Predictive Distribution of Regression Vector and Residual Sum of Squares for Normal Multiple Regression Model

Shahjahan Khan
Department of Mathematics & Computing
University of Southern Queensland
Toowoomba, Queensland, Australia
Email: khans@usq.edu.au

Abstract

This paper proposes predictive inference for the multiple regression model with independent normal errors. The distributions of the sample regression vector (SRV) and the residual sum of squares (RSS) for the model are derived by using invariant differentials. Also the predictive distributions of the future regression vector (FRV) and the future residual sum of squares (FRSS) for the future regression model are obtained. Conditional on the realized responses, the future regression vector is found to follow a multivariate Student-t distribution, and that of the residual sum of squares follows a scaled beta distribution. The new results have been applied to the market return and accounting rate data to illustrate its application.

Keywords: Multiple regression model; regression vector and residual sum of squares; non-informative prior; future regression model; predictive inference; future regression vector; multivariate normal, Student-t, beta and gamma distributions.

AMS 1991 Subject Classification : Primary 62H10, secondary 62J12.

1 Introduction

The predictive inference had been the oldest form of statistical inference used in real life. In general, predictive inference is directed towards inference involving the observables, rather than the parameters. The predictive method had been the most popular statistical tool before the diversion of interest in the inferences on parameters of the models. Predictive inference uses the realized responses from the *performed experiment* to make inference about the behavior of the unobserved responses of the *future experiment* (cf. Aitchison and Dunsmore, 1975, p.1). The outcomes of the two experiments are connected through the same structure of the model and indexed by the same set of parameters. For details on the predictive inference methods and its wide range of applications readers may refer to Aitchison and Dunsmore (1975) and Geisser (1993). Predictive inference for a set of future responses of the model, conditional on the realized responses from the same model, has been derived by many authors including Aitchison and Scalthorpe (1965), Fraser and Haq (1969), and Haq and Khan (1990). The prediction distribution of a set of future responses from the model has been used by Guttman (1970), Haq and Rinco (1973) and Khan (1992) to derive β-expectation tolerance region. There are many other kinds of applications of the prediction distributions available in the literature (see Geisser, 1993, for instance).

Like almost every other branches of statistics, there has been many studies in the area of predictive inference mainly for the independent and normal error model. The pioneering work in this area includes Fraser and Guttman (1956) Aitchison (1964), and Aitchison and Sculthorpe (1965), Fraser and Haq (1969), and Guttman (1970). Aitchison and Dunsmore (1975) provide an excellent account of the theory and application of the prediction methods. Fraser and Haq (1969) obtained prediction distribution for the multivariate normal model by using the structural distribution, instead of the Bayes posterior distribution. Haq (1982) used the structural relations, rather than the structural distribution, to derive the prediction distribution. Geisser (1993) discussed the Bayesian approach to predictive inference and included a wide range of real-life applications of the method. This includes model selection, discordancy, perturbation analysis, classification, regulation, screening and interim analysis. The predictive inference for the linear model has been dealt with by Lieberman and Miller (1963), Bishop (1976) and Ng (2000). Haq and Rinco (1976) derived the β-expectation tolerance region for generalized linear model with multivariate normal errors using the prediction distribution obtained by structural approach. Unlike the above normal theory based studies, Khan (1992), Khan and Haq (1994); and Fang and Anderson (1990), Khan (1996) and Ng (2000) provide predictive analyses of linear models with multivariate Student-t errors and spherical errors respectively.

In this paper we consider the widely used multiple regression model for the unobserved but realized responses as well as for the unobserved future responses. The two sets of errors are assumed to follow independent normal distribution. However, they are connected to one another through the common regression and scale parameters. Here, we pursue the predictive approach to derive the distribution of the regression vector and the residual sum of squares of the future responses, conditional on the set of realized responses. This is a new approach that proposes predictive inference for the regression parameters of the multiple regression model based on the future responses. The proposed predictive inference of the regression parameters depends on the realized responses, but not through the prediction distribution of the future responses. First the joint distribution of the sample regression vector and the residual sum of squares of the errors are derived from the joint distribution of the two error vectors by using the invariant differentials (cf Fraser, 1968, p.30). Then the distribution of the sample regression vector and the residual sum of squares of the realized responses are derived by using appropriate transformations. The sample regression vector is found to be a multivariate normal vector and the residual sum of squares statistic turns out to be a scaled gamma variable. These two statistics are independently distributed. Finally, the distribution of the same statistics of the future regression model, that is, the future regression vector and residual sum of squares of the future responses, conditional on the realized responses, are obtained by using the non-informative prior distribution for the parameters.

The predictive distribution of the future regression vector follows a multivariate Student-t distribution and that of the residual sum of squares of the future regression follows a scaled beta distribution. Unlike the sample regression vector and residual sum of squares of the realized regression model, the distribution of the same statistics for the future regression model, conditional on the realized responses, are dependent, and hence the joint density can't be factorized into the marginal distributions.

In many occasions the researchers may require to predict the value of the parameter, rather than the response itself. In particular, if the interest is in the predictive inference on the regression parameter, the rate of change in the response variable with unit change in the explanatory variable, we require to find the prediction distribution of the future slope vector. Given a set of realized responses and an appropriate prior distribution for the underlying parameters, this can be obtained by defining the joint distribution of the parameters and future regression vector based on the unobserved future responses. In this paper we assume that the non-informative prior distribution for the parameters of the model under consideration. Ng (2000) used an improper prior for the derivation of prediction distribution.

In the next section, we discuss the multiple regression model with normal errors. Some preliminaries are provided in section 3. Distributions of the sample regression vector and the residual sum of squares of the realized model are obtained in section 4. The multiple regression model for the future responses is introduced in section 5. The predictive distributions of the regression vector and the residual sum of squares of the future regression model, conditional on the realized responses, are derived in section 6. An illustrative example based on stock market data is provided in section 7. Some concluding remarks are included in section 8.

2 The Multiple Regression Model

Consider the commonly used linear regression equation

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

where y is the response variable, $\boldsymbol{\beta}$ is the vector of regression parameters assuming values in the p-dimensional real space $\mathcal{R}^p$, $\boldsymbol{x}$ is the vector of p regressors, σ is the scale parameter assuming values in the positive half of the real line $\mathcal{R}^+$, and e is the error variable associated with the response y. Assume that the error component, e, is normally distributed with mean 0 and variance 1, so that the variance of y is σ^2. Now, consider a set of $n > p$ independent responses, $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)$, from the above regression model that can be expressed as

$$\boldsymbol{y} = \boldsymbol{\beta} X + \sigma \boldsymbol{e} \tag{2.2}$$

where the n-dimensional row vector $\boldsymbol{y}$ is the vector of the response variable; X is the $p \times n$ dimensional matrix of the values of the p regressors; $\boldsymbol{e}$ is the $1 \times n$ row vector of the error component associated with the response vector $\boldsymbol{y}$; and the regression vector $\boldsymbol{\beta}$ and scale parameter σ are the same as defined in (2.1). Then the error vector follows a multivariate normal distribution with mean $\mathbf{0}$, a vector of n-tuple of zeros, and variance-covariance matrix, I_n. Therefore, the joint density function of the vector of errors becomes

$$f(\boldsymbol{e}) = [2\pi]^{-\frac{n}{2}} e^{-\frac{1}{2}\{\boldsymbol{e}\boldsymbol{e}'\}}. \tag{2.3}$$

Consequently, the response vector follows a multivariate normal distribution with mean vector $\boldsymbol{\beta} X$, and variance-covariance matrix, $\sigma^2 I_n$. Thus the joint density function of the response vector becomes

$$f(\boldsymbol{y}; \boldsymbol{\beta}, \sigma^2) = [2\pi\sigma^2]^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\left\{(\boldsymbol{y} - \boldsymbol{\beta} X)(\boldsymbol{y} - \boldsymbol{\beta} X)'\right\}}. \tag{2.4}$$

In this paper, we call the above multiple regression model as the the realized model of the responses from the performed experiment. The above joint density becomes the likelihood function of $\boldsymbol{\beta}$ and σ^2 when treated as a function of the parameters, rather than the sample response. The maximum likelihood estimators (m.l.e.) of the parameters as well as the likelihood ratio test can be derived, to test any hypothesis regarding the regression parameters, from the likelihood function. It is well known that the m.l.e. of the parameters of this model is the same as the ordinary least squares estimator (o.l.e.), and hence is best linear unbiased. However, in this paper we are interested to find the distribution of the sample regression vector (SRV) and the residual sum of squares (RSS) for the realized responses from the above multiple regression model as well as that of the future regression vector (FRV) and future residual sum of squares (FRSS) of the unobserved future responses from the future regression model to be defined in section 5.

3 Some Preliminaries

Some useful notations are introduced in this section to facilitate the derivation of the results in the forthcoming sections. First, we denote the sample regression vector of $\boldsymbol{e}$ on X by $\boldsymbol{b}(\boldsymbol{e})$ and the residual sum of squares of the error vector by $s^2(\boldsymbol{e})$. Then we have

$$\boldsymbol{b}(\boldsymbol{e}) = \boldsymbol{e}X'(XX')^{-1} \text{ and } s^2(\boldsymbol{e}) = [\boldsymbol{e} - \boldsymbol{b}(\boldsymbol{e})X][\boldsymbol{e} - \boldsymbol{b}(\boldsymbol{e})X]'. \tag{3.1}$$

Let $s(\boldsymbol{e})$ be the positive square root of the residual sum of squares based on the error regression, $s^2(\boldsymbol{e})$, and $\boldsymbol{d}(\boldsymbol{e}) = s^{-1}(\boldsymbol{e})[\boldsymbol{e} - \boldsymbol{b}(\boldsymbol{e})X]$ be the 'standardized' residual vector of the error regression.

Now we can write the error vector, $\boldsymbol{e}$, as a function of $\boldsymbol{b}(\boldsymbol{e})$ and $s(\boldsymbol{e})$ in the following way:

$$\boldsymbol{e} = \boldsymbol{b}(\boldsymbol{e})X + s(\boldsymbol{e})\boldsymbol{d}(\boldsymbol{e}) \text{ and hence we get } \boldsymbol{e}\boldsymbol{e}' = \boldsymbol{b}(\boldsymbol{e})XX'\boldsymbol{b}'(\boldsymbol{e}) + s^2(\boldsymbol{e}) \tag{3.2}$$

since $\boldsymbol{d}(\boldsymbol{e})\boldsymbol{d}'(\boldsymbol{e}) = 1$, inner product of two orthonormal vectors, and $X\boldsymbol{d}'(\boldsymbol{e}) = 0$, since X and $\boldsymbol{d}(\boldsymbol{e})$ are orthogonal.

From (3.2) and (2.2), the following relations (cf. Fraser, 1968, p.127) can easily be established:

$$\boldsymbol{b}(\boldsymbol{e}) = \sigma^{-1}\{\boldsymbol{b}(\boldsymbol{y}) - \boldsymbol{\beta}\}, \quad \text{and } s^2(\boldsymbol{e}) = \sigma^{-2}s^2(\boldsymbol{y}), \tag{3.3}$$

where

$$\boldsymbol{b}(\boldsymbol{y}) = \boldsymbol{y}X'(XX')^{-1} \quad \text{and } s^2(\boldsymbol{y}) = [\boldsymbol{y} - \boldsymbol{b}(\boldsymbol{y})X][\boldsymbol{y} - \boldsymbol{b}(\boldsymbol{y})X]' \tag{3.4}$$

are the sample regression vector of $\boldsymbol{y}$ on X, and the residual sum of squares of the regression based on the realized responses respectively. It may be mentioned here that both $s^2(\boldsymbol{e})$ and $s^2(\boldsymbol{y})$ have the same structure since the definitions of $s^2(\boldsymbol{e})$ in (3.3) and that of $s^2(\boldsymbol{y})$ in (3.4) ensure the same format of the two residual statistics of errors and realized responses respectively. Haq (1982) called the relation in (3.3) as the structural relations. It can easily be shown that $\boldsymbol{d}(\boldsymbol{e}) = s^{-1}(\boldsymbol{y})[\boldsymbol{y} - \boldsymbol{b}(\boldsymbol{y})X] = \boldsymbol{d}(\boldsymbol{y})$. From the above results, the density of the error vector in (2.3) can be written as a function of $\boldsymbol{b}(\boldsymbol{e})$ and $s(\boldsymbol{e})$ as follows

$$f(\boldsymbol{e}) = \psi \times e^{-\frac{1}{2}\left\{\boldsymbol{b}(\boldsymbol{e})XX'\boldsymbol{b}'(\boldsymbol{e}) + s^2(\boldsymbol{e})\right\}} \tag{3.5}$$

where ψ is an appropriate normalyzing constant. In section 5, we define similar future regression vector and future residual sum of squares for the future regression model.

4 Distribution of SRV and RSS

From the probability density of $\boldsymbol{e}$ in (2.3) and the relation (3.2) the joint probability density of $\boldsymbol{b}(\boldsymbol{e})$ and $s^2(\boldsymbol{e})$, conditional on the $\boldsymbol{d}(\boldsymbol{e})$, is obtained by using the invariant differentials (see Eaton, 1983, p.194-206 or Fraser, 1968, p.30) as follows

$$f\left(\boldsymbol{b}(\boldsymbol{e}),\ s^2(\boldsymbol{e})|\boldsymbol{d}(\boldsymbol{e})\right) = K_1(\boldsymbol{d})\ [s^2(\boldsymbol{e})]^{\frac{n-p-2}{2}} e^{-\frac{1}{2}\left\{\boldsymbol{b}(\boldsymbol{e})XX'\boldsymbol{b}'(\boldsymbol{e})+s^2(\boldsymbol{e})\right\}} \tag{4.1}$$

where $K_1(\boldsymbol{d})$ is the normalizing constant. It can be shown that the above density does not depend on $\boldsymbol{d}(\boldsymbol{e})$ (cf. Fraser, 1978, p.113) and can be written as the product of two densities in the following way:

$$f\left(\boldsymbol{b}(\boldsymbol{e}),\ s^2(\boldsymbol{e})\right) = f_1\left(\boldsymbol{b}(\boldsymbol{e})\right)\ \times\ f_2\left(s^2(\boldsymbol{e})\right) \tag{4.2}$$

where

$$f_1\left(\boldsymbol{b}(\boldsymbol{e})\right) = K_{11} e^{-\frac{1}{2}\left\{\boldsymbol{b}(\boldsymbol{e})XX'\boldsymbol{b}'(\boldsymbol{e})\right\}} \quad \text{and } f_2\left(s^2(\boldsymbol{e})\right) = K_{12}[s^2(\boldsymbol{e})]^{\frac{n-p-2}{2}} e^{-\frac{1}{2}s^2(\boldsymbol{e})} \tag{4.3}$$

in which

$$K_{11}^{-1} = [2\pi]^{\frac{p}{2}}|XX'|^{-\frac{n}{2}} \quad \text{and } K_{12}^{-1} = [2]^{\frac{n-p-2}{2}}\Gamma\left(\frac{n-p}{2}\right) \tag{4.4}$$

are the respective normalizing constants. Clearly, the joint distribution factors, and hence the marginal distributions of $\boldsymbol{b}(\boldsymbol{e})$ and $s^2(\boldsymbol{e})$ are independent of one another. Therefore, the sample regression vector based on the error regression follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $[XX']^{-1}$. That is, $\boldsymbol{b}(\boldsymbol{e}) \sim N_p\left(\mathbf{0},\ [XX']^{-1}\right)$. The residual sum of squares of the error regression, $s^2(\boldsymbol{e})$, follows a gamma distribution with shape parameter $(\frac{n-p}{2})$.

To find the distributions of the sample regression vector of the response regression, $\boldsymbol{b}(\boldsymbol{y})$, and the residual sum of squares of the response regression, $s^2(\boldsymbol{y})$, we use the following relations of the multiple regression model:

$$\boldsymbol{b}(\boldsymbol{e}) = \sigma^{-1}[\boldsymbol{b}(\boldsymbol{y}) - \boldsymbol{\beta}] \text{ and } s^2(\boldsymbol{e}) = \sigma^{-2}s^2(\boldsymbol{y}). \tag{4.5}$$

So the associated differentials can be expressed as

$$d\boldsymbol{b}(\boldsymbol{e}) = \sigma^{-p}d\boldsymbol{b}(\boldsymbol{y}) \text{ and } ds^2(\boldsymbol{e}) = \sigma^{-2}ds^2(\boldsymbol{y}). \tag{4.6}$$

Therefore, the density function of $\boldsymbol{b}(\boldsymbol{y})$ is written as

$$f\left(\boldsymbol{b}(\boldsymbol{y})\right) = [2\pi\sigma^2]^{-\frac{p}{2}}|XX'|^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\left\{\left(\boldsymbol{b}(\boldsymbol{y})-\boldsymbol{\beta}\right)XX'\left(\boldsymbol{b}(\boldsymbol{y})-\boldsymbol{\beta}\right)'\right\}} \tag{4.7}$$

and that of $s^2(\boldsymbol{y})$ is given by

$$f\left(s^2(\boldsymbol{y})\right) = \frac{1}{[2]^{\frac{n-p-2}{2}}\Gamma(\frac{n-p}{2})} \frac{[s^2(\boldsymbol{y})]^{\frac{n-p-2}{2}}}{[\sigma^2]^{\frac{n-p}{2}}} e^{-\frac{1}{2\sigma^2}s^2(\boldsymbol{y})}. \tag{4.8}$$

Thus the sample regression vector of the realized response regression follows a multivariate normal distribution with mean vector $\boldsymbol{\beta}$ and covariance matrix $\sigma^2(XX')^{-1}$, that is, $\boldsymbol{b}(\boldsymbol{y}) \sim N_p\left(\boldsymbol{\beta},\ \sigma^2(XX')^{-1}\right)$, and the residual sum of squares of the response regression, $s^2(\boldsymbol{y})$ is distributed as a scaled gamma variable with shape parameter $\frac{(n-p)}{2}$. The sample regression vector and the residual sum of squares of the realized response regression are independently distributed. This is true for both the error regression and the response regression. However, the parameters of the distributions of the statistics of the error regression are different from that of the response regression.

5 Regression Model for Future Responses

In this section we introduce the idea of predictive model for the future responses, and use both the realized sample and unobserved future sample to derive the distributions of the future regression vector as well as the future residual sum of squares. First, consider a set of $n_f \geq p$ future unobserved responses, $\boldsymbol{y}_f = (y_{f1}, y_{f2}, \cdots, y_{fn_f})$, from the multiple regression model as given in (2.1) with the same regression and scale parameters as defined in section 2. Such a set of future responses can be expressed as

$$\boldsymbol{y}_f = \boldsymbol{\beta} X_f + \sigma \boldsymbol{e}_f \tag{5.1}$$

where X_f is the $p \times n_f$ matrix of the values of the regressors that generate the future response vector $\boldsymbol{y}_f$, and $\boldsymbol{e}_f$ is the n_f-dimensional row vector of future error terms. The future responses are assumed to be generated by the same data generating process as that of the realized responses and involve the same regression and scale parameters. Thus the responses of the realized sample and the unobserved future responses are related through the same indexing parameters, $\boldsymbol{\beta}$ and σ^2. We assume non-informative prior distribution of the above parameters. Our objective here is to find the distributions of the future regression vector and the residual sum of squares of the future regression model, conditional on the realized responses.

Following the same process as in section 2, we define the following statistics based on the future regression model:

$$\boldsymbol{b}_f(\boldsymbol{e}_f) = \boldsymbol{e}_f X_f'(X_f X_f')^{-1},\ s_f^2(\boldsymbol{e}_f) = [\boldsymbol{e}_f - \boldsymbol{b}_f(\boldsymbol{e}_f)X_f][\boldsymbol{e}_f - \boldsymbol{b}_f(\boldsymbol{e}_f)X_f]' \tag{5.2}$$

in which $\boldsymbol{b}_f(\boldsymbol{e}_f)$ is the future regression vector and $s_f^2(\boldsymbol{e}_f)$ is the residual sum of squares of the future error of the future model respectively. Then we can write the future error vector, $\boldsymbol{e}_f$, in the following way:

$$\boldsymbol{e}_f = \boldsymbol{b}_f(\boldsymbol{e}_f)X_f + s_f(\boldsymbol{e}_f)\boldsymbol{d}_f(\boldsymbol{e}_f) \tag{5.3}$$

where $s_f(\boldsymbol{e}_f)$ is the positive square root of $s_f^2(\boldsymbol{e}_f)$, and hence we get

$$\boldsymbol{e}_f\boldsymbol{e}_f' = \boldsymbol{b}_f(\boldsymbol{e}_f)X_fX_f'\boldsymbol{b}_f'(\boldsymbol{e}_f) + s_f^2(\boldsymbol{e}_f) \tag{5.4}$$

since X_f and $\boldsymbol{d}(\boldsymbol{e}_f)$ are orthogonal and $\boldsymbol{d}_f(\boldsymbol{e}_f)$ is orthonormal. Moreover, the following relations can easily be observed:

$$\boldsymbol{b}_f(\boldsymbol{e}_f) = \sigma^{-1}\{\boldsymbol{b}_f(\boldsymbol{y}_f) - \boldsymbol{\beta}\}, \quad \text{and } s_f^2(\boldsymbol{e}_f) = \sigma^{-2}s_f^2(\boldsymbol{y}_f), \tag{5.5}$$

where

$$\boldsymbol{b}_f(\boldsymbol{y}_f) = \boldsymbol{y}_fX_f'(X_fX_f')^{-1} \quad \text{and } s_f^2(\boldsymbol{y}) = [\boldsymbol{y}_f - \boldsymbol{b}_f(\boldsymbol{y}_f)X_f][\boldsymbol{y}_f - \boldsymbol{b}_f(\boldsymbol{y}_f)X_f]' \tag{5.6}$$

in which $\boldsymbol{b}_f(\boldsymbol{y}_f)$ is the future regression vector of the future responses and $s_f^2(\boldsymbol{y}_f)$ is the residual sum of squares of future responses respectively. Note that the future response vector, independent of the realized responses, follows an n_f-dimensional multivariate normal distribution, that is, $\boldsymbol{y}_f \sim N_{n_f}\left(\boldsymbol{\beta}X_f,\ \sigma^2 I_{n_f}\right)$. Following the same argument as in section 2, the density function of the future error vector is given by

$$f(\boldsymbol{e}_f) = [2\pi]^{-\frac{n_f}{2}}e^{-\frac{1}{2}\left(\boldsymbol{e}_f\boldsymbol{e}_f'\right)}. \tag{5.7}$$

and hence by using the invariant differentials, as in section 4, we get the joint distribution of $\boldsymbol{b}_f(\boldsymbol{e}_f)$ and $s_f^2(\boldsymbol{e}_f)$ as follows

$$f\left(\boldsymbol{b}_f(\boldsymbol{e}_f), s_f^2(\boldsymbol{e}_f)\right) = K_2 \times [s_f^2(\boldsymbol{e}_f)]^{\frac{n_f-p-2}{2}}e^{-\frac{1}{2}\left\{\boldsymbol{b}_f(\boldsymbol{e}_f)X_fX_f'\boldsymbol{b}_f'(\boldsymbol{e}_f)+s_f^2(\boldsymbol{e}_f)\right\})} \tag{5.8}$$

where K_2 is the normalizing constant. The unconditional marginal distributions of the future regression vector and future residual sum of squares of the error regression for the future model can be obtained from the above joint density in (5.8). Since the future sample is independent of the realized sample, the join density function of the combined error vector, that is, the errors associated with the realized and that of the future responses, $(\boldsymbol{e},\ \boldsymbol{e}_f)$ can be expressed as

$$f(\boldsymbol{e},\ \boldsymbol{e}_f) = [2\pi]^{-\frac{n+n_f}{2}}e^{-\frac{1}{2}\left\{\boldsymbol{e}\boldsymbol{e}'+\boldsymbol{e}_f\boldsymbol{e}_f'\right\}}. \tag{5.9}$$

Haq and Khan (1990) used this density function to derive the prediction distribution of future responses, conditional on the realized responses. Figure 1 provides the graph of the prediction distribution for the accounting rates of stocks for the data used by Barlev and Levy (1979). Here we use this density function to derive the prediction distributions of the future regression vector and future sum of squared errors.

6 Predictive Distributions of FRV and FRSS

In this section we derive the predictive distributions of the future regression vector and the residual sum of squares for the future multiple regression model, conditional on the realized responses. Since both the realized and future regression models involve the same parameters, the joint distribution of the responses would contain the same regression and scale parameters. In the absence of any knowledge about the parameters, we consider non-informative prior distribution for the parameters as follows:

$$p(\boldsymbol{\beta}) \propto \text{ constant, and } p(\sigma^2) \propto \sigma^{-2}. \tag{6.1}$$

This prior distribution is used to derive the predictive distributions of $\boldsymbol{b}(\boldsymbol{y}_f)$ and $s^2(\boldsymbol{y}_f)$ from the joint distribution of $\boldsymbol{\beta}$, σ^2, $\boldsymbol{b}(\boldsymbol{y}_f)$ and $s^2(\boldsymbol{y}_f)$. Justification for the use of such a non-informative prior is given by Geisser (1993, p.60 & p.192), Box and Tiao (1992, p.21), Press (1989, p. 132) and Meng (1994) among any others. It is worth noting that no prior distribution is required in the structural approach (cf. Fraser, 1978) as the structural distribution, similar to the Bayes posterior distribution, can be obtained from the structural relation of the model without involving any prior distribution. Fraser and Haq (1969) discussed that for the non-informative prior, the Bayes posterior density is the same as the structural density.

6.1 Distribution of the Future Regression Vector In this sub-section we derive the prediction distribution of the future regression vector, conditional on the realized responses. The joint density function of the error statistics $\boldsymbol{b}(\boldsymbol{e})$, $s^2(\boldsymbol{e})$, $\boldsymbol{b}_f(\boldsymbol{e}_f)$ and $s_f^2(\boldsymbol{e}_f)$, for given $\boldsymbol{d}(\cdot)$, is derived from the joint density in (5.9) by applying the properties of invariant differentials, as follows:

$$p\Big(\boldsymbol{b}(\boldsymbol{e}), s^2(\boldsymbol{e}), \boldsymbol{b}_f(\boldsymbol{e}_f), s_f^2(\boldsymbol{e}_f)|\boldsymbol{d}(\cdot)\Big) = \Psi_{11}(\cdot)[s^2(\boldsymbol{e})]^{\frac{n-p-2}{2}} e^{-\frac{1}{2}g_1\left(\boldsymbol{b},X\right)}$$

$$\times\Psi_{21}(\cdot)[s_f^2(\boldsymbol{e}_f)]^{\frac{n_f-p-2}{2}} e^{-\frac{1}{2}g_2\left(\boldsymbol{b}_f,X_f\right)} \tag{6.2}$$

where $g_1\Big(\boldsymbol{b}, X\Big) = \boldsymbol{b}(\boldsymbol{e})XX'\boldsymbol{b}'(\boldsymbol{e})$; $g_2\Big(\boldsymbol{b}_f, X_f\Big) = \boldsymbol{b}_f(\boldsymbol{e}_f)X_fX_f'\boldsymbol{b}_f'(\boldsymbol{e}_f)$; and Ψ_{11} and Ψ_{21} are the normalizing constants. Since the above density does not depend on $\boldsymbol{d}(\cdot)$ so the conditioning in (6.2) can be disregarded as we need to find the joint density of $\boldsymbol{b}(\boldsymbol{y})$, $s^2(\boldsymbol{y})$, $\boldsymbol{b}_f(\boldsymbol{y}_f)$ and $s_f^2(\boldsymbol{y}_f)$ from the above joint density. The structural relation of the model yields

$$\boldsymbol{b}(\boldsymbol{e}) = \sigma^{-1}[\boldsymbol{b}(\boldsymbol{y}) - \boldsymbol{\beta}] \text{ and } s^2(\boldsymbol{e}) = \sigma^{-2}s^2(\boldsymbol{y}). \tag{6.3}$$

The joint distribution of $\boldsymbol{b}(\boldsymbol{y})$, $s^2(\boldsymbol{y})$, $\boldsymbol{b}_f(\boldsymbol{e}_f)$, and $s_f^2(\boldsymbol{e}_f)$ is then obtained by using the Jacobian of the transformation,

$$J\Big\{[\boldsymbol{b}(\boldsymbol{e}), s^2(\boldsymbol{e})] \to [\boldsymbol{b}(\boldsymbol{y}), s^2(\boldsymbol{y})]\Big\} = [\sigma^2]^{-\frac{p+2}{2}}, \tag{6.4}$$

as follows:

$$p\Big(\boldsymbol{b}(\boldsymbol{y}), s^2(\boldsymbol{y}), \boldsymbol{b}_f(\boldsymbol{e}_f), s_f^2(\boldsymbol{e}_f)\Big) = \Psi_2 \times [s^2]^{\frac{n-p-2}{2}} [s_f^2(\boldsymbol{y}_f)]^{\frac{n_f-p-2}{2}} [\sigma^2]^{-\frac{n}{2}} \times e^{-\frac{1}{2\sigma^2}\Big\{\xi_1(\boldsymbol{b},\boldsymbol{\beta})+s^2+\xi_2(\boldsymbol{b}_f(\boldsymbol{e}_f))+s_f^2(\boldsymbol{e}_f)\Big\}} \quad (6.5)$$

where $\xi_1(\boldsymbol{b}, \boldsymbol{\beta}) = (\boldsymbol{b} - \boldsymbol{\beta})XX'(\boldsymbol{b} - \boldsymbol{\beta})'$; $\xi_2(\boldsymbol{b}_f(\boldsymbol{e}_f)) = \boldsymbol{b}_f(\boldsymbol{e}_f)X_f X_f' \boldsymbol{b}_f'(\boldsymbol{e}_f)$; $\boldsymbol{b} = \boldsymbol{b}(\boldsymbol{y})$ and $s^2 = s^2(\boldsymbol{y})$. The normalizing constant Ψ_2 can be obtained by integrating the right hand side of the above function over the appropriate domains of the underlying variables. Since we are interested in the distributions of $\boldsymbol{b}_f(\boldsymbol{y}_f)$ and $s_f^2(\boldsymbol{y}_f)$, the future regression vector and residual sum of squares for the future regression, respectively, conditional on the realized responses, we don't pursue the matter any further in this paper.

To derive the joint distribution of $\boldsymbol{\beta}$, σ^2, $\boldsymbol{b}_f(\boldsymbol{y}_f)$ and $s_f^2(\boldsymbol{y}_f)$ from the above joint density, note that from the structure of the future regression equation we have

$$\boldsymbol{b}_f(\boldsymbol{e}_f) = \sigma^{-1}[\boldsymbol{b}_f(\boldsymbol{y}_f) - \boldsymbol{\beta}] \text{ and } s^2(\boldsymbol{e}_f) = \sigma^{-2}s^2(\boldsymbol{y}_f) \quad (6.6)$$

where

$$\boldsymbol{b}_f(\boldsymbol{y}_f) = \boldsymbol{y}X_f'(X_f X_f')^{-1}, \quad s^2(\boldsymbol{y}_f) = [\boldsymbol{y}_f - \boldsymbol{b}_f(\boldsymbol{y}_f)X_f'][\boldsymbol{y}_f - \boldsymbol{b}_f(\boldsymbol{y}_f)X_f]'. \quad (6.7)$$

Therefore, the Jacobian of the transformation is found to be

$$J\Big\{[\boldsymbol{b}_f(\boldsymbol{e}_f), s_f^2(\boldsymbol{e}_f)] \rightarrow [\boldsymbol{b}_f(\boldsymbol{y}_f), s^2(\boldsymbol{y}_f)]\Big\} = [\sigma^2]^{-\frac{p+2}{2}}. \quad (6.8)$$

Now, the joint density of $\boldsymbol{b}(\boldsymbol{y})$, $s^2(\boldsymbol{y})$, $\boldsymbol{b}_f(\boldsymbol{y}_f)$ and $s_f^2(\boldsymbol{y}_f)$ is obtained as

$$p\Big(\boldsymbol{b}, s^2, \boldsymbol{b}_f, s_f^2\Big) = \Psi_3(\cdot) \times [s^2]^{\frac{n-p-2}{2}} [s_f^2(\boldsymbol{y}_f)]^{\frac{n_f-p-2}{2}} [\sigma^2]^{-\frac{n+n_f}{2}} \times e^{-\frac{1}{2\sigma^2}}\Big\{\Big(\boldsymbol{b} - \boldsymbol{\beta}\Big)XX'\Big(\boldsymbol{b} - \boldsymbol{\beta}\Big)' + s^2 + \Big(\boldsymbol{b}_f - \boldsymbol{\beta}\Big)X_f X_f'\Big(\boldsymbol{b}_f - \boldsymbol{\beta}\Big)' + s_f^2\Big\} \quad (6.9)$$

where $\boldsymbol{b}_f = \boldsymbol{b}_f(\boldsymbol{y}_f)$ and $s_f^2 = s_f^2(\boldsymbol{y}_f)$ for notational convenience. From the non-informative prior distribution of the parameters of the model and the density in (6.9), we find the following joint density of $\boldsymbol{\beta}$, σ^2, $\boldsymbol{b}_f(\boldsymbol{y}_f)$ and $s_f^2(\boldsymbol{y}_f)$,

$$p\Big(\boldsymbol{\beta}, \sigma^2, \boldsymbol{b}_f, s_f^2\Big) = \Psi_3(\cdot) \times [s^2]^{\frac{n-p-2}{2}} [s_f^2(\boldsymbol{y}_f)]^{\frac{n_f-p-2}{2}} [\sigma^2]^{-\frac{n+n_f+2}{2}} \times e^{-\frac{1}{2\sigma^2}}\Big\{\Big(\boldsymbol{b} - \boldsymbol{\beta}\Big)XX'\Big(\boldsymbol{b} - \boldsymbol{\beta}\Big)' + s^2 + \Big(\boldsymbol{b}_f - \boldsymbol{\beta}\Big)X_f X_f'\Big(\boldsymbol{b}_f - \boldsymbol{\beta}\Big)' + s_f^2\Big\}. \quad (6.10)$$

A similar result can be obtained by using the structural distribution approach. In fact, the final results of this paper will be similar to that obtained by the structural distribution approach. Interested readers may refer to Fraser and Haq (1969) for details.

To evaluate the normalizing constant $\Psi_3(\cdot)$, in the above density, we go through the following steps. Let

$$\begin{aligned} I_{\sigma^2} &= \int_{\sigma^2} p\Big(\boldsymbol{\beta}, \sigma^2, \boldsymbol{b}_f,\ s_f\Big) d\sigma^2 \\ &= [s_f^2]^{\frac{n_f-p-2}{2}} \int_{\sigma^2} [\sigma^2]^{-\frac{n+n_f+2}{2}} e^{-\frac{1}{2\sigma^2}\boldsymbol{Q}} d\sigma^2 \end{aligned} \tag{6.11}$$

where

$$\boldsymbol{Q} = \Big(\boldsymbol{b}-\boldsymbol{\beta}\Big)XX'\Big(\boldsymbol{b}-\boldsymbol{\beta}\Big)' + s^2 + \Big(\boldsymbol{b}_f-\boldsymbol{\beta}\Big)X_fX_f'\Big(\boldsymbol{b}_f-\boldsymbol{\beta}\Big)' + s_f^2. \tag{6.12}$$

Therefore,

$$I_{\sigma^2} = [2]^{\frac{n+n_f}{2}} \Gamma\left(\frac{n+n_f+2}{2}\right) [s_f^2]^{\frac{n_f-p-2}{2}} [\boldsymbol{Q}]^{-\frac{n+n_f+2}{2}}. \tag{6.13}$$

To facilitate the further integrations, the terms involving the regression vector $\boldsymbol{\beta}$ in $\boldsymbol{Q}$ can be expressed as follows:

$$\begin{aligned} &\Big(\boldsymbol{b}-\boldsymbol{\beta}\Big)XX'\Big(\boldsymbol{b}-\boldsymbol{\beta}\Big)' + \Big(\boldsymbol{b}_f-\boldsymbol{\beta}\Big)X_fX_f'\Big(\boldsymbol{b}_f-\boldsymbol{\beta}\Big)' = \\ &\qquad \Big(\boldsymbol{\beta}-FA^{-1}\Big)A\Big(\boldsymbol{\beta}-FA^{-1}\Big)' + \Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)H^{-1}\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)' \end{aligned} \tag{6.14}$$

where

$$F = \boldsymbol{b}XX' + \boldsymbol{b}_fX_fX_f',\ A = XX' + X_fX_f',\ \text{and } H = [XX']^{-1} + [X_fX_f']^{-1}. \tag{6.15}$$

Then, let

$$\begin{aligned} I_{\sigma^2\boldsymbol{\beta}} &= \int_{\boldsymbol{\beta}} I_{\sigma^2}\ d\boldsymbol{\beta} = [2]^{\frac{n+n_f}{2}} \Gamma(\frac{n+n_f+2}{2})[s_f^2]^{\frac{n_f-p-2}{2}} \\ &\quad \times \int_{\boldsymbol{\beta}} \Big[\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)H^{-1}\Big(\boldsymbol{b}_f'-\boldsymbol{b}\Big) + s^2 + s_f^2 + g\Big(\boldsymbol{\beta}, A\Big)\Big]^{-\frac{n+n_f+2}{2}} d\boldsymbol{\beta} \\ &= [2]^{\frac{n+n_f}{2}} \frac{(\pi)^{\frac{p}{2}}\Gamma(\frac{n+n_f-p+2}{2})}{|A|^{\frac{1}{2}}} [s_f^2]^{\frac{n_f-p-2}{2}} \\ &\quad \times \Big[\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)H^{-1}\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)' + s^2 + s_f^2\Big]^{-\frac{n+n_f-p+2}{2}} \end{aligned} \tag{6.16}$$

where

$$g\Big(\boldsymbol{\beta}, A\Big) = \Big(\boldsymbol{\beta}-FA^{-1}\Big)A\Big(\boldsymbol{\beta}-FA^{-1}\Big)'. \tag{6.17}$$

In the same way, let

$$I_{\sigma^2\boldsymbol{\beta}\boldsymbol{b}_f} = \int_{\boldsymbol{b}_f} I_{\sigma^2\beta}\ d\boldsymbol{b}_f$$

$$
\begin{aligned}
&= [2]^{\frac{n+n_f}{2}} \frac{(\pi)^{\frac{p}{2}}\Gamma(\frac{n+n_f-p+2}{2})}{|A|^{\frac{1}{2}}}[s_f^2]^{\frac{n_f-p-2}{2}} \\
&\quad \times \int_{\boldsymbol{b}_f} \left[\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)H^{-1}\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)'+s^2+s_f^2\right]^{-\frac{n+n_f-p+2}{2}} d\boldsymbol{b}_f \\
&= [2]^{\frac{n+n_f}{2}} \frac{(\pi)^{p}\Gamma(\frac{n+n_f-2p+2}{2})}{|A|^{\frac{1}{2}}|H|^{-\frac{1}{2}}}[s_f^2]^{\frac{n_f-p-2}{2}}[s^2+s_f^2]^{-\frac{n+n_f-2p+2}{2}}. \qquad (6.18)
\end{aligned}
$$

Finally, let

$$
\begin{aligned}
I_{\sigma^2\boldsymbol{\beta}\boldsymbol{b}_f s_f^2} &= \int_{s_f^2} I_{\sigma^2\boldsymbol{\beta}\boldsymbol{b}_f}\, ds_f^2 \\
&= [2]^{\frac{n+n_f}{2}} \frac{(\pi)^{p}\Gamma(\frac{n+n_f-2p+2}{2})}{[|A|^{\frac{1}{2}}|H|^{-\frac{1}{2}}} \int_{s_f^2}[s_f^2]^{\frac{n_f-p-2}{2}}[s^2+s_f^2]^{-\frac{n+n_f-2p}{2}} ds_f^2 \\
&= [2]^{\frac{n+n_f}{2}} \frac{(\pi)^{p}\ \Gamma(\frac{n-p+2}{2})\Gamma(\frac{n_f-p}{2})}{|A|^{\frac{1}{2}}|H|^{-\frac{1}{2}}[s^2]^{\frac{n-p+2}{2}}}. \qquad (6.19)
\end{aligned}
$$

Thus, the normalizing constant for the joint distribution of $\boldsymbol{\beta}$, σ^2, $\boldsymbol{b}_f$ and s_f^2 becomes,

$$
\Psi_3(\cdot) = \frac{|A|^{\frac{1}{2}}|H|^{-\frac{1}{2}}[s^2]^{\frac{n-p}{2}}}{[2]^{\frac{n+n_f}{2}}(\pi)^{p}\ \Gamma(\frac{n-p+2}{2})\Gamma(\frac{n_f-p}{2})}. \qquad (6.20)
$$

The marginal density of $\boldsymbol{\beta}$, $\boldsymbol{b}_f$ and s_f^2, conditional on $\boldsymbol{y}$, is derived by integrating out σ^2 from the above joint density. Thus, we have,

$$
\begin{aligned}
p\Big(\boldsymbol{\beta},\boldsymbol{b}_f,s_f^2|\boldsymbol{y}\Big) &= \Psi_4 \times [s_f^2]^{\frac{n_f-p-2}{2}}\Big[s^2+\Big(\boldsymbol{\beta}-FA^{-1}\Big)A \\
&\quad \times\Big(\boldsymbol{\beta}-FA^{-1}\Big)'+\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)H^{-1}\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)+s_f^2\Big]^{-\frac{n+n_f}{2}} \qquad (6.21)
\end{aligned}
$$

where Ψ_4 is the normalizing constant.

Similarly, the marginal density of $\boldsymbol{b}_f$ and s_f^2 is obtained by integrating out $\boldsymbol{\beta}$ over $\mathcal{R}^p$ from (6.21). This gives the joint density of $\boldsymbol{b}_f$ and s_f^2, conditional on $\boldsymbol{y}$, as

$$
\begin{aligned}
p\Big(\boldsymbol{b}_f,s_f^2|\boldsymbol{y}\Big) &= \Psi_5 \times [s_f^2]^{\frac{n_f-p-2}{2}} \\
&\quad \times \Big[s^2+s_f^2+\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)H^{-1}\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)\Big]^{-\frac{n+n_f-p}{2}} \qquad (6.22)
\end{aligned}
$$

where

$$
\Psi_5 = \frac{|H|^{-\frac{1}{2}}\Gamma(\frac{n+n_f-p}{2})[s^2]^{\frac{n-p}{2}}}{(\pi)^{\frac{p}{2}}\Gamma(\frac{n-p}{2})\Gamma(\frac{n_f-p}{2})} \qquad (6.23)
$$

is the normalizing constant. The prediction distribution of the future regression vector, $\boldsymbol{b}_f = \boldsymbol{b}_f(\boldsymbol{y}_f)$, can now be obtained by integrating out s_f^2 from (6.22). The integration yields

$$
p\Big(\boldsymbol{b}_f\Big|\boldsymbol{y}\Big) = \Psi_6 \times \Big[s^2+\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)H^{-1}\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)'\Big]^{-\frac{n}{2}} \qquad (6.24)
$$

where $\Psi_6 = \Psi_4 \times B\left(\frac{n_f - p}{2}, \frac{n}{2}\right)[s^2]^{\frac{n-p}{2}}$. On simplification we get

$$\Psi_6 = \frac{\Gamma(\frac{n}{2})}{(\pi)^{\frac{p}{2}}\Gamma(\frac{n-p}{2})|H|^{\frac{1}{2}}[s^2]^{\frac{n-p}{2}}}. \tag{6.25}$$

The prediction distribution of $\boldsymbol{b}_f$ can be written in the usual multivariate Student-t distribution form as follows:

$$p\left(\boldsymbol{b}_f \middle| \boldsymbol{y}\right) = \Psi_6 \times \left[1 + \left(\boldsymbol{b}_f - \boldsymbol{b}\right)\left[s^2 H\right]^{-1}\left(\boldsymbol{b}_f - \boldsymbol{b}\right)'\right]^{-\frac{n}{2}} \tag{6.26}$$

in which $n > p$. Since the density in (6.26) is a Student-t density, the prediction distribution of the future regression vector, $\boldsymbol{b}_f$, conditional on the realized responses, follows a multivariate Student-t distribution of dimension p, with $(n-p)$ degrees of freedom. Thus, $[\boldsymbol{b}_f|\boldsymbol{y}] \sim t_p(n-p, \boldsymbol{b}, s^2 H)$ where $\boldsymbol{b}$ is the location vector and H is the scale matrix. It is observed that the degrees of freedom parameter of the prediction distribution of $\boldsymbol{b}_f$ depends on the sample size of the realized sample and the dimension of the regression parameter vector of the model. The above prediction distribution can be used to construct β-expectation tolerance region for the future regression parameter.

6.2 Distribution of Future Residual Sum of Squares

The prediction distribution of the future residual sum of squares from the future regression, $s_f^2(\boldsymbol{y}_f)$, based on the future responses, $\boldsymbol{y}_f$, conditional on the realized responses, $\boldsymbol{y}$, is obtained by integrating out $\boldsymbol{b}_f$ from (6.22) as follows:

$$p\left(s_f^2(\boldsymbol{y}_f)\middle|\boldsymbol{y}\right) = \Psi_7 \times \left[s_f^2(\boldsymbol{y}_f)\right]^{\frac{n_f - p - 2}{2}} \left[s^2 + s_f^2(\boldsymbol{y}_f)\right]^{-\frac{n + n_f - 2p}{2}}. \tag{6.27}$$

The density function in (6.27) can be written in the usual beta distribution form as follows:

$$p\left(s_f^2\middle|\boldsymbol{y}\right) = \Psi_7 \times \left[s_f^2\right]^{\frac{n_f - p - 2}{2}} \left[1 + s^{-2} s_f^2\right]^{-\frac{n + n_f - 2p}{2}} \tag{6.28}$$

where $\Psi_7 = \frac{\Gamma\left(\frac{n + n_f - 2p}{2}\right)\left[s^2\right]^{\frac{n-p}{2}}}{\Gamma\left(\frac{n-p}{2}\right)\Gamma\left(\frac{n_f - p}{2}\right)}$. This is the prediction distribution of the future residual sum of squares based on the future response $\boldsymbol{y}_f$, conditional on the realized responses, from the multiple regression model with normal error variable. The density in (6.28) is a modified form of beta density of the second kind with $(n_f - p)$ and $(n - p)$ degrees of freedom.

7 An Illustration

To illustrate how the method works, we consider a real life data set from Barlev and Levy (1979). The simple regression model fitted to this data is a special case of

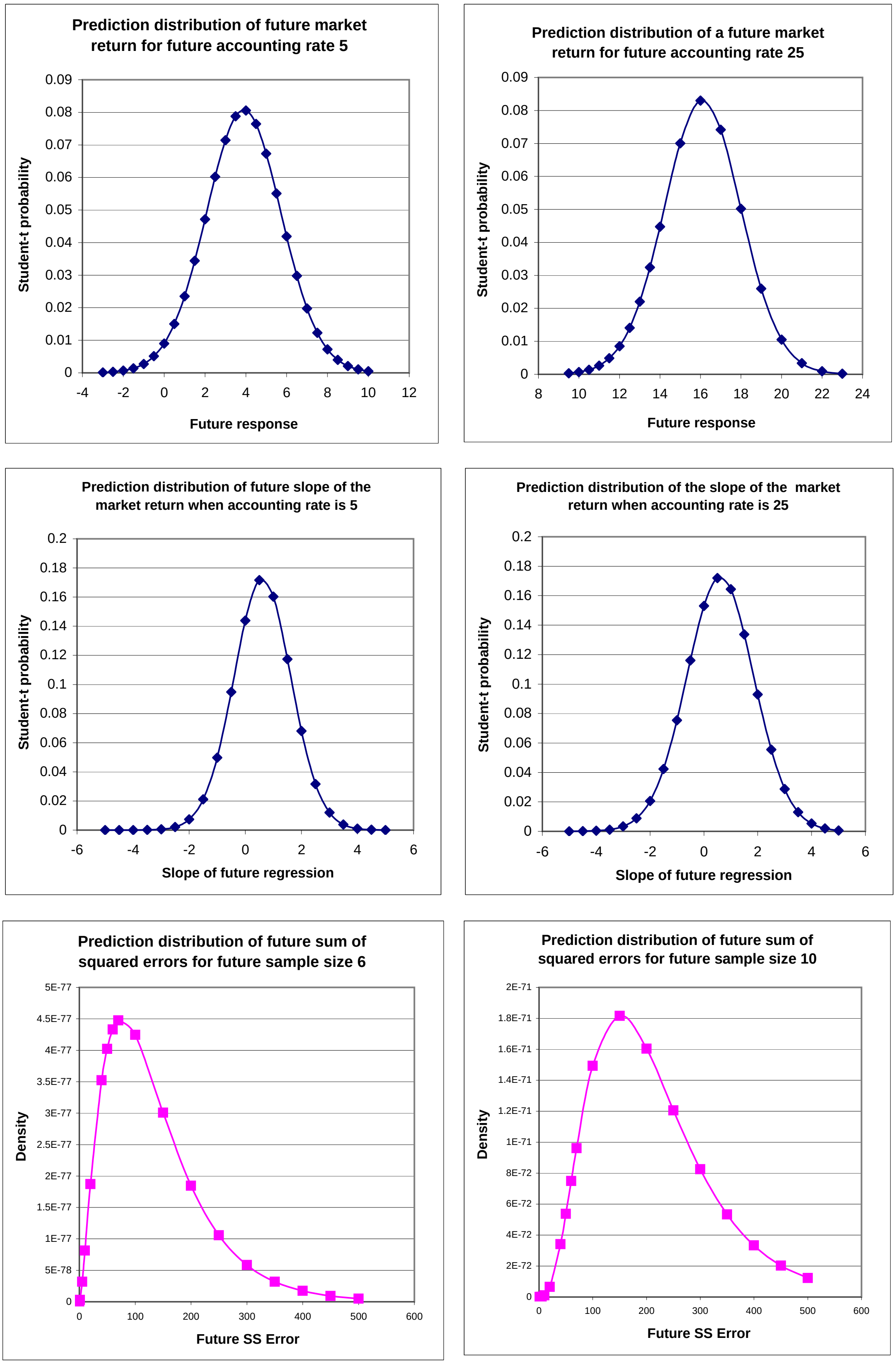

Figure 1: **Graph of prediction distribution of the future regression parameter**

the model considered in our paper. The data set was used in the study of relationship between accounting rates on stocks and market returns. It provides information on the two variables for 54 large size companies in the USA. Considering the market return to be the response and the accounting rate as the explanatory variables the fitted model becomes $\hat{y} = 0.084801 + 0.61033x$ with $\bar{x} = 12.9322$, $\bar{y} = 8.7409$, $\sum_{j=1}^{54} x_j^2 = 10293.10893$, $\sum_{j=1}^{54} x_j y_j = 6874.4131$ and $s^2 = 25.864$, the mean squared error. The prediction distribution of the regression parameter involves H which in this special case becomes $[\sum_{j=1}^{54} x_j^2]^{-1} + [{x_f}^2]^{-1}$.

The top two graphs in Figure 1 display the prediction distributions of future responses for future accounting rate 5 and 25 respectively. Both the distributions are Student-t distributions, but the one with higher future accounting rate has wider spread than the one with lower value of future accounting rate. The prediction distribution of the future regression (slope) parameter of the regression of future market rate on the future accounting rate is given in the middle two graphs of Figure 1. Both the graphs represent the Student-t distributions with different parameters. Although the shape of the distribution of both the graphs is roughly the same, the first graph here has a slightly more spread, but lower pick, than the second graph. The bottom two graphs of Figure 1 displays the prediction distribution of the future sum of squared errors for different sample sizes. These last two graphs in Figure 1 represent the beta distribution with varying arguments.

8 Concluding Remarks

The foregoing analyses reveal the fact that for the multiple regression model with independent normal errors the sample regression vector and the residual sum of squares are independently distributed. This is true for both the error regression and response regression of the realized model. But for the future regression model, the predictive distributions of the future regression vector and the residual sum of squares, conditional on the realized responses, are not independent. The sample regression vector of the realized model follows a multivariate normal distribution, but the future regression vector of the future model follows a multivariate Student-t distribution. Thus every element of the sample regression vector is independently distributed, but the components of the future regression vector are not independent. Moreover, the residual sum of squares of the realized multiple regression model follows a scaled gamma distribution, while that of the future regression model, conditional on the realized responses, follows a scaled beta distribution. The residual sum of squares based on the error regression and that of the response regression differs by a constant for the realized regression model.

Acknowledgement

The author thankfully acknowledges some valuable suggestions from Professor Lehana Thabane, McMaster University, Canada that improved the quality and content of the paper significantly. This work was initiated while the author was visiting the Institute of Statistical Research and Training, University of Dhaka, Bangladesh.

References

Aitchison, J. (1964). Bayesian tolerance regions. *Jou. of Royal Statistical Society*, B, **127**, 161-175.

Aitchison, 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*, *55*, 469-483.

Barlev, B. and Levy, H. (1979). On the variability of accounting income numbers. *Jou. of Accounting Research*, **16**, 305-315

Bishop, J. (1976). Parametric tolerance regions. *Unpublished PhD Thesis*, Department of Statistics, University of Toronto, Canada.

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

Eaton, M.L. (1983). Multivariate Statistics - A Vector Space Approach. Wiley, New York.

Fraser, D.A.S. (1968). The Structure of Inference, Wiley, New York.

Fraser, D.A.S. (1978). Inference and Linear Models, McGrow-Hill, New York.

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

Geisser, S. (1993). Predictive Inference: An Introduction. Chapman & Hall, London.

Guttman, I. (1970). Statistical tolerance regions: Classical and Bayesian. Griffin, London.

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.

Haq, M.S. and Rinco, S. (1976). β-expectation tolerance regions for a generalized multilinear model with normal error variables. *Jou. Multiv. Analysis*, **6**, 414-21.

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

Khan, S. (1996). Prediction inference for heteroscedastic multiple regression model with class of spherical error. *Aligarh Journal of Statistics,* **15 & 16**, 1-17.

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

Khan, S. (1992). Predictive inference for multilinear model with Student-t errors, *Unpublished PhD Thesis*, Department of Statistics & Actuarial Sciences, University of Western Ontario, Canada.

Lieberman, G.J. and Miller, R.G. Jr. (1963). Simultaneous tolerance intervals in regressions. *Biometrika*, **50**, 155-168.

Meng, X. L. (1994). Posterior predictive p-value, *Ann. of Statist.*, **22**(3), 1142-1160.

Ng, V.M. (2000). A note on predictive inference for multivariate elliptically contoured distributions. *Comm. In Statist. - Theory & Methods,* **29** (3), 477-483.

Press, S. J. (1989). Bayesian Statistics: Principles, Mehtods and Applications, Wiley, New York.