Far East J. Theo. Stat. **10**(1) (2003), 11-24

ROBUST PREDICTIVE INFERENCE FOR THE MULTIVARIATE LINEAR MODEL

B. M. GOLAM KIBRIA
Department of Statistics
Florida International University
Miami, FL 33199, U. S. A.
e-mail: kibriag@fiu.edu

Abstract

This paper considers the predictive inference for the future responses from the multivariate linear models with errors following an elliptically contoured distribution. First we derive the marginal likelihood function of unknown covariance parameters. Then we derive the predictive distribution for known covariance parameters of the model. It is observed that the predictive distribution of future responses of the model has a multivariate Student's t distribution, which is identical to that obtained under the independently distributed multivariate normal errors and dependent but uncorrelated t errors. This gives inference robustness with respect to departures from the independent sampling from normal and dependent but uncorrelated sampling from Student's t to elliptically contoured distribution.

1. Introduction

In recent years, there has been significant interest in the predictive inference for a linear or multivariate linear models with various error distributions. The predictive inference for the linear models has been

2000 Mathematics Subject Classification: Primary 62J05, 62H99; Secondary 62M20.

Key words and phrases: elliptically contoured error, marginal likelihood, multivariate linear model, predictive distribution, robustness, Student's t distribution.

Communicated by Nizam Uddin

Received November 29, 2002

considered by various researchers: Goldberger [7] and Hahn [8] used the classical approach; Geisser [6], Zellner and Chetty [18] and Chib et al. [2] used the Bayesian approach, Fraser and Haq [4], Haq and Rinco [9], and Fraser and Ng [5] used the structural approach, while Haq and Kibria [10], and Kibria and Haq [11] considered the structural relation of the model for the derivation of the predictive distribution.

Most of the researchers have assumed either normal or Student's t distribution for the error variables of the model. The normality and independency assumption may not be appropriate in many practical situation, specially when the parent distributions have heavier tails. In that case, the multivariate t has been emphasized by Zellner [19], Chib et al. [2], Sutradhar and Ali [17] and recently Kibria and Haq [11] among others.

Haq and Kibria [10] and Kibria and Haq [11], using the structural relation of the model derived the predictive distribution for the future responses under the multivariate normal and multivariate t distribution, respectively. In both cases, they obtained the predictive distribution as a multivariate Student's t distribution with appropriate degrees of freedoms. Therefore, the distribution of future responses for a multivariate linear model is unaffected by a change in the error distribution from normal to Student's t distribution. The invariance of the predictive distribution for the future responses suggest that the predictive distribution would be invariance to a wide class of error distributions, namely that error terms have an multivariate elliptically contoured distribution. The elliptically contoured distribution include various distributions: the multivariate normal, matrix T, multivariate Student's t and multivariate Cauchy (see Fang, Kotz and Ng [3] and Ng [15]).

Elliptically contoured distributions have been discussed extensively for traditional multivariate regression model by Anderson and Fang [1] and recently Kubokawa and Srivastava [13] among others. This distribution has also been considered by Chib et al. [2] for the derivation of predictive distribution from the linear model and using Bayesian approach. Kibria and Haq [12] considered this distribution for the linear model and derived the predictive distribution by using the structural relation of the model. Ng [15, 16] considered this distribution and derived the predictive distributions for simple multivariate linear model under both Bayesian and classical approaches. He concluded that the Bayesian analysis using improper prior yields the same predictive distribution as the classical analysis. However, none of the researchers has considered the elliptically contoured distribution for the derivation of predictive distribution from multivariate linear with unknown covariance matrix and by structural relation approach.

In this paper, we assumed that error terms have a multivariate elliptically contoured distribution. We considered a general covariance matrix for the error variables depending on a set of parameters. First, we derived the marginal likelihood function of unknown covariance parameters and then derived the predictive distribution of future responses. We adopted the structural relation of the model approach to derive the marginal likelihood function as well as the predictive distribution.

A plan of this paper is as follows. The multivariate linear model and the covariance parameters estimation have been discussed in the following section. The predictive distribution of future responses is derived in Section 3. Some special cases have been discussed in Section 4. Finally some concluding remarks are added in Section 5.

2. The Model and Parameters Estimation

Let us consider n observations for the p characteristics yielding the following multivariate linear model

$$\mathbf{Y} = \mathbf{XB} + \sigma\mathbf{E}, \tag{2.1}$$

where $\mathbf{Y}$ is an $n \times p$ matrix of observed responses, $\mathbf{X}$ is an $n \times r$ design matrix, $\mathbf{B}$ is an $r \times p$ regression matrix, $\mathbf{E}$ is an $n \times p$ errors matrix and $\sigma > 0$ is a scale parameter. We assume that $\mathbf{E}$ has an elliptically contoured distributions with the probability density function

$$p(\mathbf{E} \mid \mathbf{\Delta}_\gamma) = |\mathbf{\Delta}_\gamma|^{-\frac{p}{2}} g[tr(\mathbf{E}'\mathbf{\Delta}_\gamma^{-1}\mathbf{E})], \tag{2.2}$$

which is of the form given in Anderson and Fang [1], $g\{\cdot\}$ is a nonnegative function over $m \times m$ positive definite matrices such that $f(\mathbf{E})$ is a density function. The covariance matrix $\mathbf{\Delta}_{\boldsymbol{\gamma}}$, depends on the parameter $\boldsymbol{\gamma}$. The observed data can be used to estimate the covariance parameters. Since the predictive distribution depends on the covariance parameters, first we will discuss about the estimation technique of covariance parameters through the marginal likelihood function.

The marginal likelihood function of $\mathbf{\Delta}_{\boldsymbol{\gamma}}$

Consider $\hat{\mathbf{B}}_{\mathbf{E}}$ as the regression matrix of $\mathbf{E}$ on $\mathbf{X}$, $s^2_{\mathbf{E}}$ as the sum of squared residual and $\mathbf{Z}_{\mathbf{E}}$ as the standardized residual matrix, then

$$\hat{\mathbf{B}}_{\mathbf{E}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E},$$

$$s^2_{\mathbf{E}} = \sum_{i=1}^{p} \sum_{j=1}^{n} (e_{ij} - \mathbf{x}_j \mathbf{b}_{\mathbf{e}i})^2, \text{ and}$$

$$\mathbf{Z}_{\mathbf{E}} = s^{-1}_{\mathbf{E}}\{\mathbf{E} - \mathbf{X}\hat{\mathbf{B}}_{\mathbf{E}}\}. \tag{2.3}$$

The corresponding expressions for the response matrix $\mathbf{Y}$ will be denoted by $\hat{\mathbf{B}}_{\mathbf{Y}}$, $s^2_{\mathbf{Y}}$ and $\mathbf{Z}_{\mathbf{Y}}$, respectively. From (2.1) and (2.3), it follows that

$$\sigma = \frac{s_{\mathbf{Y}}}{s_{\mathbf{E}}}$$

$$\mathbf{B} = \hat{\mathbf{B}}_{\mathbf{Y}} - \frac{s_{\mathbf{Y}}}{s_{\mathbf{E}}}\hat{\mathbf{B}}_{\mathbf{E}}, \text{ and}$$

$$\mathbf{Z}_{\mathbf{Y}} = \mathbf{Z}_{\mathbf{E}}. \tag{2.4}$$

Then the relationship between the volume elements of $\mathbf{E}$ in terms of the new variables $\hat{\mathbf{B}}_{\mathbf{E}}$, $s_{\mathbf{E}}$ and $\mathbf{Z}_{\mathbf{E}}$ is

$$d\mathbf{E} = |\,\mathbf{X}'\mathbf{X}\,|^{\frac{p}{2}} s_{\mathbf{E}}^{p(n-r)-1} d\hat{\mathbf{B}}_{\mathbf{E}}\, ds_{\mathbf{E}}\, d\mathbf{Z}_{\mathbf{E}}, \tag{2.5}$$

where $\mathcal{R}^{np}$ has been expressed as the direct sum of the subspace $\mathcal{L}(\mathbf{X})$ and its orthogonal component $\mathcal{O}(\mathbf{X})$ and $d\mathbf{Z}_{\mathbf{E}}$ is the volume element on the unit sphere in $\mathcal{O}(\mathbf{X})$. Using (2.3), the quadratic expression in (2.2) leads to

$$\mathbf{E}'\mathbf{\Delta}_\gamma^{-1}\mathbf{E} = (\mathbf{X}\hat{\mathbf{B}}_\mathbf{E} + s_\mathbf{E}\mathbf{Z}_\mathbf{E})'\mathbf{\Delta}_\gamma^{-1}(\mathbf{X}\hat{\mathbf{B}}_\mathbf{E} + s_\mathbf{E}\mathbf{Z}_\mathbf{E})$$

$$= (\hat{\mathbf{B}}'_\mathbf{E} + s_\mathbf{E}\mathbf{A}^{-1}\mathbf{R})'\mathbf{A}(\hat{\mathbf{B}}'_\mathbf{E} + s_\mathbf{E}\mathbf{A}^{-1}\mathbf{R}) + s_\mathbf{E}^2\mathbf{Z}'_\mathbf{E}\mathbf{\Sigma}_\gamma\mathbf{Z}_\mathbf{E},$$

where $\mathbf{A} = \mathbf{X}'\mathbf{\Delta}_\gamma^{-1}\mathbf{X}$, $\mathbf{R} = \mathbf{X}'\mathbf{\Delta}_\gamma^{-1}\mathbf{Z}_\mathbf{E}$, and

$$\mathbf{\Sigma}_\gamma = \mathbf{\Delta}_\gamma^{-1} - \mathbf{\Delta}_\gamma^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Delta}_\gamma^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Delta}_\gamma^{-1}.$$

Note that,

$$tr(\mathbf{E}'\mathbf{\Delta}_\gamma^{-1}\mathbf{E}) = \sum_{i=1}^{p}(\mathbf{b}_{\mathbf{e}i} + s_\mathbf{e}\mathbf{A}^{-1}\mathbf{r}_i)'\mathbf{A}(\mathbf{b}_{\mathbf{e}i} + s_\mathbf{e}\mathbf{A}^{-1}\mathbf{r}_i) + s_\mathbf{E}^2\sum_{i=1}^{p}\mathbf{z}'_{\mathbf{e}i}\mathbf{\Sigma}_\gamma\mathbf{z}_{\mathbf{e}i},$$

where $\mathbf{r}_i = \mathbf{X}'\mathbf{\Delta}_\gamma^{-1}\mathbf{z}_{\mathbf{e}i}$, is the i^{th} column vector of the matrix $\mathbf{R}$, $\mathbf{b}_{\mathbf{e}i}$ is the i^{th} column vector of the matrix $\hat{\mathbf{B}}_\mathbf{E}$ and $\mathbf{z}_{\mathbf{e}i}$ is the i^{th} column vector of the matrix $\mathbf{Z}_\mathbf{E}$.

Then the joint density function of $\hat{\mathbf{B}}_\mathbf{E}$, $s_\mathbf{E}$ and $\mathbf{Z}_\mathbf{E}$ conditioned on $\mathbf{\Delta}_\gamma$ becomes

$$p(\hat{\mathbf{B}}_\mathbf{E}, s_\mathbf{E}, \mathbf{Z}_\mathbf{E} \mid \mathbf{\Delta}_\gamma) \propto s_\mathbf{E}^{np-pr-1}|\mathbf{\Delta}_\gamma|^{-\frac{p}{2}} g\Bigg[\sum_{i=1}^{p}(\mathbf{b}_{\mathbf{e}i} + s_\mathbf{e}\mathbf{A}^{-1}\mathbf{r}_i)'$$

$$\times \mathbf{A}(\mathbf{b}_{\mathbf{e}i} + s_\mathbf{e}\mathbf{A}^{-1}\mathbf{r}_i) + s_\mathbf{E}^2\sum_{i=1}^{p}(\mathbf{z}'_{\mathbf{e}i}\mathbf{\Sigma}_\gamma\mathbf{z}_{\mathbf{e}i})\Bigg]. \qquad (2.6)$$

The marginal probability element of $\mathbf{Z}_\mathbf{E}$ can easily be obtained by integrating (2.6) with respect to the variables $\hat{\mathbf{B}}_\mathbf{E}$ and $s_\mathbf{E}$ as

$$p(\mathbf{Z}_\mathbf{E} \mid \mathbf{\Delta}_\gamma) = \int_0^\infty\int_{-\infty}^\infty p(\hat{\mathbf{B}}_\mathbf{E}, s_\mathbf{E}, \mathbf{Z}_\mathbf{E} \mid \mathbf{\Delta}_\gamma)d\hat{\mathbf{B}}_\mathbf{E}ds_\mathbf{E}. \qquad (2.7)$$

To evaluate (2.7), we consider the following transformation:

$$\mathbf{A}^{\frac{1}{2}}(\mathbf{b}_{\mathbf{e}i} + s_{\mathbf{E}}\mathbf{A}^{-1}\mathbf{r}_i) = \mathbf{u}_i, \quad i = 1, \ldots, p$$

$$\sqrt{\left(\sum_{i=1}^{p} \mathbf{z}'_{\mathbf{e}i}\boldsymbol{\Sigma}_\gamma \mathbf{z}_{\mathbf{e}i}\right)} s_{\mathbf{E}} = v.$$

The Jacobian of the transformation $J(\mathbf{b}_{\mathbf{e}i}, s_{\mathbf{e}} \to \mathbf{u}_i, v)$ is equal to $|\mathbf{A}|^{-\frac{1}{2}}\left(\sum_{i=1}^{p} \mathbf{z}'_{\mathbf{e}i}\boldsymbol{\Sigma}_\gamma\mathbf{z}_{\mathbf{e}i}\right)^{-\frac{1}{2}}$. Then using (2.6) and (2.7), and taking into account the Jacobian, the marginal likelihood function of $\mathbf{Z}_{\mathbf{E}}$ for given $\boldsymbol{\Delta}_\gamma$ is obtained as

$$p(\mathbf{Z}_{\mathbf{E}} \mid \boldsymbol{\Delta}_\gamma) \propto |\boldsymbol{\Delta}_\gamma|^{-\frac{p}{2}}|\mathbf{X}'\boldsymbol{\Delta}_\gamma^{-1}\mathbf{X}|^{-\frac{p}{2}}\left[\sum_{i=1}^{p} \mathbf{z}'_{\mathbf{e}i}\boldsymbol{\Sigma}_\gamma \mathbf{z}_{\mathbf{e}i}\right]^{-\frac{p(n-r)}{2}}$$

$$\times \prod_{i=1}^{p}\int_0^\infty \int_{-\infty}^\infty g\left(\sum_{i=1}^{p} \mathbf{u}'_i\mathbf{u}_i + v^2\right) v^{p(n-r)-1} d\mathbf{u}_i dv, \tag{2.8}$$

where

$$\boldsymbol{\Sigma}_\gamma = \boldsymbol{\Delta}_\gamma^{-1} - \boldsymbol{\Delta}_\gamma^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Delta}_\gamma^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Delta}_\gamma^{-1}.$$

Then using the polar transformations and following Mathai et al. [14, p. 91], (2.8) leads to

$$p(\mathbf{Z}_{\mathbf{E}} \mid \boldsymbol{\Delta}_\gamma) \propto |\boldsymbol{\Delta}_\gamma|^{-\frac{p}{2}}|\mathbf{X}'\boldsymbol{\Delta}_\gamma^{-1}\mathbf{X}|^{-\frac{p}{2}}\left[\sum_{i=1}^{p} \mathbf{z}'_{\mathbf{e}i}\boldsymbol{\Sigma}_\gamma \mathbf{z}_{\mathbf{e}i}\right]^{-\frac{p(n-r)}{2}}.$$

The density function of $\mathbf{Z}_{\mathbf{E}}$ depends on $\boldsymbol{\Delta}_\gamma$ and $\mathbf{X}$, where the elements of $\boldsymbol{\Delta}_\gamma$ are unknown. It follows from (2.4) that the *pdf* of $\mathbf{Z}_{\mathbf{Y}}$ can easily be obtained from the *pdf* of $\mathbf{Z}_{\mathbf{E}}$. Thus the marginal likelihood function of $\boldsymbol{\Delta}_\gamma$ conditioned on $\mathbf{Z}_{\mathbf{Y}}$ is obtained as

$$L(\mathbf{\Delta}_\gamma \mid \mathbf{Z_Y}) \propto |\mathbf{\Delta}_\gamma|^{-\frac{p}{2}} |\mathbf{X}'\mathbf{\Delta}_\gamma^{-1}\mathbf{X}|^{-\frac{p}{2}} \left[\sum_{i=1}^{p} \mathbf{z}'_{\mathbf{y}i} \mathbf{\Sigma}_\gamma \mathbf{z}_{\mathbf{y}i} \right]^{-\frac{p(n-r)}{2}}. \tag{2.9}$$

It is observed from the likelihood function (2.9) that a closed form estimate of $\mathbf{\Delta}_\gamma$ may not be available. However, for a particular set of observed responses $\mathbf{Y}$ and for a given design matrix, the maximum likelihood estimates of the parameters are obtainable from (2.9). Note that the result in (2.9) is identical to that obtained under the assumption of independently distributed multivariate normal errors (see, Haq and Kibria [10]) and dependent but uncorrelated multivariate Student's t error (see, Kibria and Haq [11]).

3. The Predictive Distribution

Consider a set of n_f future responses from model (2.1) as

$$\mathbf{Y}_f = \mathbf{X}_f \mathbf{B} + \sigma \mathbf{E}_f, \tag{3.10}$$

where $\mathbf{Y}_f$ and $\mathbf{E}_f$ are the $n_f \times p$ matrices of future responses and future errors respectively and $\mathbf{X}_f$ is an $n_f \times r$ matrix of future regressors.

To derive the joint distribution of $\mathbf{E}$ and $\mathbf{E}_f$, we combine the observed and future error matrices as, $\mathbf{E}^{*'} = (\mathbf{E}', \mathbf{E}'_f)'$, where $\mathbf{E}^*$ is an $(n + n_f) \times p$ matrix. Let the covariance matrix of each column of $\mathbf{E}$ be an $(n + n_f) \times (n + n_f)$ matrix $\mathbf{\Psi}_\gamma$. Then the covariance matrix of $\mathbf{E}$ is $\mathbf{I}_p \otimes \mathbf{\Psi}_\gamma$, where

$$\mathbf{\Psi}_\gamma = \begin{bmatrix} \mathbf{\Psi}_{\gamma 11} & \mathbf{\Psi}_{\gamma 12} \\ \mathbf{\Psi}_{\gamma 21} & \mathbf{\Psi}_{\gamma 22} \end{bmatrix},$$

also $\mathbf{\Psi}_{\gamma 11}$ is an $n \times n$ covariance matrix of $\mathbf{e}_i$, $\mathbf{\Psi}'_{\gamma 12} = \mathbf{\Psi}_{\gamma 21}$ is an $n \times n_f$ matrix of covariances between the components of $\mathbf{e}_i$ and $\mathbf{e}_{fi}$ and $\mathbf{\Psi}_{\gamma 22}$ is the $n_f \times n_f$ covariance matrix of $\mathbf{e}_{fi}$. Then the inverse of $\mathbf{\Psi}_\gamma$ is

$$\mathbf{\Psi}_\gamma^{-1} = \begin{bmatrix} \mathbf{\Psi}_\gamma^{11} & \mathbf{\Psi}_\gamma^{12} \\ \mathbf{\Psi}_\gamma^{21} & \mathbf{\Psi}_\gamma^{22} \end{bmatrix},$$

where

$$\mathbf{\Psi}_\gamma^{11} = [\mathbf{\Psi}_{\gamma 11} - \mathbf{\Psi}_{\gamma 12}(\mathbf{\Psi}_{\gamma 22})^{-1}\mathbf{\Psi}_{\gamma 21}]^{-1}$$

$$\mathbf{\Psi}_\gamma^{12} = -(\mathbf{\Psi}_{\gamma 11})^{-1}\mathbf{\Psi}_{\gamma 12}(\mathbf{\Psi}_\gamma^{22})^{-1}$$

$$\mathbf{\Psi}_\gamma^{21} = -(\mathbf{\Psi}_{\gamma 22})^{-1}\mathbf{\Psi}_{\gamma 21}(\mathbf{\Psi}_\gamma^{11})^{-1}$$

$$\mathbf{\Psi}_\gamma^{22} = [\mathbf{\Psi}_{\gamma 22} - \mathbf{\Psi}_{\gamma 21}(\mathbf{\Psi}_{\gamma 11})^{-1}\mathbf{\Psi}_{\gamma 12}]^{-1}.$$

We further assume that the present and future errors have an elliptically contoured distribution with the following *pdf*:

$$p(\mathbf{E}, \mathbf{E}_f \,|\, \mathbf{\Psi}_\gamma)$$

$$\propto |\mathbf{\Psi}_\gamma|^{-\frac{p}{2}} g\{tr[\mathbf{E}'\mathbf{\Psi}_\gamma^{11}\mathbf{E} + \mathbf{E}'\mathbf{\Psi}_\gamma^{12}\mathbf{E}_f + \mathbf{E}_f'\mathbf{\Psi}_\gamma^{21}\mathbf{E} + \mathbf{E}_f'\mathbf{\Psi}_\gamma^{22}\mathbf{E}_f]\}. \tag{3.11}$$

Then using (2.3), and (3.11), and taking into account the Jacobian (2.5), the joint density function of $\hat{\mathbf{B}}_\mathbf{E}$, $s_\mathbf{E}$ and $\mathbf{E}_f$ for given $\mathbf{Z}_\mathbf{Y}$ and $\mathbf{\Psi}_\gamma$ is obtained as

$$p(\hat{\mathbf{B}}_\mathbf{E}, s_\mathbf{E}, \mathbf{E}_f \,|\, \mathbf{Z}_\mathbf{Y}, \mathbf{\Psi}_\gamma)$$

$$\propto |\mathbf{\Psi}_\gamma|^{-\frac{p}{2}} s^{p(n-r)-1} g\{tr[\hat{\mathbf{B}}_\mathbf{E}'\mathbf{X}'\mathbf{\Psi}_\gamma^{11}\mathbf{X}\hat{\mathbf{B}}_\mathbf{E} + 2s_\mathbf{E}\hat{\mathbf{B}}_\mathbf{E}'\mathbf{X}'\mathbf{\Psi}_\gamma^{11}\mathbf{Z}_\mathbf{Y}$$

$$+ s_\mathbf{E}^2\mathbf{Z}_\mathbf{Y}'\mathbf{\Psi}_\gamma^{11}\mathbf{Z}_\mathbf{Y} + 2s_\mathbf{E}\mathbf{Z}_\mathbf{Y}'\mathbf{\Psi}_\gamma^{12}\mathbf{E}_f + 2\hat{\mathbf{B}}_\mathbf{E}'\mathbf{X}'\mathbf{\Psi}_\gamma^{21}\mathbf{E}_f + \mathbf{E}_f'\mathbf{\Psi}_\gamma^{22}\mathbf{E}_f]\}. \tag{3.12}$$

Consider the following transformations:

$$\begin{cases} \mathbf{U} = s_\mathbf{E}^{-1}(\mathbf{E}_f - \mathbf{X}_f\hat{\mathbf{B}}_\mathbf{E}), \\ \mathbf{V} = \hat{\mathbf{B}}_\mathbf{E}, \\ w = s_\mathbf{E}. \end{cases}$$

The Jacobian of the transformation, $J\{(\hat{\mathbf{B}}_{\mathbf{E}}, s_{\mathbf{E}}, \mathbf{E}_f) \to (\mathbf{V}, w, \mathbf{U})\}$ is equal to w^{pn_f}. The quadratic expression in (3.12) can be expressed as

$$(\mathbf{V} + w\mathbf{D}^{-1}\mathbf{K})'\mathbf{D}(\mathbf{V} + w\mathbf{D}^{-1}\mathbf{K}) + w^2(\mathbf{A}_\gamma - \mathbf{K}'\mathbf{D}^{-1}\mathbf{K}),$$

where

$$\mathbf{A}_\gamma = \mathbf{Z}'_{\mathbf{Y}}\mathbf{\Psi}_\gamma^{11}\mathbf{Z}_{\mathbf{Y}} + \mathbf{Z}'_{\mathbf{Y}}\mathbf{\Psi}_\gamma^{12}\mathbf{V} + \mathbf{V}'\mathbf{\Psi}_\gamma^{21}\mathbf{Z}_{\mathbf{Y}} + \mathbf{V}'\mathbf{\Psi}_\gamma^{22}\mathbf{V},$$

$$\mathbf{K} = \mathbf{P}_1\mathbf{Z}_{\mathbf{Y}} + \mathbf{P}_2\mathbf{V};\ \mathbf{P}_1 = \mathbf{X}'\mathbf{\Psi}_\gamma^{11} + \mathbf{X}'_f\mathbf{\Psi}_\gamma^{21},\ \mathbf{P}_2 = \mathbf{X}'\mathbf{\Psi}_\gamma^{12} + \mathbf{X}'_f\mathbf{\Psi}_\gamma^{22} \text{ and}$$

$$\mathbf{D} = \mathbf{X}'\mathbf{\Psi}_\gamma^{11}\mathbf{X} + \mathbf{X}'\mathbf{\Psi}_\gamma^{12}\mathbf{X}_f + \mathbf{X}'_f\mathbf{\Psi}_\gamma^{21}\mathbf{X} + \mathbf{X}'_f\mathbf{\Psi}_\gamma^{22}\mathbf{X}_f.$$

Thus the joint density function of $\mathbf{V}$, w and $\mathbf{U}$ for given $\mathbf{Z}_{\mathbf{Y}}$ and $\mathbf{\Psi}_\gamma$ becomes

$$\begin{aligned} p(\mathbf{V}, w, \mathbf{U} \mid \mathbf{Z}_{\mathbf{Y}}, \mathbf{\Psi}_\gamma) &\propto |\mathbf{\Psi}_\gamma|^{-\frac{p}{2}} g\{tr[(\mathbf{V} + w\mathbf{D}^{-1}\mathbf{K})'\mathbf{D}(\mathbf{V} + w\mathbf{D}^{-1}\mathbf{K}) \\ &\quad + w^2(\mathbf{A}_\gamma - \mathbf{K}'\mathbf{D}^{-1}\mathbf{K})]\} w^{p(n+n_f-r)-1} \\ &\propto |\mathbf{\Psi}_\gamma|^{-\frac{p}{2}} g\left\{\sum_{i=1}^{p}(\mathbf{v}_i + w\mathbf{D}^{-1}\mathbf{k}_i)'\mathbf{D}(\mathbf{V}_i + w\mathbf{D}^{-1}\mathbf{k}_i)\right. \\ &\quad \left. + w^2 tr(\mathbf{A}_\gamma - \mathbf{K}'\mathbf{D}^{-1}\mathbf{K})\right\} w^{p(n+n_f-r)-1}. \end{aligned} \tag{3.13}$$

Integrating (3.13) with respect to $\mathbf{V}$ and w, we have the marginal *pdf* of $\mathbf{U}$ for given $\mathbf{Z}_{\mathbf{Y}}$ and $\mathbf{\Psi}_\gamma$ as

$$p(\mathbf{U} \mid \mathbf{Z}_{\mathbf{Y}}, \mathbf{\Psi}_\gamma) \propto |\mathbf{\Psi}_\gamma|^{-\frac{p}{2}}|\mathbf{D}|^{-\frac{p}{2}}[tr\{\mathbf{A}_\gamma - \mathbf{K}'\mathbf{D}^{-1}\mathbf{K}\}]^{-\frac{p(n-r+n_f)}{2}},$$

where

$$\mathbf{A}_\gamma - \mathbf{K}'\mathbf{D}^{-1}\mathbf{K} = \mathbf{Z}'_{\mathbf{Y}}\mathbf{T}^*\mathbf{Z}_{\mathbf{Y}} + (\mathbf{U} + \mathbf{T}_3^{-1}\mathbf{Q}'_2\mathbf{Z}_{\mathbf{Y}})'\mathbf{T}_3(\mathbf{U} + \mathbf{T}_3^{-1}\mathbf{T}'_2\mathbf{Z}_{\mathbf{Y}});$$

also $\mathbf{T}_1 = \mathbf{\Psi}_\gamma^{11} - \mathbf{P}'_1\mathbf{D}^{-1}\mathbf{P}_1$, $\mathbf{T}_2 = \mathbf{\Psi}_\gamma^{12} - \mathbf{P}'_1\mathbf{D}^{-1}\mathbf{P}_2$, $\mathbf{T}_3 = \mathbf{\Psi}_\gamma^{22} - \mathbf{P}'_2\mathbf{U}^{-1}\mathbf{P}_2$ and $\mathbf{T}^* = (\mathbf{T}_1 - \mathbf{T}_2\mathbf{T}_3^{-1}\mathbf{T}'_2)$.

It is readily seen from (2.1), (2.3), and (3.10) that

$$s_{\mathbf{Y}}^{-1}\{\mathbf{Y}_f - \mathbf{X}_f\hat{\mathbf{B}}_{\mathbf{Y}}\} = s_{\mathbf{E}}^{-1}\{\mathbf{E}_f - \mathbf{X}_f\hat{\mathbf{B}}_{\mathbf{E}}\} = \mathbf{U}.$$

Finally the predictive density of $\mathbf{Y}_f$ for given $\mathbf{Y}$ and $\mathbf{\Psi}_\gamma$ is obtained as

$$p(\mathbf{Y}_f \mid \mathbf{Y}, \mathbf{\Psi}_\gamma)$$

$$\propto [tr\{\mathbf{Z}'_{\mathbf{Y}}\mathbf{T}^*\mathbf{Z}_{\mathbf{Y}} + [s_{\mathbf{Y}}^{-1}(\mathbf{Y}_f - \mathbf{X}_f\hat{\mathbf{B}}_{\mathbf{Y}}) + \mathbf{T}_3^{-1}\mathbf{T}'_2\mathbf{Z}_{\mathbf{Y}}]'\mathbf{T}_3$$

$$\times [s_{\mathbf{Y}}^{-1}(\mathbf{Y}_f - \mathbf{X}_f\hat{\mathbf{B}}_{\mathbf{Y}}) + \mathbf{T}_3^{-1}\mathbf{T}'_2\mathbf{Z}_{\mathbf{Y}}]\}]^{-\frac{p(n+n_f-r)}{2}}$$

$$= \xi_1\left[1 + \frac{1}{p(n-r)}\sum_{i=1}^{p}(\mathbf{y}_{fi} - \boldsymbol{\eta}_{\gamma i})'\mathbf{\Phi}_\gamma(\mathbf{y}_{fi} - \boldsymbol{\eta}_{\gamma i})\right]^{-\frac{p(n-r)+pn_f}{2}}, \qquad (3.14)$$

where $\xi_1 = \dfrac{\Gamma\left(\frac{p(n-r+n_f)}{2}\right)|\mathbf{\Phi}_\gamma|^{\frac{p}{2}}}{[\pi(n-r)]^{\frac{pn_f}{2}}\Gamma\left(\frac{p(n-r)}{2}\right)}$ is a normalizing constant, $\boldsymbol{\eta}_{\gamma i} =$ $\mathbf{X}_f\mathbf{b}_{\mathbf{y}i} - \mathbf{S}_3^{-1}\mathbf{S}'_2\mathbf{z}_{\mathbf{y}i}$ and

$$\mathbf{\Phi}_\gamma = \left[\frac{s_{\mathbf{Y}}^2}{p(n-r)}\left(\sum_{i=1}^{p}\mathbf{z}'_{\mathbf{y}i}\mathbf{T}^*\mathbf{z}_{\mathbf{y}i}\right)\right]^{-1}\mathbf{T}_3.$$

It is observed from (3.14) that for known $\boldsymbol{\gamma}$ and given $\mathbf{Y}$ and $\mathbf{X}$, the future responses $\mathbf{Y}_f$ has a pn_f dimensional multivariate Student t-distribution with $p(n-r)$ degrees of freedom. It is also observed that each column of $\mathbf{Y}_f$ has an n_f dimensional multivariate Student t-distribution with $(n-r)$ degrees of freedom. The location parameter vector is $\boldsymbol{\eta}_{\gamma i}$, $i = 1, 2, \ldots, p$ and the scale parameter matrix is $\mathbf{\Phi}_\gamma^{-1}$. However, for unknown $\boldsymbol{\gamma}$, one may approximate the predictive density (3.14) by its estimates $\hat{\gamma}$, obtained from the marginal likelihood function (2.9). The marginal probability density function of a single future response or a set of future responses are obtainable from (3.14). The probability density function in (3.14) is identical to that obtained under the assumption of independently distributed multivariate normal errors (see, Haq and Kibria [10]) and dependent but uncorrelated multivariate Student's t error (see, Kibria and Haq [11]).

4. Some Special Cases

In this section we will discuss some special cases of the predictive distribution in (3.14).

Case I: Linear model

For $p = 1$, the results obtained in this paper coincides with that of Kibria and Haq [12], where they considered the elliptical linear model for the derivation of predictive distribution.

Case II: Uncorrelated error

If we consider that $\mathbf{\Psi}_{\boldsymbol{\gamma}} = \mathbf{I}_{(n+n_f)}$, and $p = 1$, then we observed from (3.14) that the predictive distribution of $\mathbf{y}_f$ for given $\mathbf{y}$ is a multivariate Student t-distribution with $(n - r)$ degrees of freedom, the location parameter vector is $\mathbf{X}_f \mathbf{b}_{\mathbf{y}}$ and precision matrix is $\frac{1}{s_{\mathbf{y}}^2}\{\mathbf{I}_{n_f} + \mathbf{X}_f[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}_f'\}^{-1}$. This result agrees with that of Zellner and Chetty [18], where they used the Bayesian approach with the Gaussian independence error.

Case III: MA(1) error

Let

$$\mathbf{\Psi}_{\boldsymbol{\gamma}} = \begin{pmatrix} 1+\theta^2 & -\theta & 0 & \cdots & 0 & 0 \\ -\theta & 1+\theta^2 & -\theta & \cdots & 0 & 0 \\ \cdots & -\theta & 1+\theta^2 & \cdots & \cdots\cdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots\cdots & \\ 0 & 0 & 0 & \cdots & -\theta & 1+\theta^2 \end{pmatrix},$$

where $\boldsymbol{\gamma} = \theta$, be the covariance matrix for the observed and future error variables of the MA(1) model. Then the result obtained coincides with that Haq and Kibria [10], in which they have considered the multivariate linear model with a Gaussian MA(1) error processes.

Case IV: Intraclass correlation

Consider

$$\mathbf{\Psi}_\gamma = \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \cdots & \cdots & \cdots & \cdots & \cdots\cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots\cdots \\ \rho & \rho & \rho & \cdots & 1 \end{bmatrix}, \tag{4.15}$$

where $\boldsymbol{\gamma} = \rho,$ be the covariance matrix for the observed and future error variables of the intraclass correlation model. Then the result agrees with that Kibria and Haq [11], in which they have considered the multivariate linear model with Student t and intraclass error processes.

Case V: Pearson type III error

Consider $p = 1$ in model (2.1), then as a special case of elliptical distribution, we will consider the Pearson type VII distribution for the error variables $\mathbf{e}$ as follows:

$$p(\mathbf{e} \mid \mathbf{\Delta}_\gamma) \propto \left[1 + \frac{\mathbf{e}'\mathbf{\Delta}_\gamma^{-1}\mathbf{e}}{\nu}\right]^{-N}, \; N > \frac{n}{2}, \; \nu > 0. \tag{4.16}$$

This distribution can be viewed as a special case of multivariate Student t distribution with $\nu = 2N - n$ degrees of freedom. Then following similar steps as before, we obtain the predictive distribution of $\mathbf{y}_f$ for given $\mathbf{y}$ and $\mathbf{\Psi}_\gamma$ as

$$p(\mathbf{y}_f \mid \mathbf{y}, \mathbf{\Psi}_\gamma) = \xi_2\left[1 + \frac{1}{(n-r)}(\mathbf{y}_f - \boldsymbol{\eta}_\gamma)'\mathbf{\Phi}_\gamma(\mathbf{y}_f - \boldsymbol{\eta}_\gamma)\right]^{-\frac{(n-r+n_f)}{2}}, \tag{4.17}$$

where $\xi_2 = \dfrac{\Gamma\left(\dfrac{(n-r+n_f)}{2}\right)|\mathbf{\Phi}_\gamma|^{\frac{1}{2}}}{[\pi(n-r)]^{\frac{n_f}{2}}\Gamma\left(\dfrac{n-r}{2}\right)}$ is the normalizing constant, $\boldsymbol{\eta}_\gamma =$

$\mathbf{X}_f \mathbf{b}_\mathbf{y} - s_\mathbf{y} \mathbf{T}_3^{-1} \mathbf{T}_2' \mathbf{z}_\mathbf{y}$ and $\boldsymbol{\Phi}_\gamma = \left[\frac{s_\mathbf{y}^2}{(n-r)} \{\mathbf{z}_\mathbf{y}' \mathbf{T}^* \mathbf{z}_\mathbf{y}\} \right]^{-1} \mathbf{T}_3$. The result in (4.17) coincides with that of (3.14) for $p = 1$, which is the predictive distribution for the elliptical linear model.

5. Concluding Remarks

The marginal likelihood function and the predictive distribution under the multivariate linear models with elliptically contoured distribution have been discussed in this paper. It is observed that for known covariance matrix, the predictive distribution of future responses follows a multivariate Student t-distribution with $p(n-r)$ degrees of freedom. It is interesting to note that the degrees of freedom of the predictive distribution does not depend on the degrees of freedom of the original distribution (follows from 4.17 and also from Kibria and Haq [11]). It is noted that the predictive distribution of future responses under the multivariate elliptically contoured errors assumption are identical to those obtained under independent normal errors or Student's t errors. This gives inference robustness with respect to departure from the reference case of independent sampling from the multivariate normal or dependent but uncorrelated sampling from Student's distributions. It is also noted that the results of Zellner [19], Chib et al. [2], Haq and Kibria [10], and Kibria and Haq [11, 12] follow as a special case of this paper. Furthermore, this paper is a generalization in the sense that it lead to results under the class of elliptically contoured distribution, such as normal, Student t, Cauchy. This paper considered the multivariate linear model, which does cover the linear model for $p = 1$.

References

[1] T. W. Anderson and Kai-Tai Fang, Inference in multivariate elliptically contoured distributions based on maximum likelihood, Statistical Inference in Elliptically Contoured and Related Distributions, K. T. Fang and T. W. Anderson, Eds., Allerton Press, New York, 1990, pp. 201-216.

[2] S. Chib, R. C. Tiwari and S. R. Jammalamadaka, Bayes prediction in regressions with elliptical errors, J. Econometrics 38 (1988), 349-360.

[3] K. T. Fang, S. Kotz and K. W. Ng, Symmetric Multivariate and Related Distributions, Chapman and Hall, London, 1990.

[4] D. A. S. Fraser and M. S. Haq, Structural probability and prediction for the multivariate model, J. Royal Stat. Soc. Ser. B 31 (1969), 317-331.

[5] D. A. S. Fraser and K. W. Ng, Multivariate regression analysis with spherical error, Multivariate Analysis, V. P. R. Krishnaiah, Ed., North-Holland Publishing Co., New York, 1980, pp. 369-386.

[6] S. Geisser, Bayesian estimation in multivariate analysis, Ann. Math. Stat. 36 (1965), 150-159.

[7] A. S. Goldberger, Best linear unbiased prediction in the generalized linear regression model, J. Amer. Stat. Assoc. 57 (1962), 369-375.

[8] G. J. Hahn, Simultaneous prediction intervals for a regression model, Technometrics 14 (1972), 203-214.

[9] M. S. Haq and S. Rinco, β-expectation tolerance region for a generalized multivariate model with normal error variable, J. Mult. Anal. 6 (1976), 414-421.

[10] M. S. Haq and B. M. G. Kibria, Predictive inference for linear and multivariate linear models with MA(1) error process, Comm. Stat. A 26(2) (1997), 331-353.

[11] B. M. G. Kibria and M. S. Haq, The multivariate linear model with multivariate t and intra-class covariance structure, Statistical Papers 40 (1999), 263-276.

[12] B. M. G. Kibria and M. S. Haq, Predictive inference for the elliptical linear model, J. Mult. Anal. 68 (1999), 235-249.

[13] T. Kubokawa and M. S. Srivastava, Robust improvement in estimation of a mean matrix in an elliptically contoured distribution, J. Mult. Anal. 76 (2001), 138-152.

[14] A. M. Mathai, S. B. Provost and T. Hayakawa, Bilinear Forms and Zonal Polynomials, Springer-Verlag, New York, 1995.

[15] M. V. Ng, A note on predictive inference for multivariate elliptically contoured distributions, Comm. Stat.-Theo. Meth. 29 (2000), 477-483.

[16] M. V. Ng, Robust Bayesian inference for seemingly unrelated regression with elliptical errors, J. Mult. Anal. 80 (2002), 122-135.

[17] B. C. Sutradhar and M. M. Ali, A generalization of the Wishart distribution for the elliptical model and its moments for the multivariate t model, J. Mult. Anal. 29 (1989), 155-162.

[18] A. Zellner and V. K. Chetty, Prediction and decision problems in regression models from the Bayesian point of view, J. Amer. Stat. Assoc. 60 (1965), 608-616.

[19] A. Zellner, Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms, J. Amer. Stat. Assoc. 71 (1976), 400-405.

■