Proc. ICCS-11, Lahore, Pakistan
December 19-22, 2011, Vol. 21, pp. 55-65

Bayesian Prediction of Matrix Elliptical Multivariate Models with Conjugate Prior

Mohammad Arashi [1] and Shahjahan Khan [2]
[1]Faculty of Mathematics, Shahrood University of Technology
P.O. Box 316-3619995161, Shahrood, Iran
Email: m_arashi_stat@yahoo.com
[2]Department of Mathematics and Computing
Australian Centre for Sustainable Catchments
University of Southern Queensland
Toowoomba, Queensland, AUSTRALIA,
Email: khans@usq.edu.au

Abstract

This paper considers a multivariate regression model with a matrix variate elliptically contoured (MEC) distribution for the responses. The MEC is defined as a mixture distribution of inverse Laplace transformation and matrix variate normal distribution. The prediction distribution of a set of matrix future responses from the same model with common indexing parameters is obtained under Bayesian framework with the conjugate prior of the parameters. The prediction distribution is found to be a matrix-T distribution. The results of the paper is a generalization of earlier results for linear models in terms of the generalization of the (i) multiple regression model, (ii) normal or Student-t distribution, and (iii) joint conjugate prior distribution.

Key words and phrases: Multivariate regression model, Elliptically contoured distribution, Inverse Laplace transform, Conjugate prior, matrix-T distribution, and Prediction.

1 Introduction and Some Preliminaries

Prediction analysis (see Aitchison and Dunsmore 1975, and Geisser 1993) is predominantly a Bayesian method. The predictive inference is usually based on the prediction distribution of future unobserved responses, conditional on the realized responses from the same model. The derivation of prediction distribution involves integration of the parameters from the product of the joint distribution of the realized and future responses and that of the underlying parameters. In the multivariate case such integrations are over the multidimensional spaces or Manifolds.

Under the Bayesian framework, there has been many systematic studies for linear regression models with non-normal errors. Interested readers may refer to the papers by Zellner (1976), Singh et al. (1995), Jammalamadaka et al. (1987), Chib et al. (1991), Khan and Haq (1994), Osiewalski and Steel (1993), and more recently Fang and Li (1999), Khan (2002, 2004), Ng (2002), Arashi (2010), Ng (2010), Vidal and Arellano-Valle (2010), and Tsukuma (2010).

In this section, we propose some necessary tools for mathematical computations with matrices. Also this paper we consider a wider class of distributions namely the matrix variate elliptically contoured (MEC) distributions. It is well known that a large number of commonly used distributions are member of the MEC family including the most popular matrix normal and matrix-T distributions.

Let $\boldsymbol{X}$ be an $n \times p$ random matrix, which can be expressed in terms of its elements, columns and rows as

$$\boldsymbol{X} = (x_{ij}) = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_p) = (\boldsymbol{x}_{(1)}, \cdots, \boldsymbol{x}_{(n)})'. \tag{1.1}$$

Here $\boldsymbol{x}_{(1)}, \cdots, \boldsymbol{x}_{(n)}$ can be regarded as a sample of size n from a p-dimensional population.

As pointed by Fang and Li (1999), there are many ways (not completely different) to define matrix variate elliptically contoured distributions. Four classes of elliptical matrix variate distributions are defined and discussed by Dawid (1977), and Anderson and Fang (1990). For the purpose of this paper we specifically consider the following situation

Definition 1.1. *An $n \times p$ random matrix $\boldsymbol{X}$ has a family of MEC distributions if its density has the form*

$$g(\boldsymbol{X}) = d_{n,p}|\boldsymbol{\Sigma}|^{-\frac{n}{2}} f\left\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{1}\boldsymbol{\mu}')'(\boldsymbol{X} - \boldsymbol{1}\boldsymbol{\mu}')\right\}, \tag{1.2}$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\boldsymbol{\Sigma}$ is a $p \times p$ semi-positive definite matrix. This distribution is denoted by $\boldsymbol{X} \sim E_{n,p}(\boldsymbol{\mu}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}, f)$. For notational convenience we may also use $\boldsymbol{X} \sim E_{n,p}(\boldsymbol{\mu}, \boldsymbol{I}_n, \boldsymbol{\Sigma}, f)$ where needed.

Definition 1.1 imposes the condition $f(\boldsymbol{AB}) = f(\boldsymbol{BA})$ on the density generator f for any $p \times p$ positive definite symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. Throughout the paper, without

loss of generality, we take tr operation from the argument of $f(.)$.
Some commonly used members of the MEC distributions are stated below.

(i) Matrix Variate Normal (MN) Distribution

A random matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ has a MN distribution, with mean $\boldsymbol{M}$, row and column covariance matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ respectively is denoted by $\boldsymbol{X} \sim MN_{n,p}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Omega})$, if its pdf is given by

$$f(\boldsymbol{X}) = \frac{|\boldsymbol{\Omega}|^{-\frac{n}{2}}|\boldsymbol{\Sigma}|^{-\frac{p}{2}}}{(2\pi)^{\frac{np}{2}}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Omega}^{-1}(\boldsymbol{X}-\boldsymbol{M})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{M})\right]\right\}.$$

(ii) Matrix Variate Student-t (MT) Distribution

A random matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ has a MT distribution, with mean $\boldsymbol{M}$, row and column scale matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ respectively and ν d.f. is denoted by $\boldsymbol{X} \sim MT_{n,p}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \nu)$, if its pdf is given by

$$f(\boldsymbol{X}) = \frac{|\boldsymbol{\Omega}|^{-\frac{n}{2}}|\boldsymbol{\Sigma}|^{-\frac{p}{2}}}{g_{n,p}} \left|\boldsymbol{I}_n + \boldsymbol{\Omega}^{-1}(\boldsymbol{X}-\boldsymbol{M})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{M})\right|^{-\frac{n+p+\nu-1}{2}},$$

where

$$g_{n,p} = \frac{(\nu\pi)^{\frac{np}{2}}\Gamma_p\left(\frac{\nu+p-1}{2}\right)}{\Gamma_p\left(\frac{\nu+n+p-1}{2}\right)}.$$

(iii) Matrix Variate Pearson Type-VII (MPVII) Distribution

A random matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ has a MPVII distribution, with mean $\boldsymbol{M}$, row and column scale matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ respectively and parameters m and q is denoted by $\boldsymbol{X} \sim MPVII_{n,p}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, m, q)$, if its pdf is given by

$$f(\boldsymbol{X}) = \frac{|\boldsymbol{\Omega}|^{-\frac{n}{2}}|\boldsymbol{\Sigma}|^{-\frac{p}{2}}}{h_{n,p,m,q}} \left|\boldsymbol{I}_n + \boldsymbol{\Omega}^{-1}(\boldsymbol{X}-\boldsymbol{M})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{M})\right|^{-m},$$

where

$$h_{n,p,m,q} = \frac{(q\pi)^{\frac{np}{2}}\Gamma_p\left(m-\frac{n}{2}\right)}{\Gamma_p(m)}.$$

(iv) Matrix Variate Power Exponential (MPE) Distribution

A random matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ has a MPE distribution, with mean $\boldsymbol{M}$, row and column scale matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ respectively and parameters r and s is denoted by $\boldsymbol{X} \sim MPE_{n,p}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, r, s)$, if its pdf is given by

$$f(\boldsymbol{X}) = \frac{|\boldsymbol{\Omega}|^{-\frac{n}{2}}|\boldsymbol{\Sigma}|^{-\frac{p}{2}}}{j_{n,p,r,s}} \exp\left\{-\frac{r}{2}\left(\operatorname{tr}\left[\boldsymbol{\Omega}^{-1}(\boldsymbol{X}-\boldsymbol{M})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{M})\right]\right)^{s}\right\},$$

where

$$j_{n,p,r,s} = \frac{(2\pi)^{\frac{np}{2}}\Gamma\left(\frac{np}{2s}\right)}{s\Gamma\left(\frac{np}{2}\right)r^{\frac{np}{2s}}}.$$

(v) Matrix Variate Laplace (ML) Distribution

A random matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ has a ML distribution, with mean $\boldsymbol{M}$, row and column scale matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ respectively is denoted by $\boldsymbol{X} \sim ML_{n,p}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Omega})$, if its pdf is given by

$$f(\boldsymbol{X}) = \frac{|\boldsymbol{\Omega}|^{-\frac{n}{2}}|\boldsymbol{\Sigma}|^{-\frac{p}{2}}}{l_{n,p}} \exp\left\{-\frac{\sqrt{2}}{2}\left(\mathrm{tr}\left[\boldsymbol{\Omega}^{-1}(\boldsymbol{X}-\boldsymbol{M})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{M})\right]\right)^{\frac{1}{2}}\right\},$$

where

$$i_{n,p} = \frac{2\pi^{\frac{np}{2}}\Gamma(np)}{\Gamma\left(\frac{np}{2}\right)}.$$

(vi) Matrix Variate Kotz-Type (MK) Distribution

A random matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ has a MK distribution, with mean $\boldsymbol{M}$, row and column scale matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ respectively is denoted by $\boldsymbol{X} \sim MK_{n,p}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, s)$, if its pdf is given by

$$f(\boldsymbol{X}) = \frac{|\boldsymbol{\Omega}|^{-\frac{n}{2}}|\boldsymbol{\Sigma}|^{-\frac{p}{2}}}{t_{n,p,r}} \exp\left\{-\frac{r}{2}\left(\mathrm{tr}\left[\boldsymbol{\Omega}^{-1}(\boldsymbol{X}-\boldsymbol{M})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{M})\right]\right)\right\},$$

where

$$t_{n,p,r} = \frac{(2\pi)^{\frac{np}{2}}\Gamma\left(\frac{np}{2}\right)}{\Gamma\left(\frac{np}{2}\right)r^{\frac{np}{2}}}.$$

It is well known that the family of MEC distributions can always be expressed in an integral form of a matrix variate normal distribution with *weight function* as given in the following theorem.

Theorem 1.1. *Let $\boldsymbol{X} \sim E_{n,p}(\boldsymbol{\mu}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}, f)$ where the pdf $g(\boldsymbol{X})$ of $\boldsymbol{X}$ is defined by*

$$g(\boldsymbol{X}) = |\boldsymbol{\Sigma}|^{-n/2}h\left[\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{1}\boldsymbol{\mu}')'(\boldsymbol{X}-\boldsymbol{1}\boldsymbol{\mu}')\right].$$

If $h(t)$, $t \in [0,\infty)$ has the inverse Laplace transform (denoted by $\mathcal{L}^{-1}[h(t)]$), then we have

$$g(\boldsymbol{X}) = \int_0^\infty w(z) f_{\mathcal{N}(\boldsymbol{\mu}, z^{-1}\boldsymbol{I}_n\otimes\boldsymbol{\Sigma})}(\boldsymbol{X})\,\mathrm{d}z, \tag{1.3}$$

where $f_{\mathcal{N}(\boldsymbol{\mu}, z^{-1}\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})}(\boldsymbol{X})$ *stands for the pdf of the* $n \times p$ *matrix* $\boldsymbol{X}$ *distributed as matrix variate normal with the mean matrix* $\mathbf{1}\boldsymbol{\mu}'$ *and the covariance matrix* $z^{-1}\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$, *and* $w(z)$ *is the weight function given by*

$$w(z) = (2\pi)^{np/2} z^{-np/2} \mathcal{L}^{-1}[h(2t)]. \tag{1.4}$$

For the proof refer to Theorem 4.2.1 of Gupta and Varga (1995).

Remark 1.1. *In Theorem 1.1 it is stated that* f *is the density of* $N_{n,p}(\boldsymbol{\mu}, z^{-1}\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$. *It is realized from the proof that* f *can even has one of the following densities*

(1) $N_{n,p}(\boldsymbol{\mu}, z^{-1}\boldsymbol{I}_n, \boldsymbol{\Sigma})$ *or*

(2) $N_{n,p}(\boldsymbol{\mu}, \boldsymbol{I}_n, z^{-1}\boldsymbol{\Sigma})$ *or*

(3) $N_{n,p}(\boldsymbol{\mu}, z^{-\frac{1}{2}}\boldsymbol{I}_n, z^{-\frac{1}{2}}\boldsymbol{\Sigma})$.

The above facts enable us to adopt each representation whenever is needed for practical use. In sequel we present some necessary tools for the derivations in forthcoming sections.

Section 2 describes the multivariate regression model as a generalization of the multiple regression model. The associated future multivariate regression model as well as the prediction distribution are covered in Section 3. Some concluding remarks are provided in Section 4.

2 Multivariate Regression Model

In this section, we consider a multivariate regression model with the MEC errors. For a precise setup consider the following regression model

$$\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{X} + \boldsymbol{E}, \tag{2.1}$$

where the n columns of the $p \times n$ response matrix $\boldsymbol{Y}$ can be regarded as a sample of size n from a p-dimensional population, $\boldsymbol{X}$ is the $k \times n$ design matrix of known values of rank k, $\boldsymbol{B}$ is the $p \times k$ matrix of unknown regression parameters, $n > p + k$. The $p \times n$ error matrix $\boldsymbol{E}$ is assumed to have a MEC distribution, $E_{p,n}(\mathbf{0}, \boldsymbol{\Phi}^{-1} \otimes \boldsymbol{I}_n, f)$.

From Theorem 1.1 we can immediately write

$$\begin{aligned} f(\boldsymbol{Y}|\boldsymbol{B}, \boldsymbol{\Phi}) &= \int_0^\infty f_{\mathcal{N}(\boldsymbol{B}\boldsymbol{X}, t^{-1}\boldsymbol{\Phi}^{-1} \otimes \boldsymbol{I}_n)}(\boldsymbol{Y}) W^*(t)dt \\ &= \int_0^\infty f(\boldsymbol{Y}|\boldsymbol{B}, \boldsymbol{\Phi}, t) W^*(t)dt, \end{aligned} \tag{2.2}$$

where $W^*(t)$ is the weight matrix. Now following Ng (2010) let $\boldsymbol{\Sigma} = t\boldsymbol{\Phi}$ with the Jacobian of transformation $J(\boldsymbol{\Sigma} \to \boldsymbol{\Phi}) = t^{\frac{p(p+1)}{2}}$. Further adopt a normal-Wishart conjugate prior for $(\boldsymbol{B}, \boldsymbol{\Sigma})$ as

$$\begin{aligned} \pi(\boldsymbol{B}|\boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{\frac{k}{2}} etr\left\{-\frac{1}{2}\,\boldsymbol{\Sigma}(\boldsymbol{B}-\boldsymbol{B}^*)\boldsymbol{\Upsilon}(\boldsymbol{B}-\boldsymbol{B}^*)'\right\}, \\ \pi(\boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{\nu-\frac{p+1}{2}} etr\left\{-\frac{1}{2}\,\boldsymbol{\Omega}\boldsymbol{\Sigma}\right\}, \end{aligned}$$

where ν, $p\times k$ matrix $\boldsymbol{B}^*$, $k\times k$ matrix $\boldsymbol{\Upsilon}$ and $p\times p$ matrix $\boldsymbol{\Omega}$ are all known hyperparameters.

Suppose

$$\begin{aligned} \pi(\boldsymbol{B}, \boldsymbol{\Phi}|t) &\propto \pi(\boldsymbol{B}|\boldsymbol{\Sigma})\pi(\boldsymbol{\Sigma})J(\boldsymbol{\Sigma} \to \boldsymbol{\Phi}) \\ &\propto t^{\frac{p(\nu+k)}{2}} |\boldsymbol{\Phi}|^{\frac{\nu+k}{2}-\frac{p+1}{2}} etr\left\{-\frac{1}{2}\,t\boldsymbol{\Phi}\left[(\boldsymbol{B}-\boldsymbol{B}^*)\boldsymbol{\Upsilon}(\boldsymbol{B}-\boldsymbol{B}^*)' + \boldsymbol{\Omega}\right]\right\}. \end{aligned} \quad (2.3)$$

Then the conjugate prior distribution for the MEC distribution can be obtained as

$$\pi(\boldsymbol{B}, \boldsymbol{\Phi}) \propto \int_0^\infty \pi(\boldsymbol{B}, \boldsymbol{\Phi}|t)W^*(t)dt. \quad (2.4)$$

3 Prediction Distribution

Suppose $\boldsymbol{Y}$ in equation (2.1) is observable and $\boldsymbol{Y}_f$ in

$$\boldsymbol{Y}_f = \boldsymbol{B}\boldsymbol{X}_f + \boldsymbol{E}_f \quad (3.1)$$

is an observable $p\times n_f$ matrix of future responses with a $k \times n_f$ design matrix $\boldsymbol{X}_f$ of known values with rank k, satisfying $n + n_f > p + k$.

In the same fashion as in Khan (2002) and Ng (2010), the joint density of $(\boldsymbol{Y}, \boldsymbol{Y}_f)$ is then given by

$$f(\boldsymbol{Y}, \boldsymbol{Y}_f|\boldsymbol{B}, \boldsymbol{\Phi}) \propto \int_0^\infty f(\boldsymbol{Y}, \boldsymbol{Y}_f|\boldsymbol{B}, \boldsymbol{\Phi}, t)W^*(t)dt,$$

where

$$f(\boldsymbol{Y}, \boldsymbol{Y}_f|\boldsymbol{B}, \boldsymbol{\Phi}, t) \propto |t\boldsymbol{\Phi}|^{\frac{1}{2}(n+n_f)}\,\mathrm{etr}\left\{-\frac{t}{2}\boldsymbol{\Phi}\left[\|\boldsymbol{Y}-\boldsymbol{B}\boldsymbol{X}\|^2 + \|\boldsymbol{Y}_f - \boldsymbol{B}\boldsymbol{X}_f\|^2\right]\right\}. \quad (3.2)$$

Consequently the Bayesian predictive density function of $\boldsymbol{Y}_f$ is defined as

$$f(\boldsymbol{Y}_f|\boldsymbol{Y}) \propto \int\int\int f(\boldsymbol{Y}, \boldsymbol{Y}_f|\boldsymbol{B}, \boldsymbol{\Phi}, t)\pi(\boldsymbol{B}, \boldsymbol{\Phi}|t)W^*(t)d\boldsymbol{B}d\boldsymbol{\Phi}dt, \quad (3.3)$$

where $\pi(\boldsymbol{B}, \boldsymbol{\Phi}|t)$ is given by (2.3). In the sequel we propose the main result of this approach.

Theorem 3.1. *For the multivariate regression model (2.1) where $\boldsymbol{Y}$ follows a family of MEC distributions, using the conjugate prior on $(\boldsymbol{B}, \boldsymbol{\Phi})$ given by (2.4), the predictive distribution of $\boldsymbol{Y}_f$, is the matrix-T distribution denoted by $\boldsymbol{Y}_f|\boldsymbol{Y} \sim T_{p,n_f}(\boldsymbol{M}, \boldsymbol{H}^{-1}, \boldsymbol{R}, n_f + p + \nu - 1)$ with the following density*

$$\begin{aligned} f(\boldsymbol{Y}_f|\boldsymbol{Y}) &= c(p, n_f, \nu)|\boldsymbol{H}|^{\frac{p}{2}}|\boldsymbol{R}^{-1}|^{\frac{n_f}{2}} \\ &\quad \times|\boldsymbol{I}_p + \boldsymbol{R}^{-1}(\boldsymbol{Y}_f - \boldsymbol{M})'\boldsymbol{H}(\boldsymbol{Y}_f - \boldsymbol{M})|^{-\frac{1}{2}(p+n_f+\nu-1)}, \end{aligned}$$

where

$$\begin{aligned} c(p, n_f, \nu)^{-1} &= \left[\Gamma\left(\frac{1}{2}\right)\right]^{pn_f} \frac{\Gamma_p\left[\frac{1}{2}(n_f + \nu - 1)\right]}{\Gamma_p\left[\frac{1}{2}(p + n_f + \nu - 1)\right]}, \\ \boldsymbol{L} &= \boldsymbol{X}\boldsymbol{X}' + \boldsymbol{X}_f\boldsymbol{X}_f' + \boldsymbol{\Upsilon}, \\ \boldsymbol{H} &= \boldsymbol{I}_{n_f} - \boldsymbol{X}_f'\boldsymbol{L}^{-1}\boldsymbol{X}_f \\ \boldsymbol{M} &= (\boldsymbol{Y}\boldsymbol{X}' + \boldsymbol{B}^*\boldsymbol{\Upsilon})\boldsymbol{L}^{-1}\boldsymbol{X}_f\boldsymbol{H}^{-1} \\ \boldsymbol{R} &= \boldsymbol{Y}\boldsymbol{Y}' + \boldsymbol{B}^*\boldsymbol{\Upsilon}\boldsymbol{B}^{*\prime} + \boldsymbol{\Omega} + (\boldsymbol{Y}\boldsymbol{X}' + \boldsymbol{B}^*\boldsymbol{\Upsilon})\boldsymbol{L}^{-1}(\boldsymbol{Y}\boldsymbol{X}' + \boldsymbol{B}^*\boldsymbol{\Upsilon})' \\ &\quad -(\boldsymbol{Y}\boldsymbol{X}' + \boldsymbol{B}^*\boldsymbol{\Upsilon})\boldsymbol{L}^{-1}\boldsymbol{X}_f\boldsymbol{M}' \end{aligned} \tag{3.4}$$

in which $\Gamma_p(\cdot)$ is the generalized gamma function defined as $\Gamma_b\left(\frac{c}{2}\right) = \pi^{\frac{b(b-1)}{2}} \prod_{i=1}^{b} \Gamma\left(\frac{c-i+1}{2}\right)$.

Proof: Let $f(\boldsymbol{Y}_f|t)$ be the density of $\boldsymbol{Y}_f$ under normal assumption, then by definition we have

$$f(\boldsymbol{Y}_f|\boldsymbol{Y}) \propto \int f(\boldsymbol{Y}_f|\boldsymbol{Y}, t)W^*(t)dt,$$

where

$$\begin{aligned} f(\boldsymbol{Y}_f|\boldsymbol{Y}, t) &= \int_{\boldsymbol{\Phi}>0}\int f(\boldsymbol{Y}, \boldsymbol{Y}_f|\boldsymbol{B}, \boldsymbol{\Phi}, t)\pi(\boldsymbol{B}, \boldsymbol{\Phi}|t)d\boldsymbol{B}d\boldsymbol{\Phi} \\ &\propto \int\int t^{\frac{p(n+n_f+\nu+k)}{2}}|\boldsymbol{\Phi}|^{\frac{n+n_f+\nu+k}{2}-\frac{p+1}{2}} \\ &\quad \times etr\left\{-\frac{1}{2}t\boldsymbol{\Phi}\left[\|\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{X}\|^2 + \|\boldsymbol{Y}_f - \boldsymbol{B}\boldsymbol{X}_f\|^2\right]\right\} \\ &\quad \times etr\left\{-\frac{1}{2}t\boldsymbol{\Phi}\left[(\boldsymbol{B} - \boldsymbol{B}^*)\boldsymbol{\Upsilon}(\boldsymbol{B} - \boldsymbol{B}^*)' + \boldsymbol{\Omega}\right]\right\}d\boldsymbol{B}d\boldsymbol{\Phi}. \end{aligned} \tag{3.5}$$

Using the fact that

$$\begin{aligned} &\|\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{X}\|^2 + \|\boldsymbol{Y}_f - \boldsymbol{B}\boldsymbol{X}_f\|^2 + (\boldsymbol{B} - \boldsymbol{B}^*)\boldsymbol{\Upsilon}(\boldsymbol{B} - \boldsymbol{B}^*)' + \boldsymbol{\Omega} \\ &\quad = (\boldsymbol{Y}_f - \boldsymbol{M})\boldsymbol{H}(\boldsymbol{Y}_f - \boldsymbol{M})' + (\boldsymbol{B} - \hat{\boldsymbol{B}})\boldsymbol{L}(\boldsymbol{B} - \hat{\boldsymbol{B}})' + \boldsymbol{R}, \end{aligned}$$

where

$$\hat{B} = (Y_f X'_f + B(XX' + \Upsilon))L^{-1}.$$

Then the expression in (3.5) simplifies to

$$\begin{aligned} f(Y_f|Y,t) \quad \propto \quad & \int\int_{\Phi>0} t^{\frac{1}{2}p(n+n_f+\nu)} |\Phi|^{\frac{n+n_f+\nu}{2}-\frac{p+1}{2}} \\ & \times \operatorname{etr}\left\{-\frac{t}{2}\Phi\left[(Y_f - M)H(Y_f - M)' + R\right]\right\} d\Phi \\ & \times t^{\frac{pk}{2}} |\Phi|^{\frac{k}{2}} \operatorname{etr}\left\{-\frac{t}{2}\Phi(B-\hat{B})L(B-\hat{B})'\right\} dB. \end{aligned}$$

Integrating over B, we get

$$\begin{aligned} f(Y_f|Y,t) \quad \propto \quad & \int_{\Phi>0} t^{\frac{1}{2}p(n+n_f+\nu)} |\Phi|^{\frac{n+n_f+\nu}{2}-\frac{p+1}{2}} \\ & \times \operatorname{etr}\left\{-\frac{t}{2}\Phi\left[(Y_f - M)H(Y_f - M)' + R\right]\right\} d\Phi \end{aligned}$$

Now make the transformation

$$Z = t\Phi\left[(Y_f - M)H(Y_f - M)' + R\right]$$

with the Jacobian of the transformation $|Z|^{-\frac{p+1}{2}}$ to get

$$\begin{aligned} f(Y_f|Y,t) \quad \propto \quad & \int_{Z>0} t^{\frac{1}{2}p(n+n_f+\nu)} |\left[(Y_f - M)H(Y_f - M)' + R\right]^{-1} t^{-1} Z|^{\frac{p+n_f+\nu}{2}-\frac{p+1}{2}} \\ & \times |Z|^{-\frac{p+1}{2}} etr\left(-\frac{1}{2}Z\right) dZ \\ \propto \quad & |\left[(Y_f - M)H(Y_f - M)' + R\right]|^{-\frac{p+n_f+\nu-1}{2}} \\ & \times \int_{Z>0} |Z|^{\frac{p+n_f+\nu}{2}-\frac{p+1}{2}} etr\left(-\frac{1}{2}Z\right) dZ \\ \propto \quad & |\left[(Y_f - M)H(Y_f - M)' + R\right]|^{-\frac{p+n_f+\nu-1}{2}}. \end{aligned}$$

Thus we conclude that

$$\begin{aligned} f(Y_f|Y) \quad \propto \quad & |\left[(Y_f - M)H(Y_f - M)' + R\right]|^{-\frac{p+n_f+\nu-1}{2}} \int W(t)dt \\ \propto \quad & |\left[(Y_f - M)H(Y_f - M)' + R\right]|^{-\frac{p+n_f+\nu-1}{2}} \\ \propto \quad & |I_p + R^{-1}\left[(Y_f - M)'H(Y_f - M)\right]|^{-\frac{p+n_f+\nu-1}{2}}. \end{aligned}$$

Remark 3.1. *The predictive distribution of $\mathbf{Y}_f$ under matrix normal responses is identical to the matrix-T distribution above (see Broemeling 1985, p.379). Thus one can carry out further inference for $\mathbf{Y}_f$ and find the highest posterior density (HPD) region of $\mathbf{Y}_f$ similar to the result of section 8.4 of Box and Tiao (1992).*

4 Concluding Remarks

The paper considers a multivariate conjugate prior as a mixture of matrix normal and Wishart distributions for the location and scale parameter matrices. All previous work in the area within the Bayesian framework are based on non-informative prior involving only the scale parameters. Thus in addition to dealing with the generalized multivariate regression model the paper assumes a wider set of matrix prior distributions for both the location and scale parameters. As such the results of the papers can be applied for any simple and multiple regression model as special cases. Also, the distributional assumption of the paper, namely the MEC family of distributions, covers a large number of popularly used multivariate distributions. In effect the results of this paper are applicable to any member of the elliptically contoured family of distributions including the normal and Student-t distributions for both vector and matrix variate cases.

It is interesting to note that the prediction distribution of the matrix of future responses, conditional on the matrix of realized responses, is a matrix variate T distribution. From the properties of the matrix-T distribution (see Khan 2002) the prediction distribution of any sub-matrix of future responses, conditional on the relevant sub-matrix of realized responses, also follows a matrix-T distribution. Likewise, prediction distribution of any row (or column) vector of the future response matrix will be a multivariate Student-t distribution. It is important to note that the prediction distribution of any row (or column) of the future response matrix is not independent of other rows (or columns). So the rows (or columns) of the future response matrix are dependent, even though they may be uncorrelated.

Acknowledgements

First author's research was supported by Shahrood University of Technology.

REFERENCES

1. Aitchison, J. and Dunsmore, I.R. (1975). "*Statistical Prediction Analysis*". Cambridge University Press, Cambridge.

2. Anderson, T. W. and Fang, K. T., (1990). Inference in multivariate elliptically contoured distribution based on maximum likelihood in "*Statistical Inference in Elliptically Contoured and Related Distribution*", (K. T. Fang and T. W. Anderson Ed.), 201-216. Allerton Press, New York.

3. Arashi, M. (2010). Idea of constructing an image of Bayes action, *Statistical Methodology*, **7**, 22-29.

4. Broemeling, L. D. (1985). *Bayesian Analysis of Linear Models*, Marcel Dekker, New York.

5. Box, G. E. P. and Tiao, G. C. (1992) *Bayesian Inference in Statistical Analysis*, John Wiley, New York.

6. Chib, S., Osiewalski, J. and Steel, M. F. J. (1991) Posterior inference on the degrees of freedom parameter in multivariate-t regression models, *Economics Lett.*, **37**(4), 391-397.

7. Dawid, A. P. (1977) Spherical matrix distributions and a multivariate model, *J. Royal Statist. Soc.*, **B 39**(2), 254-261.

8. Fang, K. T. and Li, R. (1999). Bayesian statistical inference on elliptical matrix distributions, *J. Mult. Anal.*, **70**(1), 66-85.

9. Geisser, S. (1993). "*Predictive Inference : An Introduction*". Chapman and Hall, London

10. Gupta, A. K. and Varga, T., (1995). Normal mixture representations of matrix variate elliptically contoured distributions, *Sankhyā*, **57**, 68-78.

11. Jammalamadaka, S. R., Tiwari, R. C. and Chib, S., (1987) Bayes prediction in the linear model with spherically symmetric errors, *Economics Lett.*, **24**(1), 39-44.

12. Khan, S. (2004). Predictive distribution of regression vector and residual sum of squares for normal multiple regression model. *Comm. In Statist.: Theory & Methods*, **33**(10), 2423-2443.

13. Khan, S. (2002). Distribution of sum of squares and products for the generalized multilinear matrix-T model. *J. of Mult. Anal.*, **83**, 124-140.

14. Khan, S. and Haq, M.S. (1994). Prediction distribution for the multilinear model with matrix-T error: The case of spherical error distribution. *J. of Applied Statistical Sciences*, **1**, 239-250.

15. Ng, V. M. (2010). On Bayesian inference with conjugate priors for scale mixtures of normal distributions, *Journal of Applied Probability and Statistics*, **5**, 69-76 Vee Ming Ng

16. Ng, V. M., (2002). Robust Bayesian Inference for Seemingly Unrelated Regressions with Elliptical Errors , *J. Mult. Anal.*, **82**, 409-414.

17. Osiewalski, J. and Steel, M. F. J. (1993). Robust Bayesian inference in elliptical regression models, *J. Econometrics*, **57**, 345-363.

18. Singh, R. K., Misra, S. and Pandey, S. K. (1995). A generalized class of estimators in linear regression models with multivariate-t distributed error, *Statist. Prob. Lett.*, **23**(2), 171-178.

19. Tsukuma, H. (2010). Shrinkage priors for Bayesian estimation of the mean matrix in an elliptically contoured distribution, *J. Mult. Anal.*, **101**, 1483-1492.

20. Vidal, I. and Arellano-Valle, R. B. (2010). Bayesian inference for dependent elliptical measurement error models, *J. Mult. Anal.*, **101**(10), 2587-2597.

21. Zellner, A. (1976). Bayesian and non-Bayesian Analysis of regression model with Multivariate Students t eorror terms, *J. Amer. Statist. Assoc.* **71**, 400-408.