THE PREDICTIVE DISTRIBUTION FOR THE HETEROSCEDASTIC MULTIVARIATE LINEAR MODELS WITH ELLIPTICALLY CONTOURED ERROR DISTRIBUTIONS

B. M. Golam Kibria

Department of Statistics, Florida International University
Miami, Florida 33199, USA
Email: kibriag@fiu.edu

Saralees Nadarajah

Department of Statistics, University of Nebraska
Lincoln, Nebraska 68583, USA
Email: snadaraj@unlserve.unl.edu

SUMMARY

This paper considers the heteroscedastic multivariate linear model with errors following elliptically contoured distributions. The marginal likelihood function of the unknown covariance parameters and the predictive distribution of future responses have been derived. The predictive distribution obtained is a product of multivariate Student’s $t$ distributions. It is interesting to note that when the models are assumed to have elliptically contoured distributions the marginal likelihood function of the parameters as well as the predictive distribution are identical to those obtained under independently distributed normal errors or dependent but uncorrelated Student’s $t$ errors. Therefore, the distribution of future responses is unaffected by a change in the error distribution from the multivariate normal and multivariate $t$ distributions to elliptically contoured distributions. 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 multivariate $t$ distributions to elliptically contoured distributions.

Keywords and phrases: Elliptically Contoured Distributions; Marginal likelihood; Multivariate model; Predictive distribution; Robustness

1 Introduction

The predictive inference for the linear models has been considered by various researchers: Goldberger (1962) and Hahn (1972) used the classical approach; Tiao and Zellner (1964), Geisser (1965), Zellner and Chetty (1965) and Kibria et al. (2002) used the Bayesian approach, Fraser and Haq (1969) used the structural approach, while Haq and Kibria (1997), Kibria and Haq (1998, 1999) and very recently Kibria (2006) used the structural relation of the model for the derivation of the predictive distribution. The error terms in linear models are assumed to be normally and independently distributed in most applied as well as theoretical research work. However, such assumptions may not be appropriate in many practical situations (for examples, see Zellner (1976), Gnanadesikan (1977), Kibria (1996) and Kibria and Haq (1998)). It happens particularly if the error distribution has heavier tails. One can tackle such situations by using the well known $t$ distribution as it has heavier tails than the normal distribution, especially for smaller degrees of freedom (e.g. Fama (1965) and Blatberg and Gonedes (1974)).

The literature on the predictive inference for the heteroscedastic multivariate linear model is limited. Kibria (1999) considered the predictive inference for the heteroscedastic multivariate linear model under the normality assumption and obtained the predictive distribution as a product of $m$ multivariate Student’s $t$ distributions. Kibria (2002) considered the predictive inference for future responses under the multivariate $t$ errors and obtained the predictive distribution as a product of $m$ multivariate Student’s $t$ distributions. Therefore, the distribution of future responses for a heteroscedastic model is unaffected by a change in the error distribution from the multivariate normal to the multivariate $t$ distribution. The invariance of the predictive distribution for the future responses suggests that the predictive distribution would be invariant to a wide class of error distributions. In this paper a very general assumption is employed, namely that error terms have a multivariate elliptically contoured distribution. The class of elliptically contoured distributions includes various distributions: the multivariate normal, matrix $t$, multivariate Student’s $t$, multivariate Kotz type and multivariate Cauchy. The class of mixtures of normal distributions is a subclass of elliptical distributions as well as the class of spherically symmetric distributions (Fang et al., 1990).

Elliptically contoured distributions for traditional multivariate regression models have been discussed extensively by Anderson and Fang (1990) and Kubokawa and Srivastava (2003) among others. These distributions have also been considered by Chib et al. (1988), Kibria and Haq (1999) and Kibria (2003) in the context of predictive inference for the general linear or multivariate linear model but not for the heteroscedastic multivariate model. In this paper, we show that when the errors of the model (2.1) are assumed to have an elliptically contoured distribution, the prediction distribution of future responses is a product of $m$ multivariate Student’s $t$ distributions. Therefore, the assumption of normality as well as multivariate $t$ is robust to the deviation in the direction of elliptical distributions as far as predictive inference is concerned.

The organization of this paper is as follows: The multivariate linear model and the
The Predictive Distribution for . . .

marginal likelihood function of the unknown covariance matrix are discussed in section 2. The predictive distribution for the future responses are derived in section 3. Some concluding remarks are given in section 4.

2 Multivariate Linear Model and Marginal Likelihood Function

Consider the following multivariate linear model

$$Y = BX + \Omega \Gamma,$$

(2.1)

where $Y$ and $X$ are the $m \times n$ and $p \times n$ ($n \geq p$) response and regressor matrices, respectively, $B$ is an $m \times p$ matrix of regression parameters, $E$ is an $m \times n$ errors matrix, and $\Gamma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_m)$ is an $m \times m$ scale parameter matrix. Here each of the $m$ rows of $Y$ may be viewed as a realization of the linear model. When the diagonal elements of $\Gamma$ are not equal, the model (2.1) is known as heteroscedastic multivariate linear model. Assume that the covariance matrix of each row of $E$ is an $n \times n$ matrix $\Omega \Lambda$, where $\Lambda$ is a function of a set of unknown parameters. Then the covariance matrix of $E$ is $\Omega \Lambda \otimes I_m$, where $\otimes$ is the kronecker product between two matrices and $I_m$ is the identity matrix of order $m$. The covariance matrix of $Y$ is $\Omega \Lambda \otimes \Sigma$, where $\Sigma = \Gamma \Gamma'$. The application of model (2.1) has been discussed by Parthasaradhi (1972), Delgado (1992) and Kibria (1999); to mention a few.

We assume that $E$ has an elliptically contoured distribution with the probability density function

$$f(E|\Omega \Lambda) \propto |\Omega \Lambda|^{-\frac{m}{2}} g\{\text{tr}(E \Omega \Lambda^{-1} E')\},$$

(2.2)

which is of the form given in Anderson and Fang (1990), where $g\{\cdot\}$ is a non-negative function over $m \times m$ positive definite matrices such that $f(E)$ is a density function. Here, $E'$ denotes the transpose of the matrix $E$ and $\text{tr}(M)$ denotes the trace of the matrix $M$.

To derive the marginal likelihood function of $\Omega \Lambda$, we define the following:

$$\hat{B}_E = EX'(XX')^{-1},$$

$$S_E = \text{diag}(E - \hat{B}_E X)(E - \hat{B}_E X)' = \text{diag}(s_{e1}^2, s_{e2}^2, \ldots, s_{em}^2),$$

$$Z_E = C_E^{-1}\{E - \hat{B}_E X\},$$

(2.3)

where $C_E = \text{diag}(s_{e1}, \ldots, s_{em})$. The corresponding expressions for the response matrix $Y$ will be denoted by $\hat{B}_Y$, $S_Y$ and $Z_Y$, respectively. From (2.1) and (2.3) we obtain

$$\Gamma = C_Y C_E^{-1},$$

$$B = \hat{B}_Y - C_Y C_E^{-1} \hat{B}_E,$$

$$Z_E = Z_Y,$$

(2.4)
The $i^{th}$ row vector of the matrix $E$ can be written as 

$$e_i = b_{ei}X + s_{ei}z_{ei}, \quad i = 1, 2, \ldots, m,$$  

(2.5)

where $b_{ei}$ is the $i^{th}$ row vector of $\hat{B}_E$, $z_{ei}$ is the $i^{th}$ row vector of $Z_E$ and $s_{ei}$ is the $i^{th}$ diagonal element of $C_E$. Then the relationship between the volume elements of $e_i$ can be expressed in terms of new variables $b_{ei}$, $s_{ei}$ and $z_{ei}$ as

$$de_i = |XX'|^{\frac{1}{2}}db_{ei}s_{ei}^{m-p-1}ds_{ei}dz_{ei}, \quad i = 1, 2, \ldots, m,$$  

(2.6)

(see Fraser and Ng (1980, p. 381)). Since the predictive distribution depends on $\Omega_\lambda$, we will derive the marginal likelihood function of $\Omega_\lambda$ first.

Using (2.3) and the Jacobian (2.6), we have the joint density function of $\hat{B}_E$, $C_E$ and $Z_E$ for given $\Omega_\lambda$ as

$$p(\hat{B}_E, C_E, Z_E|\Omega_\lambda) \propto |\Omega_\lambda|^{-\frac{m}{2}} \prod_{i=1}^{m} s_{ei}^{m-p-1} \left\{ \sum_{i=1}^{m} (b_{ei} + s_{ei}g_{ei}F^{-1}) \times F(b_{ei} + s_{ei}g_{ei}F^{-1})' + \sum_{i=1}^{m} s_{ei}^2z_{ei}A\lambda z_{ei}' \right\}^{-\frac{1}{2}},$$  

(2.7)

where $F = X\Omega_\lambda^{-1}X'$, $g_{ei}$ is the $i^{th}$ row vector of the matrix $G = X\Omega_\lambda^{-1}Z_E'$, and

$$A_\lambda = \Omega_\lambda^{-1} - \Omega_\lambda^{-1}X'(X\Omega_\lambda^{-1}X')^{-1}X\Omega_\lambda^{-1}.$$

Now, we integrate out $\hat{B}_E$ and $C_E$ from (2.7), and obtain the marginal probability density of $Z_E$ for given $\Omega_\lambda$. However, following equation (2.4) the pdf of $Z_Y$ can easily be obtained from the pdf of $Z_E$. Therefore, the marginal likelihood function of $\Omega_\lambda$ conditioned on $Y$ is given by

$$p(\Omega_\lambda|Z_Y) \propto \prod_{i=1}^{m} |\Omega_\lambda|^{-\frac{1}{2}} |X\Omega_\lambda X'|^{-\frac{1}{2}} \left[ z_{yi}A_\lambda \left( \Omega_\lambda^{-1} - \Omega_\lambda^{-1}X'(X\Omega_\lambda^{-1}X')^{-1}X\Omega_\lambda^{-1} \right) z_{yi}' \right]^{-\frac{n-p}{2}}.$$

(2.8)

For a given observation matrix $Y$ and known design matrix $X$, the maximum likelihood estimate of $\lambda$ and hence $\Omega_\lambda$ may be obtained by maximizing the likelihood function (2.8). It appears that a closed form estimate for the parameters may not be available from (2.8). However, for a given observation matrix $Y$, the MLE $\hat{\lambda}$ of $\lambda$ may be obtained numerically. The marginal likelihood function in (2.8) is identical to that obtained under the assumption of independently distributed multivariate normal errors (see Kibria (1997)) and Student’s $t$ errors (see Kibria (2002)).
3 Predictive Distribution

Consider a set of $n_f$ future responses from (2.1) corresponding to the design matrix $X_f$ as

$$Y_f = BX_f + \Gamma E_f,$$  \hspace{1cm} (3.1)

where $Y_f$ and $E_f$ are the $m \times n_f$ matrices of future responses and errors, respectively, and $X_f$ is an $p \times n_f$ matrix of future regressors. To derive the joint density of $E$ and $E_f$, we combine the observed and future error matrices as

$$E^* = (E, E_f),$$

where $E^*$ is an $m \times (n + n_f)$ matrix with $\text{Cov}(E^*) = \Phi_\lambda \otimes I_m$, where

$$\Phi_\lambda = \begin{pmatrix}
\Phi_{\lambda 11} & \Phi_{\lambda 12} \\
\Phi_{\lambda 21} & \Phi_{\lambda 22}
\end{pmatrix},$$

where $\Phi_{\lambda 11}$ is an $n \times n$ covariance matrix of $e_i$, $\Phi_{\lambda 12}^\prime = \Phi_{\lambda 21}$ is an $n \times n_f$ matrix of covariances between the components of $e_i$ and $e_{fi}$ and $\Phi_{\lambda 22}$ is an $n_f \times n_f$ covariance matrix of $e_{fi}, i = 1, \ldots, m$. Following Muirhead (1982), the inverse of $\Phi_\lambda$ is obtained as

$$\Phi_\lambda^{-1} = \begin{pmatrix}
\Phi_{\lambda 11}^{-1} & \Phi_{\lambda 12}^{-1} \\
\Phi_{\lambda 21}^{-1} & \Phi_{\lambda 22}^{-1}
\end{pmatrix}.$$

Assume that $E^*$ has a multivariate elliptically contoured distribution with a known covariance matrix $\Phi_\lambda$. Then we have,

$$p(E, E_f | \Phi_\lambda) \propto |\Phi_\lambda|^{-\frac{m}{2}} g \left[ \text{tr} \left\{ E \Phi_{\lambda 11}^{-1} E^\prime + E \Phi_{\lambda 12}^{-1} E_{fi} + E_f \Phi_{\lambda 21}^{-1} E_{fi} + E_f \Phi_{\lambda 22}^{-1} E_{fi} \right\} \right].$$

Using (2.3) and the Jacobian (2.6), the joint density function of $\hat{B}_E, C_E$ and $E_f$ for given $Z_Y$ and $\Phi_\lambda$ is obtained as

$$p(\hat{B}_E, C_E, E_f | Z_Y, \Phi_\lambda) \propto |\Phi_\lambda|^{-\frac{m}{2}} \prod_{i=1}^{m} s_i^{-\frac{p-1}{2}} g \left[ \text{tr} \left\{ (\hat{B}_E X + C_E Z_Y) \Phi_{\lambda 11} (\hat{B}_E X + C_E Z_Y)^\prime + (\hat{B}_E X + C_E Z_Y) \Phi_{\lambda 21} E_f + E_f \Phi_{\lambda 22} E_f^\prime \right\} \right].$$

Consider the following transformation:

$$\begin{align*}
D &= C_E^{-1}(E_f - \hat{B}_E X_f), \\
U &= \hat{B}_E, \\
V &= C_E.
\end{align*}$$

(3.2)
The Jacobian of the transformations $J\{(\hat{B}_E, C_E, E_f) \to (U, V, D)\}$ is equal to $\prod_{i=1}^{m} v_i^{n_f}$, where $v_i$ is the $i^{th}$ diagonal element of the matrix $V$. Then the joint density function of $D$, $V$ and $U$ for given $Z_Y$ and $\Phi_\lambda$ becomes

$$p(D, U, V|Z_Y, \Phi_\lambda) \propto |\Phi_\lambda|^{-m/2} |H|^{-m/2} \prod_{i=1}^{m} v_i^{n+n_f-p-1} \times g\left\{ \left( \sum_{i=1}^{m} (u_i + v_i m_i^{-1}) H(u_i + v_i m_i^{-1})' + \sum_{i=1}^{m} v_i^2 r_{ii} \right) \right\}, \quad (3.3)$$

where $u_i$ is the $i^{th}$ row vector of the matrix $U$, $m_i$ is the $i^{th}$ row vector of the matrix $M = P_1 Z_Y' + P_2 D'$, $P_1 = X \Phi^{11}_\lambda + X \Phi^{12}_\lambda$, $P_2 = X f_1 \Phi^{21}_\lambda + X f_2 \Phi^{22}_\lambda$ and $r_{ii}$ is the $i^{th}$ diagonal element of $R_\lambda = A - M'H^{-1}M$, where $A = Z_Y \Phi^{11}_\lambda Z_Y' + Z_Y \Phi^{12}_\lambda D' + D \Phi^{21}_\lambda Z_Y' + D \Phi^{22}_\lambda D'$ and $H = X \Phi^{11}_\lambda X' + X \Phi^{12}_\lambda X_f' + X f_1 \Phi^{21}_\lambda X' + X f_2 \Phi^{22}_\lambda X_f'$. After integrating out both $U$ and $V$ from (3.3), the marginal pdf of $D$ for given $Z_Y$ and $\Phi_\lambda$ is obtained as

$$p(D|Z_Y, \Phi_\lambda) \propto |\Phi_\lambda|^{-m/2} |H|^{-m/2} \prod_{i=1}^{m} r_{i}^{-n+n_f-p}. \quad (3.4)$$

The $i^{th}$ diagonal element of $R_\lambda$ can be expressed

$$z_{yi} Q' z_{yi} + (d_i + z_{yi}Q_3 Q_2^{-1}) Q_2 (d_i + z_{yi} Q_3 Q_2^{-1})', \quad (3.5)$$

where $Q_1 = \Phi^{11}_\lambda - P_1' H^{-1} P_1$, $Q_2 = \Phi^{22}_\lambda - P_2' H^{-1} P_2$, $Q_3 = \Phi^{12}_\lambda - P_1' H^{-1} P_2$ and $Q' = Q_1 - Q_3 Q_2^{-1} Q_3$. Finally, using the relationship $s_{ei}^{-1} (e_f - b_{ei} X_f) = s_{yi}^{-1} (y_f - b_{yi} X_f) = d_i$ for $i = 1, \ldots, m$, we obtain the pdf of $Y_f$ for given $Y$ and $\Phi_\lambda$ as:

$$p(Y_f|Y, \Phi_\lambda) = \frac{\Gamma\left(\frac{n-p}{2}\right)}{\Gamma\left(\frac{n-p+n_f}{2}\right)} \prod_{i=1}^{m} (\Delta_{\lambda_i})^{-\frac{n-n_f}{2}} \times \prod_{i=1}^{m} \left\{ 1 + \frac{1}{n-p} \left( (y_{fi} - \mu_{\lambda_i}) (y_{fi} - \mu_{\lambda_i})' \right) \right\}^{-\frac{n-n_f}{2}}. \quad (3.4)$$

where $\mu_{\lambda_i} = b_{yi} X_f - s_{yi} z_{yi} Q_3 Q_2^{-1}$ and $\Delta_{\lambda_i} = \left\{ \frac{1}{n-p} [z_{yi} Q' z_{yi}]^{-1} Q_2 \right\}$. This is the product of $m$ multivariate Student’s $t$ distributions. Properties of this distribution can be found in Dickey (1967) and possible applications can be found in Kibria et al. (2002).

It is observed that for known $\lambda$ and hence $\Phi_\lambda$, each row of $Y_f$ is distributed as a $n_f$ dimensional multivariate Student’s $t$ distribution with $(n-p)$ degrees of freedom. The location parameter vector is $\mu_{\lambda_i}$, and the scale parameter matrix is $\Delta_{\lambda_i}^{-1}$ for $i = 1, 2, \ldots, m$. The predictive distribution in (3.4) is identical to that obtained under the assumption of independently distributed multivariate normal errors (see Kibria (1999)) and Student’s $t$ error (see Kibria (2002)).
4 Concluding Remarks

This paper has derived the predictive distribution for future responses from the heteroscedastic multivariate linear models under the assumption of multivariate elliptically contoured error distributions. The predictive distribution obtained is a product of multivariate Student’s $t$ distributions. It is noted that the predictive distributions under elliptically contoured distributions are identical to those obtained under independent normal errors or dependent but uncorrelated Student’s $t$ errors. This gives inference robustness with respect to departures from multivariate normal or Student’s $t$ to elliptically contoured distributions. This paper has considered the multivariate heteroscedastic linear model, which also cover the linear model for $m = 1$.

References


