Bayesian Prediction for the Multiple Linear Regression Model with First Order Auto-Correlation

by

Abdulkhaleg Ali Al-Baiyat

A Thesis Presented to the

FACULTY OF THE COLLEGE OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

MATHEMATICAL SCIENCES

December, 1998

Bayesian Prediction for the Multiple Linear Regression Model with First Order Auto-Correlation

BY

Abdulkhaleg Ali Al-Baiyat

A Thesis Presented to the

FACULTY OF THE COLLEGE OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

Mathematical Sciences

December, 1998

UMI Number: 1393200

KING FAHD UNIVERSITY OF PETROLEUM AND MINERALS
DHAHRAN 31261, SAUDI ARABIA

COLLEGE OF GRADUATE STUDIES

This thesis written by Abdulkhaleg Ali Al-Baiyat under the direction of his Thesis advisor and approved by his Thesis Committee, has been presented to and accepted by the Dean of the College of Graduate Studies, in partial fulfilment of the requirements for the degree of MASTER OF SCIENCE IN MATHEMATICS.

Thesis Committee

Dr. Walid S. Al-Sabah
Chairperson of the Thesis Committee

Dr. A.H. Joarder (Member)

Dr. Mohammad Iqbal (Member)

Dr. Taqi Al-Faraj (Member)

Dr. Walid S. Al-Sabah
Department Chairman

28-12-98
Dr. Abdullah M. Al-Shehri
Dean, College of Graduate Studies

December, 1998

Acknowledgment

After finishing this work, I must greatly thank my previous academic advisor, Dr. Shahjahan Khan, who suggested the research problem for my thesis. He was the first one with whom I worked on my thesis. He was very kind and helpful. Without him, I would not have been able to finish my thesis even after he left the university to Australia since I continued my contact with him by means of e-mail facility. I thank him again and again for everything.

I am also indebted to Dr. Walid Al-Sabah, the chairman of the Mathematical Sciences Department, who kindly accepted to be my new academic advisor. I thank him for the continuous suggestions and corrections he made on my thesis.

I also thank the other members of my thesis committee both Dr. Mohammad Iqbal and Dr. Anwarul Haque Joarder, from the Mathematical Sciences Department, and Dr. Tagi Al-Faraj from the Management and Marketing Department, for their help.

I must also thank Dr. Husain Al-Jamed, from the Electrical Engineering Department. He was the one who kindly helped me to understand the MatLab software that enabled me to work out the final result of my research problem by means of numerical methods.

I should not forget my friend Mr. Abdullah Al-Ramadan who advised me to make my thesis work in statistics.

My many thanks are to KFUPM for giving me the chance to continue my study to the MS level and for providing me with the excellent facilities in the use of the e-mail and the use of the Library.

I must thank every member in the Mathematical Sciences Department. I sincerely thank all those who taught me, in particular, Dr. Adnan Al-Shakhs, Dr. Mohamed Elgebeily, Prof. Ali Al-Daffa, Dr. Zayid Abdul Hadi, Dr. Mohamed Boudjelkha and all those, too many to be mentioned, who supported me and pushed me towards progress.

I must not, at the end, forget to thank and appreciate the Directorate of Education in the Eastern Province for reducing my teaching load from twenty-four classes a week to twelve. My many thanks also go to Al-Najah Secondary School, the school that I work in. I thank its principal, Mr. Jamaan Al-Ghamdii and his assistants both Mr. Saeed Al-Awaamii and Mr. Mohammed Al-Hubail. I thank also the teachers of mathematics in this school for their cooperation.

I fully thank everyone helped by way or another to make such a work successful.

I ASK ALLAH TO BLESS THEM ALL.

Al-Baiyat, Abdulkhaleg Ali
December, 1998

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

THESIS ABSTRACT

FULL NAME OF STUDENT: Al-Baiyat, Abdulkhaleg Ali

TITLE OF THESIS: Bayesian Prediction for the Multiple Linear Regression Model with First Order Auto-Correlation

MAJOR FIELD: Mathematics

DATE OF DEGREE: December, 1998

In many occasions, we are interested in making inference about the yet unobserved set of responses $\mathbf{Y}_r$, conditional on the observed part $\mathbf{Y}_s$. Using the Bayesian approach, the probability density function for $\mathbf{Y}_r$, conditional on $\mathbf{Y}_s$ is obtained as a general procedure to derive the prediction distributions. In this work, we find the predictive probability density function for a set of responses of the Multiple Linear Regression Model with First Order Auto-Correlation and compare our results with that available in the literature where different approaches, other than the Bayesian approach, were used.

MASTER OF SCIENCE DEGREE

KING FAHD UNIVERSITY OF PETTROLEUM AND MINERALS
Dhahran, Saudi Arabia.

December, 1998

خلاصة الرسالة

اسم الطالب الكامل : عبدالخالق علي البيات

عنوان الرسالة : تنبؤ بييز لنموذج الانحدار الخطي المتعدد ذو الارتباط الذاتي من الرتبة الأولى

حقل التخصص : الرياضيات

تاريخ الشهادة : ديسمبر ١٩٩٨م

في كثير من الأحيان ، نكون بحاجة لدراسة مجموعة من الوقائع المستقبلية Y_r تحت فرضية حدوث مجموعة أخرى Y_s من الوقائع في الماضي . وبوجه عام ، استخدم مدخل بييز لإيجاد دالة التنبؤ ؛ وهي دالة الاحتمال الشرطية لمجموعة الوقائع Y_r المشروطة بحدوث مجموعة الوقائع Y_s. وفي هذه الرسالة ، نوجد دالة التنبؤ لمجموعة من الوقائع المستقبلية لما يعرف بنموذج الانحدار الخطي المتعدد ذو الارتباط الذاتي من الرتبة الأولى ، ونقارن ذلك بنتائج استخدمت فيها مداخل أخرى غير مدخل بييز .

درجة الماجستير في العلوم الرياضية

جامعة الملك فهد للبترول والمعادن

الظهران ، المملكة العربية السعودية

ديسمبر ١٩٩٨م

Chapter 1

INTRODUCTION

In many circumstances, the main objective of statistical studies is to make predictions. In some instances, experience may be used to make predictions while in others it is not. However, either way may result in false expectation, but which one is more reliable? Or in other words, which one has a smaller chance of error?

Forecast methods are classified as qualitative and quantitative (see[1]). Qualitative forecast methods which may or may not incorporate the past data, are intuitive. They mainly depend on the forecaster himself since he does not explicitly specify how previous data is used. Others may not be able to reproduce the same expectation.

On the other hand, quantitative forecast methods are based on statistical or mathematical models. Forecasts are made once a model is chosen. So, such methods do not depend on the forecaster but rather on the specified model. Hence, the forecasts in these methods are reproducible.

The problem remains, in quantitative forecasting, on how to choose a model that fits well to a particular data under consideration. Of course, different methods have been suggested to enable researchers to choose the right model. This is known as problem modeling. To choose a model, things like simplicity of the model, accuracy and other factors have to be taken into account, (see[1]).

After a model is chosen, one must follow an appropriate procedure and a sound statistical approach to make a forecast. One approach may be easier to use in a particular problem than the other available approaches. In our work, however, we follow a particular statistical approach on a given model under consideration.

We start with the multiple regression model whose error term follows a first order auto correlation (FOAC). Then we use the Bayesian approach to find the prediction distribution. By "first order auto correlation," we mean the following. In some regression problems, it is found that the error terms follow a certain pattern. That is, the errors in different periods are correlated. Such type of correlation is known as auto-correlation. Many economic problems follow this scheme. Examples include what is known as time series problems in which the data usually occur in time ordered sequences. The simplest auto correlation model is the first order auto-correlation in which each of the errors in a specific period is correlated to the one preceding it as shown in the following equation

$$u_t = \rho u_{t-1} + \epsilon_t \tag{1.1}$$

where

u_t is the error at time t

u_{t-1} is the error at time (t-1)

ρ is known as the coefficient of auto-correlation

ϵ_t is independent random error.

Why should we have autocorrelated errors in a specific model and what are the main causes of their presence? Clear answers to these questions have been given, in an excellent research paper by Cochrane and Orcutt [6]. In their work, they discussed several things regarding this topic and they emphasized the argument that most of the current formulations deal with economic variables that are highly positively autocorrelated, as well as the errors involved. They also discussed the sources of such criteria of the error terms.

One source of auto-correlation is the incorrectness of the choice of forming the relationship between the economic variables. Errors of this type will be positively autocorrelated since the economic variables are positively autocorrelated. A second source of systematic errors that may arise, is the omission of variables, both economic and non economic, from the analysis. Important variables may be omitted because their importance is not realized. Omission of non important variables may not have a strong influence on errors individually. However, as a whole, they may have substantial effects on errors since such variables are highly autocorrelated.

In this work, the prediction distribution of a set of unobserved responses conditional on a set of the observed responses is developed using the Bayesian approach. We work out this prediction distribution for the normal linear multiple regression model with first order auto-correlation and compare that with results available in the literature.

In chapter two, we introduce the Bayesian approach for prediction. Then in chapter three, we state the general linear multiple regression model with normal errors. In chapter four, we present the problem under consideration which is the problem of prediction of the normal linear multiple regression model with first order auto-correlation. In chapter five, we give the analysis of the problem following the Bayesian approach and compare our result with that obtained by Khan [12] where he followed the structural relation approach. We also mention the superpopulation approach, and we finally end our discussion, in the last chapter, with a summary and a conclusion.

Chapter 2

PRELIMINARIES

To make an inference about a specific parameter, there are different statistical approaches. The classical approach, for example, is dominated by methods such as estimation and hypothesis testing. In this work, we pursue Bayesian approach. A brief discussion on this method is included below.

2.1 Nature of Bayesian Inference

The Bayesian approach, depends mainly on Bayes' theorem. Inferences made using this approach (see[4]) provide a statistical way of explicitly including the sample information and the prior information regarding the distribution of the model parameters.

Bayes Theorem : (see[4]&[15])

Let $p(\mathbf{Y}, \boldsymbol{\theta})$ be the joint probability density function (pdf) for a set of random observation $\mathbf{Y}$ and a parameter vector $\boldsymbol{\theta}$, also considered random. If $p(\mathbf{Y}) \neq 0$, then

$p(\boldsymbol{\theta} \mid \mathbf{Y})$ is written as

$$p(\boldsymbol{\theta} \mid \mathbf{Y}) \propto p(\boldsymbol{\theta})p(\mathbf{Y} \mid \boldsymbol{\theta}) \qquad (2.1)$$
$$\propto \text{prior pdf} \times \text{likelihood function}$$

where " $\propto$ " denotes proportionality, $p(\boldsymbol{\theta} \mid \mathbf{Y})$ is the posterior pdf for the parameter vector $\boldsymbol{\theta}$ given the sample information $\mathbf{Y}$, $p(\boldsymbol{\theta})$ is the prior pdf for the parameter vector $\boldsymbol{\theta}$, and $p(\mathbf{Y} \mid \boldsymbol{\theta})$ which is viewed as a function of $\boldsymbol{\theta}$, is the likelihood function based on the sample information.

In the posterior distribution, our previous information is adopted through $p(\boldsymbol{\theta})$, the prior pdf of $\boldsymbol{\theta}$, while the sample information is incorporated through the likelihood function. In this regard, $p(\mathbf{Y} \mid \boldsymbol{\theta})$ provides the entire evidence of the experiment (see[4]). So, we may look at $p(\mathbf{Y} \mid \boldsymbol{\theta})$ as representing the information about $\boldsymbol{\theta}$ coming from the data. Or we may look at it as the function through which the data $\mathbf{Y}$ modifies previous knowledge of $\boldsymbol{\theta}$. In the Bayesian approach, the posterior pdf is used to make inferences about the parameters under consideration.

There are various ways of choosing $p(\boldsymbol{\theta})$. The problem becomes serious when $p(\boldsymbol{\theta})$ is totally unknown. However, since, in many cases, there is lack of information about $\boldsymbol{\theta}$, the parameter under consideration, prior densities are provided that reflect prior of ignorance. These are called noninformative, vague or diffuse prior. Jeffrey, (see[4] & [15]), proposed some priors and they are well known in the literature. However, if some knowledge about the parameter is present, it should be adopted through a prior density as an informative prior.

In many occasions, we are interested in making inferences about the yet unobserved set of responses $\mathbf{Y_r}$, conditional on the observed part $\mathbf{Y_s}$. In the Bayesian approach, the probability distribution function for $\mathbf{Y_r}$ is known as the predictive pdf. We should note that the word " prediction " is used in some texts (see[7]) to indicate the values of the responses obtained from the fitted model in order to differentiate them from the experimental responses. However, this word, *prediction*, is used here, in the sense of the distribution of $\mathbf{Y_r}$, conditional on the observed responses $\mathbf{Y_s}$. So, if for instance, $p(\mathbf{Y_r}, \boldsymbol{\theta} \mid \mathbf{Y_s})$ is the joint pdf for $\mathbf{Y_r}$ and a parameter vector $\boldsymbol{\theta}$, given the sample information $\mathbf{Y_s}$, then, this pdf could be written as

$$p(\mathbf{Y_r}, \boldsymbol{\theta} \mid \mathbf{Y_s}) = p(\mathbf{Y_r} \mid \boldsymbol{\theta}, \mathbf{Y_s})p(\boldsymbol{\theta} \mid \mathbf{Y_s}) \tag{2.2}$$

where $p(\mathbf{Y_r} \mid \boldsymbol{\theta}, \mathbf{Y_s})$, known as the likelihood function, is the conditional pdf for $\mathbf{Y_r}$ given $\boldsymbol{\theta}$ and $\mathbf{Y_s}$, whereas $p(\boldsymbol{\theta} \mid \mathbf{Y_s})$ is the posterior pdf for $\boldsymbol{\theta}$ given $\mathbf{Y_s}$. Hence, the predictive pdf is found by integrating (2.2) with respect to $\boldsymbol{\theta}$. That is,

$$p(\mathbf{Y_r} \mid \mathbf{Y_s}) = \int_{R_\theta} p(\mathbf{Y_r}, \boldsymbol{\theta} \mid \mathbf{Y_s}) d\boldsymbol{\theta} \tag{2.3}$$

where R_θ is the domain space of the parameter $\boldsymbol{\theta}$ where the pdf is non zero.

Chapter 3

THE LINEAR MULTIPLE REGRESSION MODEL WITH NORMAL ERRORS

The generalization of the univariate bell shaped normal distribution to several dimensions, is one of the most important developments in the multivariate analysis (see[10]). The well known *central limit theorem* (see appendix [C]), justifies its importance since it shows that the sampling distribution of the mean vector of a multivariate random vector is approximately normal, regardless of the form of the parent population, if the sample size is large enough. However, the importance is not just because of that but rather, some populations, by nature, follow the multivariate normal distribution. Real data (see[10]) are seldom exactly multivariate normal. Nevertheless, it is often useful to approximate the true population distribution by the normal density.

Let X_1, X_2,..., X_k be a set of independent variables that are related to a response variable Y, also called the dependent variable. The linear regression model states that the mean of Y depends in a linear fashion on the independent variables. However, because of measurement error and the effects of not explicitly considering all possible independent variables in the model, as explained before, a random error term, u, is added to the model. This error term is then looked at as a random variable, and so is the response, whose behavior is characterized by a set of distributional assumptions.

The linear regression model with a single response has the following form

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k + u. \tag{3.1}$$

Note that Y is linear in both the β's and u and hence the term linear is introduced. The independent variables may or may not enter the model as first order terms.

With n independent observations on Y and the given values of X's, the complete model is written as

$$\mathbf{Y_s} = \mathbf{X_s} \boldsymbol{\beta} + \mathbf{u_s} \tag{3.2}$$

where,

$\mathbf{Y}'_s = (y_1, y_2, ..., y_n)$ is the vector of observations,

$\boldsymbol{\beta}' = (\beta_1, \beta_2, ..., \beta_k)$ is a $(1 \times k)$ vector of regression coefficient,

$$\mathbf{X_s} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix}$$ is an $n \times k$ matrix, with rank k, of observations on k

independent variables,

$\mathbf{u}'_s = (u_1, u_2, ..., u_n)$ is a $(1 \times n)$ vector of disturbance or error terms.

By adopting the common assumption of independent normal errors, we have $E(\mathbf{u_s}) = \mathbf{0}$, and $Cov(\mathbf{u_s}) = E(\mathbf{u_s u'_s}) = \sigma^2 \mathbf{I}_n$, where $\boldsymbol{\beta}$ and σ^2 are unknown parameters and $\mathbf{I}_n$ is the $n \times n$ identity matrix.

If the regression equation is assumed to have a nonzero intercept, the elements of the first column of $\mathbf{X_s}$, the design matrix, must all equal one. That is the first column of $\mathbf{X_s}$ is $\boldsymbol{\tau}$, where $\boldsymbol{\tau}' = (1, 1, ..., 1)$. The remaining elements of $\mathbf{X_s}$ may be stochastic or nonstochastic. However, if they are stochastic, it is assumed that they are distributed independently of $\mathbf{u_s}$ with parameters that do not involve $\boldsymbol{\beta}$ nor σ.

When we assume that u_i, $i = 1, 2, ..., n$, are normally and independently distributed, each with mean zero and variance σ^2, the model is called *normal linear multiple regression model.* With this assumption (see[15]), the joint distribution of $\mathbf{Y_s}$, given $\mathbf{X_s}$, $\boldsymbol{\beta}$, and σ, is

$$p(\mathbf{Y_s} \mid \mathbf{X_s}, \boldsymbol{\beta}, \sigma) \propto \tfrac{1}{\sigma^n} \exp[-\tfrac{1}{2\sigma^2}(\mathbf{Y_s} - \mathbf{X_s}\boldsymbol{\beta})'(\mathbf{Y_s} - \mathbf{X_s}\boldsymbol{\beta})]. \tag{3.3}$$

The above model (3.2) is the simplest of all multiple regression models and may

The above model (3.2) is the simplest of all multiple regression models and may not represent many real life situations. It can be made more complicated if the error term $\mathbf{u}_s$ follows the first order auto correlation scheme introduced in (1.1). The model then becomes

$$\mathbf{Y}_s = \mathbf{X}_s\beta + \mathbf{u}_s \tag{3.4a}$$

$$\mathbf{u}_s = \rho\mathbf{u}_{-s1} + \epsilon_s \tag{3.4b}$$

or alternatively,

$$\mathbf{Y}_s = \rho\mathbf{Y}_{-s1} + (\mathbf{X}_s - \rho\mathbf{X}_{-s1})\beta + \epsilon_s. \tag{3.4c}$$

With the above specification, the above model is known as *normal linear multiple regression model with first order auto-correlation (FOAC).*

Note that

$\mathbf{Y}'_s = (y_1, y_2, ..., y_n)$ and $\mathbf{Y}'_{-s1} = (y_0, y_1, ..., y_{n-1})$ are $(1 \times n)$ vectors of observations,

ρ is a scalar,

$\beta' = (\beta_1, \beta_2, ..., \beta_k)$ is a $(1 \times k)$ vector of regression coefficient,

$\epsilon'_s = (\epsilon_1, \epsilon_2, ..., \epsilon_n)$ is a $(1 \times n)$ vector of random errors that are independent and normally distributed each with mean zero and constant variance σ^2,

$\mathbf{u}'_s = (u_1, u_2, ..., u_n)$ and $\mathbf{u}'_{-s1} = (u_0, u_1, ..., u_{n-1})$ are $(1 \times n)$ vectors of autocorrelated errors,

$$\mathbf{X_s} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix} \;\&\; \mathbf{X_{-s1}} = \begin{bmatrix} x_{01} & \cdots & x_{0k} \\ \vdots & \ddots & \vdots \\ x_{(n-1)1} & \cdots & x_{(n-1)k} \end{bmatrix} \text{ are } n \times k \text{ matrices,}$$

with rank k, of observations on k independent variables.

Remarks:

- Since y_0 appears in $\mathbf{Y}'_{-s1}$, we have to impose some assumptions on it. We may assume for simplicity, that y_0 is fixed and known. This could happen as in the case where the observations stand for the price of some goods; y_0 is then the price at time t_0 when such price was fixed by some means.

- As stated before, the model (3.2) can be made more complicated if we add the assumption that the error term $\mathbf{u_s}$ follows the first order auto correlation scheme introduced in (1.1). Considering the model in (3.4a & 3.4b) where the error term $\mathbf{u_s}$ follows the first order auto correlation scheme, it was proven (see[13]) that the variance covariance matrix of $\mathbf{u_s}$ is given by

$$Cov(\mathbf{u_s}) = \tfrac{\sigma^2}{1-\rho^2}\mathbf{\Omega}_1 \text{ where } \mathbf{\Omega}_1 = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix} \text{ is an } n \times n$$

matrix with $|\rho| < 1$. However, if we let $\mathbf{u_s} = \delta \mathbf{w_s}$, where $\delta = \frac{\sigma}{\sqrt{1-\rho^2}}$, the model becomes

$$\mathbf{Y_s} = \mathbf{X_s}\beta + \delta \mathbf{w_s}. \tag{3.5}$$

In this representation, the variance covariance matrix of $\mathbf{w_s}$ is given by $\mathbf{\Omega}_1$ only and δ is called a scale parameter. The auto-correlation coefficient between the $(i,j)^{th}$ components of $\mathbf{w_s}$ is then given by $\text{cov}(w_i, w_j) = \rho^{|i-j|}$ for $i, j = 1, 2, ..., n$. Note that $\text{cov}(\mathbf{Y_s}) = \delta^2 \mathbf{\Omega}_1$ is the unknown variance-covariance matrix of the response vector $\mathbf{Y_s}$. The model in (3.5) was used by Khan [12] and it is the one we follow to write down his result in finding the predictive pdf, but after taking the *transpose* of the whole equation in (3.5). Khan used the model as

$$\mathbf{Y_s} = \beta \mathbf{X_s} + \delta \mathbf{w_s}. \tag{3.6}$$

- Since it is assumed that the errors, and hence the responses follow a joint multivariate normal distribution, the probability density function of $\mathbf{w_s}$ is given by,

$$p(\mathbf{w_s} \mid \rho) = (2\pi)^{-(n/2)} \mid \mathbf{\Omega}_1 \mid^{-(1/2)} \exp[(-1/2)\mathbf{w}_s' \mathbf{\Omega}_1^{-1} \mathbf{w_s}], \tag{3.7}$$

- If ρ is restricted to the values of $|\rho| < 1$, the situation is called nonexplosive. However, our model in (3.4a & 3.4b), as used by Zellner [15], is broad enough to be valid even for the explosive case i.e. when $|\rho| \geq 1$. So, we go through the assumption that $-\infty < \rho < \infty$.

- If in (3.4b) $\rho = 0$, (3.4a & 3.4b) would then reduce to the simple regression model described in (3.2).

- In (3.4c), when $\rho = 1$, the intercept term disappears. So, we assume that there is no intercept in the regression. Otherwise, the value that $\rho = 1$ must be precluded from the prior assumptions.

- For our analysis in this work, we follow the original model presented in (3.4a & 3.4b) since we will be using some of the results obtained by Zellner ([14] & [15]). Zellner, in his work, followed the model given by (3.4a & 3.4b) to find the posterior distribution of some parameters, namely $\boldsymbol{\beta}$ and ρ.

Chapter 4

PROBLEM UNDER CONSIDERATION

Consider an unobserved but realized error vector $\mathbf{u_s}$ in $\Re^n$ from the multiple regression model presented in (3.4a, 3.4b & 3.4c) where the assumption on the parameters involved in the three equations above are as stated in chapter three.

We assume this process stays the same for future responses. In other words, future responses follow the same model with same parameters as follows

$$\mathbf{Y_r} = \mathbf{X_r}\beta + \mathbf{u_r} \tag{4.1a}$$

$$\mathbf{u_r} = \rho \mathbf{u_{-r1}} + \epsilon_{\mathbf{r}} \tag{4.1b}$$

or alternatively,

$$\mathbf{Y_r} = \rho\mathbf{Y_{-r1}} + (\mathbf{X_r} - \rho\mathbf{X_{-r1}})\boldsymbol{\beta} + \epsilon_{\mathbf{r}} \tag{4.1c}$$

where $\mathbf{Y}'_{\mathbf{r}} = (y_{n+1}, y_{n+2}, ..., y_{n+q})$ and $\mathbf{Y}'_{-\mathbf{r1}} = (y_n, y_{n+1}, ..., y_{n+q-1})$ are $(1 \times q)$ vectors of the unobserved responses,

$\epsilon'_{\mathbf{r}} = (\epsilon_{n+1}, \epsilon_{n+2}, ..., \epsilon_{n+q})$ is a $(1 \times q)$ vector of random error,

$\mathbf{u}'_{\mathbf{r}} = (u_{n+1}, u_{n+2}, ..., u_{n+q})$ and $\mathbf{u}'_{-\mathbf{r1}} = (u_n, u_{n+1}, ..., u_{n+q-1})$ are $(1 \times q)$ vectors of future autocorrelated errors among themselves and those of $\mathbf{u}'_{\mathbf{s}}$,

ρ and β are as described before, and

the future design matrices of order $k \times q$ are

$$\mathbf{X_r} = \begin{bmatrix} x_{(n+1)1} & \cdots & x_{(n+1)k} \\ \vdots & \ddots & \vdots \\ x_{(n+q)1} & \cdots & x_{(n+q)k} \end{bmatrix}, \text{ and } \mathbf{X_{-r1}} = \begin{bmatrix} x_{n1} & \cdots & x_{nk} \\ \vdots & \ddots & \vdots \\ x_{(n+q-1)1} & \cdots & x_{(n+q-1)k} \end{bmatrix}$$

Remark:

Khan's [12] model of the problem of future responses is as follows. Starting with the model in (3.6) for the observed responses, it is assumed to continue on the same model for a set of q future responses as follows

$$\mathbf{Y_r} = \boldsymbol{\beta}\mathbf{X_r} + \delta\mathbf{w_r} \tag{4.2}$$

where $\mathbf{X_r}$ is the future design matrix of order $k \times q$; $\mathbf{w_r}$ is the future error vector associated with the future response vector $\mathbf{Y_r}$; and $\boldsymbol{\beta}$ and $\delta > 0$ are as defined above. The components of $\mathbf{w_r}$ are assumed to have auto-correlation of first order among themselves as well as with those of the realized but unobserved error vector $\mathbf{w_s}$. Thus, the covariance matrix of the combined error vector $\mathbf{w}^* = (\mathbf{w_s}, \mathbf{w_r})$, an $(n+q) \times 1$ row vector, is an $(n+q) \times (n+q)$ matrix, $\boldsymbol{\Omega}_2$, whose diagonal elements are unity and the $(i,j)^{th}$ off-diagonal elements are $\rho^{|i-j|}$ for $i,j = 1, 2, ..., n+q$. The covariance matrix could be partitioned as follows

$$\boldsymbol{\Omega}_2 = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix} \tag{4.3}$$

where $\boldsymbol{\Omega}_{11}$ is the $n \times n$ covariance matrix for the realized error; $\boldsymbol{\Omega}_{22}$ is the $q \times q$ covariance matrix for the nonrealized error; $\boldsymbol{\Omega}_{12} = \boldsymbol{\Omega}'_{21}$ is the $n \times q$ variance covariance matrix for the realized and nonrealized error. Khan [12] found the predictive pdf for the above model as will be shown later, following the structural relation method.

Our problem is to find the predictive pdf for $\mathbf{Y_r}$ (unobserved responses), conditional on the observed responses $\mathbf{Y_s}$ by using Bayesian approach.

4.1 OBJECTIVE

Our objectives in this study are the following.

- Apply the Bayesian approach to find the predictive pdf for the multiple regression model with first order auto correlation which is described in (3.4a & 3.4b). That is to find the predictive pdf for the unobserved responses $\mathbf{Y_r}$ conditional on the observed responses $\mathbf{Y_s}$.

- Compare our result with that obtained by Khan [12] where he made some assumption on ρ, the auto correlation coefficient.

In our study, we follow the Bayesian approach to get the posterior distribution of all the parameters. Then, we find the predictive pdf of the vector of the unobserved sample $\mathbf{Y_r}$, conditional on $\mathbf{Y_s}$.

Chapter 5

ANALYSIS AND RESULT

In this chapter, we develop the solution of the problem under consideration. The result developed by Khan [12], in which he followed the structural relation approach, is presented first in this chapter. Then, we present our own analysis of the problem in which we follow the Bayesian approach. Finally, we explain the superpopulation approach, which is another suggested approach to solve the problem.

5.1 Structural Relation Approach

The objective of the structural relation approach is to develop the structural distribution which is analogous to the posterior distribution in the Bayesian approach with flat prior. To find this distribution for our problem, Khan [12] found the predictive pdf for the set of unobserved responses conditional on the set of observed responses. Starting with model (3.6) for the observed responses and with model (4.2) for the unobserved responses, he first obtained the maximum likelihood estimate of ρ from the likelihood function which is given by

$$L(\rho \mid a) = K^{*} \times \mid \Omega_1 \mid^{-\frac{1}{2}} \mid \mathbf{X}_s \Omega_1^{-1} \mathbf{X}_s' \mid^{-\frac{1}{2}} [\mathbf{a}(\mathbf{w}_s)' \Omega_1^{-1} \mathbf{a}(\mathbf{w}_s)]^{-\frac{n-k}{2}} \tag{5.1}$$

where K^{*} is a multiplicative constant that does not depend on ρ.

However, before finding (5.1), he first defined the following statistics

$$\begin{aligned} \mathbf{b}(\mathbf{w}_s) &= \mathbf{w}_s \mathbf{X}_s' (\mathbf{X}_s \mathbf{X}_s')^{-1} \\ s^2(\mathbf{w}_s) &= [\mathbf{w}_s - \mathbf{b}(\mathbf{w}_s)\mathbf{X}_s][\mathbf{w}_s - \mathbf{b}(\mathbf{w}_s)\mathbf{X}_s]' \\ \mathbf{a}(\mathbf{w}_s) &= s^{-1}(\mathbf{w}_s)[\mathbf{w}_s - \mathbf{b}(\mathbf{w}_s)\mathbf{X}_s]. \end{aligned} \tag{5.2}$$

where, $\mathbf{b}(\mathbf{w}_s)$ is the regression of $\mathbf{w}_s$ on $\mathbf{X}_s$, $s^2(\mathbf{w}_s)$ is the residual sum of squares, and $\mathbf{a}(\mathbf{w}_s)$ is the normalized residual vector. Moreover, by finding the sample auto-correlation coefficient r which is given by

$$r = \sum_{j=1}^{n-1} a_j(\mathbf{w}_s) a_{j+1}(\mathbf{w}_s), \tag{5.3}$$

where $a_j(\mathbf{w}_s)$ is the j^{th} element of $\mathbf{a}(\mathbf{w}_s)$, the marginal likelihood function of ρ becomes

$$\begin{aligned} L(\rho \mid \mathbf{a}(\mathbf{w}_s)) =& \mid \Omega_1 \mid^{\frac{-1}{2}} \mid \mathbf{X}_s \Omega_1^{-1} \mathbf{X}_s' \mid^{\frac{-1}{2}} \times \\ & \{1 + \rho^2[1 - a_1(\mathbf{w}_s) - a_n(\mathbf{w}_s)] - 2\rho r\}^{-\frac{n-k}{2}} \end{aligned} \tag{5.4}$$

Therefore, for a particular choice of $\mathbf{X_s}$, the value of ρ that maximizes $L(\rho \mid \mathbf{a}(\mathbf{w_s}))$ is the maximum likelihood estimate of ρ, and is denoted by $\hat{\rho}$. The unknown auto correlation coefficient ρ, is then replaced by its maximum likelihood estimator $\hat{\rho}$. This estimator is used for the estimation of the covariance matrix Ω_2 given in (4.3). The estimated Ω_2 is then denoted by Σ and its inverse is then partitioned as

$$\Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \tag{5.5}$$

where,

Σ^{11} is an $n \times n$ matrix corresponding to the covariance of the elements of $\mathbf{w_s}$,

Σ^{22} is a $q \times q$ matrix corresponding to the covariance of the elements of $\mathbf{w_r}$, and

$\Sigma^{12} = (\Sigma^{21})'$ is an $n \times q$ matrix corresponding to the covariance among the elements of $\mathbf{w_s}$ and $\mathbf{w_r}$.

Then, the prediction distribution of $\mathbf{Y_r}$, for given $\mathbf{Y_s}$, was obtained to be multivariate Student-t whose degrees of freedom depend on the size of the observed sample and the dimensionality of the regression parameter vector. The result is given by

$$p(\mathbf{Y_r} \mid \mathbf{Y_s}) = \Phi(\hat{\rho}, n, q, \rho) \times$$
$$\{\mathbf{a}\Sigma^{11}\mathbf{a}' + s^{-2}(\mathbf{Y_s})[\mathbf{Y_r} - \mathbf{b}(\mathbf{w_s})\mathbf{X_r}]\mathbf{M}[\mathbf{Y_r} - \mathbf{b}(\mathbf{w_s})\mathbf{X_r}]'\}^{-\frac{n+q-k}{2}} \tag{5.6}$$

Or,

$$p(\mathbf{Y_r} \mid \mathbf{Y_s}) = \Phi(\hat{\rho}, n, q, \rho)\times$$

$$\{\mathbf{a}\Sigma^{11}\mathbf{a}' + s^{-2}(\mathbf{Y_s})[\mathbf{Y_r} - \mathbf{b}(\mathbf{w_s})\mathbf{X_r}]\mathbf{M}[\mathbf{Y_r} - \mathbf{b}(\mathbf{w_s})\mathbf{X_r}]'\}^{-\frac{\nu+q}{2}} \tag{5.7}$$

where

$$\Phi\,(\hat{\rho}, n, q, \rho) = s^{-q}(\mathbf{Y_s})\times\frac{|\mathbf{M}|^{\frac{1}{2}}\Gamma(\frac{\nu+q}{2})[\mathbf{a}\Sigma^{11}\mathbf{a}']^{\frac{\nu}{2}}}{\pi^{\frac{q}{2}}\Gamma(\frac{\nu}{2})} \tag{5.8}$$

$$\begin{aligned}
\nu &= n - k \\
\mathbf{M} &= \Sigma^{22} - \mathbf{H}'\mathbf{G}\mathbf{H} \\
\mathbf{H} &= \mathbf{X_s}\Sigma^{12} + \mathbf{X_r}\Sigma^{22} \\
\mathbf{G} &= \mathbf{X_s}\Sigma^{11}\mathbf{X}'_\mathbf{s} + \mathbf{X_r}\Sigma^{22}\mathbf{X}'_\mathbf{r} + 2\mathbf{X_s}\Sigma^{12}\mathbf{X}'_\mathbf{r} \\
\mathbf{a} &= \mathbf{a}(\mathbf{w_s}) \\
\mathbf{b}(\mathbf{Y_s}) &= \beta + \delta\mathbf{b}(\mathbf{w_s}) \quad \text{or} \\
\mathbf{b}(\mathbf{Y_s}) &= \mathbf{Y_s}\mathbf{X}'_\mathbf{s}(\mathbf{X_s}\mathbf{X}'_\mathbf{s})^{-1} \\
s(\mathbf{Y_s}) &= \delta s(\mathbf{w_s}) \quad \text{or} \\
s^2(\mathbf{Y_s}) &= [\mathbf{Y_s} - \mathbf{b}(\mathbf{Y_s})\mathbf{X_s}][\mathbf{Y_s} - \mathbf{b}(\mathbf{Y_s})\mathbf{X_s}]'
\end{aligned} \tag{5.9}$$

5.2 Bayesian Approach

In the Bayesian approach, we start with the model in (3.4a & 3.4b) for the observed responses and with those in (4.1a & 4.1b) for the unobserved responses to develop the predictive pdf. However, we stated in the introduction that in order to apply the Bayesian approach, we have to have some prior distributions of the parameters involved. For the predictive problem the prior of the parameters could be developed

developed to be informative. In other words, we start with a joint non-informative prior for such parameters and then derive the posterior distribution for them with a suitable likelihood function. This posterior distribution is then used as an informative prior for the parameters to obtain the joint predictive distribution for a set of the unobserved responses conditional on the observed ones.

In our problem, the parameters involved are ρ, σ and $\boldsymbol{\beta}$ and we need some prior distributions for them. We assume a joint non-informative prior distribution given by

$$p(\rho, \boldsymbol{\beta}, \sigma) \propto \frac{1}{\sigma}. \tag{5.10}$$

A suitable likelihood function (see[14] & [15]) is given by

$$L(\rho, \boldsymbol{\beta}, \sigma \mid \mathbf{Y_s}) \propto \frac{1}{\sigma^n} \exp\{-\frac{1}{2\sigma^2}[\mathbf{Y_s} - \rho\mathbf{Y}_{-s1} - (\mathbf{X_s} - \rho\mathbf{X}_{-s1})\boldsymbol{\beta}]' \times [\mathbf{Y_s} - \rho\mathbf{Y}_{-s1} - (\mathbf{X_s} - \rho\mathbf{X}_{-s1})\boldsymbol{\beta}]\}, \tag{5.11}$$

where $\mathbf{Y_s}$, $\mathbf{Y}_{-s1}$, $\mathbf{X}_{-s1}$ and $\mathbf{X_s}$ are as defined before in chapter three. By applying Bayes theorem, the posterior distribution for $\boldsymbol{\beta}$, σ, ρ, conditional on $\mathbf{Y_s}$, is given by

$$p(\rho, \boldsymbol{\beta}, \sigma \mid \mathbf{Y_s}, \mathbf{X_s}) \propto \frac{1}{\sigma^{n+1}} \exp\{-\frac{1}{2\sigma^2}[\mathbf{Y_s} - \rho\mathbf{Y}_{-s1} - (\mathbf{X_s} - \rho\mathbf{X}_{-s1})\boldsymbol{\beta}]' \times [\mathbf{Y_s} - \rho\mathbf{Y}_{-s1} - (\mathbf{X_s} - \rho\mathbf{X}_{-s1})\boldsymbol{\beta}]\}. \tag{5.12}$$

Then, we use this posterior distribution as an informative prior for the parameters to find the predictive distribution.

Since we assume that the process would continue under the same conditions and hence follows the same model, a suitable likelihood function for the unobserved responses is given by

$$L(\mathbf{Y_r} \mid \rho, \boldsymbol{\beta}, \sigma, \mathbf{Y_r}, \mathbf{X_s}, \mathbf{X_r}) \propto \tfrac{1}{\sigma^q} \exp\{-\tfrac{1}{2\sigma^2}[\mathbf{Y_r} - \rho\mathbf{Y_r} - (\mathbf{X_r} - \rho\mathbf{X_r})\boldsymbol{\beta}]' \times$$
$$[\mathbf{Y_r} - \rho\mathbf{Y_r} - (\mathbf{X_r} - \rho\mathbf{X_r})\boldsymbol{\beta}]\}, \qquad (5.13)$$

where $\mathbf{Y_r}$, $\mathbf{Y_{-r1}}$, $\mathbf{X_{-r1}}$ and $\mathbf{X_r}$ are as defined before in chapter three. Hence, by Bayes theorem, the joint posterior distribution for $\mathbf{Y_r}$, ρ, $\boldsymbol{\beta}$, σ is given by

$$p(\mathbf{Y_r}, \rho, \boldsymbol{\beta}, \sigma \mid \mathbf{Y_s}, \mathbf{X_s}, \mathbf{X_r}) \propto L(\mathbf{Y_r} \mid \rho, \boldsymbol{\beta}, \sigma, \mathbf{Y_s}, \mathbf{X_s}, \mathbf{X_r}) p(\rho, \boldsymbol{\beta}, \sigma \mid \mathbf{Y_s}). \qquad (5.14)$$

We want to find $p(\mathbf{Y_r} \mid \mathbf{Y_s})$. We do so, by integrating (5.14) with respect to σ, $\boldsymbol{\beta}$ and ρ. The joint posterior distribution in (5.14) can be written as

$$p(\mathbf{Y_r}, \rho, \sigma, \boldsymbol{\beta} \mid ..) \propto \tfrac{1}{\sigma^{n+q+1}} \exp\{-\tfrac{1}{2\sigma^2}[(\mathbf{w_1} - \mathbf{H_1}\boldsymbol{\beta})'(\mathbf{w_1} - \mathbf{H_1}\boldsymbol{\beta}) +$$
$$(\mathbf{w_2} - \mathbf{H_2}\boldsymbol{\beta})'(\mathbf{w_2} - \mathbf{H_2}\boldsymbol{\beta})]\} \qquad (5.15)$$

where

$$\mathbf{w_1} = \mathbf{Y_s} - \rho\mathbf{Y_{-s1}}, \;\; \mathbf{H_1} = \mathbf{X_s} - \rho\mathbf{X_{-s1}}$$

$$\mathbf{w_2} = \mathbf{Y_r} - \rho\mathbf{Y_{-r1}}, \;\; \mathbf{H_2} = \mathbf{X_r} - \rho\mathbf{X_{-r1}}.$$

We integrate (5.15) with respect to σ, see result (B.1) in appendix B, to get

$$p(\mathbf{Y_r}, \rho, \boldsymbol{\beta} \mid ...) \propto [(\mathbf{w}_1 - \mathbf{H}_1\boldsymbol{\beta})'(\mathbf{w}_1 - \mathbf{H}_1\boldsymbol{\beta}) + (\mathbf{w}_2 - \mathbf{H}_2\boldsymbol{\beta})'(\mathbf{w}_2 - \mathbf{H}_2\boldsymbol{\beta})]^{-\frac{n+q}{2}} \quad (5.16)$$

Now, in (5.16), $(\mathbf{w}_1 - \mathbf{H}_1\boldsymbol{\beta})'(\mathbf{w}_1 - \mathbf{H}_1\boldsymbol{\beta})$ and $(\mathbf{w}_1 - \mathbf{H}_1\boldsymbol{\beta})'(\mathbf{w}_1 - \mathbf{H}_1\boldsymbol{\beta})$ could be written as

$$(\mathbf{w}_1 - \mathbf{H}_1\boldsymbol{\beta})'(\mathbf{w}_1 - \mathbf{H}_1\boldsymbol{\beta}) = \mathbf{w}_1'\mathbf{w}_1 + \boldsymbol{\beta}'\mathbf{H}_1'\mathbf{H}_1\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{H}_1'\mathbf{w}_1 \quad (5.17a)$$

and similarly,

$$(\mathbf{w}_2 - \mathbf{H}_2\boldsymbol{\beta})'(\mathbf{w}_2 - \mathbf{H}_2\boldsymbol{\beta}) = \mathbf{w}_2'\mathbf{w}_2 + \boldsymbol{\beta}'\mathbf{H}_2'\mathbf{H}_2\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{H}_2'\mathbf{w}_2. \quad (5.17b)$$

Adding the left hand side of (5.17a & 5.17b), the right hand side could be written as

$$\mathbf{w}_1'\mathbf{w}_1 + \mathbf{w}_2'\mathbf{w}_2 + \boldsymbol{\beta}'\mathbf{E}_1\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{L}_1 \quad (5.17c)$$

where $\mathbf{E}_1 = \mathbf{H}_1'\mathbf{H}_1 + \mathbf{H}_2'\mathbf{H}_2$ and $\mathbf{L}_1 = \mathbf{H}_1'\mathbf{w}_1 + \mathbf{H}_2'\mathbf{w}_2$. By substituting (5.17c) into (5.16) we obtain

$$p(\mathbf{Y_r}, \rho, \boldsymbol{\beta} \mid ..) \propto [\mathbf{w}_1'\mathbf{w}_1 + \mathbf{w}_2'\mathbf{w}_2 + \boldsymbol{\beta}'\mathbf{E}_1\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{L}_1]^{-\frac{n+q}{2}}. \quad (5.18)$$

Let $\mathbf{b}_1 = \mathbf{E}_1^{-1}\mathbf{L}_1$ and put it in (5.18) which then becomes

$$p(\mathbf{Y}_r, \rho, \boldsymbol{\beta} \mid ..) \propto [\mathbf{w}_1'\mathbf{w}_1 + \mathbf{w}_2'\mathbf{w}_2 - \mathbf{b}_1'\mathbf{E}_1\mathbf{b}_1 + (\boldsymbol{\beta} - \mathbf{b}_1)'\mathbf{E}_1(\boldsymbol{\beta} - \mathbf{b}_1)]^{-\frac{n+q}{2}} \tag{5.19}$$

Now we use properties of the multivariate Student t-distribution, see appendix C, to integrate (5.19) with respect to the k elements of $\boldsymbol{\beta}$ to yield

$$p(\mathbf{Y}_r, \rho \mid ..) \propto |\mathbf{E}_1|^{\frac{-1}{2}} [\mathbf{w}_1'\mathbf{w}_1 + \mathbf{w}_2'\mathbf{w}_2 - \mathbf{b}_1'\mathbf{E}_1\mathbf{b}_1]^{-\frac{\nu+q}{2}} \tag{5.20}$$

where $\nu = n - k$.

It is shown in appendix (B) that $\mathbf{E}_1$ is invertible and symmetric. We show here that

$$\mathbf{b}_1'\mathbf{E}_1\mathbf{b}_1 = \mathbf{L}_1'\mathbf{E}_1^{-1}\mathbf{L}_1. \tag{5.21a}$$

This is because $\mathbf{b}_1 = \mathbf{E}_1^{-1}\mathbf{L}_1$ which implies that

$$\mathbf{b}_1' = \mathbf{L}_1'(\mathbf{E}_1^{-1})' = \mathbf{L}_1'(\mathbf{E}_1')^{-1} = \mathbf{L}_1'\mathbf{E}_1^{-1}. \tag{5.21b}$$

We put (5.21a) in (5.20) to obtain

$$p(\mathbf{Y_r}, \rho \mid ..) \propto \mid \mathbf{E}_1 \mid^{\frac{-1}{2}} [\mathbf{w}_1'\mathbf{w}_1 + \mathbf{w}_2'\mathbf{w}_2 - \mathbf{L}_1'\mathbf{E}_1^{-1}\mathbf{L}_1]^{-\frac{\nu+q}{2}} . \tag{5.21}$$

Recall that $\mathbf{L}_1 = \mathbf{H}_1'\mathbf{w}_1 + \mathbf{H}_2'\mathbf{w}_2$. If we let $\mathbf{f}_1 = \mathbf{H}_1'\mathbf{w}_1$, then $\mathbf{L}_1 = \mathbf{f}_1 + \mathbf{H}_2'\mathbf{w}_2$ and so (5.21) becomes

$$p(\mathbf{Y_r}, \rho \mid ..) \propto \mid \mathbf{E}_1 \mid^{\frac{-1}{2}} [\mathbf{w}_1'\mathbf{w}_1 + \mathbf{w}_2'\mathbf{w}_2 - (\mathbf{f}_1 + \mathbf{H}_2'\mathbf{w}_2)'\mathbf{E}_1^{-1}(\mathbf{f}_1 + \mathbf{H}_2'\mathbf{w}_2)]^{-\frac{\nu+q}{2}} \tag{5.22}$$

or

$$p(\mathbf{Y_r}, \rho \mid ..) \propto \mid \mathbf{E}_1 \mid^{\frac{-1}{2}} [s_1 + \mathbf{w}_2'\mathbf{E}_2\mathbf{w}_2 - 2\mathbf{w}_2'\mathbf{f}_2]^{-\frac{\nu+q}{2}} \tag{5.23}$$

where $s_1 = \mathbf{w}_1'\mathbf{w}_1 - \mathbf{f}_1'\mathbf{E}_1^{-1}\mathbf{f}_1$, $\mathbf{E}_2 = \mathbf{I}_q - \mathbf{H}_2\mathbf{E}_1^{-1}\mathbf{H}_2'$ and $\mathbf{f}_2 = \mathbf{H}_2\mathbf{E}_1^{-1}\mathbf{f}_1$.

Now consider $\mathbf{w}_2$ again.

$$\mathbf{w}_2 = \mathbf{Y_r} - \rho\mathbf{Y}_{-\mathbf{r}1} = \begin{bmatrix} y_{n+1} \\ y_{n+2} \\ \vdots \\ y_{n+q} \end{bmatrix} - \rho \begin{bmatrix} y_n \\ y_{n+1} \\ \vdots \\ y_{n+q-1} \end{bmatrix} = \begin{bmatrix} y_{n+1} \\ y_{n+2} - \rho y_{n+1} \\ \vdots \\ y_{n+q} - \rho y_{n+q-1} \end{bmatrix} - \rho \begin{bmatrix} y_n \\ 0 \\ \vdots \\ 0 \end{bmatrix} .$$

This $\mathbf{w}_2$ could be written as $\mathbf{AY_r} - \mathbf{f}_3$ where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & 0 & \cdots & 0 \\ \vdots & & & & & \vdots \\ \vdots & & & & & \vdots \\ 0 & 0 & \cdots & \cdots & -\rho & 1 \end{bmatrix} \text{ is a } q \times q \text{ matrix, and } \mathbf{f}_3 = \rho \begin{bmatrix} y_n \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

which is of dimension $q \times 1$. Note that A is invertible as will be shown later in appendix (B). By putting the new expression of $\mathbf{w}_2$ into (5.23), we to get

$$p(\mathbf{Y}_r, \rho \mid ..) \propto \mid \mathbf{E}_1 \mid^{\frac{-1}{2}} [s_1 + (\mathbf{A}\mathbf{Y}_r - \mathbf{f}_3)'\mathbf{E}_2(\mathbf{A}\mathbf{Y}_r - \mathbf{f}_3) - 2(\mathbf{A}\mathbf{Y}_r - \mathbf{f}_3)'\mathbf{f}_2]^{-\frac{\nu+q}{2}} \quad (5.24)$$

or

$$p(\mathbf{Y}_r, \rho \mid ..) \propto \mid \mathbf{E}_1 \mid^{\frac{-1}{2}} [s_2 + \mathbf{Y}_r'\mathbf{E}_3\mathbf{Y}_r - 2\mathbf{Y}_r'\mathbf{L}_2]^{-\frac{\nu+q}{2}} \quad (5.25)$$

where,

$s_2 = s_1 + \mathbf{f}_3'\mathbf{E}_2\mathbf{f}_3 + 2\mathbf{f}_3'\mathbf{f}_2$,

$\mathbf{E}_3 = \mathbf{A}'\mathbf{E}_2\mathbf{A}$, and

$\mathbf{L}_2 = \mathbf{A}'\mathbf{E}_2\mathbf{f}_3 + \mathbf{A}'\mathbf{f}_2$.

Now we complete the square on $\mathbf{Y_r}$ by letting $\mathbf{b}_2 = \mathbf{E}_2^{-1}\mathbf{L}_2$ so that (5.25) becomes

$$p(\mathbf{Y_r}, \rho \mid ..) \propto \mid \mathbf{E}_1 \mid^{\frac{-1}{2}} [s_2 - \mathbf{b}_2'\mathbf{E}_3\mathbf{b}_2 + (\mathbf{Y_r} - \mathbf{b}_2)'\mathbf{E}_3(\mathbf{Y_r} - \mathbf{b}_2)]^{-\frac{\nu+q}{2}} \tag{5.26}$$

Let $s_3 = s_2 - \mathbf{b}_2'\mathbf{E}_3\mathbf{b}_2$, we have

$$p(\mathbf{Y_r}, \rho \mid ...) \propto \mid \mathbf{E}_1 \mid^{\frac{-1}{2}} [s_3 + (\mathbf{Y_r} - \mathbf{b}_2)'\mathbf{E}_3(\mathbf{Y_r} - \mathbf{b}_2)]^{-\frac{\nu+q}{2}} \tag{5.27}$$

Assume that we have adopted the maximum likelihood estimator of ρ, $\hat{\rho}$, which was suggested by Khan [12]. Assume also that we have substituted this estimate in the functions starting from (5.10) till (5.27). Then, the predictive pdf for $\mathbf{Y_r}$ conditional on $\mathbf{Y_s}$ is obtained in the following from

$$p(\mathbf{Y_r} \mid \mathbf{Y_s}) \propto \mid \bar{\mathbf{E}}_1 \mid^{\frac{-1}{2}} [\bar{s}_3 + (\mathbf{Y_r} - \bar{\mathbf{b}}_2)'\bar{\mathbf{E}}_3(\mathbf{Y_r} - \bar{\mathbf{b}}_2)]^{-\frac{\nu+q}{2}} \tag{5.28a}$$

or

$$p(\mathbf{Y_r} \mid \mathbf{Y_s}) \propto [\bar{s}_3 + (\mathbf{Y_r} - \bar{\mathbf{b}}_2)'\bar{\mathbf{E}}_3(\mathbf{Y_r} - \bar{\mathbf{b}}_2)]^{-\frac{\nu+q}{2}} \tag{5.28b}$$

where $\mid \bar{\mathbf{E}}_1 \mid^{\frac{-1}{2}}$ is absorbed in the factor of proportionality since it is no more variable. In this regard, (5.28b) could be written as

$$p(\mathbf{Y_r} \mid \mathbf{Y_s}) = \Phi[\bar{s}_3 + (\mathbf{Y_r} - \bar{\mathbf{b}}_2)'\bar{\mathbf{E}}_3(\mathbf{Y_r} - \bar{\mathbf{b}}_2)]^{-\frac{\nu+q}{2}} \tag{5.29}$$

where

$\bar{s}_3$ is obtained from s_3 by replacing ρ with $\hat{\rho}$,

$\bar{\mathbf{b}}_2$ is obtained from $\mathbf{b}_2$ by replacing ρ with $\hat{\rho}$,

$\bar{\mathbf{E}}_1$ is obtained from $\mathbf{E}_1$ by replacing ρ with $\hat{\rho}$,

$\bar{\mathbf{E}}_2$ is obtained from $\mathbf{E}_2$ by replacing ρ with $\hat{\rho}$, and

Φ is the normalizing constant given by

$$\Phi = \frac{(\bar{s}_3)^{\frac{\nu}{2}} |\bar{\mathbf{E}}_2|^{\frac{1}{2}} \Gamma(\frac{\nu+q}{2})}{\pi^{\frac{q}{2}} \Gamma(\frac{\nu}{2})} .$$

The above pdf is the density function of well known multivariate t-distribution. It is of the same form obtained by Khan [12] when using the structural relation approach. This ensures that the Bayesian approach with non-informative prior, does indeed produce the same results. However, we notice the simplicity in this approach when compared with others.

We should notice that the joint predictive distribution of ρ and $\mathbf{Y}_r$ (5.27) is too complicated as a function of ρ and an explicit analytical result for the predictive pdf is not apparent. This could be a result of the many operations on matrices that are involved in the procedure of developing the predictive distribution. As an alternative to the analytical result, we apply numerical integration methods to find the prediction distribution. However, to apply numerical methods, we must have data.

5.2.1 Simulation Method

We state here the method of simulation we followed to generate the data we used to find the predictive distribution numerically. We start with the following model to generate the data

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u_t \tag{5.30}$$

$$u_t = \rho u_{t-1} + \epsilon_t$$

where u_{t-1} is u_t of the previous period.

Procedure:

First:

We generated ϵ's (independent error terms with mean zero and variance of one) using MiniTab, statistical package.

Second:

We used the values of X_1, X_2, X_3 and Y_t from an example in Bowerman & O'-Connell [5].

Note that,

$X_1 =$ Price

$X_2 =$ Average Industry price

$X_3 =$ Advertising Expenditure

$Y_t =$ Demand.

Third:

To get a good and realistic estimate of β's, we run the data obtained in the 2nd step in a regression process by using MiniTab again. We obtained estimated values of β's as follows

$\beta_0 = 7.589$, $\beta_1 = -2.358$, $\beta_2 = 1.612$, $\beta_3 = 0.501$

So we obtained the following estimated model

$$Y_t = 7.589 - 2.358X_1 + 1.612X_2 + 0.501X_3 + random\ error\ component \tag{5.31}$$

for the true response.. This error term is the difference between the fitted line of the response and that of the true response. However, up to this step, no autocorrelation is introduced.

Fourth:

In order to make the model obtained in step(3) follow a FOAC, we introduce the following

$$Y_t = 7.589 - 2.358X_1 + 1.612X_2 + 0.501X_3 + u_t \tag{5.32}$$

$$u_t = \rho u_{t-1} + \epsilon$$

where u_{t-1} is u_t of the previous period.

Fifth:

We regenerate the responses Y_t using the model introduced in step(4) with different assumed value of ρ. We generated the responses Y_t with $\rho = 0.5$, $\rho = 1.25$, $\rho = -1.25$ and $\rho = -0.5$. Now, this new model follows a FOAC.

The following data is generated for the FOAC model.

t	X1	X2	X3	Eps.	Ut for 0.50	Y for 0.50	Ut for 1.25	Y for 1.25
0	3.85	3.80	5.50	-	-1.53	5.86	-1.53	5.86
1	3.75	4.00	6.75	0.31	-0.46	8.12	-1.60	6.97
2	3.70	4.30	7.25	0.77	0.54	9.97	-1.23	8.20
3	3.70	3.70	5.50	-0.26	0.01	7.60	-1.80	5.78
4	3.60	3.85	7.00	1.26	1.27	10.08	-0.99	7.82
5	3.60	3.80	6.56	-0.36	0.27	8.79	-1.60	6.91
6	3.60	3.75	6.75	-0.46	-0.32	8.20	-2.46	6.07
7	3.80	3.85	5.25	-0.96	-1.12	6.34	-4.03	3.43
8	3.80	3.65	5.25	0.06	-0.50	6.64	-4.98	2.16
9	3.85	4.00	6.00	0.35	0.10	8.06	-5.88	2.09
10	3.90	4.10	6.50	-1.42	-1.37	6.89	-8.77	-0.51
11	3.90	4.00	6.25	1.36	0.67	8.65	-9.60	-1.63
12	3.70	4.10	7.00	0.47	0.81	9.79	-11.53	-2.55
13	3.75	4.20	6.90	0.15	0.55	9.53	-14.26	-5.29
14	3.75	4.10	6.80	0.54	0.82	9.58	-17.29	-8.53
15	3.80	4.10	6.80	0.00	0.41	9.05	-21.61	-12.97
16	3.70	4.20	7.10	1.40	1.60	10.80	-25.62	-16.42
17	3.80	4.30	7.00	1.13	1.93	11.00	-30.89	-21.82
18	3.70	4.10	6.80	0.34	1.31	10.19	-38.27	-29.39
19	3.80	3.75	6.50	-1.59	-0.94	6.99	-49.43	-41.50
20	3.80	3.75	6.25	1.64	1.17	8.98	-60.15	-52.34
21	3.75	3.65	6.00	-1.06	-0.47	7.16	-76.24	-68.61
22	3.70	3.90	6.50	0.57	0.33	8.74	-94.73	-86.33
23	3.55	3.65	7.00	-0.32	-0.15	8.46	-118.74	-110.13
24	3.60	4.10	6.80	-1.90	-1.98	7.14	-150.32	-141.21
25	3.65	4.25	6.80	0.24	-0.75	8.49	-187.66	-178.42
26	3.70	3.65	6.50	-1.05	-1.42	6.58	-235.63	-227.62
27	3.75	3.75	5.75	0.18	-0.53	7.14	-294.36	-286.68
28	3.80	3.85	5.80	-0.77	-1.04	6.70	-368.71	-360.97
29	3.70	4.25	6.80	0.72	0.20	9.32	-460.17	-451.05

Table(1) simulated data for $\rho = 0.5$ & 1.25

t	X1	X2	X3	Eps.	Ut for -0.50	Y for -0.50	Ut for -1.25	Y for -1.25
0	3.85	3.80	5.50	-	-1.53	5.86	-1.53	5.86
1	3.75	4.00	6.75	0.31	1.08	9.65	2.22	10.80
2	3.70	4.30	7.25	0.77	0.23	9.66	-2.01	7.42
3	3.70	3.70	5.50	-0.26	-0.38	7.21	2.25	9.83
4	3.60	3.85	7.00	1.26	1.45	10.26	-1.55	7.26
5	3.60	3.80	6.56	-0.36	-1.08	7.43	1.58	10.09
6	3.60	3.75	6.75	-0.46	0.08	8.61	-2.44	6.09
7	3.80	3.85	5.25	-0.96	-1.00	6.46	2.09	9.55
8	3.80	3.65	5.25	0.06	0.56	7.70	-2.55	4.60
9	3.85	4.00	6.00	0.35	0.07	8.03	3.53	11.50
10	3.90	4.10	6.50	-1.42	-1.45	6.80	-5.84	2.42
11	3.90	4.00	6.25	1.36	2.09	10.06	8.66	16.63
12	3.70	4.10	7.00	0.47	-0.57	8.41	-10.35	-1.37
13	3.75	4.20	6.90	0.15	0.44	9.41	13.09	22.06
14	3.75	4.10	6.80	0.54	0.32	9.08	-15.82	-7.06
15	3.80	4.10	6.80	0.00	-0.16	8.48	19.77	28.42
16	3.70	4.20	7.10	1.40	1.48	10.67	-23.32	-14.12
17	3.80	4.30	7.00	1.13	0.39	9.46	30.27	39.34
18	3.70	4.10	6.80	0.34	0.15	9.03	-37.50	-28.62
19	3.80	3.75	6.50	-1.59	-1.66	6.27	45.29	53.22
20	3.80	3.75	6.25	1.64	2.47	10.28	-54.97	-47.17
21	3.75	3.65	6.00	-1.06	-2.30	5.34	67.65	75.29
22	3.70	3.90	6.50	0.57	1.72	10.13	-84.00	-75.59
23	3.55	3.65	7.00	-0.32	-1.18	7.43	104.68	113.29
24	3.60	4.10	6.80	-1.90	-1.31	7.81	-132.75	-123.63
25	3.65	4.25	6.80	0.24	0.90	10.14	166.17	175.41
26	3.70	3.65	6.50	-1.05	-1.50	6.51	-208.77	-200.76
27	3.75	3.75	5.75	0.18	0.93	8.60	261.14	268.81
28	3.80	3.85	5.80	-0.77	-1.23	6.51	-327.19	-319.45
29	3.70	4.25	6.80	0.72	1.34	10.46	409.71	418.83

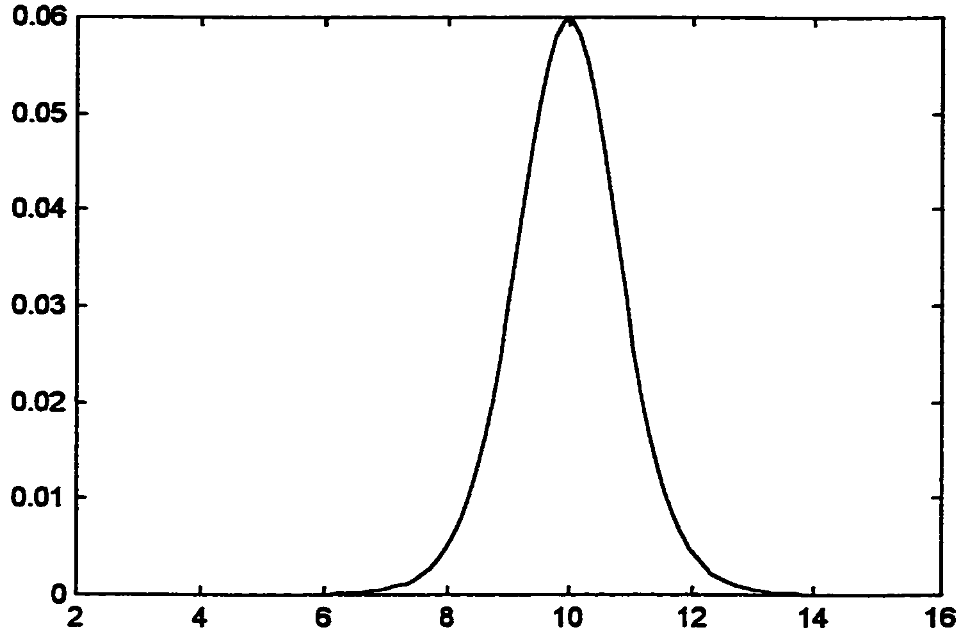
Table(2) simulated data for $\rho = -0.5$ & -1.25

For illustration purpose, we have performed numerical integration for equations (5.27) to find the normalizing constant using the first 16 data of the above table. After running the MatLab program, test1, (see appendix D), we found the normalizing constant for this joint distribution to be equal to 4.7204e+007. Regarding equation (5.28), we used the MatLab program, test2, to find the normalizing constat as well as to plot the prediction distribution. The following table gives the normalizing constant corresponding to different assumed estimated values of ρ.

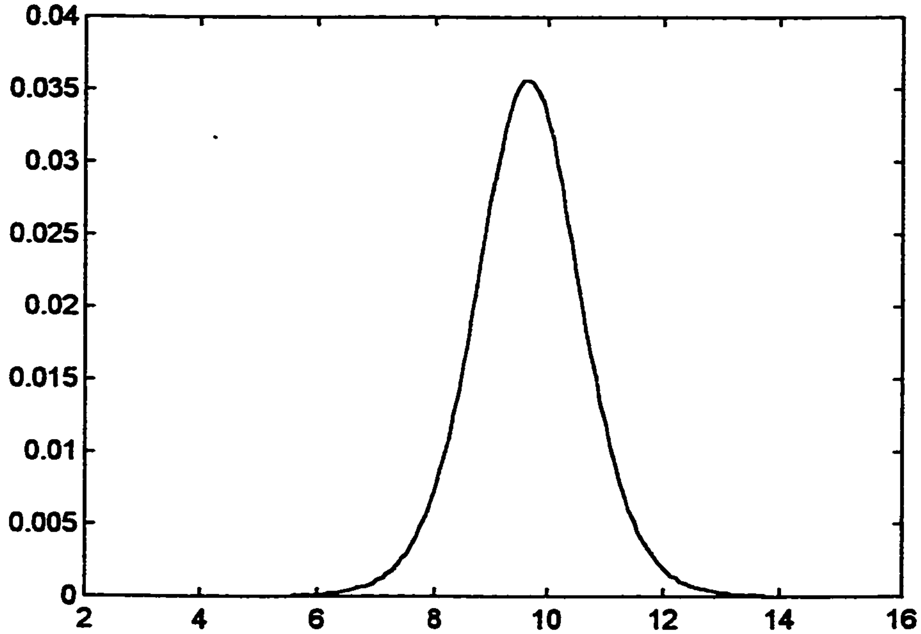
Estimated value of ρ	Normalizing constant
0.5	2.3197E+04
-0.5	6.0448E+04
1.25	1.1602E+04
-1.25	7.8221E+04

Table(3) Normalizing constants for equ.(5.28) with diffirent values of ρ

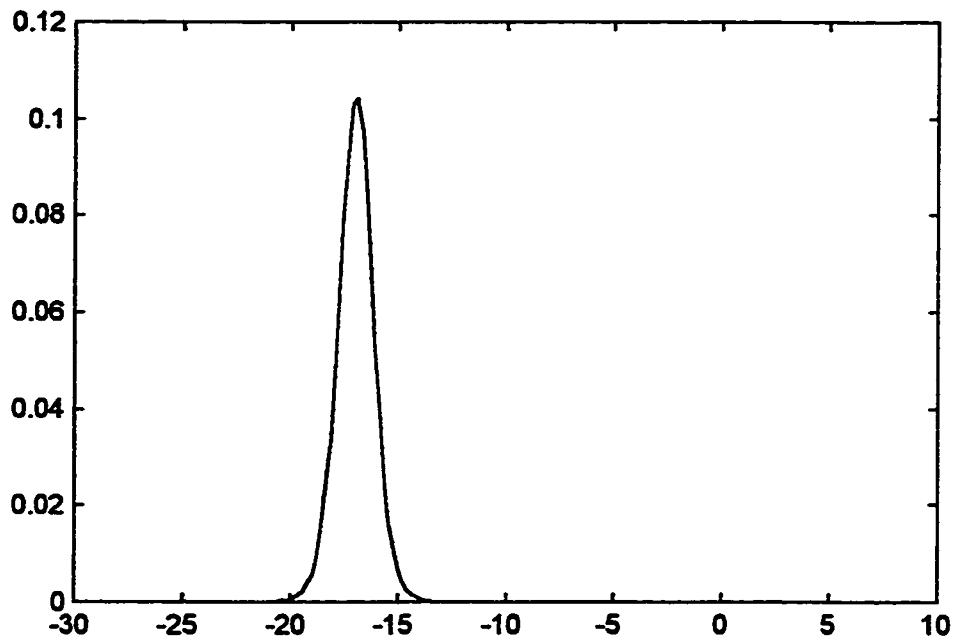
The followings are plots of equ.(5.28), the prediction distribution, corresponding to different values of ρ.



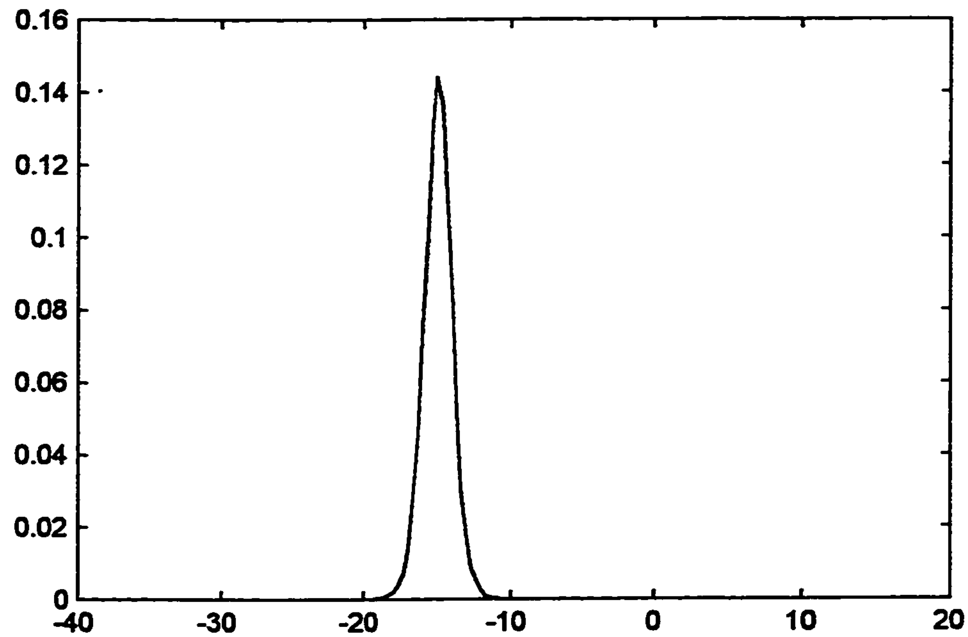
Figure(1) plot of the predictive pdf for equ.(5.28) with $\rho = 0.5$



Figure(2) plot of the predictive pdf for equ.(5.28) with $\rho = -0.5$



Figure(3) plot of the predictive pdf for equ.(5.28) with $\rho = 1.25$



Figure(4) plot of the predictive pdf for equ.(5.28) with $\rho = -1.25$

From the graphs of the prediction distribution given in figures one to four, it is to note that a particular choice of ρ would effectively affect both the center and the spread of the prediction distribution. For $|\rho| \geq 1$, the explosive case, the graphs are much more sharper than the situation of the nonexplosive case where $|\rho| < 1$. The other thing to be noted is that the predicted mean values, shown in the graphs, are very close to the real (simulated) values given in table(1).

5.3 Superpopulation Approach

Another suggested approach is the superpopulation approach. Superpopulation models consider the values of the population elements as random variables having joint distributions which may be specified completely or partially. In this approach, a model is to be introduced regardless of the procedure of how the sample is collected. Bolfarine & Zacks [3] discussed the problem of prediction, using this approach. The superpopulation model assumes that the value of the variable of interest, associated with the j^{th} unit of the population, Y_j, $j = 1, 2, ..., N$, is comprised of a deterministic element η_j and a random element e_j; that is,

$$Yj = \eta_j + e_j, \quad j = 1, 2, ..., N \tag{5.33a}$$

The random vector $\mathbf{e} = (e_1, ..., e_N)$ is assumed to have zero mean and a positive definite covariance matrix, $\mathbf{V}$. For the linear regression model, the deterministic elements η_j may be modeled as linear functions of the auxiliary variables; that is,

$$\eta_j = \sum_{k=1}^{n} \beta_k x_{jk}\,, \quad j = 1, 2, ..., N. \tag{5.33b}$$

Let $\mathbf{X}_s$ denote the $N \times n$ matrix, and let $\eta' = (\eta_1, \eta_2, ..., \eta_N)$. Then, the linear regression model can be expressed as

$$\mathbf{Y}_s = \mathbf{X}_s \boldsymbol{\beta} + \mathbf{e} \tag{5.34}$$

in which the random vector **e** is as described above. The superpopulation linear regression model (5.33) is represented by the parameter $\Psi = (\boldsymbol{\beta}, \mathbf{V})$, and is called the Ψ-model.

Concentrating our view to the multivariate normal model, consider (see[3]) the superpopulation regression model $\Psi = (\boldsymbol{\beta}, \mathbf{V})$ with normal distribution for the errors. In this regard, the joint distribution of Ψ given $\mathbf{Y}_s$ is the N-variate normal,

$$\mathbf{Y}_s \mid \Psi, \mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}; \mathbf{V}) \tag{5.34}$$

It is assumed here that **V** is known and $\boldsymbol{\beta}$ is unknown. The Bayes model assumes that $\boldsymbol{\beta}$ is a normal random vector, with mean vector **b** and covariance matrix B; that is,

$$\boldsymbol{\beta} \sim N(\mathbf{b}; \mathbf{B}) \tag{5.35}$$

The model defined by equations (5.34) & (5.35) is designated in the sequel as the Ψ_B model. The next theorem specifies the Bayes predictive distribution of $\mathbf{Y_r}$ given $\mathbf{Y_s}$, for the case where the covariance matrix $\mathbf{V}$ is known.

Theorem 1: (see[3])

Under the Bayesian model Ψ_B , the Bayes predictive distribution of $\mathbf{Y_r}$, given $\mathbf{Y_s}$, is multivariate normal, with mean vector

$$E_{\Psi_B}[\mathbf{Y_r} \mid \mathbf{Y_s}] = \mathbf{X_r}\hat{\boldsymbol{\beta}}_\mathbf{B} + \mathbf{V_{rs}}\mathbf{V_s^{-1}}(\mathbf{Y_s} - \mathbf{X_s}\hat{\boldsymbol{\beta}}_\mathbf{B}) \tag{5.36}$$

and covariance matrix,

$$\begin{aligned}\mathbf{Var}_{\Psi_B}[\mathbf{Y_r} \mid \mathbf{Y_s}] = \mathbf{V_r} - \mathbf{V_{rs}}\mathbf{V_s^{-1}}\mathbf{V_{sr}} + \\ [(\mathbf{X_r} - \mathbf{V_{rs}}\mathbf{V_s^{-1}}\mathbf{X_s})(\mathbf{X_s'}\mathbf{V_s^{-1}}\mathbf{X_s} + \mathbf{B}^{-1})^{-1}(\mathbf{X_r} - \mathbf{V_{rs}}\mathbf{V_s^{-1}}\mathbf{X_s})']\end{aligned} \tag{5.37}$$

where

$$\mathbf{V} = \begin{bmatrix} \mathbf{V_s} & \mathbf{V_{sr}} \\ \mathbf{V_{rs}} & \mathbf{V_r} \end{bmatrix}$$

and

$$\hat{\beta}_{\mathbf{B}} = (\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s + \mathbf{B}^{-1})^{-1}(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{Y}_s + \mathbf{B}^{-1}\mathbf{b}).$$

This general theorem could be specialized to the case where $\mathbf{V} = \sigma^2\mathbf{W}$, where $\mathbf{W}$ is known and $\mathbf{W}_{rs} = 0$, but σ^2 is unknown i.e. to specialize it to the case of the uncorrelated responses. The following theorem gives the detailed discussion of this aspect.

Theorem 2: (see[3])

Consider the normal model $\Psi_{\mathbf{B}}$ with $\mathbf{V} = \sigma^2\mathbf{W}$, where $\mathbf{W}$ is known and $\mathbf{W}_{rs} = 0$, but σ^2 is unknown. We also consider noninformative prior distribution on $(\beta; \sigma^2)$, according to which,

$$\zeta(\beta; \sigma^2) \propto \frac{1}{\sigma} \tag{5.38}$$

The posterior distribution of $\mathbf{Y}_r$ given $\mathbf{Y}_s$ is such that

$$E_{\Psi_{\mathbf{B}}}[\mathbf{Y}_r \mid \mathbf{Y}_s] = \mathbf{X}_r\hat{\beta}_s \tag{5.39}$$

and

$$\mathbf{Var}_{\Psi_{\mathbf{B}}}[\mathbf{Y}_r \mid \mathbf{Y}_s] = \frac{\nu}{\nu-2}\hat{\sigma}_s^2[\mathbf{W}_r + \mathbf{X}_r(\mathbf{X}_s'\mathbf{W}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_r'], \tag{5.40}$$

where $\hat{\sigma}_s^2 = (\mathbf{Y}_s - \mathbf{X}_s\hat{\boldsymbol{\beta}}_s)'\mathbf{W}_s^{-1}(\mathbf{Y}_s - \mathbf{X}_s\hat{\boldsymbol{\beta}}_s)/\nu$ and

$\hat{\boldsymbol{\beta}}_s = (\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{Y}_s$ with

$\nu = n - p.$

Our problem which was presented in chapter three could be solved using this approach. We could start with theorem 2, but with $\mathbf{W} = \boldsymbol{\Omega}_2$ where $\boldsymbol{\Omega}_2$ is the variance covariance matrix of both the observed and unobserved error terms. We also assume either informative or non informative prior distribution for ρ. Then the predictive distribution is derived. We conjecture that it would be of a multivariate t-distribution as obtained by Khan [12] and as shown here in our work using the Bayesian approach. This shows that how this problem could be tackled from different ways and that all methods lead to the same result, which suggests the equivalence of all of these different approaches. However, one of them appears to be easier than the other.

Chapter 6

SUMMARY AND CONCLUSION

In this work, we studied the problem of prediction distribution using the Bayesian approach. In particular, we applied it to the problem of prediction for the multiple regression model with FOAC. We compared our result with that obtained by Khan [12] where he followed the structural relation approach. Using the Bayesian approach we got the same result.

The difficulty in the Bayesian approach arises when performing the integration part on the parameter ρ. However, this problem has been tackled by means of numerical integration in which we used the MatLab software that can deal with the extensive matrix operations involved. To perform the numerical integration we generated data that fit our problem and then we used some of the simulated data to test our result in finding the predictive pdf. We were successful in that.

It is worth mentioning here that the problem of prediction has been discussed by Haq and Khan [8] for the linear regression model with dependent but uncorrelated multivariate Student-t error distribution. They used the structural relations of the model and they found that the prediction distribution is multivariate Student-t distribution with degrees of freedom that does not depend on the degrees of freedom of the error distribution. The form of the prediction distribution is the same as that obtained for the case of auto-correlated normal error term (see[12]).

However, the problem of prediction for the model discussed in this work could have been tackled with different approaches other than the structural relation or Bayesian. The superpopulation approach has been suggested here. This approach can be explored in the future.

APPENDIX (A)

Some Important Results on Matrices

For the detailed discussion of these results, (see[2],[9] & [10]).

Definition (A1):

A matrix $\mathbf{B}$ is said to be symmetric if it equals to its transpose i.e. $\mathbf{B} = \mathbf{B}'$.

Definition (A2):

Let $\mathbf{A}$ and $\mathbf{B}$ be two matrices, then $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.

Definition (A3):

A matrix $\mathbf{B}$ is said to be invertible if there exists another matrix $\mathbf{C}$ such that $\mathbf{BC} = \mathbf{CB} = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix.

Result(A1):

If $\mathbf{A}$, $\mathbf{B}$ are symmetric matrices, then $\mathbf{C} = \mathbf{A} + \mathbf{B}$ is also symmetric.

Result(A2):

If $\mathbf{A}$ and $\mathbf{B}$ are invertible matrices, then if $\mathbf{C} = \mathbf{A} + \mathbf{B}$ is invertible, then $\mathbf{C}^{-1} = \mathbf{A}^{-1} + \mathbf{B}^{-1}$.

Result(A3):

If $\mathbf{A}$ and $\mathbf{B}$ are invertible matrices, then if $\mathbf{C} = \mathbf{AB}$, then it is invertible and $\mathbf{C}^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

Result(A4):

If $\mathbf{A}$ is invertible and symmetric, then $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$.

Theorem (A1): (see[9])

Let $\mathbf{B}$ be a non-singular matrix and $\mathbf{U}$, $\mathbf{C}$ and $\mathbf{V}$ may be rectangular. Then

i) $(\mathbf{B} + \mathbf{UCV})$ is invertible and its inverse is given by

$$(\mathbf{B}+\mathbf{UCV})^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{U}(\mathbf{I}+\mathbf{CVB}^{-1}\mathbf{U})^{-1}\mathbf{CVB}^{-1} \tag{A1.1}$$

and

ii) $(\mathbf{B} - \mathbf{UD}^{-1}\mathbf{V})$ is also invertible and its inverse is given by

$$(\mathbf{B}-\mathbf{UD}^{-1}\mathbf{V})^{-1} = \mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{U}(\mathbf{D}-\mathbf{VB}^{-1}\mathbf{U})^{-1}\mathbf{VB}^{-1} \tag{A1.2}$$

Note that (A1.2) is obtained from (A1.1) by letting $\mathbf{C} = -\mathbf{D}^{-1}$.

APPENDIX (B)

Proofs of Some of the Used Results

In the derivation of the predictive distribution starting from equation (5.10) to (5.27), we have assumed implicitly that the matrices $\mathbf{E}_1, \mathbf{E}_2$, $\mathbf{A}$ and $\mathbf{E}_3$ are invertible. We give here proofs of some of the used results.

Note: $\mathbf{H}_1$ is an $n \times k$ matrix; $\mathbf{H}_2$ is a $q \times k$ matrix; $\mathbf{E}_1$ is a $k \times k$ matrix; $\mathbf{E}_2$ is a $q \times q$ matrix; $\mathbf{E}_3$ is a $q \times q$ matrix; $\mathbf{A}$ is a $q \times q$ matrix

First: to prove that $\mathbf{E}_1$ is invertible

$\mathbf{E}_1 = \mathbf{H}_1'\mathbf{H}_1 + \mathbf{H}_2'\mathbf{H}_2$ is invertible and symmetric but why?!. Obviously $\mathbf{E}_1$ is symmetric and to prove that it is invertible, we use theorem (A1). Now, $\mathbf{H}_1'\mathbf{H}_1$ is invertible. This is proven by Zellner [14]. We apply (A1.1) of theorem one above. To do so, we just let $\mathbf{B} = \mathbf{H}_1'\mathbf{H}_1$, $\mathbf{U} = \mathbf{I}_k$, $\mathbf{C} = \mathbf{H}_2'\mathbf{H}_2$ and $\mathbf{V} = \mathbf{I}_k$. Hence, $\mathbf{E}_1$ is invertible.■

Second: to prove that $\mathbf{E}_2$ is invertible

Recall that $\mathbf{E}_2 = \mathbf{I}_q - \mathbf{H}_2\mathbf{E}_1^{-1}\mathbf{H}_2'$. As we have done for $\mathbf{E}_1$, we do here for $\mathbf{E}_2$. We apply theorem (A1) again using (A1.2). To do so, we just let $\mathbf{B} = \mathbf{I}_q$, $\mathbf{U} = \mathbf{H}_2$, $\mathbf{D} = \mathbf{E}_1$ and $\mathbf{V} = \mathbf{H}_2'$. Hence, $\mathbf{E}_2$ is invertible.■

Third: to prove that $\mathbf{E}_3$ is invertible

Recall that $\mathbf{E}_3 = \mathbf{A}'\mathbf{E}_2\mathbf{A}$, where $\mathbf{A}$ is a $q \times q$ matrix given by

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & 0 & \cdots & 0 \\ \vdots & & & & & \vdots \\ \vdots & & & & & \vdots \\ 0 & 0 & \cdots & \cdots & -\rho & 1 \end{bmatrix}$$

$\mathbf{A}$ is similar to $\mathbf{I}_q$. This is obvious and it is easily found by doing some row operation. So, $\mathbf{A}$ is invertible. Then by results A3 & A4 given in appendix A, we have $\mathbf{E}_3$ is invertible and symmetric.■

Fourth: to prove the result used to integrate equation (5.15)

with respect to σ

We used the following result, (see [15]).

$$\int_0^\infty \sigma^{-(n+1)} \exp(-a/2\sigma^2) d\sigma = 2^{(n-2)/2}\Gamma(n/2)/a^{n/2} \tag{B.1}$$

proof:

Let $x = a/2\sigma^2$. Then the integral becomes

$$(2^{(n-2)/2}/a^{n/2}) \int_0^\infty x^{(n-2)/2} e^{-x} dx = 2^{(n-2)/2}\Gamma(n/2)/a^{n/2}$$

where Γ denotes the gamma function shown in (C.3).■

We used this result in (5.15), where

$$a = [(\mathbf{w}_1 - \mathbf{H}_1\boldsymbol{\beta})'(\mathbf{w}_1 - \mathbf{H}_1\boldsymbol{\beta}) + (\mathbf{w}_2 - \mathbf{H}_2\boldsymbol{\beta})'(\mathbf{w}_2 - \mathbf{H}_2\boldsymbol{\beta})]$$

and the factor $2^{(n-2)/2}\Gamma(n/2)$ is absorbed in the factor of proportionality..

APPENDIX (C)

Some Important Distributions and Theorems(see[11]&[15])

1. The Univariate Normal Distribution

Let X be a random variable. Then, X is said to have a normal distribution with mean θ and variance σ^2 if the density function of X has the following form

$$p(X \mid \theta, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2\sigma^2}(X-\theta)^2], \qquad -\infty < X < \infty \tag{C.1}$$

2. The Univariate Student-t Distribution

A random variable X is said to have a univariate student-t distribution if its pdf is as follows

$$p(X \mid \theta, h, \nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(1/2)\Gamma(\nu/2)}(\tfrac{h}{\nu})^{1/2}[1+\tfrac{h}{\nu}(X-\theta)^2]^{-(\nu+1)/2}, \quad -\infty < X < \infty \qquad (C.2)$$

with, $-\infty < \theta < \infty$, $0 < h < \infty$, $\nu > 0$ and Γ denotes the gamma function which is given by

$$\Gamma(n) = \int_{-\infty}^{\infty} u^{n-1}e^{-u}du, \qquad 0 < n < \infty . \qquad (C.3)$$

For $\nu > 1$, X has mean θ and for $\nu > 2$, it has variance $\frac{\nu}{h}\frac{1}{\nu-2}$.

3. The Multivariate Normal Distribution

Let $\mathbf{X} = (X_1, X_2, ..., X_n)$ be a vector. We say that the joint distribution of the elements of $\mathbf{X}$ is multivariate normal with mean vector θ and a covariance matrix Σ, if it is written as

$$p(\mathbf{X} \mid \theta, \Sigma) = \frac{|\Sigma|^{-1/2}}{(2\pi)^{n/2}} \exp[-\tfrac{1}{2}(\mathbf{X}-\theta)'\Sigma^{-1}(\mathbf{X}-\theta)] \qquad (C.4)$$

where $-\infty < X_i < \infty$; Σ is an $n \times n$ positive definite symmetric (PDS) matrix; $\theta' = (\theta_1, \theta_2, ..., \theta_n)$ with $-\infty < \theta_i < \infty$ for $i = 1, 2, ..., n$.

4. The Multivariate Student-t Distribution

We say that the elements of a random vector $\mathbf{X} = (X_1, X_2, ..., X_n)$ are distributed jointly as Multivariate Student-t if and only if they have the following pdf

$$p(\mathbf{X} \mid \theta, \mathbf{V}, \nu, n) = \frac{|\mathbf{V}|^{1/2}\nu^{\nu/2}\Gamma[(\nu+n)/2]}{\pi^{n/2}\Gamma(\nu/2)}[\nu+(\mathbf{X}-\theta)'\mathbf{V}(\mathbf{X}-\theta)]^{-(n+\nu)/2} \tag{C.5}$$

where $\nu > 0$; $\mathbf{V}$ is an $n \times n$ PDS matrix; $-\infty < X_i < \infty$; $\theta' = (\theta_1, \theta_2, ..., \theta_n)$ with $-\infty < \theta_i < \infty$, $i = 1, 2, ..., n$.

This distribution has mean at $\mathbf{X} = \theta$ and it has a variance of $[\nu/(\nu-2)]\mathbf{V}^{-1}$for $\nu > 2$.

5. The Central Limit Theorem

One of the most important theorems in statistics and probability is the central limit theorem. Its importance (see[11]) appears in showing that the distribution of the sample mean $\overline{Y}$ of a random sample of size n from any population with mean μ and variance σ^2, is approximated by the normal distribution with mean, $E(\overline{Y}) = \mu$ and standard error, $SE(\overline{Y}) = \frac{\sigma}{\sqrt{n}}$. With any sample size, if the population is normal, then the distribution of $\overline{Y}$ is exactly normal. However, it is approximated by a normal distribution for sufficiently large n.

Theorem (C1) [Central Limit Theorem]: (see[11])

Let p(.) (discrete or continuous) be the probability function of a random variable Y with mean μ and finite variance σ^2. Let $\overline{Y}_n$ be the sample mean of a random sample of size n, $Y_1, Y_2, ..., Y_n$ from p(.). Then the distribution of the random variable (or statistic)

$$Z_n = \frac{\overline{Y}_n - E(\overline{Y}_n)}{\sqrt{Var(\overline{Y}_n)}} = \frac{\overline{Y}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}(\overline{Y}_n - \mu)}{\sigma} \tag{C.6}$$

approaches the standard normal distribution as $n \rightarrow \infty$.

Proof: (see[11])

APPENDIX (D)

MatLab Program

We are providing here two MatLab programs to find the normalizing constant. Test(1) will find the normalizing constant for the joint distribution of ρ(rou) & $\mathbf{Y}_r$ given in (5.27). Test(2) will provide us with both the normalizing constant as well as a plot for the predictive pdf given in (5.28). The advantage of our programs is that they are easily extended to multiple integration since we are using the basic definitions of integration rather than trapezoidal or Simpson's rules. We just need to provide the data as well as the number of divisions we want for intervals.

Test(1)

```
% Test 1
% Note that this program finds the normalizing constant
% of the predictive pdf given in equ (5.27).
% Enter the basic parameters
roumin=-2;
roumax=2;
Yrmin=2;
Yrmax=16;
numberOfDevisions=170;
numberObservations=15;
numberUnobservations=1;
numberOfRegression=4;
ys=[5.86
8.12
9.97
7.59
10.07
8.76
```

8.21

6.35

6.65

8.07

6.88

8.64

9.78

9.52

9.57];

Ys=[8.12

9.97

7.59

10.07

8.76

8.21

6.35

6.65

8.07

6.88

8.64

9.78

9.52

9.57

```
9.05];
xs=[1 3.85 3.80 5.50
1 3.75 4.00 6.75
1 3.70 4.30 7.25
1 3.70 3.70 5.50
1 3.60 3.85 7.00
1 3.60 3.80 6.50
1 3.60 3.75 6.75
1 3.80 3.85 5.25
1 3.80 3.65 5.25
1 3.85 4.00 6.00
1 3.90 4.10 6.50
1 3.90 4.00 6.25
1 3.70 4.10 7.00
1 3.75 4.20 6.90
1 3.75 4.10 6.80];
Xs=[1 3.75 4.00 6.75
1 3.70 4.30 7.25
1 3.70 3.70 5.50
1 3.60 3.85 7.00
1 3.60 3.80 6.50
1 3.60 3.75 6.75
1 3.80 3.85 5.25
```

```
1 3.80 3.65 5.25
1 3.85 4.00 6.00
1 3.90 4.10 6.50
1 3.90 4.00 6.25
1 3.70 4.10 7.00
1 3.75 4.20 6.90
1 3.75 4.10 6.80
1 3.80 4.10 6.80];
xr=[1 3.80 4.10 6.80];
Xr=[1 3.70 4.20 7.10];
yr=[Ys(15)];
% Data entry finished
N=numberObservations;
q=numberUnobservations;
kk=numberOfRegression;
power=(N-kk+q);
umin=roumin;
umax=roumax;
ymin=Yrmin;
ymax=Yrmax;
ndiv=numberOfDevisions;
inty=(ymax-ymin)/ndiv;
intu=(umax-umin)/ndiv;
```

```
y=ymin;
u=umin;
Yr=[y];
C=[1];
A=[1];
f3=u*[Ys(15)];
Q=[ ];
ppr=[ ];
for ky1=1:ndiv;
for ku1=1:ndiv;
w1=Ys-u*ys;
w2=Yr-u*yr;
H1=Xs-u*xs;
H2=Xr-u*xr;
E1=H1'*H1+H2'*H2;
l1=H1'*w1+H2'*w2;
b1=inv(E1)*l1;
f1=H1'*w1;
s1=w1'*w1-f1'*inv(E1)*f1;
E2=C-H2*inv(E1)*H2';
f2=H2*inv(E1)*f1;
s2=s1+f3'*E2*f3+2*f3'*f2;
E3=A'*E2*A;
```

```
l2=A'*E2*f3+A'*f2;
b2=inv(E3)*l2;
s3=s2-b2'*E3*b2;
w3=Yr-b2;
s4=w3'*E3*w3;
mm1=det(E1);
mm2=mm1^(-1/2);
mm3=s3+s4;
mm4=mm3^(-power/2);
f=mm2*mm4;
pr(ky1,ku1)=f;
u=u+intu;
end
u=umin;
y=y+inty;
Yr=[y];
Q=[Q y];
end
for n=1:ndiv;
domu(1,n)=abs(intu);
domy(n,1)=abs(inty);
end
domu;
```

```
domy;
domain=domy*domu;
for k1=1:ndiv;
for k2=1:ndiv;
ppr(k1,k2)=pr(k1,k2);
end
end
integ=domain.*ppr;
norm=sum(sum(integ))
NormConst=(1/norm)
```

END OF TEST 1

Test(2)

```
% Test 2

% Note that this program plots the predictive pdf where we use an esimated

% value of rou.

% Enter the basic parameters

Yrmin=2;

Yrmax=16;

EstimatedValueOfRou=0.5;

numberOfDevisions=110;

numberObservations=15;

numberUnobservations=1;

numberOfRegression=4;


ys=[5.86

    8.12

    9.97

    7.59

   10.07

    8.76

    8.21
```

```
 6.35
 6.65
 8.07
 6.88
 8.64
 9.78
 9.52
 9.57];

Ys=[8.12
  9.97
  7.59
 10.07
   8.76
  8.21
  6.35
  6.65
  8.07
  6.88
  8.64
  9.78
   9.52
  9.57
```

```
9.05];

xs=[1 3.85 3.80 5.50
1 3.75 4.00 6.75
1 3.70 4.30 7.25
1 3.70 3.70 5.50
1 3.60 3.85 7.00
1 3.60 3.80 6.50
1 3.60 3.75 6.75
1 3.80 3.85 5.25
1 3.80 3.65 5.25
1 3.85 4.00 6.00
1 3.90 4.10 6.50
1 3.90 4.00 6.25
1 3.70 4.10 7.00
1 3.75 4.20 6.90
1 3.75 4.10 6.80];

Xs=[1 3.75 4.00 6.75
1 3.70 4.30 7.25
1 3.70 3.70 5.50
1 3.60 3.85 7.00
1 3.60 3.80 6.50
```

```
1 3.60 3.75 6.75
1 3.80 3.85 5.25
1 3.80 3.65 5.25
1 3.85 4.00 6.00
1 3.90 4.10 6.50
1 3.90 4.00 6.25
1 3.70 4.10 7.00
1 3.75 4.20 6.90
1 3.75 4.10 6.80
1 3.80 4.10 6.80];

xr = [1 3.80 4.10 6.80];

Xr =[1 3.70 4.20 7.10];

yr=[Ys(15)];

% Data entry finished

N=numberObservations;
q=numberUnobservations;
kk=numberOfRegression;
power=(N-kk+q);
```

```
ymin=Yrmin;
ymax=Yrmax;
ndiv=numberOfDevisions;
inty=(ymax-ymin)/ndiv;
intu=(umax-umin)/ndiv;
y=ymin;
u=EstimatedValueOfRou;
Yr=[y];
C=[1];
A=[1];
f3=u*[Ys(15)];
Q=[ ];
pr=[ ];
for ky1=1:ndiv;
w1=Ys-u*ys;
w2=Yr-u*yr;
H1=Xs-u*xs;
H2=Xr-u*xr;
E1=H1'*H1+H2'*H2;
l1=H1'*w1+H2'*w2;
b1=inv(E1)*l1;
f1=H1'*w1;
s1=w1'*w1-f1'*inv(E1)*f1;
```

```
E2=C-H2*inv(E1)*H2';
f2=H2*inv(E1)*f1;
s2=s1+f3'*E2*f3+2*f3'*f2;
E3=A'*E2*A;
l2=A'*E2*f3+A'*f2;
b2=inv(E3)*l2;
s3=s2-b2'*E3*b2;
w3=Yr-b2;
s4=w3'*E3*w3;
s5=s3+s4;
f=s5^(-power/2);
pr=[pr f];
y=y+inty;
Yr=[y];
Q=[Q y];
end
for n=1:ndiv;
domy(1,n)=abs(inty);
end
domy;
pr2=inty*pr;
norm=sum(pr2);
result=(1/norm)*pr2;
```

```
NormConst=(1/norm)

plot(Q, restult)
```

END OF TEST 2

REFERENCES

[1] Abraham, B., and Ledolter, J., Statistical Methods for Forecasting, first edition. USA: John Wiley & Sons, Inc., 1983.

[2] Anderson, T. W., An Introduction to Multivariate Statistical Analysis, second edition. USA: John Wiley & Sons, Inc., 1984.

[3] Bolfarine, H., and Zack, S., Prediction Theory for Finite Populations, first edition. New York: Springer-Verlag, 1992.

[4] Bowerman, B. L., and O'Connell, R. T., Time Series and Forecasting: An Applied Approach, first edition. USA: Wadsworth, Inc., 1979.

[5] Box, G. E. P., and Tiao G. C., Bayesian Inference in Statistical Analysis, first edition. USA: Addison-Wesley Publishing Company, Inc., 1973.

[6] Cochrane, D. and Orcutt, G. H., "Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms," Journal of the American Statistical Association, **44** (1949), 32-61.

[7] Cryer, J. D., and Miller, R. B., Statistics for Business: Data Analysis and Modeling, second edition. USA: Wadsworgh, Inc., 1994.

[8] Haq, M. H., and Khan, S., "Prediction Distribution for a Linear Regression Model with Multivariate Student-t Error Distribution," Commun. Statist.- Theory Meth., 19 (1990), 4705-4712.

[9] Hderson, H. V., and Searle, S. R., "On Deriving the Inverse of a Sum of Matrices," SAIM Review, 23 No 1(1981), 53-60.

[10] Johnson, A., and Wichern, W., Applied Multivariate Statistical Analysis, first edition. USA: Prentice-Hall, Inc., 1982.

[11] Khan, S., Distribution Theory: Study Book, first edition. Australia: The University of Southern Queensland, Distance Education Centre (DEC), 1996.

[12] Khan, S., "Marginal Likelihood Estimation and Prediction Distribution for Some Linear Models with First Order Auto-Correlation," Journal of Statistical Research, 44 (1997), 103-116.

[13] Sen, A. and Srivastava, M., Regression Analysis: Theory, Methods, and Applications, second edition. USA: Springer-Verlag, 1994.

[14] Zellner, A., and Tiao, G. C., "Bayesian Analysis of the Regression Model with Auto-Correlated Errors," J. Am. Statist. Assoc., 59 (1964), 763-778.

[15] Zellner, A., An Introduction to Bayesian Inference in Econometrics, first edition. USA: John Wiley & Sons, Inc., 1971.

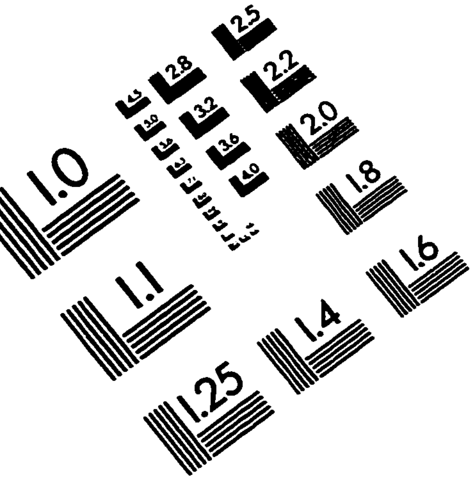
Vita

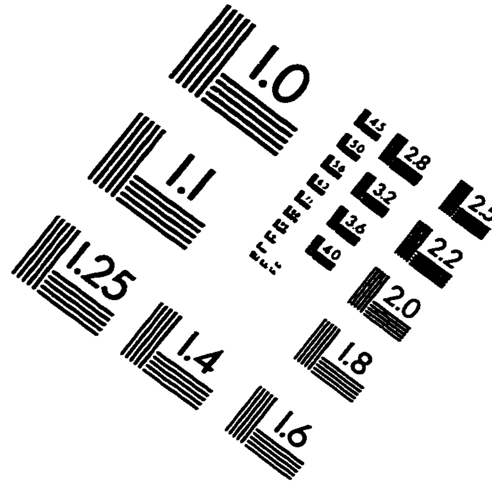
I finished my secondary school from Al-Qaudih secondary school in 1989-1990. After that, I entered the KFUPM in 1990 and I specialized in the electrical engineering. However, in 1993, I shifted my major of specialization to mathematics and I finished my BS in mathematics in 1995. I entered the MS program directly after my graduation as a full time student. After one year of my study in the MS, I joined the Education Ministry as a secondary school teacher who teaches mathematics. Meanwhile, I have continued on my study in the MS program as a part-time student under my own expense.

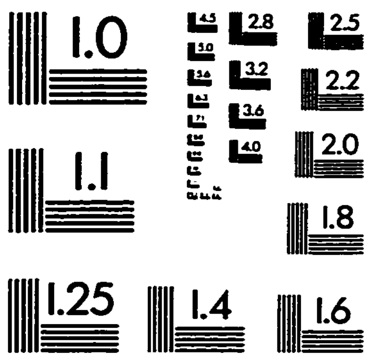
Al-Baiyat, Abdulkhaleg Ali

December, 1998

IMAGE EVALUATION TEST TARGET (QA-3)

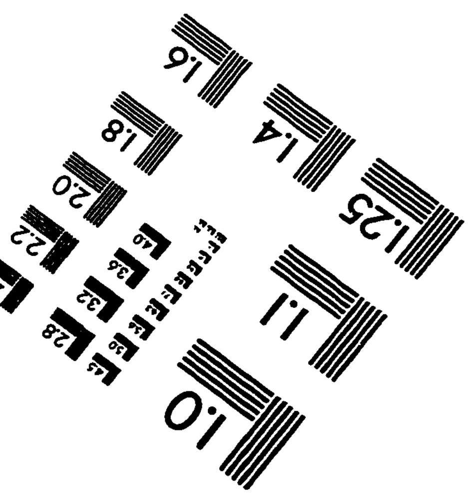


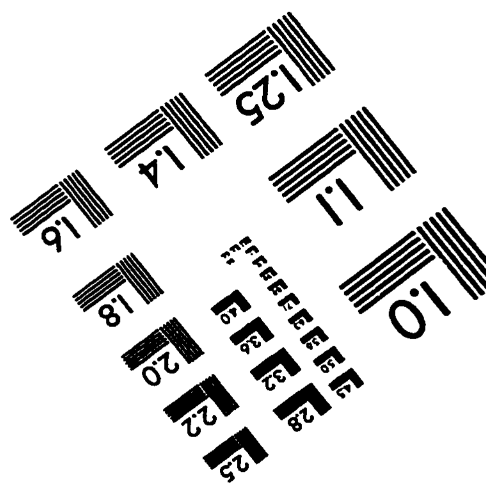




150mm

6"





APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989