$, and $q \times q$, respectively. Then $A_1 Z A_2 + C \sim N_{n,q}(A_1 M A_2 + C, A_1\Sigma A_1^T, A_2\Omega A_2^T)$.*

In the following, we assume that in the system (1.1) the random error matrices E_i $(i = 1,2)$ follow the matrix-variate normal distribution $N_{n,q}(0, I_n, V_i)$, which indicate that the rows of E_i are iid random vectors with common distribution $N_q(0, V_i)$ $(i = 1, 2)$, respectively. Thus, the rows of $E = (E_1, E_2)$ are iid random vectors with common distribution $N_{2q}(0, V)$, i.e., $E \sim N_{n,2q}(0, I_n, V)$. Hence, by Lemmas 1 and 2 we know that

$$\mathrm{Vec}(Y) = \mathrm{Vec}(Y_1, Y_2) \sim N_{2nq}\left(\begin{bmatrix} I_q \otimes X_1 & 0 \\ 0 & I_q \otimes X_2 \end{bmatrix}\mathrm{Vec}(B), V \otimes I_n\right). \tag{3.5}$$

Denote $Y_i = (y_1^{(i)}, y_2^{(i)}, \ldots, y_q^{(i)})$ $(i = 1, 2)$. Then the matrix $\hat{D} = [n - R(\tilde{X})]^{-1}(\hat{d}_{ij})_{q\times q}$ with the element

$$\hat{d}_{ij} = \left(y_i^{(1)}\right)^T [I_n - P_{\tilde{X}}]\, y_j^{(2)} = (\mathrm{Vec}(Y))^T \left[O_{i,q+j}(2q \times 2q) \otimes (I_n - P_{\tilde{X}})\right]\mathrm{Vec}(Y), \tag{3.6}$$

where the matrix $O_{i,q+j}(2q \times 2q)$ with order $2q \times 2q$ consists of all zeros only except the element in the i^{th} row and the $(q + j)^{th}$ column is one. Similarly, the $(i, j)^{th}$ element of $\hat{V}_2$ is equal to

$$(\mathrm{Vec}(Y))^T \left[O_{q+i,q+j}(2q \times 2q) \otimes (I_n - P_{\tilde{X}})\right]\mathrm{Vec}(Y),$$

where the $2q \times 2q$ order matrix $O_{q+i,q+j}(2q \times 2q)$ consists of all zeros only; except the element in the $(q + i)^{th}$ row and the $(q + j)^{th}$ column is one.

Note that $[I_q \otimes (X_1^T X_1)^{-1} X_1^T N_2]\mathrm{Vec}(Y_2) = [\mathbf{0}_{qp_1\times nq}, I_q \otimes (X_1^T X_1)^{-1} X_1^T N_2]\mathrm{Vec}(Y)$. Hence, using $X_1^T N_2[I_n - P_{\tilde{X}}] = 0$ and following from the discriminant condition of independence of the linear function and quadratic function of normal variables and the following easily verified facts:

$$\left[\mathbf{0}_{qp_1\times nq}, I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T N_2\right][V \otimes I_n]\left[O_{i,q+j}(2q \times 2q) \otimes (I_n - P_{\tilde{X}})\right] = 0, \tag{3.7}$$

and

$$\left[\mathbf{0}_{qp_1\times nq}, I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T N_2\right][V \otimes I_n]\left[O_{q+i,q+j}(2q \times 2q) \otimes (I_n - P_{\tilde{X}})\right] = 0. \tag{3.8}$$

We know that

$$\begin{aligned} E\left[\overline{\mathrm{Vec}}(B_1)_{s,\text{2-stage}}\right] &= \left[I_q \otimes \left(X_1^T X_1\right)^{-1} X_1^T\right](I_q \otimes X_1)\mathrm{Vec}(B_1) \\ &\quad + \left[\left(E\left(-\hat{D}\hat{V}_2^{-1}\right)\right) \otimes \left(X_1^T X_1\right)^{-1} X_1^T N_2\right]\left(I_q \otimes X_2\right)\mathrm{Vec}(B_2) \\ &= \mathrm{Vec}(B_1). \end{aligned} \tag{3.9}$$

Thus, we obtain the following theorem, which states the unbiasedness of the two-stage estimator.

Theorem 3. *Under the assumptions that $E_i \sim N_{n,q}(0, I_n, V_i)$ $(i = 1, 2)$, the two-stage estimator $\overline{Vec}(B_1)_{s,\text{2-stage}}$ is unbiased, i.e., $E[\overline{Vec}(B_1)_{s,\text{2-stage}}] = Vec(B_1)$.*

In the following, we refer to Grunfeld's data in Maddala (1977) and present two simulation studies to compare the performances of $\overline{\mathrm{Vec}}(B_1)_{s,\text{2-stage}}$ with those of $\widehat{\mathrm{Vec}}(B_1)$ under the conditions that there are some known relationships between the design matrices X_1 and X_2 and no relationships between X_1 and X_2, respectively.

Table 1: Comparisons between the two-stage estimator and the least square estimator

ρ	n	trace($S(\hat{B}_1)$)	trace($S(\bar{B}_{1,s,\text{2-stage}})$)	trace($S(\bar{B}_{1,s})$)
	10	12.7118	12.8775	12.7718
0.2	20	30.0456	30.2155	30.0873
	50	141.1215	141.9674	141.3323
	10	15.3561	15.4368	15.5200
0.5	20	35.3589	35.3747	35.3775
	50	108.2523	108.2895	108.2894
	10	7.7018	8.0020	7.9504
0.7	20	42.4375	45.8153	44.6528
	50	103.9155	104.3501	104.2045
	10	18.6055	18.6758	18.6689
0.9	20	35.6630	37.5563	37.2077
	50	119.9993	122.4540	122.2700

(I) The case that $X_1 = (X_2, L)$

Where the system (1.1) is of the form $Y_i = X_i B_i + E_i$ $(i = 1, 2)$ with $E = (E_1, E_2) \sim N_{n,4}(0, I_n, V)$, and

$$B_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 1 & 6 \\ -3 & 2 \end{pmatrix}, \quad V = \begin{pmatrix} 1 & 0 & \rho & 0 \\ 0 & 1 & 0 & \rho \\ \rho & 0 & 1 & 0 \\ 0 & \rho & 0 & 1 \end{pmatrix}. \tag{3.10}$$

Set $S(B_1) = (Y_1 - X_1B_1)^T(Y_1 - X_1B_1)$. Note that the estimator $\hat{B}_1 = (X_1^T X_1)^{-1} X_1^T Y_1$ given by (1.3), which corresponds to the LSE $\widehat{\text{Vec}}(B_1)$, actually makes the residual sum of squares (in the sense of nonnegative definite), trace of $S(B_1)$, determinant of $S(B_1)$ and the largest eigenvalue of $S(B_1)$ achieve their minimums (Muirhead, 1982). Therefore, under the four different criteria of measurement, if only the first equation $Y_1 = X_1B_1 + E_1$ is used then the LSE of the regression coefficient B_1 are completely identical (Fang and Zhang, 1990). Thus, without loss of generality, we illustrate the superiorities of $\overline{\text{Vec}}(B_1)_{s,\text{2-stage}}$ by comparing trace($S(\hat{B}_1)$) with trace($S(\bar{B}_{1,s,\text{2-stage}})$), where $\bar{B}_{1,s,\text{2-stage}} = (X_1^T X_1)^{-1} X_1^T Y_1 + (X_1^T X_1)^{-1} X_1^T N_2 Y_2 (-\hat{D}\hat{V}_2^{-1})^T$, which corresponds to $\overline{\text{Vec}}(B_1)_{s,\text{2-stage}}$. We also present the values of trace($S(\bar{B}_{1,s})$) for contrast, where $\bar{B}_{1,s} = (X_1^T X_1)^{-1} X_1^T Y_1 + (X_1^T X_1)^{-1} X_1^T N_2 Y_2 (-DV_2^{-1})^T$ corresponds to (2.4).

In Table 1, based on different combinations of the correlation ρ and sample size, we present some numerical demonstrations to compare trace($S(\bar{B}_{1,s,\text{2-stage}})$) with trace($S(\hat{B}_1)$) and trace($S(\bar{B}_{1,s})$), which exhibit the performances of the simplified two-stage estimator $\overline{\text{Vec}}(B_1)_{s,\text{2-stage}}$ when the sample size is relatively small and moderate. Consequently, we find that the performance of the two-stage estimator tends to improve as the sample size increases. However, it also depends on the correlation ρ, and especially when $n \geq 20$ and $\rho \geq 0.5$, we easily see that $|\text{trace}(S(\bar{B}_{1,s,\text{2-stage}})) - \text{trace}(S(\bar{B}_{1,s}))| < |\text{trace}(S(\hat{B}_1)) - \text{trace}(S(\bar{B}_{1,s}))|$, which shows that the two-stage estimator $\overline{\text{Vec}}(B_1)_{s,\text{2-stage}}$ is closer to $\overline{\text{Vec}}(B_1)_s$.

(II) The case that there are no relationships between X_1 and X_2

In this case we assume that the system (1.1) has the same form as (3.10) but there are no relationships between X_1 and X_2. The simulations are presented below. From Table 2, we see that trace($S(\bar{B}_{1,s,\text{2-stage}})$) is getting closer to trace($S(\bar{B}_{1,s})$), which implies that the two-stage estimator $\overline{\text{Vec}}(B_1)_{s,\text{2-stage}}$ is becoming better than the LSE $\widehat{\text{Vec}}(B_1)$ as the sample size goes large ($n \geq 20$ or

Table 2: Comparisons between the two-stage estimator and the least square estimator

ρ	n	trace($S(\hat{B}_1)$)	trace($S(\bar{B}_{1,s,2\text{-stage}})$)	trace($S(\bar{B}_{1,s})$)
	10	16.4822	16.7363	16.5106
0.2	20	45.1567	45.3013	45.1656
	50	90.6848	90.7475	91.0296
	10	11.7975	12.4334	12.0509
0.5	20	30.6931	31.4717	31.2451
	50	91.1112	91.1782	91.2101
	10	15.0877	15.4275	15.2887
0.7	20	34.0316	34.9262	34.6478
	50	78.8424	79.1979	79.4065
	10	7.7416	8.8334	9.0277
0.9	20	37.1639	39.6293	38.6530
	50	106.1452	106.7684	106.7741

larger), also the fact depends on the value of the correlation ρ (≥ 0.5). This is because that from the viewpoint of covariance adjustment the one-step covariance adjustment estimator $\widehat{\text{Vec}}(B_1)^{(1)}$, which is exactly equal to $\overline{\text{Vec}}(B_1)_s$, is superior to $\widehat{\text{Vec}}(B_1)$ in the sense of having less covariance even though there are no relationships between X_1 and X_2. Hence, the simulation study discloses the tendency of $\overline{\text{Vec}}(B_1)_{s,2\text{-stage}}$ performing better, which is consistent with a two-stage estimator that incorporates more information.

4. An illustrating example

The GCM is a generalized multivariate analysis-of-variance model, which is useful especially for investigating growth problems on short time series in economics, biology and medical research (see Lee and Geisser 1972, Pan and Fang 2007). The seemingly unrelated GCMs are defined as

$$\begin{cases} Y_1 = X_1 B_1 Z_1 + E_1, \\ Y_2 = X_2 B_2 Z_2 + E_2, \end{cases} \tag{4.1}$$

where Y_i are $n \times q$ observation matrices, X_i and Z_i are known design matrices of full column rank and full row rank, respectively, and the regression parameters B_1 and B_2 are unknown. The assumptions on E_1 and E_2 are the same as those in the system (1.1).

Therefore, without considering the interactions between the two equations, we obtain the LSE of B_1 from the first equation as

$$\hat{B}_1 = \left(X_1^T X_1\right)^{-1} X_1^T Y_1 V_1^{-1} Z_1^T \left(Z_1 V_1^{-1} Z_1^T\right)^{-1}, \tag{4.2}$$

which is unbiased and the corresponding covariance $\text{Cov}(\hat{B}_1) = \text{Cov}(\text{Vec}(\hat{B}_1)) = (Z_1 V_1^{-1} Z_1^T)^{-1} \otimes (X_1^T X_1)^{-1}$. However, combining the information of the second equation and the assumption $X_1^T X_2 = 0$, we obtain the system LSE for B_1 as

$$\bar{B}_1 = \left(X_1^T X_1\right)^{-1} X_1^T \left(Y_1 V^{11} + Y_2 V^{21}\right) Z_1^T \left(Z_1 V^{11} Z_1^T\right)^{-1}, \tag{4.3}$$

which is unbiased and with less covariance

$$\text{Cov}\left(\bar{B}_1\right) = \text{Cov}\left(\text{Vec}\left(\bar{B}_1\right)\right) = \left(Z_1 V^{11} Z_1^T\right)^{-1} \otimes \left(X_1^T X_1\right)^{-1}, \tag{4.4}$$

which is less than $\text{Cov}(\hat{B}_1)$ since $V_1^{-1} \leq (V_1 - D V_2^{-1} D^T)^{-1} = V^{11}$ and correspondingly $(Z_1 V_1^{-1} Z_1^T)^{-1} \geq (Z_1 V^{11} Z_1^T)^{-1}$.

Table 3: Comparisons between several estimators under the matrix 2 norm

ρ	n	norm$(\hat{B}_1 - B_1)$	norm$(\bar{B}_{1,2\text{-stage}} - B_1)$	norm$(\bar{B}_1 - B_1)$
	10	0.1931	0.2668	0.1879
0.2	20	0.0183	0.0193	0.0182
	50	0.2259	0.2266	0.2205
	10	0.2013	0.2283	0.1679
0.5	20	0.0186	0.0177	0.0171
	50	0.2557	0.2271	0.2198
	10	0.1896	0.1872	0.1383
0.7	20	0.0184	0.0134	0.0127
	50	0.2532	0.1820	0.1763
	10	0.1923	0.1138	0.0847
0.9	20	0.0173	0.0085	0.0079
	50	0.2471	0.1133	0.1103

In the case that the covariance matrix V is unknown, under the assumption that $E = (E_1, E_2) \sim N_{n,2q}(0, I_n, V)$, we use the same form estimator as that of the equation (3.1) to estimate V, which is easily shown to be unbiased. Hence, a two-stage estimator for B_1 is defined as

$$\bar{B}_{1,2\text{-stage}} = \left(X_1^T X_1\right)^{-1} X_1^T \left(Y_1 \hat{V}^{11} + Y_2 \hat{V}^{21}\right) Z_1^T \left(Z_1 \hat{V}^{11} Z_1^T\right)^{-1}, \tag{4.5}$$

where $\hat{V}^{11} = (\hat{V}_1 - \hat{D}\hat{V}_2^{-1}\hat{D}^T)^{-1}$ and $\hat{V}^{21} = -\hat{V}_2^{-1}\hat{D}^T\hat{V}^{11}$. Analogous to the previous discussions, we can establish the unbiasedness of the estimator $\bar{B}_{1,2\text{-stage}}$.

In the following, we illustrate a simulation study to compare the performances of $\bar{B}_{1,2\text{-stage}}$ with those of $\hat{B}_1$ under the matrix 2-norm criterion, where the 2-norm of a matrix A is given by $||A||_2 = ||\text{Vec}(A)||_2 = (\sum_i \sum_j a_{ij}^2)^{1/2}$. The performances of $\bar{B}_1$ are also presented as a contrast. In each simulation, a sample of size n observations is randomly generated from a $2q$-variate normal distribution with mean zero and covariance matrix V, which is considered as the error matrix $E_{n\times 2q} = (E_1, E_2)$. Next, $\hat{B}_1$, $\bar{B}_{1,2\text{-stage}}$, and $\bar{B}_1$ are calculated in each simulation. Simulations are repeated 500 times and the matrix 2-norms of the average values of $\hat{B}_1 - B_1$, $\bar{B}_{1,2\text{-stage}} - B_1$, and $\bar{B}_1 - B_1$ are given in Table 3.

Three cases are studied. The first of them corresponds to $n = 10$, the second one considers the case of $n = 20$ and the third one corresponds to the case of $n = 50$. All cases adopt the same V as (3.10), but with the correlation ρ having a number of alternative values.

Simulations for the case (i) with

$$X_1^T = \begin{pmatrix} 616 & -302 & 1 & 9 & 23 & 19 & 25 & 26 & 17 & 11 \\ 614 & -302 & 27 & 15 & 20 & 26 & 6 & 24 & 25 & 8 \\ 37 & -28 & 1 & 2 & 5 & -4 & -2 & 9 & 1 & 4 \end{pmatrix},$$

$$X_2^T = \begin{pmatrix} 1 & 4 & 2 & 5 & 6 & 3 & 5 & 2 & 5 & 8 \\ 2 & 6 & 3 & 4 & 2 & 1 & 6 & 8 & 5 & 3 \end{pmatrix},$$

$$B_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 1 & 6 \\ -3 & 2 \end{pmatrix}, \quad Z_1 = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad Z_2 = \begin{pmatrix} 1 & 5 \\ 6 & 0.5 \end{pmatrix}.$$

Simulations for the case (ii) with X_1^T being

$$\begin{pmatrix} -948 & 15 & -2654 & 5 & 25 & 35 & 10 & 30 & 45 & 0 & 20 & 5 & 15 & 70 & 65 & 75 & 60 & 90 & 60 & 20 \\ -1069 & 30 & -3407 & 15 & 80 & 55 & 35 & 85 & 70 & 40 & 25 & 95 & 75 & 5 & 10 & 20 & 90 & 65 & 45 & 100 \\ -1766 & 60 & -2713 & 100 & 10 & 70 & 75 & 50 & 65 & 90 & 40 & 45 & 5 & 25 & 85 & 55 & 35 & 30 & 80 & 15 \end{pmatrix}$$

and

$$X_2^T = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 \\ 2 & 4 & 1 & 6 & 8 & 7 & 12 & 4 & 8 & 4 & 18 & 5 & 3 & 9 & 4 & 7 & 3 & 11 & 6 & 2 \end{pmatrix},$$

where B_1, B_2, Z_1, and Z_2 are the same as the case (i).

Simulations for the case (iii) with $X_1 = [a_1, a_2, a_3]_{50\times 3}$ being randomly generated and $X_2 = [a_4, a_5]_{50\times 2}$ being obtained from the null space of X_1^T, and in this case B_1, B_2, Z_1, and Z_2 remain the same as those of the case (i).

From Table 3, except the situations that $\rho = 0.2$ and $\rho = 0.5$ with $n = 10$, we find that norm$(\bar{B}_{1,2\text{-stage}} - B_1)$ is uniformly smaller than norm$(\hat{B}_1 - B_1)$, which shows that the two-stage estimator $\bar{B}_{1,2\text{-stage}}$ is closer to the true value B_1 than the LSE $\hat{B}_1$.

5. Concluding remarks

In summary, we have investigated regression coefficients estimation and inference for the system of two multivariate SURs. Note that we focus on the estimation problem of B_1 since the positions of B_1 and B_2 are equipotent. In Section 1, we find that together with another equation information the estimator of regression coefficients can be presented as a matrix power series via the method of covariance adjustment. In Section 2, we further indicate that the matrix series has exactly one simpler form which is just the one-step covariance adjustment estimator of the regression coefficients. In Section 3, in the case that the covariance matrix of the system is unknown, we illustrate that the degeneration form of the two-stage estimator sequence is unique, and an unbiased two-stage estimator is proposed and numerical simulations are also presented to verify its superiority. The results established in the present paper enrich the existing results since they include Zellner's univariate SURs as a special case.

Acknowledgements

Partly supported by NNSF of China (11371051).

References

Alkhamisi MA (2010). Simulation study of new estimators combining the SUR ridge regression and the restricted least squares methodologies, *Statistical Papers*, **51**, 651–672.

Baksalary JK (1991). Covariance adjustment in biased estimation, *Computational Statistics & Data Analysis*, **12**, 221–230.

Fang KT and Zhang YT (1990). *Generalized Multivariate Analysis*, Springer-Verlag and Science Press, Berlin and Beijing.

Gupta AK and Kabe DG (1998). A note on a result for two SUR models, *Statistical Papers*, **39**, 417–421.

Lee JC and Geisser S (1972). Growth curve prediction, *Sankhya A*, **34**, 393–412.

Liu AY (2002). Efficient estimation of two seemingly unrelated regression equations, *Journal of Multivariate Analysis*, **82**, 445–456.

Liu JS (2000). MSEM dominance of estimators in two seemingly unrelated regressions, *Journal of Statistical Planning and Inference*, **88**, 255–266.

Maddala GS (1977). *Econometrics*, McGraw-Hill, New York.

Muirhead RJ (1982). *Aspects of Multivariate Statistical Theory*, Wiley and Sons, New York.

Pan JX and Fang KT (2007). *Growth Curve Models and Statistical Diagnostics*, Science Press, Beijing.

Percy DF (1992). Prediction for seemingly unrelated regressions, *Journal of the Royal Statistical Society Series B (Methodological)*, **54**, 243–252.

Rao CR (1967). Least square theory using an estimated dispersion matrix and its application to measurement of signal. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, LeCam, L. M. & Neyman, J. (eds), **1**, 355–372.

Revankar NS (1974). Some finite sample results in the context of two seemingly unrelated regression equations, *Journal of the American Statistical Association*, **69**, 187–190.

Schmidt P (1977). Estimation of seemingly unrelated regressions with unequal numbers of observations, *Journal of Econometrics*, **5**, 365–377.

Shukur G and Zeebari Z (2012). Median regression for SUR models with the same explanatory variables in each equation, *Journal of Applied Statistics*, **39**, 1765–1779.

Srivastava VK and Giles DEA (1987). *Seemingly Unrelated Regression Equations Models*, Marcel Dekker, New York.

Velu R and Richards J (2008). Seemingly unrelated reduced-rank regression model, *Journal of Statistical Planning and Inference*, **138**, 2837–2846.

Wang SG (1989). A new estimate of regression coefficients in seemingly unrelated regression system, *Science in China Series A*, **32**, 808–816.

Zellner A (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association*, **57**, 348–368.

Zellner A (1963). Estimators of seemingly unrelated regression equations: some exact finite sample results, *Journal of the American Statistical Association*, **58**, 977–992.

Zhou B, Xu Q, and You J (2011). Efficient estimation for error component seemingly unrelated nonparametric regression models, *Metrika*, **73**, 121–138.

Received February 27, 2018; Revised August 24, 2018; Accepted September 21, 2018