Prediction Distribution for Simultaneous Auto-Regressive Model with Multivariate Student-T Error under the Bayesian Approach

Md. Idris Ali[1], Md. Shajedul Islam[2], Md. Ahsan Ahamed[3]

[1,2]Assistant Professor, Department of Business Studies, North Bengal International University, Rajshahi, Bangladesh.
[3]Lecturer, Department of Business Studies, North Bengal International University, Rajshahi, Bangladesh.

Abstract:
Prediction distribution is a basis for predictive inferences applied in many real world situations. The Bayesian approach under uniform prior is employed in this paper to derive the prediction distribution for Simultaneous Auto-regressive model with multivariate Student-t error distribution. Conditional on a set of realized responses, a single and a set of future responses have a univariate and multivariate Student-t distributions respectively, whose degrees of freedom depend on the size of the realized sample and the dimension of the auto-regression parameters' vector but do not depend on the degrees of freedom of the error distribution. Results are identical to those obtained under normal error distribution by a range of statistical approaches such as the normal distribution, autocorrelations and classical methods. This indicates not only the inference robustness with respect to departures from normal error to multivariate Student-t error distributions, but also indicates that the Bayesian approach with uniform prior is competitive with other statistical methods in the derivation of prediction distribution.

Keywords: Student's t-distribution, Chi-square distribution, F-distribution, Auto-regressive model.

Introduction:

Simultaneous autoregressive (SAR) models have been used to describe the spatial variation of quantities of interest in the form of summaries or aggregates over regions, and have been applied for the analysis of data in diverse areas such as demography, economy and geography. SAR models are used for the general goal of unveiling and quantifying spatial relations present among the data, in particular, for detecting spatial clustering and assessing how quantities of interest are inuenced by explanatory variables. General accounts of SAR models have appeared in Anselin (1988), Haining (1990) and Cressie (1993).

The prediction distribution of future response(s) can be derived from the simultaneous auto-regressive model for statistical predictive inferences. Predictive inference uses the observations from a realized experiment to make inference about the performance of the future observation(s) of a future experiment. Many authors have considered the linear regression model in prediction problems and they have been used different methods to derive the prediction distribution. General prediction problems have been discussed by Jeffreys (1961). Fraser and Haq (1970) used the structural distribution approach, Aitchison and Dunsmore (1975) and Geisser (1993) used the Bayesian approach, and Haq (1982) and Haq and Khan (1990) used the structural relations approach to obtain the prediction distribution from the linear model to mention a few. For details of predictive inferences and applications of prediction distribution interested readers may refer to Geisser (1993) and Khan (2004), and references therein.

Most of the authors have contributed to solving the prediction problem by using linear models with independent and normal errors. Unlike others Rahman and Khan (2007) obtained prediction distribution for linear regression model with multivariate student-t errors under the bayesian approach and Haq and Khan (1990) obtained prediction distribution for the linear regression model with multivariate Student-t error terms by using the structural relation approach. In real life situations when the underlying distributions have heavier tails, linear models with multivariate Student-t errors have been emphasized and used by Zellner (1976), and Sutradhar and Ali (1989) among others. This study assumes that the error terms of the performed as well as the future simultaneous auto-regressive models have a joint multivariate Student-t distribution, and obtains the prediction distribution(s) of future response(s) by the Bayesian method under a uniform prior distribution.

Simultaneous Auto-regressive Model (SAR models):

Description: Consider a geographic region partitioned into sub-regions indexed by the integers 1, 2, . . . , n. This collection of sub regions is assumed to be endowed with a neighborhood system, $\{N_i: i = 1,2,3, \dots \dots, n\}$,where N_i denotes the collection of sub regions that, in a well defined sense, are neighbors of sub region i. This neighborhood system satisfies that for any I, $j = 1,2, \dots \dots, n, j \epsilon N_i$ if and only if $i \epsilon N_i$and $\notin N_i$, and it is key in determining the dependence structure of the SAR model.

An emblematic example commonly used in applications is the neighborhood system defined in terms of geographic adjacency

$N_i = \{j: \text{sub regions i and j share a boundary}\}, i = 1,2, \dots \dots, n.$

Other examples include neighborhood systems defined based on distance from the cancroids of sub regions or based on similarity of an auxiliary variable; see Cressie (1993, p. 554) and Case, Rosen and Hines (1993). This kind of specification is quite natural for modeling summary or aggregate data where similarity between sub regions often depends on similarity of shared features.

For each sub region it is observed the variable of interest, Y_i , and a set of p < n explanatory variables, $X_i = (x_{i1}, x_{i2}, \dots \dots, x_{ip})^T$. The SAR model for the responses, $Y = (Y_1, Y_2, \dots \dots, Y_p)^T$, is formulated by specifying a form of spatial

autoregression given by

$$Y_i = X^T\beta + \sum_{j=1}^{n} b_{ij} (Y_j - X_j^T\beta) + \varepsilon_i , \ i = 1,2, \dots, n. \quad (2.1)$$

Where $\beta = (\beta_1, \beta_2, \dots \dots, \beta_p)^T \epsilon R^p$ are unknown regression parameters, $\varepsilon_i \sim N(0, \sigma_i^2)$ are independent and b_{ij} and σ_i^2 are covariance parameters; let $B = (b_{ij})_{n\times n}$ and $= \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$. Provided $I_n - B$ is nonsingular, the n scalar equations in (2.1) can be written as

$$Y = X\beta + (I_n - B)^{-1}\varepsilon , \qquad (2.2)$$

Where X is the $n \times p$ matrix with i^{th} row X_i^T, assumed to have full rank, and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T$, so $Y \sim N_n\{X\beta,\ (I_n - B)^{-1}M(I_n - B^T)^{-1}\}$.

Now let us consider the simultaneous auto-regressive model for an $n \times 1$ dimensional vector y

$$Y = X\beta + (I_n - B)^{-1}\varepsilon, \qquad (2.3)$$

Where X is the design matrix of order an $n \times p\ (n > p)$; $\beta = (\beta_1, \beta_2, \ldots\ldots, \beta_p)^T$ is a vector of p auto-regressive parameters, and ε the errors vector associated with the responses vector y. Assume that each elements of ε is uncorrelated but not independent with others and has the same univariate Student-t distribution with location 0, scale $\sigma > 0$ and v degrees of freedom (d.f.). Thus the joint probability density function (p.d.f.) of ε is

$$\therefore f(\varepsilon) = (\sigma^2)^{-\frac{n}{2}} \prod_{i=1}^{n} \frac{1}{\sqrt{v}\, B\left(\frac{1}{2}, \frac{v}{2}\right)} \left[\frac{v^{\frac{v+n}{2}}}{[v+\sigma^{-2}\varepsilon^T\varepsilon]^{\frac{v+n}{2}}}\right] \qquad (2.4)$$

We obtained from the above equation, mean of ε and variance of ε are $E(\varepsilon) = 0$, a vector of 0's and $cov(\varepsilon) = (v-2)^{-1}v\sigma^2 I_n$ for $v > 0$ in which I_n in an identity matrix. Hence the probability density function of the realized response vector y is $y \sim t_n(X\beta, \sigma, v)$ with the probability density function of y is (Joint/Multivariate t-distribution) given by

$$\therefore\ f(y|\beta, \sigma^2) = \frac{(\sigma^2)^{-\frac{n}{2}}}{v^{\frac{n}{2}}\, \beta\left(\frac{v+n}{2}\right)} \frac{v^{\frac{v+n}{2}}}{[v+\sigma^{-2}(y-X\beta)^T(y-X\beta)]^{\frac{v+n}{2}}} \qquad (2.5)$$

Prior Distribution and Posterior Distribution:

Another class of prior distribution is called Non-information prior distributions. Which are called the priors of ignorance. Suppose that we are now in a situation where we have no definite (subject or object) prior information. To make use of Bayesian methods of inference we are compelled to express our prior knowledge in quantitative terms, we need a numerical specification of the state of prior ignorance. A common approach is to invoke the "Bayes'-Lapalce principle of insufficient reason" expressed by Jeffreys (1961) in the way: If there is a reason to believe on hypothesis rather than another, the probabilities are equal to say that the probabilities are equal is a precise way of saying that we have no ground for choosing between alternatives. Rather than a state of complete ignorance, the Non-informative prior refers to the case when relative little (or very limited) information is available a priori. It frequently means that there exists a set of parameter values that the experimenter believes to be frequently likely choices for the parameter, as described by the principle of insufficient reason. The goal here is to select a prior distribution that is locally uniform, that is a prior that approximately uniform distributed over the interval of interest. Now adopting the invariance theory (Jeffreys, 1961), the joint prior density of parameters can be written as

$$g(\beta, \sigma^2) \propto \sigma^{-2}. \qquad (2.6)$$

The posterior distribution of parameters for a set of sample observations is typically the major objective of the Bayesian statistical analysis. To obtain a posterior distribution using the Bayes's Theorem a prior distribution of unknown parameters is essential. From the equation (2.6) the joint posterior density of β and σ^2 for the realized responses y can be written as

$$\therefore\ f(\beta,\sigma^2|y) \propto \frac{(\sigma^2)^{-\frac{n+2}{2}}}{v^{\frac{n}{2}}\,\beta\left(\frac{v+n}{2}\right)} \frac{v^{\frac{v+n}{2}}}{[v+\sigma^{-2}(y-X\beta)^T(y-X\beta)]^{\frac{v+n}{2}}} \quad (2.7)$$

Inference about unknown parameters β and σ^2 from the above linear model has been considered in other studies (Zellner, 1976; Fraser and Ng, 1980). In this case, the study is concerned to derive the prediction distribution of future response(s) from the future model, conditional on the observed responses y from the realized model.

The Bayesian Prediction Rule:

If X^* be an unobserved future response from a future simultaneous auto-regressive model with the same simultaneous auto-regressive parameters and assumption of the realized model but with different design matrix, then under the Bayesian approach the prediction distribution of X^* given y can be obtained by solving the following integral

$$f(X^*|y) \propto \int_\beta \int_{\sigma^2>0} f(\beta,\sigma^2|y)\, f(X^*) d\sigma^2\, d\beta \quad (2.8)$$

where $f(\beta,\sigma^2|y)$ is the joint posterior density of unknown parameters β and σ^2 that is provided in (2.7) and $f(X^*)$ is a probability density of the future response X^* from the future model. This principle is appropriate when the future response X^* is independently distributed from the observed responses y that means X^* and y are not dependent to each other. However, in this study the responses from the realized as well as the future models are dependent but uncorrelated.

Prediction Distribution of a Set of Future Responses

Let y_f be a set of n_f future responses from the model in (2.3) corresponding to the $n_f \times p$ order design matrix X_f and $n_f \times 1$ dimensional errors vector ε_f . Thus the future model can be expressed as

$$y_f = X_f\beta + (I_n - B)^{-1}\varepsilon_f \quad (2.9)$$

where, $\varepsilon_f \sim t_{n_f}(0,\sigma,v)$, and hence $y_f \sim t_{n_f}(X_f\beta,\sigma,v)$.

It is noted that mean and variance of y_f are respectively $X_f\beta$ and $\sigma^2 = (I_n - B)^{-1}M(I_n - B^T)^{-1}$

According to the assumption, the observed errors vector ε and the unobserved future errors vector ε_f are uncorrelated but not independent then their respective observed responses y from the realized model and unobserved future responses y_f from the future model are also dependent but uncorrelated. Thus the combined joint (Bivariate t-distribution) probability density function (p.d.f.) of y and y_f is written by

$$f(y,y_f|\beta,\sigma^2) = \frac{\Gamma\left(\frac{v+n+n_f}{2}\right)(\sigma^2)^{-\frac{n+n_f}{2}}}{v^{\frac{n+n_f}{2}}\,\Gamma\left(\frac{v}{2}\right)\Gamma\left(\frac{n+n_f}{2}\right)} \frac{1}{\left[1+\frac{\sigma^{-2}(y-X\beta)^T(y-X\beta)+(y_f-X_f\beta)^T(y_f-X_f\beta)}{v}\right]^{\frac{v+n+n_f}{2}}}$$

(univariate n=1, bivariate n=2 and multivariate n=n here bivariate $n=n+n_f$)

$\therefore\ f(y, y_f|\beta, \sigma^2) \propto (\sigma^2)^{-\frac{n+n_f}{2}} [v + \sigma^{-2}\theta]^{-\frac{v+n+n_f}{2}}$. (2.10)

Where $\theta = (y - X\beta)^T(y - X\beta) + (y_f - X_f\beta)^T(y_f - X_f\beta)$ and v is the d.f. of the errors distribution.

Prior and Posterior Distribution:

Using the prior density of the equation (2.6) and the joint p.d.f. of the combined responses y_f and y in (2.10), the joint posterior density of β and σ^2 for the combined responses and can be easily obtained by the Bayes's Theorem, as

$f(\beta, \sigma^2|y, y_f) = g(\beta, \sigma^2) \times f(y, y_f|\beta, \sigma^2)$

$$\Rightarrow f(\beta, \sigma^2|y, y_f) \propto \frac{(\sigma^2)^{-\frac{n+n_f+2}{2}}}{v^{\frac{n+n_f}{2}}\beta\left(\frac{v}{2}, \frac{n+n_f}{2}\right)} \frac{1}{\left[1+\frac{\sigma^{-2}(y-X\beta)^T(y-X\beta)+(y_f-X_f\beta)^T(y_f-X_f\beta)}{v}\right]^{\frac{v+n+n_f}{2}}} \quad (2.11)$$

As y_f and y are not independent so they are not independently distributed. Thus the usual idea in (2.8) is not appropriate here. In this situation since the density function of a set of future responses from the future model is linked with the density function of the set of observed responses from the realized model within the combined joint p.d.f. in equation (2.10), the prediction distribution of can be obtained by solving the following integral

$$\therefore\ f(y_f|y) \propto \int_\beta \int_{\sigma^2>0} f(\beta, \sigma^2|y, y_f)\, d\sigma^2\, d\beta . \quad (2.12)$$

That means, in this case the prediction distribution of future responses can be derived from the joint posterior density function of the parameters for the combined responses y and y_f.

Now equation (2.11) can be expressed as the following form

$$f(\beta, \sigma^2|y, y_f) \propto \frac{(\sigma^2)^{-\frac{n+n_f+2}{2}}}{v^{\frac{n+n_f}{2}}\beta\left(\frac{v}{2}, \frac{n+n_f}{2}\right)} \frac{1}{\left[1+\frac{\sigma^{-2}(y-X\beta)^T(y-X\beta)+(y_f-X_f\beta)^T(y_f-X_f\beta)}{v}\right]^{\frac{v+n+n_f}{2}}}$$

$$\Rightarrow f(\beta, \sigma^2|y, y_f) = \frac{(\sigma^2)^{-\frac{n+n_f+2}{2}}}{v^{\frac{n+n_f}{2}}\beta\left(\frac{v}{2}, \frac{n+n_f}{2}\right)} \frac{1}{\left[1+\frac{\sigma^{-2}\theta}{v}\right]^{\frac{v+n+n_f}{2}}}$$

Where $\theta = (y - X\beta)^T(y - X\beta) + (y_f - X_f\beta)^T(y_f - X_f\beta)$

$$\Rightarrow f(\beta, \sigma^2|y, y_f) = \frac{(\sigma^2)^{-\frac{n+n_f+2}{2}}}{\beta\left(\frac{v}{2}, \frac{n+n_f}{2}\right)} \frac{v^{\frac{v}{2}}}{[v+\sigma^{-2}\theta]^{\frac{v+n+n_f}{2}}}$$

Let us assume that $C = \theta^{-1}v\sigma^2 \ \Rightarrow \sigma^2 = \frac{C\theta}{v} \ \Rightarrow \theta^{-1}\sigma^2 = \frac{v}{C}$ and putting this value in the above equation and then we can write

$$f(\beta,\sigma^2|y,y_f)\propto \frac{\left(\frac{C\theta}{v}\right)^{-\frac{n+n_f+2}{2}}}{\beta\left(\frac{v}{2},\frac{n+n_f}{2}\right)}\frac{v^{\frac{v}{2}}}{\left[v+\frac{v}{C}\right]^{\frac{v+n+n_f}{2}}}$$

$$\Rightarrow f(\beta,\sigma^2|y,y_f)=\frac{v\,.Q^{-\frac{n+n_f}{2}}}{\beta\left(\frac{v}{2},\frac{n+n_f}{2}\right)}\frac{\theta^{-1}.C^{\frac{v}{2}-1}}{[1+C]^{\frac{v+n+n_f}{2}}} \qquad (2.13)$$

Considering the transformation $F=\theta^{-1}(n+n_f)\sigma^2$ and $C=\theta^{-1}v\sigma^2$

Now putting this value in (2.13)

$$f(\beta,\sigma^2|y,y_f)=\frac{v\,.\theta^{-\frac{n+n_f}{2}}}{\beta\left(\frac{v}{2},\frac{n+n_f}{2}\right)}\frac{\theta^{-1}.C^{\frac{v}{2}-1}}{[1+C]^{\frac{v+n+n_f}{2}}}$$

$$\Rightarrow f(\beta,\sigma^2|y,y_f)\propto \theta^{-\frac{n+n_f}{2}-1}\frac{\left(\frac{v}{n+n_f}\right)^{\frac{v}{2}}}{\beta\left(\frac{v}{2},\frac{n+n_f}{2}\right)}\frac{F^{\frac{v}{2}-1}}{\left[1+\left(\frac{v}{n+n_f}F\right)\right]^{\frac{v+n+n_f}{2}}} \qquad (2.14)$$

And then after using the results, equation (2.12) can be written as

$$f(y_f|y)\propto \int_\beta\int_F \theta^{-\frac{n+n_f}{2}-1}\frac{\left(\frac{v}{n+n_f}\right)^{\frac{v}{2}}}{\beta\left(\frac{v}{2},\frac{n+n_f}{2}\right)}\frac{F^{\frac{v}{2}-1}}{\left[1+\left(\frac{v}{n+n_f}F\right)\right]^{\frac{v+n+n_f}{2}}}\,d\sigma^2\,d\beta \qquad (2.15)$$

It is clear that F has an F-distribution with v and $n+n_f$ degrees of freedom that is, $\sim F_{v,n+n_f}$. Employing that the F integral in (2.15) to integrating over F, the prediction density of future responses becomes,

$$\therefore \int_F \frac{\left(\frac{v}{n+n_f}\right)^{\frac{v}{2}}}{\beta\left(\frac{v}{2},\frac{n+n_f}{2}\right)}\frac{F^{\frac{v}{2}-1}}{\left[1+\left(\frac{v}{n+n_f}F\right)\right]^{\frac{v+n+n_f}{2}}}dF=\int_0^\infty \frac{\left(\frac{v}{n+n_f}\right)^{\frac{v}{2}}}{\beta\left(\frac{v}{2},\frac{n+n_f}{2}\right)}\frac{F^{\frac{v}{2}-1}}{\left[1+\left(\frac{v}{n+n_f}F\right)\right]^{\frac{v+n+n_f}{2}}}dF$$

To evaluate the integral, Let $\frac{v}{n+n_f}F=Q\Rightarrow F=\frac{n+n_f}{v}Q$ so that $dF=\frac{n+n_f}{v}dQ$

When F=0 then Q=0 and when $F=\infty$ *then* $Q=\infty$

$$=\frac{1}{\beta\left(\frac{v}{2},\frac{n+n_f}{2}\right)}\beta\left(\frac{v}{2},\frac{n+n_f}{2}\right) \quad \text{[Using Beta distribution of Second kind]}$$

$$\therefore \int_F \frac{\left(\frac{v}{n+n_f}\right)^{\frac{v}{2}}}{\beta\left(\frac{v}{2},\frac{n+n_f}{2}\right)}\frac{F^{\frac{v}{2}-1}}{\left[1+\left(\frac{v}{n+n_f}F\right)\right]^{\frac{v+n+n_f}{2}}}dF=1\ (unity)$$

Using the results in equation (2.15). Then we can write

$$f(y_f|y)\propto \int_\beta \theta^{-\frac{n+n_f}{2}-1}d\beta \qquad (2.16)$$

Now $\theta = (y - X\beta)^T(y - X\beta) + (y_f - X_f\beta)^T(y_f - X_f\beta)$ can be expressed as the following form $\theta = A + (\beta - P)^T W(\beta - P)$.

where $A = y^T y + y_f^T y_f - (y^T X^T + y_f^T X_f^T)W^{-1}(Xy + X_f y_f)$ is free from the parameters' vector β , $W = X^T X + X_f^T X_f$ and $P = W^{-1}(Xy + X_f y_f)$.

Putting this value in (2.16) and integrating over β by using the multivariate Student-t integral, it is easy to obtain the following probability density function of y_f given y and then hence the prediction distribution of a set of future responses y_f, conditional on a set of realized responses y is obtained as

$$f \Rightarrow f(y_f|y) = \frac{\Gamma\left(\frac{n-k+n_f}{2}\right)\left|I_{n_f} - X_f^T W^{-1} X_f\right|^{1/2}}{\Gamma\left(\frac{n-k}{2}\right)\left[S^2(n-k)\pi^{n_f}\right]^{1/2}} \times \left[1 + \frac{(y_f - X_f^T\hat{\beta})^T \Sigma (y_f - X_f^T\hat{\beta})}{n-k}\right]^{-\frac{n-k+n_f}{2}} \quad (2.17)$$

Hence the prediction distribution of a set of future responses y_f, conditional on a set of realized responses y, is obtained as

$f(y_f|y) = \Psi \times \left[1 + \frac{(y_f - X_f^T\hat{\beta})^T \Sigma (y_f - X_f^T\hat{\beta})}{n-k}\right]^{-\frac{n-k+n_f}{2}}$,and the normalizing constant of the prediction distribution is given by

$$\Phi_f = \frac{\Gamma\left(\frac{n-k+n_f}{2}\right)\left|I_{n_f} - X_f^T M^{-1} X_f\right|^{\frac{1}{2}}}{\Gamma\left(\frac{n-k}{2}\right)\left[\pi^{n_f}(n-k)S^2\right]^{\frac{1}{2}}}$$

where $\Sigma = \left[S^{-2}\left(I_{n_f} - X_f^T W^{-1} X_f\right)\right]^{-\frac{1}{2}}$, $\hat{\beta} = (XX^T)^{-1}Xy$ is the OLS estimator of the simultaneous auto-regressive vector β **,**

$$S^2 = (n-k)^{-1}\left[(y - X^T\hat{\beta})^T(y - X^T\hat{\beta})\right]$$
$$= (n-k)^{-1} y^T[I_n - X^T(XX^T)^{-1}X]y,$$

Here, it is clear thaty_f , the vector of a set of future responses, has an n_f-dimensional multivariate Student-t distribution with the location $X_f^T\hat{\beta}$, scale $\left[S^{-2}\left(I_{n_f} - X_f^T W^{-1} X_f\right)\right]^{-\frac{1}{2}}$, and the shape parameter n - k . This result is identical with the results obtained for the same model Rahman and Khan (2007), and for the multiple regression model with independent and normal errors by Zellner (1971) and Geisser (1993) among others. Therefore, it is noted that the prediction distribution is unaffected by departures from the model with independent and normal errors to multivariate Student-t errors distribution.

Prediction Distribution of a Single Future Response:

For $n_f = 1$ the set of future responses vector y_f becomes a single future response, and hence if denotes a single future response, then the future simultaneous auto-regressive model in (2.9) becomes the following form

$$y_f = X_f\beta + (I_n - B)^{-1}\varepsilon_f \quad (2.18)$$

WhereX_f , is a $1 \times p$ order design vector, β is the same simultaneous auto-regressive coefficients vector of order $p \times 1$ and ε_f is the error term associated with y_f and ε_f has a univariate Student-t distribution as $\varepsilon_f \sim t_1(0, \sigma, v)$. By the same operations as used in previous section for the derivation of prediction distribution of a set of future responses, it can be easily obtained the joint posterior density of parameters for a single response y_f and the realized responses vector y under the same prior distribution $g(\beta, \sigma^2) \propto \sigma^{-2}$ and can be easily obtained by the Bayes's Theorem, as

$$f(\beta, \sigma^2 | y, y_f) = g(\beta, \sigma^2) \times f(y, y_f | \beta, \sigma^2)$$

$$\Rightarrow f(\beta, \sigma^2 | y, y_f) \propto \frac{(\sigma^2)^{-\frac{n+3}{2}}}{v^{\frac{n+1}{2}} \beta\left(\frac{v}{2}, \frac{n+1}{2}\right)} \frac{1}{\left[1 + \frac{\sigma^{-2}(y - X\beta)^T(y - X\beta) + \left(y_f - X_f\beta\right)^T\left(y_f - X_f\beta\right)}{v}\right]^{\frac{v+n+1}{2}}} \quad (2.19)$$

The prediction distribution of a single future response y_f can be derived from the joint posterior density function for the combined responses y_f and y in (2.19) by integrating over the parameters β and σ^2 . For completing the derivation of prediction distribution of a single future response, the same operational steps are used as considered in the previous section. At first the joint posterior density can be expressed as its convenient form as like in equation

$$f(\beta, \sigma^2 | y, y_f) = \frac{v . Q^{-\frac{n+n_f}{2}}}{\beta\left(\frac{v}{2}, \frac{n+n_f}{2}\right)} \frac{\theta^{-1} . C^{\frac{v}{2}-1}}{[1+C]^{\frac{v+n+n_f}{2}}},$$

And then using an appropriate transformation

$$F = \theta^{-1}(n + n_f)\sigma^2 \text{ and} C = \theta^{-1} v \sigma^2,$$

The parameter σ^2 can be eliminated by the $F_{v,n+1}$ integral. After that θ , can be expressed as a quadratic form of β and then the properties of the multivariate Student-t distribution can be used to complete the integration over β . Finally, the prediction distribution of a single future response y_f, conditional on a set of realized responses y, is obtained as

$$f(y_f | y) = \frac{\Gamma\left(\frac{n-k+1}{2}\right) \left|I_{n_f} - X_f^T W^{-1} X_f\right|^{1/2}}{\Gamma\left(\frac{n-k}{2}\right) [S^2(n-k)\pi]^{1/2}} \times \left[1 + \frac{\left(y_f - X_f^T \hat{\beta}\right)^T \Sigma \left(y_f - X_f^T \hat{\beta}\right)}{n-k}\right]^{-\frac{n-k+1}{2}} \quad (2.20)$$

Hence the prediction distribution of a set of future responses y_f, conditional on a set of realized responses y, is obtained as

$$f(y_f|y) = \Psi \times \left[1 + \frac{(y_f - X_f^T\hat{\beta})^T \Sigma (y_f - X_f^T\hat{\beta})}{n-k}\right]^{-\frac{n-k+n_f}{2}}$$

,and the normalizing constant of the prediction distribution is given by

$$\Phi_f = \frac{\Gamma\left(\frac{n-k+1}{2}\right)\left|I_{n_f} - X_f^T M^{-1} X_f\right|^{\frac{1}{2}}}{\Gamma\left(\frac{n-k}{2}\right)\left[\pi(n-k)S^2\right]^{\frac{1}{2}}}$$

where $\sum = \left[S^{-2}\left(1 - X_f^T W^{-1} X_f\right)\right]^{-\frac{1}{2}}$, $\hat{\beta} = (XX^T)^{-1}Xy$ is the OLS estimator of the simultaneous auto-regressive vector β ,

$$S^2 = (n-k)^{-1}\left[\left(y - X^T\hat{\beta}\right)^T\left(y - X^T\hat{\beta}\right)\right]$$
$$= (n-k)^{-1} y^T [I_n - X^T(XX^T)^{-1}X] y,$$

Thus, the prediction distribution of a single future response for the simultaneous auto-regressive model with multivariate Student-t error terms is a univariate Student-t distribution with appropriate parameters.

Conclusion:

The prediction distribution of future response(s), conditional on a set of observed responses has been derived for the simultaneous auto-regressive model having multivariate Student-t errors by the improper Bayesian method. Results reveal that the prediction distribution of a single future response and a set of future responses are a univariate Student-t distribution and a multivariate Student-t distribution respectively. It has been shown that the prediction distributions for the simultaneous auto-regressive model remains identical by a change in the error distribution from normal to multivariate Student-t distribution.

Furthermore, 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. However, the shape parameter of the prediction distribution does not depend on the d.f. of the errors distribution.

References:

1. Bhattacharjss S.K. (1987). Influence of variables in Bayesian prediction, unpublished Ph.D. Thesis, University of Sheffield, U.K.
2. Bhattacharjss S.K. and Dunsmore, I.R. (1991), The Influence of variables in a logistic model, Biometrika, Vol. 78, pp.851-856.
3. Bhattacharjss S.K. and Dunsmore, I.R. (1991). The Influence of variables in a normal regression model, accepted in JIOS
4. Bhattacharjss S.K. and M. A. Idris (2005). The prediction Analysis of a second-order auto-regressive model, Unpublished M. Sc, Thesis, University of Rajshahi.
5. Aitchison, J and Dunsmore, I R (1975). Statistical prediction analysis. Cambridge University Press, Cambridge.
6. Berger, J.O., De Oliveira, V. and Sans_o, B. (2001). Objective Bayesian Analysis of Spatially Correlated Data. J. Amer. Statist. Assoc., 96, 1361-1374.
7. Fraser, D A S and Haq, M S (1970). Inference and prediction for multilinear model. Journal of Statistical Research, University of Dhaka, Dhaka 4: 93-109.
8. Bernardo, J.M (1979). Reference Posterior Distributions for Bayes Inference (with discussion). J Roy. Statist. Soc. Ser. B, 41, 113-147.
9. Fraser, D A S and Ng, K W (1980). Multivariate regression analysis with spherical error. In V.P.R. Krishnaiah, (ed.): Multivariate Analysis, North-Holland Publishing Co., New York, 369-386.
10. Geisser, S (1993). Predictive inference: An introduction. Chapman and Hall, London.
11. Haq, M S (1982). Structural relation and prediction for multivariate models. Statistische Hefte, 23: 218-228.
12. Haq, M S and Khan, S (1990). Prediction distribution for a linear regression model with multivariateStudent-t error distribution. Communications in Statistics -Theory and Methods, 19(12): 4705-4712.
13. LeSage, J.P. and Parent, O. (2007). Bayesian Model Averaging for Spatial Econo-metric Models. Geographical Analysis, 39, 241-267.
14. Jeffreys, H (1961). Theory of probability. 3rd Edn., Oxford University Press, Oxford.
15. 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): 2423-2441.
16. Sutradhar, B C and Ali, M M (1989). A generalization of the Wishart distribution for the elliptical model and its moments for the multivariate-t model. Journal of Maltivariate Analysis, 29: 155-162.
17. Zellner, A (1971). An introduction to Bayesian inference in econometrics. John Wiley and Sons, NY.
18. 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(354): 400-405.