Statistica Sinica 3(1993), 421-434

AN EFFICIENT ESTIMATION OF SEEMINGLY UNRELATED MULTIVARIATE REGRESSION MODELS WITH APPLICATION TO GROWTH CURVES ANALYSIS

Aiyi Liu

University of Science and Technology of China

Abstract: In this paper we propose an efficient method for estimating seemingly unrelated multivariate regression models. The gain in efficiency can be partially assessed by Hotelling's canonical correlations. We apply this method to the estimation problem and the concomitants selection problem in growth curve models.

Key words and phrases: Seemingly unrelated multivariate regression models, gain in efficiency, Wishart distribution, seemingly unrelated growth curve models, concomitants selection, two-stage estimate.

1. Introduction and Notation

1.1. Introduction

We study the following seemingly unrelated multivariate regression models (SUMR Model):

$$\mathbf{Y}_i = \mathbf{X}_i\mathbf{B}_i + \mathbf{E}_i \quad i = 1, 2 \tag{1.1}$$

where $\mathbf{X}_i$ $(n \times r_i)$ are possibly different matrices with rank r_i, $\mathbf{Y}_i$ and $\mathbf{E}_i$ are $(n \times p_i)$, $\mathbf{B}_i$ are $r_i \times p_i$ matrices of regression coefficients $(i = 1, 2)$, and if $\mathbf{E} = (\mathbf{E}_1, \mathbf{E}_2)$ $(n \times p; p = p_1 + p_2)$, then the rows of $\mathbf{E}$ are independently distributed with zero mean vectors and a common covariance matrix:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \tag{1.2}$$

Various authors have considered the estimation problems for model (1.1) by treating it as Zellner's seemingly unrelated regression equations (with p equations but only two different design matrices) (see Zellner (1962, 1963) and Srivastava et al. (1987)). In this paper, however, we regard model (1.1) as an SUMR model and purpose a new method to estimate the matrices of regression coefficients. It will be shown that this method is efficient whether Σ is known or unknown. As an application we first investigate the concomitants selection problem in growth curve models in order to make adjustment when the estimation method proposed by Potthoff and Roy (1964) is inefficient. Then the estimation problem for the so called seemingly unrelated growth curve models, which apparently have not yet been studied in the literature, is considered. We believe this model is very useful in biological science, for example in analyzing two results on different animals which may have some dependence on each other.

1.2. Notation and useful results

We list the following notation and some results useful in this paper.

1. Let $\mathbf{A}$, $\mathbf{D}$ be two matrices, $\mathrm{tr}(\mathbf{A}), r(\mathbf{A}), \mathbf{A}'$ and $\vec{\mathbf{A}}$ denote respectively the trace, rank, transpose and vectorized form (by stacking the columns under each other) of $\mathbf{A}$, and $\mathbf{A}\otimes\mathbf{D} = (a_{ij}\mathbf{D})$ denotes the Kronecker product of $\mathbf{A}$ and $\mathbf{D}$.

2. Suppose $\mathbf{W}$, $\mathbf{Z}$ are two random matrices. The covariance matrix of $\mathbf{W}$ is defined as $\mathrm{cov}(\mathbf{W}) = \mathrm{cov}(\vec{\mathbf{W}}')$. (Throughout this paper, the notation $\vec{\mathbf{W}}'$ denote an operation on $\mathbf{W}$ first by a transpose then by vectorized form.) We write $\mathbf{W}\sim \mathrm{N}(\mathrm{M}, \mathbf{A}\otimes\mathbf{D})$ if $\vec{\mathbf{W}}'$ has a multivariate normal distribution $\mathrm{N}(\vec{\mathrm{M}}', \mathbf{A}\otimes\mathbf{D})$. With the above notations, the following results are easy to verify:

(1) If

$$\mathrm{cov}\begin{pmatrix}\vec{\mathbf{W}}\\ \vec{\mathbf{Z}}\end{pmatrix} = \begin{pmatrix}\mathbf{V}_{11} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix},$$

then for any nonrandom matrices $\mathbf{A}_1, \mathbf{D}_1, \mathbf{A}_2, \mathbf{D}_2$,

$$\begin{aligned}&\mathrm{cov}(\mathbf{A}_1\mathbf{W}\mathbf{D}_1 + \mathbf{A}_2\mathbf{Z}\mathbf{D}_2)\\ =\ & (\mathbf{A}_1\otimes\mathbf{D}_1')\mathbf{V}_{11}(\mathbf{A}_1'\otimes\mathbf{D}_1) + (\mathbf{A}_1\otimes\mathbf{D}_1')\mathbf{V}_{12}(\mathbf{A}_2'\otimes\mathbf{D}_2)\\ &+ (\mathbf{A}_2\otimes\mathbf{D}_2')\mathbf{V}_{21}(\mathbf{A}_1'\otimes\mathbf{D}_1) + (\mathbf{A}_2\otimes\mathbf{D}_2')\mathbf{V}_{22}(\mathbf{A}_2'\otimes\mathbf{D}_2).\end{aligned} \tag{1.3}$$

(2) If $\mathbf{X}\sim \mathrm{N}(\mathbf{M}, \mathbf{C}\otimes\mathbf{D})$ and $\mathbf{A}$, $\mathbf{B}$ are two nonrandom matrices, then

$$\mathbf{E}(\mathbf{X}\mathbf{X}') = \mathbf{M}\mathbf{M}' + (\mathrm{tr}(\mathbf{D}))\mathbf{C}, \tag{1.4}$$

$$\mathbf{AXB}\sim \mathrm{N}(\mathbf{AMB}, \mathbf{ACA}'\otimes\mathbf{B}'\mathbf{DB}). \tag{1.5}$$

3. Suppose $\mathbf{A}\sim \mathbf{W}_m(k, \mathbf{V})$, the Wishart distribution with k degrees of freedom, and $\mathbf{A}$ and $\mathbf{V}$ are partitioned as $\mathbf{A} = \begin{pmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{pmatrix}$; $\mathbf{V} = \begin{pmatrix}\mathbf{V}_{11} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix}$, where $\mathbf{A}_{11}$ and $\mathbf{V}_{11}$ are $k_1\times k_1$. Then

(1)

$$\mathbf{E}(\mathbf{A}^{-1}) = \frac{1}{k-m-1}\mathbf{V}^{-1}. \tag{1.6}$$

(2)
$$\mathbf{A}_{11} \sim \mathbf{W}_{k_1}(k, \mathbf{V}_{11}), \quad \mathbf{A}_{22} \sim \mathbf{W}_{m-k_1}(k, \mathbf{V}_{22}). \tag{1.7}$$

(3) Given $\mathbf{A}_{22}$, the conditional distribution of $\mathbf{A}_{12}$ is

$$\mathbf{A}_{12}|\mathbf{A}_{22} \sim \mathbf{N}(\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{A}_{22}, \mathbf{V}_{11.2} \otimes \mathbf{A}_{22}) \tag{1.8}$$

where $\mathbf{V}_{11.2} = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}$.

4. **Definition.** Let $\mathbf{T}_1$ and $\mathbf{T}_2$ be two estimators of a parameter matrix Θ. We say that $\mathbf{T}_1$ is superior to $\mathbf{T}_2$ if $\text{MSE}(\mathbf{T}_1) \leq \text{MSE}(\mathbf{T}_2)$, where $\text{MSE}(\mathbf{T}_i) = \mathbf{E}((\vec{\mathbf{T}}_i' - \vec{\Theta}')'(\vec{\mathbf{T}}_i' - \vec{\Theta}'))$ stands for the generalized mean square error of the estimator $\vec{\mathbf{T}}_i'$ of $\vec{\Theta}'$.

It is obvious that if $\mathbf{T}_1$ and $\mathbf{T}_2$ are two unbiased estimators, then $\text{cov}(\mathbf{T}_1) \leq \text{cov}(\mathbf{T}_2)$ (i.e. $\text{cov}(\vec{\mathbf{T}}_1') \leq \text{cov}(\vec{\mathbf{T}}_2')$) implies that $\text{MSE}(\mathbf{T}_1) \leq \text{MSE}(\mathbf{T}_2)$, that is $\mathbf{T}_1$ is superior to $\mathbf{T}_2$. For two positive definite matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$, the ordering $\mathbf{Q}_1 \leq \mathbf{Q}_2$ here means that $\mathbf{Q}_2 - \mathbf{Q}_1$ is non-negative definite.

2. Estimation with Known Σ

For simplicity we begin our discussion of the method of estimation with known Σ. In the next section the more complicated case when Σ is unknown will be investigated. Assume that $\Sigma_{12} \neq 0$.

Lemma 1. *Let* $\mathbf{T}_1$ *be an unbiased estimator of a parameter matrix* Θ, *and, suppose that* $\mathbf{T}_2$ *is a random matrix with zero mean matrix and* $\text{cov}\begin{pmatrix} \vec{\mathbf{T}}_1' \\ \vec{\mathbf{T}}_2' \end{pmatrix} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}$. *Then among the class of all linear unbiased estimators*

$$\mathbf{T}^*(\mathbf{X}_1, \mathbf{Z}_1, \mathbf{X}_2, \mathbf{Z}_2) = \mathbf{X}_1\mathbf{T}_1\mathbf{Z}_1 + \mathbf{X}_2\mathbf{T}_2\mathbf{Z}_2$$

the BLUE *is*

$$\mathbf{T}^*(\mathbf{X}_0, \mathbf{Z}_0) = \mathbf{T}_1 + \mathbf{X}_0\mathbf{T}_2\mathbf{Z}_0 \tag{2.1}$$

where $\mathbf{X}_0$ *and* $\mathbf{Z}_0$ *satisfy*

$$\mathbf{X}_0 \otimes \mathbf{Z}_0' = -\mathbf{V}_{12}\mathbf{V}_{22}^{-1}, \tag{2.2}$$

$$\text{cov}\Big(\mathbf{T}^*(\mathbf{X}_0, \mathbf{Z}_0)\Big) = \mathbf{V}_{11.2} = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}. \tag{2.3}$$

The proof of Lemma 1 follows by using the Gauss-Markov theorem in the following linear model:

$$\begin{pmatrix} \vec{\mathbf{T}}_1' \\ \vec{\mathbf{T}}_2' \end{pmatrix} = \begin{pmatrix} \mathbf{I} \\ 0 \end{pmatrix} \vec{\Theta}' + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$$

where

$$\mathbf{E}\begin{pmatrix}\epsilon_1\\ \epsilon_2\end{pmatrix}=0 \quad \text{cov}\begin{pmatrix}\epsilon_1\\ \epsilon_2\end{pmatrix}=\begin{pmatrix}\mathbf{V}_{11} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix}.$$

The BLUE of Θ is

$$\vec{\Theta}^{*\prime}=\vec{\mathbf{T}}_1'-\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\vec{\mathbf{T}}_2'. \tag{2.4}$$

If $\mathbf{X}_0$ and $\mathbf{Z}_0$ satisfy (2.2), then (2.4) is equivalent to

$$\Theta^*=\mathbf{T}_1-\mathbf{X}_0\mathbf{T}_2\mathbf{Z}_0.$$

The proof is complete.

Lemma 1 is an extension of Rao's covariance adjustment theory (see Rao (1967)). Wang (1989) investigated its application and proposed a new method of estimating the regression coefficients in the seemingly unrelated regression equations.

We now consider the SUMR model (1.1) with known Σ. As is well known, the BLUE of $\mathbf{B}_1$ obtained from the first equation of (1.1) is $\mathbf{B}_1=(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}_1$, an inefficient estimator because it does not consider the correlation between the two equations. Denote $\mathbf{T}=\mathbf{T}_2'\mathbf{Y}_2$, where $\mathbf{T}_2$ is any matrix with full-column rank such that $\mathbf{X}_2'\mathbf{T}_2=0$. Observe that $\mathbf{E}(\mathbf{T})=0$, and

$$\text{cov}\begin{pmatrix}\vec{\mathbf{B}}_1'\\ \vec{\mathbf{T}}'\end{pmatrix}=\begin{pmatrix}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\otimes\Sigma_{11} & (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{T}_2\otimes\Sigma_{12}\\ \mathbf{T}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\otimes\Sigma_{21} & (\mathbf{T}_2'\mathbf{T}_2)\otimes\Sigma_{22}\end{pmatrix}, \tag{2.5}$$

where (2.5) is obtained by straightforward manipulations utilizing

$$\text{cov}\begin{pmatrix}\vec{\mathbf{Y}}_1'\\ \vec{\mathbf{Y}}_2'\end{pmatrix}=\begin{pmatrix}\mathbf{I}\otimes\Sigma_{11} & \mathbf{I}\otimes\Sigma_{12}\\ \mathbf{I}\otimes\Sigma_{21} & \mathbf{I}\otimes\Sigma_{22}\end{pmatrix}. \tag{2.6}$$

According to Lemma 1, note that the BULE of $\mathbf{B}_1$ among the class of linear unbiased estimators having the form

$$\hat{\mathbf{B}}_1(\mathbf{T}_2,\mathbf{X}_1,\mathbf{Z}_1,\mathbf{X}_2,\mathbf{Z}_2,)=\mathbf{X}_1\hat{\mathbf{B}}_1\mathbf{Z}_1+\mathbf{X}_2\mathbf{T}\mathbf{Z}_2$$

is

$$\hat{\mathbf{B}}_1(\mathbf{T}_2)=\hat{\mathbf{B}}_1+\mathbf{X}_0\mathbf{T}\mathbf{Z}_0$$

where $\mathbf{X}_0$ and $\mathbf{Z}_0$ satisfy

$$\mathbf{X}_0\otimes\mathbf{Z}_0'=-(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{T}_2(\mathbf{T}_2'\mathbf{T}_2)^{-1}\otimes\Sigma_{12}\Sigma_{22}^{-1};$$

that is,

$$\mathbf{X}_0=-(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{T}_2(\mathbf{T}_2'\mathbf{T}_2)^{-1},$$

$$\mathbf{Z}_0 = \Sigma_{22}^{-1}\Sigma_{21}.$$

Therefore,

$$\hat{\mathbf{B}}_1(\mathbf{T}_2) = \hat{\mathbf{B}}_1 - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{P}_{\mathbf{T}_2}\mathbf{Y}_2\Sigma_{22}^{-1}\Sigma_{21} \tag{2.7}$$

where $\mathbf{P}_{\mathbf{T}_2} = \mathbf{T}_2(\mathbf{T}_2'\mathbf{T}_2)^{-1}\mathbf{T}_2'$ and

$$\begin{aligned}\mathrm{cov}(\hat{\mathbf{B}}_1(\mathbf{T}_2)) &= (\mathbf{X}_1'\mathbf{X}_1)^{-1} \otimes \Sigma_{11} \\ &- (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{P}_{\mathbf{T}_2}\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \otimes \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.\end{aligned} \tag{2.8}$$

Obviously the best choice of $\mathbf{P}_{\mathbf{T}_2}$ is $\mathbf{N}_2 = \mathbf{I} - \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'$; this is so because for any matrix $\mathbf{T}_2$ such that $\mathbf{X}_2'\mathbf{T}_2 = 0$, $\mu(\mathbf{T}_2) \subset \mu(\mathbf{N}_2)$, where $\mu(\cdot)$ denotes the column space of a matrix. Therefore, $\mathbf{P}_{\mathbf{T}_2} \leq \mathbf{N}_2$. We have obtained an efficient estimator for $\mathbf{B}_1$:

$$\tilde{\mathbf{B}}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}_1 - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{N}_2\mathbf{Y}_2\Sigma_{22}^{-1}\Sigma_{21}. \tag{2.9}$$

This estimator reduces to Zellner's estimator when $p_1 = p_2 = 1$ (see Zellner (1962)).

Theorem 1. *For the estimator* $\tilde{\mathbf{B}}_1$, *we have*

1.
$$\mathbf{E}\tilde{\mathbf{B}}_1 = \mathbf{B}_1,$$

2.
$$\begin{aligned}\mathrm{cov}(\tilde{\mathbf{B}}_1) &= (\mathbf{X}_1'\mathbf{X}_1)^{-1} \otimes \Sigma_{11} - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{N}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \otimes \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\ &\leq \mathrm{cov}(\hat{\mathbf{B}}_1);\end{aligned}$$

that is, $\tilde{\mathbf{B}}_1$ *is superior to* $\hat{\mathbf{B}}_1$, *the* LSE *of* $\mathbf{B}_1$.

It is practical to assess how much is gained by using our efficient estimator in comparison with the less efficient least squares estimator. To do so, first define two kinds of measure to assess quantitatively the gain in efficiency.

Definition. Let $\mathbf{T}_1$ and $\mathbf{T}_2$ be two unbiased estimators of a parameter matrix Θ. Suppose $\mathbf{T}_2$ is superior to $\mathbf{T}_1$, then the gain in efficiency in using $\mathbf{T}_2$ in comparison with $\mathbf{T}_1$ is defined as

$$e_1(\mathbf{T}_2 \,|\, \mathbf{T}_1) = |\,\mathrm{cov}(\mathbf{T}_1) - \mathrm{cov}(\mathbf{T}_2)\,|_0, \tag{2.10}$$

where $|\cdot|_0$ is defined as the product of the non-zero eigenvalues of a matrix.

The second measure is

$$e_2(\mathbf{T}_2 \,|\, \mathbf{T}_1) = \mathrm{tr}(\mathrm{cov}(\mathbf{T}_1)) - \mathrm{tr}(\mathrm{cov}(\mathbf{T}_2)). \tag{2.11}$$

The definitions here seem natural and meaningful. In fact, since $\vec{\mathbf{T}}'_1$ is unbiased for $\vec{\Theta}'$, $|\text{cov}(\mathbf{T}_1)|_0$ and $\text{tr}(\text{cov}(\mathbf{T}_1))$ are respectively the generalized variance and the mean square error of $\vec{\mathbf{T}}'_1$. In general, the larger the e_is are, the more gain in efficiency by using $\mathbf{T}_2$. Now consider the estimator (2.9).

The expression of $\text{cov}(\tilde{\mathbf{B}}_1)$ in Theorem 1 can be rearranged as follows:

$$\begin{aligned}\text{cov}(\tilde{\mathbf{B}}_1) &= \left(\mathbf{I}\otimes\Sigma_{11}^{\frac{1}{2}}\right)\Big((\mathbf{X}_1'\mathbf{X}_1)^{-1}\otimes\mathbf{I}-(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{N}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\\&\quad\otimes\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\Big)\left(\mathbf{I}\otimes\Sigma_{11}^{\frac{1}{2}}\right).\end{aligned} \tag{2.12}$$

The expression (2.12) seems more meaningful than that in Theorem 1. Let $\rho_1^2\geq\cdots\geq\rho_k^2$ be the k non-zero eigenvalues of the matrix

$$\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}.$$

It is well known that their positive square roots $\rho_1\geq\cdots\geq\rho_k>0$ are Hotelling's population canonical correlation coefficients (see Muirhead (1982)) arising from the matrix Σ defined in (1.2). Let $\delta_1^2\geq\cdots\geq\delta_m^2$ be the m non-zero eigenvalues of the matrix

$$(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{N}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}.$$

Then, from the definitions (2.10) and (2.11), it follows that

$$e_1(\tilde{\mathbf{B}}_1|\hat{\mathbf{B}}_1)=\left(\prod_{i=1}^{m}\delta_i^2\right)\left(\prod_{i=1}^{k}\rho_i^2\right)|\Sigma_{11}|, \tag{2.13}$$

where $|\cdot|$ stands for the determinant of a matrix.

It is clear from (2.13) that the gain in efficiency using $\tilde{\mathbf{B}}_1$ depends on three parts, of which only the third part (i.e. $|\Sigma_{11}|$) depends solely on the response matrix $\mathbf{Y}_1$. The first part (i.e. $\prod_{i=1}^m\delta_i^2$) reflects the dependencies between the two design matrices $\mathbf{X}_1$ and $\mathbf{X}_2$. In fact, if the two matrices are least independent (this is the case when $\mathbf{X}_1=\mathbf{X}_2$ and hence $\prod_{i=1}^m\delta_i^2=0$ and $\tilde{\mathbf{B}}_1=\hat{\mathbf{B}}_1$), there is no gain in using $\tilde{\mathbf{B}}_1$ in comparison with $\hat{\mathbf{B}}_1$. On the other hand, if the two matrices are extremely independent (this is the case when the two spaces generated respectively by the columns of each matrix are orthogonal, that is $\mathbf{X}_1'\mathbf{X}_2=0$), this yields the maximum gain in using $\mathbf{T}_2$ for a given Σ (see Theorem 2 of this section).

The second part in (2.13) (i.e. $\prod_{i=1}^k\rho_i^2$) has a well known statistical meaning. It is Hotelling's population correlation coefficient (see Hotelling (1936)), which measures the dependencies of the two seemingly unrelated response matrices $\mathbf{Y}_1$ and $\mathbf{Y}_2$. For the given matrices $\mathbf{X}_1$ and $\mathbf{X}_2$, a larger correlation coefficient brings more gain in efficiency. On the other hand, if the canonical correlations are very weak, in which case Hotelling's correlation coefficient can be very small, the gain in efficiency may be little.

Similar statements can be drawn for the second measure of gain in efficiency defined in (2.11). We may see this from the bounds of the gain obtained below. From the definition,

$$e_2(\tilde{\mathbf{B}}_1|\hat{\mathbf{B}}_1) = \Big(\sum_{i=1}^{m} \delta_i^2\Big)\mathrm{tr}(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}).$$

Now use the fact that for any two non-negative definite matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$,

$$\underline{\lambda}(\mathbf{Q}_1)\mathrm{tr}(\mathbf{Q}_2) \le \mathrm{tr}(\mathbf{Q}_1\mathbf{Q}_2) \le \bar{\lambda}(\mathbf{Q}_1)\mathrm{tr}(\mathbf{Q}_2), \tag{2.14}$$

where $\underline{\lambda}(\cdot)$ and $\bar{\lambda}(\cdot)$ stand respectively for the smallest and largest eigenvalues of a matrix. It follows that

$$\underline{\lambda}(\Sigma_{11})\Big(\sum_{i=1}^{m} \delta_i^2\Big)\Big(\sum_{i=1}^{k} \rho_i^2\Big) \le e_2(\tilde{\mathbf{B}}_1|\hat{\mathbf{B}}_1) \le \bar{\lambda}(\Sigma_{11})\Big(\sum_{i=1}^{m} \delta_i^2\Big)\Big(\sum_{i=1}^{k} \rho_i^2\Big). \tag{2.15}$$

In (2.15) the two terms $\sum_{i=1}^{m} \delta_i^2$ and $\sum_{i=1}^{k} \rho_i^2$ also reflect respectively the dependencies between the two matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ and the dependencies between the two response matrices $\mathbf{Y}_1$ and $\mathbf{Y}_2$.

In practice Σ is often unknown. In that case we may substitute the sample canonical correlation coefficients derived from an estimator of Σ (see Section 3 of this paper) for ρ_i in (2.13) and (2.15) and evaluate roughly the gain in efficiency as consistent estimates. This is a guide for determining if the estimator $\tilde{\mathbf{B}}_1$ is worthwhile.

The following theorem gives a necessary and sufficient condition for $\tilde{\mathbf{B}}_1$ to be the BLUE of $\mathbf{B}_1$ in the SUMR model (1.1).

Theorem 2. *$\tilde{\mathbf{B}}_1$ is the* BLUE *of $\mathbf{B}_1$ in model* (1.1) *if and only if $\mathbf{P}_1\mathbf{P}_2 = \mathbf{P}_2\mathbf{P}_1$, where $\mathbf{P}_i = \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i' \quad (i = 1, 2)$.*

Proof. Model (1.1) can be rewritten as

$$\begin{pmatrix} \vec{\mathbf{Y}}_1 \\ \vec{\mathbf{Y}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I}\otimes\mathbf{X}_1 & 0 \\ 0 & \mathbf{I}\otimes\mathbf{X}_2 \end{pmatrix}\begin{pmatrix} \vec{\mathbf{B}}_1 \\ \vec{\mathbf{B}}_2 \end{pmatrix} + \begin{pmatrix} \vec{\mathbf{E}}_1 \\ \vec{\mathbf{E}}_2 \end{pmatrix}. \tag{2.16}$$

Note that $\mathrm{cov}\begin{pmatrix} \vec{\mathbf{E}}_1 \\ \vec{\mathbf{E}}_2 \end{pmatrix} = \begin{pmatrix} \Sigma_{11}\otimes\mathbf{I} & \Sigma_{12}\otimes\mathbf{I} \\ \Sigma_{21}\otimes\mathbf{I} & \Sigma_{22}\otimes\mathbf{I} \end{pmatrix}$ and $\mathbf{E}\begin{pmatrix} \vec{\mathbf{E}}_1 \\ \vec{\mathbf{E}}_2 \end{pmatrix} = 0$. Denoting $\mathbf{B}_1^*$ and $\mathbf{B}_2^*$ as the BLUE of $\mathbf{B}_1$ and $\mathbf{B}_2$, respectively, $\mathbf{B}_1^*$ and $\mathbf{B}_2^*$ satisfy the following equation

$$\begin{pmatrix} \Sigma^{11} \otimes \mathbf{X}_1'\mathbf{X}_1 & \Sigma^{12} \otimes \mathbf{X}_1'\mathbf{X}_2 \\ \Sigma^{21} \otimes \mathbf{X}_2'\mathbf{X}_1 & \Sigma^{22} \otimes \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \vec{\mathbf{B}}_1^* \\ \vec{\mathbf{B}}_2^* \end{pmatrix} = \begin{pmatrix} \Sigma^{11} \otimes \mathbf{X}_1' & \Sigma^{12} \otimes \mathbf{X}_1' \\ \Sigma^{21} \otimes \mathbf{X}_2' & \Sigma^{22} \otimes \mathbf{X}_2' \end{pmatrix} \begin{pmatrix} \vec{\mathbf{Y}}_1 \\ \vec{\mathbf{Y}}_2 \end{pmatrix} \tag{2.17}$$

where $\begin{pmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix} = \Sigma^{-1}$.

Left-multiplying both sides of (2.17) by

$$\begin{pmatrix} \mathbf{I} & -(\Sigma^{12} \otimes \mathbf{X}_1'\mathbf{X}_2)(\Sigma^{22} \otimes \mathbf{X}_2'\mathbf{X}_2)^{-1} \\ 0 & \mathbf{I} \end{pmatrix},$$

it is easy to show as a matter of straightforward calculation that $\mathbf{B}_1^*$ satisfies the following equation

$$\begin{aligned} &\left[\Sigma^{11} \otimes \mathbf{X}_1'\mathbf{X}_1 - \Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21} \otimes \mathbf{X}_1'\mathbf{P}_2\mathbf{X}_1\right]\vec{\mathbf{B}}_1^* \\ = &\ (\Sigma^{11} \otimes \mathbf{X}_1')\vec{\mathbf{Y}}_1 + (\Sigma^{12} \otimes \mathbf{X}_1'\mathbf{N}_2)\vec{\mathbf{Y}}_2 - \left[\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21} \otimes \mathbf{X}_1'\mathbf{P}_2\right]\vec{\mathbf{Y}}_1. \end{aligned} \tag{2.18}$$

After careful study of (2.18) we consider the following equation

$$(\Sigma^{11} \otimes \mathbf{X}_1'\mathbf{X}_1)\vec{\underline{\mathbf{B}}}_1^* = (\Sigma^{11} \otimes \mathbf{X}_1')\vec{\mathbf{Y}}_1 + (\Sigma^{12} \otimes \mathbf{X}_1'\mathbf{N}_2)\vec{\mathbf{Y}}_2. \tag{2.19}$$

The solution of $\vec{\underline{\mathbf{B}}}_1^*$ is then

$$\vec{\underline{\mathbf{B}}}_1^* = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1\mathbf{N}_2\mathbf{Y}_2\Sigma^{21}(\Sigma^{11})^{-1}. \tag{2.20}$$

Utilizing the result of the inverse of a partitioned matrix, we may verify that $\Sigma^{21}(\Sigma^{11})^{-1} = -\Sigma_{12}(\Sigma_{22})^{-1}$; hence $\underline{\mathbf{B}}_1^*$ is identical to the estimator $\tilde{\mathbf{B}}_1$; therefore $\tilde{\mathbf{B}}_1 = \mathbf{B}_1^*$ if and only if

$$\left(\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21} \otimes \mathbf{X}_1'\mathbf{P}_2\mathbf{X}_1\right)\vec{\mathbf{B}}_1^* = \left[\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21} \otimes \mathbf{X}_1'\mathbf{P}_2\right]\vec{\mathbf{Y}}_1. \tag{2.21}$$

Substituting the right side of (2.20) for $\vec{\mathbf{B}}_1^*$ we obtain

$$\begin{aligned} &\left[\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21} \otimes (\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1 - \mathbf{X}_1'\mathbf{P}_2)\right]\vec{\mathbf{Y}}_1 \\ = &\left[\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21}\Sigma_{12}\Sigma_{22}^{-1} \otimes \mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2\right]\vec{\mathbf{Y}}_2. \end{aligned} \tag{2.22}$$

Note that $\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1 = \mathbf{X}_1'\mathbf{P}_2$ implies $\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2 = \mathbf{X}_1'\mathbf{P}_2\mathbf{N}_2 = 0$; thus, (2.22) holds if and only if

$$\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1 = \mathbf{X}_1'\mathbf{P}_2. \tag{2.23}$$

It is easy to show that (2.23) is equivalent to $\mathbf{P}_2\mathbf{P}_1 = \mathbf{P}_1\mathbf{P}_2$, and the proof follows.

It can be concluded directly from this theorem that for a given Σ our estimator $\tilde{\mathbf{B}}_1$ gets its maximum gain in efficiency in comparison with the LSE $\hat{\mathbf{B}}_1$ if and only if $\mathbf{P}_2\mathbf{P}_1 = \mathbf{P}_1\mathbf{P}_2$.

3. Two Stage Estimation when Σ Is Unknown

When Σ is unknown, the estimator $\tilde{\mathbf{B}}_1$ is not feasible since it contains the unknown Σ. In that case we replace Σ by an observable matrix $\frac{1}{n-r(\mathbf{X})}\mathbf{S}$ and obtain the so called two stage estimator of $\mathbf{B}_1$, namely,

$$\tilde{\mathbf{B}}_1(\mathbf{S}) = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}_1 - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{N}_2\mathbf{Y}_2\mathbf{S}_{22}^{-1}\mathbf{S}_{21} \tag{3.1}$$

where $\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix} = \mathbf{Y}'\mathbf{N}\mathbf{Y}$, $\mathbf{Y} = (Y_1, Y_2)$, $\mathbf{N} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$, $\mathbf{X} = (X_1, X_2)$ and $\mathbf{A}^-$ denotes the generalized inverse of $\mathbf{A}$. Denote by $\mathbf{B}_1^*(\mathbf{S})$ the solution of the equation (2.17) with Σ replaced by $\mathbf{S}$. Similar to Theorem 2 we may show that $\tilde{\mathbf{B}}_1(\mathbf{S}) = \mathbf{B}_1^*(\mathbf{S})$ if and only if $\mathbf{P}_1\mathbf{P}_2 = \mathbf{P}_2\mathbf{P}_1$. We focus our attention on investigating the finite sample property of $\tilde{\mathbf{B}}_1(\mathbf{S})$. In order to do so assume further that $\mathbf{E} = (E_1, E_2) \sim \mathbf{N}(\mathbf{O}, \mathbf{I} \otimes \Sigma)$. Under this assumption it is easy to show that $\mathbf{S} = \mathbf{Y}'\mathbf{N}\mathbf{Y} \sim \mathbf{W}_p(n - r(X), \Sigma)$ and $\mathbf{S}$ is independent of $\mathbf{X}_1'\mathbf{Y}_1$ and $\mathbf{X}_1'\mathbf{N}_2\mathbf{Y}_2$. Moreover, since $\mathbf{E}(\mathbf{X}_1'\mathbf{N}_2\mathbf{Y}_2) = 0$, and $\mathbf{E}(\mathbf{S}_{22}^{-1}\mathbf{S}_{12}) = \Sigma_{22}^{-1}\Sigma_{21}$ (see (3.3) below), it follows that

$$\begin{aligned} \mathbf{E}(\tilde{\mathbf{B}}_1(\mathbf{S})) &= \mathbf{E}((\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}_1 - \mathbf{E}((\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{N}_2\mathbf{Y}_2\mathbf{S}_{22}^{-1}\mathbf{S}_{21}) \\ &= \mathbf{B}_1 - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{E}(\mathbf{X}_1'\mathbf{N}_2\mathbf{Y}_2)\mathbf{E}(\mathbf{S}_{22}^{-1}\mathbf{S}_{21}) = \mathbf{B}_1; \end{aligned}$$

that is, $\tilde{\mathbf{B}}_1(\mathbf{S})$ is an unbiased estimator of $\mathbf{B}_1$. We now establish the following theorem.

Theorem 3. *The two stage estimator* $\tilde{\mathbf{B}}_1(\mathbf{S})$ *is superior to* LSE $\hat{\mathbf{B}}_1$ *if* n *is sufficiently large.*

Proof. By straightforward calculation it can be shown that the conditional covariance matrix of $\tilde{\mathbf{B}}_1(\mathbf{S})$ given $\mathbf{S}$, is

$$\begin{aligned} \mathrm{cov}(\tilde{\mathbf{B}}_1(\mathbf{S})|\mathbf{S}) &= (\mathbf{X}_1'\mathbf{X}_1)^{-1} \otimes \Sigma_{11} - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{N}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \\ &\quad \otimes (\Sigma_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21} + \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\Sigma_{21} - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\Sigma_{22}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}). \end{aligned}$$

Hence, the covariance matrix of $\tilde{\mathbf{B}}_1(\mathbf{S})$ is

$$\begin{aligned} \mathrm{cov}(\tilde{\mathbf{B}}_1(\mathbf{S})) &= (\mathbf{X}_1'\mathbf{X}_1)^{-1} \otimes \Sigma_{11} - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{N}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \\ &\quad \otimes \Big[\Sigma_{12}\mathbf{E}(\mathbf{S}_{22}^{-1}\mathbf{S}_{21}) + \mathbf{E}(\mathbf{S}_{12}\mathbf{S}_{22}^{-1})\Sigma_{21} - \mathbf{E}(\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\Sigma_{22}\mathbf{S}_{22}^{-1}\mathbf{S}_{21})\Big]. \end{aligned} \tag{3.2}$$

Note that $\mathbf{S}_{22} \sim \mathbf{W}_{p_2}(n - r(X), \Sigma_{22})$ and given $\mathbf{S}_{22}$ the conditional distributions of $\mathbf{S}_{12}, \mathbf{S}_{12}\mathbf{S}_{22}^{-1}$ and $\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\Sigma_{22}^{\frac{1}{2}}$ are $\mathbf{N}(\Sigma_{12}\Sigma_{22}^{-1}\mathbf{S}_{22}, \Sigma_{11.2} \otimes \mathbf{S}_{22})$, $\mathbf{N}(\Sigma_{12}\Sigma_{22}^{-1}, \Sigma_{11.2} \otimes \mathbf{S}_{22}^{-1})$ and $\mathbf{N}(\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}, \Sigma_{11.2} \otimes \Sigma_{22}^{\frac{1}{2}}\mathbf{S}_{22}^{-1}\Sigma_{22}^{\frac{1}{2}})$, respectively, where $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Using the results listed in Section 1.2 we get

$$\mathbf{E}(\mathbf{S}_{12}\mathbf{S}_{22}^{-1}) = \mathbf{E}(\mathbf{E}(\mathbf{S}_{12}\mathbf{S}_{22}^{-1}|\mathbf{S}_{22})) = \Sigma_{12}\Sigma_{22}^{-1}, \tag{3.3}$$

$$\begin{aligned} & \mathbf{E}(\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\Sigma_{22}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}) \\ = \ & \mathbf{E}\Big(\mathbf{E}(\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\Sigma_{22}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}|\mathbf{S}_{22})\Big) \\ = \ & \mathbf{E}\Big(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} + \mathrm{tr}(\Sigma_{22}\mathbf{S}_{22}^{-1})\Sigma_{11.2}\Big) \\ = \ & \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} + (\mathrm{tr}(\Sigma_{22}\mathbf{E}\mathbf{S}_{22}^{-1}))\Sigma_{11.2} \\ = \ & \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} + \frac{p_2}{n - r(X) - p_2 - 1}\Sigma_{11.2}. \end{aligned} \tag{3.4}$$

Hence

$$\begin{aligned} \mathrm{cov}(\tilde{\mathbf{B}}_1(\mathbf{S})) &= (\mathbf{X}_1'\mathbf{X}_1)^{-1} \otimes \Sigma_{11} - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{N}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \\ & \quad \otimes \Big(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \frac{p_2}{n - r(X) - p_2 - 1}\Sigma_{11.2}\Big), \end{aligned} \tag{3.5}$$

which implies that,

$$\begin{aligned} \mathrm{MSE}(\tilde{\mathbf{B}}_1(\mathbf{S})) &= \mathrm{tr}((\mathbf{X}_1'\mathbf{X}_1)^{-1})\mathrm{tr}(\Sigma_{11}) - \mathrm{tr}\Big((\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{N}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\Big) \\ & \quad \Big[\mathrm{tr}(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) - \frac{p_2}{n - r(X) - p_2 - 1}\mathrm{tr}(\Sigma_{11.2})\Big]. \end{aligned} \tag{3.6}$$

Obviously $\mathrm{MSE}(\tilde{\mathbf{B}}_1(\mathbf{S})) \leq \mathrm{tr}((\mathbf{X}_1'\mathbf{X}_1)^{-1})\mathrm{tr}(\Sigma_{11}) = \mathrm{MSE}(\hat{\mathbf{B}}_1)$ if n is sufficiently large, and the proof is complete. The method for estimating $\mathbf{B}_1$ may also be used to estimate $\mathbf{B}_2$.

In order to draw some more practical conclusions, note that $\tilde{\mathbf{B}}_1(\mathbf{S})$ is superior to LSE $\hat{\mathbf{B}}_1$ if and only if,

$$\mathrm{tr}(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) - \frac{p_2}{n - r(X) - p_2 - 1}\mathrm{tr}(\Sigma_{11.2}) > 0,$$

which is equivalent to,

$$n - r(\mathbf{X}) - 1 > \frac{p_2\mathrm{tr}(\Sigma_{11})}{\mathrm{tr}(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})}. \tag{3.7}$$

Using (2.14) we observe that,

$$\frac{\operatorname{tr}(\Sigma_{11})}{\operatorname{tr}(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})} = \frac{\operatorname{tr}(\Sigma_{11})}{\operatorname{tr}\Big[\Big(\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\Big)(\Sigma_{11})\Big]}$$

$$\leq \frac{p_1\bar{\lambda}(\Sigma_{11})}{\underline{\lambda}(\Sigma_{11})\operatorname{tr}\Big(\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\Big)} = \frac{\bar{\lambda}(\Sigma_{11})}{\underline{\lambda}(\Sigma_{11})}\frac{p_1}{(\sum_{i=1}^k \rho_i^2)}$$

$$= \frac{\bar{\lambda}(\Sigma_{11})}{\underline{\lambda}(\Sigma_{11})}\frac{k}{(\sum_{i=1}^k \rho_i^2)}\frac{p_1}{k} \leq \frac{\bar{\lambda}(\Sigma_{11})}{\underline{\lambda}(\Sigma_{11})}\frac{p_1}{\sqrt[k]{\prod_{i=1}^k \rho_i^2}}.$$

The last inequality holds because $\frac{\sum_{i=1}^k \rho_i^2}{k} \geq \sqrt[k]{\prod_{i=1}^k \rho_i^2}$, and the ρ_is are the same as those in Section 2, therefore (3.7) holds if

$$n - r(\mathbf{X}) - 1 > p_1 p_2 \frac{\bar{\lambda}(\Sigma_{11})}{\underline{\lambda}(\Sigma_{11})}\frac{1}{\sqrt[k]{\prod_{i=1}^k \rho_i^2}}. \tag{3.8}$$

We therefore obtain the following conclusion:

Corollary. *The two stage estimator* $\tilde{\mathbf{B}}_1(\mathbf{S})$ *is superior to* LSE $\hat{\mathbf{B}}_1$ *if the inequality* (3.8) *holds.*

This is an important result which brings the sample size n, the dispersion of the response matrix $\mathbf{Y}_1$ and the population canonical correlation coefficients together. It is notable that $\frac{\bar{\lambda}(\Sigma_{11})}{\underline{\lambda}(\Sigma_{11})}$ is the condition number of Σ_{11}, while the last term $\rho_H = \prod_{i=1}^k \rho_i^2$ is Hotelling's generalized population correlation coefficient. An important conclusion is that the range of ρ_H over which $\tilde{\mathbf{B}}_1(\mathbf{S})$ is superior to $\hat{\mathbf{B}}_1$ narrows down as the sample size n becomes smaller. This means that if the canonical correlations are weak, then large n is needed for $\tilde{\mathbf{B}}_1(\mathbf{S})$ to be more efficient. In practical situations, the choice of n can be based partly on the sample canonical correlation coefficients.

4. Applications

We now investigate the growth curves analysis by SUMR model.

4.1. Selection of concomitants in growth curve model

Consider the growth curve model suggested by Potthoff and Roy (1964)

$$\mathbf{Y} = \mathbf{XBZ} + \mathbf{E} \tag{4.1}$$

where $\mathbf{Y}(n\times p)$ is a matrix of responses, $\mathbf{X}(n\times t)$ and $\mathbf{Z}(q\times p)$ are known matrices with rank t and q respectively, $\mathbf{B}$ is a $t\times q$ parameter matrix and the rows of $\mathbf{E}$

are independently distributed with mean vector zero and covariance matrix Σ. In order to estimate $\mathbf{B}$, let $\mathbf{H}(p \times q)$ be a matrix with full column rank q such that $\mathbf{ZH} = \mathbf{I}_q$, let $\mathbf{Y}_1 = \mathbf{YH}$, $\mathbf{X}_1 = \mathbf{X}$; then

$$\mathbf{EY}_1 = \mathbf{X}_1\mathbf{B}, \quad \mathrm{cov}(\mathbf{Y}_1) = \mathbf{I} \otimes (\mathbf{H}'\Sigma\mathbf{H}). \tag{4.2}$$

Since (4.2) is a standard multivariate linear model, we obtain an estimator of $\mathbf{B}$ which depends on the choice of $\mathbf{H}$,

$$\hat{\mathbf{B}}(\mathbf{H}) = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{YH}. \tag{4.3}$$

This estimation approach was first proposed by Potthoff and Roy (1964). As pointed out by Rao (1965), this approach would be inefficient as it ignores the information supplied by the concomitants defined as $\mathbf{Y}_2 = \mathbf{YT}$, where $\mathbf{T}(p \times m)$ ($m \leq p - q$) is any matrix with rank m such that $\mathbf{ZT} = 0$. Rao (1965) further suggested the use of part or all of the concomitants (the case when $m = p - q$) to make adjustment in order to improve the estimator (4.3). Note that

$$\mathbf{EY}_2 = 0, \quad \mathrm{cov}(\mathbf{Y}_2) = \mathbf{I} \otimes (\mathbf{T}'\Sigma\mathbf{T}), \tag{4.4}$$

and the rows of $(\mathbf{Y}_1, \mathbf{Y}_2) = \mathbf{Y}(\mathbf{H}, \mathbf{T})$ are independently distributed with covariance matrix $\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{H}'\Sigma\mathbf{H}' & \mathbf{H}'\Sigma\mathbf{T} \\ \mathbf{T}'\Sigma\mathbf{H} & \mathbf{T}'\Sigma\mathbf{T} \end{pmatrix}$. Hence (4.2) and (4.4) may be regarded as a special case of the SUMR model (1.1) with $\mathbf{X}_2 = 0$. Applying Theorems 1–3, we obtain the following results.

Theorem 4. 1. *When Σ is known, the* BLUE *of* $\mathbf{B}$ *among the class of linear unbiased estimators having the form* $\mathbf{A}_1\mathbf{Y}_1\mathbf{D}_1 + \mathbf{A}_2\mathbf{Y}_2\mathbf{D}_2$ *is*

$$\begin{aligned} \mathbf{B}^*(\mathbf{H}, \mathbf{T}, \Sigma) &= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}_1 - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}_2\mathbf{V}_{22}^{-1}\mathbf{V}_{21} \\ &= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{YH} - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{YT}(\mathbf{T}'\Sigma\mathbf{T})^{-1}\mathbf{T}'\Sigma\mathbf{H}. \end{aligned} \tag{4.5}$$

2. *When Σ is unknown, the two stage estimator*

$$\tilde{\mathbf{B}}(\mathbf{H}, \mathbf{T}, \mathbf{S}) = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{YH} - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{YT}(\mathbf{T}'\mathbf{ST})^{-1}\mathbf{T}'\mathbf{SH} \tag{4.6}$$

is superior to the estimator (4.3) *if n is sufficiently large, where* $\mathbf{S} = \mathbf{Y}'(\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')\mathbf{Y}$.

Theorem 4 gives an efficient method for improving the estimator (4.3) by making use of the concomitants. When all of the concomitants are considered, (4.5) leads to the BLUE of $\mathbf{B}$ in (4.1), and then (4.6) leads to the MLE of $\mathbf{B}$ obtained by Khatri (1966).

4.2. Seemingly unrelated growth curve models

During the past 30 years there have been many papers devoted to the estimation of a growth curve model. But the study of two or more growth curve models which we call seemingly unrelated growth curve models (SUGC model) has not yet received enough attention. In this paper we study the following SUGC model:

$$\mathbf{Y}_i = \mathbf{X}_i\mathbf{B}_i\mathbf{Z}_i + \mathbf{E}_i \qquad i = 1, 2 \tag{4.7}$$

where $\mathbf{Y}_i$ and $\mathbf{E}_i$ are $n \times p_i$, $\mathbf{X}_i(n \times t_i)$ and $\mathbf{Z}_i(q_i \times p_i)$ are known matrices with rank t_i and q_i respectively, $\mathbf{B}_i$ are $t_i \times q_i$ parameter matrices and the rows of $\mathbf{E} = (\mathbf{E}_1, \mathbf{E}_2)$ are independently distributed with zero mean vectors and a common covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \tag{4.8}$$

We now give an efficient method to estimate the parameter matrix $\mathbf{B}_1$, while the estimator for $\mathbf{B}_2$ may be obtained similarly.

Following the method suggested by Potthoff and Roy (1964), let $\mathbf{H}_i(p_i \times q_i)$ be a matrix with full column rank q_i such that $\mathbf{Z}_i\mathbf{H}_i = \mathbf{I}_{q_i}(i = 1, 2)$, in which case (4.7) may be reduced to a SUMR model

$$\underline{\mathbf{Y}}_i = \mathbf{X}_i\mathbf{B}_i + \underline{\mathbf{E}}_i \qquad i = 1, 2 \tag{4.9}$$

where $\underline{\mathbf{Y}}_i = \mathbf{Y}_i\mathbf{H}_i$, $\underline{\mathbf{E}}_i = \mathbf{E}_i\mathbf{H}_i$ and the rows of $\underline{\mathbf{E}} = (\underline{\mathbf{E}}_1, \underline{\mathbf{E}}_2)$ are independently distributed with zero mean vectors and a common covariance matrix

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{H}_1'\Sigma_{11}\mathbf{H}_1 & \mathbf{H}_1'\Sigma_{12}\mathbf{H}_2 \\ \mathbf{H}_2'\Sigma_{21}\mathbf{H}_1 & \mathbf{H}_2'\Sigma_{22}\mathbf{H}_2 \end{pmatrix}. \tag{4.10}$$

We obtain the following results:

Theorem 5. 1. *When Σ is known, the estimator*

$$\begin{aligned}\tilde{\mathbf{B}}_1 &= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\underline{\mathbf{Y}}_1 - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{N}_2\underline{\mathbf{Y}}_2\mathbf{V}_{22}^{-1}\mathbf{V}_{21} \\ &= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}_1\mathbf{H}_1 - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{N}_2\mathbf{Y}_2\mathbf{H}_2(\mathbf{H}_2'\Sigma_{22}\mathbf{H}_2)^{-1}\mathbf{H}_2'\Sigma_{21}\mathbf{H}_1\end{aligned} \tag{4.11}$$

is superior to the estimator obtained from the first equation of (4.9)

$$\hat{\mathbf{B}}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}_1\mathbf{H}_1, \tag{4.12}$$

where $\mathbf{N}_2 = \mathbf{I} - \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'$.

2. *When Σ is unknown, the two stage estimator*

$$\begin{aligned}\tilde{B}_1(T) &= (X_1'X_1)^{-1}X_1'Y_1H_1 \\ &\quad -(X_1'X_1)^{-1}X_1'N_2Y_2H_2(H_2'\Sigma_{22}H_2)^{-1}H_2'\Sigma_{21}H_1 \qquad (4.13)\end{aligned}$$

is superior to (4.12) *if n is sufficiently large, where* $S = Y'NY$, $Y = (Y_1, Y_2)$, *and* $N = I - X(X'X)^{-1}X$, *and* $X = (X_1, X_2)$.

Acknowledgements

I wish to thank the referee and the Associate Editor for careful reviews of the manuscript. The comments and suggestions of the referee helped the author to produce a much improved paper. The author wishes also to thank Professor Wang Songgui for helpful discussion.

References

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika* **28**, 321–377.

Khatri, C. G. (1966). A note on a MANOVA model applied to problems in growth curve. *Ann. Inst. Statist. Math.* **18**, 75–86.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* John Wiley, New York.

Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51**, 313–326.

Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika* **52**, 447–478.

Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proc. 5th Berkerly Symp. on Math. Statist.* **2**, 355–372.

Srivastava, V. K. and David, E. A. G. (1987). *Seemingly Unrelated Regression Equations Models: Estimation and Inference.* Marcel Dekker, New York.

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

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Amer. Statist. Assoc.* **57**, 348–368.

Zellner, A. (1963). Estimators for seemingly unrelated regression equations: Some exact finite sample results. *J. Amer. Statist. Assoc.* **58**, 977–992.

Department of Mathematics, University of Science and Technology of China, Hefei, Anhui 230026.

(Received August 1991; accepted December 1992)