Journal of Modern Applied Statistical Methods

Volume 7 | Issue 2 Article 19

11-1-2008

Robust Predictive Inference for Multivariate Linear Models with Elliptically Contoured Distribution Using Bayesian, Classical and Structural Approaches

B. M. Golam Kibria
Florida International University, kibriag@fiu.edu

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation
Kibria, B. M. Golam (2008) "Robust Predictive Inference for Multivariate Linear Models with Elliptically Contoured Distribution Using Bayesian, Classical and Structural Approaches," *Journal of Modern Applied Statistical Methods*: Vol. 7 : Iss. 2 , Article 19.
DOI: 10.22237/jmasm/1225513080

Journal of Modern Applied Statistical Methods
November 2008, Vol. 7, No. 2, 535-545
Copyright © 2008 JMASM, Inc.
1538 – 9472/08/$95.00

Robust Predictive Inference for Multivariate Linear Models with Elliptically Contoured Distribution Using Bayesian, Classical and Structural Approaches

B. M. Golam Kibria
Florida International University

Predictive distributions of future response and future regression matrices under multivariate elliptically contoured distributions are discussed. Under the elliptically contoured response assumptions, these are identical to those obtained under matric normal or matric-t errors using structural, Bayesian with improper prior, or classical approaches. This gives inference robustness with respect to departure from the reference case of independent sampling from the matric normal or matric t to multivariate elliptically contoured distributions. The importance of the predictive distribution for skewed elliptical models is indicated; the elliptically contoured distribution, as well as matric t distribution, have significant applications in statistical practices.

Key words: Bayesian; Classical; Elliptically Contoured Distribution; Matric Normal; Matric-t; Multivariate Linear Model; Predictive Distribution; Robustness; Structural.

Introduction

The predictive inference for multivariate regression models has been researched extensively. For example, Guttman & Hougarrd (1985) considered the classical approach, Geisser (1965) and Zellner & Chetty (1965), Kowalski, et al. (1999), Thabane (2000), Thabane and Haq (2003), and Kibria, et al. (2002) considered the Bayesian method, Fraser and Haq (1969) considered the structural approach and Haq (1982) considered the structural relation of the model approach. The predictive distributions have been derived under assumptions of multivariate normal errors, but the assumption of normality and independency for error variables may not be appropriate in many practical situations, especially when the underlying distributions have heavier tails. For such cases, multivariate t-errors with liner models have been considered by several researchers, for example: Zellner (1976), Gnanadesikan (1977), Sutradhar and Ali (1989) and Kibria and Haq (1998, 1999a). In the case of the multivariate linear model, matric-t error has been considered by Kibria and Haq (2002) and Kibria (2006).

Using the structural relation of the model, Haq (1982) derived the predictive distribution for future responses under the matric normal distribution. He obtained the predictive distributions as matric-t with appropriate degrees of freedom. Kibria and Haq (2000) considered the predictive inference for future responses under the matric-t errors and obtained the predictive distribution as a matric-t with appropriate degrees of freedoms. Therefore, the distribution of a future response matrix is not affected by a change in the error distribution from matric normal to matric-t. The invariance of the predictive distribution for the future response matrix suggests that the predictive distribution would be invariant to a wide class of error distributions. A broader assumption is

B. M. Golam Kibria is an Associate Professor in the Department of Mathematics and Statistics at the Florida International University. He is the overseas managing editor of the *Journal of Statistical Research*, coordinating editor for the *Journal of Probability and Statistical Science*. He is an elected fellow of the *Royal Statistical Society* and the *International Statistical Institute.* Email him at: kibriag@fiu.edu.

considered here: that error terms have a multivariate elliptically contoured distribution. The elliptically contoured distribution includes various distributions: the multivariate normal, matric-t, multivariate Student's t, and multivariate Cauchy (see Ng 2000). The class of of normal distribution mixtures is a subclass of the elliptical distributions as well as the class of spherically symmetric distributions (Fang, et al., 1990).

Elliptically contoured distributions have been discussed extensively for traditional multivariate regression models by Anderson and Fang (1990), Fang and Li (1999), Kubokawa and Srivastava (2001), and Arellano-Valle, et al. (2006). This distribution has also been considered by Chib, et al. (1988), Kibria and Haq (1999b), Kibria (2003), and Kibria and Nadarajah (2006) in the context of predictive inference for linear regression models. Ng (2000) considered the model under the multivariate elliptically error contoured distribution using both Bayesian and classical approaches: he obtained the same predictive distribution with both approaches.

This article reviews predictive distributions for future response and future regression matrices under multivariate elliptically contoured error distributions. When the errors of model 1 are assumed to have an elliptically contoured distribution, the prediction distribution of future response and regression matrices are also obtained as matric-t distributions under structural relation, Bayesian, and classical approaches. The assumptions of normality and matric-t are robust to deviations in the direction of elliptical distributions as far as inferences about the future regression matrix and prediction is concerned. The distribution is said to be robust if it remains the same under violations of the normality assumption.

Methodology

Consider a set of n responses from the following multivariate linear model:

$$Y = \beta X + \Gamma E, \tag{1}$$

where Y is an $m \times n$ matrix of observed responses, β is an $m \times p$ matrix of regression parameters, X is a $p \times n$ $(n \geq p)$ known design matrix, Γ is an $m \times m$ matrix of scale parameter with $\Gamma\Gamma' = \Sigma$, where $|\Gamma| > 0$ and E is an $m \times n$ random error matrix. If it is assumed that E has a spherically contoured distribution with the probability density function:

$$f(E) \propto g\{tr(EE')\}, \tag{2}$$

(Anderson & Fang, 1990), where $g\{.\}$ is a non-negative function over $m \times m$ positive definite matrices such that $f(E)$ is a density function, then the response variable Y has an elliptically contoured distribution. Here E' denotes the transpose of the matrix E, and $tr(M)$ denotes the trace of the matrix M. To derive the prediction distribution,

$$B_E = EX(XX')^{-1} \tag{3}$$

and

$$S_E = (E - B_E X)(E - B_E X)'$$

are defined as the regression matrix of E on X and the sum of squares and product (SSP) matrix respectively. Consider C_E to be a non-singular matrix such that the error SSP matrix, S_E can be expressed as $C_E C'_E = S_E$, and the standardized residual matrix is:

$$W_E = C_E^{-1}(E - B_E X). \tag{4}$$

It follows from (4) that

$$E = B_E X + C_E W_E, \tag{5}$$

and, because $W_E W'_E = I_m$:

$$EE' = B_E XX'B'_E + C_E C_{E'}. \tag{6}$$

Considering a set of n_f future responses from the multivariate linear model defined in (1) as

$$Y_f = \beta X_f + \Gamma E_f, \tag{7}$$

where Y_f and E_f are the $m \times n_f$ matrices of future responses and errors respectively, and X_f is an $p \times n_f$ ($n_f \geq p$) future design matrix. Assuming that E_f has the same distribution as E, then the joint distribution of E and E_f can be written as

$$f(E, E_f) \propto g\{tr(EE' + E_f E'_f)\}. \tag{8}$$

Defining the quantities in (3) to (6) in terms of future errors as follows:

$$B_{E_f} = E_f X_f (X_f X'_f)^{-1} \tag{9}$$

and

$$S_{E_f} = (E_f - B_{E_f} X_f)(E_f - B_{E_f} X_f)'$$

as the regression matrix of E_f on X_f and the sum of squares and product (SSP) matrix respectively. The standardized residual matrix and the future error matrix are respectively

$$W_{E_f} = C_{E_f}^{-1}(E_f - B_{E_f} X_f), \tag{10}$$

and

$$E_f = B_{E_f} X_f + C_{E_f} W'_{E_f}. \tag{11}$$

If $W_{E_f} W_{E_f}' = I_m$, then

$$E_f E_{f'} = B_{E_f} X_f X_{f'} B'_{E_f} + C_{E_f} C'_{E_f}, \tag{12}$$

where $S_{E_f} = C_{E_f}'$ are the SSP matrix for future error variables.

Derivation of Predictive Distributions: The Structural Relation Approach

Following Fraser and Ng (1980), the joint density function of error statistics B_E, S_E, and E_f for given data (D) is obtained as

$$p(B_E, S_E, E_f \mid D) \propto |S_E|^{\frac{n-m-p-1}{2}} g\left\{tr\left(B_E XX'B'_E + S_E + E_f E'_f\right)\right\}. \tag{13}$$

To obtain the desired predictive distribution, the following transformation is made:

$$\begin{aligned} R &= S_E^{-\frac{1}{2}}(E_f - B_E X_f\} \\ U &= B_E \\ V &= V. \end{aligned} \tag{14}$$

If the Jacobian of the transformation $J\{[E_f, B_E, S_E] \to [R, U, V]\}$ is equal to $|V|^{\frac{n_f}{2}}$, then the joint density of R, U, and V is

$$\begin{aligned} &p(R, U, V \mid D) \\ &\propto |V|^{\frac{n+n_f-m-p-1}{2}} g\left\{tr\left(UAU' + 2V^{\frac{1}{2}}RX'_f U' + V + V^{\frac{1}{2}}RR'V^{\frac{1}{2}}\right)\right\} \\ &\propto |V|^{\frac{n+n_f-m-p-1}{2}} g\left\{tr\left(tr(A^*) + tr(I_m + RHR')V\right)\right\}, \end{aligned} \tag{15}$$

where

$A^* = (U + V^{\frac{1}{2}}RX'_f A^{-1})A(U + V^{\frac{1}{2}}RX'_f A^{-1})'$,
$H = (I_f - X_f A^{-1} X'_f)$, and $A = XX' + X_f X'_f$ is a symmetric matrix.

Following Ng (2000) in assuming that $I_m + RHR'$ is positive definite and Q is a non-singular matrix such that $Q'Q = I_m + RHR'$. The following transformation may be made:

$$\begin{aligned} Y &= QVQ' \\ Z &= U + V^{\frac{1}{2}}RX'_f A^{-1}, \end{aligned} \tag{16}$$

the Jacobian of transformation is $|Q|^{-(m+1)}$, then the joint density function of R, Y and Z is as follows:

$$p(R,Y,Z \mid D) \propto |I_m + RHR'|^{-\frac{n+n_f-m}{2}} |Y|^{\frac{n+n_f-m-p-1}{2}} g\{tr(Y)+tr(ZAZ'\} \tag{17}$$

Integrating (17) with respect to Y and Z yields the density function of R as:

$$\begin{aligned} p(R \mid D) &\propto \iint p(R,Y,Z \mid D)dYdZ \\ &\propto |I_m + RHR|^{-\frac{n+n_f-p}{2}} . \end{aligned} \tag{18}$$

It may then be shown that:

$$\begin{aligned} R &= S_E^{-\frac{1}{2}}(E_f - B_E X_f\} \\ &= S_Y^{-\frac{1}{2}}(E_Y - B_Y X_f\}, \end{aligned} \tag{19}$$

where B_Y is the regression matrix of Y on X and $S_Y = (Y-B_Y)(Y-B_Y)'$ is the Wishart matrix. Thus, the prediction distribution of Y_f can be obtained from (18) and (19) as follows:

$$p(Y_f \mid D) \propto |I_m + S_Y^{-1}(Y_f - B_Y X_f)(I_{n_f} - X'_f A^{-1} X_f)(Y_f - B_Y X_f)'|^{\frac{n+n_f-m}{2}}, \tag{20}$$

which is a Matric-t density. The predictive distribution of the future responses for given data is an $m \times n_f$ dimensional matric-t distribution with $(n-p-m+1)$ degrees of freedom. The location parameter in the predictive density of Y_f is $B_Y X_f$ and the scale parameter matrix is $I_{n_f} - X'_f A^{-1} X_f$. This result coincides with that of Haq (1982), where he considered matric normal, and that of Kibria and Haq (2000) who considered the matric T error distribution. Thus, the predictive distribution of future responses are unaffected by departures from normality or dependent but uncorrelated assumptions to an elliptically contoured distribution. The shape parameter of the predictive distribution does not depend on the unknown parameter, instead, it depends on the sample observation and the dimension of the regression matrix.

Derivation of Predictive Distributions:

The Bayesian Approach

The density of $Y \mid \Sigma$ is given as

$$f(Y \mid \Sigma) \propto |\Sigma|^{-\frac{n}{2}} g\{tr(\Sigma^{-1}(Y-BX)(Y-BX)')\}, \tag{21}$$

Following Ng (2000), the Bayesian predictive distribution for future responses is obtained as follows. Suppose Y_f is an unobserved $m \times n_f$ of future observations, then the density function of (Y, Y_f) is given by:

$$\begin{aligned} &f(Y, Y_f \mid B, \Sigma) \propto \\ &|\Sigma|^{-\frac{n+n_f}{2}} g\{tr(\Sigma^{-1}[(Y-BX)(Y-BX)') \\ &+(Y_f - BX_f)(Y_f - BX_f)')]\}. \end{aligned} \tag{22}$$

The Bayesian predictive density of Y_f for given Y is defined as:

$$f(Y_f \mid Y) \propto \iint f(Y, Y_f \mid B, \Sigma) p(B, \Sigma^{-1}) dBd\Sigma^{-1}, \tag{23}$$

where $p(B, \Sigma^{-1})$ is the non-informative prior density function of (B, Σ^{-1}) and is,

$$p(B, \Sigma^{-1}) \propto |\Sigma^{-1}|^{-\frac{m+1}{2}} . \tag{24}$$

The predictive density is obtained as

$$
\begin{aligned}
f(Y_f \mid Y) \quad &\propto \quad \iint |\Sigma|^{-\frac{n+n_f-m-1}{2}} \\
&\times \quad g\{tr(\Sigma^{-1}[(Y-BX)(Y-BX)') \\
&+(Y_f - BX_f)(Y_f - BX_f)')]\} dBd\Sigma^{-1}.
\end{aligned}
\tag{25}
$$

And the matrix expression in (25) can be rewritten as:

$$
\begin{aligned}
&(Y-BX)(Y-BX)')(Y_f - BX_f)(Y_f - BX_f)' = \\
&S_Y + (Y_f - \hat{B}X_f)H(Y_f - \hat{B}X_f)' \\
&+(B-B^*)A(B-B^*)'
\end{aligned}
\tag{26}
$$

where $B^* = (YX' + Y_f X_f)A^{-1}$. The matrices A and H are defined under equation (15). From the following transformation,

$$
\begin{aligned}
D &= \Sigma^{-\frac{1}{2}}(B-B^*) \\
G &= K\Sigma^{-1}K'
\end{aligned}
\tag{27}
$$

where $KK' = S_Y + (Y_f - \hat{B}X_f)(Y_f - \hat{B}X_f)'$ and the Jacobian of the transformation $J[(B,\Sigma^{-1}) \to (D,G)$ is equal to $|G|^{-\frac{p}{2}}|K'K|^{-\frac{m-p+1}{2}}$, then (25) becomes

$$
\begin{aligned}
f(Y \mid Y_f) \quad &\propto \quad \iint |S_Y + (Y_f - \hat{B}X_f)H(Y_f - \hat{B}X_f)'|^{-\frac{n+n_f-k}{2}} \\
&|G|^{-\frac{n+n_f-m-p-1}{2}} \quad g\{tr(G) + tr(DAD') + \} dDdG \\
&\propto \quad |I_m + S_Y^{-1}(Y_f - B_Y X_f)(I_{n_f} - X'_f A^{-1} X_f)(Y_f - B_Y X_f)'|^{\frac{n+n_f-m}{2}}.
\end{aligned}
\tag{28}
$$

Hence Y_f has a matric-t distribution with $n_f - m - p + 1$ degrees of freedom. Thus, the predictive distribution under the structural relation and the Bayesian approaches are the same.

Derivation of Predictive Distributions:
The Classical Approach

To obtain the predictive density of Y_f, it follows from Ng (2000) that $R = S_Y^{-\frac{1}{2}}(Y_f - \hat{B}X_f)$ is the studentized variable, and $S_Y^{-\frac{1}{2}}$ is the symmetric square root of S_Y^{-1}. Since R is invariant under the transformations $Y \to BX + CY$, $Y_f \to BX_f + CY_f$, for any non-singular square matrix C, it can be assumed, without loss of generality, that $B = 0$ and $\Sigma = I_m$ to derive the predictive distribution of Y_f. With this assumption, the joint density function of (Y, Y_f) becomes

$$
f(Y, Y_f) \quad \propto \quad g\{tr(YY' + Y_f Y'_f)\} \tag{29}
$$

Because $YY' = S_Y + \hat{B}XX'\hat{B}'$ and, using the invariant differential in Fraser and Ng (1980), the joint density function of $\hat{B}_Y$, S_Y and Y_f is obtained from (29) as:

$$
f(\hat{B}_Y, S_Y, Y_f) \propto |S_Y|^{-\frac{n-p-k-1}{2}} \; g\{tr(S_Y + \hat{B}_Y XX'\hat{B}'_Y + Y_f Y'_f)\}
\tag{30}
$$

Making the transformation $R = S_Y^{-\frac{1}{2}}(Y_f - \hat{B}_Y X_f)$, followed by the Jacobian of the transformation is $|S_Y|^{\frac{n_f}{2}}$, the joint density of $\hat{B}_Y$, S_Y, R is:

$$
\begin{aligned}
&f(\hat{B}_Y, S_Y, R) \propto \\
&|S_Y|^{-\frac{n+n_f-p-k}{2}} \quad g\{tr(S_Y + \hat{B}_Y XX'\hat{B}'_y \\
&+(S_Y^{\frac{1}{2}}R + \hat{B}_Y X_f)(S_Y^{\frac{1}{2}}R + \hat{B}_Y X_f)')\}
\end{aligned}
\tag{31}
$$

The matrix expression in (31) can be rewritten as:

$$
\begin{aligned}
&S_Y + \hat{B}_Y XX'\hat{B}_Y \\
&+(S_Y^{\frac{1}{2}}R + \hat{B}_Y X_f)(S_Y^{\frac{1}{2}}R + \hat{B}_Y X_f)' = \\
&(I_m + RHR')S_Y \\
&+tr(\hat{B}_Y + S_Y^{\frac{1}{2}}RX'_f A^{-1})A(\hat{B}_Y + S_Y^{\frac{1}{2}}RX'_f A^{-1})'.
\end{aligned}
\tag{32}
$$

Making the following transformation

$$Y = QVQ'$$
$$Z = U + V^{\frac{1}{2}} W X'_f A^{-1}, \quad (33)$$

and following procedures similar to the Bayesian Approach, the the joint density function of R, Y and Z is obtained as follows:

$$p(R,Y,Z \mid D) \propto$$
$$|I_m + RHR'|^{\frac{n+n_f-m}{2}} |Y|^{\frac{n+n_f-m-p-1}{2}} g\{tr(Y) + tr(ZAZ')\} \quad (34)$$

Integrating (34) with respect to Y and Z yields the density function of Y_f as:

$$p(Y_f \mid D) \propto$$
$$|I_m + S_Y^{-1}(Y_f - B_Y X_f)(I_{n_f} - X_{f'} A^{-1} X_f)(Y_f - B_Y X_f)'|^{\frac{n+n_f-m}{2}}, \quad (35)$$

which is a Matric-t density. The predictive distribution of the future responses for given data is an $m \times n_f$ dimensional matric-t distribution with $(n-p-m+1)$ degrees of freedom. Thus, the predictive distribution under the structural relation, Bayesian and classical approaches are the same.

Predictive Distribution of Future Regression Matrix

Based on the results in Kibria (2006), the joint density function of error statistics B_E, S_E, B_{E_f} and S_{E_f} are obtained as:

$$p(B_E, S_E, B_{E_f}, S_{E_f} \mid E, X, X_f) \propto$$
$$|S_E|^{\frac{n-m-p-1}{2}} |S_{E_f}|^{\frac{n_f-m-p-1}{2}}$$
$$\times \; g\left\{tr\left(B_E XX'B'_E + S_E + B_{E_f} X_f X_{f'} B_{E_f}' + S_{E_f}\right)\right\}. \quad (36)$$

The structural relation of model (1) yields

$$B_E = \Sigma^{-\frac{1}{2}}(B_Y - \beta) \quad and \quad S_E = \Sigma^{-1} S_Y, \quad (37)$$

and the Jacobian of the transformation $J\{[B_E, S_E] \to [\beta, \Sigma]\}$ is equal to $|S_Y|^{\frac{m+1}{2}} |\Sigma|^{-\left(\frac{p}{2}+m+1\right)}$. Thus, the joint density of β, Σ, B_{E_f}, and S_{E_f} is obtained as:

$$p(\beta, \Sigma, B_{E_f}, S_{E_f} \mid E, X, X_f) \propto$$
$$|S_{E_f}|^{\frac{n_f-m-p-1}{2}} |\Sigma|^{-\frac{n+m+1}{2}} g\{tr\Sigma^{-1}((B-\beta)XX'(B-\beta)'$$
$$+ \; S + B_{E_f} X_f X_{f'} B_{E_f}' + S_{E_f})\}, \quad (38)$$

where $B_Y = B$ and $S_Y = S$ for notational convenience. Similarly, the structural relation of the model (7) yields

$$B_{E_f} = \Sigma^{-\frac{1}{2}}(B_{Y_f} - \beta)$$

and

$$S_{E_f} = \Sigma^{-1} S_{Y_f}, \quad (39)$$

where B_{Y_f} is the regression matrix for the future model, and S_{Y_f} is the Wishart matrix for the future responses. If the Jacobian of the transformation $J\{[B_{E_f}, S_{E_f}] \to [B_f, S_f]\}$ is equal to $|\Sigma|^{-\frac{p+m+1}{2}}$, then the joint density function of β, Σ, B_f, and S_f is obtained as

$$p(\beta, \Sigma, B_f, S_f \mid Y, X, X_f) \propto$$
$$|S_f|^{\frac{n_f-m-p-1}{2}} |\Sigma|^{-\frac{n+n_f+m+1}{2}}$$
$$g\left\{tr\left(\Sigma^{-1}\left[(B-\beta)XX'(B-\beta)'\right.\right.\right.$$
$$\left.\left.\left. + S + (B_f - \beta) X_f X_{f'} (B_f - \beta)' + S_f\right]\right)\right\},$$

(40)

where $B_{Y_f} = B_f$ and $S_{Y_f} = S_f$.

The marginal density function of β, B_f and S_f is obtained from (40) as

$$p(\beta, B_f, S_f \mid Y, X, X_f) \propto |S_f|^{\frac{n_f - m - p - 1}{2}} \int_\Sigma |\Sigma|^{-\frac{n + n_f + m + 1}{2}} g\left\{tr\left(\Sigma^{-1}\left[(B-\beta)XX'(B-\beta)' + S + (B_f-\beta)X_f X'_f (B_f-\beta)' + S_f\right]\right)\right\} d\Sigma. \tag{41}$$

To evaluate the integral in (41), let $\Sigma^{-1} = \Lambda$, then

$$d\Sigma = |\Lambda|^{-(m+1)}\, d\Lambda,$$

therefore,

$$p(\beta, B_f, S_f \mid Y, X, X_f) \propto |S_f|^{\frac{n_f - m - p - 1}{2}} \int_\Lambda |\Lambda|^{\frac{n + n_f - m - 1}{2}} g\left\{tr\left(\Lambda\left[(B-\beta)XX'(B-\beta)' + S + (B_f-\beta)X_f X'_f (B_f-\beta)' + S_f\right]\right)\right\} d\Lambda, \tag{42}$$

Following Ng (2002), consider G to be a nonsingular matrix of order m such that

$$G^T G = \begin{bmatrix} (B-\beta)XX'(B-\beta)' + S \\ +(B_f-\beta)X_f X_{f'}(B_f-\beta)' + S_f \end{bmatrix}.$$

The transformation, $W = G\Lambda G^T$ has the Jacobian of the transformation as $|G^T G|^{-\frac{m+1}{2}}$, and integrating the above with respect to W yields the marginal density of β, B_f and S_f as,

$$p(\beta, B_f, S_f \mid Y, X, X_f) \propto |S_f|^{\frac{n_f - m - p - 1}{2}} \begin{bmatrix} (B-\beta)XX'(B-\beta)' + S \\ +(B_f-\beta)X_f X_{f'}(B_f-\beta)' + S_f \end{bmatrix}^{-\frac{n+n_f}{2}} \int_\Sigma g\{tr(W)\} |W|^{\frac{n + n_f - m - 1}{2}} dW$$

$$\propto |S_f|^{\frac{n_f - m - p - 1}{2}} \times \begin{bmatrix} (B-\beta)XX'(B-\beta)' + S \\ +(B_f-\beta)X_f X'_f(B_f-\beta)' + S_f \end{bmatrix}^{-\frac{n+n_f}{2}}. \tag{43}$$

The density function in (43) can further be expressed as

$$p(\beta, B_f, S_f \mid Y, X, X_f) \propto |S_f|^{\frac{n_f - m - p - 1}{2}} \left[(\beta - FA^{-1})A(\beta - FA^{-1})' + S + (B_f - B)H^{-1}(B_f - B)' + S_f\right]^{-\frac{n+n_f}{2}}, \tag{44}$$

where $F = BXX' + B_f X_f X'_f$, $A = XX' + X_f X'_f$ and $H = [XX']^{-1} + [X_f X'_f]^{-1}$.

The marginal density function of B_f and S_f are obtained by integrating β using matric-t argument (Press, 1982) from (44) as

$$
\begin{aligned}
p(B_f, S_f \mid Y, X, X_f) &\propto \int_\beta p(\beta, B_f, S_f \mid D) d\beta \\
&\propto |S_f|^{\frac{n_f - m - p - 1}{2}} \int_\beta \big[(\beta - FA^{-1})A(\beta - FA^{-1})' \\
&+ S + (B_f - B)H^{-1}(B_f - B)' + S_f\big]^{-\frac{n+n_f}{2}} d\beta \\
&\propto |S_f|^{\frac{n_f - m - p - 1}{2}} \big[S + (B_f - B)H^{-1}(B_f - B)' + S_f\big]^{-\frac{n+n_f - p}{2}} .
\end{aligned} \tag{45}
$$

Finally, the predictive distribution of the future regression matrix B_f is obtained as

$$
\begin{aligned}
p(B_f \mid Y, X, X_f) &\propto \\
\int_{S_f} |S_f|^{\frac{n_f - m - p - 1}{2}} &\left[S + (B_f - B)H^{-1}(B_f - B)' + S_f\right]^{-\frac{n+n_f - p}{2}} dS_f \\
&= \frac{\Gamma_m\left(\frac{n}{2}\right)|H|^{\frac{m}{2}}}{\pi^{\frac{mp}{2}}\Gamma_m\left(\frac{n-p}{2}\right)} |S|^{-\frac{p}{2}} |I_m + S^{-1}(B_f - B)H^{-1}(B_f - B)'|^{-\frac{n}{2}},
\end{aligned} \tag{46}
$$

which is a Matric-t density. Thus the predictive distribution of the future regression matrix for given data is an $m \times p$ dimensional matric-t distribution with $(n - p - m + 1)$ degrees of freedom. That is

$$B_f \sim t_{m \times n_f}(B, H, S_Y, n - p - m + 1).$$

The predictive distribution of B_f is identical to that obtained under the assumption of matric normal error (Haq 1982). Thus, the predictive distribution of the future regression matrix is unaffected by departures from normality, or are dependent but uncorrelated assumptions to the elliptically contoured distribution. It may be concluded that the predictive distributions of a future regression matrix under structural, Bayesian and classical approaches are the same.

Conclusions

The predictive distribution of future responses for observed information under assumptions of multivariate elliptically contoured error distributions were considered, and the structural, Bayesian and classical approaches all resulted in the same predictive distributions. The predictive distributions under the elliptical errors assumption are identical to those obtained under independent normal errors or matric-t errors, thus showing robustness with respect to departure from the reference case of independent sampling from the matric normal or dependent, but uncorrelated sampling from matric-t distributions to elliptically contoured distributions. In the Classical approach, mild restictions were adopted, whereas the structural relation did not need those restrictions. The predictive distribution of the future regression matrix was also obtained as matric t. When $n_f = 1$, the predictive distribution of a single future response from a multivariate elliptically contoured distribution is obtained as a multivariate t distribution with $n - p - m + 1$ degrees of freedom. Findings in this article are more general, and include a linear model as a special case, as well as a variety of symmetric distributions. It is also noted that using the predictive distribution one can construct the β expectation tolerance regions for future response(s). In both application and theoretical aspects, these findings have potential applications in many areas of statistics.

There is great interest in the statistical literature toward robust statistical methods to represent strongly asymmetric data as adequately as possible and, at the same time, reduce the unrealistic ordinary normal or Student t assumptions. In scientific fields, such as gold concentration in soil samples (Galea-Rojas, et al., 2003), arsenate in water samples (Ripley & Thomson, 1987), cholesterol in blood samples (Lachos & Bolfarine, 2007) and many other situations, the data follow asymmetric distributions.

In such cases, normal or t distributions do not work well. Instead, certain types of skewed distributions are proposed in the literature to study the skewed data. These distributions allow for skewness and contain the normal or t distribution as a proper member or as a limiting case. Various kinds of skew distributions exist in the literature: skew-symmetric distributions (Gomez, et al., 2007), skew normal distribution (Azzalini, 1985, 1986), multivariate skew normal (Azzalini & Dalla Valle, 1996; Azzalini & Capitanio, 1999; Gupta, et al., 2004), skew t distribution (Jones & Faddy, 2003), generalized skew-t distribution (Theodossion, 1998), skew multivariate t (Azzalini & Capitanio, 2003; Gupta 2003), skew elliptical distribution (Branco & Dey, 2001; Dey & Liu, 2005; Fang 2003, 2005a, 2005b; Sahu & Chai, 2005), generalized skew elliptical distribution (Genton & Loperfido, 2005). The location and scale parameters of skewed elliptical distributions control the skewness and maintains the symmetry of the elliptical distributions.

They also provide an opportunity to study the robustness of normal theory procedures when both skewness and kurtosis are different from the normal. The skewed elliptical distributions are more useful to fit real data (Arnold & Beaver, 2000). Genton and Genton (2004) give an excellent review about skew-elliptical distributions and provide many new developments, including theoretical results and applications of skewed-elliptical distributions with real life data. Regression analysis with skewed elliptical distributions have been considered by Sahu, et al., (2003), for example. Unfortunately, predictive inferences with skewed elliptical models are limited or not available in the literature. It is necessary and to derive the predictive distribution when the error of the model follows the skewed elliptical distribution.

References

Anderson, T. W., & 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 & T. W. Andeson (*Eds.*), 201-216. NY: Allerton Press.

Arellano-Valle, R. B., Pino, G., & Iglesias, P. (2006). Bayesian inference in spherical linear models: robustness and conjugate analysis. *Journal of Multivariate Analysis*, *97*, 179-197.

Arnold, B. C., & Beaver, R. J. (2000). The skew-Cauchy distribution. *Statistical Probabilty. Letters*, *49*, 285-290.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand.J.Statist.* 12, 171-178.

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica* 46, 199-208

Azzalini, A., & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society*, *61*(B), 579-602.

Azzalini, A., & Capitanio, A. (2003). Distributions generate by perturbation of symmetry with emphasis on a multivariate skew t distribution. *Journal of the Royal Statistical Society*, *65*(B), 367-389.

Azzalini, A., & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, *83*, 715-726.

Branco, M. D., & Dey, D. K., (2001.) A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, *79*, 99-113.

Chib, S., Tiwari, R. C., & Jammalamadaka, S. R. (1988). Bayes prediction in regressions with elliptical errors. *Journal of Econometrics*, *38*, 349-360.

Dey, D. K. & Liu, J. (2005). A new construction for skew multivariate distributions. *Journal of Multivariate Analysis*, *95*, 323-344.

Fang, B. Q. (2003). The skew elliptical distributions and their quadratic forms. *Journal of Multivariate Analysis*, *87*, 298-314.

Fang, B. Q. (2005a). Noncentral quadratic forms of the skew elliptical variables. *Journal of Multivariate Analysis*, *95*, 410-430.

Fang, B. Q. (2005b). Invariant distribution of the multivariate tests in the skew elliptical model. *Statistical Methodology, 2*, 285-296.

Fang, K-T, Kotz, S., & Ng, V. M. (1990). *Symmetric Multivariate and related distributions*. London: Chapman and Hall.

Fang, K-T, & Li, R. (1999). Bayesian statistical inference on elliptical matrix distributions. *Journal of Multivariate Analysis*, *70*, 66-85.

Fraser, D. A. S., & Haq, M. S. (1969). Structural probability and prediction for the multivariate model. *Journal of the Royal Statistical Society*, *31*(B), 317-331.

Fraser, D. A. S., & K. W. Ng (1980). Multivariate regression analysis with spherical error. In *Multivariate analysis*, V. P. R. Krishnaiah (*Ed.*), 369-386. NY: North-Holland Publishing Co.

Galea-Rojas M., de Castilho M., Bolfariane, H., & De Castro M. (2003). Detection of analytical bias. *Analyst*, *128*, 1073-1081.

Geisser, S. (1965). Bayesian estimation in multivariate analysis. *Annals of Mathematical Statistics*, *36*, 150-159.

Genton, G. G. and and Genton, M. G. G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. NY: Chapman and Hall.

Genton, M. G., & Loperfido, N. M. R. (2005). Generalized skew-elliptical distributions and their quadratic forms. *Annals of the Institute of Statistical Mathematics*, 57 (*2*), 389-401.

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations.* NY: Wiley.

Gomez, H. W., Venegas, O., & Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics*, *18*, 395-407.

Gupta, A. K. (2003). Multivariate Skew t-Distribution. *Statistics*, *37*, 359-363.

Gupta, A. K., Gonzalez-Fariars, G., & Dominguez-Molina, J. A. (2004). A multivariate skew normal distribution. *Journal of Multivariate Analysis*, *89*, 181-190.

Guttman, I., & Hougaard, P. (1985). Studentization and prediction problems in multivariate multiple regression. *Communications in Statistics-Theory and Methods*, *14*, 1251-1258.

Haq, M. S. (1982). Structural relationships and prediction for the multivariate models. *Statistische Hifte*, *23*, 218-227.

Jones, M. C., & Faddy, M. J. (2003). A skew extension of the t-distribution, with application. *Journal of the Royal Statistical Society*, *65*(B), 159-174.

Kibria, B. M. G. (2003). Robust predictive inference for the multivariate linear models with elliptically contoured distribution. *Far East Journal of Theoretical Statistics*, *10*(1), 11-24.

Kibria, B. M. G. (2006). The matric t distribution and its applications in predictive inference. *Journal of Multivariate Analysis*, *97*, 785-795.

Kibria, B. M. G., & Haq, M. S. (1998). Marginal likelihood and prediction for the multivariate ARMA(1,1) linear models with multivariate t error distribution. *Journal of Statistical Research*, *32*(1), 71-80.

Kibria, B. M. G., & Haq, M. S. (1999a). The multivariate linear model with multivariate t and intra-class covariance structure, *Statistical Paper*, *40*, 263-276.

Kibria, B. M. G., & Haq, M. S. (1999b). Predictive inference for the elliptical linear model. *Journal of Multivariate Analysis*, *68*, 235-249.

Kibria, B. M. G., & Haq, M. S. (2000). The multivariate linear model with matric T error variables. *Journal of Applied Statistical Sciences*, *4*, 277-290.

Kibria, B. M. G., & Nadarajah, S. (2006). The predictive distribution for the heteroscedastic multivariate linear model with elliptically contoured distributions. *Journal of Statistical Research*, *40*(1), 35-43.

Kibria, B. M. G., Sun, L., Zidek, J. V., & Nhu, L. (2002). Bayesian spatial prediction of random space-time fields with application to mapping $PM_{2.5}$ exposure. *Journal of the American Statistical Association*, *97*, 112-124.

Kowalski, J., Mendoza-Blanco, J. R., Tu, X. M., & Gleser, E. J. (1999). On the difference in inference and prediction between the joint and independent t-error models for seemingly unrelated regression. *Communications in Statistics- Theory and Methods*, *29*, 2119-2140.

Kubokawa, T., & Srivastava, M. S. (2001). Robust improvement in estimation of a mean matrix in an elliptically contoured distribution. *Journal of Multivariate Analysis*, *76*, 138-152.

Lachos, V.H. and Bolfarine H. (2007). Skew-Probit Measurement Error Models. *Statistical Methodology*, 4, 1-12.

Ng, V. M. (2000). A note on predictive inference for multivariate elliptically contoured distributions. *Communications in Statistics-Theory and Methods*, *29*, 477-2000.

Ng, V. M. (2002). Robust Bayesian Inference for seemingly unrelated regression with elliptical errors. *Journal of Multivariate Analysis*, *80*, 122-135.

Press, J. (1982). *Applied multivariate analysis: Using bayesian and frequentist methods of inference*. FL: Robert E. Krieger Publishing Company.

Ripley, B., & Thompson, M. (1987). Regression techniques for detection of analytic bia. *Analyst*, *122*, 377-383.

Sahu, S. K., Dey, D. K., & Branco, M. D. (2003). A new class of multivariate skew distributions with applications to bayesian regression models. *The Canadian Journal of Statistics*, *31*(2), 129-150.

Sahu, S. K., & Chai, H. S. (2005) A new skew-elliptical distribution and its properties. S3RI Methodology Working Papers, M05/19). Southampton, UK, Southampton Statistical Sciences Research Institute, 33pp. http://eprints.soton.ac.uk/18180/.

Sutradhar, B. C., & M. M. Ali (1988). A generalization of the Wishart distribution for the elliptical model and its moments for the multivariate *t* model. *Journal of Multivariate Analysis*, *29*, 155-162.

Thabane, L. (2000). *Contributions to bayesian statistical inference*. Unpublished Ph. D. Thesis. Department of Statistical and Actuarial Sciences. London, Ontario, Canada.

Thabane, L. and Haq, M. S. (2003). The generalized multivariate modified bessel distributionsand its Bayesian applications. *Journal of Applied Statistical Science*, 11(3), 255-267.

Theodossion, P. (1998). Financial data and the skewed generalized *t* distribution. *Management Science*, *44*, 1650-1661.

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*, 400-405.

Zellner, A., & Chetty, V. K. (1965). Prediction and decision problems in regression models from the Bayesian point of view. *Journal of the American Statistical Association*, *60*, 608-616.