COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2022, VOL. 51, NO. 6, 1531–1548
https://doi.org/10.1080/03610926.2021.1890125

REVIEW ARTICLE

Linear approximate Bayes estimator for regression parameter with an inequality constraint

Jie Jiang[a], Lichun Wang[a], and Liqun Wang[b]

[a]Department of Mathematics, Beijing Jiaotong University, Beijing, China; [b]Department of Statistics, University of Manitoba, Manitoba, Canada

ABSTRACT
In this paper, a linear Bayes procedure is suggested to estimate the regression parameter of the linear model with an inequality constraint. The superiority of the proposed linear approximate Bayes estimator (LABE) over the inequality constrained least square estimator (CLSE) is investigated in terms of the mean square error matrix (MSEM) criterion. Also, the simulation results and a numerical example show that the LABE is a good approximation to the usual Bayes estimator (BE).

ARTICLE HISTORY
Received 24 August 2020
Accepted 9 February 2021

KEYWORDS
Linear approximate Bayes estimator; inequality constraint; mean square error matrix; Bayes estimator

1. Introduction

Linear model is a kind of common model and representative model in statistical model. Many phenomena in the fields of medicine, biology, economy, finance, engineering and so on can be described approximately by linear models, which is one of the most widely used models in modern statistics. The research on its parameter estimation without constraints has been very mature. However, sometimes simple linear models can not describe the relationship between variables very well. In many cases, variables themselves are constrained. For instance, in applied econometrics, the hypothesis testing problem and other scientific applications, the linear regression model with inequality constraints is often involved.

Let us consider the following inequality constrained linear model

$$y = X\beta + \varepsilon, \quad E(\varepsilon) = 0, \quad Cov(\varepsilon) = \sigma^2 I_n, \tag{1.1}$$

$$\text{and} \qquad c'\beta \geq d, \tag{1.2}$$

where y is an observation vector of $n \times 1$, X is an $n \times p$ full column rank matrix, β is the regression parameter, ε denotes a random error vector, c is a known column vector of $p \times 1$ and d is a scalar.

In fact, a large number of literatures have studied the regression parameter in linear model with inequality constraints. Chong (1976) applies Dantzig-Cottle algorithm to obtain an inequality constrained least square estimator and discusses its statistical properties by simulations. Lius and Bradley (1984) gives the closed form of the inequality constrained least square estimator by using the Kuhn-Tucker condition. Hans (1990)

CONTACT Lichun Wang wlc@amss.ac.cn Department of Mathematics, Beijing Jiaotong University, Beijing, China.

studies the generalized least square estimator with inequality constraints and obtains the closed form of the estimator. Under the balanced loss function, Alan (1994) calculates the risk function of the inequality constrained least square estimator and the pretest estimator. Further, many researchers have improved the inequality constrained least square estimator and obtain some new inequality constrained estimators. For example, Geroge and Thomas (1984) derives the expression of a Stein-like inequality estimator and gives the sufficient condition, which is the risk function of the Stein-like inequality constrained estimator to be less than that of the inequality maximum likelihood restricted estimator. Alan and Kazuhiro (2000) investigates the adaptive least mean square error estimator with an inequality constraint and shows numerical calculations of the risks. Following the idea of Chong (1976), Selma, Gulesen, and Selahattin (2013) constructs the expression of inequality constrained ridge regression estimator and exhibits its superiorities by simulations.

From a Bayesian perspective, Zhang, Wei, and Yang (2005) derives an empirical Bayes estimator of the estimable functions of the regression parameter in normal linear model. Ma (2008) obtains the least square kernel estimator and the best Bayes estimator in the partial linear regression model with inequality constraints by using the optimization technique and Bayesian method. Since the linear Bayesian thought is put forward by Rao (1973), many Bayesian scholars have developed and extended this method to many fields. Lamotte (1978) obtains a class of the Bayes linear estimator by searching among all linear estimators. Goldstein (1983) discusses the problem of modifying the linear Bayes estimator by using an estimate of sample variance. Samanigo and Vetrup (1999) constructs the linear empirical Bayes estimators and establishs their superiorities over the standard and traditional estimators. Recently, Qiu, Luo, and Zheng (2014) compares the performances of Bayes linear unbiased estimator under different loss functions. On the other hand, Wang and Singh (2014) applies the linear Bayes method to the problem of parameter estimation in the two-parameter distributions. However, the above papers limit the application of the linear Bayes method to the case of parameters without constraints. In this paper, we focus on how to apply the linear Bayesian method to the linear model with an inequality constraint, and construct a linear approximate Bayes estimator (LABE) of the regression parameter. It is shown that the proposed LABE is superior to the inequality constrained least square estimator (CLSE).

The paper is organized as follows. In Section 2, the expression of the proposed LABE is obtained by minimizing the Bayes risk. In Section 3, its superiority is exhibited under the MSEM criterion. In Section 4, numerical simulations are carried out to compare the proposed LABE with the inequality CLSE. A numerical example is discussed in Section 5. Finally, we make some conclusions and remarks. Some proofs are given in Appendix.

2. Linear approximate bayes estimator

Suppose that the prior distribution of the parameter β in (1.1) is $G(\beta)$ and the likelihood function is $f(y|\beta)$. Then the posterior density of β would be

$$dH(\beta|y) \propto f(y|\beta)I(c'\beta \geq d)dG(\beta),$$

where $I(\cdot)$ denotes the indictor function.

Under the loss function

$$L(\delta, \beta) = (\delta - \beta)' D(\delta - \beta), \tag{2.1}$$

where δ denotes an estimator of β and D is a positive define matrix. The Bayes estimator (BE) of the parameter β, say $\hat{\beta}_{BE}$, can be obtained by $\hat{\beta}_{BE} = E(H(\beta|y))$. However, it is not easy to deal with the integration related to $H(\beta|y)$. Simulation-based methods such as the Gibbs sampling procedure and Metropolis method maybe work, but they are somewhat inconvenient to use since they have no explicit expressions. In this Section we propose a linear Bayes procedure to estimate the parameter β, which replaces the completely specified prior $G(\beta)$ by an assumption about just a few moments of the prior.

Considering the model (1.1), it is well known that the least square estimator (LSE) of β is $\hat{\beta} = (X'X)^{-1}X'y$. When we take into account the restriction $c'\beta = d$, the equality constrained LSE of β would be $\tilde{\beta} = \hat{\beta} - (X'X)^{-1}c(c'(X'X)^{-1}c)^{-1}(c'\hat{\beta} - d)$.

According to Geroge and Thomas (1981), we transform the model (1.1) and the constraint (1.2) into the following expression

$$y = Z\theta + \varepsilon, \quad E(\varepsilon) = 0, \quad Cov(\varepsilon) = \sigma^2 I, \tag{2.2}$$

$$\text{and} \qquad h_1\theta_1 \geq d, \tag{2.3}$$

where $Z = X(X'X)^{-1/2}Q, \theta = Q'(X'X)^{1/2}\beta$, h_1 and θ_1 are the first component of $h = Q'(X'X)^{-\frac{1}{2}}c$ and θ respectively, and Q is an orthogonal matrix such that

$$Q'(X'X)^{-\frac{1}{2}}c(c'(X'X)^{-1}c)^{-1}c'(X'X)^{-\frac{1}{2}}Q = \begin{bmatrix} 1 & 0_{1\times(p-1)} \\ 0_{(p-1)\times 1} & 0_{(p-1)\times(p-1)} \end{bmatrix}_{p\times p}.$$

Note that the LSE of θ in model (2.2) is $\hat{\theta} = (Z'Z)^{-1}Z'y = Z'y$, then under the equality constraint $h_1\theta_1 = d$, the equality constrained LSE of θ is $\tilde{\theta} = (d_1, \hat{\theta}'_{(p-1)})'$, where $d_1 = d/h_1$ and $\hat{\theta}_{(p-1)}$ is a $(p-1)\times 1$ column vector composed of the last $p-1$ components of $\hat{\theta}$. Hence, under (2.3) the inequality CLSE of θ is defined as

$$\hat{\theta}_C = \begin{cases} \tilde{\theta}, & h_1\hat{\theta}_1 < d, \\ \hat{\theta}, & h_1\hat{\theta}_1 \geq d, \end{cases} \tag{2.4}$$

where $\hat{\theta}_1$ is the first component of $\hat{\theta}$.

Alternatively, it can be expressed as

$$\begin{aligned} \hat{\theta}_C &= \begin{bmatrix} I_{(-\infty,d)}(h_1\hat{\theta}_1)d_1 + I_{[d,+\infty)}(h_1\hat{\theta}_1)\hat{\theta}_1 \\ \hat{\theta}_{(p-1)} \end{bmatrix} \\ &= \hat{\theta} + \begin{bmatrix} I_{(-\infty,d)}(h_1\hat{\theta}_1)(d_1 - \hat{\theta}_1) \\ 0 \end{bmatrix}, \end{aligned} \tag{2.5}$$

where $I_{(a,b)}(\cdot)$ denotes the indicator function.

Define the class of linear estimators of θ as $\mathcal{F} = \{\bar{\theta}_L : \bar{\theta}_L = B\hat{\theta}_C + b\}$, where B and b are $p\times p$ and $p\times 1$ undetermined matrices respectively. Under the loss function (2.1), we propose a LABE of θ, say $\hat{\theta}_{LB}$, satisfying the following conditions

$$R(\hat{\theta}_{LB},\theta) = \min R(\bar{\theta}_L,\theta) = \min EL(\bar{\theta}_L,\theta), \tag{2.6}$$

$$E(\hat{\theta}_{LB}-\theta) = 0, \tag{2.7}$$

where E denotes the expectation with respect to the joint distribution of y and θ.

Given the parameter θ, we write the expectation of $\hat{\theta}_C$ as

$$\begin{aligned} E(\hat{\theta}_C|\theta) &= \theta + E\left[\begin{matrix} I_{(-\infty,d)}(h_1\hat{\theta}_1)(d_1-\hat{\theta}_1) \\ 0 \end{matrix}\middle|\theta\right] \\ &= \theta + w \end{aligned} \tag{2.8}$$

with

$$w = E\left[\begin{matrix} I_{(-\infty,d)}(h_1\hat{\theta}_1)(d_1-\hat{\theta}_1) \\ 0 \end{matrix} |\theta\right].$$

We first derive the expression of $\hat{\theta}_{LB}$ defined by the formulas (2.6) and (2.7).

Then we obtain the following equation

$$E(\hat{\theta}_C) = E\left[E(\hat{\theta}_C|\theta)\right] = E(\theta+w). \tag{2.9}$$

From the formula (2.7), we know $b = E\theta - BE(\theta+w)$. Hence, we have

$$\begin{aligned} R(\bar{\theta}_L,\theta) &= E(\bar{\theta}_L-\theta)'D(\bar{\theta}_L-\theta) = tr\left[DE(\bar{\theta}_L-\theta)(\bar{\theta}_L-\theta)'\right] \\ &= tr\left(DE\left\{\left[B\hat{\theta}_C + E\theta - BE(\theta+w) - \theta\right]\left[B\hat{\theta}_C + E\theta - BE(\theta+w) - \theta\right]'\right\}\right) \\ &= tr\left(DE\left\{B\left[\hat{\theta}_C - E(\theta+w)\right] - (\theta - E\theta)\right\}\left\{B\left[\hat{\theta}_C - E(\theta+w)\right] - (\theta - E\theta)\right\}'\right) \\ &= tr\Bigg(DBE\left\{\left[\hat{\theta}_C - E(\theta+w)\right]\left[\hat{\theta}_C - E(\theta+w)\right]'\right\}B' \\ &\quad + DE\left[(\theta-E\theta)(\theta-E\theta)'\right] - DE\left\{(\theta-E\theta)\left[\hat{\theta}_C - E(\theta+w)\right]'\right\}B' \\ &\quad - DBE\left\{\left[\hat{\theta}_C - E(\theta+w)\right](\theta-E\theta)'\right\}\Bigg) \\ &= tr\Bigg(DBE\left\{\left[\hat{\theta}_C - E(\theta+w)\right]\left[\hat{\theta}_C - E(\theta+w)\right]'\right\}B' \\ &\quad + DE\left[(\theta-E\theta)(\theta-E\theta)'\right] - 2DE\left\{(\theta-E\theta)\left[\hat{\theta}_C - E(\theta+w)\right]'\right\}B'\Bigg) \\ &= tr\left\{DBCov(\hat{\theta}_C)B' + DCov(\theta) - 2DMB'\right\}, \end{aligned} \tag{2.10}$$

where $M = Cov(\theta) + Cov(\theta, w)$.

Differentiating $R(\bar{\theta}_L,\theta)$ with respect to B and setting the result to zero gives $DBCov(\hat{\theta}_C) - DM = 0$, which yields

$$B = M(Cov(\hat{\theta}_C))^{-1}. \tag{2.11}$$

Together with $b = E\theta - M(Cov(\hat{\theta}_C))^{-1}E(\theta + w)$, we obtain

$$\hat{\theta}_{LB} = M(Cov(\hat{\theta}_C))^{-1}\hat{\theta}_C + E\theta - M(Cov(\hat{\theta}_C))^{-1}E(\theta + w). \tag{2.12}$$

Note that even if the unbiased constraint $E(\hat{\theta}_{LB} - \theta) = 0$ is removed, the same result as (2.12) can be obtained. This indicates that the LABE $\hat{\theta}_{LB}$ not only satisfies the unbiased condition $E(\hat{\theta}_{LB} - \theta) = 0$, but also performs the best among all the linear Bayes estimators.

From the above, we know $\beta = (X'X)^{-\frac{1}{2}}Q\theta$, so we figure out the inequality CLSE of β as $\hat{\beta}_C = (X'X)^{-\frac{1}{2}}Q\hat{\theta}_C$. Our aim is to construct the LABE of β, which is superior to $\hat{\beta}_C$. Let the LABE of β be of the form $\hat{\beta}_{LB} = A\hat{\beta}_C + a$, which is also characterized by the equations (2.6) and (2.7). By $\hat{\beta}_C = (X'X)^{-\frac{1}{2}}Q\hat{\theta}_C$ and mimicking the derivation process of $\hat{\theta}_{LB}$ we obtain

$$A = (X'X)^{-\frac{1}{2}}QM\left[(X'X)^{-\frac{1}{2}}Q\right]'\left[Cov(\hat{\beta}_C)\right]^{-1},$$
$$a = E(\beta) - (X'X)^{-\frac{1}{2}}QM\left[(X'X)^{-\frac{1}{2}}Q\right]'\left[Cov(\hat{\beta}_C)\right]^{-1}\left[E(\beta) + (X'X)^{-\frac{1}{2}}QE(w)\right].$$

Thus, we integrate the above discussions into the following theorem.

Theorem 2.1. *Assume that the prior covariance of β exists. Thus under the model (1.1) and the constraint (1.2), the expression of $\hat{\beta}_{LB}$ is*

$$\begin{aligned}\hat{\beta}_{LB} &= (X'X)^{-\frac{1}{2}}QM\left[(X'X)^{-\frac{1}{2}}Q\right]'\left[Cov(\hat{\beta}_C)\right]^{-1}\hat{\beta}_C + E(\beta)\\ &\quad - (X'X)^{-\frac{1}{2}}QM\left[(X'X)^{-\frac{1}{2}}Q\right]'\left[Cov(\hat{\beta}_C)\right]^{-1}\left[E(\beta) + (X'X)^{-\frac{1}{2}}QE(w)\right].\end{aligned} \tag{2.13}$$

3. The superiority of Labe

In this Section, we compare the MSEMs of $\hat{\beta}_C$ and $\hat{\beta}_{LB}$ from Bayesian viewpoint on the basis of studying the MSEMs of $\hat{\theta}_C$ and $\hat{\theta}_{LB}$.

Lemma 3.1. *Let $\hat{\theta}_C$ and $\hat{\theta}_{LB}$ be given by (2.5) and (2.12), respectively. Then*

$$\mathrm{MSEM}(\hat{\theta}_C) - \mathrm{MSEM}(\hat{\theta}_{LB}) \geq 0. \tag{3.1}$$

Proof. See the Appendix.

Thus, we come to the following result.

Theorem 3.1. *Let $\hat{\beta}_{LB}$ be given in (2.13) and $\hat{\beta}_C = (X'X)^{-\frac{1}{2}}Q\hat{\theta}_C$. Then $\hat{\beta}_{LB}$ is superior to $\hat{\beta}_C$ in terms of MSEM criterion, i.e., $\mathrm{MSEM}(\hat{\beta}_{LB}) \leq \mathrm{MSEM}(\hat{\beta}_C)$.*

Proof. The relationship between $\hat{\beta}_{LB}$ and $\hat{\theta}_{LB}$ is as follows

$$\begin{aligned}\hat{\beta}_{LB} &= (X'X)^{-\frac{1}{2}}QM\left[(X'X)^{-\frac{1}{2}}Q\right]'\left[Cov(\hat{\beta}_C)\right]^{-1}\hat{\beta}_C + E(\beta)\\ &\quad - (X'X)^{-\frac{1}{2}}QM\left[(X'X)^{-\frac{1}{2}}Q\right]'\left[Cov(\hat{\beta}_C)\right]^{-1}\left[E(\beta) + (X'X)^{-\frac{1}{2}}QE(w)\right]\\ &= (X'X)^{-\frac{1}{2}}QM\left[Cov(\hat{\theta}_C)\right]^{-1}\left[(X'X)^{-\frac{1}{2}}Q\right]^{-1}\hat{\beta}_C + E(\beta)\\ &\quad - (X'X)^{-\frac{1}{2}}QM\left[Cov(\hat{\theta}_C)\right]^{-1}\left[(X'X)^{-\frac{1}{2}}Q\right]^{-1}\left[E(\beta) + (X'X)^{-\frac{1}{2}}QE(w)\right]\\ &= (X'X)^{-\frac{1}{2}}Q\left[M(Cov(\hat{\theta}_C))^{-1}\hat{\theta}_C + E\theta - M(Cov(\hat{\theta}_C))^{-1}E(\theta + w)\right]\\ &= (X'X)^{-\frac{1}{2}}Q\hat{\theta}_{LB}.\end{aligned} \tag{3.2}$$

Using (3.1), we have

$$\text{MSEM}(\hat{\beta}_C) - \text{MSEM}(\hat{\beta}_{LB}) = (X'X)^{-\frac{1}{2}}Q\left[\text{MSEM}(\hat{\theta}_C) - \text{MSEM}(\hat{\theta}_{LB})\right]Q'(X'X)^{-\frac{1}{2}} \geq 0.$$

The proof of the Theorem 3.1 is complete.

4. Numerical comparisons

In this Section we use some simulations results under different prior distributions to exhibit the properties of $\hat{\beta}_{LB}$. To see the effect of the number of observations, n is chosen to be $5, 10, \ldots, 100$. Then let $p=3$, $\sigma=3$, $d=2$. According to Selma, Gulesen, and Selahattin (2013), those elements of the matrix X are simulated by $x_{ij} = (1-\gamma^2)^{\frac{1}{2}}z_{ij} + \gamma z_{i,p+1}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, p$ where z_{ij} subject to independent standard normal distribution and γ is the correlation coefficient, whose value is taken as 0.75. The parameter vector β is generated from the prior distribution. Once X and β are generated, the observed vector y is simulated by $N(X\beta, \sigma^2 I)$.

Suppose that the random error vector ε in (1.1) is normally distributed, i.e.,

$$\varepsilon \sim N_n(0, \sigma^2 I_n). \tag{4.1}$$

We first calculate $E(\hat{\theta}_C)$ and $Cov(\hat{\theta}_C)$, as follows

$$E(\hat{\theta}_C) = E\theta + E\left[\begin{matrix}(d_1 - \theta_1)\Phi(r) + \dfrac{\sigma}{\sqrt{2\pi}}\dfrac{|h_1|}{h_1}\exp\left\{-\dfrac{(d_1-\theta_1)^2}{2\sigma^2}\right\}\\ 0\end{matrix}\right], \tag{4.2}$$

where $r = \frac{(d_1-\theta_1)}{\sigma}\frac{h_1}{|h_1|}$ and Φ is the cumulative distribution function of standard normal distribution.

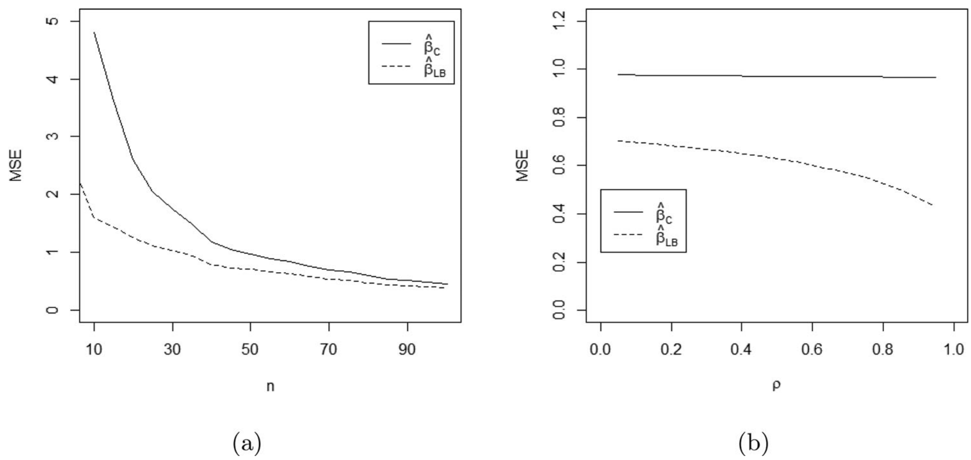


Figure 1. The changes of the MSEs under the normal prior.

And

$$\begin{aligned} Cov(\hat{\theta}_C) &= E\Big[Cov(\hat{\theta}_C|\theta)\Big] + Cov\Big[E(\hat{\theta}_C|\theta)\Big] \\ &= \sigma^2\begin{bmatrix} 0 & 0 \\ 0 & I_{(p-1)} \end{bmatrix} + E\begin{bmatrix} I_1 - I_2^2 & 0 \\ 0 & 0 \end{bmatrix} \\ &\quad + Cov(\theta + w), \end{aligned} \tag{4.3}$$

where

$$\begin{aligned} I_1 &= (\theta_1 - d_1)^2(1 - \Phi(r)) + \sigma^2\left[\frac{r}{\sqrt{2\pi}}\exp\left(-\frac{r^2}{2}\right) + (1 - \Phi(r))\right] \\ &\quad + 2(\theta_1 - d_1)\frac{|h_1|}{h_1}\sigma\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{r^2}{2}\right) \end{aligned} \tag{4.4}$$

and

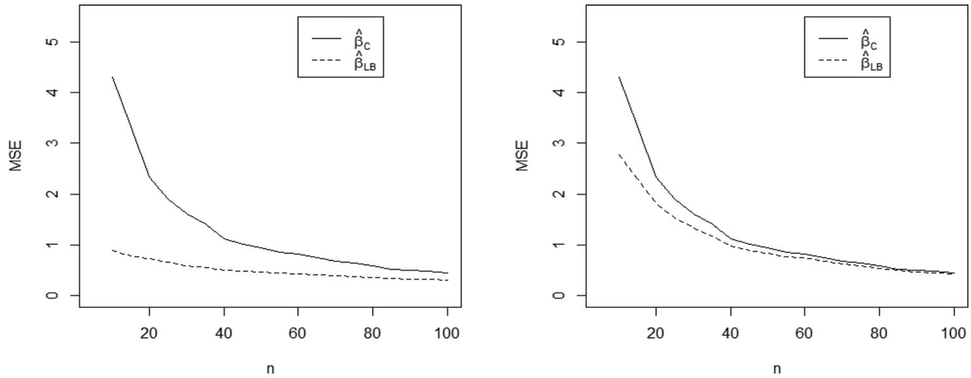
$$I_2 = (\theta_1 - d_1)(1 - \Phi(r)) + \frac{|h_1|}{h_1}\sigma\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{r^2}{2}\right). \tag{4.5}$$

The proofs of (4.2)-(4.5) are given in Appendix.

Case 1: We assume that β follows a truncated normal prior distribution $N_p(\mu_0, \Sigma_0)\mathrm{I}(c'\beta \geq d)$, then the posterior distribution of β would be

$$h(\beta|y) \propto \exp\left\{-\frac{1}{2}(\beta - \mu_1)'\left(\frac{X'X}{\sigma^2} + \Sigma_0^{-1}\right)(\beta - \mu_1)\right\}I(c'\beta \geq d),$$

where $\mu_1 = \Sigma_1\left(\Sigma_0^{-1}\mu_0 + \frac{X'y}{\sigma^2}\right)$ and $\Sigma_1 = \left(\frac{X'X}{\sigma^2} + \Sigma_0^{-1}\right)^{-1}$. Let hyper parameters $\mu_0 = (1, ..., 1)'$ and $\Sigma_0 = \rho 1_p 1_p' + (1 - \rho)I_p$, where $1_p = (1, ..., 1)'$ and ρ is correlation coefficient, which is chosen as $0.1, 0.2, ..., 1$.



(a) (b)

Figure 2. The situation of the components of β are independent.

First, 5000 samples of β are generated from $N_p(\mu_0, \Sigma_0)\mathrm{I}(c'\beta \geq d)$. Using the fact that $\theta = Q'(X'X)^{\frac{1}{2}}\beta$, we obtain θ. Then the values of $w, E(\theta)$ and $Cov(\hat{\theta}_C)$ are obtained by Monte Carlo simulation, so that the MSEMs of $\hat{\theta}_C$ and $\hat{\theta}_{LB}$ can be calculated and MSEs are obtained by the trace of MSEM. Finally, by the relationships of $\hat{\beta}_C$ and $\hat{\theta}_C$, similarly $\hat{\beta}_{LB}$ and $\hat{\theta}_{LB}$, we obtain the MSEs of $\hat{\beta}_C$ and $\hat{\beta}_{LB}$. The process is repeated 500 times and the average of MSEs is taken.

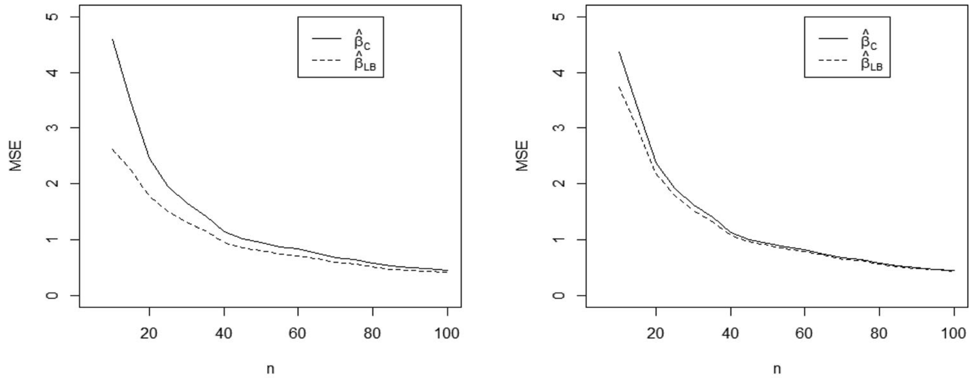
As n and ρ change, Figure 1 presents the corresponding variations of the MSEs of $\hat{\beta}_{LB}$ and $\hat{\beta}_C$ respectively. For $\rho = 0$, Figure 1(a) exhibits that the MSEs are decreasing with the increasing of n, and the MSEs of $\hat{\beta}_C$ are always larger than those of $\hat{\beta}_{LB}$. In particular, both the MSEs of $\hat{\beta}_{LB}$ and $\hat{\beta}_C$ are greatly different for small n. For $n = 50$, Figure 1(b) points out the change of ρ has little influence on the MSE of $\hat{\beta}_C$, while the MSE of $\hat{\beta}_{LB}$ decreases gradually as ρ grows larger. Note that when ρ is not equal to 0 the prior distribution has a uniform correlation structure (see Žežula (2006), Klein and Žežula (2007), etc.), so we can find that for this covariance structure the MSEs of $\hat{\beta}_{LB}$ are always less than those of $\hat{\beta}_C$

Case 2: When β has a three-dimension uniform prior over the region D and satisfies $c'\beta \geq d$, we obtain the posterior distribution of β as

$$h(\beta|y) \propto \exp\left\{ -\frac{1}{2}(\beta - \mu_2)'\Sigma_2(\beta - \mu_2)\right\} I_D(\beta) I(c'\beta \geq d),$$

where $\mu_2 = (X'X)^{-1}X'y$ and $\Sigma_2 = \left(\frac{X'X}{\sigma^2}\right)^{-1}$.

Figure 2 exhibits the MSE performances of $\hat{\beta}_C$ and $\hat{\beta}_{LB}$ under the same scenarios as Figure 1 except that β is simulated from the uniform distribution. The two graphs of Figure 2 present the MSEs of $\hat{\beta}_C$ and $\hat{\beta}_{LB}$ are changing with n in the case that these components of β are independent. Specifically, Figure 2(a) exhibits the influence of n on MSE when the three components of β follow the uniform distributions $U(10, 12)$,



(a) (b)

Figure 3. The situation of the components of β are dependent.

$U(13, 15)$ and $U(14, 17)$, respectively. Figure 2(b) is plotted under the condition that the prior distributions are $U(9, 13)$, $U(12, 16)$ and $U(13, 19)$, where the prior information is getting more dispersed. Comparing with Figure 2(a) we see that the two lines in Figure 2(b) are closer to each other as the sample size n gets larger. It also indicates that the more scattered prior provides information the less, and if the information about the regression parameter comes mainly from the sample, then the difference between the MSEs of $\hat{\beta}_{LB}$ and those of $\hat{\beta}_C$ is becoming small. Figure 3 is plotted when β follows an uniform distribution on the sphere, i.e., $\beta_1^2 + \beta_2^2 + \beta_3^2 \leq R_0^2$, where the radius of ball R_0 is assumed to be 5 and 10, and the larger R_0 implies the variance is larger. Figures 3(a) and 3(b) present the influence of n on the MSE when $R_0 s$ are 5 and 10 respectively, and we find that the difference between the MSEs of $\hat{\beta}_{LB}$ and $\hat{\beta}_C$ is getting smaller as R_0 turns larger.In all of the above Figures, the MSEs of the two estimators are compared under the same prior conditions. It is readily seen that the MSEs of $\hat{\beta}_{LB}$ are always less than those of $\hat{\beta}_C$, which is consistent with the theoretical results, i.e., $\hat{\beta}_{LB}$ is superior to $\hat{\beta}_C$.

In order to examine the approximate performances of $\hat{\beta}_{LB}$ and $\hat{\beta}_C$ to $\hat{\beta}_{BE}$, the following formula is defined by

$$\| \bar{\beta} - \hat{\beta}_{BE} \| = \frac{1}{500} \sum_{j=1}^{500} \sqrt{\sum_{i=1}^{p} (\bar{\beta}_{ij} - \hat{\beta}_{BEij})^2}, \tag{4.6}$$

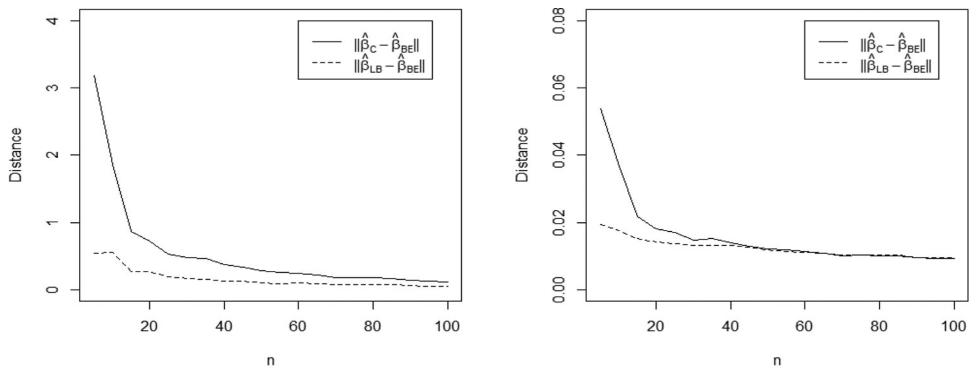
where $\bar{\beta}$ can be $\hat{\beta}_{LB}$ or $\hat{\beta}_C$. The above formula indicates 500 times experiments are repeated. $\hat{\beta}_{LB}$ would be computed by the formula (3.2) and $\hat{\beta}_C$ can be computed by using the expression $\hat{\beta}_C = (X'X)^{-\frac{1}{2}} Q \hat{\theta}_C$. It is noted that the $\hat{\beta}_{BE}$ does not have an explicit expression, which needs to be calculated via the Monte Carlo method.

In this part, $\| \hat{\beta}_{LB} - \hat{\beta}_{BE} \|$ and $\| \hat{\beta}_C - \hat{\beta}_{BE} \|$ are calculated for the several priors in Table 1 respectively. Note that β_1, β_2 and β_3 in pr2 and pr4 also satisfy $c'\beta \geq d$.

In Figures 4 and 5, we plot the curves of $||\hat{\beta}_C - \hat{\beta}_{BE}||$ and $||\hat{\beta}_{LB} - \hat{\beta}_{BE}||$ changing with sample size n for priors 1, 2, 3 and 4, respectively. From Figures 4 and 5 we see

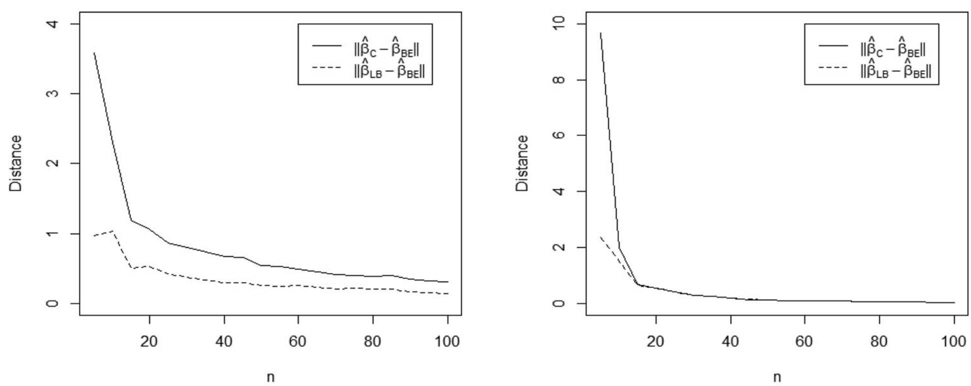
Table 1. The priors of β.

independent priors	dependent priors
pr1: $\beta \sim N_3(\mu, \tau^2 I_3) I(c'\beta \geq d)$	pr3: $\beta \sim N_3(\mu, \Sigma) I(c'\beta \geq d)$
pr2: $\beta_1 \sim U(6, 14)$; $\beta_2 \sim U(8, 17)$; $\beta_3 \sim U(9, 20)$	pr4: $\beta_1^2 + \beta_2^2 + \beta_3^2 < R_0^2$



(a) (b)

Figure 4. The changes of $||\hat{\beta}_C - \hat{\beta}_{BE}||$ and $||\hat{\beta}_{LB} - \hat{\beta}_{BE}||$ under the pr1 and the pr2.



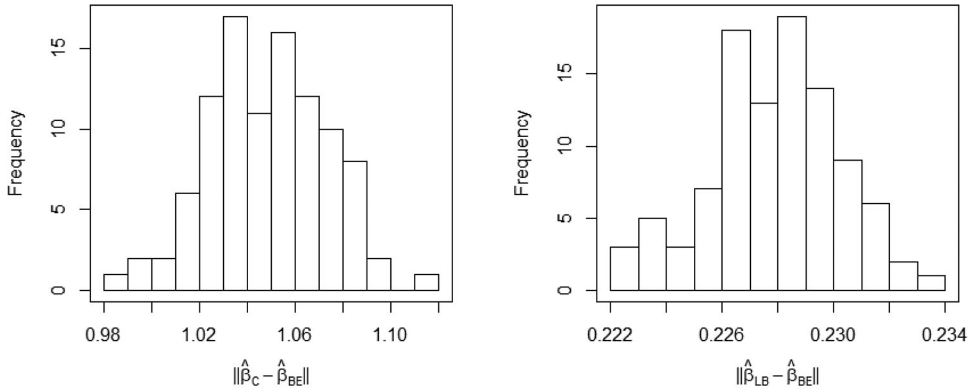
(a) (b)

Figure 5. The changes of $||\hat{\beta}_C - \hat{\beta}_{BE}||$ and $||\hat{\beta}_{LB} - \hat{\beta}_{BE}||$ under the pr3 and the pr4.

that $||\hat{\beta}_C - \hat{\beta}_{BE}||s$ are less than $||\hat{\beta}_{LB} - \hat{\beta}_{BE}||s$ for almost all n, which implies that as an approximation $\hat{\beta}_{LB}$ outperforms $\hat{\beta}_C$.

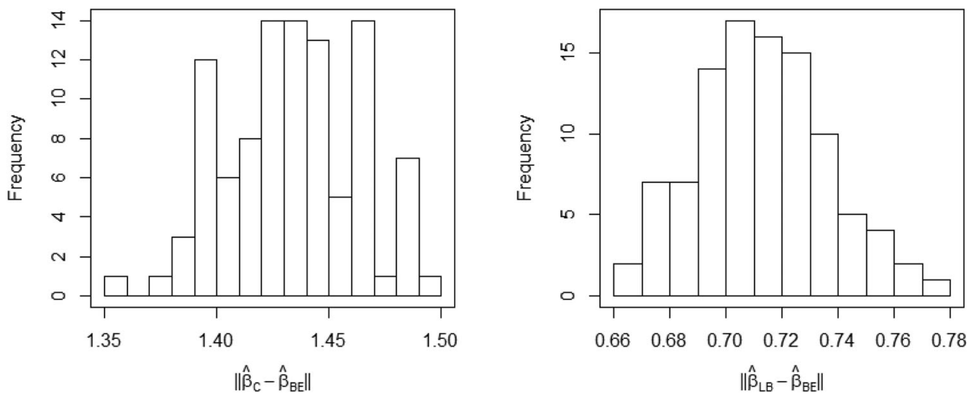
In particular, the pr4 is defined on the sphere which is more complex than the other three prior distributions. But, it can be found that $\hat{\beta}_{LB}$ still behaves better than $\hat{\beta}_C$.

Figures 6 and 7 further exhibit the frequency histograms of $||\hat{\beta}_C - \hat{\beta}_{BE}||$ and $||\hat{\beta}_{LB} - \hat{\beta}_{BE}||$ for the priors 1 and 3, which are obtained by repeating the simulations 100 times. From Figure 6 we see that $||\hat{\beta}_C - \hat{\beta}_{BE}||s$ are concentrated in the interval (1.02, 1.08), while $||\hat{\beta}_{LB} - \hat{\beta}_{BE}||s$ are always concentrated in the interval (0.226, 0.23). Also, from Figure 7 we know that $||\hat{\beta}_C - \hat{\beta}_{BE}||s$ are concentrated in the interval (1.39, 1.47), and $||\hat{\beta}_{LB} - \hat{\beta}_{BE}||s$ are concentrated in the interval (0.685, 0.725). Obviously, it can be found



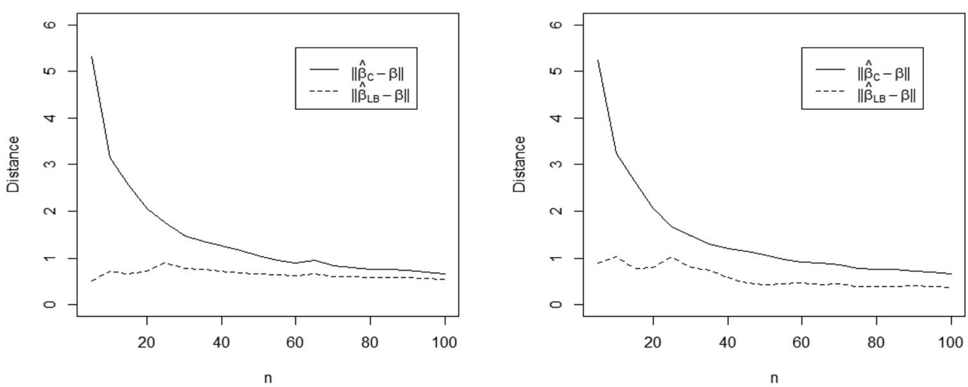
(a) (b)

Figure 6. The histograms of $||\hat{\beta}_C - \hat{\beta}_{BE}||$ and $||\hat{\beta}_{LB} - \hat{\beta}_{BE}||$ under the pr1.



(a) (b)

Figure 7. The histograms of $||\hat{\beta}_C - \hat{\beta}_{BE}||$ and $||\hat{\beta}_{LB} - \hat{\beta}_{BE}||$ under the pr3.



(a) (b)

Figure 8. The changes of $||\hat{\beta}_C - \beta||$ and $||\hat{\beta}_{LB} - \beta||$ under the pr1 and the pr3.

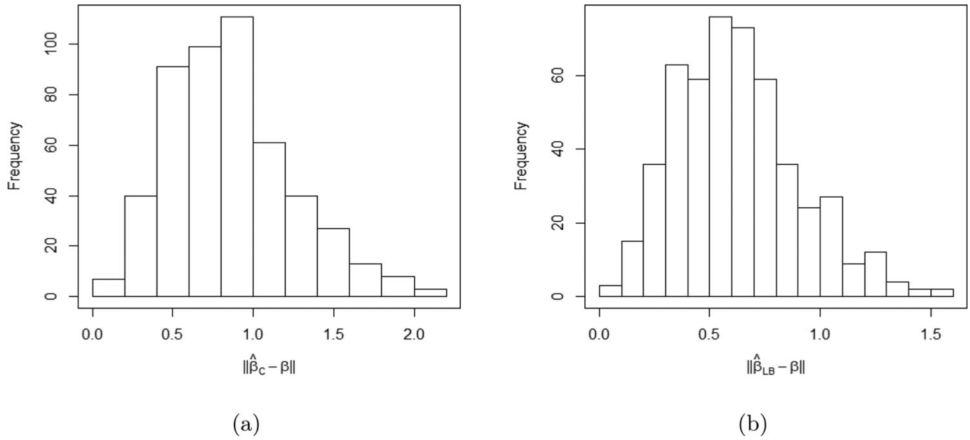


Figure 9. The histograms of $||\hat{\beta}_C - \beta||$ and $||\hat{\beta}_{LB} - \beta||$ under the pr1.

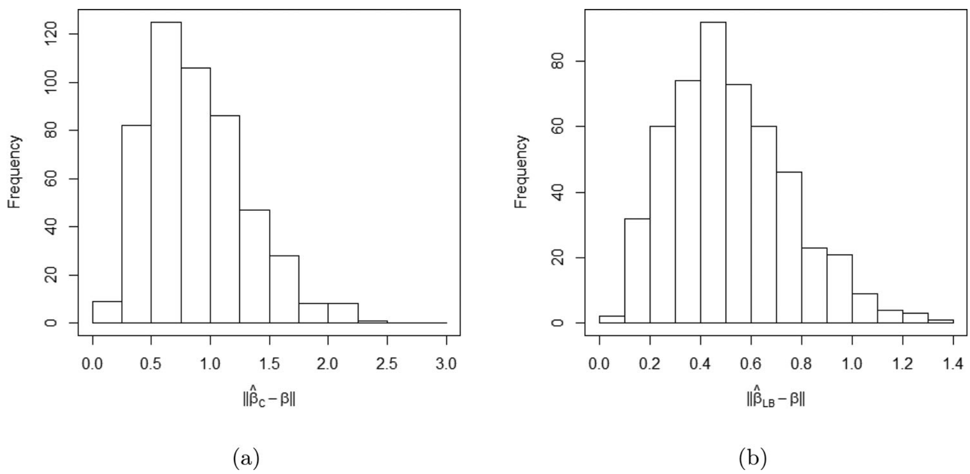


Figure 10. The histograms of $||\hat{\beta}_C - \beta||$ and $||\hat{\beta}_{LB} - \beta||$ under the pr3.

from Figures 6 and 7 that the mode of $||\hat{\beta}_{LB} - \hat{\beta}_{BE}||$ is smaller than that of $||\hat{\beta}_C - \hat{\beta}_{BE}||$, which provides further evidence that $\hat{\beta}_{LB}$ has a better approximate representation for $\hat{\beta}_{BE}$.

On the other hand, in order to study the estimation performance of $\hat{\beta}_C$ and $\hat{\beta}_{LB}$ for β, under the normal prior we discuss $||\hat{\beta}_C - \beta||$ and $||\hat{\beta}_{LB} - \beta||$, which are defined by replacing $\hat{\beta}_{BE}$ in (4.6) with the truth value of parameter β. Figures 8(a) and 8(b) show the variations of $||\hat{\beta}_C - \beta||$ and $||\hat{\beta}_{LB} - \beta||$ with n for the pr1 and the pr3, respectively. From Figure 8, it is clear that all $||\hat{\beta}_{LB} - \beta||s$ are relatively small. In Figures 9 and 10, we draw the frequency histograms of $||\hat{\beta}_C - \beta||$ and $||\hat{\beta}_{LB} - \beta||$ for the pr1 and the pr3, respectively. In Figure 9, we can see that the shapes of the two graphs are roughly the same, but the mode of $||\hat{\beta}_{LB} - \beta||$ is still smaller than that of $||\hat{\beta}_C - \beta||$. In Figure 10, $||\hat{\beta}_C - \beta||s$ are substantially concentrated in the interval (0.5, 1.25), while $||\hat{\beta}_{LB} - \beta||s$ are substantially concentrated in the interval (0.2, 0.5), so the mode of $||\hat{\beta}_{LB} - \beta||$ is

Table 2. The values of α_0, α_1 and λ for different correlation coefficients ρ.

ρ	α_0	α_1	λ
0.50	0.8357	0.1643	4.5163
0.75	0.7636	0.2364	3.2298
0.95	0.6716	0.3284	2.0449

Table 3. The values of α_0, α_1 and λ for different prior mean vectors μ.

μ	α_0	α_1	λ
$(1,1,1,1,1)'$	0.7106	0.2894	2.4549
$(-1,1,2,2,1)'$	0.7167	0.2833	2.5298
$(-0.5,1.5,2,1.5,1)'$	0.7900	0.2100	3.7615

significantly smaller than that of $||\hat{\beta}_C - \beta||$. In summary, Figures 8–10 all show that as an estimator of β the performances of $\hat{\beta}_{LB}$ are better.

5. Numerical examples

In this Section we employ an example to demonstrate the application of the proposed LABE. The Portland cement data are used (Hald and Friedman 1952). The data mainly focuses on the relationship between the heat produced by the silicate cement in the process of solidification and hardening and the percentage of the four compounds. These four components are tricalcium aluminate(X_1), tricalcium silicate(X_2), tetracalcium ferricaluminate(X_3) and dicalcium silicate(X_4). The heat generated after 180 days of curing is calculated by the calories(y) per gram of cement. The details are as follows: $X_1 = (7,1,11,11,7,11,3,1,2,21,1,11,10)'$, $X_2 = (26,25,56,31,52,55,71,31,54,47,40,66,68)'$, $X_3 = (6,15,8,8,6,9,17,22,18,4,23,9,8)'$, $X_4 = (60,52,20,47,33,22,6,44,22,26, 34,12, 12)'$, $y = (78.5,74.3,104.3,87.6,95.9,109.2,102.7,72.5,93.1,115.9,83.8,113.3,109.4)'$.

Denote $X_{13\times5} = (\mathbf{1}, X_1, X_2, X_3, X_4)$ and $\beta = (\beta_0,\beta_1,\beta_2,\beta_3,\beta_4)'$, where $\mathbf{1}$ is a 13×1 vector consisting of 1. We present the linear model:

$$y = X\beta + e, \qquad e \sim N_{13}(0, \sigma^2 I_{13}). \tag{5.1}$$

We intend to add the constraint condition $\beta_1 - \beta_2 + \beta_3 - \beta_4 \geq 0$ or $c'\beta \geq 0$, where $c = (0,1,-1,1,-1)'$. We test the hypothesis $H_0 : c'\beta \geq 0$ and alternative hypothesis $H_1 : c'\beta < 0$ in the framework of the unconstrained linear model (5.1) from the view of Bayesian inference.

We use the estimator $||y - X\hat{\beta}_C||^2/(n-4)$ to estimate σ^2 which is approximately equal to 5.67 and assume that the parameter β follows $N(\mu,\Sigma)$ with $\Sigma = \rho 1_p 1_p' + (1-\rho)I_p$. Then the posterior distribution of β is $N(\mu^*,\Sigma^*)$, where $\Sigma^* = \left(\frac{X'X}{\sigma^2} + \Sigma^{-1}\right)^{-1}$ and $\mu^* = \Sigma^*\left(\Sigma^{-1}\mu + \frac{X'y}{\sigma^2}\right)$. Thus we calculate the posterior probability of the hypothesis H_0, say $\alpha_0 = P(c'\beta \geq 0|y)$, which is obtained by using 50000 samples from $N(\mu^*,\Sigma^*)$. Note that the posterior probability of the hypothesis H_1 would be $\alpha_1 = 1 - \alpha_0$.

Firstly, we fix $\mu = (1,1,1,1,1)'$ and set $\rho = 0.5, 0.75, 0.95$. For these three different values of ρ we present α_0, α_1 and the posterior probability ratio λ in Table 2.

Table 4. Distances under different correlation coefficients ρ $(\mu = (1,1,1,1,1)')$.

ρ	$\|\|\hat{\beta}_{LB} - \hat{\beta}_{BE}\|\|$	$\|\|\hat{\beta}_C - \hat{\beta}_{BE}\|\|$	$\|\|\hat{\beta}_{LB} - \beta\|\|$	$\|\|\hat{\beta}_C - \beta\|\|$
0.50	0.3038	106.2768	1.6038	6.2768
0.75	0.4599	106.5145	1.4268	4.5951
0.95	0.1865	106.5099	1.3652	2.5529

Table 5. Distances under different prior mean vectors μ $(\rho = 0.85)$.

μ	$\|\|\hat{\beta}_{LB} - \hat{\beta}_{BE}\|\|$	$\|\|\hat{\beta}_C - \hat{\beta}_{BE}\|\|$	$\|\|\hat{\beta}_{LB} - \beta\|\|$	$\|\|\hat{\beta}_C - \beta\|\|$
$(1,1,1,1,1)'$	0.3494	106.5161	1.8788	2.4384
$(-1,1,2,2,1)'$	0.3396	109.0290	2.5930	4.0804
$(-0.5,1.5,2,1.5,1)'$	0.2703	108.5066	1.9862	3.6210

Then, let $\mu = (1,1,1,1,1)', (-1,1,2,2,1)'$ and $(-0.5,1.5,2,1.5,1)'$ and fix $\rho = 0.85$, imitating the above calculation, the results of α_0, α_1 and λ are shown in Table 3.

From Tables 2 and 3, it is easily found that $\lambda > 1$ for all priors, which means that the hypothesis H_0 is accepted from Bayesian viewpoint.

In what follows, under the constraint condition we exhibit $||\hat{\beta}_{LB} - \hat{\beta}_{BE}||, ||\hat{\beta}_C - \hat{\beta}_{BE}||, ||\hat{\beta}_{LB} - \beta||$ and $||\hat{\beta}_C - \beta||$ for different ρ and μ in Tables 4 and 5, where we set the true value of β be $(0.6,1.3,1.3,0.05,0.6)'$.

In Table 4, it is easily seen that $||\hat{\beta}_{LB} - \hat{\beta}_{BE}||$ and $||\hat{\beta}_C - \hat{\beta}_{BE}||$ slightly fluctuate when ρ takes different values and $||\hat{\beta}_{LB} - \hat{\beta}_{BE}||$ is consistently less than $||\hat{\beta}_C - \hat{\beta}_{BE}||$. In Table 5, we find that $||\hat{\beta}_C - \hat{\beta}_{BE}||$ fluctuates for different prior means μ, which implies $||\hat{\beta}_C - \hat{\beta}_{BE}||$ is sensitive to the choice of μ, while the different values of μ have little influence on $||\hat{\beta}_{LB} - \hat{\beta}_{BE}||$. On the other hand, we see that both $||\hat{\beta}_{LB} - \beta||$ and $||\hat{\beta}_C - \beta||$ tend to be smaller as ρ gets larger. Also, $||\hat{\beta}_{LB} - \beta|| \le ||\hat{\beta}_C - \beta||$ can be found from all the numerical results in Tables 4 and 5. These show that $\hat{\beta}_{LB}$ outperforms $\hat{\beta}_C$ in this case, which further indicates that the LABE is applicable and feasible.

6. Conclusions

This paper employs the linear Bayes procedure to estimate the regression parameter with an inequality constraint since the usual Bayes estimator has no explicit form which causes it not easy to be used. We obtain the expression of the linear approximate Bayes estimator (LABE) without specifying the form of prior distribution and only make an assumption on the second moment of the prior instead. It is further proved that the proposed LABE is superior to the inequality constrained least square estimator (CLSE) under the MSEM criterion. Also, the simulation results show that the LABE has not only small MSEs but also good approximation performances compared with the inequality CLSE. Finally, we present a real example to verify the availability of the LABE.

Funding

This work was supported by Supported by NNSF of China (11371051).

References

Alan, T. K. W. 1994. Risk comparison of the inequality constrained least squares and other related estimators under balanced loss. *Economics Letters* 46:203–10.

Alan, T. K. W., and O. Kazuhiro. 2000. Minimum mean-squared error estimation in linear regression with an inequality constraint. *Journal of Statistical Planning and Inference* 86: 157–73.

Chong, K. L. 1976. Inequality constrained least-squares estimation. *Journal of the American Statistical Association* 71:746–51.

Geroge, G. J., and A. Y. Thomas. 1981. Sampling performance of an inequality restricted estimator. *Economics Letters* 7:327–33.

Geroge, G. J., and A. Y. Thomas. 1984. The non-optimality of the inequality restricted estimatior under squared error loss. *Journal of Econometrics* 25:165–77.

Goldstein, M. 1983. General variance modifications for linear Bayes estimators. *Journal of the American Statistical Association* 78 (383):616–18. doi:10.1080/01621459.1983.10478019.

Hald, A., and B. Friedman. 1952. *Statistical theory with engineering applications*. New York: John Wiley.

Hans, J. W. 1990. On inequality constrained generalized least-squares estimation. *Linear Algebra and Its Applications* 127:379–92.

Klein, D., and I. Žežula. 2007. *On uniform correlation structure*. Mathematical Methods in Economocs and Industry. Conference presented at Mathematical Methods in Economics and Industry, Slovakia; p. 94–100.

Lamotte, L. R. 1978. Bayes linear estimators. *Technometrics* 20 (3):281–90. doi:10.1080/00401706.1978.10489673.

Lius, A. E., and S. Bradley. 1984. A closed form solution for the least squares regression problem with linear inequality constraints. *Communications in Statistics-Theory and Methods* 13: 1127–34.

Ma, H. Q. 2008. Estimation of parameters for the partial linear regression models with inequality constraints. *Mathematical Theory and Applications* 28:5–8.

Qiu, H., J. Luo, and S. Zheng. 2014. On Bayes linear unbiased estimator under the balanced loss function. *Communications in Statistics - Theory and Methods* 43 (24):5175–94. doi:10.1080/03610926.2012.731128.

Rao, C. R. 1973. *Linear statistical inference and its application*. 2nd ed. New York: John Wiley and Sons.

Samanigo, F. J., and E. Vetrup. 1999. On improving standard estimators via linear empirical Bayes methods. *Statistics and Probability Letters* 44:309–18.

Selma, T., U. S. Gulesen, and K. Selahattin. 2013. Inequality constrained ridge regression estimator. *Statistics and Probability Letters* 83:2391–98.

Wang, L. C., and R. S. Singh. 2014. Linear Bayes estimator for the two-parameter exponential family under type II censoring. *Computational Statistics & Data Analysis* 71:633–42. doi:10.1016/j.csda.2013.07.020.

Žežula, I. 2006. Secial variance structures in the growth curve model. *Journal of Multivariate Analysis* 97 (3):606–18. doi:10.1016/j.jmva.2005.10.001.

Zhang, W. P., L. S. Wei, and Y. N. Yang. 2005. The superiorities of empirical bayes estimation of parameters in linear model. *Statistics & Probability Letters* 72 (1):43–50. doi:10.1016/j.spl.2004.12.001.

Appendix

The proof of Lemma 3.1

The MSEM of $\hat{\theta}_{LB}$ is

$$\begin{aligned}
\text{MSEM}(\hat{\theta}_{LB}) &= E(\hat{\theta}_{LB} - \theta)(\hat{\theta}_{LB} - \theta)' \\
&= E\{[M(Cov(\hat{\theta}_C))^{-1}\hat{\theta}_C + E\theta - M(Cov(\hat{\theta}_C))^{-1}E(\theta + w) - \theta] \\
&\quad [M(Cov(\hat{\theta}_C))^{-1}\hat{\theta}_C + E\theta - M(Cov(\hat{\theta}_C))^{-1}E(\theta + w) - \theta]'\} \\
&= E\{M(Cov(\hat{\theta}_C))^{-1}[\hat{\theta}_C - E(\theta + w)] - (\theta - E\theta)\} \\
&\quad \{M(Cov(\hat{\theta}_C))^{-1}[\hat{\theta}_C - E(\theta + w)] - (\theta - E\theta)\}' \\
&= M(Cov(\hat{\theta}_C))^{-1}M' - M(Cov(\hat{\theta}_C))^{-1}M' \\
&\quad - M(Cov(\hat{\theta}_C))^{-1}M' + Cov(\theta) \\
&= Cov(\theta) - M(Cov(\hat{\theta}_C))^{-1}M'.
\end{aligned} \tag{A1}$$

On the other hand, we know that the MSEM of $\hat{\theta}_C$ is

$$\begin{aligned}
\text{MSEM}(\hat{\theta}_C) &= E(\hat{\theta}_C - \theta)(\hat{\theta}_C - \theta)' \\
&= E(\hat{\theta}_C\hat{\theta}_C') - E(\hat{\theta}_C\theta') - E(\theta\hat{\theta}_C') + E(\theta\theta') \\
&= E(\hat{\theta}_C\hat{\theta}_C') - (M' + E\hat{\theta}_C E\theta') - (M + E\theta E\hat{\theta}_C') + E\theta E\theta' + Cov(\theta) \\
&= E(\hat{\theta}_C\hat{\theta}_C') - E\hat{\theta}_C E\hat{\theta}_C' + (E\hat{\theta}_C - E\theta)(E\hat{\theta}_C - E\theta)' \\
&\quad + Cov(\theta) - M - M' \\
&= Cov(\hat{\theta}_C) + EwEw' + Cov(\theta) - M - M',
\end{aligned} \tag{A2}$$

where note that

$$\begin{aligned}
E\hat{\theta}_C(\theta - E\theta)' &= E\{E\left[\hat{\theta}_C(\theta - E\theta)'|\theta\right]\} \\
&= E(\theta + w)(\theta - E\theta)' \\
&= E[\theta - E\theta + w - Ew + E(\theta + w)](\theta - E\theta)' \\
&= Cov(\theta) + Cov(w, \theta) = M',
\end{aligned}$$

therefore, it is easy to see $E(\hat{\theta}_C\theta') = M' + E\hat{\theta}_C E\theta'$ and $E(\theta\hat{\theta}_C') = M + E\theta E\hat{\theta}_C'$.

Comparing (A1) with (A2), we have

$$\begin{aligned}
&\text{MSEM}(\hat{\theta}_C) - \text{MSEM}(\hat{\theta}_{LB}) \\
&= Cov(\hat{\theta}_C) + EwEw' - M - M' + M(Cov(\hat{\theta}_C))^{-1}M' \\
&\quad \geq Cov(\hat{\theta}_C) - M - M' + M(Cov(\hat{\theta}_C))^{-1}M' \\
&= Cov(\hat{\theta}_C) - M(Cov(\hat{\theta}_C))^{-1}Cov(\hat{\theta}_C) - M' + M(Cov(\hat{\theta}_C))^{-1}M' \\
&= Cov(\hat{\theta}_C) + M(Cov(\hat{\theta}_C))^{-1}\left[M' - Cov(\hat{\theta}_C)\right] - M' \\
&= \left[M(Cov(\hat{\theta}_C))^{-1} - I\right]\left[M' - Cov(\hat{\theta}_C)\right] \\
&= \left[M - Cov(\hat{\theta}_C)\right](Cov(\hat{\theta}_C))^{-1}\left[M' - Cov(\hat{\theta}_C)\right] \geq 0.
\end{aligned} \tag{A3}$$

The proofs of (4.2)-(4.5) in Section 4.

Note that

$$w = E\left[\begin{matrix} I_{(-\infty,d)}(h_1\hat{\theta}_1)(d_1-\hat{\theta}_1) \\ 0 \end{matrix}\middle|\theta\right], \tag{A4}$$

and set $u_1 = \frac{h_1\hat{\theta}_1 - h_1\theta_1}{\sqrt{h_1^2\sigma^2}}$, we have $u_1|\theta \sim N(0,1)$ and $I_{(-\infty,d)}(h_1\hat{\theta}_1)(d_1-\hat{\theta}_1) = I_{(-\infty,r)}(u_1)\left(d_1-\theta_1-\frac{|h_1|}{h_1}u_1\sigma\right)$, where $r = \frac{(d_1-\theta_1)}{\sigma}\frac{h_1}{|h_1|}$.

Let w_1 be the first element of the vector w. We obtain

$$\begin{aligned} w_1 &= E\left\{I_{(-\infty,r)}(u_1)\left(d_1-\theta_1-\frac{|h_1|}{h_1}u_1\sigma\right)\middle|\theta\right\} \\ &= \int_{-\infty}^{r}\frac{1}{\sqrt{2\pi}}\left(d_1-\theta_1-\frac{|h_1|}{h_1}u_1\sigma\right)\exp\left\{-\frac{u_1^2}{2}\right\}du_1 \\ &= (d_1-\theta_1)\Phi(r) + \frac{\sigma}{\sqrt{2\pi}}\frac{|h_1|}{h_1}\exp\left\{-\frac{(d_1-\theta_1)^2}{2\sigma^2}\right\}, \end{aligned} \tag{A5}$$

where Φ is the cumulative distribution function of standard normal distribution.

Hence, $E(\hat{\theta}_C)$ can be expressed as

$$E(\hat{\theta}_C) = E\theta + E\left[\begin{matrix}(d_1-\theta_1)\Phi(r) + \frac{\sigma}{\sqrt{2\pi}}\frac{|h_1|}{h_1}\exp\left\{-\frac{(d_1-\theta_1)^2}{2\sigma^2}\right\} \\ 0\end{matrix}\right]. \tag{A6}$$

Also, we have

$$\begin{aligned} Cov(\hat{\theta}_C) &= E[Cov(\hat{\theta}_C|\theta)] + Cov[E(\hat{\theta}_C|\theta)] \\ &= E\left\{Cov\left(\hat{\theta}_0 + \left[I_{(-\infty,r)}(u_1)d_1 + I_{[r,+\infty)}(u_1)(\theta_1+\frac{|h_1|}{h_1}u_1\sigma)\ 0\right]\middle|\theta\right)\right\} \\ &\quad + Cov(\theta+w) = \sigma^2\begin{bmatrix}0 & 0 \\ 0 & I_{(p-1)}\end{bmatrix} + E\begin{bmatrix}Var(I_{[r,+\infty)}(u_1)\left(\theta_1+\frac{|h_1|}{h_1}u_1\sigma-d_1\right)\middle|\theta) & 0 \\ 0 & 0\end{bmatrix} \\ &\quad + Cov(\theta+w), \end{aligned} \tag{A7}$$

where $\hat{\theta}_0$ is a $p \times 1$ vector only with 0 replacing the first component of $\hat{\theta}$.

Furthermore, we write $Var[I_{(r,+\infty)}(u_1)\left(\theta_1+\frac{|h_1|}{h_1}u_1\sigma-d_1\right)|\theta] = I_1 - I_2^2$ with

$$\begin{aligned} I_1 &= E[I_{[r,+\infty)}(u_1)(\theta_1+\frac{|h_1|}{h_1}u_1\sigma-d_1)^2|\theta] \\ &= \int_{r}^{+\infty}(\theta_1-d_1+\frac{|h_1|}{h_1}u_1\sigma)^2\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_1^2}{2}\right)du_1 \\ &= (\theta_1-d_1)^2(1-\Phi(r)) + \sigma^2\left[\frac{r}{\sqrt{2\pi}}\exp\left(-\frac{r^2}{2}\right) + (1-\Phi(r))\right] \\ &= (\theta_1-d_1)^2(1-\Phi(r)) + \sigma^2\left[\frac{r}{\sqrt{2\pi}}\exp\left(-\frac{r^2}{2}\right) + (1-\Phi(r))\right] \\ &\quad + 2(\theta_1-d_1)\frac{|h_1|}{h_1}\sigma\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{r^2}{2}\right) \end{aligned} \tag{A8}$$

and

$$\begin{aligned} I_2 &= E\left[I_{[r,+\infty)}(u_1)\left(\theta_1 + \frac{|h_1|}{h_1}u_1\sigma - d_1\right)\middle|\theta\right] \\ &= \int_r^{+\infty}\left(\theta_1 - d_1 + \frac{|h_1|}{h_1}u_1\sigma\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_1^2}{2}\right)du_1 \\ &= (\theta_1 - d_1)(1 - \Phi(r)) + \frac{|h_1|}{h_1}\sigma\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{r^2}{2}\right). \end{aligned} \tag{A9}$$