⊠연구논문

On a Bayesian Estimation of Multivariate Regression Models with Constrained Coefficient Matrix

Hea-Jung Kim
Dept. of Statistics, Dongguk University

Abstract

Consider the linear multivariate regression model $Y = X_1B_1 + X_2B_2 + U$, where $\mathrm{Vec}(U) \sim N(0, \Sigma \otimes I_N)$. This paper is concerned with Bayes inference of the model when it is suspected that the elements of B_2 are constrained in the form of intervals. The use of the Gibbs sampler as a method for calculating Bayesian marginal posterior densities of the parameters under a generalized conjugate prior is developed. It is shown that the approach is straightforward to specify distributionally and to implement computationally, with output readily adopted for required inference summaries. The method developed is applied to a real problem.

1. Introduction

Let $y_1, \ldots, y_p$ be $N\times 1$ vectors representing N independent observations on each of p correlated dependent random variables and let X be a functional or a conditional regressor matrix of q independent variables. Suppose we want to investigate the linear relationship between the dependent and the independent variables using a linear regression model

$$\begin{aligned} Y &= X_1 B_1 + X_2 B_2 + U, \\ &= XB + U, \end{aligned} \tag{1.1}$$

where $Y = [y_1, \ldots, y_p]$, $X = [X_1 : X_2]$ is a $N \times q$ regressor matrix on q independent variables, $B = [B_1' : B_2']'$ is a $q \times p$ matrix of unknown regression parameters with B_ℓ a $q_\ell \times p$ matrix ($\ell = 1,2$), $q = q_1 + q_2$, and U is a $N \times p$ matrix of error vectors such that the rows of U, say $v_1, v_2, \ldots, v_N$ are independent normal random vectors with mean vector zero and unknown $p \times p$ covariance matrix Σ. The multivariate linear regression model is completely defined if we impose the condition, $p + q \le N$, on the quantities in (1.1). This assumption is used so that there will be a sufficient number of observations available to estimate all of the parameters.

Some of the early uses of such model are due to Barten(1964) in industrial applications, Box and Tiao(1972) in quality control and Zellner(1971) in the evaluation of annual investment. The model will be estimated by the methods of least squares and maximum likelihood. For a general discussion of the methods and their properties the reader is referred to the books of Press(1982, Chapter 8) and Johnson and Wichern(1992, Chapter 7). A completely different approach is Bayesian estimation method initiated by Tiao and Zellner(1964). Box and Tiao (1973, Chapter 8) and Press(1989, Chapter 5) have given comprehensive reviews of Bayesian inference of the multivariate regression model. For a constrained parameter problem, Ghosh et. al.(1989) and Merwe and Merwe(1992) considered empirical and hierarchical Bayesian estimation of B_1 under the constraint that $B_2 = 0$, respectively, but Bayesian estimation of B_1 under other type of constraints on B_2 has not been seen in the literatures yet.

Constrained parameter problem may arise in a wide variety of applications of the multivariate regression model. The parametric Bayes perspective is attractive for examining such models. For example consider interval constrained regression coefficients model, which in a classical setting might employ isotonic regression of maximum likelihood estimates to obtain point estimates. A more satisfying analysis would develop and compare posterior distributions arising from priors that reflect the interval restrictions. However, one problem of the Bayesian estimation is that analytic approaches (exact or approximate) for carrying out required multi-dimensional integrations in this case will be well-nigh impossible. Another problem involved in the Bayesian analysis is that noted by Rothenberg(1963).

Rothenberg noted that if we use a natural conjugate prior or a vague prior for the model (1.1), this will involve placing restrictions on the parameters, namely, the variances and covariances of coefficients appearing in the equations of the system (1.1). This is due to the fact that the matrix $(X'X)^{-1}$ enters the covariance structure in the following way, $\Sigma \otimes (X'X)^{-1}$. Thus, for example, the ratios of variances of corresponding coefficients in the first and second equations will all be equal if we use a natural conjugate prior or a vague prior(see Section 2 for more restrictions). The restrictions involved in the Bayes estimators are not reasonable in most situations.

The purpose of this paper is to eliminate the two problems (say, the former and the latter problem) in the Bayesian estimation of B in (1.1) when it is suspected that each element of B_2 is constrained to an interval. For resolving the latter problem, we develop a Bayesian analysis of the model which avoids placing the restrictions on the variances and covariances of the posterior distribution of B. This is done by use of a generalized natural conjugate prior density (so called by Press) where B and Σ are independent and B_1 has normal density, B_2 a truncated normal and $\Lambda = \Sigma^{-1}$ a Wishart. It is shown that the joint and marginal posterior densities of B_1, B_2 and Λ are not standard, and hence the former problem naturally arises in the estimation of B. We resolve it by suggesting an estimation scheme via the Gibbs sampler. It is shown that the Bayesian calculations, involved in the former problem, can be implemented routinely for interval constrained parameters by means of the Gibbs sampler.

2. Constrained Multivariate Regression Models

As stated before, in the standard multivariate models, the $N \times p$ response matrix Y is related to $N \times q$ regressor matrix X and $q \times p$ coefficient matrix B by the equation

$$Y = XB + U, \ N \geq p + q. \tag{2.1}$$

Let $U' = (v_1, \ldots, v_N)$, so that $U'U = \sum_{i=1}^{N} v_i v_i'$. Assume $v_i \overset{ind}{\sim} N_p(0, \Sigma)$, and each element of B_2 is constrained to an interval where $B = (B_1', B_2')'$. Under the assumptions, the rows of the Y matrix are independent normal random vectors, and hence the likelihood of parameters is

$$\begin{aligned} L(B,\Lambda|Y,X) &\propto |\Lambda|^{N/2}\exp\left\{-\frac{1}{2}\operatorname{tr}(Y-XB)'(Y-XB)\Lambda\right\} I(Q), \\ &= |\Lambda|^{N/2}\exp\left\{-\frac{1}{2}\operatorname{tr}\Lambda V\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}(\beta-\hat{\beta})'\Lambda\otimes X'X(\beta-\hat{\beta})\right\} I(Q), \end{aligned} \tag{2.2}$$

where B is a $q\times p$ real matrix, $\beta=\mathrm{Vec}(B)$, i.e. $\beta=(\beta_1,\dots,\beta_{pq})'$, $I(Q)$ denotes the indicator function for the set Q

$$Q = \{a_j\le\beta_j\le c_j\ ;\ j=kq-q_2+1,\dots,kq\ ,\ \ k=1,\dots,p\},$$

$\Lambda=\Sigma^{-1}$ is an unknown $p\times p$ positive definite symmetric matrix, $V=(Y-X\hat{B})'(Y-X\hat{B})$ is the residual matrix, $\hat{B}=(X'X)^{-1}X'Y$ is the maximum likelihood estimator (also is the least squares estimator) of B, and $\hat{\beta}=\mathrm{Vec}(\hat{B})$. Here $\mathrm{Vec}(S)$ is the vec operator denoting the vector formed by stacking columns of S, one underneath the other.

When B_2 is not constrained, sampling distribution of the maximum likelihood and the least square estimators of B is

$$\mathrm{Vec}(\hat{B}) \sim N(\mathrm{Vec}(B),\Sigma\otimes(X'X)^{-1}).$$

From the distribution, one can easily see that the variance-covariance matrix of $\hat{B}$ places restrictions on the variances and covariances of $\hat{B}$. Specifically, if $\hat{\beta}_{ij}$ denotes ijth element of $\hat{B}$, the variance-covariance matrix of $\mathrm{Vec}(\hat{B})$, $\Sigma\otimes(X'X)^{-1}$, yields the restrictions

$$\frac{Var(\hat{\beta}_{ij})}{Var(\hat{\beta}_{ij'})} = \frac{Var(\hat{\beta}_{i'j})}{Var(\hat{\beta}_{i'j'})} = \frac{Cov(\hat{\beta}_{ij},\hat{\beta}_{i'j})}{Cov(\hat{\beta}_{ij'},\hat{\beta}_{i'j'})} = \frac{\sigma_{ii}}{\sigma_{i'i'}},$$

for $i\neq i'$, $i,i'=1,\dots,p$, $j\neq j'$, $j,j'=1,\dots q$, where $\Sigma=\{\sigma_{ij}\}$.

Broemeling(1985) showed that the same restrictions are also placed in the Bayesian estimate of the variance-covariance matrix of B when one uses usual priors(a natural conjugate prior and the Jeffrey's vague prior).

To eliminate the above problem and to take the constraints of B_2 into account, we proceed with the analysis by putting a generalized natural conjugate prior densities (cf. Press 1982) on B and Λ, and then applying Bayes' theorem.

Apply the procedure, described in Press(1982), to (2.2) to find a generalized natural conjugate prior for β and Λ. Assuming Λ is known and interchanging the role of (Y, X) and β gives, after enrichment and normalization,

$$\beta \sim N_{I(Q)}(\phi, F), \quad F > 0, \tag{2.3}$$

where $\beta = (\beta_1, \ldots, \beta_{pq})'$, ϕ and F are arbitrary, $N_{I(Q)}(\cdot, \cdot)$ denotes the multivariate normal distribution with truncated intervals

$$Q = \{ a_j \leq \beta_j \leq c_j \ ; \ j = kq - q_2 + 1, \ldots, kq, \ \ k = 1, \ldots, p \}.$$

This defines the constrained intervals of all the elements in B_2. Similar procedure gives the prior

$$\Lambda \sim W(G, p, m), \quad m > p - 1, \tag{2.4}$$

where $W(G, p, m)$ denote a p-dimensional Wishart distribution with scale parameter G and m degrees of freedom.

Combining (2.3) and (2.4), we get the joint generalized natural conjugate prior, a truncated normal-Wishart prior. This prior density does not have the restriction problem on B associated with ordinary natural conjugate prior and the Jeffrey's vague prior for the problem.

The joint posterior density for β and Λ is found by multiplying the truncated normal-Wishart prior and likelihood function (2.2) (note $X'X$ can be singular),

$$p(\beta, \Lambda | Y, X) \propto |\Lambda|^{(N+m-p-1)/2} \exp\Big\{ -\frac{1}{2} \operatorname{tr}[\Lambda(G^{-1} + V + (B - \hat{B})'X'X(B - \hat{B})] \\ - \frac{1}{2}(\beta - \phi)'F^{-1}(\beta - \phi) \Big\} \, I(Q) \tag{2.5}$$

This is again seen to be a truncated normal-Wishart density.

The marginal posterior density of β can be found by integrating (2.5) with respect to Λ, using the Wishart normalizing constant, to get

$$p(\beta \mid Y, X) \propto \frac{\exp\left\{-\frac{1}{2}(\beta-\phi)' F^{-1}(\beta-\phi)\right\} I(Q)}{|G^{-1}+V+(B-\hat{B})'X'X(B-\hat{B})|^{(N+m)/2}}. \tag{2.6}$$

The posterior density for β is seen to be the product of a factor in the generalized multivariate Student t form and a factor in the multivariate normal form. Therefore, the exact posterior density of β is rather complicated as it stands. Moreover, exact marginal posterior density of Λ is analytically intractable due to the truncated normal distribution of B_2. On the other hand, we can see, from (2.5), that full conditionals reduce analytically to closed-form distributions. Highly efficient sampling routines are available for these distributions, see for example Spiegelhalter et al.(1996).

3. The Gibbs Sampler

Although exact marginal posterior distributions for the parameters are not available, the conjugate structure of the joint prior specification allows for painless calculation of the Gibbs conditionals(cf. Gilks 1996). Thus the Gibbs sampler may be used to explore the posterior distributions without having established property of the marginal posteriors.

A value for each constrained coefficient β_i is drawn in turn from its distribution conditional on $\beta_\ell(\ell \neq i)$ and Λ, and a value for Λ is drawn conditional on β. In the algorithm the Gibbs sampler moves from any point in the support of β and Λ to any nondegenerate neighborhood of any other point in the support with positive probability in one step. Convergence of the continuous state Markov chain induced by the Gibbs sampler to the posterior distribution may therefore be demonstrated following the argument of Tierney(1994).

Theorem 1. The full conditional posterior distributions of β and Λ are

$$\beta \mid \Lambda \sim N_{I(Q)}(\beta_0, M^{-1}), \tag{3.1}$$

$$\Lambda \mid \beta \sim W((G^{-1} + V + (B - \widehat{B})' X' X (B - \widehat{B}))^{-1},\ p,\ N + m), \tag{3.2}$$

where $\beta = (\beta_1, \ldots, \beta_{pq})'$,

$\beta_0 = (\beta_{01}, \ldots, \beta_{0pq})' = (F^{-1} + \Lambda \otimes X'X)^{-1}(F^{-1}\phi + [\Lambda \otimes X'X]\widehat{\beta})$,

$M^{-1} = (F^{-1} + \Lambda \otimes X'X)^{-1}$.

Here $N_{I(Q)}(\cdot, \cdot)$ denotes a multivariate normal distribution truncated at an interval $a_i \le \beta_i \le b_i$ for $i = kq - q_2 + 1, \ldots, kq$, and $k = 1, \ldots, p$. Otherwise, $a_i = -\infty$ and $b_i = \infty$.

Proof. From (2.5), we see that, given Λ, the density of β is

$$p(\beta \mid \Lambda, Y, X) \propto \exp\left\{-\frac{1}{2}(\beta - \widehat{\beta})' \Lambda \otimes X'X(\beta - \widehat{\beta}) - \frac{1}{2}(\beta - \phi)' F^{-1}(\beta - \phi)\right\} I(Q). \tag{3.3}$$

Completing the square on β gives

$$p(\beta \mid \Lambda, Y, X) \propto \exp\left\{-\frac{1}{2}(\beta - \beta_0)' M (\beta - \beta_0)\right\} I(Q). \tag{3.4}$$

Using the definition in Gilks(1996) of the full conditional posterior distribution, the conditional of β given Λ is obtained from (3.4). Conditional on β, (2.5) gives the distribution of Λ.

Noticing that it is computationally inefficient to invert the $pq \times pq$ matrix $M = (F^{-1} + \Lambda \otimes X'X)$ in (3.4), instead we may use following result.

Result 1. Let L be a factor of F, $LL' = F$. Let $L'(\Lambda \otimes X'X)L$ have diagonalization PDP', i.e. D is a diagonal matrix of eigenvalues of $L'(\Lambda \otimes X'X)L$, and the columns of P are the corresponding, ordered eigenvectors normalized so that $P'P = PP' = I_{pq}$. Then

$$(F^{-1} + \Lambda \otimes X'X)^{-1} = H(I_{pq} + D)^{-1} H', \tag{3,5}$$

where $H = LP$.

Proof. By making use of properties of the symmetric matrices, F and $\Lambda \otimes X'X$, and matrix algebra, we have the result.

The Gibbs sampling algorithm based upon the above conditionals is attractive because the Gibbs sampler is so straightforward in this situation. Simple matrix calculations and the ability to simulate normal and truncated normal variables and Wishart random matrix are all that required. See Smith and Hocking(1972) for the algorithm for generating Wishart random matrix. The computations in Result 1 are only performed once in each pass of the above Gibbs sampling algorithm.

An algorithm for generating from the truncated univariate normal distribution, described in Devroye(1986), can be employed to generate β from the truncated multivariate normal variate in (3.1). This can be done by following process. Given the full conditional distribution of β as $N_{I(Q)}(\beta_0, \Omega)$, $\Omega = M^{-1}$, we may rearrange the elements of β to get $\beta^* = (\beta_1^{*\prime}, \beta_2^{*\prime})'$, where $\beta_1^* = \mathrm{Vec}(B_1)$ and $\beta_2^* = \mathrm{Vec}(B_2)$ that is the $pq_2 \times 1$ vector consisting of the constrained elements, i.e. $a_j \le \beta_j \le b_j$ for $j = 1, \ldots, pq_2$. Let denote β_0^* and Ω_* be mean vector and covariance matrix of β^* correspondingly arranged and define

$$\beta_0^* = \begin{pmatrix} \beta_{01}^* \\ \beta_{02}^* \end{pmatrix}, \quad \Omega^* = \begin{pmatrix} \omega'_{(1)} \\ \vdots \\ \omega'_{(pq)} \end{pmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}.$$

Then the full conditional distribution of β^* given Λ is $N_{I(Q)}(\beta_0^{*}, \Omega^*)$ and the following results hold.

Result 2. The full conditional distribution of β^* can be decomposed by

$$\beta_1^* \mid \beta_2^*, \Lambda \sim N(\beta_{01}^* + \Omega_{12}\Omega_{22}^{-1}(\beta_2^* - \beta_{02}^*)\ ,\ \Omega_{11.2}), \tag{3.6}$$

$$\beta_2^*(j) \mid \beta_1^*, \beta_2^*(\ell), \ell \neq j, \Lambda \sim N_{I(Q_j)}(\beta_{02}^*(j) + \omega_{(12)}(j)'\Omega_{22}(j)^{-1} [\beta^*(j) - \beta_0^*(j)], \Omega_{11.2}(j)),\ j = 1, \ldots, pq_2, \tag{3.7}$$

where $\Omega_{11.2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$, $\beta_2^* = (\beta_2^*(1), \ldots, \beta_2^*(pq_2))'$, $Q_j = \{a_j \le \beta_j \le b_j\}$, $\beta_{02}^* = (\beta_{02}^*(1), \ldots, \beta_{02}^*(pq_2))'$, $\omega_{(12)}(j)$ is a $(pq-1)\times 1$ vector obtained by deleting $(pq_1 + j)$ th element from $\omega_{(pq_1+j)}$, $\Omega_{22(j)}$ denotes the $(pq-1)\times(pq-1)$ matrix derived from Ω^* by crossing out the row and column containing $(pq_1 + j, pq_1 + j)$ th element, say ω_{pq_1+j, pq_1+j}, $\Omega_{11.2}(j) = \omega_{pq_1+j, pq_1+j} - \omega_{12}(j)' \Omega_{22}(j)\omega_{12}(j)$, and $\beta^*(j)$ and $\beta_0^*(j)$ denote $(pq-1)\times 1$ vectors obtained by deleting $(pq_1 + j)$ th elements from β^* and β_0^*, respectively. Here $N_{I(Q_j)}(\cdot\,,\cdot)$ denotes a univariate normal distribution truncated at an interval $a_j \le \beta_j \le b_j$, $j = 1, \ldots, pq_2$.

Proof. Applying the properties of the conditional normal distribution to the distribution of $\beta^* \mid \Lambda$, we have the result.

When one uses Result 2 for generating B(or equivalently β), drawings from (3.7) can be generated through the one-for-one sampling method(see, Devroye 1986) using a truncated normal density. Once we get a Gibbs sample of β^* from (3.6) and (3.7), then we reconstruct B from the value of β^* and proceed to generated Λ from the conditional distribution (3.3). This Gibbs sampling algorithm gives considerably faster sampler than that adopting a naive rejection method for the generation of B via (3.1) (see, Devroye 1986 for the naive methods).

4. An Illustrative Example

In order to investigate the performance of the suggested Gibbs sampler, Bayesian analysis was conducted to a chemical process data listed in <Table 1> and analyzed by Box and Tiao (1972, P. 454).

< Table 1 > Yields of Product and By-product of a Chemical Process

Temp. F	Product Y_1	By-product Y_2
161.3	63.7	20.3
164.0	59.5	24.2
165.7	67.9	18.0
170.1	68.8	20.5
173.9	66.1	20.1
176.2	70.4	17.5
177.6	70.0	18.2
181.7	73.7	15.4
185.6	74.1	17.8
189.0	79.6	13.3
193.4	77.1	16.7
195.7	82.8	14.8

Our interest is to analyze the chemical process with two dependent variables, yield of product and yield of by-product, and one independent variable, temperature, thus the model is

$$Y_{i1} = \theta_{11}X_{i1} + \theta_{21}X_{i2} + \varepsilon_{i1},$$
$$Y_{i2} = \theta_{12}X_{i1} + \theta_{22}X_{i2} + \varepsilon_{i2}, \quad i=1,\dots,12,$$

where $X_{i1}=1$ and $X_{i2}=(T_i-\overline{T})/100$ is the ith setting of transposed temperature corresponding to product yield Y_{i1} and by-product yield Y_{i2}, for $i=1,2,\dots,N$, $N=12$. The regression matrix is

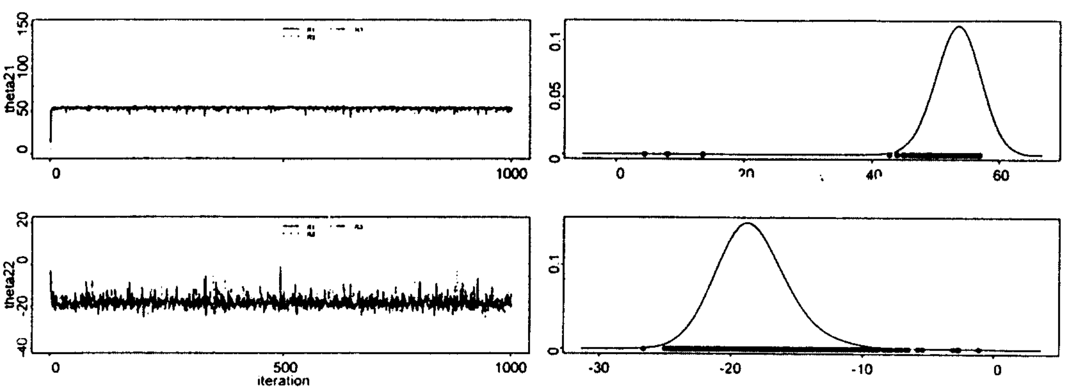
$$B = (B_1', B_2')' = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix},$$

where $B_2=(\theta_{21}, \theta_{22})$ having constraints $\theta_{21}\geq 0$ and $\theta_{22}\leq 0$. The N independent bivariate error vector $\varepsilon_i'=(\varepsilon_{i1}, \varepsilon_{i2})$ are normal with mean zero and unknown 2×2 precision matrix Λ. We used the generalized natural conjugate prior $\beta=\mathrm{Vec}(B) \sim N_{I(Q)}(0, F)$, where $F=10I_4+e_{(2)}e_{(1)'}+2e_{(3)}e_{(1)}'+e_{(3)}e_{(2)}'$

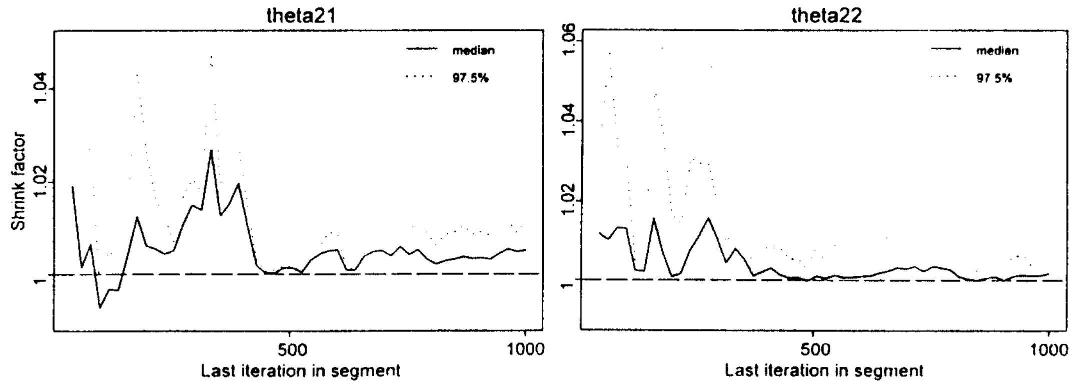
$+3e_{(4)}e_{(1)}' + 2e_{(4)}e_{(2)}' + e_{(4)}e_{(3)}'$ with $e_{(k)}$ being the kth column of I_4, $Q=\{\theta_{21}\geq 0, \theta_{22}\leq 0\}$, and $\Lambda \sim W(G,2,2)$, with $G=\{g_{ij}\}$ with $g_{ii}=.01$ and $g_{ij}=0$, $i,j=1,2$, reflecting lack of prior information about β and Λ. The truncated intervals in Q are obvious upon investigating the data set.

To diagnose the convergence of the Gibbs sampler, as advocated by Cowles and Carlin(1996), we ran three parallel chains with starting points drawn from what we believe is a distribution overdispersed with respect to the stationary distribution. Then we visually inspected convergence of these chains by overlapping their sampled values on common graphs(so called traces) for θ_{21} and θ_{22}, and estimating kernel densities of their sampled values(See, Figure 1). <Figure 2> annotates each graph with the Gelman and Rubin(1992) statistic.

As seen in <Figure 1> and <Figure 2>, the convergence diagnostic shows that theGibbs sampler mixes fast and converges to a stationary distribution within 1000 iterations. Those figures for the chains of θ_{11} and θ_{12}(not presented here) showed the same convergence.



< Figure 1 > Traces and Kernel Densities of θ_{21} and θ_{22} (1000 values per trace and kernel density).



< Figure 2 > Gelman and Rubin Shrinkage Factors
(shrinkage factor approaching to 1 implies the convergence).

Taking the number of burn-in iterations M to discard as $M=1000$, Gibbs sample is collected from $M+1$ to iteration $M^*=M+1000$, the last iteration generated. If the chosen summary statistic is designated $\mu=T(\theta_{ij})$, then the estimate of its mean based on the retained iterates θ_{ij}^{ℓ}, $\ell=M+1,\ldots,M^*$ is

$$\hat{\mu} = \frac{1}{M^*-M-1}\sum_{\ell=M+1}^{M^*} T(\theta_{ij}^{\ell}).$$

To assess the performance of the Gibbs sampler, the resulting estimates of θ_{ij}'s are compared to the corresponding unconstrained ML estimates (since constrained ML has not been proposed yet). We also compared 95% interval for each θ_{ij} with that of the ML estimate. As expected, <Table 2> notes that the constrained estimation of θ_{ij}'s is more efficient than the unconstrained estimation in the sense that the former yields shorter 95% intervals and standard deviations of θ_{ij}'s than the latter does.

< Table 2 > Comparison of the Bayes Estimates and Unconstrained ML Estimates. (standard deviation of each estimate is noted in the parenthesis)

parameter	method	estimate	95% interval	
			lower	upper
θ_{11}	Bayes	71.134 (0.123)	70.932	71.425
	ML	69.738 (0.715)	69.738	72.581
θ_{12}	Bayes	18.070 (0.242)	17.614	18.627
	ML	18.066 (0.553)	16.981	19.711
θ_{21}	Bayes	53.452 (1.490)	49.918	55.536
	ML	54.435 (6.497)	41.699	67.171
θ_{22}	Bayes	-18.306 (2.410)	-22.241	12.714
	ML	-20.093 (5.024)	-29.940	10.245

5. Concluding Remarks

A constrained Bayesian estimation of the multivariate linear regression model is pursued in this paper. It is done by use of a generalized natural conjugate prior density where $\mathrm{Vec}(B)$ and Λ are independent and $\mathrm{Vec}(B)$ has a truncated normal density and Λ a Wishart. However, the joint and marginal posterior densities are not of standard distributions and one must use approximations to obtain standard forms, leading to small sample problem. Instead, we use the Gibbs sampler to estimate the model. It is seen that the approach is straightforward to specify distributionally and computationally, with output readily adopted for required inference summaries. Thus the content of this paper solves two major problems attached to the estimation of the multivariate linear regression model. First, the suggested estimation scheme avoids well known restrictions on the variances covariances of the usual Bayesian and frequentist estimates of $\mathrm{Vec}(B)$.

Second, the scheme provides a simple method for estimating a constrained regression coefficient matrix.

Convergence of the Gibbs sampler proposed is examined by an illustrative example in Section 4. Various diagnostic tools for testing the convergence showed that the Gibbs sequence safely converges without regard to the choice of starting points. Therefore, other inferences such as prediction of dependent vectors and hypothesis testing of B can be immediately conducted by the Gibbs sampler. The result in this paper can be readily extended to the analysis of vector time series models and MANOVA. A study pertaining to the extension is not unimportant and is left as a future study of interest.

References

[1] Barten, A.P.(1964), Consumer demand functions under conditions of almost additive preferences, *Econometrica*, 32, pp. 1-38.

[2] Box, G.E.P. and Tiao, G.C.(1972), *Bayesian Inference in Statistical Analysis*, Massachusetts: Addison-Wesley.

[3] Broemeling, L.D.(1985), *Bayesian Analysis of Linear Models*, New York: Marcel Dekker.

[4] Cowles, M.K. and Carlin, B.P.(1996), Markov chain Monte Carlo convergence diagnostics: a comparative review, *Journal of the American Statistical Association*, 91, pp. 883-904.

[5] Devroye, L.(1986), *Non-uniform Random Variate Generation*, New York : Springer.

[6] Gelman, A., and Rubin, D.B.(1992), Inference from iterative simulation using multiple sequences, *Statistical Science*, 7, pp. 457-511.

[7] Gilks, W.R.(1996), Full conditional distributions, *Markov Chain Monte Carlo in Practice*, ed. by Gilks et al., New York: Capman and Hall.

[8] Ghosh, M., Saleh, A.K. Md. E. and Sen, P.K.(1989), Empirical Bayes subset estimation in regression models, *Statistics and Decisions*, 7, pp. 15-35.

[9] Johnson, R.A. and Wichern, D.W.(1992), *Applied Multivariate Statistical Analysis*, London: Prentice Hall

[10] Press, S.J.(1989), *Bayesian Statistics: Principles, Models and Applications*, New York: Wiley.

[11] Press, S.J.(1982), *Applied Multivariate Analysis*, Florida : Krieger Publishing Co.

[12] Rothenberg, T.J.(1963), A Bayesian analysis of simultaneous equations system, Report 6315, Econometric Institute, Netherland School of Economics, Rotterdam.

[13] Smith, W.B. and Hocking, R.R.(1972), Wishart variate generation, *Applied Statistics*, 21, pp. 341-345.

[14] Spiegelhalter, A., Thomas, A. and Best, N.G.(1996), Computation on Bayesian Graphical models, *Bayesian Statistics V*, Eds. by Bernardo, J.M. et al., Oxford University Press.

[15] Tiao, G.C. and Zellner, A.(1964), On the Bayesian estimation of multivariate regression, *Journal of the Royal Statistical Society, Series B*, 26, pp. 277-285.

[16] Tierney, L.(1991), Markov chains for exploring posterior distributions, *Annals of Statistics*, 22, pp. 1701-1727.

[17] Van der Merwe, A.J. and Van der Merwe, C.A.(1992), Empirical and hierarchical Bayes estimation in multivariate regression models, *Bayesian Statistics IV*, Eds. by Bernardo, J.M. et al., Oxford University Press.

[18] Zellner, A.(1971), *An Introduction to Bayesian Inference in Econometrics*, New York : John Wiley & Sons.