Proceedings of The Third Conferences On
Research And Education In Mathematics 2007

PREDICTION DISTRIBUTION FOR LINEAR REGRESSION MODEL WITH MULTIVARIATE STUDENT-*T* ERRORS UNDER THE BAYESIAN APPROACH

Azizur Rahman, Shahjahan Khan
University of Southern Queensland, Australia
rahmana@usq.edu.au, khan@usq.edu.au

Abstract:

Prediction distribution is a basis for predictive inferences applied in many real world situations. It is a distribution of the unobserved future response(s) conditional on a set of realized responses from an informative experiment. Various statistical approaches can be used to obtain prediction distributions for different models. This study derives the prediction distribution(s) for multiple linear regression model using the Bayesian method when the error components of both the performed and future models have a multivariate Student-t distribution. The study observes that the prediction distribution(s) of future response(s) has a multivariate Student-t distribution whose degrees of freedom depends on the size of the realized sample and the dimension of the regression parameters' vector but does not depend on the degrees of freedom of the errors distribution.

Introduction

The prediction distribution of future response(s) can be derived from the regression 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 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 multiple regression 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.

The Multiple Regression model

Consider the linear regression model for an $n\times 1$ dimensional responses vector $\boldsymbol{y}$

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e} \tag{1}$$

where $\boldsymbol{X}$ is the design matrix of order an $n\times k$ $(n>k)$; $\beta=(\beta_0,\beta_1,\ldots,\beta_{k-1})'$ is a vector of k regression parameters; and $\boldsymbol{e}$ is the errors vector associated with the responses vector $\boldsymbol{y}$. Assume that each elements of $\boldsymbol{e}$ is uncorrelated but not independent with others and has the same univariate Student-*t* distribution with location 0, scale $\sigma>0$ and ν degrees of freedom (d.f.). Thus the joint probability density function (p.d.f.) of $\boldsymbol{e}$ is

$$f(\boldsymbol{e}) \propto (\sigma^2)^{-\frac{n}{2}}[\nu+\sigma^{-2}\boldsymbol{e}'\boldsymbol{e}]^{-\frac{\nu+n}{2}} \tag{2}$$

It is noted that $E(\boldsymbol{e})$ = $\mathbf{0}$, a vector of 0's and $Cov(\boldsymbol{e})$ = $(\nu-2)^{-1}\nu\sigma^2 I_n$ for $\nu>2$ in which I_n is an identity matrix . Hence the p.d.f. of the realized responses vector $\boldsymbol{y}$ is $\boldsymbol{y}\sim t_n(\boldsymbol{X}\beta,\ \sigma,\nu)$ with density

$$f(\boldsymbol{y}|\beta,\sigma^2) \propto (\sigma^2)^{-\frac{n}{2}}[\nu+\sigma^{-2}(\boldsymbol{y}-\boldsymbol{X}\beta)^{/}(\boldsymbol{y}-\boldsymbol{X}\beta)]^{-\frac{\nu+n}{2}} \quad (3)$$

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. Adopting the invariance theory (Jeffreys, 1961), the joint uniform prior density of parameters can be written as $g(\beta,\sigma^2) \propto \sigma^{-2}$. Under this uniform prior the joint posterior density of β and σ^2 for the realized responses $\boldsymbol{y}$ can be obtained as

$$f(\beta,\sigma^2|\boldsymbol{y}) \propto (\sigma^2)^{-\frac{n+2}{2}}[\nu+\sigma^{-2}(\boldsymbol{y}-\boldsymbol{X}\beta)^{/}(\boldsymbol{y}-\boldsymbol{X}\beta)]^{-\frac{\nu+n}{2}} \quad (4)$$

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 $\boldsymbol{y}$ from the realized model.

The Bayesian Prediction rule

If z* be an unobserved future response from a future regression model with the same regression parameters and assumption of the realized model but with different design matrix, then under the Bayesian approach the prediction distribution of z* given $\boldsymbol{y}$ can be obtained by solving the following integral

$$f(z^*|\boldsymbol{y}) \propto \int_{\beta}\int_{\sigma^2>0} f(\beta,\sigma^2|\boldsymbol{y})\, f(z^*)\, d\sigma^2\, d\beta \quad (5)$$

where $f(\beta,\sigma^2|\boldsymbol{y})$ is the joint posterior density of unknown parameters β and σ^2 that is provided in (4) and $f(z^*)$ is a probability density of the future response z* from the future model. This principle is appropriate when the future response z* is independently distributed from the observed responses $\boldsymbol{y}$, that means z* and $\boldsymbol{y}$ 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 $\boldsymbol{y}_f$ be a set of n_f future responses from the model in (1) corresponding to the $n_f \times k$ order design matrix $\boldsymbol{X}_f$ and $n_f \times 1$ dimensional errors vector $\boldsymbol{e}_f$. Thus the future model can be expressed as

$$\boldsymbol{y}_f = \boldsymbol{X}_f\beta + \boldsymbol{e}_f \quad (6)$$

where, $\boldsymbol{e}_f \sim t_{n_f}(\boldsymbol{0}, \sigma,\nu)$, and hence $\boldsymbol{y}_f \sim t_{n_f}(\boldsymbol{X}_f\beta, \sigma,\nu)$.

According to the assumption, the observed errors vector $\boldsymbol{e}$ and the unobserved future errors vector $\boldsymbol{e}_f$ are uncorrelated but not independent then their respective observed responses $\boldsymbol{y}$ from the realized model and unobserved future responses $\boldsymbol{y}_f$ from the future model are also dependent but uncorrelated. Thus the combined joint p.d.f. of $\boldsymbol{y}$ and $\boldsymbol{y}_f$ is

$$f(\boldsymbol{y},\boldsymbol{y}_f|\beta,\sigma^2) \propto (\sigma^2)^{-\frac{n+n_f}{2}}[\nu+\sigma^{-2}Q]^{-\frac{\nu+n+n_f}{2}} \quad (7)$$

where $Q = (\boldsymbol{y}-\boldsymbol{X}\beta)^{/}(\boldsymbol{y}-\boldsymbol{X}\beta)+(\boldsymbol{y}_f-\boldsymbol{X}_f\beta)^{/}(\boldsymbol{y}_f-\boldsymbol{X}_f\beta)$ and ν is the d.f. of the errors distribution.

Using the prior density $g(\beta,\sigma^2) \propto \sigma^{-2}$ and the joint p.d.f. of the combined responses $\boldsymbol{y}_f$ and $\boldsymbol{y}$ in (7), the joint posterior density of β and σ^2 for the combined responses $\boldsymbol{y}_f$ and $\boldsymbol{y}$ can be easily obtained by the Bayes's Theorem, as

$$f(\beta,\sigma^2\,|\,\boldsymbol{y},\boldsymbol{y}_f) \propto (\sigma^2)^{-\frac{n+n_f+2}{2}}\,[\nu+\sigma^{-2}\,\mathrm{Q}]^{-\frac{\nu+n+n_f}{2}} \qquad (8)$$

As $\boldsymbol{y}_f$ and $\boldsymbol{y}$ are not independent so they are not independently distributed. Thus the usual idea in (5) is not appropriate here. In this situation since the density function of a set of future responses $\boldsymbol{y}_f$ from the future model is linked with the density function of the set of observed responses $\boldsymbol{y}$ from the realized model within the combined joint p.d.f. in equation (7), the prediction distribution of $\boldsymbol{y}_f$ can be obtained by solving the following integral

$$f(\mathrm{y}_f\,|\,\boldsymbol{y}) \propto \int_\beta\int_{\sigma^2>0} f(\boldsymbol{y},\boldsymbol{y}_f\,|\,\beta,\sigma^2)\; g(\beta,\sigma^2)\; d\sigma^2\; d\beta$$
$$\propto \int_\beta\int_{\sigma^2>0} f(\beta,\sigma^2\,|\,\boldsymbol{y},\boldsymbol{y}_f)\;\; d\sigma^2\; d\beta \qquad (9)$$

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 $\boldsymbol{y}$ and $\boldsymbol{y}_f$ **.**

Now equation (8) can be expressed as the following convenient form

$$f(\beta,\sigma^2\,|\,\boldsymbol{y},\boldsymbol{y}_f) \propto (Q)^{-\frac{n+n_f}{2}}\;\frac{Q^{-1}C^{\frac{\nu}{2}-1}}{[1+C]^{-\frac{\nu+n+n_f}{2}}} \qquad (10)$$

where $C = Q^{-1}\nu\sigma^2$. Considering the transformation $r = Q^{-1}(n+n_f)\sigma^2$ in (10) and then after using the results, equation (9) can be written as

$$f(\mathrm{y}_f\,|\,\boldsymbol{y}) \propto \int_\beta\int_r (Q)^{-\frac{n+n_f}{2}}\;\frac{r^{\frac{\nu}{2}-1}}{\left[1+\dfrac{\nu}{n+n_f}\,r\right]^{-\frac{\nu+n+n_f}{2}}}\quad dr\;d\beta \qquad (11)$$

It is clear that r has an F distribution with ν and $n+n_f$ degrees of freedom that is, $r \sim F_{\nu,n+n_f}$. Employing the F integral in (11) to integrating over r, the prediction density of future responses becomes

$$f(\mathrm{y}_f\,|\,\boldsymbol{y}) \propto \int_\beta (Q)^{-\frac{n+n_f}{2}}\; d\beta \qquad (12)$$

Now $Q = (\boldsymbol{y} - \boldsymbol{X}\beta)^{/}(\boldsymbol{y} - \boldsymbol{X}\beta) + (\boldsymbol{y}_f - \boldsymbol{X}_f\,\beta)^{/}(\boldsymbol{y}_f - \boldsymbol{X}_f\,\beta)$ can be expressed as the following form

$$Q = A + (\beta\text{-}P)^{/}\,M\,(\beta\text{-}P) \qquad (13)$$

where $\mathrm{A} = \boldsymbol{y}^{/}\boldsymbol{y} + \boldsymbol{y}_f^{/}\,\boldsymbol{y}_f - (\boldsymbol{y}^{/}\boldsymbol{X} + \boldsymbol{y}_f^{/}\,\boldsymbol{X}_f)\,M^{-1}(\boldsymbol{X}^{/}\boldsymbol{y} + \boldsymbol{X}_f^{/}\,\boldsymbol{y}_f)$ is free from the parameters' vector β, $M = \boldsymbol{X}^{/}\boldsymbol{X} + \boldsymbol{X}_f^{/}\boldsymbol{X}_f$ and $P = M^{-1}(\boldsymbol{X}^{/}\boldsymbol{y} + \boldsymbol{X}_f^{/}\,\boldsymbol{y}_f)$.

Using the expression of Q in (13) to equation (12) and then integrating over β by using the multivariate Student-*t* integral, it is easy to obtain the following p.d.f. of $\boldsymbol{y}_f$ given $\boldsymbol{y}$

$$f(\boldsymbol{y}_f\,|\,\boldsymbol{y}) \propto [\boldsymbol{y}^{/}\boldsymbol{y} + \boldsymbol{y}_f^{/}\,\boldsymbol{y}_f - (\boldsymbol{y}^{/}\boldsymbol{X} + \boldsymbol{y}_f^{/}\,\boldsymbol{X}_f)\,M^{-1}(\boldsymbol{X}^{/}\boldsymbol{y} + \boldsymbol{X}_f^{/}\,\boldsymbol{y}_f)]^{-\frac{n-k+n_f}{2}}$$

Hence the prediction distribution of a set of future responses $\boldsymbol{y}_f$, conditional on a set of realized responses $\boldsymbol{y}$, is obtained as

$$f(\boldsymbol{y}_f \mid \boldsymbol{y}) = \Phi_f \; [(n-k) + (\boldsymbol{y}_f - \boldsymbol{X}_f \hat{\beta})^{/} H (\boldsymbol{y}_f - \boldsymbol{X}_f \hat{\beta})]^{-\frac{n-k+n_f}{2}} \qquad (14)$$

where $H = s^{-2}(I_{n_f} - \boldsymbol{X}_f M^{-1} \boldsymbol{X}_f^{/})$, $\hat{\beta} = (\boldsymbol{X}^{/}\boldsymbol{X})^{-1}\boldsymbol{X}^{/}\boldsymbol{y}$ is the OLS estimator of the regression vector β, $s^2 = (n-k)^{-1}[(\boldsymbol{y} - \boldsymbol{X}\hat{\beta})^{/}(\boldsymbol{y} - \boldsymbol{X}\hat{\beta})] = \boldsymbol{y}^{/}[\, I_n - \boldsymbol{X}(\boldsymbol{X}^{/}\boldsymbol{X})^{-1}\boldsymbol{X}^{/}\,]\boldsymbol{y}$, and the normalizing constant of the prediction distribution is given by $\Phi_f = \dfrac{\Gamma\left(\frac{n-k+n_f}{2}\right)\left|I_{n_f} - X_f M^{-1} X_f^{/}\right|^{\frac{1}{2}}}{\Gamma\left(\frac{n-k}{2}\right)\left[\pi^{n_f}(n-k)s^2\right]^{\frac{1}{2}}}$.

Here, it is clear that $\boldsymbol{y}_f$, the vector of a set of future responses, has an n_f-dimensional multivariate Student-*t* distribution with the location $\boldsymbol{X}_f \hat{\beta}$, scale $[s^{-2}(I_{n_f} - \boldsymbol{X}_f M^{-1}\boldsymbol{X}_f^{/})]^{-\frac{1}{2}}$, and the shape parameter $n-k$. This result is identical with the results obtained for the same model by Haq and Khan (1990), 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 indepedent 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 $\boldsymbol{y}_f$ becomes a single future response, and hence if y_f denotes a single future response, then the future regression model in (6) becomes the following form

$$y_f = \boldsymbol{x}_f \beta + e_f \qquad (15)$$

where, $\boldsymbol{x}_f$ is a $1 \times k$ order design vector, β is the same regression coefficients vector of order $k \times 1$ and e_f is the error term associated with y_f and e_f has a univariate Student-*t* distribution as $e_f \sim t_1\,(0, \sigma, \nu)$. 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 $\boldsymbol{y}$ under the same prior distribution $g(\beta, \sigma^2) \propto \sigma^{-2}$, as

$$f(\beta, \sigma^2 \mid \boldsymbol{y}, y_f) \propto (\sigma^2)^{-\frac{n+3}{2}} [\nu + \sigma^{-2} Q]^{-\frac{\nu+n+1}{2}} \qquad (16)$$

where $Q = (\boldsymbol{y} - \boldsymbol{X}\beta)^{/}(\boldsymbol{y} - \boldsymbol{X}\beta) + (y_f - \boldsymbol{x}_f \beta)^2$ in which y_f is scalar.

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 $\boldsymbol{y}$ in (16) 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 is used as considered in the previous section. At first the joint posterior density can be expressed as its convenient form as like in equation (10), and then using an appropriate transformation $r = Q^{-1}(n+1)\sigma^2$, the parameter σ^2 can be eliminated by the $F_{\nu,n+1}$ integral. After that, Q 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 $\boldsymbol{y}$, is obtained as

$$f(y_f \mid \boldsymbol{y}) = \Psi_f \left[(n - k) + (y_f - \boldsymbol{x}_f \hat{\beta})^{/} H (y_f - \boldsymbol{x}_f \hat{\beta})\right]^{-\frac{n-k+1}{2}} \quad (17)$$

where $H = s^{-2}(1 - \boldsymbol{x}_f M^{-1} \boldsymbol{x}_f^{/})$ in which $M = \boldsymbol{X}^{/}\boldsymbol{X} + \boldsymbol{x}_f^{/}\boldsymbol{x}_f$, $\hat{\beta} = (\boldsymbol{X}^{/}\boldsymbol{X})^{-1}\boldsymbol{X}^{/}\boldsymbol{y}$ is the OLS estimator of the regression vector β, $s^2 = (n-k)^{-1}[(\boldsymbol{y} - \boldsymbol{X}\hat{\beta})^{/}(\boldsymbol{y} - \boldsymbol{X}\hat{\beta})] = \boldsymbol{y}^{/}[I_n - \boldsymbol{X}(\boldsymbol{X}^{/}\boldsymbol{X})^{-1}\boldsymbol{X}^{/}]\boldsymbol{y}$, and Ψ_f represents the normalizing constant which is given by $\Psi_f = \dfrac{\Gamma\left(\frac{n-k+1}{2}\right)\left|1 - x_f M^{-1} x_f^{/}\right|^{\frac{1}{2}}}{\Gamma\left(\frac{n-k}{2}\right)\left[\pi (n-k) s^2\right]^{\frac{1}{2}}}$.

Thus, the prediction distribution of a single future response for the multiple regression 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 multiple linear regression 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 multiple regression 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:

Aitchison, J and Dunsmore, I R (1975). *Statistical prediction analysis*. Cambridge University Press, Cambridge.
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.
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.
Geisser, S (1993). *Predictive inference: An introduction*. Chapman and Hall, London.
Haq, M S (1982). Structural relation and prediction for multivariate models. *Statistische Hefte*, 23: 218-228.
Haq, M S and Khan, S (1990). Prediction distribution for a linear regression model with multivariate Student-*t* error distribution. *Communications in Statistics -Theory and Methods*, 19(12): 4705-4712.
Jeffreys, H (1961). *Theory of probability*. 3rd *Edn*., Oxford University Press, Oxford.
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.
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.
Zellner, A (1971). *An introduction to Bayesian inference in econometrics.* John Wiley and Sons, NY.
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.