University of Southern Queensland

# Bayesian Prediction Distributions for Some Linear Models under Student-$t$ Errors

A Dissertation submitted by

**Azizur Rahman**
B.Sc.(Honours), M.Sc.

For the degree of
**Master of Philosophy**

June, 2007

# Abstract

This thesis investigates the prediction distributions of future response(s), conditional on a set of realized responses for some linear models having Student-$t$ error distributions by the Bayesian approach under the uniform priors. The models considered in the thesis are the multiple regression model with multivariate-$t$ errors and the multivariate simple as well as multiple regression models with matrix-$T$ errors. For the multiple regression model, results reveal that the prediction distribution of a single future response and a set of future responses are a univariate and multivariate Student-$t$ distributions respectively with appropriate location, scale and shape parameters. The shape parameter of these prediction distributions depend on the size of the realized responses vector and the dimension of the regression parameters' vector, but do not depend on the degrees of freedom of the error distribution. In the multivariate case, the distribution of a future responses matrix from the future model, conditional on observed responses matrix from the realized model for both the multivariate simple and multiple regression models is matrix-$T$ distribution with appropriate location matrix, scale factors and shape parameter. The results for both of these models indicate that prediction distributions depend on the realized responses only through the sample regression matrix and the sample residual sum of squares and products matrix. The prediction distribution also depends on the design matrices of the realized as well as future models. The shape parameter of the prediction distribution of the future responses matrix depends on size of the realized sample and the number of regression parameters of the multivariate model. Furthermore, the prediction distributions are derived by the Bayesian method as multivariate-$t$ and matrix-$T$ are identical to those obtained under normal errors' distribution by the different statistical methods such as the classical, structural distribution and structural relations of the model approaches. This indicates not only the inference robustness with respect to departures from normal error to Student-$t$ error distributions, but also indicates that the Bayesian approach with a uniform prior is competitive with other statistical methods in the derivation of prediction distribution.

# Certification of Dissertation

I certify that the ideas, mathematical derivation of the formulas, findings and conclusions reported in this dissertation are the result of my own work, except where otherwise acknowledged. I also certify that the thesis is original and has not been previously submitted for any other award to any other university, except where otherwise acknowledged.

..................................... ......................

Signature of Candidate Date

ENDORSEMENT

..................................................... ......................

Signature of Principal Supervisor Date

..................................................... ......................

Signature of Associate Supervisor Date

# Acknowledgements

I express my first and foremost gratefulness to *Almighty Allah*, Who has created us for engaging in good deeds. With His infinite blessings and divine support I have been able to accomplish this dissertation, *Alhamdulillah*.

I wish to thank my supervisor Associate Professor Dr Shahjahan Khan for his advice and guidance throughout the course of my studies at USQ. I would also like to thank him for many academic discussions that we had during the course of this research and unbelievable support. I must also thank Dr Peter Dunn, the associate supervisor of my dissertation, for his encouragement, advice and support, specially for lectures on R.

I would like to thank the Head of Mathematics and Computing Department-Asso. Prof. Ron Addie, as well as Dr Ashley Plank and Mr Paul Fahey for providing me with a part time work at the department as well as their valuable advices and supports. I also acknowledge my gratefulness to all staff of the Department of Mathematics and Computing, particularly to Prof. A.J. Roberts, Dr Jiuyong Li , Dr Yan Li , Asso. Prof. Jesus A. Lopez, Dr Zhongwei Zhang, Dr Richard Watson, Ms Helen Nkansah and Mr Peter Greenup to mention a few, for fruitful discussions, cordial help and co-operation. I thankfully acknowledge the excellent support of the staff at the Office of Research and Higher Degrees at USQ.

I am thankful to all of my good friends at USQ, especially to Prof. Don Gorman, Mr Ronald Davis, Dr H. Barbara, Dr H. Henk, Dr E. Robert, Dr Delwar Hossain, Ms N. Annette, Mr Mafiz Rahman, Dr Khorshed Alam, Mr Enamul Kabir, Mr JD. Gupta, Mr Rezaul Bashar and Ms H. Queen to mention a few, all of you made my last one year so interesting and enjoyable. My sincere thanks are extended to Dr J. Bhowmik , Mr Owen and Mrs Tilly

Brosnan, Mr D. Kesteven, Ms R. Maureen, Mr B. David, Mr Rahim, Mr Saiful and Mr Jabid, some of my friends in Australia and overseas, for their great supports in varieties of ways.

I would like to thank all staff in the Department of Statistics, University of Chittagong, Bangladesh, particularly to the Chairperson of the Department Prof. Soma Chowdhury for all kinds of cordial advice and support. To my respective teacher and good friend Dr Zahirul Hoque- I gratefully acknowledge all of your cordial support and valuable advice.

Finally, I wish to express my deepest sense of gratitude to my beloved parents, brothers and sister for their immeasurable sacrifice, blessing, continuous inspiration and all kind of supports which led me to reach this stage of my academic studying and success. To Aleya, my wife- your sincere care and sacrifice encouraged me to complete my research work.

# Contents

# Chapter 1

# Introduction

The prediction distribution is a fundamental tool for all kinds of predictive inferences. There are many practical situations in real life where we need to make inferences about the behavior of the unobserved future responses for a model such as the linear regression model, based on the observed responses from the model. A linear regression model is used to represent the linear relationship between the response variable and a single or a set of explanatory variables. Different linear models are commonly used in many practical situation. This study considers the simple, multiple, multivariate simple and multivariate multiple regression models. The prediction distribution of future response(s) can be derived for these regression models in statistical predictive inferences.

To deal with linear regression models, some statistical assumptions for the error variables are prominent. Traditionally in prediction problems, the regression models with independent and normal error distribution are commonly considered by researchers. In many practical situations, the assumption of normality and independency for error variables of a linear model may not be appropriate. In particular, when the underlying distributions have heavier tails, the normal errors assumption fails to allow sufficient probability in the tail areas to make allowance for any extreme value or outliers. Also, the normality assumption can not deal with the uncorrelated but not independent observations which are often common in time series and econometric studies. For these cases the Student-$t$ errors assumption is suitable in which the error variables of the linear models are assumed to follow a Student-$t$ distribution with appropriate parameters.

A number of statistical methods can lead to prediction distribution. The commonly used statistical approaches are the classical approach, the structural distribution approach, the structural relations of the model approach and the Bayesian approach. The classical method is one of the oldest tools for statistical inference which is based on the likelihood principle and deals with the parameters of a model through sample observations. Under the structural distribution approach a linear model can be expressed in a reduced form, which has an error probability distribution and a structural equation between predictor variables and unknown parameters related to error constants - that are also unknown in the structural equation. The error probability distribution can generate a structural distribution of unknown parameters that is used to derive a prediction of future response(s). But the method of structural relations of the model is based on a composite structural equation of the model rather than the structural distribution, in which an observed errors value from the error distribution provides a relation between the observed responses and the unknown parameters. Moreover, the Bayesian approach is based on the Bayes's Theorem introduced by Bayes (1763) where a probabilistic information or prior probability distribution of unknown parameters is essential. The prior distribution of parameters has a type of statistical form such as a noninformative or uniform or flat prior and a class of conjugate or informative prior distributions. In practical situations, it is realistic to assume that there is some information given in previous studies and it is more convenient to use that prior information in predictive inferences.

This thesis deals with the derivation problems of prediction distribution for some linear models having Student-$t$ errors under the Bayesian approach with uniform prior distribution. The widely used multiple regression model and its special case the simple linear regression model, the multivariate simple regression model and the multivariate multiple regression model are considered in this study to obtain the prediction distributions of the future response(s) for the models.

The Bayesian prediction rule is very straightforward and mainly based on the Bayes's posterior distribution of unknown parameters. Let $\boldsymbol{y}$ be a set of observed responses from a performed or realized experiment with a joint probability density $f(\boldsymbol{y}|\boldsymbol{\theta})$, in which $\boldsymbol{\theta}$ is a set of unknown parameters. Again let a prior density of $\boldsymbol{\theta}$ be $g(\boldsymbol{\theta})$, the posterior density of $\boldsymbol{\theta}$ for given $\boldsymbol{y}$ can be obtained by Bayes's theorem and defined as $f(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta})g(\boldsymbol{\theta})$. Now, if $z^*$ is an unobserved future response from a future experiment with the same parameters $\boldsymbol{\theta}$ and assumption of the performed experiment but with different given values for the predictors, then under the Bayesian approach the prediction distribution of $z^*$, conditional on $\boldsymbol{y}$, can be obtained by solving the following integral

$$f(z^*|\boldsymbol{y}) \propto \int_{\boldsymbol{\theta}} f(\boldsymbol{\theta}|\boldsymbol{y}) f(z^*|\boldsymbol{\theta}) d\boldsymbol{\theta}$$

where $f(z^*|\boldsymbol{\theta})$ is the probability density function of the future response $z^*$ from the future model. This principle is appropriate when the future response $z^*$ is independently distributed with the observed responses $\boldsymbol{y}$, that means $z^*$ and $\boldsymbol{y}$ are not dependent to each other. However, in this study the responses for the realized as well as the future models are considered uncorrelated but not independent for the multiple regression model, and correlated for the multivariate models.

The outline of the dissertation is as follows. In Chapter 2 the realized and the future multiple linear regression model are specified with multivariate Student $t$-error distribution. The prediction distribution of a single future response as well as a set of future responses are derived by the Bayesian approach under uniform prior. As a special case, the simple linear regression model with multivariate Student-$t$ errors is also considered in this chapter to illustrate the results of prediction distribution. Chapter 3 introduces the multivariate simple regression model with matrix-$T$ errors assumption and obtains the prediction distribution of the future responses matrix, conditional on the realized responses matrix. This chapter also defines a matrix-$T$ distribution and briefly discuss the properties and application of the matrix-$T$ distribution. In chapter 4, the multivariate multiple regression model under matrix-$T$ error distribution is defined. Then the prediction distribution of the future responses matrix for the multivariate multiple regression model is derived. Furthermore, each of the above three chapters contains an introduction section with a relevant literature review and another section for the chapters' concluding remarks. The final chapter contains the summary of the results and the final conclusions.

# Chapter 2

# Multiple Regression Model

## 2.1 Introduction

The widely used multiple linear regression model represents the linear relationship between the response variable and a set of explanatory variables. The prediction distribution of future response(s) for the regression model can derive for statistical predictive inferences. In general, predictive inference uses the observed responses from a *performed experiment* to make inferences about the behavior of the unobserved future response(s) of a *future experiment* (Aitchison and Dunsmore, 1975). The details of predictive inference methods and applications of prediction distribution can be found elsewhere (Aitchison and Sculthorpe, 1965; Geisser, 1993).

Different statistical methods can lead to prediction distribution and many authors have considered prediction problems in the linear regression model. General prediction problems have been discussed by Jeffreys (1961), Aitchison and Sculthorpe (1965), and Faulkenberry (1973). Goldberger (1962), Wilson (1967) and Hahn (1972) considered prediction from linear models by using the classical method. Fraser and Haq (1970) obtained the prediction distribution for the multilinear model by using the structural distribution approach. Some authors used the structural relations, rather than the structural density function, to drive the prediction distribution from the multiple regression model (Haq 1982, Haq and Khan 1990, and Khan 2004). Khan and Haq (1994) proposed predictive inference for the auto-correlated multilinear regression model. Zellner and Chetty (1965), Aitchison and Dunsmore (1975) and Geisser (1993) discussed the prediction problem from the Bayesian viewpoint, and its applications in many areas has been discussed by Roberts (1965), Geisser (1993) and Khan (2002).

Most of the authors have contributed to study the prediction problem for linear models with independent and normal errors. But in many practical situations when the underlying distributions have heavier tails, the normal errors assumption may not be appropriate, and for such case the multivariate student-$t$ errors assumption for linear models is suitable. Unlike others Haq and Khan (1990), and Khan and Haq (1994) obtained prediction distribution for the linear regression model with multivariate Student-$t$ error terms by using the structural relation approach. The linear models with multivariate $t$-errors have also been considered by Zellner (1976) and Sutradhar and Ali (1989). In this chapter, the thesis assumes that the error terms of the performed as well as the future regression model have a joint multivariate Student-$t$ distribution with a zero location vector. Under this assumption, the marginal distribution of each error component has an univariate Student-$t$ distribution that includes the Cauchy as well as normal distributions as special cases.

The layout of this chapter is as follows. In Section 2.2, the multiple linear regression model is specified with $t$-errors. The future model is introduced in Section 2.3. The prediction distribution of a single future response has been obtained using the Bayesian approach with uniform prior distribution in Section 2.4. Then the prediction distribution of a set of future responses is derived in Section 2.5. As a special case, the simple linear regression model with multivariate Student-$t$ error is considered in Section 2.6 to illustrate the results of prediction distribution. Section 2.7 contains a summary of results and some concluding remarks.

## 2.2 The model

Let the multiple linear regression model for $n$ responses, $\boldsymbol{y} = (y_1, y_2, ..y_n)'$, be given by

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e} \tag{2.1}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, .., \beta_{k-1})'$, a $k \times 1$ dimensional regression parameters vector; $\boldsymbol{e}$, a $n \times 1$ error vector; and $\boldsymbol{X}$, a $n \times k$ dimensional design matrix of explanatory variables $(n > k)$.

Assume that each of the $n$ components in $\boldsymbol{e}$ is uncorrelated but not independent of the others and has the same univariate Student-$t$ distribution with location 0, scale $\sigma > 0$ and $\nu$ degrees of freedom. Here $\nu > 0$ represents the shape parameter of the $t$-distribution. Therefore, the joint probability density function (pdf) for the $n$ elements of $\boldsymbol{e}$ is an $n$-dimensional multivariate Student-$t$ pdf

$$f(\boldsymbol{e}) \propto (\sigma^2)^{-\frac{n}{2}} \left[1 + \frac{1}{\nu\sigma^2}\boldsymbol{e}'\boldsymbol{e}\right]^{-\frac{\nu+n}{2}}. \tag{2.2}$$

It is to be noted that the mean vector and covariance matrix of $\boldsymbol{e}$ are $\mathrm{E}(\boldsymbol{e}) = \mathbf{0}$, and $\mathrm{Cov}(\boldsymbol{e}) = \frac{\nu\sigma^2}{(\nu-2)} I_n$ for $\nu > 2$. Thus the elements of $\boldsymbol{e}$ and hence those of $\boldsymbol{y}$ are uncorrelated but not independent. Therefore the probability density function of the realized vector $\boldsymbol{y}$ becomes

$$f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2, \nu) \propto (\sigma^2)^{-\frac{n}{2}} \left[1 + \frac{1}{\nu\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right]^{-\frac{\nu+n}{2}}. \tag{2.3}$$

A Bayesian analysis of the multiple regression model with multivariate Student-$t$ errors has been discussed by Zellner (1976). The posterior distribution of parameters for a set of observations is typically the major objective of the Bayesian statistical analysis. A posterior distribution implies a marginal distribution known as prediction distribution for outcomes of any future sample observations (Roberts, 1965). Fraser and Ng (1980) discussed details about the inference of parameters $\boldsymbol{\beta}$ and $\sigma$ of the linear model. This chapter derives the prediction distribution of unobserved responses from the future model given the observed responses $\boldsymbol{y}$ from the realized model, using the Bayesian methodology with a noninformative prior distribution of parameters.

## 2.3 The future model

Let $y_f$ be an unobserved future response from the model (2.1) corresponding to the $1 \times k$ dimensional design vector $\boldsymbol{x}_f$. Then the future multiple linear regression model for a single response can be defined as

$$y_f = \boldsymbol{x}_f\boldsymbol{\beta} + e_f \tag{2.4}$$

where $\boldsymbol{\beta}$ is a $k \times 1$ dimensional parameters vector for future response $y_f$, and $e_f$ is the scaler error component associated with $y_f$. According to the assumption $e_f$ has univariate Student-$t$ distribution with $\nu$ degrees of freedom i.e., $e_f \sim t_1(0, \sigma, \nu)$.

Since the errors are uncorrelated but not independent, the realized error $\boldsymbol{e}$ and the future error $e_f$ have been combined to form a $n+1$ dimensional multivariate Student-$t$ distribution with $\nu$ degrees of freedom. Hence the observed responses $\boldsymbol{y}$ from the performed model and the unobserved response $y_f$ from the future model are also not independent but uncorrelated, then the joint density function of the combined responses $\boldsymbol{y}$ for the *performed experiment* and $y_f$ for the *future experiment* becomes

$$p(\boldsymbol{y}, y_f|\boldsymbol{\beta}, \sigma^2) \propto (\sigma^2)^{-\frac{n+1}{2}} \left[1 + \frac{1}{\nu\sigma^2}(Q_{\boldsymbol{y}} + Q_{y_f})\right]^{-\frac{\nu+n+1}{2}} \tag{2.5}$$

where $Q_{\boldsymbol{y}} = (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})$ and $Q_{y_f} = (y_f - \boldsymbol{x}_f\boldsymbol{\beta})^2$.

## 2.4 Prediction of a single future response

In the Bayesian method a prior distribution of unknown parameters is customary. Let a noninformative joint prior distribution of unknown parameters $\boldsymbol{\beta}$ and $\sigma^2$ be

$$p(\boldsymbol{\beta}, \sigma^2) \propto \frac{1}{\sigma^2}. \tag{2.6}$$

It is assumed that the degrees of freedom of the error distribution is unknown and the elements of $\boldsymbol{\beta}$ and $\log\sigma^2$ are independently and uniformly distributed.

On combining the prior density function of parameters in (2.6) with the joint density function of the combined responses from both the *performed* and *future* models in equation (2.5) by means of Bayes' Theorem, we have the following joint posterior density of $\boldsymbol{\beta}$ and $\sigma^2$ for given $\boldsymbol{y}$ and $y_f$:

$$\begin{aligned} p(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}, y_f) &\propto p(\boldsymbol{y}, y_f | \boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}, \sigma^2) && (2.7) \\ &\propto (\sigma^2)^{-\frac{n+3}{2}} \left[1 + \frac{Q}{\nu\sigma^2}\right]^{-\frac{\nu+n+1}{2}}, && (2.8) \end{aligned}$$

where $Q = Q_{\boldsymbol{y}} + Q_{y_f}$.

Since the density function of a future response from the future model join with the density function of a set of observed responses $\boldsymbol{y}$ within the joint density in (2.5), the prediction distribution of a future response can be obtained by solving the following integral

$$f(y_f|\boldsymbol{y}) \propto \int_{\boldsymbol{\beta}} \int_{\sigma^2} p(\boldsymbol{y}, y_f|\boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}, \sigma^2) d\sigma^2 d\boldsymbol{\beta} \qquad (2.9)$$

or

$$f(y_f|\boldsymbol{y}) \propto \int_{\boldsymbol{\beta}} \int_{\sigma^2} p(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}, y_f) d\sigma^2 d\boldsymbol{\beta}. \qquad (2.10)$$

That means in this case we can obtain the prediction distribution of future response(s) from the joint posterior density of unknown parameters for given combined responses generated from the performed and future models.

The joint posterior density in (2.8) can be written as the following convenient form

$$f(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}, y_f) \propto (\sigma^2)^{\frac{\nu}{2}-1} \left[Q + \nu\sigma^2\right]^{-\frac{\nu+n+1}{2}}. \qquad (2.11)$$

For the transformation

$$Q + \nu\sigma^2 = t^{-1},$$

the Jacobian of the transformation is $|J| = \frac{1}{\nu t^2}$ with the range of $t$ from 0 to $\frac{1}{Q}$.

Now equation (2.10) can be written as

$$f(y_f|\boldsymbol{y}) \propto \int_{\boldsymbol{\beta}} \int_0^{\frac{1}{Q}} \left[\frac{1}{\nu}\left(\frac{1}{t} - Q\right)\right]^{\frac{\nu}{2}-1} \left[\frac{1}{t}\right]^{-\frac{\nu+n+1}{2}} \frac{1}{\nu}\left(\frac{1}{t}\right)^2 dtd\boldsymbol{\beta}$$

or

$$f(y_f|\boldsymbol{y}) \propto \nu^{-\frac{\nu}{2}} \int_{\boldsymbol{\beta}} \int_0^{\frac{1}{Q}} [1 - Qt]^{\frac{\nu}{2}-1} t^{\frac{n+1}{2}-1} dtd\boldsymbol{\beta}. \tag{2.12}$$

If we put $z = Qt$, then the prediction density in (2.12) becomes

$$f(y_f|\boldsymbol{y}) \propto \int_{\boldsymbol{\beta}} [Q]^{\frac{n+1}{2}} \nu^{-\frac{\nu}{2}} \int_0^1 z^{\frac{n+1}{2}-1} [1 - z]^{\frac{\nu}{2}-1} dzd\boldsymbol{\beta}. \tag{2.13}$$

Equation (2.13) confirms that $z$ has a *beta* distribution, that is, $z \sim B(\frac{n+1}{2}, \frac{\nu}{2})$. After integrating with respect to $z$, the equation (2.13) becomes

$$f(y_f|\boldsymbol{y}) \propto \int_{\boldsymbol{\beta}} [Q]^{-\frac{n+1}{2}} d\boldsymbol{\beta}. \tag{2.14}$$

Now $Q = Q_{\boldsymbol{y}} + Q_{y_f}$ can be expressed as the following quadratic form of the parameters' vector $\boldsymbol{\beta}$

$$Q = A + (\boldsymbol{\beta} - B)' M (\boldsymbol{\beta} - B) \tag{2.15}$$

where $M = \boldsymbol{X}'\boldsymbol{X} + \boldsymbol{x}_f'\boldsymbol{x}_f$, $B = M^{-1}(\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{x}_f' y_f)$ and

$$A = \boldsymbol{y}'\boldsymbol{y} + y_f^2 - (\boldsymbol{y}'\boldsymbol{X} + y_f \boldsymbol{x}_f) M^{-1} (\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{x}_f' y_f).$$

It is noted that $A$ is free from the unknown regression parameters' vector $\boldsymbol{\beta}$.

Using relation (2.15), the probability density function in (2.14) can be expressed as

$$f(y_f|\boldsymbol{y}) \propto \int_{\boldsymbol{\beta}} [A + (\boldsymbol{\beta} - B)' M (\boldsymbol{\beta} - B)]^{-\frac{n+1}{2}} d\boldsymbol{\beta}. \tag{2.16}$$

The prediction density for $y_f$ can be obtained by integrating the above equation with respect to the elements of $\boldsymbol{\beta}$ using the multivariate Student-$t$ integral. Hence the prediction distribution of $y_f$, given a set of observed responses $\boldsymbol{y}$, is obtained as

$$f(y_f|\boldsymbol{y}) \propto [A]^{-\frac{n-k+1}{2}}$$

or

$$f(y_f|\boldsymbol{y}) \propto [\boldsymbol{y}'\boldsymbol{y} + y_f^2 - (\boldsymbol{y}'\boldsymbol{X} + y_f \boldsymbol{x}_f) M^{-1} (\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{x}_f' y_f)]^{-\frac{n-k+1}{2}}. \tag{2.17}$$

Applying matrix multiplication that has been discussed by Zellner (1971, p.73) on the quantity within the square brackets in (2.17), the prediction distribution of a single future response $y_f$ is obtained as

$$f(y_f|\boldsymbol{y}) = \Psi_f \left[ (n-k) + (y_f - \boldsymbol{x}_f \hat{\boldsymbol{\beta}})' H (y_f - \boldsymbol{x}_f \hat{\boldsymbol{\beta}}) \right]^{-\frac{n-k+1}{2}} \tag{2.18}$$

where

$$H = \frac{(1 - \boldsymbol{x}_f M^{-1} \boldsymbol{x}_f')}{s^2}, \ \hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{y},$$

$$s^2 = \frac{1}{n-k} [(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})] = \boldsymbol{y}'[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}$$

and

$$\Psi_f = \frac{\Gamma(\frac{n-k+1}{2}) |1 - \boldsymbol{X}_f M^{-1} \boldsymbol{x}_f'|^{\frac{1}{2}}}{\Gamma(\frac{n-k}{2}) \left[\pi (n-k) s^2\right]^{\frac{1}{2}}}$$

is the normalizing constant.

Therefore, under the Bayesian approach with uniform prior the prediction distribution of a future response $y_f$, given a set of realized responses $\boldsymbol{y}$, for the multiple linear regression model with multivariate Student-$t$ errors is a univariate Student-$t$ distribution with $n-k$ degrees of freedom, mean $\boldsymbol{x}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ and variance $\frac{(n-k)s^2}{(n-k-2)}\left(1-\boldsymbol{x}_f M^{-1}\boldsymbol{x}_f'\right)^{-1}$. Also note that the degrees of freedom of the prediction distribution does not depend on $\nu$, the degrees of freedom in the error distribution of the model. This result also coincides with the result obtained by Haq and Khan (1990) and Khan (1992), that is, the prediction distribution of future response for multivariate $t$-errors model using the *structural relation* approach.

## 2.5 Prediction of a set of future responses

The future multiple linear regression model for a set of $n^f$ responses can be defined as

$$\boldsymbol{y}_f = \boldsymbol{X}_f\boldsymbol{\beta} + \boldsymbol{e}_f \tag{2.19}$$

where $\boldsymbol{\beta}$ is a $k\times 1$ dimensional parameters vector for future response; $\boldsymbol{y}_f$ and $\boldsymbol{e}_f$ both are $n^f\times 1$ dimensional responses and error values respectively; and $\boldsymbol{X}_f$ is a $n^f\times k$ dimensional design matrix of future model explanatory variables. Also in the future model we assume that $\boldsymbol{e}_f$ has multivariate Student-$t$ distribution with $\nu$ degrees of freedom, i.e., $\boldsymbol{e}_f\sim t_{n^f}(0,\sigma,\nu)$.

Thus, the observed responses for the realized model and the unobserved future responses for the future model are combined to define as the joint density function of the set of realized responses $\boldsymbol{y} = (y_1, y_2, .., y_n)'$ and a set of future responses $\boldsymbol{y}_f = (y_1, y_2, ....y_{n^f})'$, by the following way

$$p(\boldsymbol{y}, \boldsymbol{y}_f|\boldsymbol{\beta}, \sigma^2) \propto (\sigma^2)^{-\frac{n+n^f}{2}} \left[1 + \frac{1}{\nu\sigma^2}(Q\boldsymbol{y} + Q\boldsymbol{y}_f)\right]^{-\frac{\nu+n+n^f}{2}} \tag{2.20}$$

where,

$$Q\boldsymbol{y} = (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})$$

and

$$Q\boldsymbol{y}_f = (\boldsymbol{y}_f - \boldsymbol{X}_f\boldsymbol{\beta})'(\boldsymbol{y}_f - \boldsymbol{X}_f\boldsymbol{\beta}).$$

Considering the joint prior density function of parameters in (2.6) and the combined joint density function of $\boldsymbol{y}$ and $\boldsymbol{y}_f$ in (2.20), the joint posterior density function of unknown parameters $\boldsymbol{\beta}$ and $\sigma^2$ for the responses $\boldsymbol{y}$ and $\boldsymbol{y}_f$ is obtained as

$$f(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}, \boldsymbol{y}_f) \propto (\sigma^2)^{-\frac{n+n^f+2}{2}} \left[1 + \frac{Q}{\nu\sigma^2}\right]^{-\frac{\nu+n+n^f}{2}} \tag{2.21}$$

where $Q = Q\boldsymbol{y} + Q\boldsymbol{y}_f$.

Proceeding as before in section 2.4 and applying the appropriate transformations as $Q + \nu\sigma^2 = t^{-1}$ and then $z = Qt$ in equation (2.21), the prediction distribution of $\boldsymbol{y}_f$, conditional on $\boldsymbol{y}$, can be expressed as

$$f(\boldsymbol{y}_f|\boldsymbol{y}) \propto \int_{\boldsymbol{\beta}} [Q]^{\frac{n+n^f}{2}} \nu^{-\frac{\nu}{2}} \int_0^1 z^{\frac{n+n^f}{2}-1}[1-z]^{\frac{\nu}{2}-1} dz d\boldsymbol{\beta} \tag{2.22}$$

where $z \sim B(\frac{n+n^f}{2}, \frac{\nu}{2})$.

Now at first using *beta* integral to integrating out $z$ from the above joint density, and then after expressing $\boldsymbol{\beta}$ in the quadratic form and using multivariate Student-$t$ distribution properties, the prediction distribution of a set of future responses $\boldsymbol{y}_f$ is obtained as

$$f(\boldsymbol{y}_f|\boldsymbol{y}) \propto [\boldsymbol{y}'\boldsymbol{y} + \boldsymbol{y}_f'\boldsymbol{y}_f - (\boldsymbol{y}'\boldsymbol{X} + \boldsymbol{y}_f'\boldsymbol{X}_f)M^{-1}(\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{X}_f'\boldsymbol{y}_f)]^{-\frac{n-k+n^f}{2}} \quad (2.23)$$

where $M = \boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}_f'\boldsymbol{X}_f$.

Hence, the prediction distribution of a set of future responses $\boldsymbol{y}_f$ given $\boldsymbol{y}$ is derived as

$$f(\boldsymbol{y}_f|\boldsymbol{y}) = \Phi_f \left[(n-k) + (\boldsymbol{y}_f - \boldsymbol{X}_f\hat{\boldsymbol{\beta}})'H(\boldsymbol{y}_f - \boldsymbol{X}_f\hat{\boldsymbol{\beta}})\right]^{-\frac{n-k+n^f}{2}} \quad (2.24)$$

where

$$H = \frac{(\boldsymbol{I}_{n^f} - \boldsymbol{X}_f M^{-1}\boldsymbol{X}_f')}{s^2},$$

$\hat{\boldsymbol{\beta}}$ and $s^2$ are the same as defined in section 2.4, and

$$\Phi_f = \frac{\Gamma(\frac{n-k+n^f}{2})|\boldsymbol{I}_{n^f} - \boldsymbol{X}_f M^{-1}\boldsymbol{X}_f'|^{\frac{1}{2}}}{\Gamma(\frac{n-k}{2})\left[\pi^{n^f}(n-k)s^2\right]^{\frac{1}{2}}}$$

is the normalizing constant.

Under the Bayesian theory with a noninformative prior, conditional on a set of realized responses $\boldsymbol{y}$, a set of future responses $\boldsymbol{y}_f$ has the prediction distribution of multivariate Student-$t$ distribution with $n-k$ degrees of freedom. This indicates that the shape parameter of the prediction distribution depends on the size of the observed sample $n$ and the dimension of $\boldsymbol{\beta}$. The location and scale of the prediction distribution are $\boldsymbol{X}_f\hat{\boldsymbol{\beta}}$ and $\left\{s^{-2}\left|\boldsymbol{I}_{n^f} - \boldsymbol{X}_f M^{-1}\boldsymbol{X}'_f\right|\right\}^{-\frac{1}{2}}$ respectively. This result coincides with that of Zellner (1971) and Hahn (1972), where they considered the normal error terms and operated through *classical* treatment. Geisser (1993) obtained a similar result for the linear model with normal errors using the *Bayesian* technique under a diffuse prior distribution, and Fraser and Haq (1970) obtained an identical result by using *structural distribution* of the model instead of the Bayesian approach. The result also conforms that obtained for the model having multivariate Student-$t$ error distribution through the *structural relation* approach (Haq and Khan 1990 and Khan 1992). Thus the prediction distribution is unaffected by departures from the model with indepedent and normal errors to multivariate Student-$t$ errors distribution under different statistical methods.

## 2.6 Special case: The Simple Regression Model

This section provides the results of prediction distribution of a single future response as well as a set of future responses for the simple linear regression model with multivariate Student-$t$ errors.

Since the simple linear regression model is an special case of the multiple linear regression model, the prediction distribution for a single future response as well as a set of future responses for the simple linear regression model can be easily obtained from that for the multiple regression model. Note that for $k = 2$, equation (2.1) in Section 2.2 represents a simple linear regression model with multivariate Student-$t$ error distribution. Following the appropriate operational process as used for the multiple regression model the prediction distribution of a single future response for the simple regression model can be derived as

$$f(y_f|\boldsymbol{y}) = \Psi_f \left[(n-2) + (y_f - \boldsymbol{x}_f\hat{\boldsymbol{\beta}})' H (y_f - \boldsymbol{x}_f\hat{\boldsymbol{\beta}})\right]^{-\frac{n-1}{2}} \tag{2.25}$$

where

$$H = \frac{(1 - \boldsymbol{x}_f M^{-1} \boldsymbol{x}'_f)}{s^2}$$

in which $M = \boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}'_f\boldsymbol{X}_f$, $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$,

$$s^2 = \frac{1}{n-2}[(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})]$$

and the normalizing constant is given by

$$\Psi_f = \frac{\Gamma(\frac{n-1}{2})|1 - \boldsymbol{X}_f M^{-1} \boldsymbol{x}'_f|^{\frac{1}{2}}}{\Gamma(\frac{n-2}{2})\left[\pi(n-2)s^2\right]^{\frac{1}{2}}}.$$

The above density is a univariate Student-$t$ distribution with $n-2$ degrees of freedom, mean $\boldsymbol{x}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ and variance $\frac{(n-2)s^2}{(n-4)}\left(1 - \boldsymbol{x}_f M^{-1} \boldsymbol{x}'_f\right)^{-1}$.

Moreover, in this special case for $k = 2$ in the multiple regression model, we can obtain the prediction distribution of a set of future responses for the simple regression model as

$$f(\boldsymbol{y}_f|\boldsymbol{y}) = \Phi_f \left[(n-2) + (\boldsymbol{y}_f - \boldsymbol{X}_f\hat{\boldsymbol{\beta}})' H (\boldsymbol{y}_f - \boldsymbol{X}_f\hat{\boldsymbol{\beta}})\right]^{-\frac{n+n^f-2}{2}} \tag{2.26}$$

where

$$H = \frac{(\boldsymbol{I}_{n^f} - \boldsymbol{X}_f M^{-1} \boldsymbol{X}'_f)}{s^2}$$

$M$, $\hat{\boldsymbol{\beta}}$ and $s^2$ are the same as defined above in (2.25), and

$$\Phi_f = \frac{\Gamma(\frac{n+n^f-2}{2})|\boldsymbol{I}_{n^f} - \boldsymbol{X}_f M^{-1} \boldsymbol{X}'_f|^{\frac{1}{2}}}{\Gamma(\frac{n-2}{2}) \left[\pi^{n^f}(n-2)s^2\right]^{\frac{1}{2}}}$$

is the normalizing constant.

It is clear that under the Bayesian method with a uniform prior distribution, a set of future responses for the simple linear regression model has the prediction distribution of multivariate Student-$t$ distribution with $n-2$ degrees of freedom. The location and scale of this prediction distribution are $\boldsymbol{X}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ and $\left[s^{-2}\left|\boldsymbol{I}_{n^f} - \boldsymbol{X}_f M^{-1} \boldsymbol{X}'_f\right|\right]^{-\frac{1}{2}}$ respectively.

## 2.7 Concluding remarks

In this chapter, the prediction distribution for unobserved future response(s), conditional on a set of observed responses is derived from the multiple linear regression model by using the Bayesian method with a noninformative prior distribution, and under the assumption that the error terms of the model are uncorrelated but not independent as well as having the joint multivariate Student-$t$ distribution. The results reveal that the prediction distribution of a single future response and a set of future responses are univariate Student-$t$ distribution and multivariate Student-$t$ distribution respectively. These prediction distributions coincide with the results as derived by other statistical methods as well as for the model with normal errors. Therefore, it is noted that the prediction distribution for multiple regression model with multivariate Student-$t$ error distribution as well as independent and normal errors under the Bayesian, classical, structural relation and structural distribution approach is the same. Furthermore, the simple regression model is considered as a special case of the multiple regression model, and the results reveal that, like the response(s) for the multiple regression model, a single future response as well as a set of future responses for the simple regression model follow the univariate and multivariate Student-$t$ distributions respectively with $n-2$ degrees of freedom. As well, the prediction distribution depends on the observed responses and the design matrices of the realized model as well as the future model. The shape parameter of the prediction distribution depends on the size of the realized sample and the dimension of parameters vector of the model. However the shape parameter of the prediction distribution does not depend on the degrees of freedom of the error distribution.

# Chapter 3

# Multivariate Simple Regression Model

## 3.1 Introduction

The multivariate simple regression model represents the relationship between a set of values of several dependent variables and a single value of an independent or explanatory variable. This model is more general than the simple linear regression model which represents the linear relationship between a single value of dependent variable and a specific value of the independent variable. It is noted that the multivariate simple linear regression model is an extension of the commonly used simple linear regression model in a multivariate setup. The multivariate simple regression model is used to analyze data from different experimental situations where more than one response variables are observed for a single value of the explanatory variable.

There are many experimental situations in real life where we need to study on a set of responses from more than one dependent variable corresponding to a single value of the independent variable. For example, if several patients are given the same dose of a medicine to observe any response from the subjects, then for one particular value of the explanatory variable, there will be several values of the response variable from different subjects. The model can also be applied to any other experimental or observational studies where multiple responses are generated for one particular value of the independent variable. For details on the multivariate simple regression model interested readers may refer to Khan (2005, 2006) and Saleh (2006).

In studying multivariate regression models, most researchers have considered the normal errors model. Geisser (1965) studied the multivariate linear model under independent normal errors and obtained the prediction distribution by a classical approach. Fraser and Haq (1969) considered a structural distribution approach, and Haq (1982) used the method of structural relationships of the model to obtain the prediction distribution from the normal errors multivariate model. The Bayesian method has been considered by Zellner and Chetty (1965), Zellner (1971) and Kibria et al. (2002) among others to deal with prediction problems from the traditional multivariate model under the multivariate normal errors assumption. Furthermore, the multivariate simple regression model with independent and normal errors is considered by Khan (2006). He uses the structural relation of the model approach and the Bayesian approach to obtain the prediction distribution.

The matrix-$T$ errors regression models are considered by few researchers. Khan and Haq (1994b) investigated the predictive inference for the future responses from a multilinear model with matrix-$T$ errors by using the structural relationships of the model. In addition, the prediction distribution for future responses from the multivariate linear model with matrix-$T$ errors has been studied by Kibria and Haq (2000). They also used the structural relation approach. Khan (2002) considered this structural relation method to obtain the prediction distribution of the regression matrix under the matrix-$T$ errors but not for the future responses matrix.

In this chapter, the widely used multivariate simple regression model with matrix-$T$ error distribution is introduced. The prediction distribution for the future responses matrix is derived by the Bayesian approach under the uniform or non-informative prior distribution of the unknown parameters of the model. It is shown that the prediction distribution of the future responses matrix is a matrix-$T$ distribution with appropriate degrees of freedom that depends on the size of the realized sample and the dimension of the regression parameter. Furthermore, since the errors matrix, the responses matrix and the prediction distribution of the future responses matrix follow matrix-$T$ distribution, a definition of matrix-$T$ distribution is provided here and its properties as well as applications are also briefly addressed.

The layout of this chapter is as follows. Section 3.2 defines a matrix-$T$ distribution and also addresses some proprieties, and application of matrix-$T$ distribution in predictive inference. The multivariate simple regression model with matrix-$T$ errors is defined in Section 3.3, and the uniform prior distribution of unknown parameters is provided in Subsection 3.3.1. The future multivariate simple regression model is defined in Section 3.4. In Section 3.5, the prediction distribution of the future responses matrix is derived by the Bayesian method under uniform prior. Some concluding remarks are presented in Section 3.6.

## 3.2 Matrix-$T$ distribution

To deal with multivariate models having matrix-$T$ errors, the definition of matrix-$T$ density is essential. The errors matrix as well as the responses matrix for matrix-$T$ error models will obviously follow matrix-$T$ distributions.

Let $\Gamma_m(b)$ be a generalized gamma function introduced by Siegel (1935) and defined as

$$\boldsymbol{\Gamma}_m(b) = \left[\Gamma\left(\frac{1}{2}\right)\right]^{\frac{1}{2}m(m-1)} \prod_{\alpha=1}^{m} \Gamma\left(b + \frac{\alpha - m}{2}\right); b > \frac{m-1}{2} \qquad (3.1)$$

for nonzero positive integers $m$ and $b$ with $\alpha = 1, 2, ....., m$.

Then a random matrix $\boldsymbol{X}$ of order $p \times m$ has a matrix-$T$ distribution with location parameter $\boldsymbol{\mu} \in \Re^{p\times m}$, scale factors $\boldsymbol{\Omega}_{p\times p}$ and $\boldsymbol{\Sigma}_{m\times m}$ both being positive definite symmetric matrices, and the shape parameter $\nu > 0$, defined by $\boldsymbol{X} \sim T_{pm}\ (\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}, \nu)$, if its density function is given by:

$$f(\boldsymbol{X}|\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}) = c\frac{|\boldsymbol{\Omega}^{-1}|^{\frac{m}{2}}}{|\boldsymbol{\Sigma}|^{\frac{p}{2}}}|I_m + \boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})|^{-\frac{1}{2}(\nu+p+m-1)} \qquad (3.2)$$

where $c = [\Gamma(\frac{1}{2})]^{mp}\frac{\Gamma_m[\frac{1}{2}(\nu+m-1)]}{\Gamma_m[\frac{1}{2}(\nu+p+m-1)]}$ is the normalizing constant.

The matrix-$T$ distribution was first introduced by Dickey (1967). He derived the density of matrix-$T$ as a logical generalization of the multivariate Student-$t$ distribution to deal with matrix variate problems. The matrix-$T$ distribution and some of its properties can be found in Zellner (1971, pp. 396-99), Box and Tiao (1992, sec. 8.4), Press (1982), Bauwens et al. (1999, pp. 305-9), and Loschi et al. (2003) among others. For example, when $\boldsymbol{X}$ has the above matrix-$T$ distribution the mode is equal to $\boldsymbol{\mu}$ and $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu}$ if $\nu > 1$, while the covariance matrix of $\boldsymbol{X}$ is $\mathrm{Cov}(\boldsymbol{X}) = \frac{\nu}{\nu-2}[\boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}]$ if $\nu > 2$, and $\otimes$ represents the Kronecker product between two matrices.

Moreover, the distribution of the transpose of $\boldsymbol{X}$ can be written as $\boldsymbol{X}' \sim T_{mp}\left(\boldsymbol{\mu}', \boldsymbol{\Sigma}^{-1}, \boldsymbol{\Omega}^{-1}, \nu\right)$ when $\boldsymbol{X} \sim T_{pm}\left(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}, \nu\right)$. This follows from a well known result for the determinant, that is, if $A$ and $B$ are $p \times m$ matrices, then $|I_p + AB'| = |I_m + B'A|$ (see Magnus and Neudecker, 1988, sec. 12), and also, from the properties of the generalized gamma function $\frac{\Gamma_p(\nu+p+m)}{\Gamma_p(\nu+p)} = \frac{\Gamma_m(\nu+p+m)}{\Gamma_m(\nu+m)}$ for all $p, m \geq 1$ and $\nu > 0$; (see, e.g., Phillips, 1989).

According to Khan (2002), since the degrees of freedom $\nu$ of matrix-$T$ distribution is a positive real number, for different values of $\nu$ we get different distributions, and hence the matrix-$T$ model represents a class of distributions with varying shape. If $\nu$ tends to infinity, the matrix-$T$ density approaches to matric-variate normal density, and when $\nu = 1$, the matrix-$T$ distribution becomes matrix variate Cauchy distribution.

The distribution of individual elements of $\boldsymbol{X}$ can be obtained directly from the respective components in the parameters of the above matrix-$T$ distribution in (3.2). For instance, if $X_{ij}$ and $\mu_{ij}$ denote the element in $i^{th}$ row and $j^{th}$ column of $\boldsymbol{X}$ and $\boldsymbol{\mu}$, respectively, and if $\sigma_{jj}$ is the $j^{th}$ diagonal element of $\boldsymbol{\Sigma}$, while $\omega_{ii}$ denotes the $i^{th}$ diagonal element of $\boldsymbol{\Omega}$, then it can be shown that $X_{ij} \sim t(\mu_{ij}, \omega_{ii}, \sigma_{jj}, \nu)$ when $\boldsymbol{X}$ has the above matrix-$T$ distribution. That means $X_{ij}$ has an univariate Student-$t$ distribution. Similarly, the marginal distribution of any column (or row) as well as the conditional distribution of one column (or row), given another, follows a multivariate Student-$t$ distribution with appropriate parameters. As well, the marginal and conditional distributions of any sub-matrix of $\boldsymbol{X}$ and one sub-matrix, given another, follow matrix-$T$ distribution with appropriate parameters and can also be determined directly from the parameters of the distribution of the random matrix $\boldsymbol{X}$.

The matrix-$T$ distribution has a wide range of applications in multivariate statistical inference, especially in Bayesian analysis. Kibria (2006) demonstrates some applications of matrix-$T$ distribution in spatial prediction problems. He reveals that the matrix-$T$ distribution has significant contributions to interpolate the air pollutants and thereby to the field of environmental risk analysis. Further applications of matrix-$T$ distribution have been discussed in Press (1986), Box and Tiao (1992, sec. 8.4) among many others.

## 3.3 The Model

Let $\boldsymbol{y}_j$ be a row vector of order $1 \times p$ of the values of the $j^{th}$ responses associated with a single value of the independent variable $x_j$ from a multivariate simple regression model. Then $n$ realization of $\boldsymbol{y}_j$ can be expressed as the set of linear equations

$$\boldsymbol{y}_j = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_j + \boldsymbol{e}_j, \text{ for } j = 1, 2, ...., n \tag{3.3}$$

where $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are the $p$-dimensional intercept and slope parameters respectively, and $\boldsymbol{e}_j$ is the error vector of order $1 \times p$ associated with the responses vector $\boldsymbol{y}_j$.

Assume that each of the $p$ components in $\boldsymbol{e}_j$ is correlated with the others and $\boldsymbol{e}_j$ has the multivariate Student-$t$ distribution with $\nu$ degrees of freedom, location $\mathbf{0}$ and scale factor $\boldsymbol{\Sigma}$, where $\mathbf{0}$ is a $1 \times p$ dimensional vector of zero's and $\boldsymbol{\Sigma}$ is a $p \times p$ order non-singular scale parameter matrix. Thus, the joint probability density function of $\boldsymbol{e}_j$ is a $p$-dimensional multivariate Student-$t$

density, which is defined as

$$f(\boldsymbol{e}_j) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}} \left[1 + \frac{1}{\nu}\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_j{}'\boldsymbol{e}_j\right]^{-\frac{\nu+p}{2}}. \tag{3.4}$$

It is to be noted that the mean vector and covariance matrix of $\boldsymbol{e}_j$ are $\mathrm{E}(\boldsymbol{e}_\mathrm{j}) = \mathbf{0}$ for $\nu > 1$, and $\mathrm{Cov}(\boldsymbol{e}_\mathrm{j}) = \frac{\nu}{(\nu-2)}\boldsymbol{\Sigma}$ for $\nu > 2$.

The linear model in (3.3) can be expressed in the following form

$$\boldsymbol{y}_j = \boldsymbol{x}_j\boldsymbol{\beta} + \boldsymbol{e}_j \tag{3.5}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)'$, a $2 \times p$ dimensional parameters vector; and $\boldsymbol{x}_j = (1, x_j)$, a $1 \times 2$ dimensional design matrix of the explanatory variable for $j = 1, 2, ...., n$. Since the elements of $\boldsymbol{e}_j$ are correlated, the elements of responses vector $\boldsymbol{y}_j$ are also correlated. Therefore, the joint density function of a set of $p$ elements of $\boldsymbol{y}_j$ can be expressed as

$$f(\boldsymbol{y}_j|\boldsymbol{\beta}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}} \left[1 + \frac{1}{\nu}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_j - \boldsymbol{x}_j\boldsymbol{\beta})'(\boldsymbol{y}_j - \boldsymbol{x}_j\boldsymbol{\beta})\right]^{-\frac{\nu+p}{2}}. \tag{3.6}$$

Now, a set of $n(> p)$ vector responses $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_n)'$ is a $n \times p$ matrix, that has been generated from the above multivariate simple regression model in (3.5), then the multivariate simple regression model can be expressed in the following way

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{E} \tag{3.7}$$

where $\boldsymbol{Y}$ is a $n \times p$ matrix of observed responses, $\boldsymbol{X}$ a $n \times 2$ dimensional design matrix of explanatory variables for $(n > 2)$, $\boldsymbol{\beta}$ is a $2 \times p$ matrix of regression parameters and $\boldsymbol{E}$ is a $n \times p$ random error matrix associated with the response matrix. Since each of the $p$ dimensional row vector of errors

$\boldsymbol{e}_j$ follows a multivariate Student-$t$ distribution with density given in (3.4), the errors matrix $\boldsymbol{E}$ has a matrix-$T$ distribution with the probability density function

$$f(\boldsymbol{E}) \propto |\boldsymbol{I}_n^{-1}|^{\frac{p}{2}}|\boldsymbol{\Sigma}^{-1}|^{\frac{n}{2}}|\boldsymbol{I}_p + \boldsymbol{\Sigma}^{-1}\boldsymbol{E}'\boldsymbol{E}|^{-\frac{1}{2}(\nu+p+n-1)} \tag{3.8}$$

where $\boldsymbol{I}_n$ is an $n \times n$ order identity matrix.

It is noted that, within the errors matrix $\boldsymbol{E}$, $p$ elements of each row are correlated and $n$ rows are uncorrelated.

Thus, the responses matrix $\boldsymbol{Y}$ also has a matrix-$T$ distribution with probability density function as

$$f(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\beta},\boldsymbol{\Sigma}) = \boldsymbol{c}\frac{|\boldsymbol{I}_n^{-1}|^{\frac{p}{2}}}{|\boldsymbol{\Sigma}|^{\frac{n}{2}}}|\boldsymbol{I}_p + \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})|^{-\frac{1}{2}(\nu+p+n-1)} \tag{3.9}$$

i.e.,

$$\boldsymbol{Y} \sim T_{np}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{I}_{n\times n}, \boldsymbol{\Sigma}_{p\times p}, \nu)$$

where the normalizing constant is given by

$$\boldsymbol{c} = \left[\Gamma\left(\frac{1}{2}\right)\right]^{np}\frac{\Gamma_p[\frac{1}{2}(\nu+p-1)]}{\Gamma_p[\frac{1}{2}(\nu+p+n-1)]}$$

in which $\Gamma_p(.)$ is a generalized gamma function as defined in (3.1), $\boldsymbol{I}_n$ is an $n\times n$ positive definite identity matrix, $\boldsymbol{\Sigma}$ is a $p\times p$ positive definite symmetric matrix and $\nu$ is degrees of freedom of the matrix-$T$ distribution.

### 3.3.1 The prior and posterior distributions

A prior distribution of unknown parameters is an important element of the Bayesian statistical method. Assume that the joint prior distribution of the regression matrix $\boldsymbol{\beta}$ and the $\frac{1}{2}p(p+1)$ distinct elements of $\boldsymbol{\Sigma}$ is uniform, which means in the experimental situation that a little is known about these unknown parameters. Adopting the invariance theory (Jeffreys, 1961, p.179), we consider the following prior distribution

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = p(\boldsymbol{\beta})p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{p+1}{2}}, \tag{3.10}$$

where

$$p(\boldsymbol{\beta}) = \text{ constant,}$$

$$p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{p+1}{2}},$$

and $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are independently distributed.

Let $\boldsymbol{\theta}$= [$\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$] be a set of unknown parameters. Then the Bayes' posterior density of $\boldsymbol{\theta}$ can be defined as

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{p(\boldsymbol{\theta})p(\boldsymbol{Y}|\boldsymbol{\theta})}{p(\boldsymbol{Y})} \tag{3.11}$$

where

$$p(\boldsymbol{Y}) = \int_{\boldsymbol{\theta}} p(\boldsymbol{\theta})p(\boldsymbol{Y}|\boldsymbol{\theta})d\boldsymbol{\theta}$$

in which $p(\boldsymbol{Y}|\boldsymbol{\theta})$ is the joint probability function or likelihood function for $\boldsymbol{Y}$ and $p(\boldsymbol{\theta})$ is a prior density of $\boldsymbol{\theta}$.

Now, the posterior density of parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ for the observed responses matrix $\boldsymbol{Y}$ can be defined as

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) \propto p(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{\Sigma})p(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \tag{3.12}$$

where $p(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{\Sigma})$ is the joint probability function or likelihood function for $\boldsymbol{Y}$, which is provided in equation (3.9) and $p(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ is the prior density of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ is defined in equation (3.10). Hence we get the following posterior density of unknown parameters' matrices $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ for the realized responses matrix $\boldsymbol{Y}$ as

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) \propto |\boldsymbol{\Sigma}|^{-\frac{n+p+1}{2}} \left|\boldsymbol{I}_p + \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})\right|^{-\frac{\nu+p+n-1}{2}} \tag{3.13}$$

in which the normalizing constant of the posterior distribution can be easily obtained by integrating over the density function with respect to the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$.

To evaluate the normalizing constant the above density in (3.13) can be written as

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = \boldsymbol{\Phi}|\boldsymbol{\Sigma}|^{-\frac{n+p+1}{2}} \left|\boldsymbol{I}_p + \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})\right|^{-\frac{\nu+p+n-1}{2}} \tag{3.14}$$

where $\boldsymbol{\Phi}$ represents the normalizing constant. The value of $\boldsymbol{\Phi}$ can be obtained by solving the following equation

$$\begin{aligned} 1 &= \boldsymbol{\Phi} \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\Sigma}} |\boldsymbol{\Sigma}|^{-\frac{n+p+1}{2}} \\ &\times \left|\boldsymbol{I}_p + \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})\right|^{-\frac{\nu+p+n-1}{2}} d\boldsymbol{\Sigma} d\boldsymbol{\beta} \end{aligned} \tag{3.15}$$

or

$$\begin{aligned} \boldsymbol{\Phi}^{-1} &= \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\Sigma}} |\boldsymbol{\Sigma}|^{-\frac{n+p+1}{2}} \\ &\times \left|\boldsymbol{I}_p + \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})\right|^{-\frac{\nu+p+n-1}{2}} d\boldsymbol{\Sigma} d\boldsymbol{\beta} \end{aligned} \tag{3.16}$$

Using the matrix transformation $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Lambda}$ with the Jacobian of the transformation

$$|J| = \frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\Lambda}} = |\boldsymbol{\Lambda}^{-1}|^{p+1},$$

equation (3.16) can be expressed as

$$\begin{aligned}
\mathbf{\Phi}^{-1} &= \int_{\boldsymbol{\beta}}\int_{\mathbf{\Lambda}} |\mathbf{\Lambda}^{-1}|^{-\frac{n-p-1}{2}} \, |\boldsymbol{I}_p + \mathbf{\Lambda}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})|^{-\frac{\nu+p+n-1}{2}} \, d\mathbf{\Lambda} d\boldsymbol{\beta} \\
&= \int_{\boldsymbol{\beta}}\int_{\mathbf{\Lambda}} |\mathbf{\Lambda}|^{\frac{n}{2}-\frac{p+1}{2}} \\
&\times \ |\boldsymbol{I}_p + \mathbf{\Lambda}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})|^{-(\frac{n}{2}+\frac{\nu+p-1}{2})} \, d\mathbf{\Lambda} d\boldsymbol{\beta}. \qquad (3.17)
\end{aligned}$$

Applying the generalized beta integral for the matrix variables (cf. Khan, 2000) over $\mathbf{\Lambda}$ the above equation becomes

$$\begin{aligned}
\mathbf{\Phi}^{-1} &= \int_{\boldsymbol{\beta}} |(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})|^{-\frac{n}{2}} \, B_p\left(\frac{n}{2}, \frac{\nu+p-1}{2}\right) d\boldsymbol{\beta} \\
&= B_p\left(\frac{n}{2}, \frac{\nu+p-1}{2}\right) \\
&\times \int_{\boldsymbol{\beta}} \left|\boldsymbol{S}_Y + (\boldsymbol{\beta}-\boldsymbol{X}\hat{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\beta}-\boldsymbol{X}\hat{\boldsymbol{\beta}})\right|^{-\frac{n}{2}} d\boldsymbol{\beta} \qquad (3.18)
\end{aligned}$$

where, $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ is the OLS of $\boldsymbol{\beta}$ and $\boldsymbol{S}_Y = (\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})$ of order $p \times p$.

Now equation (3.18) can be written as

$$\begin{aligned}
\mathbf{\Phi}^{-1} &= B_p\left(\frac{n}{2}, \frac{\nu+p-1}{2}\right) |\boldsymbol{S}_Y|^{-\frac{n}{2}} \\
&\times \int_{\boldsymbol{\beta}} \left|\boldsymbol{I}_p + \boldsymbol{S}_Y^{-1}(\boldsymbol{\beta}-\boldsymbol{X}\hat{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\beta}-\boldsymbol{X}\hat{\boldsymbol{\beta}})\right|^{-\frac{\delta+p+2-1}{2}} d\boldsymbol{\beta} \quad (3.19)
\end{aligned}$$

where $\delta = n - p - 1$.

By using the properties of the matrix-$T$ distribution to the integration with respect to $\boldsymbol{\beta}$, the result can be expressed as

$$\begin{aligned}
\mathbf{\Phi}^{-1} &= B_p\left(\frac{n}{2}, \frac{\nu+p-1}{2}\right) |\boldsymbol{S}_Y|^{-\frac{n}{2}} |\boldsymbol{X}'\boldsymbol{X}|^{-\frac{p}{2}} |\boldsymbol{S}_Y| \left[\Gamma\left(\frac{1}{2}\right)\right]^{2p} \frac{\Gamma_p\left(\frac{\delta+p-1}{2}\right)}{\Gamma_p\left(\frac{\delta+p+1}{2}\right)} \\
&= B_p\left(\frac{n}{2}, \frac{\nu+p-1}{2}\right) |\boldsymbol{S}_Y|^{-\frac{n-2}{2}} |\boldsymbol{X}'\boldsymbol{X}|^{-\frac{p}{2}} \left[\Gamma\left(\frac{1}{2}\right)\right]^{2p} \frac{\Gamma_p\left(\frac{\delta+p-1}{2}\right)}{\Gamma_p\left(\frac{\delta+p+1}{2}\right)} \\
&= \left[\Gamma\left(\frac{1}{2}\right)\right]^{2p} |\boldsymbol{S}_Y|^{-\frac{n-2}{2}} |\boldsymbol{X}'\boldsymbol{X}|^{-\frac{p}{2}} \frac{\Gamma_p\left(\frac{\delta+p-1}{2}\right)}{\Gamma_p\left(\frac{\delta+p+1}{2}\right)} B_p\left(\frac{n}{2}, \frac{\nu+p-1}{2}\right) (3.20)
\end{aligned}$$

Now putting $\delta = n - p - 1$ and expressing $B_p\left(\frac{n}{2}, \frac{\nu+p-1}{2}\right)$ as the generalized gamma function, the above equation has a form

$$\boldsymbol{\Phi}^{-1} = \left[\Gamma\left(\frac{1}{2}\right)\right]^{2p} |\boldsymbol{S}_Y|^{-\frac{n-2}{2}} |\boldsymbol{X}'\boldsymbol{X}|^{-\frac{p}{2}} \frac{\Gamma_p\left(\frac{n-2}{2}\right)}{\Gamma_p\left(\frac{n}{2}\right)} \frac{\Gamma_p\left(\frac{n}{2}\right)\Gamma_p\left(\frac{\nu+p-1}{2}\right)}{\Gamma_p\left(\frac{\nu+n+p-1}{2}\right)}$$

and hence the normalizing constant in $p(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$ is obtained as

$$\boldsymbol{\Phi} = \frac{|\boldsymbol{X}'\boldsymbol{X}|^{\frac{p}{2}} |\boldsymbol{S}_Y|^{\frac{n-2}{2}}}{\left[\Gamma\left(\frac{1}{2}\right)\right]^{2p}} \frac{\Gamma_p\left(\frac{\nu+n+p-1}{2}\right)}{\Gamma_p\left(\frac{n-2}{2}\right)\Gamma_p\left(\frac{\nu+p-1}{2}\right)}. \tag{3.21}$$

## 3.4 The Future Model

Let $\boldsymbol{Y}_f$ be an unobserved future responses matrix generated from the model provided in equation (3.7) corresponding to the $n_f \times 2$ dimensional design matrix $\boldsymbol{X}_f$. Then the future multivariate simple regression model for $\boldsymbol{Y}_f$ can be defined as

$$\boldsymbol{Y}_f = \boldsymbol{X}_f\boldsymbol{\beta} + \boldsymbol{E}_f \tag{3.22}$$

where $\boldsymbol{\beta}$ is a $2 \times p$ dimensional regression parameters matrix for future responses, and $\boldsymbol{E}_f$ is an $n^f \times p$ dimensional errors matrix associated the responses matrix $\boldsymbol{Y}_f$. Also in the future model, $\boldsymbol{E}_f$ has a Matrix-$T$ distribution with $\nu$ degrees of freedom, that is,

$$\boldsymbol{E}_f \sim T_{n_f p}(\mathbf{0}, \boldsymbol{I}_{n_f \times n_f}, \boldsymbol{\Sigma}_{p\times p}, \nu).$$

The components in each row of the realized error matrix $\boldsymbol{E}$ and the future error matrix $\boldsymbol{E}_f$ are correlated, their respective components in each row of the observed responses matrix $\boldsymbol{Y}$ for the realized model and that of the unobserved future responses matrix $\boldsymbol{Y}_f$ from the future model are also correlated.

Thus the joint likelihood function of parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ for the combined responses $\boldsymbol{Y}$ from the realized model and $\boldsymbol{Y}_f$ from the future model can be expressed as

$$p(\boldsymbol{Y}, \boldsymbol{Y}_f | \boldsymbol{\beta}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n+n^f}{2}} \left| \boldsymbol{I}_p + \boldsymbol{\Sigma}^{-1} \boldsymbol{Q} \right|^{-\frac{\nu+p+n+n^f-1}{2}} \tag{3.23}$$

where

$$\boldsymbol{Q} = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + (\boldsymbol{Y}_f - \boldsymbol{X}_f\boldsymbol{\beta})'(\boldsymbol{Y}_f - \boldsymbol{X}_f\boldsymbol{\beta}).$$

## 3.5 Prediction of future responses

Since prior distribution of unknown parameters is a part of the Bayesian method, a uniform prior density is considered in (3.10) which has been used by Zellner (1971), Bernardo and Rueda (2002) and Khan (2006) among others. Employing this prior distribution in conjunction with the joint likelihood function for the response matrices of both the performed and future models as given in equation (3.23), the following posterior density of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ can be obtained by the Bayes' Theorem

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma} | \boldsymbol{Y}, \boldsymbol{Y}_f) \propto |\boldsymbol{\Sigma}|^{-\frac{n+n^f+p+1}{2}} \left| \boldsymbol{I}_p + \boldsymbol{\Sigma}^{-1} \boldsymbol{Q} \right|^{-\frac{\nu+p+n+n^f-1}{2}} \tag{3.24}$$

Then the prediction distribution of a future response can be derived by solving the following integral

$$f(\boldsymbol{Y}_f | \boldsymbol{Y}) \propto \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\Sigma}} p(\boldsymbol{Y}, \boldsymbol{Y}_f | \boldsymbol{\beta}, \boldsymbol{\Sigma}) p(\boldsymbol{\beta}, \boldsymbol{\Sigma}) d\boldsymbol{\Sigma} d\boldsymbol{\beta} \tag{3.25}$$

or

$$f(\boldsymbol{Y}_f | \boldsymbol{Y}) \propto \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\Sigma}} p(\boldsymbol{\beta}, \boldsymbol{\Sigma} | \boldsymbol{Y}, \boldsymbol{Y}_f) d\boldsymbol{\Sigma} d\boldsymbol{\beta} \tag{3.26}$$

or

$$f(\boldsymbol{Y}_f|\boldsymbol{Y}) \propto \int_{\boldsymbol{\beta}} \int_{|\boldsymbol{\Sigma}|>\boldsymbol{0}} |\boldsymbol{\Sigma}|^{-\frac{n+n^f+p+1}{2}} \left|\boldsymbol{I}_p + \boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\right|^{-\frac{\nu+p+n+n^f-1}{2}} d\boldsymbol{\Sigma} d\boldsymbol{\beta}. \quad (3.27)$$

Consider the following matrix transformation

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Lambda}.$$

The Jacobian of the transformation is

$$|J| = \frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\Lambda}} = |\boldsymbol{\Lambda}^{-1}|^{p+1}.$$

Employing the result of the above transformation, the equation (3.27) can be written as

$$f(\boldsymbol{Y}_f|\boldsymbol{Y}) \propto \int_{\boldsymbol{\beta}} \int_{|\boldsymbol{\Lambda}|>\boldsymbol{0}} |\boldsymbol{\Lambda}^{-1}|^{-\frac{n+n^f-p-1}{2}} |\boldsymbol{I}_p + \boldsymbol{\Lambda}\boldsymbol{Q}|^{-\frac{\nu+p+n+n^f-1}{2}} d\boldsymbol{\Lambda} d\boldsymbol{\beta} \quad (3.28)$$

or

$$f(\boldsymbol{Y}_f|\boldsymbol{Y}) \propto \int_{\boldsymbol{\beta}} \int_{|\boldsymbol{\Lambda}|>\boldsymbol{0}} |\boldsymbol{\Lambda}|^{\frac{n+n^f}{2}-\frac{p+1}{2}} |\boldsymbol{I}_p + \boldsymbol{\Lambda}\boldsymbol{Q}|^{-\left(\frac{n+n^f}{2}+\frac{\nu+p-1}{2}\right)} d\boldsymbol{\Lambda} d\boldsymbol{\beta}. \quad (3.29)$$

Now using an extension of the generalized beta integral with the matrix variable introduced by Khan (2000) as

$$I = \int_{|\boldsymbol{W}|>\boldsymbol{0}} |\boldsymbol{W}|^{a-\frac{m+1}{2}} |\boldsymbol{I}_m + \boldsymbol{W}\boldsymbol{D}|^{-(a+b)} d\boldsymbol{W} = |\boldsymbol{D}|^{-1} B_m(a,b); \quad (3.30)$$

Following result is obtained from equation (3.29) by integrating with respective to $\boldsymbol{\Lambda}$

$$f(\boldsymbol{Y}_f|\boldsymbol{Y}) \propto \int_{\boldsymbol{\beta}} |\boldsymbol{Q}|^{-\frac{n+n^f}{2}} B_p\left(\frac{n+n^f}{2}, \frac{\nu+p-1}{2}\right) d\boldsymbol{\beta} \quad (3.31)$$

or

$$f(\boldsymbol{Y}_f|\boldsymbol{Y}) \propto \int_{\boldsymbol{\beta}} |\boldsymbol{Q}|^{-\frac{n+n^f}{2}} d\boldsymbol{\beta}. \tag{3.32}$$

Now $\boldsymbol{Q}$ can be expressed as the following convenient form

$$\boldsymbol{Q} = \boldsymbol{R} + (\boldsymbol{\beta} - \boldsymbol{P})'\boldsymbol{M}(\boldsymbol{\beta} - \boldsymbol{P}), \tag{3.33}$$

where $\boldsymbol{R} = \boldsymbol{Y}'\boldsymbol{Y} + \boldsymbol{Y}_f'\boldsymbol{Y}_f - \boldsymbol{P}'\boldsymbol{M}\boldsymbol{P}$, $\boldsymbol{P} = \boldsymbol{M}^{-1}(\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{X}_f'\boldsymbol{Y}_f)$ and $\boldsymbol{M} = \boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}_f'\boldsymbol{X}_f$.

Applying this convenient form of $\boldsymbol{Q}$ in (3.33) to equation (3.32) and then after integrating with respect to $\boldsymbol{\beta}$ by matrix-$T$ integral, the prediction density can be expressed as

$$f(\boldsymbol{Y}_f|\boldsymbol{Y}) \propto |\boldsymbol{R}|^{-\frac{n+n^f-2}{2}}. \tag{3.34}$$

It is noted that $\boldsymbol{R}$ is free from the unknown regression matrix $\boldsymbol{\beta}$.

To simplify the expression of $\boldsymbol{R}$ into a convenient quadratic form for a prediction distribution of $\boldsymbol{Y}_f$, the quadratic form in $\boldsymbol{Y}_f$ can be established by the following way

$$\begin{aligned}
\boldsymbol{R} &= \boldsymbol{Y}'\boldsymbol{Y} + \boldsymbol{Y}_f'\boldsymbol{Y}_f - \boldsymbol{P}'\boldsymbol{M}\boldsymbol{P} \\
&= \boldsymbol{Y}'\boldsymbol{Y} + \boldsymbol{Y}_f'\boldsymbol{Y}_f - [\boldsymbol{M}^{-1}(\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{X}_f'\boldsymbol{Y}_f)]'\boldsymbol{M}[\boldsymbol{M}^{-1}(\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{X}_f'\boldsymbol{Y}_f)] \\
&= \boldsymbol{Y}'\boldsymbol{Y} + \boldsymbol{Y}_f'\boldsymbol{Y}_f - (\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{X}_f'\boldsymbol{Y}_f)'\boldsymbol{M}^{-1}\boldsymbol{M}\boldsymbol{M}^{-1}(\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{X}_f'\boldsymbol{Y}_f) \\
&= \boldsymbol{Y}'\boldsymbol{Y} + \boldsymbol{Y}_f'\boldsymbol{Y}_f - (\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{X}_f'\boldsymbol{Y}_f)'\boldsymbol{M}^{-1}(\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{X}_f'\boldsymbol{Y}_f) \\
&= \boldsymbol{Y}'\boldsymbol{Y} + \boldsymbol{Y}_f'\boldsymbol{Y}_f - \boldsymbol{Y}'\boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{X}'\boldsymbol{Y} - \boldsymbol{Y}_f'\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}_f'\boldsymbol{Y}_f \\
&\quad - \boldsymbol{Y}'\boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{X}_f'\boldsymbol{Y}_f - \boldsymbol{Y}_f'\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}'\boldsymbol{Y}
\end{aligned}$$

$$\begin{aligned}
&= \boldsymbol{Y}'(\boldsymbol{I}-\boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{X}')\boldsymbol{Y}+\boldsymbol{Y}_f'(\boldsymbol{I}-\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}_f')\boldsymbol{Y}_f \\
&\quad -\boldsymbol{Y}'\boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{X}_f'\boldsymbol{Y}_f-\boldsymbol{Y}_f'\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}'\boldsymbol{Y} \\
&= \boldsymbol{Y}'(\boldsymbol{I}-\boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{X}')\boldsymbol{Y}+\boldsymbol{Y}_f'\boldsymbol{H}\boldsymbol{Y}_f-\boldsymbol{Y}'\boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{X}_f'\boldsymbol{Y}_f \\
&\quad -\boldsymbol{Y}_f'\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}'\boldsymbol{Y} \qquad (3.35)
\end{aligned}$$

where

$$\boldsymbol{H}=\boldsymbol{I}-\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}_f'.$$

As shown by Zellner (1971, p.235)

$$\boldsymbol{H}^{-1}=(\boldsymbol{I}-\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}_f')^{-1}=\boldsymbol{I}+\boldsymbol{X}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_f'$$

which can be verified by the following matrix multiplication

$$\begin{aligned}
\boldsymbol{H}\boldsymbol{H}^{-1} &= (\boldsymbol{I}-\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}_f')[\boldsymbol{I}+\boldsymbol{X}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_f'] \\
&= \boldsymbol{I}-\boldsymbol{X}_f[\boldsymbol{M}^{-1}-(\boldsymbol{X}'\boldsymbol{X})^{-1}+\boldsymbol{M}^{-1}\boldsymbol{X}_f'\boldsymbol{X}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}]\boldsymbol{X}_f' \\
&= \boldsymbol{I}-\boldsymbol{X}_f\boldsymbol{M}^{-1}[\boldsymbol{X}'\boldsymbol{X}-\boldsymbol{M}+\boldsymbol{X}_f'\boldsymbol{X}_f](\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_f' \\
&= \boldsymbol{I} \qquad (3.36)
\end{aligned}$$

since $\boldsymbol{X}'\boldsymbol{X}-\boldsymbol{M}+\boldsymbol{X}_f'\boldsymbol{X}_f=\boldsymbol{0}$, by the definition of $\boldsymbol{M}=\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_f'\boldsymbol{X}_f$.

Therefore, $\boldsymbol{R}$ can be expressed as

$$\boldsymbol{R}=\boldsymbol{Y}'[\boldsymbol{I}-Z_1(\boldsymbol{M},\boldsymbol{H})]\boldsymbol{Y}+[\boldsymbol{Y}_f-Z_2(\boldsymbol{M},\boldsymbol{H})]'\boldsymbol{H}[\boldsymbol{Y}_f-Z_2(\boldsymbol{M},\boldsymbol{H})] \quad (3.37)$$

where

$$Z_1(\boldsymbol{M},\boldsymbol{H})=\boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{X}'+\boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{X}_f'\boldsymbol{H}^{-1}\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}'$$

and

$$Z_2(\boldsymbol{M},\boldsymbol{H})=\boldsymbol{H}^{-1}\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}'\boldsymbol{Y}.$$

Now the following relationships (see. e.g., Zellner, 1971, p.235) can be expressed as

$$
\begin{aligned}
Z_1(\boldsymbol{M},\boldsymbol{H}) &= \boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{X}' + \boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{X}_f'\boldsymbol{H}^{-1}\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}' \\
&= \boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{X}' + \boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{X}_f'[\boldsymbol{I} + \boldsymbol{X}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_f']\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}' \\
&= \boldsymbol{X}\boldsymbol{M}^{-1}[\boldsymbol{X}' + \boldsymbol{X}_f'\{\boldsymbol{I} + \boldsymbol{X}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_f'\}\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}'] \\
&= \boldsymbol{X}\boldsymbol{M}^{-1}[\boldsymbol{X}' + \boldsymbol{X}_f'\boldsymbol{X}_f\{\boldsymbol{M}^{-1}\boldsymbol{X}' + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_f'\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}'\}] \\
&= \boldsymbol{X}\boldsymbol{M}^{-1}[\boldsymbol{X}' + \boldsymbol{X}_f'\boldsymbol{X}_f\boldsymbol{F}] \qquad (3.38)
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{F} &= \boldsymbol{M}^{-1}\boldsymbol{X}' + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_f'\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}' \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{M}^{-1}\boldsymbol{X}' + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_f'\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}' \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}[\boldsymbol{X}'\boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{X}' + \boldsymbol{X}_f'\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}'] \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}[\{\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}_f'\boldsymbol{X}_f\}\boldsymbol{M}^{-1}\boldsymbol{X}'] \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}[\boldsymbol{M}\boldsymbol{M}^{-1}\boldsymbol{X}'] \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'
\end{aligned}
$$

Using this value of $\boldsymbol{F}$ to equation (3.38), $Z_1(\boldsymbol{M},\boldsymbol{H})$ is obtained as

$$
\begin{aligned}
Z_1(\boldsymbol{M},\boldsymbol{H}) &= \boldsymbol{X}\boldsymbol{M}^{-1}[\boldsymbol{X}' + \boldsymbol{X}_f'\boldsymbol{X}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'] \\
&= \boldsymbol{X}\boldsymbol{M}^{-1}[(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' + \boldsymbol{X}_f'\boldsymbol{X}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'] \\
&= \boldsymbol{X}\boldsymbol{M}^{-1}[\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}_f'\boldsymbol{X}_f](\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \\
&= \boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{M}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \\
&= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'. \qquad (3.39)
\end{aligned}
$$

And also

$$
Z_2(\boldsymbol{M},\boldsymbol{H}) = \boldsymbol{H}^{-1}\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}'\boldsymbol{Y}
$$

$$\begin{aligned}
&= [\boldsymbol{I} + \boldsymbol{X}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_f']\boldsymbol{X}_f\boldsymbol{M}^{-1}\boldsymbol{X}'\boldsymbol{Y} \\
&= \boldsymbol{X}_f[\boldsymbol{I} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_f'\boldsymbol{X}_f]\boldsymbol{M}^{-1}\boldsymbol{X}'\boldsymbol{Y} \\
&= \boldsymbol{X}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}_f'\boldsymbol{X}_f)\boldsymbol{M}^{-1}\boldsymbol{X}'\boldsymbol{Y} \\
&= \boldsymbol{X}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{M}\boldsymbol{M}^{-1}\boldsymbol{X}'\boldsymbol{Y} \\
&= \boldsymbol{X}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}. \qquad (3.40)
\end{aligned}$$

Applying the relationships in (3.39) and in (3.40) to the equation (3.37), $\boldsymbol{R}$ can be expressed as

$$\begin{aligned}
\boldsymbol{R} &= \boldsymbol{Y}'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{Y} \\
&+ [\boldsymbol{Y}_f - \boldsymbol{X}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}]'\boldsymbol{H}[\boldsymbol{Y}_f - \boldsymbol{X}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}] \\
&= \boldsymbol{Y}'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{Y} + (\boldsymbol{Y}_f - \boldsymbol{X}_f\hat{\boldsymbol{\beta}})'\boldsymbol{H}(\boldsymbol{Y}_f - \boldsymbol{X}_f\hat{\boldsymbol{\beta}}) \qquad (3.41)
\end{aligned}$$

where $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ is the OLS of $\boldsymbol{\beta}$.

Again, since $\boldsymbol{Y}'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{Y} = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{S}_Y$, the expression in (3.41) has the following form

$$\boldsymbol{R} = \boldsymbol{S}_Y + (\boldsymbol{Y}_f - \boldsymbol{X}_f\hat{\boldsymbol{\beta}})'\boldsymbol{H}(\boldsymbol{Y}_f - \boldsymbol{X}_f\hat{\boldsymbol{\beta}}), \qquad (3.42)$$

Using this relation of $\boldsymbol{R}$ in (3.42) to the equation (3.34), the prediction distribution of the future responses matrix $\boldsymbol{Y}_f$, conditional on a set of realized responses matrix $\boldsymbol{Y}$, is finally obtained as

$$f(\boldsymbol{Y}_f|\boldsymbol{Y}) = C(\boldsymbol{Y}, \boldsymbol{H})\left[\boldsymbol{S}_Y + (\boldsymbol{Y}_f - \boldsymbol{X}_f\hat{\boldsymbol{\beta}})'\boldsymbol{H}(\boldsymbol{Y}_f - \boldsymbol{X}_f\hat{\boldsymbol{\beta}})\right]^{-\frac{n+n^f-2}{2}} \qquad (3.43)$$

where $C(\boldsymbol{Y}, \boldsymbol{H})$ is the normalizing constant which is given by

$$C(\boldsymbol{Y}, \boldsymbol{H}) = \frac{(\pi)^{-\frac{pn^f}{2}} \Gamma_p\left(\frac{n-2}{2}\right) |\boldsymbol{H}|^{-\frac{p}{2}}}{\Gamma_p\left(\frac{n+n^f-2}{2}\right) |\boldsymbol{S}_Y|^{\frac{n-2}{2}}}. \tag{3.44}$$

It is clear that the above density in (3.43) is the probability density function of an $n^f p$-dimensional matrix-$T$ with location matrix $\boldsymbol{X}_f\hat{\boldsymbol{\beta}}$, scale factors $\boldsymbol{S}_Y$ and $\boldsymbol{H}$ and degrees of freedom $(n-p-1)$. Thus, the matrix of future responses $Y_f$, has a matrix-$T$ distribution, and hence the the location of the prediction distribution is $\boldsymbol{X}_f\hat{\boldsymbol{\beta}}$ and the covariance matrix is given by $\text{Cov}(\boldsymbol{Y}_f|\boldsymbol{Y}) = \frac{(n-p-1)}{n-p-3}[\boldsymbol{S}_Y \otimes \boldsymbol{H}]$, where $\otimes$ denotes the Kronecker product.

Note $\boldsymbol{Y}_f|\boldsymbol{Y} \sim T_{n^f p}(\boldsymbol{X}_f\hat{\boldsymbol{\beta}}, \boldsymbol{S}_Y, H, \omega = n-p-1)$, that is, $E(\boldsymbol{Y}_f|\boldsymbol{Y}) = \boldsymbol{X}_f\hat{\boldsymbol{\beta}}$, $Cov(\boldsymbol{Y}_f|\boldsymbol{Y}) = \frac{(n-p-1)}{n-p-3}[\boldsymbol{S}_Y \otimes \boldsymbol{H}]$ and the degrees of freedom (df) $\omega = n-p-1$.

## 3.6 Concluding remarks

In this chapter, the prediction distribution for future responses matrix $\boldsymbol{Y}_f$, conditional on an observed responses matrix $\boldsymbol{Y}$ for the multivariate simple regression model with matrix-$T$ errors has been derived by the Bayesian method under uniform prior distribution. The results reveal that the prediction distribution of the future responses matrix is a matrix-$T$ distribution with $n-p-1$ degrees of freedom, location matrix $\boldsymbol{X}_f(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ and scale factors $\boldsymbol{S}_Y$ and $\boldsymbol{H}$. The shape parameter or degrees of freedom of the prediction distribution of the future responses matrix depends on the size of the realized sample and the number of the regression parameters of the model, but it does not depend on the degrees of freedom of the errors matrix distribution. In addition, the prediction distribution depends on the observed responses matrix $\boldsymbol{Y}$ only through the sample regression matrix and the sample residual sum of squares and products matrices as well as the design matrices $\boldsymbol{X}$ and $\boldsymbol{X}_f$ of both the realized and future multivariate simple regression model.

By the properties of the matrix-$T$ distribution, the distribution of individual element, the marginal distribution of any column (or row), the conditional distribution of one column (or row), given another, and the marginal as well as conditional distributions of any sub-matrix, and one sub-matrix, given another, of the future responses matrix $\boldsymbol{Y}_f$ can be determined directly from the parameters of the matrix-$T$ distribution of $\boldsymbol{Y}_f$. Also it is interesting to note that, an individual element of $\boldsymbol{Y}_f$ has a univariate Student-$t$ distribution, the marginal distribution of any column (or row) as well as the conditional distribution of one column (or row), given another, follows a multivariate Student-$t$ distribution and the marginal as well as the conditional distributions of any sub-matrix of $\boldsymbol{Y}_f$ and one sub-matrix, given another, follow matrix-$T$ distribution with appropriate parameters.

# Chapter 4

# Multivariate Multiple Regression Model

## 4.1 Introduction

The multivariate multiple regression model is an extension of the multiple linear regression model in a multivariate or matrix-variate setup. This model represents the linear relationship between a set of values of several response variables and a single set of values of some explanatory or predictor variables. Let $\boldsymbol{Y}_1, \boldsymbol{Y}_2, ..., \boldsymbol{Y}_n$ be a set of vectors of $n$ responses and $x_1, x_2, ..., x_{k-1}$ be a single set of $k-1$ predictor variables. Then a relation between these set of vector responses and the single set of values of predictor variables can be written as a multivariate multiple regression model when each of the $n$ responses is assumed to follow its own multiple linear regression model.

The multivariate multiple regression model is used to analyze data from different situations in econometrics as well as in many other experimental circumstances to deal with a set of linear regression equations. There are many experimental situations in real life where we need to study a set of responses from more than one dependent variables corresponding to a set of values from several independent variables. For example, in an agricultural farm, if several varieties in different blocks are given a single set of a treatments (such as kinds of fertilizers, water, spray, medicine etc.) to observe any responses to the products, then for a set of values of the predictor variables, there will be several set of values of the response variables from different products. More about this model can be found in Tiao and Zellner (1964).

Kibria (2006) considers the multivariate model and derives the prediction distribution of future responses and regression matrices by the structural relation approach introduced by Fraser and Haq (1969). Haq (1982) used their developed method, and Guttman and Hougaard (1985) considered the classical approach to obtain the prediction distribution from the multivariate multiple regression model with independent and normal errors distribution. Also, from the Bayesian point of view, the multivariate regression model under independent and normal errors assumption have been studied by Tiao and Zellner (1964), Geisser (1965), Zellner (1971), Kibria et al. (2002) and Khan (2006).

Furthermore, in the case of multiple regression model, the matrix$T$ error has been considered by Khan and Haq (1994), Kibria and Haq (2000)and Khan (2002). Khan and Haq (1994b) derived the prediction distribution of the future responses matrix by the structural relation model, and Khan (2002) obtained the prediction distribution of regression matrix using the same method. In both cases they obtained the prediction distribution as matrix-$T$ distribution with appropriate parameters.

The multivariate multiple linear regression model is presented in this chapter. The general assumption that the errors matrix of the model has a matrix-$T$ distribution is considered. The Bayesian approach is applied to derive the prediction distribution for the future responses matrix. Since a prior distribution is an element of the Bayesian method and it can be obtained from previous studies, a non-informative or uniform prior distribution of the unknown parameters matrices is considered here (cf. Jeffreys, 1961). This non-informative prior has been used by many researchers such as Zellner (1971), Bernardo and Rueda (2002) and Rowe (2003). It is observed that by the Bayesian method the prediction distribution of the future responses matrix follows a matrix-$T$ distribution with appropriate degrees of freedom and location as well as scale factors.

The layout of this chapter is as follows. In Section 4.2 the multivariate multiple regression model with matrix-$T$ errors is introduced. A uniform prior distribution is defined as well as the posterior density is obtained in Subsection 4.2.1. The future multivariate simple regression model is defined in section 4.3. Section 4.4 derives the prediction distribution of the future responses matrix under the Bayesian method with uniform prior. Some concluding remarks are presented in Section 4.5.

## 4.2 The Model

Suppose $\boldsymbol{Y}_1, \boldsymbol{Y}_2, ..., \boldsymbol{Y}_n$ is a set of vectors of $n$ responses and $x_1, x_2, ..., x_{k-1}$ is a set of $k-1$ predictor variables. Also let each of the $n$ responses follows its own multiple linear regression model, as

$$\boldsymbol{Y}_j = \boldsymbol{\beta}_{0j} + \boldsymbol{\beta}_{1j} x_1 + \boldsymbol{\beta}_{2j} x_2 + ... + \boldsymbol{\beta}_{k-1j} x_{k-1} + \boldsymbol{e}_j, \text{ for } j = 1, 2, .., n$$

where, $\boldsymbol{Y}_j$ and $\boldsymbol{e}_j$ are the $j^{th}$ response and its associated error vectors respectively each of order $1 \times p$. $\boldsymbol{\beta}_{.j}$ is a $k \times 1$ dimensional regression parameters vector on the regression line of $j^{th}$ response. Moreover, $[x_1, x_2, ..., x_{k-1}]$ is a single set of values of the $k-1$ predictor variables. Assume $\boldsymbol{e}_j \sim t_p(\mathbf{0}, \boldsymbol{\Sigma}, \nu)$.

Let $x_{j.} = [x_{j1}, x_{j2}, ..., x_{j(k-1)}]$, $\boldsymbol{Y}_j = [y_{j1}, y_{j2}, ..., y_{jp}]'$ and $\boldsymbol{e}_j = [e_{j1}, e_{j2}, ..., e_{jp}]'$ denote the values of the explanatory, response and error variables respectively for the $j^{th}$ trial. Then there is an $n \times k$ order design matrix

$$\tilde{\boldsymbol{X}} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1(k-1)} \\ 1 & x_{21} & x_{22} & \cdots & x_{2(k-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{n(k-1)} \end{bmatrix}.$$

Also, the following matrix notations may consider

$$\tilde{\boldsymbol{\beta}} = \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0p} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{(k-1)1} & \beta_{(k-1)2} & \cdots & \beta_{(k-1)p} \end{bmatrix},$$

the unknown regression coefficients matrix of order $k \times p$,

$$\tilde{\boldsymbol{Y}} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{bmatrix},$$

the $n \times p$ order matrix of responses and

$$\tilde{\boldsymbol{E}} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{np} \end{bmatrix}$$

the $n \times p$ order errors matrix.

The multivariate multiple linear regression model can be expressed as

$$\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{E}} \tag{4.1}$$

where $\text{rank}(\tilde{\boldsymbol{X}}) = k$ and $n \geq k$.

Now, assume that the components of each row in $\tilde{\boldsymbol{E}}$ of the model are correlated, and jointly follow a multivariate Student-$t$ distribution. Since each row of the errors matrix $\tilde{\boldsymbol{E}}$ are uncorrelated with others, the covariance of the errors matrix becomes $\frac{\nu}{\nu-2}[\boldsymbol{\Sigma} \otimes I_n]$; where $\nu$ is the shape parameter or degrees of freedom of the matrix-$T$ distribution for the errors matrix, $\otimes$ denotes the Kronecker product between two matrices $\boldsymbol{\Sigma}$ and $I_n$ in which $\boldsymbol{\Sigma}$ is a $p \times p$ order positive definite symmetric matrix and $\boldsymbol{I}_n$ is an identity matrix of order $n \times n$. The errors matrix $\tilde{\boldsymbol{E}}$ has a matrix-$T$ distribution with the density

$$f(\tilde{\boldsymbol{E}}) \propto |\boldsymbol{I}_n^{-1}|^{\frac{p}{2}} |\boldsymbol{\Sigma}^{-1}|^{\frac{n}{2}} |\boldsymbol{I}_p + \boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{E}}'\tilde{\boldsymbol{E}}|^{-\frac{1}{2}(\nu+p+n-1)}. \tag{4.2}$$

Hence the responses matrix $\tilde{\boldsymbol{Y}}$ from the model (4.1) has also a matrix-$T$ distribution, that is, $\tilde{\boldsymbol{Y}} \sim T_{np}(\tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}}, \boldsymbol{I}_{n\times n}, \boldsymbol{\Sigma}_{p\times p}, \nu)$ with the following probability density function

$$f(\tilde{\boldsymbol{Y}}|\tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}, \nu) = \boldsymbol{C}\frac{|\boldsymbol{I}_n^{-1}|^{\frac{p}{2}}}{|\boldsymbol{\Sigma}|^{\frac{n}{2}}} |\boldsymbol{I}_p + \boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})'(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})|^{-\frac{1}{2}(\nu+p+n-1)} \tag{4.3}$$

where the normalizing constant

$$\boldsymbol{C} = \left[\Gamma\left(\frac{1}{2}\right)\right]^{pn} \frac{\Gamma_p\left[\frac{1}{2}(\nu+p-1)\right]}{\Gamma_p\left[\frac{1}{2}(\nu+p+n-1)\right]}$$

in which $\Gamma_p(.)$ is a generalized gamma function introduced by Siegel (1935), and defined in equation (3.1) in Section 3.2.

### 4.2.1 The prior and posterior distributions

Assume that the joint prior distribution of the regression matrix $\tilde{\boldsymbol{\beta}}$ and the $\frac{1}{2}p(p+1)$ distinct elements of $\boldsymbol{\Sigma}$ is noninformative or uniform. Also, assume that the elements $\tilde{\boldsymbol{\beta}}$ as well as that of $\boldsymbol{\Sigma}$ are independently distributed. Which means if $p(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma})$ is a joint prior density of $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}$, then

$$p(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}) = p(\tilde{\boldsymbol{\beta}})p(\boldsymbol{\Sigma}).$$

Adopting the invariance theory due to Jeffreys (1961, p. 179), this study consider

$$p(\tilde{\boldsymbol{\beta}}) = \text{constant},$$

and

$$p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{p+1}{2}}.$$

Thus, the joint prior density of unknown parameters matrices $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}$ becomes

$$p(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{p+1}{2}}. \tag{4.4}$$

This uniform prior distribution has been used by many researchers such as Zellner (1971), Bernardo and Rueda (2002), and Khan (2006).

Using the Bayes's theory described in Subsection 3.3.1, the joint posterior density function of $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}$ for the given responses matrix $\tilde{\boldsymbol{Y}}$ can be obtained as

$$p(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}|\tilde{\boldsymbol{Y}}) = \tilde{\boldsymbol{\Phi}}|\boldsymbol{\Sigma}|^{-\frac{n+p+1}{2}} \left|\boldsymbol{I}_p + \boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})'(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})\right|^{-\frac{\nu+p+n-1}{2}} \tag{4.5}$$

where the normalizing constant is given by

$$\tilde{\mathbf{\Phi}} = \frac{|\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}|^{\frac{p}{2}}|\tilde{\boldsymbol{S}}_Y|^{\frac{n-k}{2}}}{\left[\Gamma\left(\frac{1}{2}\right)\right]^{kp}} \frac{\Gamma_p\left(\frac{\nu+n+p-1}{2}\right)}{\Gamma_p\left(\frac{n-k}{2}\right)\Gamma_p\left(\frac{\nu+p-1}{2}\right)} \tag{4.6}$$

and this value of $\tilde{\mathbf{\Phi}}$ is evaluated in Appendix A.1.

## 4.3 The Future Model

Let $\tilde{\boldsymbol{Y}}_f$ be an unobserved future response matrix from the model provided in equation (4.1) to the $n^f \times k$ dimensional design matrix $\tilde{\boldsymbol{X}}_f$. Then the future multivariate multiple linear regression model for $\tilde{\boldsymbol{Y}}_f$ can be defined as

$$\tilde{\boldsymbol{Y}}_f = \tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{E}}_f \tag{4.7}$$

where $\tilde{\boldsymbol{\beta}}$ is a $k \times p$ dimensional regression parameters matrix for future responses; $\tilde{\boldsymbol{Y}}_f$ and $\tilde{\boldsymbol{E}}_f$ both are $n^f \times p$ dimensional response matrix and associated error matrix respectively. In the future model it is assumed that, $\tilde{\boldsymbol{E}}_f$ has an $n^f p$ dimensional matrix-$T$ distribution with $\nu$ degrees of freedom, which is expressed as

$$\tilde{\boldsymbol{E}}_f \sim T_{n^f p}(\mathbf{0}, \boldsymbol{I}_{n^f \times n^f}, \mathbf{\Sigma}_{p\times p}, \nu).$$

Since the elements in each row of the realized errors matrix $\tilde{\boldsymbol{E}}$ and the future errors matrix $\tilde{\boldsymbol{E}}_f$ are correlated, and $n$ rows in $\tilde{\boldsymbol{E}}$ as well as $n^f$ rows in $\tilde{\boldsymbol{E}}_f$ are uncorrelated, the joint density function of $\tilde{\boldsymbol{\beta}}$ and $\mathbf{\Sigma}$ for the realized responses matrix $\tilde{\boldsymbol{Y}}$ from the performed experiment and the unobserved future responses matrix $\tilde{\boldsymbol{Y}}_f$ from the future experiment can be expressed as

$$p(\tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{Y}}_f|\tilde{\boldsymbol{\beta}}, \mathbf{\Sigma}) \propto |\mathbf{\Sigma}|^{-\frac{n+n^f}{2}} \left|\boldsymbol{I}_p + \mathbf{\Sigma}^{-1}\tilde{\boldsymbol{Q}}\right|^{-\frac{\nu+p+n+n^f-1}{2}} \tag{4.8}$$

where

$$\tilde{\boldsymbol{Q}} = (\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})'(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}}) + (\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{\beta}})'(\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{\beta}}).$$

## 4.4 Prediction of future responses

Using the uniform prior density for $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}$ as given in equation (4.4) and the joint density function for the response matrices $\tilde{\boldsymbol{Y}}$ and $\tilde{\boldsymbol{Y}}_f$ in (4.8), the joint posterior density of unknown parameters $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}$ for the responses matrices $\tilde{\boldsymbol{Y}}$ and $\tilde{\boldsymbol{Y}}_f$ can be determined as

$$p(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}|\tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{Y}}_f) \propto |\boldsymbol{\Sigma}|^{-\frac{n+n^f+p+1}{2}} \left|\boldsymbol{I}_p + \boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{Q}}\right|^{-\frac{\nu+p+n+n^f-1}{2}}. \qquad (4.9)$$

Now the prediction distribution of a set of $n^f$ future responses can be obtained by solving the following integral

$$f(\tilde{\boldsymbol{Y}}_f|\tilde{\boldsymbol{Y}}) \propto \int_{\tilde{\boldsymbol{\beta}}}\int_{|\boldsymbol{\Sigma}|>\boldsymbol{0}} p(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}|\tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{Y}}_f) d\boldsymbol{\Sigma} d\tilde{\boldsymbol{\beta}}. \qquad (4.10)$$

or

$$f(\tilde{\boldsymbol{Y}}_f|\tilde{\boldsymbol{Y}}) \propto \int_{\tilde{\boldsymbol{\beta}}}\int_{|\boldsymbol{\Sigma}|>\boldsymbol{0}} |\boldsymbol{\Sigma}|^{-\frac{n+n^f+p+1}{2}} \left|\boldsymbol{I}_p + \boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{Q}}\right|^{-\frac{\nu+p+n+n^f-1}{2}} d\boldsymbol{\Sigma} d\tilde{\boldsymbol{\beta}}. \qquad (4.11)$$

Applying the appropriate matrix transformation $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Lambda}$, as considered in Chapter 3 (sec. 3.5) the probability density in (4.11) can be written as

$$f(\tilde{\boldsymbol{Y}}_f|\tilde{\boldsymbol{Y}}) \propto \int_{\tilde{\boldsymbol{\beta}}}\int_{|\boldsymbol{\Lambda}|>\boldsymbol{0}} |\boldsymbol{\Lambda}^{-1}|^{-\frac{n+n^f-p-1}{2}} \left|\boldsymbol{I}_p + \boldsymbol{\Lambda}\tilde{\boldsymbol{Q}}\right|^{-\frac{\nu+p+n+n^f-1}{2}} d\boldsymbol{\Lambda} d\tilde{\boldsymbol{\beta}}. \qquad (4.12)$$

or

$$f(\tilde{\boldsymbol{Y}}_f|\tilde{\boldsymbol{Y}}) \propto \int_{\tilde{\boldsymbol{\beta}}}\int_{|\boldsymbol{\Lambda}|>\boldsymbol{0}} |\boldsymbol{\Lambda}|^{\frac{n+n^f}{2}-\frac{p+1}{2}} \left|\boldsymbol{I}_p + \boldsymbol{\Lambda}\tilde{\boldsymbol{Q}}\right|^{-\left(\frac{n+n^f}{2}+\frac{\nu+p-1}{2}\right)} d\boldsymbol{\Lambda} d\tilde{\boldsymbol{\beta}}. \qquad (4.13)$$

By using the properties of the generalized beta integral for the matrix variables (cf. Khan, 2000), the following prediction density can be obtained from the equation in (4.13) after integrating with respective to $\boldsymbol{\Lambda}$

$$f(\tilde{\boldsymbol{Y}}_f|\tilde{\boldsymbol{Y}}) \propto \int_{\tilde{\boldsymbol{\beta}}} |\tilde{\boldsymbol{Q}}|^{-\frac{n+n^f}{2}} B_p\left(\frac{n+n^f}{2}, \frac{\nu+p-1}{2}\right) d\tilde{\boldsymbol{\beta}}, \tag{4.14}$$

or

$$f(\tilde{\boldsymbol{Y}}_f|\tilde{\boldsymbol{Y}}) \propto \int_{\tilde{\boldsymbol{\beta}}} |\tilde{\boldsymbol{Q}}|^{-\frac{n+n^f}{2}} d\tilde{\boldsymbol{\beta}}. \tag{4.15}$$

Now $\tilde{\boldsymbol{Q}}$ can be expressed as a quadratic form in $\tilde{\boldsymbol{\beta}}$ by the following way

$$\tilde{\boldsymbol{Q}} = \tilde{\boldsymbol{R}} + (\tilde{\boldsymbol{\beta}} - \boldsymbol{P})' M (\tilde{\boldsymbol{\beta}} - \boldsymbol{P}), \tag{4.16}$$

where $\tilde{\boldsymbol{R}} = \tilde{\boldsymbol{Y}}'\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{Y}}_f'\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{P}}'\tilde{\boldsymbol{M}}\tilde{\boldsymbol{P}}$, $\tilde{\boldsymbol{P}} = \tilde{\boldsymbol{M}}^{-1}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{Y}}_f)$ and $\tilde{\boldsymbol{M}} = \tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f$.

Using this representation of $\tilde{\boldsymbol{Q}}$ in (4.16) to equation (4.15) and then integrating with respect to $\tilde{\boldsymbol{\beta}}$ by matrix-$T$ integral, the prediction distribution of the future responses matrix is obtained as

$$f(\tilde{\boldsymbol{Y}}_f|\tilde{\boldsymbol{Y}}) \propto |\tilde{\boldsymbol{R}}|^{-\frac{n+n^f-k}{2}} \tag{4.17}$$

where $\tilde{\boldsymbol{R}}$ is free from the unknown parameters.

For the completion of the derivation $f(\tilde{\boldsymbol{Y}}_f|\tilde{\boldsymbol{Y}})$, the prediction density of the unobserved future responses matrix $\tilde{\boldsymbol{Y}}_f$, given the realized responses matrix $\tilde{\boldsymbol{Y}}$, the relation $\tilde{\boldsymbol{R}} = \tilde{\boldsymbol{Y}}'\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{Y}}_f'\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{P}}'\tilde{\boldsymbol{M}}\tilde{\boldsymbol{P}}$ in (4.16) can be expressed as the following quadratic from of the future responses matrix $\tilde{\boldsymbol{Y}}_f$

$$\tilde{\boldsymbol{R}} = \tilde{\boldsymbol{S}}_Y + (\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{X}}_f\hat{\tilde{\boldsymbol{\beta}}})'\tilde{\boldsymbol{H}}(\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{X}}_f\hat{\tilde{\boldsymbol{\beta}}}) \tag{4.18}$$

where $\hat{\tilde{\boldsymbol{\beta}}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}}$ is the OLS of $\tilde{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{H}} = \tilde{\boldsymbol{I}} - \tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'_f$ and $\tilde{\boldsymbol{S}}_Y = (\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\hat{\tilde{\boldsymbol{\beta}}})'(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\hat{\tilde{\boldsymbol{\beta}}})$. See, Appendix A.2 for the derivation of the quadratic form of $\tilde{\boldsymbol{Y}}_f$, which follows somewhat similar operational steps as described in Chapter 3 (sec. 3.5).

Applying the expression of $\tilde{\boldsymbol{R}}$ in (4.18) to (4.17), finally the prediction distribution of the future responses matrix $\tilde{\boldsymbol{Y}}_f$, conditional on the realized responses matrix $\tilde{\boldsymbol{Y}}$, is obtained as

$$f(\tilde{\boldsymbol{Y}}_f|\tilde{\boldsymbol{Y}}) = C(\tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{H}}) \left[\boldsymbol{S}_Y + (\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{X}}_f\hat{\tilde{\boldsymbol{\beta}}})' H (\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{X}}_f\hat{\tilde{\boldsymbol{\beta}}})\right]^{-\frac{n+n^f-k}{2}} \tag{4.19}$$

where the normalizing constant is given by

$$C(\tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{H}}) = \frac{(\pi)^{-\frac{n^f p}{2}} \Gamma_p\left(\frac{n-k}{2}\right) |\tilde{\boldsymbol{H}}|^{-\frac{p}{2}}}{\Gamma_p\left(\frac{n+n^f-k}{2}\right) |\tilde{\boldsymbol{S}}_Y|^{\frac{n-k}{2}}}. \tag{4.20}$$

The density in (4.19) is the probability density function of a $n^f p$-dimensional matrix-$T$ with location matrix $\tilde{\boldsymbol{X}}_f\hat{\tilde{\boldsymbol{\beta}}}$, scale factors $\tilde{\boldsymbol{S}}_Y$, $\tilde{\boldsymbol{H}}$ and shape parameter $n-p-k+1$. Therefore, the future responses matrix $\tilde{\boldsymbol{Y}}_f$ for the multivariate multiple regression model has a $n^f p$ dimensional matrix-$T$ distribution. The location of the prediction distribution is $\tilde{\boldsymbol{X}}_f(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}}$ and the covariance matrix is $\frac{(n-p-k+1)}{(n-p-k-1)}[\tilde{\boldsymbol{S}}_Y \otimes \tilde{\boldsymbol{H}}]$, where $\otimes$ represents the Kronecker product of the matrices $\tilde{\boldsymbol{S}}_Y$ and $\tilde{\boldsymbol{H}}$.

Moreover, from the properties of the matrix-$T$ distribution, the marginal prediction distribution of any column or row vector of the future responses matrix $\tilde{\boldsymbol{Y}}_f$ is a multivariate Student-$t$ distribution with appropriate parameters. An individual observation of the future responses matrix follows a prediction distribution as a univariate Student-$t$ distribution, and any sub-matrix of $\tilde{\boldsymbol{Y}}_f$ has a prediction distribution as a matrix-$T$ distribution with appropriate location, scale and shape parameters. It is also noted that, the prediction distributions of any sub-matrix, column (or row) vector and individual element of the future responses matrix $\tilde{\boldsymbol{Y}}_f$ can be obtained directly from the parameters of the matrix-$T$ distribution for the future responses matrix $\tilde{\boldsymbol{Y}}_f$.

## 4.5 Concluding remarks

The prediction distribution of future responses matrix for the multivariate multiple regression model with matrix-$T$ errors has been derived in this chapter. The Bayesian method under uniform prior is considered here to obtain the prediction distribution. It has been seen that the prediction distribution of the matrix of future responses $\tilde{\boldsymbol{Y}}_f$, conditional on the realized responses matrix $\tilde{\boldsymbol{Y}}$, is a matrix-$T$ distribution with appropriate shape, location and scale parameters. The shape parameter of the prediction distribution of the future responses matrix depends on the size of the realized sample and the number of the regression parameters of the model. However, the shape parameter does not depend on the degrees of freedom of the distribution of errors or responses matrix. The prediction distribution of the future responses matrix depends on the observed responses matrix $\tilde{\boldsymbol{Y}}$ and both of the design matrices $\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{X}}_f$ of the realized and future models. Furthermore, it is noted that the marginal prediction distributions of any sub-matrix, any column or row vector and an individual observation of the future responses matrix $\tilde{\boldsymbol{Y}}_f$ are a matrix-$T$, a multivariate Student-$t$ and a univariate Student-$t$ distributions respectively with appropriate location, scale and shape parameters; and all of these marginal distributions can be obtained directly from the prediction distribution of $\tilde{\boldsymbol{Y}}_f$.

# Chapter 5

# Conclusions

In this thesis, the multiple, multivariate simple and multivariate multiple regression models with Student-$t$ errors are considered. The Bayesian approach under the noninformative prior distribution has been employed to derive the prediction distribution of the future responses for these models. It is revealed that the prediction distributions of a single future response and a set of future responses, conditional on a set of observed responses for the multiple linear regression model are univariate Student-$t$ distribution and multivariate Student-$t$ distribution respectively with appropriate parameters. Results have shown that these prediction distributions coincide with the results as derived by other statistical methods such as the classical, structural distributions, structural relations of the model as well as for the normal errors model. Therefore, the prediction distribution for the multiple regression model with multivariate Student-$t$ error distribution as well as independent and normal errors under the Bayesian approach with uniform prior and the classical, structural relation as well as structural distribution approaches is the same. Furthermore, the shape parameter of the prediction distribution depends on the size of the realized sample and the dimension of the parameters vector of the model. However the shape parameter of the prediction distribution does not depend on the degrees of freedom of the errors distribution.

Moreover, an special case of multiple regression model, the simple linear regression model has been also considered in this dissertation. Results demonstrate that, a single and a set of future response(s) for the simple regression model with multivariate-$t$ errors have a univariate and multivariate Student-$t$ distribution respectively with $n-2$ degrees of freedom and appropriate location as well as scale parameters.

The multivariate simple regression model with matrix-$T$ errors distribution has been defined in this study, and the prediction distribution of the future responses matrix, conditional on an observed responses matrix for the model is derived. It has been shown that by the Bayesian approach under the uniform prior distribution, the prediction distribution of the future responses matrix for the multivariate regression model is a matrix-$T$ distribution with appropriate location, scale and shape parameters. In addition, the prediction distribution depends on the observed responses matrix $\boldsymbol{Y}$ only through the sample regression matrix and the sample residual sum of squares and products matrices as well as the design matrix $\boldsymbol{X}_f$ of the future multivariate simple regression model.

Finally, we have also derived the prediction distribution for the future responses matrix, conditional on an observed responses matrix for the multivariate multiple regression model with matrix-$T$ errors distribution. The same Bayesian method is employed here that has been considered for the multivariate simple regression model to obtain the prediction distribution. The results of the derivations reveal that the prediction distribution of the matrix of future responses for the multivariate multiple regression model is also a matrix-$T$ distribution with appropriate shape, location and scale parameters. Like the prediction distribution for the multivariate simple regression model, the shape parameter of the prediction distribution of the future responses matrix for the multivariate multiple regression model depends on the size of the realized sample and the number of the regression parameters of the model. However, the shape parameter does not depend on the degrees of freedom of the distribution of errors matrix. In addition, the prediction distribution of the future responses matrix depends on the observed responses matrix only through the sample regression matrix and the sample residual sum of squares and products matrix, and it also depends on the design matrices of the realized and future regression models. It is noted that the marginal prediction distributions of any sub-matrix, any column or row vector and an individual response observation of the future responses matrix for the multivariate simple as well as multiple regression models are a matrix-$T$, a multivariate Student-$t$ and a univariate Student-$t$ distributions respectively with appropriate location, scale and shape parameters; and all of these marginal distributions can be obtained directly from the parameters of the prediction distributions of the future responses matrices for these two models.

A number of possibilities for future work into the Bayesian predictive inference for regression models are described below.

1. This thesis considered the noninformative or uniform prior distribution of unknown parameters to derive prediction distribution of future response(s)

for the multiple, multivariate simple and multiple regression models. Further work should consider another type of prior distribution such as conjugate prior, mixture prior, shrinkage prior, or Stein's prior distribution.

2. The Bayesian statistical inference for these linear models, in particular the Bayesian optimal tolerance regions or beta-expectation tolerance regions for the future responses may be addressed in the future research.

3. This research considered the multiple regression model with multivariate Student-$t$ errors where the elements of the error vector are uncorrelated but not independent with others. What will be the prediction distribution of future response(s) for correlated error components?

4. In the multivariate models this study considered the assumption that within the row the elements of error vectors are correlated and between rows the error vectors are uncorrelated but not independent. Further study should check other assumptions to obtain the prediction distribution by the Bayesian approach.

5. In econometrics and many other practical situations in real life the dynamic linear regression model is of interest. The prediction distributions for the dynamic regression models should be studied by the Bayesian method in further research.

# Bibliography


itchison, J. and Dunsmore, I.R. (1975). Statistical prediction analysis, *Cambridge University Press*, Cambridge.

Aitchison, J. and Sculthorpe, D. (1965). Some problems of statistical prediction, *Biometrika*, **52**,469-83.

Bauwens, L., Lubrano, M. and Richard, J.F. (1999). Bayesian inference in dynamic econometric models, *Oxford University Press*, Oxford.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances, *Philosophical Transactions of the Royal Society of London*, **53**, 370-418.

Bernardo, J. and Rueda, R. (2002). Bayesian hypothesis testing: a reference approach, *International Statistical Review*, **70**, 351-372.

Box, G.E.P. and Tiao, G.C. (1992). Bayesian inference in statistical analysis, *John Wiley and Sons, Inc.*, New York.

Dickey, J.M. (1967). Matricvariate generalizations of the multivariate-$t$ distribution and the inverted multivariate-$t$ distribution, *The Annals of Mathematical Statistics*, **38**(2), 511-518.

Faulkenberry, G.D. (1973). A method of obtaining prediction intervals, *Journal of the American Statistical Association*, **68**, 433-435.

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

Fraser, D.A.S. and Haq, M.S. (1970). Inference and prediction for multilineal model, *Journal of Statistical Research*, **4**, 93-109.

Fraser, D.A.S. and Ng, K.W. (1980). Multivariate regression analysis with spherical error, *in V.P.R. Krishnaiah, (ed.)* Multivariate Analysis, *North-Holland Publishing Co., New York*, 369-386.

Geisser, S. (1965). Bayesian estimation in multivariate analysis, *The Annals of Mathematical Statistics*, **36**(1), 150-159.

Geisser, S. (1993). Predictive inference: An introduction, *Chapman and Hall*, London.

Goldberger, A.S. (1962). Best linear unbiased prediction in the generalized linear regression model, *Journal of the American Statistical Association*, **57**, 369-375.

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

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

Haq, M.S. (1982). Structural relation and prediction for multivariate models, *Statistische Hefte*, **23**, 218-228.

Haq, M.S. and Khan, S. (1990). Prediction distribution for a linear regression model with multivariate Student-$t$ error distribution, *Communications In Statistics -Theory and Methods*, **19**(12), 4705-4712.

Jeffreys, H. (1961). Theory of probability, 3rd Edn. *Oxford University*

*Press*, Oxford.

Khan, S. (1992). Predictive inference for multilineal models with Student-*t* errors, *Unpublished PhD Thesis*, Department of Statistical and Actuarial Sciences, The University of Western Ontario,Canada.

Khan, S. (2000). On the generalized beta integral with matrix argument, *Pak. Jou. Statist.*, Special Volume in honour of Saleh and Aly, **16** (2), 163-167.

Khan, S. (2002). Distribution of sum of squares and products for the generalized multilineal matrix-T model, *Journal of Multivariate Analysis*, **83**, 124-140.

Khan, S. (2004). Predictive distribution of regression vector and residual sum of squares for normal multiple regression model, *Communications In Statistics: Theory and Methods*, **33**(10), 2423-2441.

Khan, S. (2005). Estimation of parameters of the simple multivariate linear model with Student-*t* errors, *Journal of Statistical Research*, **39**(2), 79-94.

Khan, S. (2006). Prediction distribution of future regression and residual sum of squares matrices for multivariate simple regression model with correlated normal responses, *Journal of Applied Probability and Statistics*, **l**, 15-30.

Khan, S. and Haq, M.S. (1994b). Prediction distribution for the multilineal model with Matrix-T error: The case of spherical error distribution, *Journal of Applied Statistical Science*, **1**, 239-250.

Khan, S. and Haq, M.S. (1994). Prediction inference for multilineal model with errors having multivariate Student-*t* distribution and first-order autocorrelation structure, *Sankhya, Part B: The Indian Journal of Statistics*, **56**, 95-106.

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

Kibria, B.M.G. and Haq, M.S. (2000). The multivariate linear model with matrix-$t$ error variables, *Journal of Applied Statistical Science*, **9**, 266-277.

Kibria, B.M.G., Sun, L., Zidek, J.V. and Le, N. (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.

Loschi, R.H., Iglesian, P.L. and Valle, R.B.A. (2003). Predictivistic characterization of multivariate Student-$t$ model, *Journal of Multivariate Analysis*, **85**, 10-23.

Magnus, J. and Neudecker, H. (1988). Matrix differential calculus with applications in statistics and econometrics, *John Wiley and Sons, Inc.*, Chichester.

Phillips, P.C.B. (1989). Spherical matrix distributions and Cauchy quotients, *Statistics and Probability Letters*, **8**, 51-53.

Press, J. (1982). Applied multivariate analysis: using Bayesian and Frequentist methods of inference, *Robert E. Krieger Publishing Company*, Florida.

Press, J. (1986). Applied multivariate analysis, *Holt, Rinehart and Winston, Inc.*, New York.

Roberts, H.V. (1965). Probabilistic prediction, *Journal of the American*

*Statistical Association*, **60**(309), 50-62.

Rowe, D.B. (2003). Multivariate Bayesian statistics, *Chapman and Hall, CRC*, London.

Saleh, A.K.M.E. (2006). Theory of preliminary test and Stein-type estimation with application, *John Wiley and Sons, Inc.*, New York.

Siegel, C.L. (1935). Ueber die analyticche theorie der quadratischen formen, *Annals of Mathematics*, **36**, 527-606.

Sutradhar, B.C. and Ali, M.M. (1989). 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.

Tiao, G.C. and Zellner, A. (1964). On the Bayesian estimation of multivariate regression, *Journal of Royal Statistical Society*, Part. B, **26**(2), 277-285.

Wilson, A.L. (1967). An approach to simultaneous tolerance interval in regression, *The Annals of Mathematical Statistics*, **38**, 1536-40.

Zellner, A. (1971). An introduction to Bayesian inference in econometrics, *John Wiley and Sons, Inc., New York*.

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**(354), 400-405.

Zellner, A. and 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.

# Appendix A

# Appendix

## A.1

This section of appendix derives the normalizing constant of the joint posterior density function in (4.5) that has provided in (4.6).

The density in (4.5) can be written as

$$p(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}|\tilde{\boldsymbol{Y}}) = \tilde{\boldsymbol{\Phi}}|\boldsymbol{\Sigma}|^{-\frac{n+p+1}{2}} \left|\tilde{\boldsymbol{I}}_p + \boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})'(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})\right|^{-\frac{\nu+p+n-1}{2}} \tag{A.1}$$

where $\tilde{\boldsymbol{\Phi}}$ represents the normalizing constant.

To evaluate the value of $\tilde{\boldsymbol{\Phi}}$, the following characteristic of the probability density function is used

$$\begin{aligned} 1 &= \tilde{\boldsymbol{\Phi}} \int_{\tilde{\boldsymbol{\beta}}} \int_{\boldsymbol{\Sigma}} |\boldsymbol{\Sigma}|^{-\frac{n+p+1}{2}} \\ &\quad \times \left|\tilde{\boldsymbol{I}}_p + \boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})'(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})\right|^{-\frac{\nu+p+n-1}{2}} d\boldsymbol{\Sigma} d\tilde{\boldsymbol{\beta}} \end{aligned} \tag{A.2}$$

The value of $\tilde{\tilde{\boldsymbol{\Phi}}}$ is obtained by solving the following equation

$$\begin{aligned} \tilde{\boldsymbol{\Phi}}^{-1} &= \int_{\tilde{\boldsymbol{\beta}}} \int_{\boldsymbol{\Sigma}} |\boldsymbol{\Sigma}|^{-\frac{n+p+1}{2}} \\ &\quad \times \left|\tilde{\boldsymbol{I}}_p + \boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})'(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})\right|^{-\frac{\nu+p+n-1}{2}} d\boldsymbol{\Sigma} d\tilde{\boldsymbol{\beta}} \end{aligned} \tag{A.3}$$

Considering the matrix transformation $\mathbf{\Sigma}^{-1} = \mathbf{\Lambda}$ with the Jacobian of the transformation

$$|J| = \frac{d\mathbf{\Sigma}}{d\mathbf{\Lambda}} = |\mathbf{\Lambda}^{-1}|^{p+1},$$

in (A.3)

$$\begin{aligned}\tilde{\mathbf{\Phi}}^{-1} &= \int_{\tilde{\boldsymbol{\beta}}}\int_{\mathbf{\Lambda}} |\mathbf{\Lambda}^{-1}|^{-\frac{n-p-1}{2}} \left|\tilde{\boldsymbol{I}}_p + \mathbf{\Lambda}(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})'(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})\right|^{-\frac{\nu+p+n-1}{2}} d\mathbf{\Lambda} d\tilde{\boldsymbol{\beta}} \\ &= \int_{\tilde{\boldsymbol{\beta}}}\int_{\mathbf{\Lambda}} |\mathbf{\Lambda}|^{\frac{n}{2}-\frac{p+1}{2}} \\ &\quad \times \left|\tilde{\boldsymbol{I}}_p + \mathbf{\Lambda}(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})'(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})\right|^{-(\frac{n}{2}+\frac{\nu+p-1}{2})} d\mathbf{\Lambda} d\tilde{\boldsymbol{\beta}}. \end{aligned} \tag{A.4}$$

Using the generalized beta integral for the matrix variables discussed by Khan (2000) for integrating with respect to $\mathbf{\Lambda}$, the results is

$$\begin{aligned}\tilde{\mathbf{\Phi}}^{-1} &= \int_{\tilde{\boldsymbol{\beta}}} \left|(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})'(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})\right|^{-\frac{n}{2}} B_p\left(\frac{n}{2}, \frac{\nu+p-1}{2}\right) d\tilde{\boldsymbol{\beta}} \\ &= B_p\left(\frac{n}{2}, \frac{\nu+p-1}{2}\right) \\ &\quad \times \int_{\tilde{\boldsymbol{\beta}}} \left|\tilde{\boldsymbol{S}}_Y + (\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{X}}\hat{\tilde{\boldsymbol{\beta}}})'\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{X}}\hat{\tilde{\boldsymbol{\beta}}})\right|^{-\frac{n}{2}} d\tilde{\boldsymbol{\beta}} \end{aligned} \tag{A.5}$$

where, $\hat{\tilde{\boldsymbol{\beta}}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}}$ is the OLS of $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{S}}_Y = (\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\hat{\tilde{\boldsymbol{\beta}}})'(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\hat{\tilde{\boldsymbol{\beta}}})$ of order $p \times p$.

Equation (A.5) can be written as

$$\begin{aligned}\tilde{\mathbf{\Phi}}^{-1} &= B_p\left(\frac{n}{2}, \frac{\nu+p-1}{2}\right) |\tilde{\boldsymbol{S}}_Y|^{-\frac{n}{2}} \\ &\quad \times \int_{\tilde{\boldsymbol{\beta}}} \left|\tilde{\boldsymbol{I}}_p + \boldsymbol{S}_Y^{-1}(\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{X}}\hat{\tilde{\boldsymbol{\beta}}})'\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{X}}\hat{\tilde{\boldsymbol{\beta}}})\right|^{-\frac{\delta+k+p-1}{2}} d\tilde{\boldsymbol{\beta}} \end{aligned} \tag{A.6}$$

where $\delta = n - k - p + 1$.

Using the properties of the matrix-$T$ distribution for integrating with respect to $\tilde{\boldsymbol{\beta}}$ we have

$$\tilde{\mathbf{\Phi}}^{-1} = B_p\left(\frac{n}{2}, \frac{\nu+p-1}{2}\right) |\tilde{\boldsymbol{S}}_Y|^{-\frac{n}{2}} \frac{|\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}|^{-\frac{p}{2}}}{|\tilde{\boldsymbol{S}}_Y|^{-\frac{k}{2}}} \left[\Gamma\left(\frac{1}{2}\right)\right]^{kp} \frac{\Gamma_p\left(\frac{\delta+p-1}{2}\right)}{\Gamma_p\left(\frac{\delta+k+p-1}{2}\right)}$$

$$= B_p\left(\frac{n}{2},\frac{\nu+p-1}{2}\right)|\tilde{\boldsymbol{S}}_Y|^{-\frac{n-k}{2}}|\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}|^{-\frac{p}{2}}\left[\Gamma\left(\frac{1}{2}\right)\right]^{kp}\frac{\Gamma_p\left(\frac{\delta+p-1}{2}\right)}{\Gamma_p\left(\frac{\delta+k+p-1}{2}\right)}$$

$$= \left[\Gamma\left(\frac{1}{2}\right)\right]^{kp}\frac{|\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}|^{-\frac{p}{2}}}{|\tilde{\boldsymbol{S}}_Y|^{\frac{n-k}{2}}}\frac{\Gamma_p\left(\frac{\delta+p-1}{2}\right)}{\Gamma_p\left(\frac{\delta+k+p-1}{2}\right)}B_p\left(\frac{n}{2},\frac{\nu+p-1}{2}\right) \tag{A.7}$$

Considering the value of $\delta$ and also expressing $B_p\left(\frac{n}{2},\frac{\nu+p-1}{2}\right)$ as the generalized gamma function, the equation in (A.7) can be expressed as

$$\tilde{\boldsymbol{\Phi}}^{-1} = \left[\Gamma\left(\frac{1}{2}\right)\right]^{kp}|\tilde{\boldsymbol{S}}_Y|^{-\frac{n-k}{2}}|\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}|^{-\frac{p}{2}}\frac{\Gamma_p\left(\frac{n-k}{2}\right)}{\Gamma_p\left(\frac{n}{2}\right)}\frac{\Gamma_p\left(\frac{n}{2}\right)\Gamma_p\left(\frac{\nu+p-1}{2}\right)}{\Gamma_p\left(\frac{\nu+n+p-1}{2}\right)} \tag{A.8}$$

Therefore, the normalizing constant is obtained as

$$\tilde{\boldsymbol{\Phi}} = \frac{|\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}|^{\frac{p}{2}}|\tilde{\boldsymbol{S}}_Y|^{\frac{n-k}{2}}}{\left[\Gamma\left(\frac{1}{2}\right)\right]^{kp}}\frac{\Gamma_p\left(\frac{\nu+n+p-1}{2}\right)}{\Gamma_p\left(\frac{n-k}{2}\right)\Gamma_p\left(\frac{\nu+p-1}{2}\right)}. \tag{A.9}$$

# A.2

In this section, the expression of $\tilde{\boldsymbol{R}}$ in (4.18) that has been used to derive the prediction distribution of future responses for multivariate multiple regression model is derived.

The quadratic form of $\tilde{\boldsymbol{Y}}_f$ in (4.18) can be established by the following way.

Using $\tilde{\boldsymbol{P}} = \tilde{\boldsymbol{M}}^{-1}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{Y}})$, the relation $\tilde{\boldsymbol{R}}$ in equation (4.16) can be expressed as

$$\begin{aligned}
\tilde{\boldsymbol{R}} &= \tilde{\boldsymbol{Y}}'\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{Y}}_f'\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{P}}'\tilde{\boldsymbol{M}}\tilde{\boldsymbol{P}} \\
&= \tilde{\boldsymbol{Y}}'\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{Y}}_f'\tilde{\boldsymbol{Y}}_f - [\tilde{\boldsymbol{M}}^{-1}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{Y}}_f)]'\tilde{\boldsymbol{M}}[\tilde{\boldsymbol{M}}^{-1}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{Y}}_f)] \\
&= \tilde{\boldsymbol{Y}}'\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{Y}}_f'\tilde{\boldsymbol{Y}}_f - (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{Y}}_f)'\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}^{-1}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{Y}}_f) \\
&= \tilde{\boldsymbol{Y}}'\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{Y}}_f'\tilde{\boldsymbol{Y}}_f - (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{Y}}_f)'\tilde{\boldsymbol{M}}^{-1}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{Y}}_f)
\end{aligned}$$

$$
\begin{aligned}
&= \tilde{\boldsymbol{Y}}'\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{Y}}_f'\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{Y}}'\tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{Y}}_f'\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{Y}}_f \\
&\quad - \tilde{\boldsymbol{Y}}'\tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{Y}}_f'\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} \\
&= \tilde{\boldsymbol{Y}}'(\boldsymbol{I} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}')\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{Y}}_f'(\boldsymbol{I} - \tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}_f')\tilde{\boldsymbol{Y}}_f \\
&\quad - \tilde{\boldsymbol{Y}}'\tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{Y}}_f'\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} \\
&= \tilde{\boldsymbol{Y}}'(\boldsymbol{I} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}')\tilde{\boldsymbol{Y}} + \tilde{\boldsymbol{Y}}_f'\tilde{\boldsymbol{H}}\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{Y}}'\tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{Y}}_f \\
&\quad - \tilde{\boldsymbol{Y}}_f'\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} \qquad (A.10)
\end{aligned}
$$

where

$$\tilde{\boldsymbol{H}} = \boldsymbol{I} - \tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}_f'$$

and hence

$$\tilde{\boldsymbol{H}}^{-1} = (\boldsymbol{I} - \tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}_f')^{-1} = \boldsymbol{I} + \tilde{\boldsymbol{X}}_f(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}_f'$$

which can be verified by the following matrix multiplication (Zellner, 1971, p.235)

$$
\begin{aligned}
\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^{-1} &= (\boldsymbol{I} - \tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}_f')[\boldsymbol{I} + \tilde{\boldsymbol{X}}_f(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}_f'] \\
&= \boldsymbol{I} - \tilde{\boldsymbol{X}}_f[\tilde{\boldsymbol{M}}^{-1} - (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1} + \tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}]\tilde{\boldsymbol{X}}_f' \\
&= \boldsymbol{I} - \tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}[\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} - \tilde{\boldsymbol{M}} + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f](\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}_f' \\
&= \boldsymbol{I} \qquad (A.11)
\end{aligned}
$$

since $\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} - \tilde{\boldsymbol{M}} + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f = \boldsymbol{0}$, by the definition of $\tilde{\boldsymbol{M}} = \tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f$.

Now, $\tilde{\boldsymbol{R}}$ can be written with the following functional forms of $\tilde{\boldsymbol{M}}$ and $\tilde{\boldsymbol{H}}$

$$\tilde{\boldsymbol{R}} = \tilde{\boldsymbol{Y}}'[\boldsymbol{I} - Z_1(\tilde{\boldsymbol{M}}, \tilde{\boldsymbol{H}})]\tilde{\boldsymbol{Y}} + [\tilde{\boldsymbol{Y}}_f - Z_2(\tilde{\boldsymbol{M}}, \tilde{\boldsymbol{H}})]'\tilde{\boldsymbol{H}}[\tilde{\boldsymbol{Y}}_f - Z_2(\tilde{\boldsymbol{M}}, \tilde{\boldsymbol{H}})] \qquad (A.12)$$

where

$$Z_1(\tilde{\boldsymbol{M}}, \tilde{\boldsymbol{H}}) = \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}' + \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{H}}^{-1}\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'$$

and

$$Z_2(\tilde{\boldsymbol{M}},\tilde{\boldsymbol{H}}) = \tilde{\boldsymbol{H}}^{-1}\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}}.$$

Now the function $Z_1(\tilde{\boldsymbol{M}},\tilde{\boldsymbol{H}})$ can be expressed as its useful form by the following way

$$\begin{aligned}
Z_1(\tilde{\boldsymbol{M}},\tilde{\boldsymbol{H}}) &= \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}' + \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{H}}^{-1}\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}' \\
&= \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}' + \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}_f'[\boldsymbol{I} + \tilde{\boldsymbol{X}}_f(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}_f']\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}' \\
&= \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}[\tilde{\boldsymbol{X}}' + \tilde{\boldsymbol{X}}_f'\{\boldsymbol{I} + \tilde{\boldsymbol{X}}_f(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}_f'\}\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'] \\
&= \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}[\tilde{\boldsymbol{X}}' + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}' + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'] \\
&= \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}[\tilde{\boldsymbol{X}}' + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f\{\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}' + (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'\}] \\
&= \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}[\tilde{\boldsymbol{X}}' + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{F}}] \qquad \text{(A.13)}
\end{aligned}$$

where

$$\begin{aligned}
\tilde{\boldsymbol{F}} &= \tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}' + (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}' \\
&= (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}' + (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}' \\
&= (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}[\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}' + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'] \\
&= (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}[\{\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f\}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'] \\
&= (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}[\tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'] \\
&= (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'
\end{aligned}$$

Employing this value of $\tilde{\boldsymbol{F}}$ to equation (A.13), $Z_1(\tilde{\boldsymbol{M}},\tilde{\boldsymbol{H}})$ is obtained as

$$\begin{aligned}
Z_1(\tilde{\boldsymbol{M}},\tilde{\boldsymbol{H}}) &= \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}[\tilde{\boldsymbol{X}}' + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'] \\
&= \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}[(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}' + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'] \\
&= \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}[\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f](\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'
\end{aligned}$$

$$= \tilde{\boldsymbol{X}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{M}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'$$
$$= \tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'. \tag{A.14}$$

Also $Z_2(\tilde{\boldsymbol{M}}, \tilde{\boldsymbol{H}})$ can be expressed as the following form

$$\begin{aligned} Z_2(\tilde{\boldsymbol{M}}, \tilde{\boldsymbol{H}}) &= \tilde{\boldsymbol{H}}^{-1}\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} \\ &= [\boldsymbol{I} + \tilde{\boldsymbol{X}}_f(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}_f']\tilde{\boldsymbol{X}}_f\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} \\ &= \tilde{\boldsymbol{X}}_f[\boldsymbol{I} + (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f]\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} \\ &= \tilde{\boldsymbol{X}}_f(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \tilde{\boldsymbol{X}}_f'\tilde{\boldsymbol{X}}_f)\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} \\ &= \tilde{\boldsymbol{X}}_f(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} \\ &= \tilde{\boldsymbol{X}}_f(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}}. \end{aligned} \tag{A.15}$$

Applying the useful expressions in (A.14) and in (A.15) to the equation (A.12), $\tilde{\boldsymbol{R}}$ can be written as

$$\begin{aligned} \tilde{\boldsymbol{R}} &= \tilde{\boldsymbol{Y}}'[\boldsymbol{I} - \tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}']\tilde{\boldsymbol{Y}} \\ &\quad + [\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{X}}_f(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}}]'\tilde{\boldsymbol{H}}[\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{X}}_f(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}}] \\ &= \tilde{\boldsymbol{Y}}'[\boldsymbol{I} - \tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}']\tilde{\boldsymbol{Y}} + (\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{X}}_f\hat{\tilde{\boldsymbol{\beta}}})'\tilde{\boldsymbol{H}}(\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{X}}_f\hat{\tilde{\boldsymbol{\beta}}}) \end{aligned} \tag{A.16}$$

where $\hat{\tilde{\boldsymbol{\beta}}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}}$ is the OLS of $\tilde{\boldsymbol{\beta}}$.

Again, using the well known relation $\tilde{\boldsymbol{Y}}'[\boldsymbol{I} - \tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}']\tilde{\boldsymbol{Y}} = (\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\hat{\tilde{\boldsymbol{\beta}}})'(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\hat{\tilde{\boldsymbol{\beta}}}) = \tilde{\boldsymbol{S}}_Y$, the equation in (A.16) can be expressed as the following convenient quadratic form of $\tilde{\boldsymbol{Y}}_f$

$$\tilde{\boldsymbol{R}} = \tilde{\boldsymbol{S}}_Y + (\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{X}}_f\hat{\tilde{\boldsymbol{\beta}}})'\tilde{\boldsymbol{H}}(\tilde{\boldsymbol{Y}}_f - \tilde{\boldsymbol{X}}_f\hat{\tilde{\boldsymbol{\beta}}}). \tag{A.17}$$