

# An Assessment of the Effects of Prior Distributions on the Bayesian Predictive Inference

Azizur Rahman<sup>1</sup>, Junbin Gao<sup>2</sup>, Catherine D'Este<sup>3</sup> & Syed Ejaz Ahmed<sup>4</sup>

<sup>1</sup> School of Computing and Mathematics, Charles Sturt University, Wagga Wagga, Australia

<sup>2</sup> The University of Sydney Business School, University of Sydney, Sydney, Australia

<sup>3</sup> NCEPH, The Australian National University, Canberra, Australia

<sup>4</sup> Department of Mathematics and Statistics, Brock University, Ontario, Canada

Correspondence: Azizur Rahman, School of Computing and Mathematics, Charles Sturt University, Wagga Wagga, NSW 2678, Australia. E-mail: azrahman@csu.edu.au

Received: May 2, 2016 Accepted: May 18, 2016 Online Published: August 16, 2016

doi:10.5539/ijsp.v5n5p31

URL: <http://dx.doi.org/10.5539/ijsp.v5n5p31>

## Abstract

Predictive inference is one of the oldest methods of statistical analysis and it is based on observable data. Prior information plays an important role in the Bayesian methodology. Researchers in this field are often subjective to exercise noninformative prior. This study tests the effects of a range of prior distributions on the Bayesian predictive inference for different modelling situations such as linear regression models under normal and Student- $t$  errors. Findings reveal that different choice of priors not only provide different prediction distributions of the future response(s) but also change the location and/or scale or shape parameters of the prediction distributions.

**Keywords:** Bayesian method, normal and Student- $t$  error, prediction distributions, prior distributions, regression models

## 1. Introduction

The posterior distribution for parameters of a set of observations is typically the major objective of the Bayesian statistical analysis. A posterior distribution also implies a marginal density known as prediction distribution of any future observations from a model. Predictive inference is one of the oldest and useful methods of statistical inference. A prediction density of future response(s) can be derived from various statistical models. In general, predictive inference uses the observed responses from a *performed experiment* to make inferences about the behavior of the unobserved future response(s) of a *future experiment* (Aitchison & Dunsmore, 1975). A detail of predictive inference methods and applications of prediction density can be found elsewhere (Geisser, 1993; Rahman, 2008, Rahman et al., 2010; Rahman & Upadhyay, 2015; Rahman & Harding, 2016).

Various methods can lead to prediction density and different researchers considered different approaches in prediction problems. General prediction problems have been discussed by Jeffreys (1961), Aitchison and Sculthorpe (1965), and Faulkenberry (1973). Goldberger (1962), Wilson (1967) and Hahn (1972) studied prediction problems by the classical method. Fraser and Haq (1970) obtained the prediction distribution by using the structural density function, and later on Haq (1982) develops the structural relations approach. Most of these authors have contribution in prediction problems under the independent normal errors assumption. Some researchers have been discussed the prediction problems and its applications in many areas from the Bayesian viewpoint (Zellner & Chetty, 1965; Aitchison & Dunsmore, 1975; Zellner, 1976; Sutradhar & Ali, 1989; Geisser, 1993; Rahman, 2008). Unlike others Rahman (2009, 2011) obtained prediction distribution using the Bayesian method for a range of statistical models under the  $t$  errors assumption. In many practical situations when the underlying distributions have heavier tails, models with Student  $t$ -errors are most appropriate.

A prior distribution of unknown parameters is an essential component in the Bayesian predictive inference. The prior distribution describes researcher's subjective belief about unknown parameter(s) before observing given data. When the distributions of data and the prior information are significantly different from each other, an inferential conflict exists between the sources of information. For example, the posterior distribution may be strongly affected by the prior information for models under the normal distribution assumption (O'Hagan & Forster, 2004). Some conflicts may cause by the data (for instance, outliers) or by the prior knowledge. In such situations, if the models are with a light-tailed distribution, the conflict may strongly influence the posterior distribution and potentially lead to inappropriate statistical inferences (Andrade & O'Hagan, 2006). A number of researchers suggest using models under heavy-tailed distributions such as Student- $t$  distributions (e.g., see Dawid, 1973; O'Hagan, 1979; Fernandez et al., 1995; Le & O'Hagan, 1998; Haro-Lopez & Smith 1999).

In the Bayesian predictive inference an important issue is how to define the prior distribution and examine its effect on statistical inference. In a model when both the prior and posterior distributions come from the same family of distributions, the prior distribution is considered as conjugate (Seber & Lee, 2003). That means, a conjugate prior is one for which application of Bayes' technique results in a posterior having the same family of distributions as prior. On the other hand, if there is no prior information available for the unknown parameters researchers want a prior distribution with minimal influence on the inference. Such a prior is called a noninformative prior or uniform prior (Bernardo & Smith, 1994).

Kass and Wasserman (1996) stated two different interpretations of noninformative priors: i) noninformative priors are formal representations of ignorance; ii) there is no objective, unique prior that represents ignorance, instead noninformative priors are chosen by agreement. In the second interpretation, noninformative priors are the default to use when there is insufficient information to otherwise define the prior. However, nowadays the first interpretation is somewhat useless, so focus on considering different priors to see if any is preferable in some sense.

Box and Tiao (1973) define a noninformative prior as a prior which provides little information relative to the experiment. Pericchi and Walley (1991) have a quite different view. They say that no single probability distribution can model ignorance satisfactory, therefore large classes of distributions are needed. They use the first interpretation of Kass and Wasserman (1996) but they realise that a single distribution is not enough. Therefore noninformative priors are also classified as vague prior, flat prior, and reference prior etc (e.g. see Berger & Bernardo, 1989).

The common form of noninformative prior is also known as the Jeffreys' prior (Jeffreys, 1946) which is based on the Fisher information criterion. For a model with parameter space  $\Theta \subseteq \mathbf{R}$ , the Fisher information is  $I(\theta) = E_{\theta} \left( \frac{d \log(f(x|\theta))}{d\theta} \right)^2$  where  $f(x | \theta)$  is the sampling distribution and the expectation is taken over  $f(x | \theta)$ . Under regularity conditions,  $I(\theta) = -E_{\theta} \left( \frac{d^2 \log(f(x|\theta))}{d\theta^2} \right)$ . In such a setting, the Jeffreys prior for  $\theta$  is defined by

$$g(\theta) \propto I(\theta)^{1/2}$$

to be proportional to the square root of the Fisher Information at  $\theta$ . In general the Jeffreys prior may be improper. If  $\theta$  has the Jeffreys prior and  $h$  is a monotone differentiable function of  $\theta$ , the prior induced on  $h(\theta)$  by the Jeffreys prior on  $\theta$  is

$$g(h(\theta)) = g(\theta) \left| \frac{dh}{d\theta} \right|^{-1} \propto I(\theta)^{1/2} \left| \frac{dh}{d\theta} \right|^{-1} = I(h(\theta))^{1/2}.$$

So the Jeffreys priors are invariant under reparameterization. If the prior favors values of  $\theta$  for which  $I(\theta)$  is large, the effect is to minimise the effect of the prior relative to the information in the data. In this sense the Jeffreys prior as attempting to be noninformative about  $\theta$ .

This study examines the effects of a range of prior distributions on predictive inference for different modelling situations under the normal and Student- $t$  errors. It is obvious that different choice of priors not only provides different prediction distributions but also changes the location and/or scale or shape parameters of the prediction distributions. Ultimately these also affect on measures of Bayesian credible intervals for the predictive inference.

The remainder of paper is organised as follow. In Section 2, the prediction densities of a future response have been derived using the Bayesian approach under different situations of the conjugate prior distributions. In Section 3, the prediction distribution has been obtained under the non-conjugate prior information. In Section 4, the prediction densities of a single and a set of unobserved responses have been produced for the simple linear regression model with the *Normal* errors and the *Student - t* errors. Finally, a summary of the significant results and the concluding remarks are provided in Section 5.

## 2. Prediction Under Conjugate Prior

The probability function of prior information is called *prior distribution* which illustrates the subjective belief of a researcher about unknown parameter(s) before observing data. In the Bayesian inference when both the prior and posterior distributions come from the same family of distributions, the prior distribution is regarded as conjugate prior. Let  $f(x|\theta)$  be the probability model,  $f(\mathbf{x}|\theta)$  be likelihood function for the observed data  $\mathbf{x}$  given the unknown parameter(s)  $\theta$ ,  $g(\theta)$  be the prior distribution model for  $\theta$  and  $f(\theta|\mathbf{x})$  be the posterior distribution model for  $\theta$  given  $\mathbf{x}$ . If  $g(\theta)$  and  $f(\theta|\mathbf{x})$  both belong to the same distributions family then they are called conjugate distributions, and  $g(\theta)$  is the conjugate prior for  $f(\mathbf{x}|\theta)$ . The term conjugate priors and its application appeared in a study by Raiffa and Schlaifer (1961). Before them the same theory was developed by Barnard (1954) as considered the distributions involved as being *closed under sampling*. If there exist, conjugate priors are often used for its computational convenience (Seber & Lee, 2003), and they provide the most transparent view of the relationship between prior and posterior. This section analyses the effects of a conjugate prior on predictive inference under various modelling situations.

### 2.1 Unequal but Known Variances

Let  $Y_j|\mu \sim N(\mu, \sigma_1^2)$  for  $j = 1, 2, \dots, n$ ;  $\mu \sim N(\mu_0, \sigma_0^2)$ ; and  $\sigma_1, \mu_0$  and  $\sigma_0$  are known. That is

$$p(\mathbf{y}|\mu, \sigma_1) \propto e^{-\frac{1}{2\sigma_1^2} \sum_{j=1}^n (y_j - \mu)^2}$$

is the likelihood function for a set of observations  $\mathbf{y} = (y_1, y_2, \dots, y_n)'$  and

$$g(\mu|\mu_0, \sigma_0) \propto e^{-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2}$$

is a prior density function of the parameter  $\mu$ . The posterior density function of  $\mu$  given  $\mathbf{y}$  can be obtained from the relationship

$$\begin{aligned} p(\mu|\mathbf{y}) &\propto p(\mathbf{y}|\mu, \sigma_1)g(\mu|\mu_0, \sigma_0) \\ &\propto e^{-\frac{1}{2} \left[ \frac{1}{\sigma_0^2} (\mu - \mu_0)^2 + \frac{1}{\sigma_1^2} \sum_{j=1}^n (y_j - \mu)^2 \right]} \end{aligned} \tag{1}$$

The enclosed exponential term  $\frac{1}{\sigma_0^2} (\mu - \mu_0)^2 + \frac{1}{\sigma_1^2} \sum_{j=1}^n (y_j - \mu)^2 = \xi(\text{say})$  can be expressed as more convenient form

$$\xi = \xi_1 + \left( \frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2} \right) \left( \mu - \frac{n\bar{y}\sigma_0^2 + \mu_0\sigma_1^2}{n\sigma_0^2 + \sigma_1^2} \right)^2, \tag{2}$$

where  $\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j$ ,  $s^2 = \frac{1}{n-1} \sum_{j=1}^n (y_j - \bar{y})^2$  and  $\xi_1 = \frac{\mu_0^2}{\sigma_0^2} + \frac{1}{\sigma_1^2} \{ (n-1)s^2 + n\bar{y}^2 \}$  does not contain the unknown parameter  $\mu$ . Using expression (2) in (1) the posterior density function of  $\mu$  becomes

$$p(\mu|\mathbf{y}) = \Psi(n, \sigma_0, \sigma_1) e^{-\frac{1}{2} \left( \frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2} \right) \left( \mu - \frac{n\bar{y}\sigma_0^2 + \mu_0\sigma_1^2}{n\sigma_0^2 + \sigma_1^2} \right)^2}; -\infty < \mu < \infty \tag{3}$$

where the normalizing constant is given by,  $\Psi(n, \sigma_0, \sigma_1) = \frac{\left( \frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2} \right)^{\frac{1}{2}}}{\sqrt{2\pi}}$ .

Let  $m'$  and  $s'^2$  are the mean and variance of the posterior distribution of  $\mu$ . Then

$$\mu|\mathbf{y} \sim N \left( m' = \frac{n\bar{y}\sigma_0^2 + \mu_0\sigma_1^2}{n\sigma_0^2 + \sigma_1^2}, s'^2 = \left[ \frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2} \right]^{-1} \right).$$

The posterior mean  $m'$  can be expressed as

$$m' = \frac{\frac{1}{\sigma_0^2}}{\frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2}} \times \mu_0 + \frac{\frac{n}{\sigma_1^2}}{\frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2}} \times \bar{y}$$

i.e., the posterior mean equals the weighted average of the prior mean and  $\bar{y}$  where the weights are the proportions of the posterior precision. In addition, the posterior variance  $s'^2$  has a relationship  $\frac{1}{s'^2} = \frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2}$  i.e., the posterior precision equals the precision of  $\bar{y}$  plus the prior precision.

Consider a future response  $y_f$  from the same normal model with mean  $\mu$  and known variance  $\sigma_1^2$  i.e.,  $y_f|\mu \sim N(\mu, \sigma_1^2)$ . Then the *pdf* of  $y_f$  is

$$p(y_f|\mu) = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{1}{2\sigma_1^2} (y_f - \mu)^2}; -\infty < y_f < \infty. \tag{4}$$

Using the Bayesian method the prediction distribution of  $y_f$  for a given  $\mathbf{y}$  can be derived as

$$\begin{aligned} f(y_f|\mathbf{y}) &= \int_{\mu} p(\mu|\mathbf{y})p(y_f|\mu)d\mu \\ &= C \int_{\mu} e^{-\frac{1}{2}\xi_f} d\mu, \end{aligned} \tag{5}$$

where  $C = \frac{1}{\sqrt{2\pi}\sigma_1} \frac{\left( \frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2} \right)^{\frac{1}{2}}}{\sqrt{2\pi}}$  and  $\xi_f = \left( \frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2} \right) \left( \mu - \frac{n\bar{y}\sigma_0^2 + \mu_0\sigma_1^2}{n\sigma_0^2 + \sigma_1^2} \right)^2 + \frac{1}{\sigma_1^2} (y_f - \mu)^2$  come from equations (3) and (4). Now,  $\xi_f$  has a convenient expression

$$\xi_f = M + \left( \frac{n+1}{\sigma_1^2} + \frac{1}{\sigma_0^2} \right) \left[ \mu - \frac{Q}{\left( \frac{n+1}{\sigma_1^2} + \frac{1}{\sigma_0^2} \right)} \right]^2 \tag{6}$$

where  $M = \frac{y_f^2}{\sigma_1^2} + \frac{\left(\frac{n\bar{y} + \mu_0}{\sigma_1^2 + \sigma_0^2}\right)^2}{\left(\frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2}\right)}$  which does not contain  $\mu$ , and  $Q = \frac{y_f}{\sigma_1^2} + \frac{n\bar{y}}{\sigma_1^2} + \frac{\mu_0}{\sigma_0^2}$ . Using equation (6) in (5) and then integrating over  $\mu$  we get,

$$f(y_f|\mathbf{y}) = C\sqrt{2\pi}\left[\frac{n+1}{\sigma_1^2} + \frac{1}{\sigma_0^2}\right]^{-\frac{1}{2}} e^{-\frac{1}{2}M} = \frac{\left[\frac{n\sigma_0^2 + \sigma_1^2}{\sigma_1^2\{\sigma_1^2 + (n+1)\sigma_0^2}\right]^{\frac{1}{2}}}{\sqrt{2\pi}} e^{-\frac{1}{2}M}. \tag{7}$$

Applying the convenient form  $M = b(y_f - a)^2$  where  $a = \frac{n\bar{y}\sigma_0^2 + \mu_0\sigma_1^2}{n\sigma_0^2 + \sigma_1^2}$  and  $b = \frac{n\sigma_0^2 + \sigma_1^2}{\sigma_1^2\{\sigma_1^2 + (n+1)\sigma_0^2\}}$ , the prediction density of  $y_f$  can be written as

$$f(y_f|\mathbf{y}) = \Psi_f(\cdot) e^{-\frac{1}{2}\left[\frac{n\sigma_0^2 + \sigma_1^2}{\sigma_1^2\{\sigma_1^2 + (n+1)\sigma_0^2}\right] \left(y_f - \frac{n\bar{y}\sigma_0^2 + \mu_0\sigma_1^2}{n\sigma_0^2 + \sigma_1^2}\right)^2}; -\infty < y_f < \infty \tag{8}$$

where  $\Psi_f(\cdot) = \Psi_f(n, \sigma_0, \sigma_1) = \frac{\left[\frac{n\sigma_0^2 + \sigma_1^2}{\sigma_1^2\{\sigma_1^2 + (n+1)\sigma_0^2}\right]^{\frac{1}{2}}}{\sqrt{2\pi}}$  is the normalizing constant.

Hence the prediction density of a future response  $y_f$  follows a normal distribution with mean equals to the posterior mean  $\frac{n\bar{y}\sigma_0^2 + \mu_0\sigma_1^2}{n\sigma_0^2 + \sigma_1^2}$  and with variance  $\left[\frac{n\sigma_0^2 + \sigma_1^2}{\sigma_1^2\{\sigma_1^2 + (n+1)\sigma_0^2}\right]^{-1}$ , that is greater than the posterior variance  $\left[\frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2}\right]^{-1}$ .

### 2.2 Equal but Unknown Variances

This subsection consider the common unknown variance  $\sigma^2$  for both the model and prior distribution. Let  $Y_j|\mu, \sigma \sim N(\mu, \sigma^2)$  for  $j = 1, 2, \dots, n$ ; that is, the likelihood function for  $\mathbf{y}$  is  $p(\mathbf{y}|\mu, \sigma) \propto \sigma^{-n} e^{-\frac{1}{2\sigma^2} \sum_{j=1}^n (y_j - \mu)^2}$ . Consider a normal prior for the location parameter  $\mu$  with known mean  $\mu_0$  and common unknown variance from the model such that  $\mu|\sigma \sim N(\mu_0, \sigma^2)$ , and  $g(\mu, \sigma|\mu_0) \propto \sigma^{-1} e^{-\frac{1}{2\sigma^2}(\mu - \mu_0)^2}$  is a joint prior density function of the parameters  $\mu$  and  $\sigma$ . Thus the joint posterior density function of  $\mu$  and  $\sigma$  can be expressed as

$$p(\mu, \sigma|\mathbf{y}) \propto \sigma^{-(n+1)} e^{-\frac{1}{2\sigma^2} [\sum_{j=1}^n (y_j - \mu)^2 + (\mu - \mu_0)^2]}. \tag{9}$$

The normalizing constant of (9) can be obtained by integrating the joint density function with respect to  $\sigma$  and  $\mu$ . If the normalizing constant is denoted by  $\Psi(c)$  then we have

$$\Psi^{-1}(c) = \frac{2^{\frac{n}{2}-1} \Gamma(\frac{n-1}{2}) \Gamma(\frac{1}{2})}{(n+1)^{\frac{1}{2}} [(n-1)s^2]^{\frac{n-1}{2}}} \left[ 1 + \frac{\frac{n}{n+1}(\mu_0 - \bar{y})^2}{(n-1)s^2} \right]^{-\frac{n-1}{2}}$$

where  $\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j$  and  $s^2 = \frac{1}{n-1} \sum_{j=1}^n (y_j - \bar{y})^2$ .

The integration of equation (9) with respect to  $\sigma$  yields the marginal posterior density function of  $\mu$  as

$$p(\mu|\mathbf{y}) = \Psi_1(\cdot) \left[ 1 + \frac{(n+1)(\mu - \frac{n\bar{y} + \mu_0}{n+1})^2}{(n-1)s^2 \left\{ 1 + \frac{\frac{n}{n+1}(\mu_0 - \bar{y})^2}{(n-1)s^2} \right\}} \right]^{-\frac{n}{2}}; -\infty < \mu < \infty \tag{10}$$

where the normalizing constant is given by

$$\Psi_1(n, s^2, \bar{y}) = \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2}) \Gamma(\frac{1}{2})} \frac{(n+1)^{\frac{1}{2}}}{[(n-1)s^2]^{\frac{1}{2}}} \left[ 1 + \frac{\frac{n}{n+1}(\mu_0 - \bar{y})^2}{(n-1)s^2} \right]^{-\frac{1}{2}}.$$

Let  $U = \left[ \frac{(n+1)(\mu - \frac{n\bar{y} + \mu_0}{n+1})^2}{s^2 \left\{ 1 + \frac{\frac{n}{n+1}(\mu_0 - \bar{y})^2}{(n-1)s^2} \right\}} \right]^{\frac{1}{2}}$  then  $U$  has a Student- $t$  distribution with  $n - 1$  df i.e.,  $U \sim t_{n-1}$ .

Now if  $y_f|\mu, \sigma$  be a future response from  $N(\mu, \sigma^2)$  i.e.,  $p(y_f|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(y_f - \mu)^2}$ . Then the prediction distribution of  $y_f$  for given  $\mathbf{y}$  can be obtained from the following equation

$$f(y_f|\mathbf{y}) = \int_{\mu} \int_{\sigma} p(\mu, \sigma|\mathbf{y}) p(y_f|\mu, \sigma) d\sigma d\mu. \tag{11}$$

Using appropriate integration over  $\sigma$  and  $\mu$  we get the density function of future response as

$$f(y_f|\mathbf{y}) = \Psi_f(.) \left[ 1 + \frac{\frac{n+1}{n+2}(y_f - \frac{n\bar{y} + \mu_0}{n+1})^2}{(n-1)s^2 \left\{ 1 + \frac{\frac{n}{n+1}(\mu_0 - \bar{y})^2}{(n-1)s^2} \right\}} \right]^{-\frac{n}{2}} ; -\infty < y_f < \infty \tag{12}$$

where

$$\Psi_f(n, s^2, \bar{y}) = \sqrt{\frac{n+1}{n+2}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2}) \sqrt{\pi}} \frac{1}{[(n-1)s^2]^{\frac{1}{2}}} \left[ 1 + \frac{\frac{n}{n+1}(\mu_0 - \bar{y})^2}{(n-1)s^2} \right]^{-\frac{1}{2}}$$

is the normalizing constant.

If  $V_f = \left[ \frac{\frac{n+1}{n+2}(y_f - \frac{n\bar{y} + \mu_0}{n+1})^2}{s^2 \left\{ 1 + \frac{\frac{n}{n+1}(\mu_0 - \bar{y})^2}{(n-1)s^2} \right\}} \right]^{\frac{1}{2}}$ , then  $V_f \sim t_{n-1}$ .

In this case, although the joint posterior distribution follows a normal distribution, the marginal posterior distribution of location parameter  $\mu$  and the prediction distribution of a future response  $y_f$  both have a Student- $t$  distribution with  $n-1$   $df$  and the same location  $\frac{n\bar{y} + \mu_0}{n+1}$ , but different scales such as  $\frac{(n+1)}{s^2} \left\{ 1 + \frac{\frac{n}{n+1}(\mu_0 - \bar{y})^2}{(n-1)s^2} \right\}^{-1}$  and  $\frac{(n+1)}{(n+2)s^2} \left\{ 1 + \frac{\frac{n}{n+1}(\mu_0 - \bar{y})^2}{(n-1)s^2} \right\}^{-1}$  respectively.

### 3. Prediction under Nonconjugate Prior

This section deals with the predictive inference of normal model under noninformative prior information (Jeffreys, 1961).

Let  $Y_j|\mu, \sigma \sim N(\mu, \sigma^2)$ , for  $j = 1, 2, \dots, n$ . A nonconjugate prior distribution of parameters  $\mu$  and  $\sigma$  is  $p(\mu, \sigma) \propto \frac{1}{\sigma}$ . Hence the joint posterior density function of parameters is as follows,

$$p(\mu, \sigma|\mathbf{y}) \propto \sigma^{-(n+1)} e^{-\frac{1}{2\sigma^2} \sum_{j=1}^n (y_j - \mu)^2}; -\infty < \mu < \infty, \sigma > 0$$

and the normalizing constant say  $\Psi^{-1}(\cdot)$  can be expressed as a relation,

$$\Psi(\cdot) = \int_{\mu} \int_{\sigma} (\sigma^2)^{-\frac{(n+1)}{2}} e^{-\frac{1}{2\sigma^2} \sum_{j=1}^n (y_j - \mu)^2} d\sigma d\mu.$$

Now after taking integration over  $\sigma$  we get

$$\Psi(\cdot) = 2^{\frac{n}{2}-1} \Gamma(\frac{n}{2}) \int_{\mu} \left[ \sum_{j=1}^n (y_j - \mu)^2 \right]^{-\frac{n}{2}} d\mu.$$

Again by taking integration over  $\mu$  we have the following result,

$$\Psi(\cdot) = \frac{2^{\frac{n}{2}-1}}{\Gamma(\frac{n-1}{2})} \sqrt{\frac{\pi}{n}} [(n-1)s^2]^{-\frac{n-1}{2}}$$

where  $s^2 = \frac{1}{n-1} \sum_{j=1}^n (y_j - \bar{y})^2$ .

Now if  $y_f|\mu, \sigma \sim N(\mu, \sigma^2)$  and  $p(y_f|\mu, \sigma)$ , then the prediction distribution of  $y_f$  for given  $\mathbf{y}$  can be obtained from the following equation

$$f(y_f|\mathbf{y}) = \int_{\mu} \int_{\sigma} p(\mu, \sigma|\mathbf{y}) p(y_f|\mu, \sigma) d\sigma d\mu.$$

Using appropriate integration with respect to  $\sigma$  and  $\mu$  we get the prediction density of future response  $y_f$  as

$$f(y_f|\mathbf{y}) = \Psi_f(.) \left[ 1 + \frac{\frac{n}{n+1}(y_f - \bar{y})^2}{(n-1)s^2} \right]^{-\frac{n}{2}} ; -\infty < y_f < \infty \tag{13}$$

where  $\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j$ , and  $\Psi_f(\cdot) = \sqrt{\frac{n}{n+1}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2}) \sqrt{\pi}} \frac{1}{[(n-1)s^2]^{\frac{1}{2}}}$  is the normalizing constant.

Hence  $t_f = \sqrt{\frac{n}{n+1}} \frac{(y_f - \bar{y})}{s}$  is a Student- $t$  variate with  $n-1$   $df$ . That is, under the noninformative prior the prediction density of a future response  $y_f$  has a Student- $t$  distribution with  $n-1$   $df$ .

### 4. Simple Linear Regression Model

The simple linear regression model involves a single explanatory variable (x) to a single response variable (y) linearly in the parameters. The model can be represented as

$$y_j = \beta_0 + \beta_1 x_j + e_j \text{ for } j = 1, 2, \dots, n$$

where  $y_j$  is the  $j$ th observation on the response variable,  $x_j$  is the  $j$ th observation on the explanatory variable,  $e_j$  is the  $j$ th error term associated with  $y_j$ th response, and  $\beta_0$  and  $\beta_1$  are the intercept and slope parameters respectively. This linear model can be expressed as a convenient form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{14}$$

where  $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ , a  $2 \times 1$  dimensional parameters vector;  $\mathbf{y}$  and  $\mathbf{e}$ , both are  $n \times 1$  dimensional response and error vectors respectively; and  $\mathbf{X}$ , a  $n \times 2$  dimensional design matrix of explanatory variable.

#### 4.1 Prediction under the Normal Distribution Error

Assume  $e_j$  is identically and independently distributed as a normal variable with mean zero and variance  $\sigma^2$  for  $j = 1, 2, \dots, n$ , that is,  $e_j|\sigma \sim N(0, \sigma^2)$ . Then,  $\mathbf{e}|\sigma \sim N(\mathbf{0}, \mathbf{I}_n\sigma^2)$ . The likelihood function of the response vector  $\mathbf{y}$  is

$$f(\mathbf{y}|\boldsymbol{\beta}, \sigma) \propto \sigma^{-n} e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}$$

Consider a noninformative prior distribution of unknown parameters  $\boldsymbol{\beta}$  and  $\sigma$  as  $p(\boldsymbol{\beta}, \sigma) \propto \frac{1}{\sigma}$ . Thus the joint posterior density function of parameters for given  $\mathbf{y}$  can be expressed as

$$p(\boldsymbol{\beta}, \sigma|\mathbf{y}) \propto \sigma^{-(n+1)} e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}; -\infty < \boldsymbol{\beta} < \infty, \sigma > 0$$

with the normalizing constant say  $\Psi^{-1}(\cdot)$ , and it can be obtained from the equation

$$\Psi(\cdot) = \int_{\boldsymbol{\beta}} \int_{\sigma} \sigma^{-(n+1)} e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} d\sigma d\boldsymbol{\beta}$$

Integration with respect to  $\sigma$  yields the result as

$$\Psi(\cdot) = \frac{\Gamma(\frac{n}{2})}{2^{-\frac{n-2}{2}}} \int_{\boldsymbol{\beta}} [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]^{-\frac{n}{2}} d\boldsymbol{\beta}, \tag{15}$$

Now

$$\begin{aligned} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\ &= [(n-2)s^2] \left[ 1 + \frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{(n-2)s^2} \right] \end{aligned} \tag{16}$$

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

Using (16) in (15) and then integrating over  $\boldsymbol{\beta}$  we get  $\Psi(\cdot) = \frac{\Gamma(\frac{n-2}{2})\pi[(n-2)s^2]^{-\frac{n-2}{2}}}{2^{-\frac{n-2}{2}}|\mathbf{X}'\mathbf{X}|^{\frac{1}{2}}}$ . Hence the normalizing constant of the joint posterior density function of parameters is as follows

$$\Psi^{-1}(\cdot) = \frac{|\mathbf{X}'\mathbf{X}|^{\frac{1}{2}} [(n-2)s^2]^{-\frac{n-2}{2}}}{2^{-\frac{n-2}{2}} \pi \Gamma(\frac{n-2}{2})}$$

Moreover, the future simple linear regression model can be expressed as

$$y_f = \mathbf{X}_f\boldsymbol{\beta} + e_f \tag{17}$$

where  $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ , a  $2 \times 1$  dimensional parameters vector for future response;  $y_f$  and  $e_f$ , both are  $1 \times 1$  dimensional response and error values respectively; and  $\mathbf{X}_f$ , a  $1 \times 2$  dimensional design matrix of future model explanatory variable. Also in future model we assume that  $e_f$  is identically and independently distributed as a normal variable with mean zero and variance  $\sigma^2$  that is,  $e_f|\sigma \sim N(0, \sigma^2)$ . Thus, the *pdf* of  $y_f$  given  $\boldsymbol{\beta}$  and  $\sigma$  is

$$f(y_f|\boldsymbol{\beta}, \sigma) = \frac{\sigma^{-1}}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_f-\mathbf{X}_f\boldsymbol{\beta})'(y_f-\mathbf{X}_f\boldsymbol{\beta})} \tag{18}$$

Now the prediction distribution of  $y_f$  for given  $\mathbf{y}$ ,  $\mathbf{X}$  and  $\mathbf{X}_f$  can be obtained from the following equation

$$f(y_f|\mathbf{y}, \mathbf{X}, \mathbf{X}_f) = \int_{\boldsymbol{\beta}} \int_{\sigma} p(\boldsymbol{\beta}, \sigma|\mathbf{y})p(y_f|\boldsymbol{\beta}, \sigma)d\sigma d\boldsymbol{\beta} = \frac{\Psi^{-1}(\cdot)}{\sqrt{2\pi}} \int_{\boldsymbol{\beta}} \int_{\sigma} \sigma^{-(n+2)} e^{-\frac{1}{2\sigma^2}[(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})+(\mathbf{y}_f-\mathbf{X}_f\boldsymbol{\beta})'(\mathbf{y}_f-\mathbf{X}_f\boldsymbol{\beta})]} d\sigma d\boldsymbol{\beta} \tag{19}$$

By integrating with respect to  $\sigma$  provides the result as

$$f(y_f|\mathbf{y}, \mathbf{X}, \mathbf{X}_f) = \frac{\Psi^{-1}(\cdot)}{\sqrt{2\pi}} \frac{\Gamma(\frac{n+1}{2})}{2^{-\frac{n-1}{2}}} \int_{\boldsymbol{\beta}} [\mathbf{D}]^{-\frac{n+1}{2}} d\boldsymbol{\beta} \tag{20}$$

where,  $\mathbf{D} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + (\mathbf{y}_f - \mathbf{X}_f\boldsymbol{\beta})'(\mathbf{y}_f - \mathbf{X}_f\boldsymbol{\beta})$ . Now  $\mathbf{D}$  can be expressed as a convenient form

$$\begin{aligned} \mathbf{D} &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + (\mathbf{y}_f - \mathbf{X}_f\boldsymbol{\beta})'(\mathbf{y}_f - \mathbf{X}_f\boldsymbol{\beta}) \\ &= \mathbf{y}'\mathbf{y} + y_f^2 + \boldsymbol{\beta}'\mathbf{Q}\boldsymbol{\beta} - 2\boldsymbol{\beta}'(\mathbf{X}'\mathbf{y} + \mathbf{X}'_f y_f) \\ &= \mathbf{A} [1 + (\boldsymbol{\beta} - \mathbf{P})'\mathbf{Q}_1(\boldsymbol{\beta} - \mathbf{P})] \end{aligned} \tag{21}$$

where  $\mathbf{A} = \mathbf{y}'\mathbf{y} + y_f^2 - (\mathbf{y}'\mathbf{X} + y_f\mathbf{X}'_f)\mathbf{Q}^{-1}(\mathbf{X}'\mathbf{y} + \mathbf{X}'_f y_f)$  is free from unknown  $\boldsymbol{\beta}$ ,  $\mathbf{Q} = \mathbf{X}'\mathbf{X} + \mathbf{X}'_f\mathbf{X}_f$ ,  $\mathbf{P} = \mathbf{Q}^{-1}(\mathbf{X}'\mathbf{y} + \mathbf{X}'_f y_f)$  and  $\mathbf{Q}_1 = \mathbf{Q}\mathbf{A}^{-1}$ .

Using (21) in (20) and then integrating over  $\boldsymbol{\beta}$ , we obtain the following prediction density function of  $y_f$  as

$$f(y_f|\mathbf{y}, \mathbf{X}, \mathbf{X}_f) = \Psi_f(\cdot) \left[ 1 + \frac{z'_f w_1 z_f}{(n-2)s^2} \right]^{-\frac{n-1}{2}} ; -\infty < y_f < \infty \tag{22}$$

where  $z_f = (\mathbf{y}_f - \mathbf{X}_f\hat{\boldsymbol{\beta}})$ ,  $w_1 = (1 - \mathbf{X}_f\mathbf{Q}^{-1}\mathbf{X}'_f)$  and  $\Psi_f(\cdot) = \frac{\Gamma(\frac{n-1}{2})|w_1|^{\frac{1}{2}}}{\Gamma(\frac{n-2}{2})[\pi(n-2)s^2]^{\frac{1}{2}}}$  is the normalizing constant.

Moreover, according to the same approach as considered above the prediction density function of a set of future response  $\mathbf{y}_f = (y_1, y_2, \dots, y_{n_f})'$  for given  $\mathbf{y}$ ,  $\mathbf{X}$  and  $\mathbf{X}_f$  can be obtain as

$$f(\mathbf{y}_f|\mathbf{y}, \mathbf{X}, \mathbf{X}_f) = \Psi_{n_f}(\cdot) \left[ 1 + \frac{\mathbf{z}'_f \mathbf{w}_2 \mathbf{z}_f}{(n-2)s^2} \right]^{-\frac{n+n_f-2}{2}} ; -\infty < y_f < \infty \tag{23}$$

where  $\mathbf{z}_f = (\mathbf{y}_f - \mathbf{X}_f\hat{\boldsymbol{\beta}})$ ,  $\mathbf{w}_2 = [\mathbf{I}_{n_f} - \mathbf{X}_f\mathbf{Q}^{-1}\mathbf{X}'_f]$  and  $\Psi_{n_f}(\cdot) = \frac{\Gamma(\frac{n+n_f-2}{2})|\mathbf{w}_2|^{\frac{1}{2}}}{\Gamma(\frac{n-2}{2})[\pi(n-2)]^{\frac{n_f}{2}}(s^2)^{\frac{1}{2}}}$  is the normalizing constant, with  $s^2 = \frac{1}{n-2}[(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})]$ ,  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ , and  $\mathbf{Q} = \mathbf{X}'\mathbf{X} + \mathbf{X}'_f\mathbf{X}_f$ .

#### 4.1 Prediction under the Student-t Error

Assume that each of the  $n$  components in  $\mathbf{e}$  for the simple linear regression model in (14) is uncorrelated but not independent of the others and has the same univariate Student- $t$  distribution with location 0, scale  $\sigma > 0$  and  $\nu$  degrees of freedom. Therefore, the joint  $pdf$  for the  $n$  elements of  $\mathbf{e}|\sigma$  is an  $n$ -dimensional multivariate Student- $t$  distribution

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

Here  $E(\mathbf{e}) = \mathbf{0}$ , and  $Cov(\mathbf{e}) = \frac{\nu\sigma^2}{(\nu-2)}\mathbf{I}_n$  for  $\nu > 2$ . Thus the elements of  $\mathbf{e}$  and hence those of  $\mathbf{y}$  are uncorrelated but not independent. Therefore the probability density function of the realized vector  $\mathbf{y}$  becomes

$$f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \nu) \propto (\sigma^2)^{-\frac{n}{2}} \left[ \nu + \frac{1}{\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right]^{-\frac{\nu+n}{2}} . \tag{24}$$

Let a single response from the future model in equation (17) follows the assumption that  $e_f|\sigma$  has univariate Student- $t$  distribution with  $\nu$  degrees of freedom i.e.,  $e_f|\sigma \sim t_1(0, \sigma, \nu)$ . Hence, the realized error  $\mathbf{e}$  and the future error  $e_f$  have been combined to form a  $(n + 1)$  dimensional multivariate Student- $t$  distribution with  $\nu$  degrees of freedom. Accordingly

the joint density function of the combined responses  $\mathbf{y}$  from the *performed experiment* and  $y_f$  from the *future experiment* becomes

$$p(\mathbf{y}, y_f | \boldsymbol{\beta}, \sigma^2) \propto (\sigma^2)^{-\frac{n+1}{2}} \left[ \nu + \frac{1}{\sigma^2} (Q_y + Q_{y_f}) \right]^{-\frac{\nu+n+1}{2}} \tag{25}$$

where  $Q_y = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$  and  $Q_{y_f} = (y_f - \mathbf{X}_t\boldsymbol{\beta})'(y_f - \mathbf{X}_t\boldsymbol{\beta})$ .

Let a noninformative joint prior distribution of unknown parameters  $\boldsymbol{\beta}$  and  $\sigma^2$  be  $p(\boldsymbol{\beta}, \sigma^2) \propto \sigma^{-2}$ . It is assumed that the degrees of freedom of the error distribution are given and the elements of  $\boldsymbol{\beta}$  and  $\log \sigma^2$  are independently and uniformly distributed. Combining the prior information with the joint density function in (25) by means of Bayes' Theorem, we have the following joint posterior density of  $\boldsymbol{\beta}$  and  $\sigma^2$  for  $\mathbf{y}$  and  $y_f$

$$p(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, y_f) \propto p(\mathbf{y}, y_f | \boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}, \sigma^2) \propto (\sigma^2)^{-\frac{n+3}{2}} \left[ \nu + \frac{Q}{\sigma^2} \right]^{-\frac{\nu+n+1}{2}} \tag{26}$$

where  $Q = Q_y + Q_{y_f}$ .

The prediction distribution of a future response can be obtained by solving the following integral

$$f(y_f | \mathbf{y}) \propto \int_{\boldsymbol{\beta}} \int_{\sigma^2} p(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, y_f) d\sigma^2 d\boldsymbol{\beta}. \tag{27}$$

That means in this case we can obtain the prediction distribution of future response(s) from the joint posterior density of unknown parameters for combined responses generate from both the performed and future models. Now the joint posterior density in (26) can be expressed as the following convenient form

$$p(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, y_f) \propto (\sigma^2)^{\frac{\nu}{2}-1} [Q + \nu\sigma^2]^{-\frac{\nu+n+1}{2}} \tag{28}$$

Using a transformation as  $Q + \nu\sigma^2 = t^{-1}$  we have the Jacobian of the transformation  $|J| = \frac{1}{\nu t^2}$  with the range of  $t$  from 0 to  $Q^{-1}$ . Hence equation (28) can be written as

$$\begin{aligned} f(y_f | \mathbf{y}) &\propto \int_{\boldsymbol{\beta}} \int_0^{Q^{-1}} \left[ \frac{1}{\nu} \left( \frac{1}{t} - Q \right) \right]^{\frac{\nu}{2}-1} \left[ \frac{1}{t} \right]^{-\frac{\nu+n+1}{2}} \frac{1}{\nu} \left( \frac{1}{t} \right)^2 dt d\boldsymbol{\beta} \\ &\propto \nu^{-\frac{\nu}{2}} \int_{\boldsymbol{\beta}} \int_0^{Q^{-1}} [1 - Qt]^{\frac{\nu}{2}-1} t^{\frac{n+1}{2}-1} dt d\boldsymbol{\beta}. \end{aligned} \tag{29}$$

If we put  $z = Qt$ , then the joint *pdf* in (29) becomes

$$f(y_f | \mathbf{y}) \propto \int_{\boldsymbol{\beta}} [Q]^{-\frac{n+1}{2}} \nu^{-\frac{\nu}{2}} \int_0^1 z^{\frac{n+1}{2}-1} [1 - z]^{\frac{\nu}{2}-1} dz d\boldsymbol{\beta}. \tag{30}$$

Equation (30) confirms that  $z$  has a *beta* distribution and hence  $z \sim B\left(\frac{n+1}{2}, \frac{\nu}{2}\right)$ . After integrating with respect to  $z$ , the above equation become to the following form

$$f(y_f | \mathbf{y}) \propto \int_{\boldsymbol{\beta}} [Q]^{-\frac{n+1}{2}} d\boldsymbol{\beta}. \tag{31}$$

Now,  $Q = Q_y + Q_{y_f}$  can be expressed as the convenient quadratic form of the parameters vector  $\boldsymbol{\beta}$ , i.e.,  $Q = A + (\boldsymbol{\beta} - B)'M(\boldsymbol{\beta} - B)$ . Where  $M = \mathbf{X}'\mathbf{X} + \mathbf{X}'_t\mathbf{X}_t$ ,  $B = M^{-1}(\mathbf{X}'\mathbf{y} + \mathbf{X}'_ty_f)$  and  $A = \mathbf{y}'\mathbf{y} + y_f^2 - (\mathbf{y}'\mathbf{X} + y_f\mathbf{X}'_t)M^{-1}(\mathbf{X}'\mathbf{y} + \mathbf{X}'_ty_f)$ . It is noted that  $A$  is free from the unknown regression parameters' vector  $\boldsymbol{\beta}$ .

Using the convenient form of  $Q$ , the probability density function in (31) can be expressed as

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



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

$$f(y_f|\mathbf{y}) \propto [A]^{-\frac{n-1}{2}} \propto [\mathbf{y}'\mathbf{y} + y_f^2 - (\mathbf{y}'\mathbf{X} + y_f\mathbf{X}_f)\mathbf{M}^{-1}(\mathbf{X}'\mathbf{y} + \mathbf{X}'_f y_f)]^{-\frac{n-1}{2}} \tag{32}$$

Now,  $\mathbf{y}'\mathbf{y} + y_f^2 - (\mathbf{y}'\mathbf{X} + y_f\mathbf{X}_f)\mathbf{M}^{-1}(\mathbf{X}'\mathbf{y} + \mathbf{X}'_f y_f) = \mathbf{y}'(1 - \mathbf{X}\mathbf{M}^{-1}\mathbf{X}')\mathbf{y} + y_f'(1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)y_f - \mathbf{y}'\mathbf{X}\mathbf{M}^{-1}\mathbf{X}'_f y_f + y_f\mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'\mathbf{y} = \mathbf{y}'(1 - \mathbf{X}\mathbf{M}^{-1}\mathbf{X}')\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{M}^{-1}\mathbf{X}'_f(1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)^{-1}\mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'\mathbf{y} + [y_f - (1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)^{-1}\mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'\mathbf{y}](1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)[y_f - (1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)^{-1}\mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'\mathbf{y}]$ .

But using appropriate simplification we can obtain  $(1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)^{-1} = [1 + \mathbf{X}_f(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_f]$  and then  $(1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)^{-1}\mathbf{X}_f\mathbf{M}^{-1} = \mathbf{X}_f(\mathbf{X}'\mathbf{X})^{-1}$ . So the expression in (32) becomes  $\mathbf{y}'(1 - \mathbf{X}\mathbf{M}^{-1}\mathbf{X}')\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{M}^{-1}\mathbf{X}'_f\mathbf{X}_f(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} + (y_f - \mathbf{X}_f(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})'(1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)(y_f - \mathbf{X}_f(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})$ .

Applying the OLS estimate  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$  we get  $\mathbf{y}'(1 - \mathbf{X}\mathbf{M}^{-1}\mathbf{X}')\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{M}^{-1}\mathbf{X}'_f\mathbf{X}_f\hat{\beta} + (y_f - \mathbf{X}_f\hat{\beta})'(1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)(y_f - \mathbf{X}_f\hat{\beta}) = \mathbf{y}'[1 - \mathbf{X}(\mathbf{M}^{-1} + \mathbf{M}^{-1}\mathbf{X}'_f\mathbf{X}_f(\mathbf{X}'\mathbf{X})^{-1})\mathbf{X}']\mathbf{y} + (y_f - \mathbf{X}_f\hat{\beta})'(1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)(y_f - \mathbf{X}_f\hat{\beta}) = \mathbf{y}'[1 - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} + (y_f - \mathbf{X}_f\hat{\beta})'(1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)(y_f - \mathbf{X}_f\hat{\beta})$ . Also since  $\mathbf{y}'[1 - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\beta} = (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) = (n - 2)s^2$ , the expression in (32) becomes the convenient form as  $(n - 2)s^2 + (y_f - \mathbf{X}_f\hat{\beta})'(1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)(y_f - \mathbf{X}_f\hat{\beta})$ . Thus, using the ultimate form the prediction distribution of a single future response  $y_f$  is obtained as

$$f(y_f|\mathbf{y}) = \Psi_f(\cdot) \left[ (n - 2) + (y_f - \mathbf{X}_f\hat{\beta})' H (y_f - \mathbf{X}_f\hat{\beta}) \right]^{-\frac{n-1}{2}} \tag{33}$$

where  $H = \frac{(1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)}{s^2}$ ,  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ ,  $s^2 = \frac{1}{n-2}[(\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})]$  and  $\Psi_f(\cdot) = \frac{\Gamma(\frac{n-1}{2})|1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f|^{\frac{1}{2}}}{\Gamma(\frac{n-2}{2})[\pi(n-2)s^2]^{\frac{1}{2}}}$  is the normalizing constant.

Therefore, under the Student- $t$  errors the prediction distribution of a future response has an univariate Student- $t$  distribution of  $n - 2$   $df$ , with location  $\mathbf{X}_f(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$  and variance  $\frac{(n-2)s^2}{(n-4)}(1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)^{-1}$ . It is also revealed that the  $df$  of the prediction distribution does not depend on the  $df$  parameter  $\nu$  of the  $t$ -distribution for error terms.

Furthermore, employing the same procedures the prediction distribution of a set of future responses  $\mathbf{y}_f = (y_1, y_2, \dots, y_{n_f})'$  from simple linear regression model [i.e.,  $\mathbf{y}_f = \mathbf{X}_f\beta + \mathbf{e}_f$ , with  $\mathbf{e}_f|\sigma \sim t_{n_f}(0, \sigma, \nu)$ ] for the given  $\mathbf{y}$ ,  $\mathbf{X}$  and  $\mathbf{X}_f$  can be obtain as

$$f(\mathbf{y}_f|\mathbf{y}) = \Psi_{n_f}(\cdot) \left[ (n - 2) + (\mathbf{y}_f - \mathbf{X}_f\hat{\beta})' H (\mathbf{y}_f - \mathbf{X}_f\hat{\beta}) \right]^{-\frac{n+n_f-2}{2}} \tag{34}$$

for  $H = \frac{(\mathbf{I}_{n_f} - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)}{s^2}$ ,  $\hat{\beta}$  and  $s^2$  are same as of (33) and  $\Psi_{n_f}(\cdot) = \frac{\Gamma(\frac{n+n_f-2}{2})|\mathbf{I}_{n_f} - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f|^{\frac{1}{2}}}{\Gamma(\frac{n-2}{2})[\pi^{n_f}(n-2)s^2]^{\frac{1}{2}}}$ .

Therefore, under the Student- $t$  errors a set of future responses  $\mathbf{y}_f$  has the prediction density of an  $n_f$ -dimensional Student- $t$  distribution with  $n - 2$   $df$ . This indicate that the shape parameter of the prediction distribution depends on the size of the observed sample  $n$  and the dimension of regression parameters  $\beta$ . The location and scale of the prediction distribution are  $\mathbf{X}_f\hat{\beta}$  and  $\left\{s^{-2}(\mathbf{I}_{n_f} - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}'_f)\right\}^{-\frac{1}{2}}$  respectively. This result coincides with that of Zellner (1971) and Hahn (1972), where they had considered the normal error terms and obtained the distribution using the *classical* method. Consistant results had also obtained for the normal errors regression model by using the *structural distribution* (Fraser & Haq, 1970) and *structural relations of the model* (Fraser & Ng, 1980) approaches instead of the Bayesian approach. So, the prediction density is unaffected by departures from the model with independent and normal errors to the Student- $t$  errors under the classical, structural distribution, structural relations of the model and Bayesian methods.

### 5. Conclusion

The posterior distribution for parameters of a set of observations is typically the major objective of the Bayesian statistical analysis. A posterior distribution implies a marginal distribution known as prediction distribution for outcomes of any future sample observations from a medel, and prediction distributions have many applications in statistical inference. In this paper, the effects of a range of conjugate and non-conjugate prior distributions on the Bayesian predictive inference have been tested. The prediction distributions for future response(s), conditional on a set of observed responses have been also derived for the simple linear regression model with the normal and Student- $t$  errors situations. Findings reveal that if a conjugate prior for the normal error model with known but unequal variances are used, then the joint posterior distribution of parameters, the marginal posterior distribution of the location parameter  $\mu$  and the prediction distribution

of future response  $y_f$  all have normal distributions. The mean of the prediction distribution is same as the posterior mean and the variance is greater than the posterior variance.

In another case, when we have considered a conjugate prior with the same unknown variance in the model the joint posterior distribution of parameters has revealed a normal distribution. However, the results have clearly demonstrated that the marginal posterior density of location parameter and the prediction distribution of future response both have a Student- $t$  distribution with  $n - 1$   $df$ , and with the same mean but different scale. In this specific situation, the variance of the prediction distribution is much smaller (i.e., at the minimum of four times smaller) than the variance of the marginal posterior distribution of  $\mu$ . And if we use a noninformative prior for the unknown parameters of the model, the finding reveals that the prediction density of future response follows a Student- $t$  distribution with  $n - 1$   $df$ .

Moreover, the results for the simple linear regression model show that conditional on a set of realised responses, the prediction distribution of a future observation has an univariate Student- $t$  distribution with  $n - 2$   $df$ , and with mean  $\mathbf{X}_f(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$  and variance  $\frac{(n-2)s^2}{(n-4)}(1 - \mathbf{X}_f\mathbf{M}^{-1}\mathbf{X}_f')^{-1}$  for both of the Normal and Student- $t$  errors situations. It is also obvious that the  $df$  of a prediction distribution does not depend on the degrees of freedom parameter  $\nu$  of the  $t$ -distribution of disturbance terms. Similarly, under the Bayesian method with the uniform prior, a set of future responses  $\mathbf{y}_f$  has the prediction distribution of an  $n_f$ -dimensional multivariate Student- $t$  distribution with  $n - 2$   $df$ . This indicates that the shape parameter of the prediction distribution depends on the size of the observed sample  $n$  and the dimension of  $\beta$ . The location and scale of the prediction distributions are same for the model under the Normal errors and the Student- $t$  errors situations.

In conclusion, findings reveal that the prediction distributions of a single future response and a set of future responses from the simple linear regression model unaffected by departures from the model with the Normal errors to the Student- $t$  errors distribution. The prediction distributions by the Bayesian method are also coincide with the results derived by the *classical, structural relations of the model* and *structural distribution* approaches. Ultimately, the prediction distribution depends on the observed responses and the design matrices of the realized model as well as the future model. The shape parameter of the prediction distribution depends on the size of the realized sample and the dimension of parameters vector of the model, but it is completely independent to the degrees of freedom of the error's distribution of the model.

### Acknowledgements

The authors sincerely thank the anonymous referees and an editor for their helpful comments which were used to improve the manuscript. An initial version of the paper was presented at the 2014 Australian Statistical Conference (ASC) in conjunction with the Institute of Mathematical Statistics (IMS) Annual Meeting in Sydney, Australia. We also thank several attendees including Richard Gerlach, Gary Glonek and Chong You for their valuable comments and suggestions which had contributed greatly to design this paper.

### References

- Aitchison, J., & Dunsmore, I. R. (1975). *Statistical prediction analysis*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511569647>
- Aitchison, J., & Sculthorpe, D. (1965). Some problems of statistical prediction. *Biometrika*, 52, 469-83. <http://dx.doi.org/10.2307/2333699>
- Andrade, J. A. A., & O'Hagan, A. (2006). Bayesian Robustness Modeling Using Regularly Varying Distribution. *Bayesian Analysis*, 1, 169-188. <http://dx.doi.org/10.1214/06-BA106>
- Barnard, G. A. (1954). Sampling Inspection and Statistical Decisions. *Journal of the Royal Statistical Society: Series B*, 16, 151-174.
- Berger, J., & Bernardo, J. (1989). Estimating a product of means: Bayesian analysis with reference priors. *Journal of the American Statistical Association*, 89, 200-207. <http://dx.doi.org/10.1080/01621459.1989.10478756>
- Bernardo, J., & Smith, A. (1994). *Bayesian theory*. New York: John Wiley and Sons. <http://dx.doi.org/10.1002/9780470316870>
- Box, G., & Tiao, G. (1973). *Bayesian Inference in Statistical Analysis*. New York: John Wiley and Sons.
- Dawid, A. P. (1973). Posterior Expectations for Large Observations. *Biometrika*, 60, 664C667. <http://dx.doi.org/10.1093/biomet/60.3.664>
- Faulkenberry, G. D. (1973). A method of obtaining prediction intervals. *Journal of the American Statistical Association*, 68, 433-435. <http://dx.doi.org/10.1080/01621459.1973.10482450>
- Fernandez, C., Osiewalski, J., & Steel, M. F. J. (1995). Modelling and Inference with V-spherical Distributions. *Journal*

- of the American Statistical Association, 90, 1331C1340. <http://dx.doi.org/10.2307/2291523>
- Fraser, D. A. S., & Ng, K. W. (1980). Multivariate regression analysis with spherical error. In V. P. R. Krishnaiah (ed.), *Multivariate Analysis* (pp. 369-386). New York: North-Holland Publishing Co.
- Fraser, D. A. S., & Haq, M.S. (1970). Inference and prediction for multilinear model. *Journal of Statistical Research*, 4, 93-109.
- Geisser, S. (1993). *Predictive inference: An introduction*. London: Chapman and Hall. <http://dx.doi.org/10.1007/978-1-4899-4467-2>
- Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, 57, 369-375. <http://dx.doi.org/10.1080/01621459.1962.10480665>
- Hahn, G. J. (1972). Simultaneous prediction intervals for a regression model. *Technometrics*, 14, 203-214. <http://dx.doi.org/10.1080/00401706.1972.10488896>
- Haq, M. S. (1982). Structural relation and prediction for multivariate models. *Statistische Hefte*, 23, 218-228. <http://dx.doi.org/10.1007/BF02933051>
- Haro-Lopez, R., & Smith, A. (1999). On robust Bayesian analysis for location and scale parameters. *Journal of Multivariate Analysis*, 70, 30C56. <http://dx.doi.org/10.1006/jmva.1999.1820>
- Jeffreys, H. (1961). *Theory of probability* (3rd ed). Oxford: Oxford University Press.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London: Series A, Mathematical and Physical Sciences*, 186, 453-461. <http://dx.doi.org/10.1098/rspa.1946.0056>
- Kass, R., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91, 1343-1370. <http://dx.doi.org/10.1080/01621459.1996.10477003>
- Le, H. & O'Hagan, A. (1998). A Class of Bivariate Heavy-tailed Distributions. *Sankya: The Indian Journal of Statistics*, B-60, 82C100.
- O'Hagan, A. (1979). On Outliers Rejection Phenomena in Bayes Inference. *Journal of the Royal Statistical Society: Series B*, 41, 358C367.
- O'Hagan, A. & Forster, J. J. (2004). *Bayesian Inference - Kendall's Advanced Theory of Statistics* (volume 2B). London: Arnold.
- Pericchi, L., & Walley, P. (1991). Robust Bayesian credible intervals and prior ignorance. *International Statistical Review*, 58, 1-23. <http://dx.doi.org/10.2307/1403571>
- Rahman, A. (2008). *Bayesian predictive inference for some linear models under Student-t errors*. Saarbrücken: VDM-Verlag.
- Rahman, A. (2009). Objective Bayesian prediction for the matrix-T error regression model. Paper presented at the 2009 *International Workshop on Objective Bayes Methodology (O-Bayes09)*, The Wharton School at the University of Pennsylvania, Philadelphia USA, pp. 1-13.
- Rahman, A. (2011). Bayesian predictive inference for multivariate simple regression model with matrix-T error. *Pioneer Journal of Theoretical and Applied Statistics*, 2, 99-112.
- Rahman, A., & Harding, A. (2016). *Small Area Estimation and Microsimulation Modelling*. Florida: Chapman and Hall or CRC Press.
- Rahman, A., Harding, A., Tanton, R., & Liu, S. (2010). Methodological Issues in Spatial Microsimulation Modelling for Small Area Estimation. *International Journal of Microsimulation*, 3, 3-22.
- Rahman, A., & Upadhyay, S. K. (2015). A Bayesian Reweighting Technique for Small Area Estimation. In S. K. Upadhyay et al., (eds), *Current Trend in Bayesian Methodology with Applications* (pp. 503-520). Florida: Chapman and Hall or CRC Press. <http://dx.doi.org/10.1201/b18502-25>
- Raiffa, H., & Schlaifer, R. (1961). *Applied statistical decision theory*. Boston: Clinton Press Inc.
- Seber, G. A. F., & Lee, A. J. (2003). *Linear regression analysis*. New York: Wiley. <http://dx.doi.org/10.1002/9780471722199>
- Sutradhar, B. C., & Ali, M. M. (1989). A generalization of the Wishart distribution for the elliptical model and its moments for the multivariate-t model. *Journal of Multivariate Analysis*, 29, 155-162. [41](http://dx.doi.org/10.1016/0047-</a></p></div><div data-bbox=)

259X(89)90082-1

- Wilson, A. L. (1967). An approach to simultaneous tolerance interval in regression. *Annals of Mathematical Statistics*, 38, 1536-40. <http://dx.doi.org/10.1214/aoms/1177698707>
- Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-*t* error terms. *Journal of the American Statistical Association*, 71, 400-405. <http://dx.doi.org/10.2307/2285322>
- Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. New York: John Wiley and Sons.
- Zellner, A., & Chetty, V. K. (1965). Prediction and decision problems in regression models from the Bayesian point of view. *Journal of the American Statistical Association*, 60, 608-616. <http://dx.doi.org/10.1080/01621459.1965.10480817>

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).