Noname manuscript No.
(will be inserted by the editor)

Shahjahan Khan

Optimal tolerance regions for future regression vector and residual sum of squares of multiple regression model with multivariate spherically contoured errors

Received: date / Accepted: date

Abstract This paper considers multiple regression model with multivariate spherically symmetric errors to determine optimal β-expectation tolerance regions for the future regression vector (FRV) and future residual sum of squares (FRSS) by using the prediction distributions of some appropriate functions of future responses. The prediction distribution of the FRV, conditional on the observed responses, is multivariate Student-t distribution. Similarly, the prediction distribution of the FRSS is a beta distribution. The optimal β-expectation tolerance regions for the FRV and FRSS have been obtained based on the F-distribution and beta distribution respectively. The results in this paper are applicable for multiple regression model with normal and Student-t errors.

Keywords Multiple regression model; prediction distribution; optimal β-expectation tolerance region; invariant differential; non-informative prior; spherical/elliptical distributions; multivariate Student-t, beta and F distributions.

Mathematics Subject Classification (2000) Primary 62H10, · Secondary 62J12.

Shahjahan Khan
Department of Mathematics and Computing
Australian Centre for Sustainable Catchments
University of Southern Queensland
Toowoomba, Queensland 4350, AUSTRALIA
Tel.: +61 7 4631 5532
Fax: +61 7 4631 5550
E-mail: khans@usq.edu.au

1 Introduction

The predictive inference had been the oldest form of statistical inference used in real life. Predictive inference uses the realized responses from the *performed experiment* to make inference about the behavior of the unobserved future 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 common set of parameters. The prediction distribution forms the basis of all predictive inference. For details on the predictive inference methods and wide range of applications of prediction distribution interested readers may refer to Aitchison and Sculthorpe (1965) and Geisser (1993). Prediction distribution for a set of future responses, or appropriate functions of future responses, of a model, conditional on the realized responses from the same model, has been derived by many authors including Fraser and Haq (1969), Aitchison and Dunsmore (1975), and Haq and Khan (1990). The prediction distribution of a set of future responses from the multilinnear model has been used by Haq and Rinco (1976) to derive β-expectation tolerance region. Guttman (1970) and Aitchison and Dunsmore (1975) obtained different kinds of tolerance regions from the prediction distribution.

A statistical tolerance region (interval in one dimension) is a region, defined on the sample space, that contains a specified proportion of the future responses, or any suitable function of future responses of a random variable under study with a preassigned level of probability. There are several kinds of tolerance regions available in the literature (cf. Guttman, 1970, and Aitchison and Dunsmore, 1975). Geisser (1993) discussed the Bayesian approach to predictive inference and discussed a wide range of real-life applications in many areas. This includes model selection, discordancy, perturbation analysis, classification, regulation, screening and interim analysis. The β-expectation tolerance region is a special type of tolerance region when the expected probability of the region to contain a set of future responses or an appropriate function of future responses is a known value β, a real number, usually not too far from 1. It is a problem under the broader area of the predictive inference and can be solved by using the prediction distribution.

In the recent years, there has been a growing interest in the non-normal and robust models. Nevertheless, Fisher (1956) discarded the normal distribution as a sole model for the distribution of errors. Fraser (1979, p.41) showed that the results based on the Student-t errors for linear models are applicable to those of normal models, but not the vice-versa. Prucha and Kelejian (1984) critically described the problems of normal distribution and recommended the Student-t distribution as a better alternative for many problems. The failure of the normal distribution to model the fat-tailed distributions has led to the use of some other members of the spherical/elliptical class of distributions. Some of the well known members of the spherically/elliptically contoured family of distributions are the multivariate normal, Kotz Type, Pearson Type VII, Multivariate t, Multivariate Cauchy, Pearson Type II, Logistic, Multivariate Bassel, Scale mixture and Stable laws. Extensive work on this area of non-normal models has been done in recent years. A brief summary of such literature has been given by Chmielewiski (1981), and other notable references include Fang and Zhang (1990), Haq and Khan (1990), and Fang and Anderson (1990). Zellner (1976) first considered the linear regression model with Student-t errors. However, Fang and Anderson (1990), Anderson (1993), Khan (1996) and Ng (2000) provide predictive analyses of future responses for linear models with spherically contoured errors. Recently Khan (2004, 2005, 2006) has derived the prediction distributions of the future regression vector and future residual sum of squares for the multiple regression model with normal and multivariate Student-t errors respectively. This is a new approach that provides predictive inference for the future regression parameters as a function of future responses, rather than that of the future responses themselves.

This paper considers the widely used multiple regression model with a more general assumption of a family of spherically contoured errors for the realized as well as the future responses. The two sets of responses are connected through the common set of regression and scale parameters. The family of spherically contoured distributions encompass a range of symmetrical distributions including the normal and Student-t distributions as special cases. The distribution of the FRV and FRSS of the future responses, conditional on the realized responses, are obtained under the non-informative prior for the parameters. Identical prediction distributions are obtained by the structural approach without assuming any prior distribution. The predictive distribution of the FRV follows a multivariate Student-t distribution, and the FRSS of the future regression follows a scaled beta distribution. The distribution of the statistics for the future regression model, conditional on the realized responses, are dependent, and hence their joint density can't be factorized.

The multiple regression model with spherically contoured errors is provided in section 2. The β-expectation tolerance region and its optimality are introduced in section 3. Some preliminaries are included in section 4. The predictive distributions of the FRV and FRSS, conditional on the realized sample, are provided in section 5. The β-expectation tolerance regions for the FRV and FRSS are derived in section 6. Some concluding remarks are included in section 7.

2 The Multiple Regression Model with Spherical Errors

Consider the multiple regression equation

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

where y is the response variable, $\boldsymbol{\delta}$ 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 with known values, 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 distributed as a spherically contoured variable with location 0 and scale 1. Now, consider a set of $n > p$ responses, $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)$, from the above regression model that can be expressed as

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

where the n-dimensional row vector $\boldsymbol{y}$ is a vector of the response variable; X is a $p \times n$ dimensional matrix of fixed known values of the p regressors; $\boldsymbol{e}$ is a $1 \times n$ vector of the error component associated with the response vector $\boldsymbol{y}$; and the regression parameter vector $\boldsymbol{\delta}$ and scale parameter σ are the same as defined in (2.1). Then the error vector follows a multivariate spherically contoured distribution with location $\underline{0}$, a vector of n-tuple of zeros, and scale matrix, I_n. Therefore, the joint density function of the vector of errors can be written as

$$f(\boldsymbol{e}) \propto g\{\boldsymbol{e}\boldsymbol{e}'\} \tag{2.3}$$

such that the $f(\boldsymbol{e})$ is a proper density function. See Fang and Zhang (1980) for details on spherical distributions and their properties. Note that when $g(\boldsymbol{e}\boldsymbol{e}') = e^{-\frac{1}{2}\boldsymbol{e}\boldsymbol{e}'}$ the spherical density becomes the normal density, and for $g(\boldsymbol{e}\boldsymbol{e}') = \left(1 + \frac{1}{\nu}\boldsymbol{e}\boldsymbol{e}'\right)^{-\frac{\nu+p}{2}}$ we get the Student-t density with ν degrees of freedom. It is well known that when the error vector is spherically distributed then the response vector follows a family of multivariate elliptically symmetric contoured distributions with location vector $\boldsymbol{\delta}X$, scale matrix, $\sigma^2 I_n$, and density function

$$f(\boldsymbol{y}; \boldsymbol{\delta}, \sigma^2) \propto \left[\frac{1}{\sigma^2}\right]^{\frac{n}{2}} g\left\{\frac{1}{\sigma^2}\left[(\boldsymbol{y} - \boldsymbol{\delta}X)(\boldsymbol{y} - \boldsymbol{\delta}X)'\right]\right\} \tag{2.4}$$

so that $f(\cdot)$ is a proper density function. In this paper, the above multiple regression model represents the realized model of the responses from the *performed experiment.*

Now, consider another set of n_f $(\geq p)$ unobserved future responses, $\boldsymbol{y}_f = (y_{f1}, y_{f2}, \cdots, y_{fn_f})$, from the same multiple regression model in (2.1) with the same regression and scale parameters. Such a set of future responses can be expressed as

$$\boldsymbol{y}_f = \boldsymbol{\delta}X_f + \sigma\boldsymbol{e}_f \tag{2.5}$$

where X_f is a $p \times n_f$ matrix of the values of p regressors that generate the future response vector $\boldsymbol{y}_f$, and $\boldsymbol{e}_f$ is a vector of future error terms. Similar to the error vector of the realized model the future error vector from the *future experiment* follows a family of multivariate spherically contoured distributions, and as such the future responses follow a family of multivariate elliptically contoured distributions.

From the specifications of the model, the future sample is not independent of the realized sample. However, for the joint density function of the combined error vector, that is, the errors associated with the realized and that of the future responses, $\boldsymbol{e}^* = [\boldsymbol{e},\ \boldsymbol{e}_f]$, we can write $\boldsymbol{e}^*\boldsymbol{e}^{*'} = \sum_{j=1}^{n} e_j^2 + \sum_{j'=1}^{n_f} e_{fj'}^2 = \boldsymbol{e}\boldsymbol{e}' + \boldsymbol{e}_f\boldsymbol{e}_f'$. Then the join density function of the combined error vector, $(\boldsymbol{e},\ \boldsymbol{e}_f)$, can be expressed as

$$f(\boldsymbol{e},\ \boldsymbol{e}_f) \propto g\left\{\boldsymbol{e}\boldsymbol{e}' + \boldsymbol{e}_f\boldsymbol{e}_f'\right\} \tag{2.6}$$

where $g\{\cdot\}$ is such that $f(\cdot)$ is a proper density. This joint density function is used to derive the prediction distributions of the functions of the future errors as well as the responses of the future model in the next section. For the β-expectation tolerance region we need the prediction distribution of the future regression vector (FRV) and future residual sum of squares (FRSS) of the future regression model.

3 β-Expectation Tolerance Region

In the literature, a tolerance region $R(Y)$ is defined on a probability space $(\mathcal{X}, \mathcal{A}, \mathrm{P}_\theta)$ where $\mathcal{X}$ is the sample space of the responses in the random sample $(\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_n)$; $\mathcal{A}$ is a σ-field defined on the sample space; and P_θ is the probability measure such that $\theta = [\boldsymbol{\delta} X, \sigma]$ (see the multiple regression model in the next section) is an element of the joint parameter space Ω. Thus a tolerance region $R(Y)$ is a statistic defined on the sample space $\mathcal{X}$ and takes values in the σ-field $\mathcal{A}$. The probability content of the region $R(Y)$ is called the coverage of the tolerance region and is denoted by $C(R) = P_Y^\theta[R(Y)]$. Note that $C(R)$ being a function of $R(Y)$, a random variable, is itself a random variable whose probability measure is induced by the measure P_θ.

Of different kinds of tolerance regions available in the literature, here we consider a particular kind of tolerance region that has an expected probability of $0 < \beta < 1$. A tolerance region $R(Y)$ is called a β-expectation tolerance region if the expectation of its coverage probability is equal to a preassigned value β. Thus for a given set of observed responses $\boldsymbol{y}$, a β-expectation tolerance region $R(Y)$ must satisfy

$$E[C(R)|\boldsymbol{y}] = \beta. \tag{3.1}$$

Let $p(\boldsymbol{\lambda}_f \mid \boldsymbol{y})$ denote the prediction distribution of $\boldsymbol{\lambda}_f$, a function of a set of future responses $\boldsymbol{Y}_f$, for the given set of observed responses $\boldsymbol{y}$. Then we can write,

$$\int_R p(\boldsymbol{\lambda}_f \mid \boldsymbol{y}) d\boldsymbol{\lambda}_f = \int_R \int_\Omega p(\boldsymbol{\lambda}_f, \theta \mid \boldsymbol{y}) d\theta d\boldsymbol{\lambda}_f \tag{3.2}$$

where $p(\boldsymbol{\lambda}_f, \theta \mid \boldsymbol{y})$ is the joint density function of $\boldsymbol{\lambda}_f$ and Θ for any given $\boldsymbol{y}$. Since, in general, $\boldsymbol{\lambda}_f$ and Θ may not necessarily be independent, so $\boldsymbol{\lambda}_f$ and Θ are assumed to be not independent, and hence the density can't be factored. However, by applying the rule of conditional probability and assuming that the conditions of Fubini's theorem hold (to be able to change the order of integration), we can write,

$$\begin{aligned}
\int_R p(\boldsymbol{\lambda}_f \mid \boldsymbol{y})\, d\boldsymbol{\lambda}_f &= \int_R \int_\Omega p(\theta \mid \boldsymbol{y})\, p(\boldsymbol{\lambda}_f \mid \theta, \boldsymbol{y})\, d\theta\, d\boldsymbol{\lambda}_f \\
&= \int_\Omega \int_R p(\boldsymbol{\lambda}_f \mid \theta, \boldsymbol{y})\, p(\theta \mid \boldsymbol{y})\, d\boldsymbol{\lambda}_f\, d\theta \\
&= \int_\Omega P\left[\boldsymbol{\lambda}_f \in R(\boldsymbol{Y}) \mid \theta, \boldsymbol{y}\right]\, p(\theta \mid \boldsymbol{y})\, d\theta \\
&= E_\theta\left[C(R) \mid \boldsymbol{y}\right] = \beta
\end{aligned} \tag{3.3}$$

where $p(\theta \mid \boldsymbol{y})$ is the density of the parameter Θ for any given $\boldsymbol{y}$. In the Bayesian approach this density function, $p(\theta \mid \boldsymbol{y})$ becomes the Bayes posterior density and in the structural approach it is the structural density. Fraser and Haq (1969) discussed that for the non-informative prior, the Bayes posterior density is the same as the structural density. Thus one can find a β-expectation tolerance region for any suitable function of a set of future responses by using the prediction distribution of the function of future responses. However, there are many regions on the sample space that are likely to satisfy (2.1), and hence a β-expectation tolerance region is not unique. So the search for an optimal tolerance region becomes obvious.

3.1 An Optimal Tolerance Region

There could be infinitely many tolerance regions on the same sample space having the same expected coverage. Hence we need to search for an optimal tolerance region. A β-expectation tolerance region is said to be optimal if the enclosure or the coverage of the tolerance region is the minimum subject to

$$E_{\theta|\boldsymbol{y}} \{C[R(Y)]\} \geq \beta \tag{3.4}$$

where $\theta \mid \boldsymbol{y}$ denotes the density of Θ for given $\boldsymbol{y}$. But as shown in (3.3), the relation (3.4) can be written as

$$P_{\boldsymbol{\lambda}_f|\boldsymbol{y}} \{ \boldsymbol{\lambda}_f \in R(Y) \} \geq \beta \tag{3.5}$$

where $P_{\boldsymbol{\lambda}_f|\boldsymbol{y}}$ represents the prediction density of a function of the future response $\boldsymbol{\lambda}_f$ for any given set of data, $\boldsymbol{y}$. Different approaches have been proposed to determine an optimal tolerance region in the literature. Here, we would apply the Neyman-Pearson Lemma approach to find a tolerance region that satisfies (2.4) and has a minimum enclosure.

Let us assume that the coverage $C[R(Y)]$ has an induced probability density $h(\boldsymbol{\lambda}_f)$ on the space of the future responses. Then by the Neyman-Pearson Lemma a tolerance region $R(Y)$ would be optimal if it satisfies the following:

$$R(Y) = \left\{ \boldsymbol{\lambda}_f \ : \ \frac{p(\boldsymbol{\lambda}_f \mid \boldsymbol{y})}{h(\boldsymbol{\lambda}_f)} \ > \ k_{(\boldsymbol{y})} \right\} \tag{3.6}$$

where $k_{(\boldsymbol{y})}$ is determined such that

$$P_{\boldsymbol{\lambda}_f|\boldsymbol{y}} \{ \boldsymbol{\lambda}_f \in R(Y) \} = \beta. \tag{3.7}$$

Bishop (1976, p. 99-100) shows that the β-expectation tolerance region obtained by using the prediction distribution is an optimal tolerance region. Therefore, the β-expectation tolerance region defined above would be an optimal tolerance region in the sense of having a minimum enclosure.

4 Some Preliminaries

Some useful notations are introduced in this section to facilitate the presentation 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

$$\boldsymbol{b}(\boldsymbol{e}) = \boldsymbol{e}X'(XX')^{-1} \text{ and } s^2(\boldsymbol{e}) = [\boldsymbol{e} - \boldsymbol{b}(\boldsymbol{e})X][\boldsymbol{e} - \boldsymbol{b}(\boldsymbol{e})X]'. \tag{4.1}$$

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. So

$$\boldsymbol{e} = \boldsymbol{b}(\boldsymbol{e})X + s(\boldsymbol{e})\boldsymbol{d}(\boldsymbol{e}) \text{ and hence } \boldsymbol{e}\boldsymbol{e}' = \boldsymbol{b}(\boldsymbol{e})XX'\boldsymbol{b}'(\boldsymbol{e}) + s^2(\boldsymbol{e}) \tag{4.2}$$

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 (4.2) we observe the following relationship between the error and response statistics (cf. Fraser, 1968, p.127)

$$\boldsymbol{b}(\boldsymbol{e}) = \sigma^{-1}\{\boldsymbol{b}(\boldsymbol{y}) - \boldsymbol{\delta}\}, \quad \text{and } s^2(\boldsymbol{e}) = \sigma^{-2}s^2(\boldsymbol{y}), \tag{4.3}$$

where $\boldsymbol{b}(\boldsymbol{y}) = \boldsymbol{y}X'(XX')^{-1}$ and $s^2(\boldsymbol{y}) = [\boldsymbol{y} - \boldsymbol{b}(\boldsymbol{y})X][\boldsymbol{y} - \boldsymbol{b}(\boldsymbol{y})X]'$ 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. Now define the following statistics based on the future regression model:

$$\boldsymbol{b}_f(\boldsymbol{e}_f) = \boldsymbol{e}_f X_f'(X_f X_f')^{-1}, \; s_f^2(\boldsymbol{e}_f) = [\boldsymbol{e}_f - \boldsymbol{b}_f(\boldsymbol{e}_f)X_f][\boldsymbol{e}_f - \boldsymbol{b}_f(\boldsymbol{e}_f)X_f]' \tag{4.4}$$

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

$$\boldsymbol{e}_f = \boldsymbol{b}_f(\boldsymbol{e}_f)X_f + s_f(\boldsymbol{e}_f)\boldsymbol{d}_f(\boldsymbol{e}_f) \text{ and hence } \boldsymbol{e}_f\boldsymbol{e}_f' = \boldsymbol{b}_f(\boldsymbol{e}_f)X_f X_f'\boldsymbol{b}_f'(\boldsymbol{e}_f) + s_f^2(\boldsymbol{e}_f) \tag{4.5}$$

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:

$$\boldsymbol{b}_f(\boldsymbol{e}_f) = \sigma^{-1}\{\boldsymbol{b}_f(\boldsymbol{y}_f) - \boldsymbol{\delta}\}, \quad \text{and } s_f^2(\boldsymbol{e}_f) = \sigma^{-2}s_f^2(\boldsymbol{y}_f), \tag{4.6}$$

where $\boldsymbol{b}_f(\boldsymbol{y}_f) = \boldsymbol{y}_f X_f'(X_f X_f')^{-1}$ and $s_f^2(\boldsymbol{y}) = [\boldsymbol{y}_f - \boldsymbol{b}_f(\boldsymbol{y}_f)X_f][\boldsymbol{y}_f - \boldsymbol{b}_f(\boldsymbol{y}_f)X_f]'$ 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.

5 Predictive Distribution of FRV and FRSS

The joint density function of the error statistics $\boldsymbol{b}(\boldsymbol{e})$, $s^2(\boldsymbol{e})$, $\boldsymbol{b}_f(\boldsymbol{e}_f)$ and $s_f^2(\boldsymbol{e}_f)$, given $\boldsymbol{d}(\cdot)$, is derived from the above joint density of the combined error vector by applying the properties of invariant differentials (see Eaton, 1983, p.194-206) as follows:

$$p\Big(\boldsymbol{b}(\boldsymbol{e}), s^2(\boldsymbol{e}), \boldsymbol{b}_f(\boldsymbol{e}_f), s_f^2(\boldsymbol{e}_f)|\boldsymbol{d}(\cdot)\Big) \propto [s^2(\boldsymbol{e})]^{\frac{n-p-2}{2}}[s_f^2(\boldsymbol{e}_f)]^{\frac{n_f-p-2}{2}} \times g\left\{\boldsymbol{b}(\boldsymbol{e})XX'\boldsymbol{b}'(\boldsymbol{e}) + \boldsymbol{b}_f(\boldsymbol{e}_f)X_fX_f'\boldsymbol{b}_f'(\boldsymbol{e}_f)\right\}. \quad (5.1)$$

Note that the above density does not depend on $\boldsymbol{d}(\cdot)$ (cf. Fraser, 1968, p.132) so the conditional distribution is the same as the unconditional distribution. Using the Jacobian of the transformation,

$$J\Big\{[\boldsymbol{b}_f(\boldsymbol{e}_f), s_f^2(\boldsymbol{e}_f)] \rightarrow [\boldsymbol{b}_f(\boldsymbol{y}_f), s_f^2(\boldsymbol{y}_f)]\Big\} = [\sigma^2]^{-\frac{p+2}{2}}, \quad (5.2)$$

and the non-informative prior distribution,

$$p\left(\boldsymbol{\delta}, \frac{1}{\sigma^2}\right) \propto \left\{\frac{1}{\sigma^2}\right\}^{-1}, \quad (5.3)$$

for the parameters of the model, the joint posterior density of $\boldsymbol{\delta}$, σ, $\boldsymbol{b}_f(\boldsymbol{y}_f)$ and $s_f^2(\boldsymbol{y}_f)$ is obtained as

$$p\Big(\boldsymbol{\delta}, \sigma^2, \boldsymbol{b}_f, s_f^2\Big) \propto [s^2]^{\frac{n-p-2}{2}}[s_f^2(\boldsymbol{y}_f)]^{\frac{n_f-p-2}{2}}[\sigma^2]^{-\frac{n+n_f-p}{2}} g\left\{\frac{1}{\sigma^2}\Big[\Big(\boldsymbol{b}-\boldsymbol{\delta}\Big)XX'\Big(\boldsymbol{b}-\boldsymbol{\delta}\Big)' + s^2 + \Big(\boldsymbol{b}_f-\boldsymbol{\delta}\Big)X_fX_f'\Big(\boldsymbol{b}_f-\boldsymbol{\delta}\Big)' + s_f^2\Big]\right\} \quad (5.4)$$

where $\boldsymbol{b}_f = \boldsymbol{b}_f(\boldsymbol{y}_f)$ and $s_f^2 = s_f^2(\boldsymbol{y}_f)$ for notational convenience. Such results can also be obtained by using the structural approach. However, the final results of this paper will be the same as that obtained by the structural approach. Interested readers may refer to Fraser and Haq (1969) for details.

To find the predictive distribution of the FRV and FRSS we need to integrate out $\boldsymbol{\delta}$ and σ^2 from the above joint density. Following Ng (2000), to integrate out σ^2, let $\frac{1}{\sigma^2} = \lambda$. So $d\sigma^2 = \lambda^{-2}d\lambda$. Thus the join density function of $\boldsymbol{\delta}$, $\boldsymbol{b}_f(\boldsymbol{y}_f)$ and $s_f^2(\boldsymbol{y}_f)$ can be written as

$$p\Big(\boldsymbol{\delta}, \boldsymbol{b}_f, s_f^2\Big) \propto \left[s_f^2(\boldsymbol{y}_f)\right]^{\frac{n_f-p-2}{2}} \int_{\lambda>0} [\lambda]^{-\frac{n+n_f-2}{2}} g\left\{\lambda\Big[Q_y + s^2 + Q_{y_f} + s_f^2\Big]\right\} d\lambda \quad (5.5)$$

where $Q_y = \Big(\boldsymbol{b}-\boldsymbol{\delta}\Big)XX'\Big(\boldsymbol{b}-\boldsymbol{\delta}\Big)'$ and $Q_{y_f} = \Big(\boldsymbol{b}_f-\boldsymbol{\delta}\Big)X_fX_f'\Big(\boldsymbol{b}_f-\boldsymbol{\delta}\Big)'$. Now let $\psi^2 = Q + s^2 + s_f^2$ in which $Q = Q_y + Q_{y_f}$. Then set $w = \lambda\psi^2$, and hence $d\lambda = \left[\psi^2\right]^{-1}dw$. Completion of the integration leads to the join density function of $\boldsymbol{\delta}$, $\boldsymbol{b}_f(\boldsymbol{y}_f)$ and $s_f^2(\boldsymbol{y}_f)$ to be

$$p\Big(\boldsymbol{\delta}, \boldsymbol{b}_f, s_f^2\Big) \propto \left[s_f^2(\boldsymbol{y}_f)\right]^{\frac{n_f-p-2}{2}}\Big[Q + s^2 + s_f^2\Big]^{-\frac{n+n_f}{2}}. \quad (5.6)$$

Expressing the terms involving $\boldsymbol{\delta}$ in Q as

$$\begin{aligned} Q &= \Big(\boldsymbol{b}-\boldsymbol{\delta}\Big)XX'\Big(\boldsymbol{b}-\boldsymbol{\delta}\Big)' + \Big(\boldsymbol{b}_f-\boldsymbol{\delta}\Big)X_fX_f'\Big(\boldsymbol{b}_f-\boldsymbol{\delta}\Big)' \\ &= \Big(\boldsymbol{\delta}-FA^{-1}\Big)A\Big(\boldsymbol{\delta}-FA^{-1}\Big)' + \Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)H^{-1}\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)' \end{aligned} \tag{5.7}$$

where

$$F = \boldsymbol{b}XX' + \boldsymbol{b}_fX_fX_f', \; A = XX' + X_fX_f', \text{ and } H = [XX']^{-1} + [X_fX_f']^{-1} \tag{5.8}$$

the joint density of $\boldsymbol{b}_f$ and s_f^2 becomes

$$p\Big(\boldsymbol{b}_f, s_f^2\Big) = \Psi_{12} \times \big[s_f^2\big]^{\frac{n_f-p-2}{2}} \Big[s^2 + s_f^2 + \Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)H^{-1}\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)\Big]^{-\frac{n+n_f-p}{2}} \tag{5.9}$$

where $\Psi_{12} = \{|H|^{-\frac{1}{2}}\Gamma(\frac{n+n_f-p}{2})[s^2]^{\frac{n-p}{2}}\}\{(\pi)^{\frac{p}{2}}\Gamma(\frac{n-p}{2})\Gamma(\frac{n_f-p}{2})\}^{-1}$ is the normalizing constant. Integrating out s_f^2 from the above joint density, the prediction distribution of the future regression vector, $\boldsymbol{b}_f = \boldsymbol{b}_f(\boldsymbol{y}_f)$, becomes

$$p\Big(\boldsymbol{b}_f\Big|\boldsymbol{y}\Big) = \Psi_1 \times \Big[s^2 + \Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)H^{-1}\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)'\Big]^{-\frac{n}{2}} \tag{5.10}$$

where $\Psi_1 = \{\Gamma(\frac{n}{2})[s^2]^{\frac{n-p}{2}}\}\{(\pi)^{\frac{p}{2}}\Gamma(\frac{n-p}{2})|H|^{\frac{1}{2}}\}^{-1}$. The prediction distribution of $\boldsymbol{b}_f$ can be written in the usual multivariate Student-t distribution form as follows:

$$p\Big(\boldsymbol{b}_f\Big|\boldsymbol{y}\Big) = \Psi_6 \times \Big[1 + \Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)\Big[s^2H\Big]^{-1}\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)'\Big]^{-\frac{n}{2}} \tag{5.11}$$

in which $n > p$. Thus, $[\boldsymbol{b}_f|\boldsymbol{y}] \sim t_p(n-p,\; \boldsymbol{b},\; \frac{n-p}{n-p-2}Hs^2)$ where $\boldsymbol{b}$ is the sample regression vector of realized responses and H is the scale matrix. Khan (2005) obtained the same result for the multiple regression model with normal errors. Khan (2004) noted that the prediction distribution of the FRV of the future regression model does not depend on the shape parameter, ν of the multiple regression model with multivariate Student-t errors.

5.1 Distribution of Future Residual Sum of Squares

The prediction distribution of the FRSS for the future regression model, $s_f^2(\boldsymbol{y}_f)$, conditional on the realized responses, $\boldsymbol{y}$, is obtained as

$$p\Big(s_f^2(\boldsymbol{y}_f)\Big|\boldsymbol{y}\Big) = \Psi_2 \times [s_f^2(\boldsymbol{y}_f)]^{\frac{n_f-p-2}{2}} \big[s^2 + s_f^2(\boldsymbol{y}_f)\big]^{-\frac{n+n_f-2p}{2}}. \tag{5.12}$$

The above density function can be written in the usual form of beta distribution of the second kind as follows

$$p\Big(s_f^2\Big|\boldsymbol{y}\Big) = \Psi_7 \times [s_f^2]^{\frac{n_f-p-2}{2}} \big[1 + s^{-2}s_f^2\big]^{-\frac{n+n_f-2p}{2}} \tag{5.13}$$

where $\Psi_2 = \{\Gamma(\frac{n+n_f-2p}{2})[s^2]^{-\frac{n-p}{2}}\}\{\Gamma(\frac{n-p}{2})\Gamma(\frac{n_f-p}{2})\}^{-1}$ is the normalizing constant. This is the prediction distribution of the FRSS based on the future response $\boldsymbol{y}_f$, conditional on the realized responses $\boldsymbol{y}$, from the multiple regression model with a family of multivariate spherically contoured errors. The density in (5.13) is a modified form of beta density. However, it can be shown that s_f^2/s^2 is a beta variable with arguments $(n_f-p)/2$ and $(n-p)/2$. Obviously, for the existence of the above distribution of s_f^2 we must have $n_f > p$ in addition to $n > p$. Khan (2004) obtained the same prediction distribution of the FRSS, conditional on the realized responses, for the multiple regression model with multivariate normal errors.

6 Optimal β-Expectation Tolerance Region

Since the tolerance regions based on prediction distributions are optimal in the sense of having minimum closure, we use the prediction distributions of the FRV and FRSS to find optimal β-expectation tolerance regions for them. In order to obtain the tolerance regions, we need to determine the sampling distribution of some appropriate functions involved in the prediction distribution of the statistics of the future responses.

From the definition of the β-expectation tolerance region for any future statistic, $R^*(\boldsymbol{y}) = \{\boldsymbol{\tau} : \boldsymbol{\tau} < \boldsymbol{\tau}^*\}$ is a β-expectation tolerance region for $\boldsymbol{\tau} > 0$ if $\boldsymbol{\tau}^*$ is the β^{th} quantile of the sampling distribution of the future statistic $\boldsymbol{\tau}$. That is, $R^*(\boldsymbol{y})$ is a β-expectation tolerance region for the future statistic $\boldsymbol{\tau}$ if $\boldsymbol{\tau}^*$ is such that

$$\int_{\boldsymbol{\tau}=0}^{\boldsymbol{\tau}^*} f(\boldsymbol{\tau}) d\boldsymbol{\tau} = \beta \qquad (6.1)$$

where $f(\boldsymbol{\tau})$ is the pdf of the future statistic $\boldsymbol{\tau}$.

6.1 Tolerance Region for the FRV

For the future statistic $\boldsymbol{b}_f(\boldsymbol{y})$ the prediction distribution is a multivariate Student-t distribution. To find an optimal β-expectation tolerance region for the FRV, we use the prediction distribution of the FRV to determine the prediction distribution of an appropriate quadratic form of the FRV. The following result is useful to derive the β-expectation tolerance region for the FRV.

Theorem 5.1: *If a p dimensional random vector $\boldsymbol{\eta}$ follows a multivariate Student-t distribution with location vector $\boldsymbol{\zeta}$, scale matrix $\boldsymbol{\Omega}$, and shape parameter ν then the scaled quadratic form $\frac{1}{\nu}(\boldsymbol{\eta}-\boldsymbol{\zeta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\eta}-\boldsymbol{\zeta})'$ follows an F distribution with p and ν degrees of freedom.*
The proof is straightforward.

Since the prediction distribution of the FRV $\boldsymbol{b}_f$ is a p-variate Student-t distribution we use the above theorem to assert that the distribution of the quadratic form

$$\frac{1}{(n-p)}\Big(\boldsymbol{b}_f - \boldsymbol{b}\Big)\Big[s^2 H\Big]^{-1}\Big(\boldsymbol{b}_f - \boldsymbol{b}\Big)' \qquad (6.2)$$

is an F distribution with p and $n-p$ degrees of freedom. Then an optimal β-expectation tolerance region that will enclose 100β percent of the future regression vectors from the multiple regression model with multivariate spherically symmetric errors is given by the ellipsoidal region:

$$R_1(\boldsymbol{b}_f|\boldsymbol{y}) = \left\{\boldsymbol{y}_f : \left[\frac{1}{n-p}\Big(\boldsymbol{b}_f - \boldsymbol{b}\Big)\Big[s^2 H\Big]^{-1}\Big(\boldsymbol{b}_f - \boldsymbol{b}\Big)'\right] \leq F_{p,n-p,\beta}\right\} \quad (6.3)$$

where $F_{p,\ n-p,\beta}$ is the $\beta \times 100$ percentile point of a central F distribution with p and $n-p$ degrees of freedom such that $P(F_{p,\ n-p} < F_{p,\ n-p,\beta}) = \beta$. As noted by Bishop (1976) the region given by $R_1(\boldsymbol{b}_f|\boldsymbol{y})$ in the foregoing expression is an optimal β-expectation tolerance region, and among all such tolerance regions it has the minimum enclosure. Note that $R_1(\boldsymbol{b}_f|\boldsymbol{y})$ depends on the sample responses through H, a function of observed and future regressors, $\boldsymbol{b} = \boldsymbol{b}(\boldsymbol{y})$ and $s = s(\boldsymbol{y})$. Moreover, it depends on the size of the observed sample as well as the dimension of the regression vector.

6.2 Tolerance Region for the FRSS

The β-expectation tolerance region for the FRSS can be based on its prediction distribution. From the previous section, the prediction distribution of the FRSS is known to be a beta distribution. So an optimal β-expectation tolerance region for the FRSS can be defined using an appropriate beta distribution. A region on the sample space of the responses is a β-expectation tolerance region if it encloses 100β percent of the future residual sum of squares from the multiple regression model and is given by the ellipsoidal region

$$R_2(s_f|\boldsymbol{y}) = \left\{\boldsymbol{y}_f : \left[s_f^2(s_{\boldsymbol{y}}^2)^{-1}\right] \leq B_\beta\left(\frac{n_f-p}{2}, \frac{n-p}{2}\right)\right\} \quad (6.4)$$

where $B_\beta(\frac{n_f-p}{2}, \frac{n-p}{2})$ is the $\beta \times 100$ percentile point of a beta distribution with arguments $\left(\frac{n_f-p}{2}\right)$ and $\left(\frac{n-p}{2}\right)$ such that

$$P\left[B\left(\frac{n_f-p}{2},\ \frac{n-p}{2}\right) < B_\beta\left(\frac{n_f-p}{2},\ \frac{n-p}{2}\right)\right] = \beta.$$

Using the following relationship between the inverse beta distribution and F distribution, the above β-expectation tolerance region for the FRSS can be based on an F distribution with (n_f-p) and $(n-p)$ degrees of freedom.

Theorem 5.2: *If ψ follows a beta distribution with arguments $\frac{\lambda}{2}$ and $\frac{\tau}{2}$ then $\varphi = \frac{\tau}{\lambda}[\psi]^{-1}$ follows an F distribution with λ and τ degrees of freedom.*

In view of the above fact, since $s_f^2(s_{\boldsymbol{y}}^2)^{-1}$ follows a beta distribution with arguments $\frac{n-p}{2}$ and $\frac{n_f-p}{2}$, the statistic $\left[s_{\boldsymbol{y}}^2\{s_f^2\}^{-1}\right]$ is distributed as a scaled F variable with (n_f-p) and $(n-p)$ degrees of freedom. That is, $\left[s_{\boldsymbol{y}}^2\{s_f^2\}^{-1}\right] \sim \frac{n-p}{n_f-p}F_{n_f-p,n-p}$. Therefore an equivalent β-expectation tolerance region for

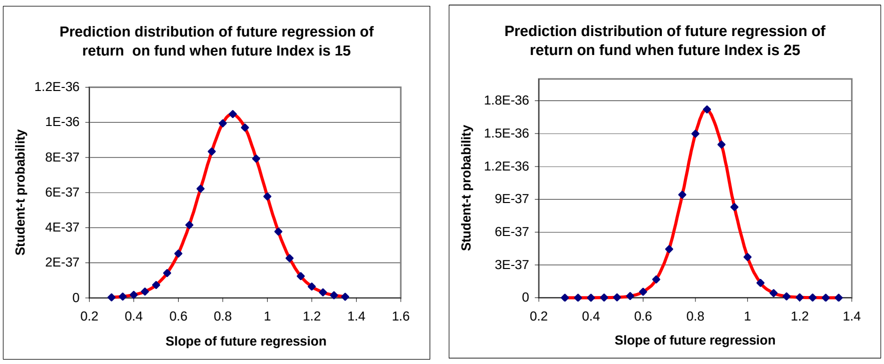

Fig. 1 Prediction distribution of the future regression parameter for selected values of future Index

the future residual sum of squares from the multiple regression model with multivariate spherically symmetric errors is given by the ellipsoidal region:

$$R_3(s_f|\boldsymbol{y}) = \left\{ \boldsymbol{y}_f : \left[\frac{s^2_{\boldsymbol{y}}}{s^2_f} \right] \leq \frac{n-p}{n_f - p} F_{n_f-p,n-p,\beta} \right\} \tag{6.5}$$

where $F_{n_f-p,\ n-p,\beta}$ is the $\beta \times 100$ percentile point of a central F distribution with $n_f - p$ and $n - p$ degrees of freedom such that

$$P\left(F_{n_f-p,\ n-p} < F_{n_f-p,\ n-p,\beta}\right) = \beta.$$

It is interesting to note that optimal β-expectation tolerance regions for both the FRV and FRSS can be based on the F distribution, of course, with appropriate degrees of freedom parameters.

7 An Example

As an example, we consider a real life data set from Moore (2003, p.98). The data contains information on the Percentage Return of fund (response variable), and Index of Overseas Stock Market in Europe, Australia and Far East (EAFE) from 1982 (the first full year of the fund's existence) to 2001. For the given data the fitted ordinary least squares model becomes $\hat{y} = 3.3670 + 0.8449x$ with adjusted value of $R^2 = 81.79\%$, $\bar{x} = 13.274$, $\bar{y} = 14.582$, $\sum_{j=1}^{20} x_j^2 = 14029.951$, and $s^2 = 86.859$, the mean squared error. The prediction distribution of the regression parameter involves H which in this special case becomes $[\sum_{j=1}^{20} x_j^2]^{-1} + [{x_f}^2]^{-1}$.

The prediction distribution of the future regression (slope) parameter of the regression of future Percentage Return on the future Index is given in the two graphs of Figure 1. The two graphs represent two Student-t distributions with different parameters. Although the shape of the distribution of both the graphs is roughly the same, the first graph has a slightly more spread, but lower pick, than the second graph.

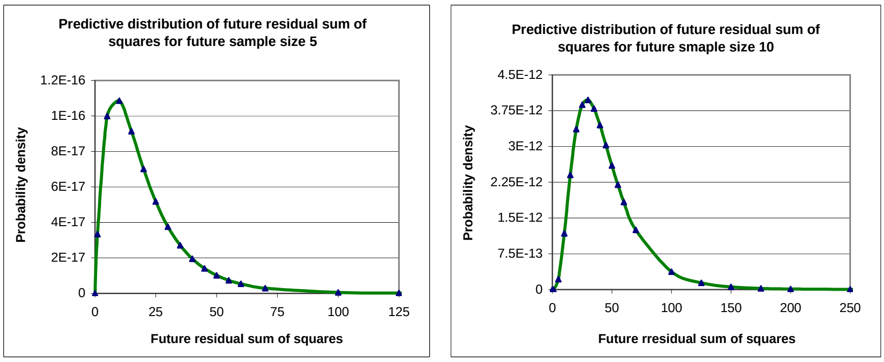

Fig. 2 Prediction distribution of the future residual sum of squares for selected future sample sizes

Figure 2 display the prediction distributions of the future residual sum of squares for different sample sizes. These two graphs represent two beta distributions with varying arguments. Therefore, tolerance regions for the FRV can be based on the quadratic form of the Student-t variable, that is, F-distribution. Similarly, tolerance regions for the FRSS would be based on the beta distribution.

8 Concluding Remarks

The multiple regression model with a family of spherically symmetric errors have been considered to derive the β-expectation tolerance region for the FRV and FRSS. Khan (2004) showed that the above statistics are not independently distributed for the multiple regression model with multivariate normal as well as multivariate Student-t errors. The prediction distribution of the FRV for any member of the family of spherically contoured multiple regression model is a Student-t distribution including the most commonly used members of the family of models, namely the multivariate normal and Student-t models. Similarly, the prediction distribution of the FRSS is a scaled beta distribution for all members of spherically contoured family of models. It is quite interesting to note that although for different member distribution of the spherically contoured family of distributions the regression models are different, but the prediction distributions of the FRV and FRSS are the same regardless of the choice of any particular member. Thus the same predictive inference, including β-expectation tolerance regions, applies for all members of the family of spherically contoured regression models.

The β-expectation tolerance region for the FRV based on the distribution of an appropriate quadratic form of the FRV. From the prediction distribution of the FRV it follows that the required quadratic form of the FRV follows an F distribution. Similarly, the tolerance region for the FRSS is based on the appropriate beta distribution or equivalently an appropriate F distribution. Since the β-expectation tolerance regions of this paper are based on the prediction distributions, they are optimal in the sense of having minimum enclosure among all such tolerance regions. The optimal β-expectation tolerance regions defined in the paper provide the criterion for the necessary and sufficient conditions that any set of future responses satisfying the rules in $R_1(\cdot)$ and $R_3(\cdot)$, given the observed responses, will produce FRV and FRSS such that $\beta \times 100\%$ of the time such tolerance regions will contain the true future regression vector and true future residual sum of squares respectively. The results in this paper are also applicable to multiple regression models with normal and Student-t errors as these two widely used distributions are also popular members of the family of elliptically contoured distributions. The same prediction distributions and hence the same β-expectation tolerance regions for the FRV and FRSS can be obtained by using the structural distribution or structural relation approach.

Acknowledgements The author is grateful to the Coordinating Editor and two unknown referees for some valuable suggestions that have improved the quality of the paper.

References

1. Aitchison, J. and Dunsmore, I.R. (1975). Statistical Prediction Analysis. Cambridge University Press, Cambridge.
2. Aitchison, J. and Sculthorpe, D. (1965). Some problems of statistical prediction. *Biometrika*, *55*, 469-483.
3. Anderson, T.W. (1993). Nonnormal multivariate distributions: Inference based on elliptically contoured distributions. In Multivariate Analysis: Future Directions. ed by C.R. Rao. North-Holland, Amsterdam, 1–24.
4. Bernardo, J and Rueda, R (2002). Bayesian hypothesis testing: A reference approach. em International Statistical Review. Vol. 70, 351-372.
5. Chew, V. (1966). Confidence, prediction and tolerance regions for multivariate normal distribution. *Journal of the American Statistical Association*, **62**, 605-617.
6. Chmielewski, M.A. (1981). Elliptically symmetric distributions. *International Statistical Review*, Vol. 49, 67–74.
7. Eaton, M.L. (1983). Multivariate Statistics - Vector Space Approach. Wiley, New York.
8. Fang, K.T. and Anderson, T.W. (1990). Statistical Inference in Elliptically Contoured and Related Distributions. Allerton Press Inc., New York.
9. Fang, K.T., Kotz, S. and Ng, K.W. (1990). Symmetric Multivariate and Related Distributions. Chapman and Hall, New York
10. Fang, K.T. and Zhang, Y. (1990). Generalized Multivariate Analysis, Science Press and Springer-Verlag, New York
11. Fisher, R.A. (1956). Statistical Methods in Scientific Inference. Oli. & Boyd, London.
12. Fraser, D.A.S. (1968). The Structure of Inference, Wiley, New York.
13. Fraser, D.A.S. (1979). Inference and Linear Models. McGraw-Hill, New York
14. Fraser, D.A.S., and Guttman, I. (1956). Tolerance regions. *Annals of Mathematical Statistics*, **27**, 162-179.
15. 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.
16. Fuchs, C. and Kenett, S. (1987). Multivariate tolerance regions and F-tests. *Journal of Quality Technology*, **19**, 122-131.
17. Geisser, S. (1993). Predictive Inference: An Introduction. Chapman & Hall, London.
18. Guttman, I. (1970a). Statistical Tolerance Regions. Griffin, London.

19. Guttman, I. (1970b). Construction of β-content tolerance regions at confidence level γ for large samples for k-variate normal distribution. *Annals of Mathematical Statistics*, **41**, 376-400.
20. 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.
21. 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.
22. Khan, S. (1996). Prediction inference for heteroscedastic multiple regression model with class of spherical error. *Aligarh Journal of Statistics,* **15 & 16**, 1-17.
23. Khan, S. (2004). Predictive distribution of regression vector and residual sum of squares for normal multiple regression model. *Communications In Statistics: Theory and Methods*, **33(10)**, 2423-2443.
24. Khan, S. (2005). Optimal Tolerance Regions for Some Functions of Multiple Regression Model with Normal Error, *Working Paper Series*, SC-MC-0504, Faculty of Sciences, University of Southern Queensland, Australia.
25. Khan, S. (2006). Optimal Tolerance Regions for Some Functions of Multiple Regression Model with Student-t Errors. *Journal of Statistics & Management Systems*, **9(3)**, 699-715
26. 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.
27. Moore, D. (2003). The Basic Practice of Statistics, 3^{rd} ed., Freeman, New York.
28. Ng, V.M. (2000). A note on predictive inference for multivariate elliptically contoured distributions. *Comm. In Statist. - Theory & Methods,* **29** (3), 477-483.
29. Prucha I.R. and Kelejian, H.H. (1984). The structure of simultaneous equation estimators: A generalization towards non-normal disturbances, *Econometrica*, Vol. 52, 721–736.
30. Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error term. *Jou. Amer. Statist. Assoc.*, Vol. 60, 601-616.