How to Improve Classical Estimators via Linear Bayes Method?

Lichun Wang[1,a]

[a]Department of Mathematics, Beijing Jiaotong University, China

Abstract

In this survey, we use the normal linear model to demonstrate the use of the linear Bayes method. The superiorities of linear Bayes estimator (LBE) over the classical UMVUE and MLE are established in terms of the mean squared error matrix (MSEM) criterion. Compared with the usual Bayes estimator (obtained by the MCMC method) the proposed LBE is simple and easy to use with numerical results presented to illustrate its performance. We also examine the applications of linear Bayes method to some other distributions including two-parameter exponential family, uniform distribution and inverse Gaussian distribution, and finally make some remarks.

Keywords: linear Bayes method, MCMC method, MSEM criterion, normal linear model, two-parameter exponential family, uniform distribution, inverse Gaussian distribution

1. Introduction

The linear Bayes method was originally proposed by Hartigan (1969), which suggests that in Bayesian statistics one can replace a completely specified prior distribution by an assumption with just a few moments of the distribution. It has been subsequently discussed by Rao (1973) from linear optimization viewpoint. Lamotte (1978) later develops a class of linear estimators, called Bayes linear estimators, by searching, among all linear estimators that have least average total mean squared error. Goldstein (1983) considers the problem of modifying the linear Bayes estimator for the mean of a distribution of unknown form using a sample variance estimate. Heiligers (1993) studies the relationship between linear Bayes estimation and minimax estimation in linear models with partial parameter restrictions. Hoffmann (1996) proposes a well-described subclass of Bayes linear estimators for the unknown parameter vector in the linear regression model with ellipsoidal parameter constraints and obtains a necessary and sufficient condition to ensure that the considered Bayes linear estimators improves the least squared estimator over the whole ellipsoid regardless of the selected generalized risk function. In the framework of empirical Bayes, Samaniego and Vestrup (1999) and Pensky and Ni (2000) respectively construct linear empirical Bayes estimators and establish their superiorities over standard and traditional estimators. In application fields, Busby *et al.* (2005) proposes the application of Bayes linear methodology to uncertainty evaluation in reservoir forecasting. Zhang and Wei (2005) also drive the unique Bayes linear unbiased estimator of estimable functions for the singular linear model. Wei and Zhang (2007) employs linear Bayes procedure to define Bayes linear minimum risk estimation in a linear model and discusses its superiorities. Recently, Zhang *et al.* (2011, 2012)

extend the research on a linear Bayes estimator to the partitioned linear model and multivariate linear models, respectively.

In this paper, along the same line as in Wang and Singh (2014), we use the normal linear model as an example to demonstrate how to apply a linear Bayes method to simultaneously estimate all the parameters involved in the model and elaborate advantages and potential disadvantages.

Let W be a known p-dimension subspace of R^n. Suppose that we observe the random vector

$$Y \sim N_n\left(\mu, \sigma^2 I_n\right), \quad \mu \in W, \ \sigma^2 > 0. \tag{1.1}$$

This model is called the normal linear model as defined by Arnold (1980) and adopted by many other authors as well. Let X be a basis matrix for W. Then X is an $n \times p$ matrix of rank p, and there exists a unique $\beta \in R^p$ such that $\mu = X\beta$. Hence, an equivalent version of the normal linear model can be presented, and we observe the random vector

$$Y \sim N_n\left(X\beta, \sigma^2 I_n\right), \quad \beta \in R^p, \ \sigma^2 > 0. \tag{1.2}$$

Define $\hat{\beta} = (X'X)^{-1}X'Y$ and $\hat{\sigma}^2 = (\|Y\|^2 - \|X\hat{\beta}\|^2)/(n - p)$ and note that the fact that $(\hat{\beta}, \hat{\sigma}^2)$ is a complete sufficient statistic for the above linear model. Hence, the classical estimators for the parameters β and σ^2 are $\hat{\beta}$ and $\hat{\sigma}^2$, which are the uniformly minimum variance unbiased estimator (UMVUE) in the sense of minimizing mean squared error.

From the Bayesian viewpoint, note that in most cases past experience about the parameters β and σ^2 are often available. Let $f_0(\beta, \sigma^2)$ be the joint prior of β and σ^2 and the loss function be

$$L\left(\hat{\theta}, \theta\right) = \left(\hat{\theta} - \theta\right)' D\left(\hat{\theta} - \theta\right), \tag{1.3}$$

where D is a positive definite matrix and $\hat{\theta}$ denotes the estimate of the vector $\theta = (\beta', \sigma^2)'$. Then, by virtue of the Bayes theorem, the usual Bayes estimators (UBE) for β and σ^2, say $\hat{\beta}_{UB}$ and $\hat{\sigma}^2_{UB}$, can be calculated by

$$\hat{\beta}_{UB} = \int \int \beta g\left(\beta, \sigma^2|y\right) d\beta d\sigma^2,$$

$$\hat{\sigma}^2_{UB} = \int \int \sigma^2 g\left(\beta, \sigma^2|y\right) d\beta d\sigma^2, \tag{1.4}$$

where $g(\beta, \sigma^2|y)$ denotes the conditional joint posterior density of β and σ^2 given Y. However, it is difficult to handle complicated or non-standard integrations. Normally, in these cases approximate Bayes estimators are suggested such as Lindley's approximation and Tierney and Kadane's approximation, see Lindley (1980) and Tierney and Kadane (1986) for details. Simulation-based methods such as the Gibbs sampling procedure and Metropolis method also have emerged in the past twenty years, see Martinez and Martinez (2007). Traditional Bayes estimators (UBE) are somewhat complicated and inconvenient to use in these situations.

In the following, enlightened by Rao (1973), we employ the linear Bayes method to propose a linear Bayes estimator (LBE) for the parameters β and σ^2 simultaneously as well as investigate superiorities. We also extend our discussions on the application of linear Bayes method to some other useful distributions.

The survey is organized as follows: In Section 2 we define the LBE for the parameter vector $\theta = (\beta', \sigma^2)'$ and establish it superiorities over the classic UMVUE and MLE. Numerical comparisons between the LBE and the usual Bayes estimator (UBE) are presented in Section 3. Extended discussions and remarks are made in Section 4.

Throughout this paper, for two nonnegative definite matrices A_1 and A_2 of the same size, we say $A_1 \geq A_2$ if and only if $A_1 - A_2$ is a nonnegative definite matrix.

2. Linear Bayes Estimator and Its Superiorities

2.1. The proposed LBE

Denote $\theta = (\beta', \sigma^2)'$. In what follows we assume that the prior $G(\theta)$ belongs to the distribution family:

$$\mathcal{G} = \left\{ G(\theta) : E\left[\|\beta\|^2 + \left(\sigma^2\right)^2 \right] < \infty \right\}. \tag{2.1}$$

Put $T = (\hat{\beta}', \hat{\sigma}^2)'$ and define the linear Bayes estimator (LBE) of θ, say $\hat{\theta}_{LB}$, be of the form $\tilde{\theta} = BT + b$ satisfying

$$R\left(\hat{\theta}_{LB}, \theta\right) = \min_{B,b} E_{(Y,\theta)} L\left(\tilde{\theta}, \theta\right) \quad \text{and} \quad E_{(Y,\theta)}\left(\hat{\theta}_{LB} - \theta\right) = 0, \tag{2.2}$$

where align and b are $(p+1) \times (p+1)$ and $(p+1) \times 1$ undetermined matrices respectively, $E_{(Y,\theta)}$ denotes the expectation with respect to the joint distribution of Y and θ and the loss function is given by (1.3).

Thus, we have the following conclusion.

Theorem 1. *Let $\hat{\theta}_{LB}$ be defined by (2.2). If $n \geq p + 1$, then*

$$\hat{\theta}_{LB} = T - W[W + Cov(\theta)]^{-1}(T - E\theta),$$

where $W = E[Cov(T|\theta)] = diag((X'X)^{-1}E\sigma^2, 2E\sigma^4/(n-p))$.

Proof: From the constraint $E_{(Y,\theta)}(\tilde{\theta} - \theta) = 0$, we know $b = E\theta - BE_{(Y,\theta)}(T)$. Note that

$$E_{(Y,\theta)}(T) = E[E(T|\theta)] = E(\beta', \sigma^2)' = E\theta. \tag{2.3}$$

Hence $b = E\theta - BE\theta$, and accordingly we have

$$
\begin{aligned}
R\left(\tilde{\theta}, \theta\right) &= E_{(Y,\theta)} L\left(\tilde{\theta}, \theta\right) \\
&= E_{(Y,\theta)}[BT + E\theta - BE\theta - \theta]' D[BT + E\theta - BE\theta - \theta] \\
&= E_{(Y,\theta)} \left\{ \text{tr}\left(D[B(T - E\theta) - (\theta - E\theta)][B(T - E\theta) - (\theta - E\theta)]'\right) \right\} \\
&= \text{tr}\left(DE_{(Y,\theta)}[B(T - E\theta) - (\theta - E\theta)][B(T - E\theta) - (\theta - E\theta)]'\right) \\
&= \text{tr}\left(DBE_{(Y,\theta)}\left[(T - E\theta)(T - E\theta)'\right]B'\right) - \text{tr}\left(DCov(\theta)B'\right) - \text{tr}(DBCov(\theta)) + \text{tr}(DCov(\theta)). \tag{2.4}
\end{aligned}
$$

For given θ, using the independence between $\hat{\beta}$ and $\hat{\sigma}^2$, we have

$$
\begin{aligned}
E_{(Y,\theta)}\left[(T - E\theta)(T - E\theta)'\right] &= E[Cov(T|\theta)] + Cov(E(T|\theta)) \\
&= W + Cov(\theta), \tag{2.5}
\end{aligned}
$$

where $W = diag((X'X)^{-1}E\sigma^2, 2E\sigma^4/(n-p))$.

Substituting (2.5) into (2.4) and letting $\partial R(\tilde{\theta}, \theta)/\partial B$ be zero, we have

$$DB[W + Cov(\theta)] - DCov(\theta) = 0,$$

which yields

$$B = I_{p+1} - W[W + \text{Cov}(\theta)]^{-1}. \qquad (2.6)$$

Together with $b = E\theta - BE\theta$ we come to the conclusion of Theorem 1.   □

Remark 1. In the definition of $\hat{\theta}_{LB}$ (2.2), if we discard the so-called unbiased constraint $E_{(Y,\theta)}(\hat{\theta}_{LB} - \theta) = 0$, then by directly computing $R(\tilde{\theta}, \theta)$ and denoting $\partial R(\tilde{\theta}, \theta)/\partial B = 0$ and $\partial R(\tilde{\theta}, \theta)/\partial b = 0$, we can obtain the same expression for the LBE $\hat{\theta}_{LB}$, which means that $\hat{\theta}_{LB}$ satisfies the unbiased condition as well as performs best among linear Bayes estimators in the sense of minimizing $E_{(Y,\theta)}L(BT + b, \theta)$.

2.2. The superiorities of LBE

Note that

$$\hat{\theta}_U = (\hat{\beta}', \hat{\sigma}^2)' = \begin{pmatrix} I_p & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{\sigma}^2 \end{pmatrix} = T, \qquad (2.7)$$

where we use $\hat{\theta}_U$ to denote the UMVUE of $\theta = (\beta', \sigma^2)'$.

Theorem 2. *Let $\hat{\theta}_{LB}$ and $\hat{\theta}_U$ be given by (2.2) and (2.7) respectively. If $n \geq p+1$, then $\hat{\theta}_{LB}$ is superior to $\hat{\theta}_U$ in terms of MSEM criterion, i.e. $MSEM(\hat{\theta}_{LB}) \leq MSEM(\hat{\theta}_U)$.*

Proof: Since $E_{(Y,\theta)}(\hat{\theta}_{LB} - \theta) = 0$, we have

$$\begin{aligned}
\text{MSEM}\left(\hat{\theta}_{LB}\right) &= E_{(Y,\theta)}\left[\left(\hat{\theta}_{LB} - \theta\right)\left(\hat{\theta}_{LB} - \theta\right)'\right] \\
&= E\left[\text{Cov}\left(\hat{\theta}_{LB} - \theta|\theta\right)\right] + \text{Cov}\left(E\left[\hat{\theta}_{LB} - \theta|\theta\right]\right). \qquad (2.8)
\end{aligned}$$

Denote $M = [W + \text{Cov}(\theta)]^{-1}$. Then by Theorem 1 we know

$$\begin{aligned}
\text{MSEM}\left(\hat{\theta}_{LB}\right) &= (I - WM)W(I - WM)' + WM\text{Cov}(\theta)(WM)' \\
&= (I - WM)W(I - MW) + WM\text{Cov}(\theta)MW \\
&= W - 2WMW + WM[W + \text{Cov}(\theta)]MW \\
&= W - WMW. \qquad (2.9)
\end{aligned}$$

However,

$$\begin{aligned}
\text{MSEM}\left(\hat{\theta}_U\right) &= E_{(Y,\theta)}\left[\left(\hat{\theta}_U - \theta\right)\left(\hat{\theta}_U - \theta\right)'\right] \\
&= E\left\{E\left[\left(\hat{\theta}_U - \theta\right)\left(\hat{\theta}_U - \theta\right)'\Big|\theta\right]\right\} \\
&= E\left\{E\left[(T - \theta)(T - \theta)'|\theta\right]\right\} \\
&= W. \qquad (2.10)
\end{aligned}$$

Comparing (2.9) with (2.10), we have

$$\text{MSEM}\left(\hat{\theta}_{LB}\right) \leq \text{MSEM}\left(\hat{\theta}_U\right). \qquad (2.11)$$

The proof of Theorem 2 is completed.   □

Moreover, note that the MLE of θ, denoted by $\hat{\theta}_{ML}$, equals to $B_0 T$ with

$$B_0 = \begin{pmatrix} I_p & 0 \\ 0 & \frac{n-p}{n} \end{pmatrix}. \tag{2.12}$$

Thus,

$$\text{MSEM}\left(\hat{\theta}_{ML}\right) = B_0 W B_0' + \left(B_0 - I_{p+1}\right) E\left(\theta\theta'\right)\left(B_0 - I_{p+1}\right)'$$

$$= \begin{pmatrix} \frac{X'X}{E\sigma^2} & 0 \\ 0 & \frac{n^2}{(2n+p^2-2p)E\sigma^4} \end{pmatrix}^{-1}. \tag{2.13}$$

Theorem 3. *Let $\hat{\theta}_{LB}$ and $\hat{\theta}_{ML}$ be given by (2.2) and (2.12) respectively. If $n \geq p + 1$, then $\hat{\theta}_{LB}$ is superior to $\hat{\theta}_{ML}$ in terms of MSEM criterion, i.e. MSEM$(\hat{\theta}_{LB}) \leq$ MSEM$(\hat{\theta}_{ML})$.*

Proof: We rewrite

$$\text{MSEM}\left(\hat{\theta}_{LB}\right) = W - WMW = \left[W^{-1} + \text{Cov}^{-1}(\theta)\right]^{-1}$$

$$= \left[\begin{pmatrix} \frac{X'X}{E\sigma^2} & 0 \\ 0 & \frac{n^2}{(2n+p^2-2p)E\sigma^4} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & c_0 \end{pmatrix} + \text{Cov}^{-1}(\theta)\right]^{-1}, \tag{2.14}$$

where $c_0 = (np^2 - 4np - p^3 + 2p^2)/\{2(2n + p^2 - 2p)E\sigma^4\}$.

Hence, in order to establish the MSEM superiority of $\hat{\theta}_{LB}$ over $\hat{\theta}_{ML}$, it suffices to show that

$$\begin{pmatrix} 0 & 0 \\ 0 & c_0 \end{pmatrix} + \text{Cov}^{-1}(\theta) \geq 0, \tag{2.15}$$

for $n \geq p + 1$.

Denote $\text{Cov}^{-1}(\theta) = S$, where $S = (S^{ij})$ is a 2×2 partition matrix and

$$S^{11} = \text{Cov}^{-1}(\beta) + \text{Cov}^{-1}(\beta)E\left(\sigma^2 - E\sigma^2\right)(\beta - E\beta)S^{22}E\left(\sigma^2 - E\sigma^2\right)(\beta - E\beta)'\text{Cov}^{-1}(\beta),$$

$$S^{12} = -\text{Cov}^{-1}(\beta)E\left(\sigma^2 - E\sigma^2\right)(\beta - E\beta)S^{22},$$

$$S^{21} = -S^{22}E\left(\sigma^2 - E\sigma^2\right)(\beta - E\beta)'\text{Cov}^{-1}(\beta),$$

$$S^{22} = \left[\text{Var}\left(\sigma^2\right) - E\left(\sigma^2 - E\sigma^2\right)(\beta - E\beta)'\text{Cov}^{-1}(\beta)E\left(\sigma^2 - E\sigma^2\right)(\beta - E\beta)\right]^{-1}. \tag{2.16}$$

Thus, to prove (2.15), it is adequate to show that

$$\left|S^{11}\right|\left|c_0 + S^{22} - S^{21}\left(S^{11}\right)^{-1}S^{12}\right| \geq 0, \tag{2.17}$$

or equivalently to show that

$$c_0 + S^{22} - S^{21}\left(S^{11}\right)^{-1}S^{12} \geq 0, \tag{2.18}$$

for $n \geq p + 1$, where we use the fact of $S^{11} \geq 0$ and accordingly the value of the determinant $|S^{11}|$ is nonnegative.

Set $\Delta = \mathrm{Var}(\sigma^2) - E(\sigma^2 - E\sigma^2)(\beta - E\beta)'\mathrm{Cov}^{-1}(\beta)E(\sigma^2 - E\sigma^2)(\beta - E\beta)$ and note that $S^{11} = \mathrm{Cov}^{-1}(\beta)[\mathrm{Cov}(\beta) + E(\sigma^2 - E\sigma^2)(\beta - E\beta)(1/\Delta)E(\sigma^2 - E\sigma^2)(\beta - E\beta)']\mathrm{Cov}^{-1}(\beta)$, hence we have

$$c_0 + S^{22} - S^{21}\left(S^{11}\right)^{-1}S^{12} = c_0 + \frac{1}{\Delta} - \frac{1}{\Delta}E\left(\sigma^2 - E\sigma^2\right)(\beta - E\beta)'$$
$$\times\left[\mathrm{Cov}(\beta) + E\left(\sigma^2 - E\sigma^2\right)(\beta - E\beta)\frac{1}{\Delta}E\left(\sigma^2 - E\sigma^2\right)(\beta - E\beta)'\right]^{-1}$$
$$\times E\left(\sigma^2 - E\sigma^2\right)(\beta - E\beta)\frac{1}{\Delta}. \tag{2.19}$$

Further, using $[\Sigma + A\Sigma_1 A']^{-1} = \Sigma^{-1} - \Sigma^{-1}A[A'\Sigma^{-1}A + \Sigma_1^{-1}]^{-1}A'\Sigma^{-1}$, we have

$$c_0 + S^{22} - S^{21}\left(S^{11}\right)^{-1}S^{12} = c_0 + \frac{1}{\Delta} - \frac{a}{\Delta^2} + \frac{a^2}{(\Delta + a)\Delta^2}, \tag{2.20}$$

where $a = E(\sigma^2 - E\sigma^2)(\beta - E\beta)'\mathrm{Cov}^{-1}(\beta)E(\sigma^2 - E\sigma^2)(\beta - E\beta)$.

Note that $\Delta = \mathrm{Var}(\sigma^2) - a$, hence

$$\frac{1}{\Delta} - \frac{a}{\Delta^2} + \frac{a^2}{(\Delta + a)\Delta^2} = \frac{1}{\mathrm{Var}(\sigma^2)}, \tag{2.21}$$

and accordingly

$$c_0 + S^{22} - S^{21}\left(S^{11}\right)^{-1}S^{12}$$
$$= c_0 + \frac{1}{\mathrm{Var}\left(\sigma^2\right)}$$
$$= \frac{\left[(n+4)p^2 - (4n+4)p - p^3 + 4n\right]E\sigma^4 + \left[4np - (n+2)p^2 + p^3\right]\left(E\sigma^2\right)^2}{2\left(2n + p^2 - 2p\right)\mathrm{Var}\left(\sigma^2\right)E\sigma^4}$$
$$\geq \frac{\left[2p^2 + 4n - 4p\right]\left(E\sigma^2\right)^2}{2\left(2n + p^2 - 2p\right)\mathrm{Var}\left(\sigma^2\right)E\sigma^4} > 0, \quad \text{for } n \geq p+1, \tag{2.22}$$

where we use the facts that $E\sigma^4 \geq (E\sigma^2)^2$ and $(n+4)p^2 - (4n+4)p - p^3 + 4n \geq 0$ for $n \geq p+1$.

Hence, Theorem 3 has been proved. $\square$

Remark 2. For the two-parameter exponential family given by

$$f(x; \mu, \lambda) = \lambda^{-1}\exp\left(\frac{x - \mu}{\lambda}\right),$$

where $x > \mu$, we assume that $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(r)}(2 \leq r \leq n)$ denote the type II censored samples. Define $Q_i = [n - (i-1)](X_{(i)} - X_{(i-1)})$, where $X_{(0)} = 0$, then Q_1 and $P = \sum_{i=2}^{r}Q_i$ are mutually independent and also (Q_1, P) is sufficient for the parameter vector (μ, λ). Set $T = (Q_1, P)'$, the classical UMVUE and MLE for $\theta = (\mu, \lambda)'$ can be defined as follows

$$\hat{\theta}_U = \begin{pmatrix} \frac{1}{n} & \frac{-1}{n(r-1)} \\ 0 & \frac{1}{r-1} \end{pmatrix}T, \qquad \hat{\theta}_{ML} = \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{r} \end{pmatrix}T.$$

Under the assumption that the prior $G(\theta)$ satisfies the condition $E\|\theta\|^2 < \infty$, we can obtain the expression of LBE $\hat{\theta}_{LB}$ for the parameter vector $\theta = (\mu, \lambda)'$ in this case and establish its superiorities over the $\hat{\theta}_U$ and $\hat{\theta}_{ML}$ by virtue of MSEM criterion similarly. The interested readers are referred to Wang and Singh (2014) for more details.

Remark 3. Let $X_1, X_2, \ldots, X_n$ be independently drawn from the uniform distribution $U(\theta_1, \theta_2)$ with density $f(x; \theta_1, \theta_2) = (\theta_2 - \theta_1)^{-1}$, where $\theta_1 < x < \theta_2$. Note that $X_{(1)} = \min_{1 \le i \le n} X_i$ and $X_{(n)} = \max_{1 \le i \le n} X_i$ are sufficient and complete statistics, hence, set $T = (X_{(1)}, X_{(n)})'$, we obtain the classic UMVUE and MLE for $\theta = (\theta_1, \theta_2)'$ in this case:

$$\hat{\theta}_U = \begin{pmatrix} \frac{n}{n-1} & \frac{-1}{n-1} \\ \frac{-1}{n-1} & \frac{n}{n-1} \end{pmatrix} T, \qquad \hat{\theta}_{ML} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} T.$$

Similarly, using the assumption that the prior $G(\theta)$ satisfies the condition $E\|\theta\|^2 < \infty$, the expression of LBE $\hat{\theta}_{LB}$ for the parameter vector $\theta = (\theta_1, \theta_2)'$ can be easily obtained and its MSEM superiorities over the $\hat{\theta}_U$ and $\hat{\theta}_{ML}$ can also be proved.

Remark 4. Let $X_1, X_2, \ldots, X_n$ be a random sample from the two-parameter inverse Gaussian distribution $IG(\alpha_1, \alpha_2)$ with pdf

$$f(x; \alpha_1, \alpha_2) = \left(\frac{\alpha_2}{2\pi x^3} \right)^{\frac{1}{2}} \exp\left(-\frac{\alpha_2(x - \alpha_1)^2}{2\alpha_1^2 x} \right),$$

where $x > 0$. It is easily shown that the statistics $\bar{X} = (1/n)\sum_{i=1}^n X_i$ and $\bar{S} = n/\{\sum_{i=1}^n (1/X_i - 1/\bar{X})\}$ are sufficient and complete. Tweedie (1957) shows that $\bar{X}$ and $\bar{S}$ are independent, $\bar{X}$ having an inverse Gaussian distribution with parameters α_1 and $n\alpha_2$, and $n\alpha_2/\bar{S}$ having a χ_{n-1}^2 distribution. Schwarz and Samanta (1991) gives a proof of these facts using an inductive argument. Hence we obtain the classic UMVUE and MLE for the parameter $\theta = (\alpha_1, \alpha_2)'$ as follows

$$\hat{\theta}_U = \begin{pmatrix} 1 & 0 \\ 0 & \frac{n-3}{n} \end{pmatrix} T, \qquad \hat{\theta}_{ML} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} T,$$

where $T = (\bar{X}, \bar{S})'$. Assume that the prior $G(\theta)$ belongs to the prior family $\mathcal{G} = \{G(\theta) : E[\alpha_1^2 + \alpha_2^2] < \infty, E[\alpha_1^3 \alpha_2^{-1}] < \infty\}$, we can obtain the expression of LBE $\hat{\theta}_{LB}$ for the parameter vector $\theta = (\alpha_1, \alpha_2)'$ and prove that it prevails over the classic UMVUE and MLE under MSEM criterion.

2.3. An illustration example

To illustrate Theorem 2 and Theorem 3 we investigate the case of two-dimensional normal linear model, i.e.

$$Y \sim N\left(\beta_0 + \beta_1 x, \sigma^2 I_n\right), \tag{2.23}$$

where we assume that $(\beta_0, \beta_1)' \sim N((1, 2)', \text{Cov}(\beta_0, \beta_1))$ with $\text{Cov}(\beta_0, \beta_1)$ having three alternative values and $\sigma^2 \sim U(a, b)$ with three different pairs a and b. We also assume that (β_0, β_1) and σ^2 are uncorrelated, i.e. $\text{Cov}(\beta_0, \beta_1, \sigma^2) = \text{diag}(\text{Cov}(\beta_0, \beta_1), \text{Var}(\sigma^2))$, and $x = (-4, -3, -2, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16)$.

Table 1: $\text{IP}(\hat{\theta}_U)$ and $\text{IP}(\hat{\theta}_{ML})$ under different priors and sample sizes

Priors	n	$\text{IP}(\hat{\theta}_U)$	$\text{IP}(\hat{\theta}_{ML})$	$\text{tr}\left(\text{Cov}\left(\beta_0, \beta_1, \sigma^2\right)\right)$
$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim N\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & \frac{1}{3} \\ \frac{1}{3} & 1 \end{pmatrix}\right)$	5	0.9712	0.9542	
	10	0.9516	0.9404	7/3
$\sigma^2 \sim U(7, 9)$	20	0.9066	0.8974	
$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim N\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 6 & 2 \\ 2 & 6 \end{pmatrix}\right)$	5	0.9234	0.8779	
	10	0.8803	0.8524	40/3
$\sigma^2 \sim U(6, 10)$	20	0.7706	0.7480	
$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim N\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 18 & 6 \\ 6 & 18 \end{pmatrix}\right)$	5	0.8458	0.7537	
	10	0.7265	0.6627	124/3
$\sigma^2 \sim U(4, 12)$	20	0.5369	0.4961	

Define the percentages of improvement of $\hat{\theta}_{LB}$ over $\hat{\theta}_U$ and $\hat{\theta}_{ML}$, respectively, by

$$\text{IP}\left(\hat{\theta}_U\right) = \frac{\text{tr}\left(\text{MSEM}\left(\hat{\theta}_U\right) - \text{MSEM}\left(\hat{\theta}_{LB}\right)\right)}{\text{tr}\left(\text{MSEM}\left(\hat{\theta}_U\right)\right)} \quad \text{and} \quad \text{IP}\left(\hat{\theta}_{ML}\right) = \frac{\text{tr}\left(\text{MSEM}\left(\hat{\theta}_{ML}\right) - \text{MSEM}\left(\hat{\theta}_{LB}\right)\right)}{\text{tr}\left(\text{MSEM}\left(\hat{\theta}_{ML}\right)\right)}.$$

For different sample size n (= 5, 10, 20), the corresponding computation results for $\text{IP}(\hat{\theta}_U)$ and $\text{IP}(\hat{\theta}_{ML})$ under the three different priors are presented in Table 1, where $\text{tr}(\text{Cov}(\beta_0, \beta_1, \sigma^2))$ is used as an index of the variation of the prior information.

As stated in Theorems 2 and 3, since the above priors belong to the family (2.1), both $\text{MSEM}(\hat{\theta}_U) - \text{MSEM}(\hat{\theta}_{LB})$ and $\text{MSEM}(\hat{\theta}_{ML}) - \text{MSEM}(\hat{\theta}_{LB})$ are always nonnegative definite. From Table 1, firstly we can see that when the sample size n is fixed, as expected, both $\text{IP}(\hat{\theta}_U)$ and $\text{IP}(\hat{\theta}_{ML})$ decrease as the variation of the prior increases (i.e, $\text{tr}(\text{Cov}(\beta_0, \beta_1, \sigma^2))$ tends to be larger); secondly, for the same prior information, it is only natural that as the sample size n grows, which means that the sample information gets more, both $\text{IP}(\hat{\theta}_U)$ and $\text{IP}(\hat{\theta}_{ML})$ decrease; finally, it seems that $\text{IP}(\hat{\theta}_U)$ is larger than $\text{IP}(\hat{\theta}_{ML})$, the reason may be due to $\text{MSEM}(\hat{\theta}_U) \geq \text{MSEM}(\hat{\theta}_{ML})$ for our case.

3. Numerical Comparisons between LBE and UBE

For the model (1.2), note that under the loss $L(\hat{\theta}, \theta)$ and the prior $G(\theta)$, the usual Bayes estimator (UBE) of θ, say $\hat{\theta}_{UB}$, would be equal to $E(\theta|Y)$. In this Section, for given priors $G(\theta)$, we present some numerical comparisons between the LBE $\hat{\theta}_{LB}$ and the UBE $\hat{\theta}_{UB}$, the latter is calculated by employing an MCMC sampling method.

Suppose $p = 2$ and let us consider the normal linear model

$$Y \sim N\left(\gamma_0 + \gamma_1 x, \sigma^2 I_n\right). \tag{3.1}$$

Denote $\beta = (\gamma_0, \gamma_1)'$. We assume that σ^2 follows an inverse-Gamma distribution with density

$$\pi\left(\sigma^2; \lambda_0, t\right) = \frac{t^{\lambda_0 - 1}}{\Gamma(\lambda_0 - 1)} \left(\frac{1}{\sigma^2}\right)^{\lambda_0} \exp\left(-\frac{t}{\sigma^2}\right)$$

and given σ^2 the conditional distribution of β is $N_2(\tilde{\beta}_0, \sigma^2 \Sigma_0)$.

Note that the posterior density of (β, σ^2) given Y is

$$f\left(\beta, \sigma^2|y\right) \propto \left(\sigma^{-2}\right)^{\lambda_0 + 1 + \frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[2\lambda_0 + (n-2)\hat{\sigma}^2 + \left(\beta - \tilde{\beta}_0\right)' \Sigma_0^{-1} \left(\beta - \tilde{\beta}_0\right) + \left(\beta - \hat{\beta}\right)' X'X \left(\beta - \hat{\beta}\right)\right]\right\},$$

where $X' = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix}$ denotes the design matrix.

However, it is almost impossible to calculate $\hat{\theta}_{UB} = (\hat{\beta}'_{UB}, \hat{\sigma}^2_{UB})'$ analytically, so we have to obtain it numerically.

Note that the posterior conditional densities of β given σ^2 and σ^2 given β are respectively proportional to

$$f_1\left(\beta|\sigma^2, y\right) \propto \exp\left\{-\frac{1}{2\sigma^2}\left(\beta - \tilde{\beta}\right)' \tilde{\Sigma}^{-1}\left(\beta - \tilde{\beta}\right)\right\}, \tag{3.2}$$

$$f_2\left(\sigma^2|\beta, y\right) \propto \sigma^{-2\left(\lambda_0 + 1 + \frac{n}{2}\right)} \exp\left(-\frac{c_1}{2\sigma^2}\right), \tag{3.3}$$

where $\tilde{\beta} = (\Sigma_0^{-1} + X'X)^{-1}(\Sigma_0^{-1}\tilde{\beta}_0 + X'X\hat{\beta})$, $\tilde{\Sigma} = \sigma^2(\Sigma_0^{-1} + X'X)^{-1}$ and $c_1 = 2\lambda_0 + (n-2)\hat{\sigma}^2 + (\beta - \tilde{\beta}_0)'\Sigma_0^{-1}(\beta - \tilde{\beta}_0) + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})$.

The Gibbs sampler was originally developed by Geman and Geman (1984) as applied to image processing and the analysis of Gibbs distributions on a lattice. It is brought into mainstream statistics through the articles of Gelfand and Smith (1990) and Gelfand *et al.* (1990). The Gibbs sampler can also be shown to be a special case of the Metropolis-Hastings algorithm, see Gilks *et al.* (1996) and Robert and Casella (1999). In describing the Gibbs sampler, we follow the treatment in Casella and George (1992).

Step 1.   Choose the initial values of β and σ^2 and denote the values of β and σ^2 at the jth step by β_j and σ^2_j, respectively.

Step 2.   Generate β_{j+1} and σ^2_{j+1} from $f_1(\beta|\sigma^2_j, y)$ and $f_2(\sigma^2|\beta_j, y)$, respectively.

Step 3.   Repeat Step 2 for N times.

Step 4.   Calculate the Bayes estimator of $l(\beta, \sigma^2)$ by $1/(N - m_0) \sum_{j=m_0+1}^{N} l(\beta_j, \sigma^2_j)$, where $l(\beta, \sigma^2)$ denotes any a function of β and σ^2 and m_0 is the burn-in period.

Note that under the above priors, it is readily seen that $E\theta = (E\beta', E\sigma^2)' = (\tilde{\beta}'_0, t/[\lambda_0 - 2])'$ and $\text{Cov}(\theta) = \text{diag}(t/(\lambda_0 - 2) \cdot \Sigma_0, t^2/\{(\lambda_0 - 2)^2(\lambda_0 - 3)\})$ since $\text{Cov}(\beta) = E\text{Cov}(\beta|\sigma^2) = \Sigma_0 E\sigma^2$ and $\text{Corr}(\beta, \sigma^2) = 0$.

In the following table we first calculate the values of $\hat{\theta}_{LB}$ and the corresponding numerical results of $\hat{\theta}_{UB}$ for different prior parameters and then present $\|\hat{\theta}_{LB} - \hat{\theta}_{UB}\| = \sqrt{\|\hat{\beta}_{LB} - \hat{\beta}_{UB}\|^2 + (\hat{\sigma}^2_{LB} - \hat{\sigma}^2_{UB})^2}$, which is defined as an index of degree of approximation between $\hat{\theta}_{LB}$ and $\hat{\theta}_{UB}$.

The above numerical comparisons indicate two trends, one is that for the same prior, $\|\hat{\theta}_{LB} - \hat{\theta}_{UB}\|$ tends to be smaller as sample size gets larger, the other is that given sample size, $\|\hat{\theta}_{LB} - \hat{\theta}_{UB}\|$ increases as the prior variance becomes larger. In the process of simulation, we find that the value of $\|\hat{\theta}_{LB} - \hat{\theta}_{UB}\|$ is affected by the value of $(\hat{\sigma}^2_{LB} - \hat{\sigma}^2_{UB})^2$; however, the value of $\|\hat{\beta}_{LB} - \hat{\beta}_{UB}\|^2$ is always small, which means the LBE $\hat{\beta}_{LB}$ is rather close to the UBE $\hat{\beta}_{UB}$ and there could be a certain difference between the LBE $\hat{\sigma}^2_{LB}$ and the UBE $\hat{\sigma}^2_{UB}$ for our cases.

Remark 5. Two cases are considered for the two-parameter exponential family. In case (I) we assume that the parameters μ and λ have independent prior distributions, where μ follows an exponential distribution and λ has an inverted Gamma prior. In case (II), we suppose that, given λ, the conditional prior of μ is an inverted Gamma density and λ follows an inverted Gamma prior. For the

Table 2: $\|\hat{\theta}_{LB} - \hat{\theta}_{UB}\|$ under different prior parameters and sample sizes

n	The prior parameters	$\|\hat{\theta}_{LB} - \hat{\theta}_{UB}\|$
	$\lambda_0 = 2,\ t = 3,\ \tilde{\beta}_0 = \begin{pmatrix} -2 \\ 3 \end{pmatrix},\ \Sigma_0 = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$	1.3369
20	$\lambda_0 = 2,\ t = 4,\ \tilde{\beta}_0 = \begin{pmatrix} -2 \\ 3 \end{pmatrix},\ \Sigma_0 = \begin{pmatrix} 5 & 3.5 \\ 3.5 & 5 \end{pmatrix}$	1.4472
	$\lambda_0 = 2,\ t = 8,\ \tilde{\beta}_0 = \begin{pmatrix} -2 \\ 3 \end{pmatrix},\ \Sigma_0 = \begin{pmatrix} 20 & 14 \\ 14 & 20 \end{pmatrix}$	1.8390
	$\lambda_0 = 2,\ t = 3,\ \tilde{\beta}_0 = \begin{pmatrix} -2 \\ 3 \end{pmatrix},\ \Sigma_0 = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$	1.0589
50	$\lambda_0 = 2,\ t = 4,\ \tilde{\beta}_0 = \begin{pmatrix} -2 \\ 3 \end{pmatrix},\ \Sigma_0 = \begin{pmatrix} 5 & 3.5 \\ 3.5 & 5 \end{pmatrix}$	1.1407
	$\lambda_0 = 2,\ t = 8,\ \tilde{\beta}_0 = \begin{pmatrix} -2 \\ 3 \end{pmatrix},\ \Sigma_0 = \begin{pmatrix} 20 & 14 \\ 14 & 20 \end{pmatrix}$	1.4855

Where $(x_1, x_2, \ldots, x_{20})$ is a subset of $(x_1, x_2, \ldots, x_{50}) = (-4\ \ 3.4\ \ 2.4\ \ 0\ \ 1\ \ 2\ \ 3\ \ 4\ \ 3.5\ \ 0.6\ \ 7\ \ 8\ \ 9\ \ 10\ \ 11\ \ 12\ \ -1.9\ \ 14\ \ 15$
$16\ \ 17\ \ 18\ \ 19\ \ 21\ \ 22\ \ 23\ \ 34\ \ 35\ \ -13\ \ 17\ \ 18.5\ \ 19.9\ \ 24\ \ 28\ \ 32\ \ 33\ \ 37\ \ 39\ \ -12\ \ -16\ \ -19\ \ 44\ \ 45\ \ 23.4\ \ 31.7\ \ 33.5\ \ 45.2$
$60.7\ \ -14.3\ \ -17)$.

above two cases, numerical simulations show that $\|\hat{\theta}_{LB} - \hat{\theta}_{UB}\|$s are small, which means that as a linear approximation of $\hat{\theta}_{UB}$, $\hat{\theta}_{LB}$ works better.

Remark 6. In the case of the uniform distribution $U(\theta_1, \theta_2)$, numerical computations show that $\hat{\theta}_{LB}$ works very good for both independent prior and non-independent prior. For example, for the single parameter uniform distribution $U(0, \theta_2)$, we assume the prior $\pi(\theta_2)$ has finite second-order moment and mimic the above discussions, then the LBE for the parameter θ_2 is $\hat{\theta}_{2,LB} = a_0 X_{(n)} + b_0$ with

$$a_0 = \frac{(n+1)(n+2)\mathrm{Var}(\theta_2)}{(n+1)^2 E\theta_2^2 - n(n+2)(E\theta_2)^2} \quad \text{and} \quad b_0 = \frac{[(1-a_0)n + 1]E\theta_2}{n+1}.$$

Specifically, let $\pi(\theta_2) = t_2^{t_1} \theta_2^{-t_1-1} \exp(-t_2/\theta_2)/\Gamma(t_1)$ and together with $f(x_{(n)}|\theta_2)$ and the squared loss, we know that the UBE $\hat{\theta}_{2,UB}$ is

$$E(\theta_2|x_{(n)}) = \frac{\int_{x_{(n)}}^{\infty} \theta_2^{-t_1-n} \exp(-t_2/\theta_2)\, d\theta_2}{\int_{x_{(n)}}^{\infty} \theta_2^{-t_1-n-1} \exp(-t_2/\theta_2)\, d\theta_2} = \frac{t_2}{t_1+n-1} \frac{P\left(\chi^2(2(t_1+n-1)) \leq 2t_2/x_{(n)}\right)}{P\left(\chi^2(2(t_1+n)) \leq 2t_2/x_{(n)}\right)},$$

where we utilize the relationship between the inverse Gamma and the χ^2 distribution (Mao and Tang, 2012). Say, let $n = 5$, $x_{(n)} = 2$ and $t_1 = 3$ and $t_2 = 8$, simple computations show that $a_0 = 1.1351$, $b_0 = 0.2163$ and $P(\chi^2(14) \leq 8) = 0.1107$ and $P(\chi^2(16) \leq 8) = 0.0511$. Hence, we have $\hat{\theta}_{2,LB} = 2.4865$ and $\hat{\theta}_{2,UB} = 2.4758$, which show that the LBE is very close to the UBE.

Remark 7. For the two-parameter inverse Gaussian distribution $IG(\alpha_1, \alpha_2)$, similarly, numerical studies show that LBE is adequate.

Remark 8. In above simulation, it should be noted that the problem of deciding when to stop the chain is an important issue and is the topic of current research in MCMC methods. If the resulting sequence has not converged to the target distribution, then the estimators and inferences we get from it are suspect. Let γ represent the characteristic of the target distribution (mean, moments, quantiles, etc.) in which we are interested. An obvious method to monitor convergence to target distribution is to run multiple sequences of the chain and plot γ versus the iteration number.

4. Conclusions and Remarks

This paper uses the normal linear model $Y \sim N(X\beta, \sigma^2 I_n)$ as an example to investigate the application of linear Bayes method, where we employ the linear Bayes method to simultaneously estimate regression parameter β and the variance parameter σ^2 as well as propose a linear Bayes estimator for the parameter vector $\theta = (\beta', \sigma^2)'$. The proposed linear Bayes estimator is shown superior to the classical estimators UMVUE and MLE, respectively, in terms of the mean squared error matrix criterion. Numerical simulations are presented to verify the validity of the linear Bayes estimator. The procedure used in this paper includes normal distribution as its special case and can be extended easily to other useful distributions (such as log-normal, inverse Gaussian distribution and two-parameter exponential family), which are frequently used parametric lifetime models in survival analysis and reliability theory. We also discuss and remark the applications of linear Bayes method to the two-parameter exponential family, uniform distribution and the inverse Gaussian distribution. Compared with the usual Bayes estimator, we find that

(1) The proposed linear Bayes estimator is simple and easy to calculate as well as a good approximation in many situations; the linear Bayes method works especially well for the case of uniform distribution.

(2) We can always define a linear Bayes estimator if there exists sufficient statistic for the parametric model; subsequently, the conclusions of Theorems 2 and 3 always hold.

However, an advantage of the usual Bayes estimator over the linear Bayes estimator is that the former allows for noninformative (improper) priors. Of note is that the linear Bayes estimator may be an inadequate approximation in some situations even for the cases of proper priors. Hence there is still scope for the linear Bayes method to be improved. For instance, for some cases a quadratic Bayes estimator would be a better alternative. However, for the case of normal linear model, one can consider to add more other statistics into the definition of T, for example, we can replace $T = (\hat{\beta}', \hat{\sigma}^2)'$ by $T_1 = (\hat{\beta}', \|\hat{\beta}\|^2, \hat{\sigma}^2)'$ or $T_2 = (\hat{\beta}', \hat{\sigma}^2, \hat{\sigma}^4)'$ to redefine a new linear Bayes estimator. We note that the loss function often plays an important role in Bayesian analysis; consequently, some interesting loss functions such as the balanced loss and the linex loss can be integrated with the linear Bayes method in future studies.

Acknowledgements

References

Arnold, S. F. (1980). *The Theory of Linear Models and Multivariate Analysis,* John Wiley & Sons, New York.

Busby, D., Farmer, C. L. and Iske, A. (2005). Uncertainty evaluation in reservoir forecasting by Bayes linear methodology, Algorithms for approximation, proceedings of the 5th international conference, Chester, 187–196.

Casella, G. and George, E. I. (1992). An introduction to Gibbs Sampling, *The American Statistician*, **46**, 167–174.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, **85**, 398–409.

Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling, *Journal of the American Statistical Association*, **85**, 972–985.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions PAMI*, **6**, 721–741.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

Goldstein, M. (1983). General variance modifications for linear Bayes estimators, *Journal of the American Statistical Association*, **78**, 616–618.

Hartigan, J. A. (1969). Linear Bayesian methods, *Journal of the Royal Statistical Society, Series B*, **31**, 440–454.

Heiligers, B. (1993). Linear Bayes and minimax estimation in linear models with partially restricted parameter space, *Journal of Statistical Planning and Inference*, **36**, 175–183.

Hoffmann, K. (1996). A subclass of Bayes linear estimators that are minimax, *Acta Applicandue Mathematicae*, **43**, 87–95.

Lamotte, L. R. (1978). Bayes linear estimators, *Technometrics*, **3**, 281–290.

Lindley, D. V. (1980). Approximate Bayesian methods, *Trabajos de Estadistica*, **21**, 223–237.

Mao, S. S. and Tang, Y. C. (2012). *Bayesian Statistics*, Second Edition, China Statistics Press, Beijing.

Martinez, W. L. and Martinez, A. R. (2007). *Computational Statistics Handbook with MATLAB*, Second Edition, Chapman & Hall/CRC, New York.

Pensky, M. and Ni, P. (2000). Extended linear empirical Bayes estimation, *Communications in Statistics - Theory and Methods*, **29**, 579–592.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Second Edition, John Wiley & Sons, New York.

Robert, C. P. and Casella, G.(1999). *Monte Carlo Statistical Methods*, Springer-Verlag, New York.

Samaniego, F. J. and Vestrup, E. (1999). On improving standard estimators via linear empirical Bayes methods, *Statistics & Probability Letters*, **44**, 309–318.

Schwarz, C. J. and Samanta, M. (1991). An inductive proof of the sampling distributions for the MLEs of the parameters in an inverse Gaussian distributions, *The American Statistician*, **45**, 223–235.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association*, **81**, 82–86.

Tweedie, M. C. K. (1957). Statistical properties of inverse Gaussian distributions, *The Annals of Mathematical Statistics*, **28**, 362–377.

Wang, L. C. and Singh, R. S. (2014). Linear Bayes estimator for the two-parameter exponential family under type II censoring, *Computational Statistics and Data Analysis*, **71**, 633–642.

Wei, L. S. and Zhang, W. P. (2007). The superiorities of Bayes linear minimum risk estimation in linear model, *Communications in Statistics-Theory and Methods*, **36**, 917–926.

Zhang, W. P. and Wei, L. S. (2005). On Bayes linear unbiased estimation of estimable functions for the singular linear model, *Science in China Series A Mathematics*, **7**, 898–903.

Zhang, W. P., Wei, L. S. and Chen, Y. (2011). The superiorities of Bayes linear unbiased estimation in partitioned linear model, *Journal of System Science and Complexity*, **5**, 945–954.

Zhang, W. P., Wei, L. S. and Chen, Y. (2012). The superiorities of Bayes linear unbiased estimation in multivariate linear models, *Acta Mathematicae Applicatae Sinica*, English Series 2, 383–394.