Economics Letters 24 (1987) 39–44
North-Holland

BAYES PREDICTION IN THE LINEAR MODEL WITH SPHERICALLY SYMMETRIC ERRORS

S. Rao JAMMALAMADAKA
University of California, Santa Barbara, CA 93106, USA

Ram C. TIWARI
University of North Carolina, Charlotte, NC 28223, USA

Siddhartha CHIB
University of Missouri, Columbia, MO 65201, USA

Received 30 December 1986
Accepted 26 March 1987

This paper is concerned with Bayes prediction in a linear regression model when the density of the observations is given by

$$f(y\,|\,\beta, \tau^2) = \int_{z>0} (2\pi)^{-n/2}(\tau^2)^{n/2}\left\{\psi(z)^{-2}\right\}^{n/2} \exp\left(-\frac{\tau^2}{2}\psi(z)^{-2}\|y - X\beta\|^2\right) \mathrm{d}G(z),$$

where $y \in R^n$, $\beta \in R^k$, $\tau^2 > 0$, Z is a positive random variable with distribution function G, $\psi(.)$ is a positive function, and $\|.\|$ denotes the Euclidean norm. We show that when prior information is objective or in the conjugate family, the Bayes prediction density is the same as that when the density of the observations is normal, *for any Z*.

1. Introduction

This paper is concerned with the consequences on Bayes predictive inferences in regression models when the error terms are spherically distributed. The consequences on parameter inferences under this assumption have been studied by several authors, e.g., Strawderman (1974), Berger (1975), Judge et al. (1985), Ullah and Zande-Walsh (1985), West (1984) and Zellner (1976). The conclusion from these papers that is most relevant for our study is reached in Zellner (1976) where it is shown that the joint parameter posterior, with objective prior information, and multivariate-t errors differs from that obtained with multivariate normal errors. In this paper we show that predictive inferences, surprisingly, are completely unaffected by departures of the normality assumption in the direction of the spherically symmetric family, thus generalizing the result in Chib et al. (1987a) obtained for the multivariate-t case. This robustness result for Bayes prediction holds for a wide class of densities since the spherical family includes the density used by Zellner – the multivariate-t – as well as others like the multivariate normal, ϵ-contaminated normal and multivariate exponential.

2. The model

Consider the linear models with non-random regressors

$$\begin{aligned} y_1 &= X_1\beta + \epsilon_1, \\ y_2 &= X_2\beta + \epsilon_2, \end{aligned} \tag{1}$$

and assume that (y_1, X_1, X_2) is observed, and interest centers on the (as yet) unobserved vector y_2. The dimensions of y_i, X_i, ϵ_i are $n_i \times 1$, $n_i \times k$, and $n_i \times 1$, respectively, $i = 1, 2$. The unknown coefficient vector β is $k \times 1$. We let $y' = (y_1', y_2')$ be the combined vector of dependent variables and similarly for the independent variables, let $X' = (X_1', X_2')$. Also $\epsilon' = (\epsilon_1', \epsilon_2')$ denotes the $n \times 1$ error vector, $n = n_1 + n_2$. Assume that the error term ϵ has a spherical distribution which implies that the distribution of the observations y is spherical with density given by

$$f(y \mid \beta, \tau^2) = \int_0^\infty (2\pi)^{-n/2}(\tau^2)^{n/2}\{\psi(z)^{-2}\}^{n/2} \exp\left(-\frac{\tau^2}{2}\psi(z)^{-2}\|y - X\beta\|^2\right) dG(z), \tag{2}$$

where Z is a positive random variable with distribution function G, $\psi(.)$ is a positive measurable function, $1/\tau^2$ is a scale parameter of the dispersion matrix, and $\|.\|$ denotes the Euclidean norm. Alternatively, if we let u denote a random vector with a multivariate normal density having mean vector 0 and covariance matrix $(\tau^2)^{-1}I_n$, then y is equal in distribution to $u \cdot \psi(Z) + X\beta$. Thus, conditionally on Z, y is normal with parameters $\mathrm{E}y \mid (Z, \beta, \tau^2) = X\beta$, and $\mathrm{E}(y - X\beta)(y - X\beta)' \mid (Z, \beta, \tau^2) = (\tau^2)^{-1}\psi(z)^2 I_n$, which is the integrand in (2). Consequently, integrating over the distribution of Z leads to the marginal distribution of y, as in (2) above.

Properties of the spherical distribution are developed in Kelker (1970), and summarized in Chmielewski (1981). We mention in passing that if Z is chi-squared with ν degrees of freedom distributed independently of u, and $\psi(Z) = (Z/\nu)^{-1/2}$, $\nu > 0$, then y has the multivariate-t distribution with density given by

$$f(y \mid \beta, \tau^2, \nu) = c \cdot (\tau^2)^{n/2}[1 + \tau^2/\nu \|y - X\beta\|^2]^{-(n+\nu)/2}, \qquad y \in R^n, \tag{3}$$

where $c = \Gamma((n + \nu)/2)/(\Gamma(\nu/2)(\pi)^{n/2})$. The case $\nu = 1$ results in the multivariate Cauchy distribution. From (2), several other distributions can be similarly generated [see also Muirhead (1982, p. 32–34)].

Given this structure, we find the Bayes prediction density of y_2 given D, $f^{\mathrm{B}}(y_2 \mid D)$, where D includes the sample and prior information. This prediction density is defined as [see Atchison and Dunsmore (1975)]

$$\begin{aligned} f^{\mathrm{B}}(y_2 \mid D) &= \int_\Theta f(y_2, y_1 \mid \theta) \cdot \pi(\theta)\, d\theta \Big/ \int_\Theta f(y_1 \mid \theta)\pi(\theta)\, d\theta, \\ &\propto \int_\Theta f(y_2, y_1 \mid \theta)\pi(\theta)\, d\theta, \end{aligned} \tag{4}$$

where $\theta = (\beta, \tau^2)$ and $\theta \in \Theta = R^k \times (0, \infty)$, and the constant of proportionality depends only on D, $f(y_2, y_1 \mid \theta)$ is the (joint) density function of (y_2, y_1) given θ, and $\pi(\theta)$ is the prior pdf of θ.

The first result of the paper is stated in the following proposition which shows that the Bayes prediction density, with spherical errors and an objective prior, is identical to that obtained with the multivariate normal errors.

Throughout we let $\hat{\beta}=(X'X)^{-1}X'y$ and $s^2=\| y-X\hat{\beta}\|^2/n-k$ denote the ordinary least squares estimators of β and $1/\tau^2$ respectively, and let $SSE=\| y-X\hat{\beta}\|^2$ denote the sum of squared errors all based on (y, X). The corresponding quantities based on (y_1, X_1) are denoted by the subscript 1; e.g.,

$$\hat{\beta}_1=(X_1'X_1)^{-1}X_1'y_1 \quad \text{and} \quad s_1^2=\left\| y_1-X_1\hat{\beta}_1\right\|^2.$$

Proposition 1. Let the prior information on (β, τ^2) be objective, with joint pdf $\pi(\beta, \tau^2)\propto(1/\tau)^2$, $\beta\in R^k$ and $\tau^2>0$. Then the Bayes prediction density of y_2 given D, for any Z, is

$$f^{\mathrm{B}}(y_2\mid D)\propto\left[1+\frac{1}{(n_1-k)s_1}2(y_2-X_2\hat{\beta}_1)'\left[I_{n_2}+X_2(X_1'X_1)^{-1}X_2'\right]^{-1}(y_2-X_2\hat{\beta}_1)\right]^{-(n_2+n_1-k)/2}. \tag{5}$$

Proof. Clearly,

$$f^{\mathrm{B}}(y_2\mid D)=\int_{z>0}f^{\mathrm{B}}(y_2\mid D, z)\,\mathrm{d}G(z),$$

where

$$f^{\mathrm{B}}(y_2\mid D, Z)\propto\int_{\Theta}f(y_2, y_1\mid\theta, z)\pi(\theta\mid z)\,\mathrm{d}\theta$$

$$\propto\left[\psi(z)^{-2}\right]^{n/2}\int_{\tau^2>0}\int_{\beta\in R^k}(\tau^2)^{n/2-1}\,\mathrm{e}^{-(\tau^2/2)\psi(z)^{-2}}$$

$$\times[SSE+(\beta-\hat{\beta})'(X'X)(\beta-\hat{\beta})]\,\mathrm{d}\beta\,\mathrm{d}\tau^2$$

$$\propto\left[\psi(z)^{-2}\right]^{(n-k)/2}\int_{\tau^2>0}(\tau^2)^{(n-k)/2-1}\,\mathrm{e}^{-\tau^2/2\psi(z)^{-2}SSE}\,\mathrm{d}\tau^2$$

$$\propto(SSE)^{-(n-k)/2},$$

where the constant of proportionality depends on D but is independent of Z. Because $f^{\mathrm{B}}(y_2\mid D, Z)=f^{\mathrm{B}}(y_2\mid D)$, the random vector $y_2\mid D$ is distributed independently of Z. Now using the result that $SSE=SSE_1+(y_2-X_2\hat{\beta}_1)'[I_{n2}+X_2(X_1'X_1)^{-1}X_2']^{-1}(y_2-X_2\hat{\beta}_1)$ [see for example, Chib et al. (1987b)], (5) follows. Finally, the density in (5) is identical to that obtained in Zellner (1971) under the assumption of multivariate normal errors.

3. Extensions

In this section we consider the case where prior information on (β, τ^2) is from the conjugate family of distributions relevant for the spherical density function in (2). First we point out the important result, which can be proved directly, that if we let $\phi^2 = \tau^2 \cdot \psi(Z)^{-2}$, then the conjugate prior density on (β, τ^2) for the spherical family is given by

$$\pi(\beta, \tau^2) = \int_{Z>0} \pi(\beta \mid \phi^2)\pi(\phi^2) \mid J \mid \mathrm{d}G(z), \tag{6}$$

where

$$\pi(\beta \mid \phi^2) \propto (\phi^2)^{k/2} \, \mathrm{e}^{-\phi^2[(\beta-\beta^*)'A^*(\beta-\beta^*)]/2} \tag{7}$$

and

$$\pi(\phi^2) \propto (\phi^2)^{\delta^*/2-1} \, \mathrm{e}^{-\phi^2\nu^*/2} \tag{8}$$

is a normal-gamma prior for (β, τ^2) with hyperparameters, β^*, A^*, $\delta^*/2$ and $\nu^*/2$, and $|J| = \psi(Z)^{-2}$ is the Jacobian of the transformation from $(\beta, \phi)^2 \to (\beta, \tau^2)$.

This is a convenient result that gives the conjugate prior for the entire spherical family of distributions. As an example, if we consider the case which leads to multivariate-t distribution, with $\psi(Z) = (Z/\nu)^{-1/2}$, and Z having a chi-squared distribution with ν degrees of freedom, eqs. (6)–(8) imply that the conjugate prior pdf for (β, τ^2) is

$$\pi(\beta, \tau^2) \propto \left[\frac{\tau^2}{\nu + \nu^*\tau^2}\right]^{k/2} \cdot \left[1 + \frac{\tau^2}{(\nu + \nu^*\tau^2)}(\beta - \beta^*)'A^*(\beta - \beta^*)\right]^{-(k+\delta^*+\nu)/2}$$

$$\cdot \left[\frac{\nu}{\nu^*\tau^2}\right]^{\nu/2+1} \cdot \left[1 + \frac{\nu}{\nu^*\tau^2}\right]^{-(\delta^*+\nu)/2}.$$

Thus $\pi(\beta, \tau^2)$ is in the form of a multivariate-t pdf for β, given τ^2, times an F-pdf for $\sigma^2 = 1/\tau^2$. This conjugate prior has been derived by Zellner (1976) using a different approach.

The next result shows that the Bayes prediction density with spherical errors and conjugate prior (6) is identical to the prediction density with multivariate normal errors and the usual normal-gamma prior.

Proposition 2. Let prior information on (β, τ^2) be as in (6) with hyperparameters known or assessed. Then the Bayes prediction density of y_2 given D, for any Z, is

$$f^{\mathrm{B}}(y_2 \mid D) \propto \left[(y_2 - X_2\beta_1^{**})'A^{-1}(y_2 - X_2\beta_1^{**}) + C\right]^{-(n_2+n_1+\delta^*)/2} \tag{9}$$

where $A = I_{n_2} + X_2(A^* + X_1'X_1)^{-1}X_2'$,

$$C = y_1'y_1 + \beta^{*\prime}A^*\beta^* - (A^*\beta^* + X_1'y_1)'(A^* + X_1'X_1)^{-1}(A^*\beta^* + X_1'y_1) + \nu^*$$

and β_1^{**} is the mean of the pdf of β prior to observing y_2, and is given by

$$\beta_1^{**} = (A^* + X_1'X_1)^{-1}(A^*\beta^* + X_1'X_1\hat{\beta}_1).$$

That is, y_2 given D has an n_2-variate multivariate-t density with $n_1 + \delta^*$ degrees of freedom, location vector $X_2\beta_1^{**}$, and precision matrix $(n_1 + \delta^*)(AC)^{-1}$.

Proof. Proceeding along the lines of Proposition 1, it can be seen that

$$\begin{aligned} f^{\mathrm{B}}(y_2 \mid D, Z) &\propto \int_{\tau^2>0}\int_{\beta\in R^k} f(y \mid \beta, \tau^2, Z)\cdot\pi(\beta, \tau^2 \mid Z)\,\mathrm{d}\beta\,\mathrm{d}\tau^2 \\ &\propto \psi(z)^{-2}\int_{\tau^2>0}\int_{\beta\in R^k} (\phi^2)^{(n+k+\delta^*)/2-1} \\ &\quad\times \mathrm{e}^{-(\phi^2/2)[\nu^*+SSE+(\beta-\beta^*)A^*(\beta-\beta^*)+(\beta-\hat{\beta})'(X'X)(\beta-\hat{\beta})]}\,\mathrm{d}\beta\,\mathrm{d}\tau^2 \\ &\propto \left[\psi(z)^{-2}\right]^{(n+k+\delta^*)/2}\int_{\tau^2>0}\int_{\beta\in R^k} (\tau^2)^{(n+k+\delta^*)/2-1} \\ &\quad\times \mathrm{e}^{-(\tau^2/2)\psi(Z)^{-2}[\nu^*+SSE+(\beta-\beta^{**})'(A^*+X'X)(\beta-\beta^{**})+R]}\,\mathrm{d}\beta\,\mathrm{d}\tau^2, \end{aligned}$$

where

$$\beta^{**} = (A^* + X'X)^{-1}(A^*\beta^* + X'X\hat{\beta}) \text{ and } R = \beta^{*\prime}A^*\beta^* + \hat{\beta}'(X'X)\hat{\beta} - \beta^{**\prime}(A^* + X'X)\beta^{**}.$$

Thus,

$$\begin{aligned} f^{\mathrm{B}}(y_2 \mid D, Z) &\propto \left[\psi(Z)^{-2}\right]^{(n+\delta^*)/2}\int_{\tau^2>0} (\tau^2)^{(n+\delta^*)/2-1}\,\mathrm{e}^{-\tau^2\psi(Z)^{-2}[\nu^*+SSE+R]/2}\,\mathrm{d}\tau^2 \\ &\propto [\nu^* + SSE + R]^{-(n_2+n_1+\delta^*)/2}, \end{aligned}$$

independent of Z. The rest follows after considerable simplification of the quantities within square brackets.

Note that in the limiting case when $A^* = 0$, $\delta^* = -k$ and $\nu^* = 0$ the prediction density in (9) reduces to (5). Further, (9) can easily be seen to be identical to the prediction density obtained with multivariate normal errors and a normal-gamma prior for (β, τ^2).

4. Concluding comments

This paper considers the Bayes prediction problem for a spherical linear model with both objectives and informative prior information and obtained the surprising result that predictive inferences are completely unaffected by departures from the normality assumption in the direction of the spherical family. In contrast, as is shown in Zellner (1976) for the multivariate-t case, parameter posteriors have to be modified.

References

Atchison, J. and I.R. Dunsmore, 1975, Statistical prediction analysis (Cambridge University Press, Cambridge).
Berger, J., 1975, Minimax estimation of location vectors for a wide class of densities, Annals of Statistics 3, 1318–1328.
Chib, S., S. Rao Jammalamadaka and Ram C. Tiwari, 1986a, Bayes and likelihood prediction in regressions with homoscedastic and heteroscedastic multivariate-t errors, Mimeo.
Chib, S., S. Rao Jammalamadaka and Ram C. Tiwari, 1986b, Another look at the recursive estimation of the general linear model, American Statistician, forthcoming.
Chmielewski, M.A., 1981, Elliptically symmetric distributions: A review and bibliography, International Statistical Review 49, 67–76.
Judge, C., S. Miyazaki and T. Yancey, 1985, Minimax estimates for the location vectors of spherically symmetric densities, Econometric Theory 1, 409–417.
Kelker, D., 1970, Distribution theory of spherical distributions and a location-scale parameter generalization, Sankya, Series A, 419–430.
Muirhead, R.J., 1982, Aspects of multivariate statistical theory (Wiley, New York).
Press, S.J., 1982, Applied multivariate analysis (Krieger, Huntington, NY).
Strawderman, W.E., 1974, Minimax estimation of location parameters for certain spherically symmetric distributions, Journal of Multivariate Analysis 4, 255–264.
Ullah, Aman and V. Zande-Walsh, 1985, Estimation and testing in a regression model with spherically symmetric errors, Economics Letters 17, 127–132.
West, M., 1984, Outlier models and prior distributions in Bayesian linear regression, Journal of the Royal Statistical Society, Series B, 46, 431–439.
Zellner, A., 1971, An introduction to Bayesian inference in econometrics (Wiley, New York).
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.