Optimal Tolerance Regions for Some Functions of Multiple Regression Model with Student-t Errors

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

Abstract

This paper considers the multiple regression model 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. It is assumed that the errors of the regression model follow a multivariate Student-t distribution with unknown shape parameter, ν. The prediction distribution of the FRV, conditional on the observed responses, is a multivariate Student-t distribution but its shape parameter does not depend on the unknown degrees of freedom of the Student-t model. 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.

AMS 2000 Subject Classification: Primary 62A25, Secondary 62J05

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

1 Introduction

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

*The author acknowledges excellent research facilities in the Department of Management & Marketing, College of Business Administration, University of Bahrain, Kingdom of Bahrain

level of probability. There are several kinds of tolerance regions available in the literature (cf. Guttman, 1970b, and Aitchison and Dunsmore, 1975). 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. A statistical tolerance region is a region defined on the sample space that contains a specified proportion of the realizations of the values of a random variable, or a suitable function of it, under study with a pre-assigned level of probability. It is a problem under the broader area of the predictive inference and can be solved by using the prediction distribution.

There has been a growing interest in statistics for the use of non-normal models to represent symmetrical, but fat or heavier tailed distributions of the errors. In this paper we consider the widely used multiple regression model with errors, for both the realized and future responses, following the multivariate Student-t distribution with unknown shape parameter, ν. The two sets of responses are connected through the common, shape, regression and scale parameters. Following Khan (2004), we pursue the Bayesian approach to derive the distribution of the FRV and FRSS for the future responses, conditional on a set of realized responses. This is a new development that deals with the predictive inference for the future regression parameters, rather than that of the future responses. The prediction distributions of the FRV and FRSS, conditional on a given set of data, have been provided first, by using the invariant differentials and non-informative prior distribution. Then β-expectation tolerance regions have been obtained for the FRV and FRSS of the multiple regression model by using these prediction distributions. It has been proved (see Bishop, 1976, p.99-100) that the β-expectation tolerance regions based on such prediction distributions are optimal in the sense of having minimum enclosures.

Many researchers conducted studies in the area of tolerance regions. The first work in this area is due to Wilks (1941). Others include Scheffe and Tukey (1944), Paulson (1943), Wald and Wolfowitz (1946), Fraser (1953), Fraser and Guttman (1956) and Guttman (1959, 1970a). A detail theory of tolerance region has been presented by Guttman (1970b). The Bayesian works include Aitchison (1964), Aitchison and Dunsmore (1975), and Geisser (1993), while Fraser and Haq (1969), Rinco (1973), Haq and Khan (1990) pursued the structural distribution approach. Aitchison and Dunsmore (1975) provide an excellent account of the theory and application of the prediction problem including various tolerance regions. Geisser (1993) discussed the Bayesain approach to predictive inference including the tolerance region and pointed out a number of real-life applications in various fields. This includes model selection, discordancy, perturbation analysis, classification, regulation, screening and interim analysis. The tolerance regions for linear models have been dealt with by Wallis (1951), Lieberman and Miller (1963), and Bishop (1976). Haq and Rinco (1976) derived the β-expectation tolerance region for generalized linear model with multivariate normal error distribution using the structural distribution approach. All the above studies deal with the prediction distribution of future responses. However, Khan (2004) proposed prediction distributions for the future regression vector (FRV) and future residual sum of squares (FRSS). Here we pursue the same approach to find the optimal β-expectation tolerance regions for the FRV and FRSS using the distribution of appropriate future statistics.

The multiple regression model with the multivariate Student-t errors is specified in the next section. The general formulation of the β-expectation tolerance region and its optimality criterion are introduced in Section 3. Section 4 presents some preliminaries and useful relationships. The prediction distributions of the FRV and FRSS are given in section 5. The β-expectation tolerance regions for the FRV and FRSS are obtained in section 6. Some concluding remarks are included in section 7.

2 Multiple Regression with Student-t Errors

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 reserchers to the use of the Student-t model in such situations. In addition to being robust, the Student-t distribution is a 'more typical' member of the elliptical/spherical class of distributions than the normal distribution. Moreover, the normal distribution is a special (limiting) case of the Student-t distribution. It also covers the Cauchy and sub-Cauchy distributions (cf. Fraser, 1979, p.27) on the other extreme. 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 Ullah and Walsh (1981). Zellner (1976) first considered the linear regression model with Student-t errors.

In this paper we consider the multiple regression equation

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

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 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 Student-t variable with unknown shape parameter ν, location 0 and scale 1, so that the variance of y is $\frac{\nu}{\nu-2}\sigma^2$. 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

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

where the n-dimensional row vector $\boldsymbol{y}$ is a vector of the response variable; X is a $p \times n$ dimensional matrix of 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 vector $\boldsymbol{\beta}$ and scale parameter σ are the same as defined in (2.1). Then the error vector follows a multivariate Student-t (cf. Khan, 2000) distribution with location $\mathbf{0}$, a vector of n-tuple of zeros, and variance-covariance matrix, $\frac{\nu}{\nu-2}I_n$. Therefore, the joint density function of the errors becomes

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

Consequently, the response vector follows a multivariate Student-t distribution with mean vector $\boldsymbol{\beta}X$, variance-covariance matrix, $\frac{\nu}{\nu-2}\sigma^2 I_n$, and density function

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

In this paper, the above multiple regression model represents the realized model of the responses from the *performed experiment*.

2.1 Multiple Regression Model for Future Responses

In this subsection we introduce the idea of predictive model for the responses from the *future experiment*. First, consider a set of $n_f > p$ 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. Such a set of future responses can be expressed as

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

where X_f is the $p \times n_f$ matrix of the values of the regressors that generates 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 shape, 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. Unlike the normal model the two sets of errors for the multivariate Student-t model are dependent and hence the joint density of the combined error vector, $(\boldsymbol{e},\ \boldsymbol{e}_f)$, can't be written as the product of the marginal densities of the realized and future errors. However, the join density function of the combined error vector, that is, the errors associated with the realized and the future responses, $(\boldsymbol{e},\ \boldsymbol{e}_f)$, can be expressed as

$$f(\boldsymbol{e},\ \boldsymbol{e}_f) = \frac{\Gamma(\frac{\nu+n+n_f}{2})}{[\pi\nu]^{\frac{n+n_f}{2}}\Gamma(\frac{\nu}{2})}\left[1+\frac{1}{\nu}\left\{\boldsymbol{e}\boldsymbol{e}'+\boldsymbol{e}_f\boldsymbol{e}_f'\right\}\right]^{-\frac{\nu+n+n_f}{2}}. \tag{2.6}$$

It may be noted here that for the multiple regression model with the normal errors, the realized error vector and future error vector are independent, and hence the joint density function of the two vectors can be written as the product of their marginal density functions. But for the Student-t model the two vectors of errors are uncorrelated but dependent, and as such can't be written as the product of their marginal density functions. However, the above joint density function is used to derive the prediction distributions of the functions of the future responses that subsequently provide the basis for the derivation of optimal β-expectation tolerance regions.

3 Formulation of 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 $\theta = [\boldsymbol{\beta}X, \sigma]$ (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,

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

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 (3.1), 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

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

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

$$P_{\boldsymbol{\lambda}_f \mid \boldsymbol{y}} \left\{ \, \boldsymbol{\lambda}_f \in R(Y) \, \right\} \geq \beta \tag{3.5}$$

where $P_{\boldsymbol{\lambda}_f \mid \boldsymbol{y}}$ 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 (3.4) 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 (3.2) we observe the following relationship between the error and response statistics (cf. Fraser, 1968, p.127)

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

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

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

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:

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

where $\boldsymbol{b}_f(\boldsymbol{y}_f) = \boldsymbol{y}_fX_f'(X_fX_f')^{-1}$ 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.

Note since the error vectors for the realized and future responses are not independent, the joint density of the combined error vector can't be written as the product of the two marginal densities for the two sets of errors. Haq and Khan (1990) used this density function to derive the prediction distribution of future responses, conditional on the realized responses, by structural distribution approach. Here we require the prediction distributions of the FRV and FRSS to find the β-expectation tolerance regions for them by using the Bayesian approach with non-informative prior distribution.

5 Predictive Distributions of FRV and FRSS

In this section we provide the predictive distributions of the future regression vector (FRV) and future residual sum of squares (FRSS) for the future multiple regression model, conditional on the realized responses. This can be done either by the Bayesian approach or structural approach. For the Bayesian approach, in the absence of any knowledge about the parameters, an appropriate non-informative prior distribution for the parameters is pursued. 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 many 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, and hence they both lead to the same prediction distribution. We use the prediction distributions of the FRV and FRSS to derive optimal β-expectation tolerance regions for the two functions of the future responses.

5.1 Distribution of the Future Regression Vector

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 of the combined error vector by applying the properties of invariant differentials (see Eaton, 1983, p.194-206) as

$$p\Big(\boldsymbol{b}(\boldsymbol{e}), s^2(\boldsymbol{e}), \boldsymbol{b}_f(\boldsymbol{e}_f), s_f^2(\boldsymbol{e}_f)|\boldsymbol{d}(\cdot)\Big) = \Psi_1(\cdot) \times [s^2(\boldsymbol{e})]^{\frac{n-p-2}{2}}[s_f^2(\boldsymbol{e}_f)]^{\frac{n_f-p-2}{2}}$$
$$\times \left[1 + \frac{1}{\nu} + \boldsymbol{b}(\boldsymbol{e})XX'\boldsymbol{b}'(\boldsymbol{e}) + g_2\Big(\boldsymbol{b}_f(\boldsymbol{e}_f)X_fX_f'\boldsymbol{b}_f'(\boldsymbol{e}_f)\Big)\right]^{-\frac{\nu+n+n_f}{2}} \quad (5.1)$$

where $\Psi_1(\cdot) = \{\Gamma(\frac{\nu+n+n_f}{2})|XX'|^{\frac{1}{2}}|X_fX_f'|^{\frac{1}{2}}\}\{[\pi]^p[\nu]^{\frac{n+n_f}{2}}\Gamma(\frac{n-p}{2})\Gamma(\frac{n_f-p}{2})\Gamma(\frac{\nu}{2})\}^{-1}$ is the normalizing constant. Since the above density does not depend on $\boldsymbol{d}(\cdot)$ the conditioning can be ignored. Now, the joint density of $\boldsymbol{\beta}$ and σ as well as the future response statistics $\boldsymbol{b}_f(\boldsymbol{y}_f)$ and $s_f^2(\boldsymbol{y}_f)$ is obtained as

$$\begin{aligned} p\Big(\boldsymbol{\beta}, \sigma^2, \boldsymbol{b}_f(\boldsymbol{e}_f), s_f^2(\boldsymbol{e}_f)\Big) &= \Psi_2(\cdot) \times [s^2]^{\frac{n-p}{2}}[s_f^2(\boldsymbol{y}_f)]^{\frac{n_f-p-2}{2}}[\sigma^2]^{-\frac{n+n_f-p}{2}} \\ &\quad \Big[1 + \frac{1}{\nu\sigma^2}\Big\{(\boldsymbol{b}-\boldsymbol{\beta})XX'(\boldsymbol{b}-\boldsymbol{\beta})' + s^2 \\ &\quad +\boldsymbol{b}_f(\boldsymbol{e}_f)X_fX_f'\boldsymbol{b}_f'(\boldsymbol{e}_f) + s_f^2(\boldsymbol{e}_f)\Big\}\Big]^{-\frac{\nu+n+n_f}{2}} \end{aligned} \quad (5.2)$$

where $\boldsymbol{b} = \boldsymbol{b}(\boldsymbol{y})$ and $s^2 = s^2(\boldsymbol{y})$. The normalizing constant $\Psi_2(\cdot)$ can be obtained by integrating the above function over the appropriate domain 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 future 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.

The joint posterior density of $\boldsymbol{\beta}$ and σ, and the future response statistics $\boldsymbol{b}_f(\boldsymbol{y}_f)$ and $s_f^2(\boldsymbol{y}_f)$ is obtained by using the Jacobian of the transformation

$$J\Big\{[\boldsymbol{b}_f(\boldsymbol{e}_f), s_f^2(\boldsymbol{e}_f)] \to [\boldsymbol{b}_f(\boldsymbol{y}_f), s^2(\boldsymbol{y}_f)]\Big\} = [\sigma^2]^{-\frac{p+2}{2}} \quad (5.3)$$

and the non-informative prior distribution of the parameters of the model and the density in (5.2), as

$$\begin{aligned} p\Big(\boldsymbol{\beta}, \sigma^2, \boldsymbol{b}_f, s_f^2\Big) &= \Psi_3(\cdot) \times [s^2]^{\frac{n-p-2}{2}}[s_f^2(\boldsymbol{y}_f)]^{\frac{n_f-p-2}{2}}[\sigma^2]^{-\frac{n+n_f-p}{2}} \\ &\quad \times \Big[1 + \frac{1}{\nu\sigma^2}\Big\{\Big(\boldsymbol{b}-\boldsymbol{\beta}\Big)XX'\Big(\boldsymbol{b}-\boldsymbol{\beta}\Big)' + s^2 \\ &\quad +\Big(\boldsymbol{b}_f-\boldsymbol{\beta}\Big)X_fX_f'\Big(\boldsymbol{b}_f-\boldsymbol{\beta}\Big)' + s_f^2\Big\}\Big]^{-\frac{\nu+n+n_f}{2}} \end{aligned} \quad (5.4)$$

where $\Psi_3(\cdot)$ is the normalizing constant, and $\boldsymbol{b}_f = \boldsymbol{b}_f(\boldsymbol{y}_f)$ and $s_f^2 = s_f^2(\boldsymbol{y}_f)$ for notational convenience.

A similar result can be obtained by using the structural distribution approach. In fact, the final results of this paper will be the same as that obtained by the structural distribution approach. Interested readers may refer to Fraser and Haq (1969) for details. The quadratic form in $\boldsymbol{\beta}$, $\boldsymbol{b}$ and $\boldsymbol{b}_f$ of the above density can be expressed as follows:

$$\Big(\boldsymbol{b}-\boldsymbol{\beta}\Big)XX'\Big(\boldsymbol{b}-\boldsymbol{\beta}\Big)'+\Big(\boldsymbol{b}_f-\boldsymbol{\beta}\Big)X_fX_f'\Big(\boldsymbol{b}_f-\boldsymbol{\beta}\Big)'= \Big(\boldsymbol{\beta}-FA^{-1}\Big)A\Big(\boldsymbol{\beta}-FA^{-1}\Big)'+\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)H^{-1}\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)' \quad (5.5)$$

where $F=\boldsymbol{b}XX'+\boldsymbol{b}_fX_fX_f'$, $A=XX'+X_fX_f'$, and $H=[XX']^{-1}+[X_fX_f']^{-1}$. Thus, the normalizing constant for the joint distribution of $\boldsymbol{\beta}$, σ^2, $\boldsymbol{b}_f$ and s_f^2 becomes, $\Psi_3(\cdot)=|A|^{\frac{1}{2}}|H|^{-\frac{1}{2}}[s^2]^{\frac{n-p}{2}}\Big\{[2]^{\frac{n+n_f}{2}}(\pi)^p\,\Gamma(\frac{n-p+2}{2})\Gamma(\frac{n_f-p}{2})\Big\}^{-1}$. The marginal density of $\boldsymbol{b}_f$ and s_f^2 is obtained as

$$p\Big(\boldsymbol{b}_f,s_f^2|\boldsymbol{y}\Big)=\Psi_4(\cdot)\times[s_f^2]^{\frac{n_f-p-2}{2}}\Big[s^2+s_f^2+\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)H^{-1}\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)\Big]^{-\frac{n+n_f-p}{2}} \quad (5.6)$$

where $\Psi_4(\cdot)=|H|^{-\frac{1}{2}}\Gamma(\frac{n+n_f-p}{2})[s^2]^{\frac{n-p}{2}}\Big\{(\pi)^{\frac{p}{2}}\Gamma(\frac{n-p}{2})\Gamma(\frac{n_f-p}{2})\Big\}^{-1}$ is the normalizing constant. Note that the above joint density of FRV and FRSS are dependent, and hence it can't be expressed as the product of two marginal densities. However, the marginal densities are obtainable by the usual method of integration. Therefore, the prediction distribution of the future regression vector, $\boldsymbol{b}_f$ can be written in the usual multivariate Student-t distribution form as

$$p\Big(\boldsymbol{b}_f\Big|\boldsymbol{y}\Big) \;=\; \Psi_5(\cdot)\times\Big[1+\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)\Big[s^2H\Big]^{-1}\Big(\boldsymbol{b}_f-\boldsymbol{b}\Big)'\Big]^{-\frac{n}{2}} \quad (5.7)$$

where $\Psi_5(\cdot)=\Gamma(\frac{n}{2})[s^2]^{\frac{n-p}{2}}\Big\{(\pi)^{\frac{p}{2}}\Gamma(\frac{n-p}{2})|H|^{\frac{1}{2}}\Big\}^{-1}$. Thus, $[\boldsymbol{b}_f|\boldsymbol{y}]\sim t_p(n-p,\boldsymbol{b},s^2H)$ 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 is used to construct β-expectation tolerance region for the future regression vector.

5.2 Distribution of Future Residual Sum of Squares

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

$$p\Big(s_f^2(\boldsymbol{y}_f)\Big|\boldsymbol{y}\Big)=\Psi_6(\cdot)\times\Big[s_f^2(\boldsymbol{y}_f)\Big]^{\frac{n_f-p-2}{2}}\Big[s^2+s_f^2(\boldsymbol{y}_f)\Big]^{-\frac{n+n_f-2p}{2}} \quad (5.8)$$

where $\Psi_6(\cdot) = \Gamma\left(\frac{n+n_f-2p}{2}\right)[s^2]^{\frac{n-p}{2}}\left\{\Gamma\left(\frac{n-p}{2}\right)\Gamma\left(\frac{n_f-p}{2}\right)\right\}^{-1}$. The density function in (5.8) can be written in the usual beta distribution form as follows:

$$p\left(s_f^2\middle|\boldsymbol{y}\right) = \Psi_7 \times \left[s_f^2\right]^{\frac{n_f-p-2}{2}}\left[1+s^{-2}s_f^2\right]^{-\frac{n+n_f-2p}{2}} \tag{5.9}$$

where $\Psi_7(\cdot) = \Psi_6(\cdot) \times [s^2]^{-\frac{n-n_f-2p}{2}}$. The above density is a modified form of the beta density of the second kind with $(n_f - p)/2$ and $(n - p)/2$ degrees of freedom. This prediction distribution is used to construct β-expectation tolerance region for the future residual sum of squares.

6 Optimal β-Expectation Tolerance Region

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 optimal 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 in section 3, $R^*(\boldsymbol{y}) = \{\boldsymbol{W} : \boldsymbol{W} < \boldsymbol{W}^*\}$ is a β-expectation tolerance region for $\boldsymbol{W} > 0$ if $\boldsymbol{W}^*$ is the β^{th} quantile of the sampling distribution of the future statistic $\boldsymbol{W}$. That is, $R^*(\boldsymbol{y})$ is a β-expectation tolerance region for the future statistic $\boldsymbol{W}$ if $\boldsymbol{W}^*$ is such that

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

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

6.1 Tolerance Region for the FRV

In the current study, to find an optimal β-expectation tolerance region of the FRV, we use the prediction distribution of the FRV which is known to be a multivariate Student-t distribution. We use the following result 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 of the theorem 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

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

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 is given by the ellipsoidal region:

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

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

For the derivation of optimal β-expectation tolerance region of the FRSS we use the prediction distribution of the FRSS. 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 as follows. A region on the sample space of the responses is an optimal β-expectation tolerance region if it encloses 100β percent of the future residual sum of squares from the multiple regression model, and it is given by the ellipsoidal region

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

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.*

The proof of the theorem is straightforward. 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

$\left[s^2_{\boldsymbol{y}}\{s^2_f\}^{-1}\right]$ is distributed as a scaled F variable with $(n_f - p)$ and $(n - p)$ degrees of freedom. That is, $\left[s^2_{\boldsymbol{y}}\{s^2_f\}^{-1}\right] \sim \frac{n-p}{n_f-p} F_{n_f-p,n-p}$. Therefore an equivalent β-expectation tolerance region for the future residual sum of squares from the multiple regression model 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 Concluding Remarks

The optimal β-expectation tolerance regions for the FRV and FRSS of a multiple regression model with multivariate Student-t errors are obtained in this paper. This study reveals the fact that conditional on the observed responses, the prediction distribution of the FRV from the multiple regression model is a multivariate Student-t distribution. Similarly, conditional on the observed responses, the prediction distribution of the FRSS is a scaled beta distribution. The shape parameter or the number of degrees of freedom of the prediction distributions depends on the size of the observed sample and the dimension of the regression vector, but not on the shape parameter of the multiple regression model. The β-expectation tolerance region for the FRV is based on the distribution of an appropriate quadratic form of the FRV that 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 from the multiple regression model 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 such tolerance regions will contain the true future regression vector and true future residual sum of squares respectively. As shown by Khan (2005), the same optimal β-expectation tolerance regions for the FRV and FRSS is valid for the multiple regression model with normal errors.

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.

Bishop, J. (1976). Parametric tolerance regions. Unpublished Ph.D. Thesis, University of Toronto, Canada

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

Chmielewski, M.A. (1981). Elliptically symmetric distributions. *International Statistical Review*, Vol. 49, 67–74.

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

Fraser, D.A.S. (1953). Nonparametric tolerance regions. *Ann. Math. Statist.*, **24**, 44-55.

Fisher, R.A. (1956). Statistical Methods in Scientific Inference. Oli. & Boyd, London.

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

Fraser, D.A.S. (1979). Inference and Linear Models. McGraw-Hill, New York

Fraser, D.A.S., and Guttman, I. (1956). Tolerance regions. *Ann Math. Statist.*, **27**, 162-171.

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

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

Guttman, I. (1957). On the power of optimum tolerance regions. *Ann. Math. Statist.*, **28**, 773-778.

Guttman, I. (1959). Optimal tolerance regions and power when sampling from some nonnormal universes. *Ann. Math. Statist.*, **30**, 926-938.

Guttman, I. (1970a). Construction of β-content tolerance regions at confidence level γ for large samples from the k-variate normal distribution. *Ann. Math. Statist.*, **41**, 376-400.

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

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. (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.

Khan, S. (2000). Improved estimation of the mean vector for Student-t model. *Communications In Statistics: Theory & Methods* **29, (3)**, 507-527.

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.

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.

Paulson, E. (1943). A note on tolerance limits. *Ann.. Math. Statist.*, **14**, 90-93.

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

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.

Rinco, S. (1973). β-expectation tolerance regions based on the structural models. *Ph.D. thesis, Unv. of Western Ontario, Canada.*

Sheffe, H. and Tukey, J.W. (1944).A formula for sample sizes for population tolerance limits. *Ann. Math. Statist.* **15**, 217

Ullah, A. and Walsh, V.Z. (1984). On the robustness of ML, LR and W test in regression. *Econometrica.* **52**, 1055-1066.

Wallis, W.A. (1951). Tolerance intervals for linear regressions. *Proc. Berkeley Symp. on Math. Statist. and Prob.*, University of California Press, Berkeley

Wald, A. and Wolfowitz, J. (1946). Tolerance limits for a normal distribution. *Ann. Math. Statist.*, **17**, 208-215.

Wilks, S.S. (1941). Determination of sample sizes for setting tolerance limits. *Ann. Math. Statist.*, **12**, 91-96.

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.