COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
https://doi.org/10.1080/03610926.2022.2138438

Approximate Bayesian estimator for the parameter vector in linear models with multivariate *t* distribution errors

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, Winnipeg, Canada

ABSTRACT
This article constructs an approximate Bayes estimator for the parameter vector consisted of regression coefficients and variance parameter in the linear model in which the error terms follow multivariate *t* distribution. Its superiorities over the classical estimators are strictly proved in terms of the mean squared error matrix (MSEM) criterion. Compared with the Bayes estimator computed via the MCMC method, the proposed Bayes estimator is simple and easy to interpret and compute, which only requires relatively little prior designation. The numerical computations further verify that the approximate Bayes estimator performs well. Also, the proposed procedure can be easily extended to other multivariate distribution cases.

ARTICLE HISTORY
Received 2 June 2022
Accepted 13 October 2022

KEYWORDS
Linear Bayes procedure; Gibbs sampling; Bayes estimator; multivariate *t* distribution

1. Introduction

There is a growing body of literatures that recognize the multivariate normal distribution as the most widely used assumption for the distribution of error terms in linear models. However, in some cases, its application effects are controversial. For example, some random phenomena with thick tail characteristics, such as stock return rate in the financial field mentioned by Fama (1965), modeled using normal assumptions can lead to some misleading conclusions. A question which has got a lot of attention for a long time is whether *t* distribution can be introduced as the error term's assumption. According to Lin (1972), *t* distribution can be defined as the mixture of a normal distribution and a inverse gamma distribution, and specifically, an intermediate random variable obeying the inverse gamma distribution is used to adjust the spread of the normal distribution. Hence, compared with the normal distribution, West (1984) and Lange et al. (1989) indicate that *t* distribution is more flexible in characterizing tail characteristics of random phenomena, which is reflected in the ability to reduce the influence of outliers in data and enhance the robustness of statistical analysis. The readers are referred to Fernandez and Steel (1998), Chib et al. (2002) and Jacquier et al. (2004) for more details, which provide some useful accounts of how random phenomena with thick tail features can be modeled using the theory of *t* distribution.

CONTACT Lichun Wang lchwang@bjtu.edu.cn, Department of Mathematics, Beijing Jiaotong University, Beijing 100044, China.

The multivariate t can be considered as a generalization of the univariate t distribution. The density function of n dimension t distribution, denoted as $T_\nu(\mu, \Sigma, n)$, is given by

$$f_n(x) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)|\Sigma|^{-\frac{1}{2}}}{(\pi\nu)^{\frac{n}{2}}\Gamma\left(\frac{\nu}{2}\right)}\left[1 + \frac{1}{\nu}(x-\mu)'\Sigma^{-1}(x-\mu)\right]^{-\frac{\nu+n}{2}}, \tag{1.1}$$

where ν is a shape parameter (degrees of freedom), μ and Σ denote mean vector and scale matrix respectively. For the study of linear models, many researchers suggest the use of multivariate t distribution as the distribution assumption of error terms. See Sutradhar and Ali (1986), Giles (1991), Singh (1991) and Liu and Rubin (1995), etc. From the Bayesian point of view, a posterior distribution of the regression coefficients and the variance parameter has been established by Zellner (1976) under a diffuse prior, and Bayesian analysis of these two parameters are well conducted. In the case that the error term obeys an independent multivariate t distribution, Fonseca et al. (2008) develop a Bayesian analysis based on two different Jeffreys priors and show that the proposed Bayesian estimator is comparable to other estimators based on priors previously used in the literature.

The main issue of the article is to conduct a Bayesian analysis of linear regression model with an uncorrelated multivariate t error term. We propose to estimate the regression coefficients and the variance parameter of the model simultaneously via a linear Bayesian procedure, which is originally suggested by Hartigan (1969) and then discussed by Rao (1973) from linear optimization viewpoint. Consider the following linear model

$$y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \varepsilon_{n\times 1}, \tag{1.2}$$

where the error term $\varepsilon_{n\times 1} \sim T_\nu(0, \sigma^2 I, n)$ with known $\nu(>2)$, $y_{n\times 1}$ denotes the observation vector and $X_{n\times p}$ is the full column rank design matrix. Following the property of multivariate t distribution, we know that the density of y is

$$g(y;\beta,\sigma^2) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\sigma^n(\pi\nu)^{\frac{n}{2}}\Gamma\left(\frac{\nu}{2}\right)}\left[1 + \frac{1}{\nu\sigma^2}(y-X\beta)'(y-X\beta)\right]^{-\frac{\nu+n}{2}}. \tag{1.3}$$

Denote $\theta = (\beta', \sigma^2)'$. Assume that the joint prior of β and σ^2 is $H(\theta)$ and adopt the following multivariate weighted quadratic loss

$$L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)'A(\hat{\theta} - \theta), \tag{1.4}$$

where A is a positive definite matrix and $\hat{\theta}$ denotes a estimation of θ. Thus, by the Bayes theorem, the usual Bayes estimators (BE) of β and σ^2 would be given by

$$\hat{\beta}_{BE} = \int\int \beta g(y;\theta)dH(\theta)/g(y), \tag{1.5}$$

$$\hat{\sigma}^2_{BE} = \int\int \sigma^2 g(y;\theta)dH(\theta)/g(y), \tag{1.6}$$

where $g(y) = \int\int g(y;\theta)dH(\theta)$.

Obviously, the above two integrals are not easy to handle, even if some seemingly convenient priors $H(\theta)$ are selected. Therefore, in such cases, approximate Bayesian methods such as Lindley (1980) and Tierney and Kadane (1986) are raised. However, calculating the third derivative of posterior is often an arduous task of the Lindley approximation, and the Tierney and Kadane approximation needs to satisfy a main condition that the product of likelihood and prior is unimodal, which more or less limits its application. Also, using the idea of MCMC such as the Gibbs sampling procedure and the Metropolis method have been suggested in the past twenty years, see Martinez and Martinez (2007), Albert (2009) and Martino and Elvira (2017), etc. Anyway, in this case the usual Bayes estimators are somewhat intricate and not convenient to use.

Employing the linear Bayesian procedure, Lamotte (1978) develops a class of Bayes linear estimators by searching, among all linear estimators, ones which have least average total mean squared error. Since then many authors have given their attention to it, such as Robbins (1983), Maritz (1989), Samaniego and Vestrup (1999), Zhang and Wei (2005) and Wang and Singh (2014). Recently, Jones et al. (2016) has considered how to employ Bayesian linear analysis to find an optimal experimental design. In particular, Weinstein et al. (2018) uses the linear Bayesian method to construct linear empirical Bayesian estimator of the normal mean vector under heteroscedasticity.

In this article, we plan to replace the completely specified prior $H(\theta)$ by an assumption about just a few moments of the prior and then employ the linear Bayes procedure to simultaneously estimate the parameters β and σ^2. In contrast to those traditional Bayes estimators highly depending on all priori information and hardly being expressed explicitly, the proposed linear approximate Bayes estimator not only has a well-defined expression but also has properties that are easy to depict only by using the priori moments, which are easily determined in most occasions.

The organization of the remaining Sections is as follows. Section 2 constructs a simultaneously linear approximate Bayes estimator for the regression coefficients and the variance parameter. Section 3 states some theoretical results which show that the approximate Bayes estimator is superior to the least squared estimator (LSE) and the maximum likelihood estimator (MLE) in terms of the mean square error matrix (MSEM) criterion. Section 4 compares the approximate Bayes estimator with the BE obtained via the MCMC method and further illustrates its superiorities by numerical computations. Finally, we make some conclusions in Section 5.

2. The proposed approximate Bayes estimator

For the model (1.2), we know that the least square estimators (LSE) of β and σ^2 are

$$\hat{\beta}_{LS} = (X'X)^{-1}X'y \quad \text{and} \quad \hat{\sigma}^2_{LS} = \frac{\nu-2}{\nu}\frac{y'(I-P_X)y}{n-p}, \tag{2.1}$$

where $P_X = X(X'X)^{-1}X'$ and $E(\hat{\sigma}^2_{LS}) = \frac{\sigma^2}{(n-p)}tr(I-P_X) = \sigma^2$.

Thus, we have the following two conclusions.

Theorem 2.1 *$\hat{\beta}_{LS}$ and $\hat{\sigma}^2_{LS}$ are conditionally uncorrelated, i.e., $Cov[(\hat{\beta}_{LS}, \hat{\sigma}^2_{LS})|\theta] = 0$.*

Proof. Because $y|\theta \sim T_\nu(X\beta, \sigma^2 I, n)$, one can write

$$y|\tau^2, \theta \sim N_n(X\beta, \tau^2\sigma^2 I), \tag{2.2}$$

$$\tau^2 \sim IG(\nu/2, \nu/2). \tag{2.3}$$

Put $C = (X'X)^{-1}X'$, then

$$\begin{aligned}
Cov\Big[(\hat\beta_{LS}, \hat\sigma^2_{LS})|\theta\Big] &= Cov\left\{\left[Cy, \frac{\nu-2}{\nu(n-p)}y'(I-P_X)y\right]\Big|\theta\right\} \\
&= C\frac{\nu-2}{\nu(n-p)}Cov\{[y, y'(I-P_X)y]|\theta\} \\
&= C\frac{\nu-2}{\nu(n-p)}\{Cov[E(y|\tau^2,\theta), E(y'(I-P_X)y|\tau^2,\theta)|\theta] + E[Cov(y, y'(I-P_X)y|\tau^2,\theta)|\theta]\} \\
&= C\frac{\nu-2}{\nu(n-p)}\{Cov[X\beta, (n-p)\sigma^2\tau^2|\theta] + E[2\sigma^2\tau^2(I-P_X)X\beta|\theta]\} \\
&= 0.
\end{aligned} \tag{2.4}$$

Therefore, given $\theta = (\beta', \sigma^2)'$, we conclude that $\hat\beta_{LS}$ and $\hat\sigma^2_{LS}$ are conditionally uncorrelated.

Theorem 2.1 is proved.

Set $\hat\theta_{LS} = (\hat\beta'_{LS}, \hat\sigma^2_{LS})'$. Define the class of linear estimators of the parameter vector θ as $\mathcal{R} = \{\hat\theta : \hat\theta = B\hat\theta_{LS} + b\}$, where B and b are unknown matrix and vector respectively. Under the loss (1.4), the best linear Bayes estimator, say $\hat\theta_{LBE}$, is searched by satisfying the following conditions

$$R(\hat\theta_{LBE}, \theta) = \min_{B,b} E_{(y,\theta)}\Big[(\hat\theta - \theta)'A(\hat\theta - \theta)\Big], \tag{2.5}$$

$$E_{(y,\theta)}(\hat\theta_{LBE} - \theta) = 0, \tag{2.6}$$

where $E_{(y,\theta)}$ denotes the joint expectation with respect to y and $\theta = (\beta', \sigma^2)'$.

Theorem 2.2. *Under the condition that the prior distribution $H(\theta)$ belongs to the family $\mathcal{H} = \{H(\theta) : E(\theta)^2 < \infty\}$, the expression of the linear Bayes estimator $\hat\theta_{LBE}$ is*

$$\hat\theta_{LBE} = Cov(\theta)[W + Cov(\theta)]^{-1}\hat\theta_{LS} + W[W + Cov(\theta)]^{-1}E\theta, \tag{2.7}$$

where for $\nu > 4$,

$$W = E\Big[Cov(\hat\theta_{LS}|\theta)\Big] = diag\left\{\frac{\nu}{\nu-2}(X'X)^{-1}E(\sigma^2), \frac{2(n-p+\nu-2)}{(n-p)(\nu-4)}E(\sigma^4)\right\}. \tag{2.8}$$

Proof. From (2.6), $b = E\theta - BE\theta$. Then

$$\begin{aligned} R(\hat{\theta}, \theta) &= E_{(y,\theta)}\left[(\hat{\theta}-\theta)'A(\hat{\theta}-\theta)\right] \\ &= E_{(y,\theta)}(B\hat{\theta}_{LS}+E\theta-BE\theta-\theta)'A(B\hat{\theta}_{LS}+E\theta-BE\theta-\theta) \\ &= tr\left\{ABE_{(y,\theta)}\left[(\hat{\theta}_{LS}-E\theta)(\hat{\theta}_{LS}-E\theta)'\right]B'\right\}+tr[ACov(\theta)] \\ &\quad - 2tr\left[ACov(\theta)B'\right]. \end{aligned} \tag{2.9}$$

For given θ, according to Theorem 2.1, $Cov[(\hat{\beta}_{LS}, \hat{\sigma}^2_{LS})|\theta] = 0$. Therefore,

$$\begin{aligned} &E_{(y,\theta)}\left[(\hat{\theta}_{LS}-E\theta)(\hat{\theta}_{LS}-E\theta)'\right] \\ &= E\left[Cov(\hat{\theta}_{LS}|\theta)\right]+Cov\left[E(\hat{\theta}_{LS}|\theta)\right] \\ &= W + Cov(\theta). \end{aligned} \tag{2.10}$$

In addition, note that

$$Var(\hat{\beta}|\theta) = \frac{\nu}{(\nu-2)}(X'X)^{-1}\sigma^2 \tag{2.11}$$

and

$$\begin{aligned} Var(\hat{\sigma}^2_{LS}|\theta) &= \frac{(\nu-2)^2}{(n-p)^2\nu^2}Var\left[y'(I-P_X)y|\theta\right] \\ &= \frac{(\nu-2)^2}{(n-p)^2\nu^2}\left\{E\left[Var(y'(I-P_X)y|\theta,\tau^2)|\theta\right]+Var\left[E(y'(I-P_X)y|\theta,\tau^2)|\theta\right]\right\} \\ &= \frac{(\nu-2)^2}{(n-p)^2\nu^2}\{2E\left[(n-p)\tau^4\sigma^4|\theta\right]+(n-p)^2Var(\tau^2\sigma^2|\theta)\} \\ &= \frac{(\nu-2)^2}{(n-p)^2\nu^2}\left\{2(n-p)\left[\frac{\nu^2/4}{(\nu/2-1)^2}+\frac{\nu^2/4}{(\nu/2-1)^2(\nu/2-2)}\right]\right. \\ &\quad \left.+(n-p)^2\left[\frac{\nu^2/4}{(\nu/2-1)^2(\nu/2-2)}\right]\right\}\sigma^4 \\ &= \frac{2(n-p+\nu-2)}{(n-p)(\nu-4)}\sigma^4. \end{aligned} \tag{2.12}$$

Thus, combining (2.11), (2.12) and Theorem 2.1 yields $W = diag\{\frac{\nu}{(\nu-2)}(X'X)^{-1} E(\sigma^2), \frac{2(n-p+\nu-2)}{(n-p)(\nu-4)}E(\sigma^4)\}$. Inserting (2.10) into (2.9) and setting $\frac{\partial R(\hat{\theta},\theta)}{\partial B} = 0$, we have

$$AB[W + Cov(\theta)] - ACov(\theta) = 0,$$

which leads to

$$B = I_{p+1} - W[W + Cov(\theta)]^{-1}. \tag{2.13}$$

Together with $b = (I - B)E\theta$, we come to the conclusion of Theorem 2.2.

Theorem 2.2 is proved.

Remark 2.1. In fact, the constraint on unbiasedness by (2.6) is unnecessary, which means that the expression of linear Bayes estimator has nothing to do with the constraint. Therefore, the requirement for unbiasedness does not impose additional restriction on the linear estimator.

Remark 2.2. It is worth noting that the proposed LBE combines the information of the observed data with the prior information. It is seen from the formula (2.7) that the expression of LBE is the weighted sum of the LSE and the prior mean.

3. The superiorities of the approximate Bayes estimator

Theorem 3.1. *Under the MSEM criterion, $\hat{\theta}_{LBE}$ is superior to $\hat{\theta}_{LS}$.*

Proof. Since $E_{(y,\theta)}(\hat{\theta}_{LBE}-\theta)=0$, we have

$$\begin{aligned}\text{MSEM}(\hat{\theta}_{LBE}) &= E_{(y,\theta)}\left[(\hat{\theta}_{LBE}-\theta)(\hat{\theta}_{LBE}-\theta)'\right]\\ &= E\left[Cov(\hat{\theta}_{LBE}-\theta|\theta)\right]+Cov\left[E(\hat{\theta}_{LBE}-\theta|\theta)\right].\end{aligned} \tag{3.1}$$

Let $M=[W+Cov(\theta)]^{-1}$, we further have

$$\begin{aligned}\text{MSEM}(\hat{\theta}_{LBE}) &= (I-WM)W(I-WM)'+WMCov(\theta)(WM)'\\ &= W-2WMW+WM[W+Cov(\theta)]MW\\ &= W-WMW.\end{aligned} \tag{3.2}$$

Otherwise

$$\begin{aligned}\text{MSEM}(\hat{\theta}_{LS}) &= E_{(y,\theta)}\left[(\hat{\theta}_{LS}-\theta)(\hat{\theta}_{LS}-\theta)'\right]\\ &= E\left\{E\left[(\hat{\theta}_{LS}-\theta)(\hat{\theta}_{LS}-\theta)'|\theta\right]\right\}=W.\end{aligned} \tag{3.3}$$

Comparing (3.2) with (3.3), we have

$$\text{MSEM}(\hat{\theta}_{LBE}) \le \text{MSEM}(\hat{\theta}_{LS}). \tag{3.4}$$

Theorem 3.1 is proved.

Lemma 3.1. *For the model (1.2), the MLE of θ is $\hat{\theta}_{ML}=\left(\hat{\beta}'_{LS},\frac{(y-X\hat{\beta}_{LS})'(y-X\hat{\beta}_{LS})}{n}\right)'$.*

Proof. See the Appendix.

Theorem 3.2. *Let $\hat{\theta}_{LBE}$ and $\hat{\theta}_{ML}$ be given by Theorem 2.2 and Lemma 3.1 respectively. If $n>\frac{\nu p}{2}$ and $\nu>4$, then MSEM$(\hat{\theta}_{LBE})\le$ MSEM $(\hat{\theta}_{ML})$.*

Proof. See the Appendix.

4. Numerical simulations

4.1. Settings of the model and the priors of the parameters

In this Section, numerical simulations are carried out to study the performances of the $\hat{\theta}_{LBE}$ and the BE $\hat{\theta}_{BE}$. We set $p=2$, $\nu=5$, 15, 35 and let the true value of $\theta = (\beta_0, \beta_1, \sigma^2)'$ be equal to $(2, -1, 3)'$. The design matrix would be

$$X = (1_n \quad x)_{n\times 2},$$

where $1'_n = (1, 1, \ldots, 1)$ and x is a constant vector which is generated from $N(0, I_n)$. Once X is generated, the observed vector y is simulated by the multivariate t distribution, $T_n(X\beta, \sigma^2 I_n, \nu)$, which can also be expressed in a hierarchical structure:

$$y|\tau^2 \sim N_n(X\beta, \tau^2\sigma^2 I_n), \tag{4.1}$$

$$\tau^2 \sim IG(\nu/2, \nu/2). \tag{4.2}$$

The following prior form is adopted. Let the prior of β be $N(\tilde{\mu}, \Sigma)$, where $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ in which ρ is correlation coefficient, and $\tilde{\mu} = (\mu_0, \mu_1)'$. σ^2 has an inverted Gamma prior with density

$$\pi(\sigma^2) \propto (\sigma^2)^{-(r+1)} \exp\left(-\frac{\lambda}{\sigma^2}\right), r > 2, \lambda > 0. \tag{4.3}$$

Therefore, the joint posterior density of β, σ^2 and τ^2 is

$$\begin{aligned}\pi(\beta, \sigma^2, \tau^2|y) \propto \tau^{2(-n/2-\nu/2-1)}(\sigma^2)^{-(n/2+r+1)} \exp\Big\{ -\frac{1}{2}\Big[(y-X\beta)'(y-X\beta)/(\tau^2\sigma^2) \\ + (\beta-\tilde{\mu})'\Sigma^{-1}(\beta-\tilde{\mu}) + \frac{\nu}{\tau^2} + 2\frac{\lambda}{\sigma^2}\Big]\Big\}.\end{aligned} \tag{4.4}$$

It is hard to calculate a posterior expectation of θ, which involves multiple integrals, so one can obtain the $\hat{\theta}_{BE}$ numerically via the MCMC method. To this end, the full conditional posterior distribution of β is given by

$$\beta|\sigma^2, \tau^2, y \sim N(\mu_{(1)}, \Sigma_{(1)}), \tag{4.5}$$

where $\Sigma_{(1)} = (X'X/(\tau^2\sigma^2) + \Sigma^{-1})^{-1}$ and $\mu_{(1)} = \Sigma_{(1)}(X'y/(\tau^2\sigma^2) + \Sigma^{-1}\tilde{\mu})$.

The full conditional posterior distribution of σ^2 is given by

$$\sigma^2|\beta, \tau^2, y \sim IG(n/2 + r, (y-X\beta)'(y-X\beta)/(2\tau^2) + \lambda). \tag{4.6}$$

Also, the full conditional posterior distribution of τ^2 is

$$\tau^2|\beta, \sigma^2, y \sim IG\left(\frac{n+\nu}{2}, (y-X\beta)'(y-X\beta)/(2\sigma^2) + \nu/2\right). \tag{4.7}$$

We employ the MCMC method to obtain the numerical solution of $\hat{\theta}_{BE}$. The steps of the Gibbs sampling (see Albert (2009)) are given below.

Table 1. Various prior distributions.

Prior	Prior distribution	$tr[Cov(\theta)]$
Pr1:	$\beta \sim N\left(\begin{pmatrix} 1 \\ -1.5 \end{pmatrix}, \begin{pmatrix} 10^2 & -75 \\ -75 & 10^2 \end{pmatrix}\right)$, $\sigma^2 \sim IG(8, 21)$	201.5
Pr2:	$\beta \sim N\left(\begin{pmatrix} 1 \\ -1.5 \end{pmatrix}, \begin{pmatrix} 10^2 & 75 \\ 75 & 10^2 \end{pmatrix}\right)$, $\sigma^2 \sim IG(3, 21)$	310.25

Step 1. Choose the initial values of β, σ^2 and τ^2, say $\beta^{(0)}$, $\sigma^{2(0)}$ and $\tau^{2(0)}$, and note the values of β, σ^2 and τ^2 at the j-th step by $\beta^{(j)}$, $\sigma^{2(j)}$ and $\tau^{2(j)}$.
Step 2. Update the (j + 1)-th iteration as follows
(a) Sample $\beta^{(j+1)}$ from distribution $\pi(\beta|\sigma^{2(j)}, \tau^{2(j)}, y)$;
(b) Sample $\sigma^{2(j+1)}$ from distribution $\pi(\sigma^2|\beta^{(j+1)}, \tau^{2(j)}, y)$;
(c) Sample $\tau^{2(j+1)}$ from distribution $\pi(\tau^2|\beta^{(j+1)}, \sigma^{2(j+1)}, y)$.
Step 3. Repeat Step 2 N times.
Step 4. Calculate the average of the N samples to get the posterior expectation.

In the above steps, we introduce the parameter τ^2 to simplify the iterations but we are not interested in it. For the following numerical studies, we take $N = 5000$.

The prior distributions involved in the above simulations are listed in Table 1. In simulations, the distance between the LBE and the BE is defined by $\| \hat{\theta}_{LBE} - \hat{\theta}_{BE} \| = \sqrt{(\hat{\beta}_{LBE} - \hat{\beta}_{BE})'(\hat{\beta}_{LBE} - \hat{\beta}_{BE}) + (\hat{\sigma^2}_{LBE} - \hat{\sigma}^2_{BE})^2}$. The smaller value indicates that the LBE and the BE are closer.

Here $tr[Cov(\theta)]$ describes the variation of the prior. Note that the large variances in the prior are considered. The purpose is that we expect the experimental results to be somewhat robust with respect to prior distributions.

4.2. Numerical computations and comparisons

Figures 1 and 2 plot the distances between the LBE and the BE for different priors, respectively. It can be seen that in the case of $v = 5$ the distances between the LBE and the BE are small already. Further, when the prior distributions are the same, taking larger values of the degrees of freedom is more favorable for the approximation effect of the LBE. The reason is that when the degrees of freedom v is getting larger the t distribution is approaching the normal distribution. At this point, in our experimental setup the LBE and the BE corresponding to the component β are very close to each other, which causes the difference between the LBE and the BE of θ to be smaller. Besides, as the prior variance becomes small, the distance gets smaller, which implies a better approximation of the LBE at this time.

Since in the theoretical part we obtain the conclusion that the LBE outperforms the MLE under the MSEM criterion when $n > \frac{pv}{2}$, we will examine the superiority of the LBE in the case of $n \leq \frac{pv}{2}$. The numerical results displayed in Table 2 show that MSEM($\hat{\theta}_{ML}$) >MSEM($\hat{\theta}_{LBE}$) for different combinations of n and v, which satisfy the condition $n \leq \frac{pv}{2}$.

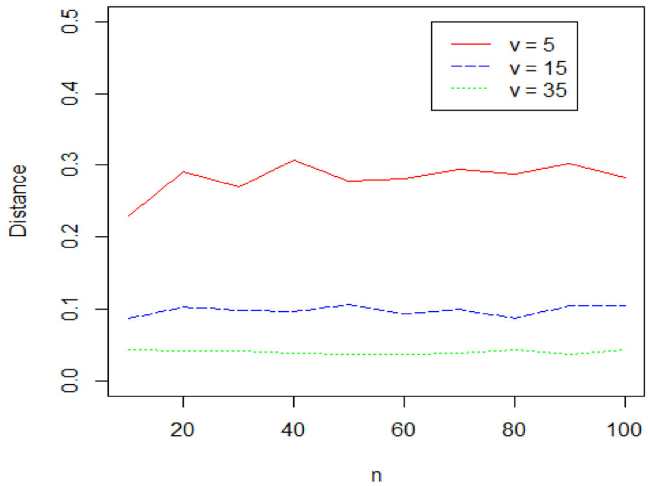

Figure 1. The distance between $\hat{\theta}_{LBE}$ and $\hat{\theta}_{BE}$ for Pr 1.

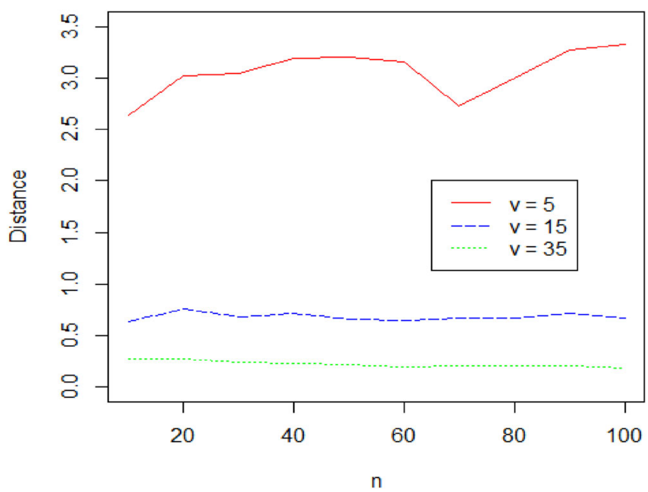

Figure 2. The distance between $\hat{\theta}_{LBE}$ and $\hat{\theta}_{BE}$ for Pr 2.

4.3. Robustness analysis of the LBE

In the previous simulations, the real data is generated from multivariate t-distribution. In this subsection, to investigate the robustness of the proposed LBE based on multivariate t regression, the data is intentionally generated from multivariate Normal and multivariate Laplace (ML) distributions. We use the distance between the estimate and the true value as a criterion, which is defined by $\| \hat{\theta} - \theta \| = \sqrt{(\hat{\beta} - \beta)'(\hat{\beta} - \beta) + (\hat{\sigma}^2 - \sigma^2)^2}$, where $\hat{\theta}$ can be $\hat{\theta}_{LBE}$, $\hat{\theta}_{LS}$ and $\hat{\theta}_{ML}$. We set $p=2$, $\nu=5$ and use Pr1 as the prior of θ.

In Figure 3, the data y is simulated from $T_n(X\beta, \sigma^2 I_n, \nu)$. Fixing the expectation and the covariance matrix, in Figures 4 and 5, the data y is sampled from $N\left(X\beta, \frac{\nu\sigma^2}{\nu-2} I_n\right)$ and $ML\left(\frac{\nu\sigma^2}{\nu-2}, X\beta, I_n\right)$, respectively. It is seen from Figures 3–5 that $\| \hat{\theta}_{LBE} - \theta \|$ is

Table 2. The MSEM($\hat{\theta}_{ML}$) and MSEM($\hat{\theta}_{LBE}$) under two priors.

Prior	n, v	MSEM($\hat{\theta}_{ML}$)–MSEM($\hat{\theta}_{LBE}$)
Pr1:	$n=5,\ v=15$	$\begin{pmatrix} 0.0421 & 0.0873 & 0.0000 \\ 0.0874 & 0.1966 & 0.0000 \\ 0.0000 & 0.0000 & 4.5667 \end{pmatrix}$
	$n=10,\ v=15$	$\begin{pmatrix} 0.0038 & 0.0037 & 0.0000 \\ 0.0037 & 0.0051 & 0.0000 \\ 0.0000 & 0.0000 & 3.1777 \end{pmatrix}$
	$n=20,\ v=35$	$\begin{pmatrix} 0.0006 & 0.0006 & 0.0000 \\ 0.0006 & 0.0009 & 0.0000 \\ 0.0000 & 0.0000 & 0.9285 \end{pmatrix}$
Pr2:	$n=5,\ v=15$	$\begin{pmatrix} 0.0880 & 0.0713 & 0.0000 \\ 0.0713 & 1.0691 & 0.0000 \\ 0.0000 & 0.0000 & 50.6143 \end{pmatrix}$
	$n=10,\ v=15$	$\begin{pmatrix} 0.0259 & -0.0172 & 0.0000 \\ -0.0172 & 0.0385 & 0.0000 \\ 0.0000 & 0.0000 & 37.1325 \end{pmatrix}$
	$n=20,\ v=35$	$\begin{pmatrix} 0.0064 & -0.0055 & 0.0000 \\ -0.0055 & 0.0099 & 0.0000 \\ 0.0000 & 0.0000 & 7.6653 \end{pmatrix}$

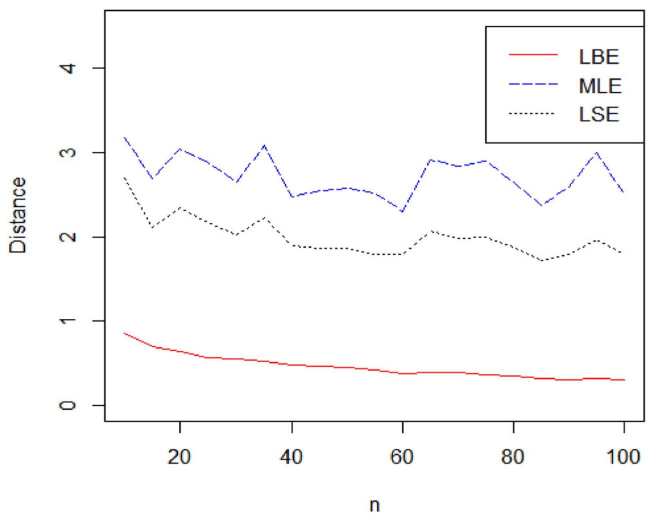

Figure 3. The distance between estimators and real value for multivariate t distribution.

always smaller than $\| \hat{\theta}_{LS} - \theta \|$ and $\| \hat{\theta}_{ML} - \theta \|$ even the error distribution is changed. Then we can conclude that the LBE is somewhat robust to the error distribution.

4.4. An application to real data

The real data set from Siegel (1977) is applied to illustrate the results presented in this Subsection. These data have also been analyzed by Sheather (2009) and Yang and Yuan (2017). We take the bid price as the dependent variable and the coupon rate as the regressor variable. As Yang analyzed, it is appropriate to analyze the data by using Student-t linear regression model with the degrees of freedom v and the scale parameter σ^2.

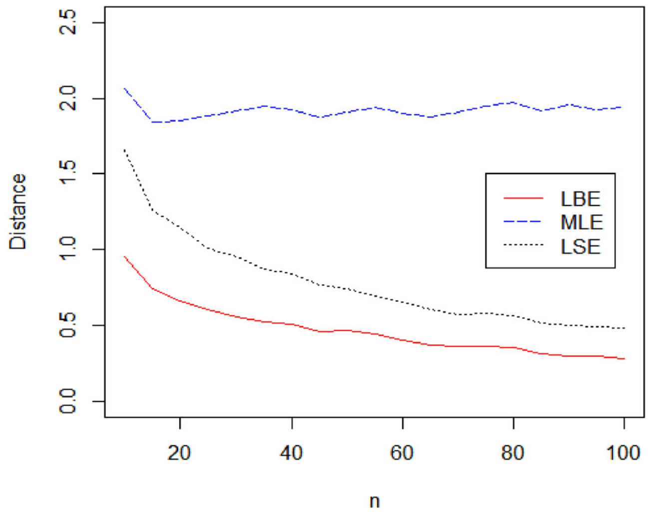

Figure 4. The distance between estimators and real value for multivariate Normal distribution.

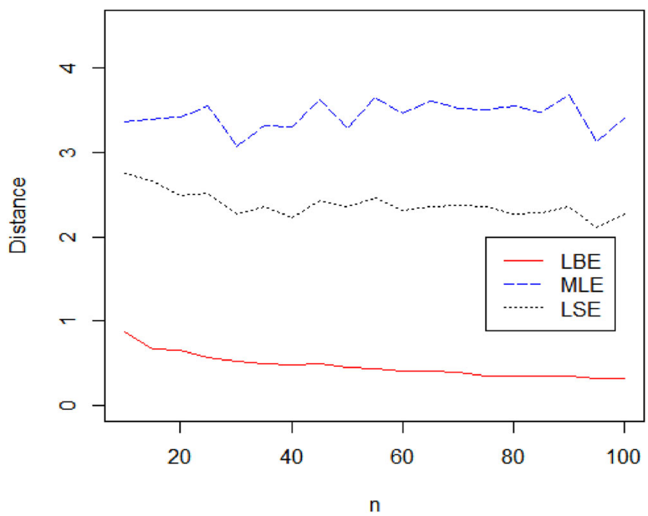

Figure 5. The distance between estimators and real value for multivariate Laplace distribution.

Table 3. Estimations for different v and (λ, r).

v	(λ, r)	$\hat{\theta}_{LBE}$	$\hat{\theta}_{BE}$	$\parallel \hat{\theta}_{LBE} - \hat{\theta}_{BE} \parallel$
5	(3)	(−16.1062, 0.2450, 1.3763)′	(−16.1091, 0.2451, 1.3113)′	0.07956
	(2.5, 11.25)	(−16.0762, 0.2447, 5.9404)′	(−16.1367, 0.2453, 2.9121)′	3.0290
20	(3)	(−16.1099, 0.2451, 1.3230)′	(−16.1078, 0.2450, 1.3430)′	0.04525
	(2.5, 11.25)	(−16.08633, 0.2448, 2.6577)′	(−16.1163, 0.2451, 2.2777)′	0.3813
50	(3)	(−16.1103, 0.2451, 1.3667)′	(−16.1008, 0.2450, 1.3714)′	0.03855
	(2.5, 11.25)	(−16.0878, 0.2449, 2.1885)′	(−16.0966, 0.2449, 2.0793)′	0.1095

We assume that $\beta \sim N((-16, 0.5)', 100I_2)$ and $\sigma^2 \sim IG(\lambda, r)$. We analyze the data by using several different values of (λ, r) and the degrees of freedom v. The values of $\hat{\theta}_{LBE}$ in Table 3 are obtained by Theorem 2.2. To visually compare the magnitudes of the

Table 4. The eigenvalues of matrices MSEM $(\hat{\theta}_{LS})$-MSEM $(\hat{\theta}_{LBE})$ and MSEM $(\hat{\theta}_{ML})$-MSEM $(\hat{\theta}_{LBE})$.

v	(λ, r)	λ_{LS}	λ_{ML}
5	(3)	(7.9877, 1.1532, 4.6362e-13)	(23.8838, 1.1532, 4.6362e-13)
	(2.5, 11.25)	(6.3877e + 02, 6.0191e + 01, 4.6327e-11)	(1.8309e + 03, 6.0191e + 01, 4.6327e-11)
20	(3)	(5.3056e-01, 2.4226e-01, 2.0616e-13)	(5.3056e-01, 3.3723e-01, 2.0616e-13)
	(2.5, 11.25)	(3.2508e + 01, 2.3922e + 01, 2.0599e-11)	(3.2508e + 01, 3.1044e + 01, 2.0599e-11)
50	(3)	(4.6837e-01, 8.4472e-02, 1.8113e-13)	(4.6837e-01, 6.8908e-02, 1.8113e-13)
	(2.5, 11.25)	(2.9361e + 01, 8.7348, 1.8101e-11)	(2.9361e + 01, 7.5676, 1.8101e-11)

matrices, we present the eigenvalues of matrices MSEM $(\hat{\theta}_{LS})$-MSEM $(\hat{\theta}_{LBE})$ and MSEM $(\hat{\theta}_{ML})$-MSEM $(\hat{\theta}_{LBE})$ in Table 4 and the eigenvalues of these two matrices are denoted as λ_{LS} and λ_{ML} respectively.

As can be seen from Table 3, except for the case of $v = 5, \lambda = 2.5, r = 11.25,$ the LBE and the BE are close to each other, also the distance $\| \hat{\theta}_{LBE} - \hat{\theta}_{BE} \|$ gradually becomes smaller as v becomes larger, which indicates that v has some influence on the LBE and the BE. Therefore, the estimation of v could be an important step in making inference with this kind of models.

On the other hand, from Table 4 we know that the MSEMs of LBE are always smaller than those of the other two classical estimates for the real data. Combined with the above analysis, the LBE $\hat{\theta}_{LBE}$ is feasible and applicable in this practical application case.

5. Conclusions

The multivariate t distribution is widely used in many fields, see Kotz and Nadarajah (2004), Kibria and Joarder (2006) and among others. This article employs a linear Bayes procedure to propose an approximate Bayes estimator for the parameter vector consisting of the regression coefficients and the variance parameter in the linear model whose error terms obey an uncorrelated multivariate t distribution. The proposed linear estimator only designates some prior moments instead of some specific priors and is simple and easy to compute. In terms of the mean square error matrix criterion, its superiorities over the classical estimators such as the least squared estimator and the maximum likelihood estimator are strictly proved. The numerical computations show that the proposed linear Bayes estimator is a good approximation, regardless of whether the degrees of freedom is large or small. Also, a real data example further verifies that the linear approximate Bayes estimator is feasible and applicable. Moreover, the procedure used in this article can be further extended to some elliptically contoured distributions such as multivariate normal and multivariate Laplace, see Fang (1987) and Eltoft et al. (2006), etc.

Acknowledgments

We would like to express our appreciations to Dr. Qi Zhou and Dr. Shaoqian Huang who provided assistance in preparing this article.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

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

References

Albert, J. 2009. *Bayesian computation with R*, New York: Springer.

Chib, S., F. Nardari, and N. Shephard. 2002. Markov chain monte carlo methods for stochastic volatility models. *Journal of Econometrics* 108 (2):281–316. doi:10.1016/S0304-4076(01)00137-3.

Eltoft, T., T. Kim, and T. W. Lee. 2006. On the multivariate Laplace distribution. *IEEE Signal Processing Letters* 13 (5):300–3. doi:10.1109/LSP.2006.870353.

Fama, E. F. 1965. The behaviour of stock market prices. *The Journal of Business* 38 (1):34–105. doi:10.1086/294743.

Fang, K. T. 1987. Introduction to generalized multivariate analysis. *Advances in Mathematics* 16: 1–15.

Fernandez, C., and M. F. J. Steel. 1998. On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* 93 (441):359–71. doi:10.2307/2669632.

Fonseca, T. C. O., M. A. R. Ferreira, and H. S. Migon. 2008. Objective Bayesian analysis for the Student-t regression model. *Biometrika* 95 (2):325–33. doi:10.1093/biomet/asn001.

Giles, A. J. 1991. Pretesting for linear restrictions in a regression model with spherically symmetric distributions. *Journal of Econometrics* 50 (3):377–98. doi:10.1016/0304-4076(91)90026-A.

Hartigan, J. A. 1969. Linear Bayes methods. *Journal of the Royal Statistical Society B* 31 (3): 446–54. doi:10.1111/j.2517-6161.1969.tb00804.x.

Jacquier, E., N. G. Polson, and P. E. Rossi. 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* 122 (1):185–212. doi:10.1016/j.jeconom.2003.09.001.

Jones, M., M. Goldstein, P. Jonathan, and D. Randell. 2016. Bayes linear analysis for Bayesian optimal experimental design. *Journal of Statistical Planning and Inference* 171:115–29. doi:10.1016/j.jspi.2015.10.011.

Kibria, B. M. G, and A. H. Joarder. 2006. A short review of multivariate t-distribution. *Journal of Statistical Research* 40:256–422.

Kotz, S, and S. Nadarajah. 2004. *Multivariate t-distributions and their applications*. Cambridge: Cambridge University Press.

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

Lange, K. L., R. J. A. Little, and J. M. G. Taylor. 1989. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association* 84 (408):881–96. doi:10.2307/2290063.

Lin, P. E. 1972. Some characterizations of the multivariate t distribution. *Journal of Multivariate Analysis* 2 (3):339–44. doi:10.1016/0047-259X(72)90021-8.

Lindley, D. V. 1980. Approximate Bayesian methods. *Trabajos de Stadistica 31* 31 (1):223–45. doi:10.1007/BF02888353.

Liu, C, and D. B. Rubin. 1995. ML estimation of the t distribution using EM and its extension ECM and ECME. *Statistica Sinica* 5:19–39.

Maritz, J. S. 1989. Linear empirical Bayes estimation of quantiles. *Statistics & Probability Letters* 8 (1):59–65. doi:10.1016/0167-7152(89)90084-9.

Martinez, W. L, and A. R. Martinez. 2007. *Computational statistics handbook with MATLAB*, 2nd ed, New York: Chapman & Hall/CRC.

Martino, L, and V. Elvira. 2017. *Metropolis sampling*, New York: Wiley StatsRef: Statistics Reference Online: 1–18.

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

Robbins, H. 1983. Some thoughts on empirical Bayes estimation. *Annals of Statistics* 3:713–23.

Samaniego, F. J, and E. Vestrup. 1999. On improving standard estimators via linear empirical Bayes methods. *Statistics & Probability Letters* 44 (3):309–18. doi:10.1016/S0167-7152(99)00022-X.
Sheather, S. 2009. *A modern approach to regression with R*, New York: Springer.
Siegel, A. 1977. *Practical business statistics*. Boston: Irwin McGraw-Hill.
Singh, R. S. 1991. James-stein rule estimators in linear regression models with multivariate-t distributed error. *Australian Journal of Statistics* 33 (2):145–58. doi:10.1111/j.1467-842X.1991.tb00422.x.
Sutradhar, B. C, and M. M. Ali. 1986. Estimation of the parameters of a regression model with a multivariate t error variable. *Communications in Statistics-Theory and Methods* 15 (2):429–50. doi:10.1080/03610928608829130.
Tierney, L, and J. B. Kadane. 1986. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* 81 (393):82–6. doi:10.1080/01621459.1986.10478240.
Wang, L. C, and R. S. Singh. 2014. Linear Bayes estimator for the two-parameter exponential family under type II censoring. *Computational Statistics and Data Analysis* 71:633–42. doi:10.1016/j.csda.2013.07.020.
Weinstein, A., Z. Ma, L. D. Brown, and C. H. Zhang. 2018. Group-linear empirical Bayes estimates for a heteroscedastic normal mean. *Journal of the American Statistical Association* 113 (522):698–710. doi:10.1080/01621459.2017.1280406.
West, M. 1984. Outlier models and prior distributions in Bayesian linear regression. *Journal of the Royal Statistical Society Series B* 46 (3):431–9. doi:10.1111/j.2517-6161.1984.tb01317.x.
Yang, F, and H. Yuan. 2017. A non-iterative Bayesian sampling algorithm for linear regression models with scale mixtures of normal distributions. *Computational Economics* 49 (4):579–97. doi:10.1007/s10614-016-9580-5.
Zellner, A. 1976. Bayesian and non-Bayesian analysis of the regression model with multivariate student-t error terms. *Journal of the American Statistical Association* 71 (354):400–5. doi:10.2307/2285322.
Zhang, W. P, and L. S. Wei. 2005. On Bayes linear unbiased estimation of estimable functions for the singular linear model. *Science in China Series A* 48 (7):898–903. doi:10.1360/04ys0043.

Appendix

The proof of Lemma 3.1.

Note that the likelihood function of (β, σ^2) can be written as

$$L\left(\beta, \sigma^2|y\right) = l(\nu)(\sigma^2)^{-\frac{n}{2}}\left\{\nu + \frac{(y-X\beta)'(y-X\beta)}{\sigma^2}\right\}^{-\frac{n+\nu}{2}}, \tag{A1}$$

where $l(\nu) = \nu^{\frac{\nu}{2}}\pi^{-\frac{n}{2}}\Gamma\left(\frac{\nu+n}{2}\right)/\Gamma\left(\frac{\nu}{2}\right)$. First, by $\frac{\partial \ln L(\beta,\sigma^2|y)}{\partial\beta} = 0$, we have

$$\hat{\beta}_{ML} = (X'X)^{-1}X'y = \hat{\beta}_{LS}. \tag{A2}$$

Next, by the derivation result that

$$\frac{\partial\left(\ln\left(a_0 + \frac{b_0}{x}\right)\right)}{\partial x} = \frac{\partial(\ln(a_0x+b_0) - \ln x)}{\partial x} = \frac{a_0}{a_0x+b_0} - \frac{1}{x}, \tag{A3}$$

where a_0 and b_0 are constants, we have

$$\frac{\partial \ln L\left(\beta, \sigma^2|y\right)}{\partial\sigma^2} = \frac{\nu}{2\sigma^2} - \frac{\nu(n+\nu)}{2\left[\nu\sigma^2 + (y-X\beta)'(y-X\beta)\right]}. \tag{A4}$$

Setting $\frac{\partial \ln L(\beta, \sigma^2|y)}{\partial \sigma^2} = 0$ we have

$$\hat{\sigma}^2_{ML} = \frac{\left(y - X\hat{\beta}_{LS}\right)'\left(y - X\hat{\beta}_{LS}\right)}{n}. \tag{A5}$$

Therefore, we obtain the MLE of θ.
Lemma 3.1 is proved.

The proof of Theorem 3.2.
Note that $\hat{\theta}_{ML} = B_0\hat{\theta}_{LS}$ with $B_0 = \mathrm{diag}(I_p, \frac{\nu(n-p)}{n(\nu-2)})$. Thus,

$$\begin{aligned}\mathrm{MSEM}\left(\hat{\theta}_{ML}\right) &= B_0 W B_0' + (B_0 - I_{p+1})E\theta\theta'(B_0 - I_{p+1})' \\ &= \begin{pmatrix} \frac{\nu}{\nu-2}(X'X)^{-1}E\sigma^2 & 0 \\ 0 & \frac{\left[2\nu^2(n-p)(n-p+\nu-2)+(2n-\nu p)^2(\nu-4)\right]E\sigma^4}{(\nu-2)^2 n^2(\nu-4)} \end{pmatrix},\end{aligned} \tag{A6}$$

where $H = \frac{\nu}{\nu-2}(X'X)^{-1}$. Denote $Q = HE\sigma^2 + Cov(\beta)$, we have

$$\begin{aligned} M &= (W + Cov(\theta))^{-1} \\ &= \begin{pmatrix} Q^{-1} + Q^{-1}Cov\left(\beta, \sigma^2\right)\Delta^{-1}Cov\left(\beta', \sigma^2\right)Q^{-1} & -Q^{-1}Cov\left(\beta, \sigma^2\right)\Delta^{-1} \\ -\Delta^{-1}Cov\left(\beta', \sigma^2\right)Q^{-1} & \Delta^{-1} \end{pmatrix},\end{aligned} \tag{A7}$$

where

$$\Delta = Var(\sigma^2) + \frac{2(n-p+\nu-2)}{(n-p)(\nu-4)}E\sigma^4 - Cov\left(\beta', \sigma^2\right)Q^{-1}Cov\left(\beta, \sigma^2\right). \tag{A8}$$

Rewrite the MSEM of the LBE as

$$\mathrm{MSEM}\left(\hat{\theta}_{LBE}\right) = W - WMW = WMCov(\theta) = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}, \tag{A9}$$

where

$$R_{11} = HE\sigma^2 Q^{-1}Cov(\beta) - HE\sigma^2 Q^{-1}Cov(\beta, \sigma^2)\Delta^{-1}Cov(\beta', \sigma^2)Q^{-1}HE\sigma^2, \tag{A10}$$

$$R_{12} = \frac{2(n-p+\nu-2)E\sigma^4}{(n-p)(\nu-4)}HE\sigma^2 Q^{-1}Cov\left(\beta, \sigma^2\right)\Delta^{-1}, \tag{A11}$$

$$R_{21} = \frac{2(n-p+\nu-2)E\sigma^4}{(n-p)(\nu-4)}\Delta^{-1}Cov(\beta', \sigma^2)Q^{-1}HE\sigma^2, \tag{A12}$$

and

$$\begin{aligned} R_{22} = &-\frac{2(n-p+\nu-2)E\sigma^4}{(n-p)(\nu-4)}\Delta^{-1}Cov\left(\beta', \sigma^2\right)Q^{-1}Cov\left(\beta, \sigma^2\right) \\ &+ \frac{2(n-p+\nu-2)E\sigma^4}{(n-p)(\nu-4)}\Delta^{-1}Var(\sigma^2).\end{aligned} \tag{A13}$$

From (A6) and (A9), we get

$$\begin{aligned} &\mathrm{MSEM}\left(\hat{\theta}_{ML}\right) - \mathrm{MSEM}\left(\hat{\theta}_{LBE}\right) \\ &= \begin{pmatrix} HE\sigma^2 - R_{11} & -R_{12} \\ -R_{21} & \frac{\left[2\nu^2(n-p)(n-p+\nu-2)+(2n-\nu p)^2(\nu-4)\right]E\sigma^4}{(\nu-2)^2 n^2(\nu-4)} - R_{22} \end{pmatrix} \hat{=} K.\end{aligned} \tag{A14}$$

To prove the MSEM superiority of the LBE, we only need to show that the matrix

$$K \geq 0. \tag{A15}$$

Let $s_0 = \frac{\left[2\nu^2(n-p)(n-p+\nu-2)+(2n-\nu p)^2(\nu-4)\right]E\sigma^4}{(\nu-2)^2 n^2(\nu-4)} - R_{22}$, we first have

$$\begin{aligned}
s_0 &= \frac{\left[2\nu^2(n-p)(n-p+\nu-2)+(2n-\nu p)^2(\nu-4)\right]E\sigma^4}{(\nu-2)^2 n^2(\nu-4)} \\
&\quad + \frac{2(n-p+\nu-2)E\sigma^4}{(n-p)(\nu-4)}\Delta^{-1}Cov(\beta',\sigma^2)Q^{-1}Cov(\beta,\sigma^2) \\
&\quad - \frac{2(n-p+\nu-2)E\sigma^4}{(n-p)(\nu-4)}\Delta^{-1}Var(\sigma^2) \\
&= \frac{\left[2\nu^2(n-p)(n-p+\nu-2)+(2n-\nu p)^2(\nu-4)\right]E\sigma^4}{(\nu-2)^2 n^2(\nu-4)} \\
&\quad + \frac{2(n-p+\nu-2)E\sigma^4}{(n-p)(\nu-4)}\Delta^{-1}\left[Cov(\beta',\sigma^2)Q^{-1}Cov(\beta,\sigma^2) - Var(\sigma^2)\right] \\
&\geq \frac{\left[2\nu^2(n-p)(n-p+\nu-2)+(2n-\nu p)^2(\nu-4)\right]E\sigma^4}{(\nu-2)^2 n^2(\nu-4)} - \frac{2(n-p+\nu-2)E\sigma^4}{(n-p)(\nu-4)} \\
&= \frac{E\sigma^4}{\nu-4}\frac{2(n-p+\nu-2)\left[\nu^2(n-p)^2-(\nu-2)^2n^2\right]+(2n-\nu p)^2(\nu-4)(n-p)}{(\nu-2)^2n^2(n-p)} \\
&= \frac{E\sigma^4}{\nu-4}\frac{c_0}{(\nu-2)^2n^2(n-p)},
\end{aligned} \tag{A16}$$

where $c_0 = \left(n - \frac{\nu p}{2}\right)[12(\nu-2)n^2 + (24p - 24\nu - 8\nu p - 2\nu^2 p + 8\nu^2 + 16)n + 2\nu^2p^2 - 4\nu^2 p - 4\nu p^2 + 8\nu p]$ and the third inequality takes advantage of the fact that $\Delta^{-1}[Cov(\beta',\sigma^2)Q^{-1}Cov(\beta,\sigma^2) - Var(\sigma^2)] > -1$. In order to prove $s_0 > 0$, it is enough to show that $c_0 > 0$. Note that the equation $c_0 = 0$ has the following three different roots

$$x_1 = \frac{\nu p}{2}, \tag{A17}$$

$$\begin{aligned}
x_2 &= \frac{p}{2} - \frac{\nu}{3} + \frac{\sqrt{\nu^2p^2 - 8\nu^2 p + 16\nu^2 - 12\nu p^2 + 8\nu p - 32\nu + 36p^2 + 48p + 16}}{12} \\
&\quad + \frac{\nu p}{12} + \frac{1}{3}
\end{aligned} \tag{A18}$$

and

$$\begin{aligned}
x_3 &= \frac{p}{2} - \frac{\nu}{3} - \frac{\sqrt{\nu^2p^2 - 8\nu^2 p + 16\nu^2 - 12\nu p^2 + 8\nu p - 32\nu + 36p^2 + 48p + 16}}{12} \\
&\quad + \frac{\nu p}{12} + \frac{1}{3}.
\end{aligned} \tag{A19}$$

Differentiating c_0 with respect to n, we obtain a polynomial about n

$$\frac{dc_0}{dn} = l_1 n^2 + l_2 n + l_3, \tag{A20}$$

where

$$l_1 = 36(\nu - 2), \tag{A21}$$

$$l_2 = 8(6p - 6\nu + \nu p - 2\nu^2 p + 2\nu^2 + 4), \tag{A22}$$

$$l_3 = \nu^3 p^2 - 4\nu^3 p + 6\nu^2 p^2 + 8\nu^2 p - 16\nu p^2. \tag{A23}$$

Note that $l_1 > 0$ for $\nu > 4$ and c_0 strictly increases when n is larger than the largest root. Using $x_1 > \frac{x_2+x_3}{2}$ and the relationship among the three roots, we have $x_1 > \max\{x_2, x_3\}$. Therefore, we conclude that $s_0 > 0$ when $n > \frac{\nu p}{2}$ and $\nu > 4$.

Note that

$$\begin{aligned} &\begin{pmatrix} I_p & s_0^{-1}R_{12} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} HE\sigma^2 - R_{11} & -R_{12} \\ -R_{21} & s_0 \end{pmatrix} \begin{pmatrix} I_p & 0 \\ s_0^{-1}R_{21} & 1 \end{pmatrix} \\ &= \begin{pmatrix} HE\sigma^2 - R_{11} - s_0^{-1}R_{12}R_{21} & 0 \\ 0 & s_0 \end{pmatrix}. \end{aligned} \tag{A24}$$

To prove the matrix $K \geq 0$, we still need to prove the following fact

$$\begin{aligned} &HE\sigma^2 - R_{11} - R_{12}\left[\frac{\left(2\nu^2(n-p)(n-p+\nu-2) + (2n-\nu p)^2(\nu-4)\right)E\sigma^4}{(\nu-2)^2 n^2(\nu-4)} - R_{22}\right]^{-1} R_{21} \\ &= HE\sigma^2 Q^{-1}\left[Q + Cov(\beta, \sigma^2)\left(\frac{1}{\Delta} - \frac{4(n-p+\nu-2)^2(E\sigma^4)^2}{\Delta^2 s_0(n-p)^2(\nu-4)^2}\right)Cov(\beta', \sigma^2)\right]Q^{-1}HE\sigma^2 \geq 0. \end{aligned} \tag{A25}$$

If

$$\frac{1}{\Delta} - \frac{4(n-p+\nu-2)^2(E\sigma^4)^2}{\Delta^2 s_0(n-p)^2(\nu-4)^2} \geq 0,$$

the inequality (A25) obviously holds. Conversely, if

$$\frac{1}{\Delta} - \frac{4(n-p+\nu-2)^2(E\sigma^4)^2}{\Delta^2 s_0(n-p)^2(\nu-4)^2} < 0,$$

let $a = Cov(\beta', \sigma^2)Q^{-1}Cov(\beta, \sigma^2)$, then

$$\begin{aligned} &\frac{1}{a} + \frac{1}{\Delta} - \frac{4(n-p+\nu-2)^2(E\sigma^4)^2}{\Delta^2 s_0(n-p)^2(\nu-4)^2} \\ &= \frac{(a+\Delta)\Delta s_0(n-p)^2(\nu-4)^2 - 4a(n-p+\nu-2)^2(E\sigma^4)^2}{a\Delta^2 s_0(n-p)^2(\nu-4)^2} \hat{=} \frac{F_1}{G}, \end{aligned} \tag{A26}$$

where we have

$$\begin{aligned} F_1 &= (a+\Delta)\Delta(n-p)^2(\nu-4)^2\left(\frac{\left[2\nu^2(n-p)(n-p+\nu-2) + (2n-\nu p)^2(\nu-4)\right]E\sigma^4}{(\nu-2)^2 n^2(\nu-4)}\right) \\ &\quad + (a+\Delta)\Delta(n-p)^2(\nu-4)^2\left(\frac{2(n-p+\nu-2)E\sigma^4}{(n-p)(\nu-4)}\Delta^{-1}a\right) \\ &\quad - (a+\Delta)\Delta(n-p)^2(\nu-4)^2\left(\frac{2(n-p+\nu-2)E\sigma^4}{(n-p)(\nu-4)}\Delta^{-1}Var(\sigma^2)\right) \\ &\quad - 4a(n-p+\nu-2)^2(E\sigma^4)^2 \end{aligned}$$

$$= (a+\Delta)\left(\frac{\left[2\nu^2(n-p)(n-p+\nu-2)+(2n-\nu p)^2(\nu-4)\right]\Delta(n-p)^2(\nu-4)^2E\sigma^4}{(\nu-2)^2n^2(\nu-4)}\right)$$
$$+ (a+\Delta)\left(2(n-p+\nu-2)(n-p)(\nu-4)aE\sigma^4\right)$$
$$- (a+\Delta)\left(2(n-p+\nu-2)(n-p)(\nu-4)Var(\sigma^2)E\sigma^4\right)$$
$$- 4a(n-p+\nu-2)^2(E\sigma^4)^2$$
$$= (a+\Delta)\left(\frac{\left[2\nu^2(n-p)(n-p+\nu-2)+(2n-\nu p)^2(\nu-4)\right]\Delta(n-p)^2(\nu-4)^2E\sigma^4}{(\nu-2)^2n^2(\nu-4)}\right)$$
$$+ (a+\Delta)\left[2(n-p+\nu-2)(n-p)(\nu-4)\left(\frac{2(n-p+\nu-2)E\sigma^4}{(n-p)(\nu-4)}-\Delta\right)E\sigma^4\right]$$
$$- 4a(n-p+\nu-2)^2(E\sigma^4)^2$$
$$= (a+\Delta)\left(\frac{\left[2\nu^2(n-p)(n-p+\nu-2)+(2n-\nu p)^2(\nu-4)\right]\Delta(n-p)^2(\nu-4)^2E\sigma^4}{(\nu-2)^2n^2(\nu-4)}\right)$$
$$+ (a+\Delta)\left[4(n-p+\nu-2)^2(E\sigma^4)^2\right] - (a+\Delta)\left[2(n-p+\nu-2)(n-p)(\nu-4)\Delta E\sigma^4\right]$$
$$- 4a(n-p+\nu-2)^2(E\sigma^4)^2$$
$$= \left[(\nu-2)^2n^2(\nu-4)\right]^{-1}(a+\Delta)2\nu^2(n-p)(n-p+\nu-2)\Delta(n-p)^2(\nu-4)^2E\sigma^4$$
$$+ \left[(\nu-2)^2n^2(\nu-4)\right]^{-1}(a+\Delta)(2n-\nu p)^2(\nu-4)\Delta(n-p)^2(\nu-4)^2E\sigma^4$$
$$- \left[(\nu-2)^2n^2(\nu-4)\right]^{-1}(a+\Delta)2(n-p+\nu-2)(n-p)(\nu-4)^2n^2(\nu-2)^2\Delta E\sigma^4$$
$$+ \left[(\nu-2)^2n^2(\nu-4)\right]^{-1}4\Delta(n-p+\nu-2)^2(\nu-2)^2n^2(\nu-4)(E\sigma^4)^2.$$

After further simplification, we come to

$$\begin{aligned}F_1 &= \Delta\left[(\nu-2)^2n^2\right]^{-1}\left[E\sigma^4-(E\sigma^2)^2+\frac{2(n-p+\nu-2)E\sigma^4}{(n-p)(\nu-4)}\right]\\&\times 2\nu^2(n-p)(n-p+\nu-2)(n-p)^2(\nu-4)E\sigma^4\\&+\Delta\left[(\nu-2)^2n^2\right]^{-1}\left[E\sigma^4-(E\sigma^2)^2+\frac{2(n-p+\nu-2)E\sigma^4}{(n-p)(\nu-4)}\right]\\&\times (2n-\nu p)^2(\nu-4)(n-p)^2(\nu-4)E\sigma^4\\&-\Delta\left[(\nu-2)^2n^2\right]^{-1}\left[E\sigma^4-(E\sigma^2)^2+\frac{2(n-p+\nu-2)E\sigma^4}{(n-p)(\nu-4)}\right]\\&\times 2(n-p+\nu-2)(n-p)(\nu-4)n^2(\nu-2)^2E\sigma^4\\&+\Delta\left[(\nu-2)^2n^2\right]^{-1}4(n-p+\nu-2)^2(\nu-2)^2n^2(E\sigma^4)^2.\end{aligned} \tag{A27}$$

Let $F_2 = \left[2\nu^2(n-p)(n-p+\nu-2)+(2n-\nu p)^2(\nu-4)\right](n-p)^2(\nu-4) - 2(n-p+\nu-2)(n-p)(\nu-4)n^2(\nu-2)^2$ and $F_3 = 4(n-p+\nu-2)^2(\nu-2)^2n^2$, then (A27) can be represented as

$$F_1 = \Delta\left[(\nu-2)^2n^2\right]^{-1}\left[F_4(E\sigma^4)^2 - F_2E\sigma^4(E\sigma^2)^2\right], \tag{A28}$$

where

$$
\begin{aligned}
F_4 &= \frac{2(n-p+\nu-2)+(n-p)(\nu-4)}{(n-p)(\nu-4)}F_2+F_3 \\
&= \left[2(n-p+\nu-2)+(n-p)(\nu-4)\right] \\
&\quad\times\left\{\left[2\nu^2(n-p)(n-p+\nu-2)+(2n-\nu p)^2(\nu-4)\right](n-p)-2(n-p+\nu-2)n^2(\nu-2)^2\right\} \\
&\quad+4(n-p+\nu-2)^2(\nu-2)^2n^2 \\
&= \left[2(n-p+\nu-2)+(n-p)(\nu-4)\right] \\
&\quad\times\left[2\nu^2(n-p)(n-p+\nu-2)+(2n-\nu p)^2(\nu-4)\right](n-p) \\
&\quad-2(n-p+\nu-2)n^2(\nu-2)^2(n-p)(\nu-4) \\
&= (n-p)\left\{(n-p+\nu-2)[J_{11}+J_{22}+J_{33}-J_{44}]+(2n-\nu p)^2(\nu-4)^2\right\},
\end{aligned} \tag{A29}
$$

with $J_{11}=4\nu^2(n-p+\nu-2)(n-p)$, $J_{22}=2\nu^2(n-p)^2(\nu-4)$, $J_{33}=2(2n-\nu p)^2(\nu-4)$ and $J_{44}=2n^2(\nu-2)^2(\nu-4)$. Let $F_5=J_{22}-J_{44}$. Then for $n>\frac{\nu p}{2}$ and $\nu>4$, obviously

$$
F_5=\left(n-\frac{\nu p}{2}\right)\left[8(\nu-4)(\nu-1)n-49\nu^2+16p\nu\right]>0. \tag{A30}
$$

Hence, we obtain $F_4>0$. Using (A28) and the fact that $E\sigma^4\ge(E\sigma^2)^2$, we obtain

$$
\begin{aligned}
F_1 &\ge \Delta\left[(\nu-2)^2n^2\right]^{-1}\left[\left(\frac{2(n-p+\nu-2)+(n-p)(\nu-4)}{(n-p)(\nu-4)}-1\right)F_2+F_3\right]E\sigma^4(E\sigma^2)^2 \\
&= 2\Delta\left[(\nu-2)^2n^2\right]^{-1}[J_{55}+J_{66}+J_{77}]E\sigma^4(E\sigma^2)^2,
\end{aligned} \tag{A31}
$$

where $J_{55}=2\nu^2(n-p+\nu-2)(n-p)^2$, $J_{66}=(2n-\nu p)^2(\nu-4)(n-p)$ and $J_{77}=2(n-p+\nu-2)(\nu-2)^2n^2$.

Then by (A26), we have

$$
\begin{aligned}
&\frac{1}{a}+\frac{1}{\Delta}-\frac{4(n-p+\nu-2)^2(E\sigma^4)^2}{\Delta^2 s_0(n-p)^2(\nu-4)^2}=\frac{F_1}{G} \\
\ge\ & 2\left[a\Delta s_0(n-p)^2(\nu-4)^2(\nu-2)^2n^2\right]^{-1} \\
\times\ & \left[2\nu^2(n-p+\nu-2)(n-p)^2+(2n-\nu p)^2(\nu-4)(n-p)+2(n-p+\nu-2)(\nu-2)^2n^2\right] \\
\times\ & E\sigma^4(E\sigma^2)^2>0.
\end{aligned}
$$

Thus, we come to

$$
a+\left(\frac{1}{\Delta}-\frac{4(n-p+\nu-2)^2(E\sigma^4)^2}{\Delta^2 s_0(n-p)^2(\nu-4)^2}\right)^{-1}<0. \tag{A32}
$$

Then, we have

$$
\begin{aligned}
&\left(Q+Cov(\beta,\sigma^2)\left(\frac{1}{\Delta}-\frac{4(n-p+\nu-2)^2(E\sigma^4)^2}{\Delta^2 s_0(n-p)^2(\nu-4)^2}\right)Cov(\beta',\sigma^2)\right)^{-1} \\
&= Q^{-1}-Q^{-1}Cov(\beta,\sigma^2)\left[a+\left(\frac{1}{\Delta}-\frac{4(n-p+\nu-2)^2(E\sigma^4)^2}{\Delta^2 s_0(n-p)^2(\nu-4)^2}\right)^{-1}\right]^{-1}Cov(\beta',\sigma^2)Q^{-1}>0.
\end{aligned}
$$

Hence, in the case that $\frac{1}{\Delta}-\frac{4(n-p+\nu-2)^2(E\sigma^4)^2}{\Delta^2 s_0(n-p)^2(\nu-4)^2}<0$, we still conclude that the inequality (A25) holds. Theorem 3.2 has been proved.