MSE Superiority of Bayes and Empirical Bayes Estimators in Two Generalized Seemingly Unrelated Regressions *

Lichun Wang † Noël Veraverbeke

Department of Mathematics, Beijing Jiaotong University, Beijing 100044, China

Center for Statistics, Hasselt University, 3590 Diepenbeek, Belgium

Abstract This paper deals with the estimation problem in a system of two seemingly unrelated regression equations where the regression parameter is distributed according to the normal prior distribution $N(\beta_0, \sigma^2_\beta \Sigma_\beta)$. Resorting to the covariance adjustment technique, we obtain the best Bayes estimator of the regression parameter and prove its superiority over the best linear unbiased estimator (BLUE) in terms of the mean square error (MSE) criterion. Also, under the MSE criterion, we show that the empirical Bayes estimator of the regression parameter is better than the Zellner type estimator when the covariance matrix of error variables is unknown.

Keywords: Bayes method; seemingly unrelated regressions, covariance adjusted approach, mean square error criterion.

2000 MSC: Primary: 62J05, 62F11; Secondary: 62C10, 62C12

1. Introduction

The seemingly unrelated regression system was first introduced by Zellner (1962, 1963) and later developed by Kementa and Gilbert (1968), Mehta and Swamy (1976) and Wang (1988), etc. Recently, in a special issue of Journal of Statistical Planning and Inference 88 (2000), Gao and Huang establish some finite sample properties of the Zellner estimator in the context of m seemingly unrelated regression equations, whereas, Liu proposes a two stage estimator and proves its superiorities over the ordinary least square estimator and Zellner type estimator under mean square error matrix criterion.

*Partly supported by NNSF of China (10571001) and IUAP(P6/03) of Belgium

†Corresponding author: wlc@amss.ac.cn

Differing from the past works, in this paper we employ the Bayes and empirical Bayes approach to construct the estimators of the regression parameter and exhibit their MSE properties. Also, differing from the above regressions, here we do not make the same dimension assumption of observation vectors.

A system of two generalized seemingly unrelated regression equations is given by

$$y_1 = X_1\beta + u_1, \quad y_2 = X_2\gamma + u_2, \tag{1.1}$$

where y_1 and y_2 are $m \times 1$ and $n \times 1$ vectors of observations ($m \neq n$, without loss of generality, let $m > n$), X_1 and X_2 are $m \times p_1$ and $n \times p_2$ matrices with full column rank, β and γ are vectors of unknown parameters, u_1 and u_2 are $m \times 1$ and $n \times 1$ vectors of error variables, and

$$\begin{aligned} E(u_1) &= 0, \quad E(u_2) = 0, \\ Cov(u_1, u_1) &= \sigma_{11} I_m, \quad Cov(u_2, u_2) = \sigma_{22} I_n, \\ Cov(u_1, u_2) &= \sigma_{12} \begin{pmatrix} I_n \\ 0 \end{pmatrix}, \quad Cov(u_2, u_1) = \sigma_{21}(I_n \vdots 0), \end{aligned}$$

where $\Sigma^* = (\sigma_{ij})$ is a 2×2 non-diagonal positive definite matrix. Such a system (usually $m = n$) appears in many research fields and has received considerable attention including the above authors and Chen (1986), Lin (1991) and so on.

Denote $y = (y_1', y_2')'$, $X = \text{diag}(X_1, X_2)$, $\alpha = (\beta', \gamma')'$, $u = (u_1', u_2')'$, $\Sigma_{ij} = Cov(u_i, u_j)$. Then (1.1) can be represented as

$$y = X\alpha + u, \quad E(u) = 0, \quad Cov(u) = \Sigma, \tag{1.2}$$

where $\Sigma = (\Sigma_{ij})_{2\times 2}$ is a partitioned matrix.

In what follows, our main concern is how to estimate β better. To adopt the Bayes and empirical Bayes approach, we assume that the prior distribution of the parameter β is

$$\beta \sim N(\beta_0, \sigma_\beta^2 \Sigma_\beta), \tag{1.3}$$

where Σ_β is a positive definite matrix (namely $\Sigma_\beta > 0$), β_0 and σ_β^2 are hyper-parameters.

Furthermore, assume

$$u|\beta \sim N(0, \Sigma). \tag{1.4}$$

It follows from (1.3) and (1.4), that the posterior density of β given y_1 is (see Wang and Chow (1994))

$$f(\beta|y_1) \propto \exp\{-\frac{1}{2\sigma_{11}}(\beta - \bar{\beta})'\bar{\Sigma}^{-1}(\beta - \bar{\beta})\}, \tag{1.5}$$

where

$$\bar{\beta} = \bar{\Sigma}(X_1'X_1\hat{\beta} + \lambda\Sigma_\beta^{-1}\beta_0), \tag{1.6}$$

$\bar{\Sigma} = (X_1'X_1 + \lambda\Sigma_\beta^{-1})^{-1}$, $\lambda = \sigma_{11}/\sigma_\beta^2$ and $\hat{\beta} = (X_1'X_1)^{-1}X_1'y_1$. Thus, under any quadratic loss, the Bayes estimator (BE) of the parameter β would be the posterior expectation of β with given y_1, i.e.,

$$\hat{\beta}_{BE} = E(\beta|y_1) = \bar{\beta}. \tag{1.7}$$

It is clear that $\hat{\beta}_{BE}$ only contains the information of the first equation in the regressions (1.1) but that it does not make most use of all information of regressions since $\sigma_{12} \neq 0$.

As we know the estimation problems arise in many situations in statistics. An important concept is the minimum variance unbiased estimation (MVUE) and an interesting result is how to judge whether an estimator is MVUE or not: Let $\hat{g}(x)$ be an unbiased estimator (UE) of $g(\theta)$, and $Var_\theta(\hat{g}(x)) < \infty$, then $\hat{g}(x)$ is MVUE if and only if $Cov_\theta(\hat{g}(x), l(x)) = 0$ for any $\theta \in \Theta$ (parameter space), where $l(x)$ denotes any UE of zero. Obviously, if there exists an UE $l_0(x)$ of zero such that $Cov_\theta(\hat{g}(x), l_0(x)) \neq 0$, then $\hat{g}(x)$ must not be the MVUE of its mean. However, a problem is how we utilize the relationship between $l_0(x)$ and $\hat{g}(x)$ to obtain the MVUE of $g(\theta)$. Rao (1967) introduced the covariance adjusted approach to propose a UE of $g(\theta)$ whose variance is less than $\hat{g}(x)$, which is a linear combination of $\hat{g}(x)$ and $l_0(x)$.

In the followings, by virtute of the covariance adjustment technique, firstly, we use an UE of zero to improve $\hat{\beta}_{BE}$ and get $\hat{\beta}_{BE}^{(1)}$, secondly, we adjust $\hat{\beta}_{BE}^{(1)}$ by another UE of zero. Repeating this process, we obtain the best BE of the parameter β, which contains all information of β in the regressions (1.1), and prove its MSE superiority over the BLUE of β. When σ_{ij} $(i, j = 1, 2)$ and the hyper-parameters are unknown, we replace them by their consistent estimators in the best BE of β and present the corresponding empirical Bayes (EB) estimator and exhibit its MSE superiority, too.

2. MSE Superiority of the Best BE

We first state the following covariance adjustment lemma.

Lemma 2.1. Assume that T_1 and T_2 are $k_1 \times 1$ and $k_2 \times 1$ statistics with $ET_1 = \theta$ and $ET_2 = 0$, where θ is an unknown parameter vector. Let

$$Cov\begin{pmatrix} T_1 \\ T_2 \end{pmatrix} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} = V.$$

If $V_{12} \neq 0$, then there exists a best linear unbiased estimator (BLUE) $\theta^* = T_1 - V_{12}V_{22}^{-}T_2$ over a class of estimators $\mathcal{A} = \{A_1T_1 + A_2T_2 | A_1, A_2 \text{ are nonrandom matrices}\}$, and

$$Cov(\theta^*) = V_{11} - V_{12}V_{22}^{-}V_{21} \leq V_{11} = Cov(T_1),$$

where V_{22}^{-} is a generalized inverse of matrix V_{22}, and $A \geq B$ denotes $A - B \geq 0$(that means $A - B$ is real positive semi-definite).

Proof. It can directly be derived from Rao (1967).

Combing Lemma 2.1 with $\hat{\beta}_{BE}$, we obtain the covariance adjustment estimator sequence for the parameter β as follows:

$$\begin{aligned} \hat{\beta}_{BE}^{(2k-1)} &= \hat{\beta}_{BE}^{(2k-2)} - Cov(\hat{\beta}_{BE}^{(2k-2)}, N_2y_2|\beta)[Cov(N_2y_2)]^{-}N_2y_2, \\ \hat{\beta}_{BE}^{(2k)} &= \hat{\beta}_{BE}^{(2k-1)} - Cov(\hat{\beta}_{BE}^{(2k-1)}, N_1y_1|\beta)[Cov(N_1y_1|\beta)]^{-}N_1y_1, \\ & k = 1, 2, ..., \end{aligned} \tag{2.1}$$

where $\hat{\beta}_{BE}^{(0)} = \hat{\beta}_{BE}$.

Simple induction computation yields

$$\begin{aligned}
\hat{\beta}_{BE}^{(2k-1)} &= \bar{\Sigma} X_1'(P_1 + \rho^2 \bar{N}_2 \bar{N}_1)^{k-1}(y_1 - \frac{\sigma_{12}}{\sigma_{22}} \bar{N}_2 y_2) + \lambda \bar{\Sigma} \Sigma_\beta^{-1} \beta_0, \\
\hat{\beta}_{BE}^{(2k)} &= \bar{\Sigma} X_1'(P_1 + \rho^2 \bar{N}_2 \bar{N}_1)^{k} y_1 - \bar{\Sigma} X_1'(P_1 + \rho^2 \bar{N}_2 \bar{N}_1)^{k-1} \frac{\sigma_{12}}{\sigma_{22}} \bar{N}_2 y_2 \\
&\quad + \lambda \bar{\Sigma} \Sigma_\beta^{-1} \beta_0,
\end{aligned} \tag{2.2}$$

where $\rho^2 = \sigma_{12}^2/(\sigma_{11}\sigma_{22})$, $\bar{N}_1 = (I_n \vdots 0) N_1$, $\bar{N}_2 = \begin{pmatrix} I_n \\ 0 \end{pmatrix} N_2$, and

$$\begin{aligned}
P_1 &= I_m - N_1 = X_1(X_1'X_1)^{-1}X_1', \\
P_2 &= I_n - N_2 = X_2(X_2'X_2)^{-1}X_2'.
\end{aligned}$$

Therefore, we have the following theorem.

Theorem 2.1. Let $\hat{\beta}_{BE}^{(2k-1)}$ and $\hat{\beta}_{BE}^{(2k)}$ be defined in (2.1), we have

$$\begin{aligned}
\hat{\beta}_{BE}^{(\infty)} &= \lim_{k\to\infty} \hat{\beta}_{BE}^{(2k-1)} = \lim_{k\to\infty} \hat{\beta}_{BE}^{(2k)} = \bar{\Sigma} X_1' \sum_{i=0}^{\infty} (\rho^2 \bar{N}_2 \bar{N}_1)^i (y_1 - \frac{\sigma_{12}}{\sigma_{22}} \bar{N}_2 y_2) + \lambda \bar{\Sigma} \Sigma_\beta^{-1} \beta_0 \\
&= \bar{\Sigma} X_1'(I_m - \rho^2 \bar{N}_2 \bar{N}_1)^{-1}(y_1 - \frac{\sigma_{12}}{\sigma_{22}} \bar{N}_2 y_2) + \lambda \bar{\Sigma} \Sigma_\beta^{-1} \beta_0.
\end{aligned}$$

Proof. Note that $X_1'(P_1 + \rho^2 \bar{N}_2 \bar{N}_1)^k = X_1' \sum_{i=0}^k (\rho^2 \bar{N}_2 \bar{N}_1)^i$ and $\lambda(\rho^2 \bar{N}_2 \bar{N}_1) < 1$, we know that Theorem 2.1 is true, where $\lambda(A)$ denotes any eigenvalue of matrix A.

Remark 2.1. Following the fact that $E(\hat{\beta}_{BE}^{(k)}|\beta) = E(\hat{\beta}_{BE}^{(0)}|\beta)$ and the monotonicity of $Cov(\hat{\beta}_{BE}^{(k)}|\beta)$, it is easy to see that in BE sequence (2.2) $Cov(\hat{\beta}_{BE}^{(k+1)}) \leq Cov(\hat{\beta}_{BE}^{(k)})$ and hence MSE$(\hat{\beta}_{BE}^{(k+1)}) \leq$ MSE $(\hat{\beta}_{BE}^{(k)})$. Obviously, $\hat{\beta}_{BE}^{(\infty)}$ is the best.

In the following, note that in the regressions (1.1) X_1 and X_2 are $m \times p_1$ and $n \times p_2$ matrices and $m > n$, we partition X_1 as $X_1 = (X_{11}' \vdots X_{12}')'$ and make the following intuitive assumption

$$\mu(X_{11}') \bigcap \mu(X_{12}') = \{0\}, \tag{2.3}$$

where X_{11} and X_{12} are $n \times p_1$ and $(m-n) \times p_1$ matrices, respectively, and $\mu(A)$ denotes the space generated by the column vector of matrix A.

Lemma 2.2. If $\mu(X'_{11})\bigcap\mu(X'_{12})=\{0\}$, then

(a) $X_{11}(X'_1X_1)^{-1}X'_{12}=0;$

(b) $X'_{11}X_{11}(X'_1X_1)^{-1}X'_{11}=X'_{11},\quad X'_{12}X_{12}(X'_1X_1)^{-1}X'_{12}=X'_{12};$

(c) $X'_1\bar{N}_2\bar{N}_1=-X'_1\bar{P}_2\bar{N}_1,\quad \bar{P}_2\bar{N}_1\bar{N}_2\bar{N}_1=\bar{P}_2\bar{P}_1\bar{P}_2\bar{N}_1,$

where $\bar{P}_1=(I_n\vdots 0)P_1$, $\bar{P}_2=\begin{pmatrix} I_n \\ 0 \end{pmatrix}P_2$.

Proof. (a) Set

$$D=X'_{11}X_{11}(X'_1X_1)^{-1}X'_{12}X_{12}, \tag{2.4}$$

thus $\mu(D)\subset\mu(X'_{11})$. Note that $X'_1X_1=X'_{11}X_{11}+X'_{12}X_{12}$, we can represent D as

$$D=X'_{11}X_{11}-X'_{11}X_{11}(X'_1X_1)^{-1}X'_{11}X_{11}, \tag{2.5}$$

that means $D=D'$. Hence, $\mu(D)\subset\mu(X'_{12})$. Since $\mu(X'_{11})\bigcap\mu(X'_{12})=\{0\}$, $D=0$. Then

$$X_{11}(X'_1X_1)^{-1}X'_{12}=0. \tag{2.6}$$

(b) It follows from (a),

$$X'_{11}X_{11}(X'_1X_1)^{-1}X'_{11}X_{11}=X'_{11}X_{11}(X'_1X_1)^{-1}(X'_{11}X_{11}+X'_{12}X_{12})=X'_{11}X_{11}.$$

Hence, $X'_{11}X_{11}(X'_1X_1)^{-1}X'_{11}=X'_{11}$. Similarly, we can prove the other conclusion of (b).

(c) The conclusions of (c) are direct results of (a) and (b).

Based on Lemma 2.2, we present Theorem 2.2.

Theorem 2.2. In the regressions (1.1), the BLUE of the parameter β is

$$\hat{\beta}_{BLUE}=(X'_1X_1)^{-1}X'_1(I_m-\rho^2\bar{N}_2\bar{N}_1)^{-1}(y_1-\frac{\sigma_{12}}{\sigma_{22}}\bar{N}_2y_2).$$

Proof. From the expression of (1.2), when Σ is known, we know

$$\hat{\alpha}_{BLUE}=\begin{pmatrix} \hat{\beta}_{BLUE} \\ \hat{\gamma}_{BLUE} \end{pmatrix}=(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y. \tag{2.7}$$

Denote $\Sigma^{-1} = (\Sigma^{ij})_{2\times 2}$ and $(X'\Sigma^{-1}X)^{-1} = (W^{ij})_{2\times 2}$, we obtain

$$\hat{\beta}_{BLUE} = (W^{11}X_1'\Sigma^{11} + W^{12}X_2'\Sigma^{21})y_1 + (W^{11}X_1'\Sigma^{12} + W^{12}X_2'\Sigma^{22})y_2. \tag{2.8}$$

By simple algebra and induction computation, we have

$$\hat{\beta}_{BLUE} = (X_1'X_1)^{-1}X_1'\{I_m - \rho^2\sum_{i=0}^{\infty}(\rho^2\bar{P}_2\bar{P}_1)^i\bar{P}_2\bar{N}_1\}(y_1 - \frac{\sigma_{12}}{\sigma_{22}}\bar{N}_2y_2). \tag{2.9}$$

By the conclusion (c) of Lemma 2.2, we know

$$X_1'\{I_m - \rho^2\sum_{i=0}^{k-1}(\rho^2\bar{P}_2\bar{P}_1)^i\bar{P}_2\bar{N}_1\} = X_1'\sum_{i=0}^{k}(\rho^2\bar{N}_2\bar{N}_1)^i. \tag{2.10}$$

Together with $\lambda(\rho^2\bar{N}_2\bar{N}_1) < 1$ Theorem 2.2 ' conclusion holds.

Especially, if $P_{11}P_2 = P_2P_{11}$, where $P_{11} = X_{11}(X_1'X_1)^{-1}X_{11}'$, then we have the following clear and succinct results for $\hat{\beta}_{BE}^{(\infty)}$ and $\hat{\beta}_{BLUE}$.

Theorem 2.3. If $P_{11}P_2 = P_2P_{11}$, then

$$\begin{aligned}
\hat{\beta}_{BE}^{(\infty)} &= \bar{\Sigma}X_1'(y_1 - \frac{\sigma_{12}}{\sigma_{22}}\bar{N}_2y_2) + \lambda\bar{\Sigma}\Sigma_\beta^{-1}\beta_0 = \hat{\beta}_{BE}^{(1)}. \\
\hat{\beta}_{BLUE} &= \hat{\beta} - \frac{\sigma_{12}}{\sigma_{22}}(X_1'X_1')^{-1}X_1'\bar{N}_2y_2.
\end{aligned}$$

Proof. Using the fact that

$$\begin{aligned}
X_1'\bar{N}_2\bar{N}_1 &= (X_{11}' \vdots X_{12}')\begin{pmatrix} I_n - P_2 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} I_n - P_{11} & 0 \\ 0 & I_{m-n} - X_{12}(X_1'X_1)^{-1}X_{12}' \end{pmatrix} \\
&= X_{11}' - X_{11}'P_2 - X_{11}'P_{11} + X_{11}'P_2P_{11}
\end{aligned}$$

and $X_{11}'P_{11} = X_{11}'$, by $P_{11}P_2 = P_2P_{11}$, Theorem 2.3 is obvious.

Now we state the comparison result of MSE$(\hat{\beta}_{BE}^{(\infty)})$ and MSE $(\hat{\beta}_{BLUE})$.

Theorem 2.4. Let $(\hat{\beta}_{BE}^{(\infty)})$ and $(\hat{\beta}_{BLUE})$ be defined in Theorem 2.1 and Theorem 2.2, respectively, then MSE$(\hat{\beta}_{BE}^{(\infty)}) <$ MSE$(\hat{\beta}_{BLUE})$.

Proof. Firstly, simple calculation shows

$$Cov(\hat{\beta}_{BE}^{(\infty)}|\beta) = \sigma_{11}\bar{\Sigma}C\bar{\Sigma}, \tag{2.11}$$

where $C = X_1'(I_m - \rho^2 \bar{N}_2 \bar{N}_1)^{-1}[I_m - \rho^2 \bar{N}_2 \bar{N}_2'](I_m - \rho^2 \bar{N}_1' \bar{N}_2')^{-1} X_1$. Similarly,

$$Cov(\hat{\beta}_{BLUE}|\beta) = \sigma_{11}(X_1'X_1)^{-1}C(X_1'X_1)^{-1}. \tag{2.12}$$

Secondly, we have

$$Cov(E(\hat{\beta}_{BE}^{(\infty)}|\beta)) = \sigma_\beta^2 \bar{\Sigma} X_1'X_1 \Sigma_\beta X_1'X_1 \bar{\Sigma}, \tag{2.13}$$

and

$$Cov(E(\hat{\beta}_{BLUE}|\beta)) = \sigma_\beta^2 \Sigma_\beta. \tag{2.14}$$

Hence

$$\begin{aligned} Cov(\hat{\beta}_{BE}^{(\infty)}) &= ECov(\hat{\beta}_{BE}^{(\infty)}|\beta) + Cov(E(\hat{\beta}_{BE}^{(\infty)}|\beta)) \\ &= \sigma_{11}\bar{\Sigma}C\bar{\Sigma} + \sigma_\beta^2 \bar{\Sigma} X_1'X_1 \Sigma_\beta X_1'X_1 \bar{\Sigma}, \end{aligned} \tag{2.15}$$

and also

$$Cov(\hat{\beta}_{BLUE}) = \sigma_{11}(X_1'X_1)^{-1}C(X_1'X_1)^{-1} + \sigma_\beta^2 \Sigma_\beta. \tag{2.16}$$

Note that $X_1'X_1 + \lambda \Sigma_\beta^{-1} > X_1'X_1 > 0$, hence $(X_1'X_1)^{-1} > \bar{\Sigma}$. Thus by $C \geq 0$, we have

$$\sigma_{11}(X_1'X_1)^{-1}C(X_1'X_1)^{-1} \geq \sigma_{11}\bar{\Sigma}C\bar{\Sigma}. \tag{2.17}$$

Similarly, from $(X_1'X_1)^{-1} > \bar{\Sigma}$ and $X_1'X_1\Sigma_\beta X_1'X_1 > 0$, we have

$$\sigma_\beta^2 \Sigma_\beta > \sigma_\beta^2 \bar{\Sigma} X_1'X_1 \Sigma_\beta X_1'X_1 \bar{\Sigma}. \tag{2.18}$$

It follows from (2.15)-(2.18),

$$Cov(\hat{\beta}_{BE}^{(\infty)}) < Cov(\hat{\beta}_{BLUE}). \tag{2.19}$$

Thus

$$\begin{aligned} \mathrm{MSE}(\hat{\beta}_{BE}^{(\infty)}) &= trace[Cov(\hat{\beta}_{BE}^{(\infty)})] + ||E\hat{\beta}_{BE}^{(\infty)} - \beta||^2 \\ &< trace[Cov(\hat{\beta}_{BLUE})] + ||E\hat{\beta}_{BLUE} - \beta||^2 \\ &= \mathrm{MSE}(\hat{\beta}_{BLUE}), \end{aligned} \tag{2.20}$$

since $E\hat{\beta}_{BE}^{(\infty)} = E\{E[\hat{\beta}_{BE}^{(\infty)}|\beta]\} = \beta_0 = E\hat{\beta}_{BLUE}$.

The proof of Theorem 2.4 is complete.

3. MSE Superiority of the EB Estimator

However, in many situations, the covariance of errors Σ may be unknown, so that $\hat{\beta}_{BE}^{(\infty)}$ is unavailable to use. In this Section we use observations y_i $(i = 1, 2)$ to construct an estimator for σ_{ij} $(i, j = 1, 2)$ and present the corresponding EB estimator and its MSE superiority.

Denote $Y = (y_{11} \vdots y_2)$, where y_{11} is the sub-vector containing the first n observations of y_1. We define the estimator for $\Sigma^* = (\sigma_{ij})$ as follows:

$$\hat{\Sigma}^* = (\hat{\sigma}_{ij}) = \frac{1}{n-r} Y' N^* Y, \tag{3.1}$$

where $N^* = I_n - X^*[(X^*)'X^*]^-(X^*)'$, $X^* = (X_{11} \vdots X_2)$ and $\text{rank}(X^*) \hat{=} r$.

Note that $N^* EY = 0$, hence $(n-r)\hat{\Sigma}^*|\beta \sim \text{Wishart}(n-r, \Sigma^*)$. Thus, there exist $n-r$ i.i.d. random variables $Z_i \sim N_2\,(0, \Sigma^*)$ such that $Y'N^*Y = \sum_{i=1}^{n-r} Z_i Z_i'$. By the law of large numbers of Kolmogorov, we have $\hat{\Sigma}^*|\beta \xrightarrow{a.s.} \Sigma^*$, as $R \to \infty$, where $R = n - r$.

3.1 β_0 is known

We define the EB estimator for the parameter β as follows:

$$\tilde{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*) = \bar{\Sigma} X_1' \{I_m - \hat{\rho}^2 \sum_{i=0}^{\infty} (\hat{\rho}^2 \bar{P}_2 \bar{P}_1)^i \bar{P}_2 \bar{N}_1\}(y_1 - \frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}} \bar{N}_2 y_2) + \lambda_0 \bar{\Sigma} \Sigma_\beta^{-1} \beta_0, \tag{3.2}$$

where $\hat{\rho}^2 = \hat{\sigma}_{12}^2 / (\hat{\sigma}_{11} \hat{\sigma}_{22})$.

Also define the estimator of the BLUE as

$$\tilde{\beta}_{BLUE}(\hat{\Sigma}^*) = (X_1'X_1)^{-1} X_1' \{I_m - \hat{\rho}^2 \sum_{i=0}^{\infty} (\hat{\rho}^2 \bar{P}_2 \bar{P}_1)^i \bar{P}_2 \bar{N}_1\}(y_1 - \frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}} \bar{N}_2 y_2), \tag{3.3}$$

which is a Zellner type estimator.

It is necessary to notice that in this subsection we take λ as a constant λ_0 for simplicity. That means $\sigma_\beta^2 = \sigma_{11}/\lambda_0$, i.e., $\beta \sim N(\beta_0, \lambda_0^{-1}\sigma_{11}\Sigma_\beta)$. In fact if $\lambda = \sigma_{11}/\sigma_\beta^2$ is unknown, we must define a suitable estimator, such as $\hat{\beta}$ since $\hat{\beta} \sim N(\beta_0, \sigma_{11}(X_1'X_1)^{-1} + \sigma_\beta^2 \Sigma_\beta)$, for the hyper-parameter σ_β^2. Unfortunately, it is very very difficult to separate σ_β^2 from the covariance structure $\sigma_{11}(X_1'X_1)^{-1} + \sigma_\beta^2\Sigma_\beta$. Also, Arnold (1981) suggests taking $\Sigma_\beta = (X_1'X_1)^{-1}$ as a convenient choice. Although it makes the above problem easy, it is more or less unusual or unreasonable.

Theorem 3.1. Let $\tilde{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*)$ and $\tilde{\beta}_{BLUE}(\hat{\Sigma}^*)$ be given in (3.2) and (3.3), respectively. If $R \to \infty$, then $\text{MSE}(\tilde{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*)) < \text{MSE}(\tilde{\beta}_{BLUE}(\hat{\Sigma}^*))$.

Proof. Denote $B = X_1'\{I_m - \hat{\rho}^2 \sum_{i=0}^{\infty} (\hat{\rho}^2 \bar{P}_2 \bar{P}_1)^i \bar{P}_2 \bar{N}_1\}$, we have

$$\begin{aligned} Cov(\tilde{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*)|\beta) &= ECov(\tilde{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*)|\beta, \hat{\sigma}_{ij}) + Cov(E(\tilde{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*)|\beta, \hat{\sigma}_{ij})) \\ &= \sigma_{11} E[\bar{\Sigma} B(I_m + \bar{N}_2(\frac{\sigma_{22}}{\sigma_{11}} \frac{\hat{\sigma}_{12}^2}{\hat{\sigma}_{22}^2} - 2\frac{\sigma_{12}}{\sigma_{11}} \frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}})\bar{N}_2')B'\bar{\Sigma}]. \qquad (3.4) \end{aligned}$$

Similarly, we have

$$\begin{aligned} &Cov\ (\tilde{\beta}_{BLUE}(\hat{\Sigma}^*)|\beta) = ECov(\tilde{\beta}_{BLUE}(\hat{\Sigma}^*)|\beta, \hat{\sigma}_{ij}) + Cov(E(\tilde{\beta}_{BLUE}(\hat{\Sigma}^*)|\beta, \hat{\sigma}_{ij})) \\ &= \ \sigma_{11} E[(X_1'X_1)^{-1} B(I_m + \bar{N}_2(\frac{\sigma_{22}}{\sigma_{11}} \frac{\hat{\sigma}_{12}^2}{\hat{\sigma}_{22}^2} - 2\frac{\sigma_{12}}{\sigma_{11}} \frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}})\bar{N}_2')B'(X_1'X_1)^{-1}]. \qquad (3.5) \end{aligned}$$

By the fact that $\hat{\sigma}_{ij}|\beta \xrightarrow{a.s.} \sigma_{ij}$ as $R \to \infty$, it is easy to see that

$$Cov(\tilde{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*)|\beta) \leq Cov(\tilde{\beta}_{BLUE}(\hat{\Sigma}^*)|\beta), \quad \text{as } R \to \infty. \qquad (3.6)$$

Note that $X_1'\bar{P}_2\bar{N}_1 N^* = X_1'\bar{N}_2 N^* = X_1'\bar{P}_2\bar{N}_1\bar{N}_2 N^* = 0$, $\hat{\sigma}_{ij}$ $(i, j = 1, 2)$ are independent of $X_1'\bar{P}_2\bar{N}_1 y_1$, $X_1'\bar{N}_2 y_2$ and $X_1'\bar{P}_2\bar{N}_1\bar{N}_2 y_2$. Therefore, we have

$$E(\tilde{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*)|\beta) = E(\hat{\beta}_{BE}^{(\infty)}|\beta), \qquad (3.7)$$

and

$$E(\tilde{\beta}_{BLUE}(\hat{\Sigma}^*)|\beta) = E(\hat{\beta}_{BLUE}|\beta). \qquad (3.8)$$

It follows from (3.6)-(3.8), (2.13)-(2.14) and (2.18),

$$\begin{aligned} Cov(\tilde{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*)) &= E(Cov(\tilde{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*)|\beta)) + Cov(E(\tilde{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*)|\beta)) \\ &< E(Cov(\tilde{\beta}_{BLUE}(\hat{\Sigma}^*)|\beta)) + Cov(E(\tilde{\beta}_{BLUE}(\hat{\Sigma}^*)|\beta)) \\ &= Cov(\tilde{\beta}_{BLUE}(\hat{\Sigma}^*), \quad \text{as } R \to \infty. \qquad (3.9) \end{aligned}$$

Together with (3.7) and (3.8), we obtain

$$\mathrm{MSE}(\tilde{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*)) < \mathrm{MSE}(\tilde{\beta}_{BLUE}(\hat{\Sigma}^*)), \quad \text{as } R \to \infty. \tag{3.10}$$

Theorem 3.1 has been proved.

3.2 β_0 is unknown

In this subsection, we do not make the assumption that λ is equal to a constant λ_0.

Since $\hat{\beta} \sim N(\beta_0, \sigma_{11}(X_1'X_1)^{-1} + \sigma_\beta^2 \Sigma_\beta)$, we first estimate β_0 by $\hat{\beta}$ in the second term of the right-hand side of the expression (3.2) and hence the EB estimator of β is

$$\hat{\hat{\Sigma}} X_1' \{I_m - \hat{\rho}^2 \sum_{i=0}^{\infty} (\hat{\rho}^2 \bar{P}_2 \bar{P}_1)^i \bar{P}_2 \bar{N}_1\}(y_1 - \frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}} \bar{N}_2 y_2) + \hat{\lambda} \hat{\hat{\Sigma}} \Sigma_\beta^{-1} \hat{\beta}, \tag{3.11}$$

where $\hat{\sigma}_\beta^2$ is a suitable estimator for σ_β^2, $\hat{\lambda} = \hat{\sigma}_{11}/\hat{\sigma}_\beta^2$ and $\hat{\hat{\Sigma}} = (X_1'X_1 + \hat{\lambda}\Sigma_\beta^{-1})^{-1}$.

However, note that $Cov(\hat{\beta}, N_2 y_2 | \beta) = \sigma_{12}(X_1'X_1)^{-1} X_{11}' N_2 \neq 0$ if $X_{11}' N_2 \neq 0$ (In fact $P_{11} P_2 \neq P_2 P_{11} \Longrightarrow X_{11}' N_2 \neq 0$). Thereby, by Lemma 2.1, we adjust $\hat{\beta}$ by $N_2 y_2$ and obtain a better estimator $\hat{\beta}_1(\Sigma^*)$. Similarly, we can use $N_1 y_1$ to improve $\hat{\beta}_1(\Sigma^*)$ and get $\hat{\beta}_2(\Sigma^*)$. Repeating above steps, finally we obtain

$$\hat{\beta}_\infty(\Sigma^*) = (X_1'X_1)^{-1} X_1' \{I_m - \rho^2 \sum_{i=0}^{\infty} (\rho^2 \bar{P}_2 \bar{P}_1)^i \bar{P}_2 \bar{N}_1\}(y_1 - \frac{\sigma_{12}}{\sigma_{22}} \bar{N}_2 y_2). \tag{3.12}$$

Replacing Σ^* by $\hat{\Sigma}^*$ in (3.12) and substituting $\hat{\beta}_\infty(\hat{\Sigma}^*)$ into (3.11), we define the following EB estimator for the parameter β in this subsection,

$$\begin{aligned} \bar{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*) &= \hat{\hat{\Sigma}} X_1' \{I_m - \hat{\rho}^2 \sum_{i=0}^{\infty} (\hat{\rho}^2 \bar{P}_2 \bar{P}_1)^i \bar{P}_2 \bar{N}_1\}(y_1 - \frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}} \bar{N}_2 y_2) \\ &\quad + \hat{\lambda} \hat{\hat{\Sigma}} \Sigma_\beta^{-1} \hat{\beta}_\infty(\hat{\Sigma}^*). \end{aligned} \tag{3.13}$$

It is interesting to see $\bar{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*) = \tilde{\beta}_{BLUE}(\hat{\Sigma}^*)$ at this time. Hence, we have the following obvious result.

Theorem 3.2. If $R \to \infty$, then the EB estimator equals to the estimator of BLUE, i.e., $\bar{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*) = \tilde{\beta}_{BLUE}(\hat{\Sigma}^*)$, and

$$\mathrm{MSE}(\bar{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*)) = \mathrm{MSE}(\tilde{\beta}_{BLUE}(\hat{\Sigma}^*)).$$

Similar to Theorem 2.3, Theorem 3.2 has the following corollary.

Corollary 3.1. If $P_{11}P_2 = P_2P_{11}$ and $R \to \infty$, then

$$\bar{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*) = \tilde{\beta}_{BLUE}(\hat{\Sigma}^*) = \hat{\beta} - \frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}}(X_1'X_1')^{-1}X_1'\bar{N}_2y_2.$$

Also, it is not difficult to see that $\text{MSE}(\bar{\beta}_{EB}^{(\infty)}(\hat{\Sigma}^*)) = \text{MSE}(\tilde{\beta}_{BLUE}(\hat{\Sigma}^*)) \le \text{MSE}(\hat{\beta})$.

4. Conclusions

The covariance adjustment technique is a very effective approach. Combing it with the Bayes method, under the assumption that the prior is normal, it presents the best BE of the regression parameter in the sense of covariance. And under the MSE criterion, the best BE performs better than the BLUE.

If the normal prior mean β_0 is known, based on a covariance condition, we show that the corresponding EB estimator is better under MSE criterion. Even though the hyper-parameter β_0 is unknown, the EB estimator can still work as good as the Zellner type estimator. In fact, due to estimating β_0 by $\hat{\beta}$ in $\hat{\beta}_{BE}$, following the covariance adjustment approach, the best BE equals to the BLUE as well as the EB estimator is the same as Zellner estimator. Also, we find the BLUE of the regression parameter can be obtained using the covariance adjustment approach to improve $\hat{\beta}$.

Acknowledgement The authors would like to thank the Editor and an anonymous referee for helpful comments which improved the presentation of the paper.

References

[1] Arnold, S. F., 1981. The theory of linear model and multivariate analysis. Wiley, New York.

[2] Chen, C. H., 1986. Optimality of two-stage estimators of seemingly unrelated regression equations. Chinese Journal of Applied Probability and Statistics, 2, 112-119.

[3] Gao, D.D., Huang, R. B., 2000. Some finite sample properties of Zellner estimator in the context of m seemingly unrelated regression equations. Journal of Statistical Planning and Inference, 88, 267-283.

[4] Kementa, J., Gilbert, R. F., 1968. Small sample properties of alternative estimators of seemingly unrelated regressions. J. Amer. Statist. Assoc., 63, 1180-1200.

[5] Lin, C. T., 1991. The efficiency of least squares estimator of a seemingly unrelated regression model. Comm. Statist. Simulation, 20, 919-925.

[6] Liu, J. S., 2000. MESM dominance of estimators in two seemingly unrelated regressions. Journal of Statistical Planning and Inference, 88, 255-266.

[7] Mehta, J. S., Swamy, P.A.V.B., 1976. Further evidence on the relative efficiencies of Zellner's seemingly unrelated regressions estimator. J. Amer. Statist. Assoc., 71, 634-639.

[8] Rao, C. R., 1967. Least squares theory using estimated dispersion matrix and its application to measurement of signals, In Proceedings of the 5th Berkeley Symposium on Math. Statist. and Prob., L. Le Cam and J. Neyman. eds., 1, 355—372.

[9] Wang, S. G., 1988. A new estimator of regression coefficients in linear regression system. Chinese Sciences, Ser. A, 10, 1033-1040.

[10] Wang, S. G., Chow, S. C., 1994. Advanced linear models. Marcel Dekker Inc., New York.

[11] Zellner, A., 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. J. Amer. Statist. Assoc., 57, 348-368.

[12] Zellner, A., 1963. Estimators of seemingly unrelated regressions: some exact finite sample results. J. Amer. Statist. Assoc., 58, 977-992.