BAYES PREDICTION IN REGRESSIONS WITH ELLIPTICAL ERRORS*

Siddhartha CHIB

University of Missouri, Columbia, MO 65211, USA

Ram C. TIWARI

University of North Carolina, Charlotte, NC 28223, USA

S. Rao JAMMALAMADAKA

University of California, Santa Barbara, CA 93106, USA

In this paper the prediction problem is considered for linear regression models with elliptical errors when the Bayes prior is non-informative. We show that the Bayes prediction density under the elliptical errors assumption is exactly the same as that obtained with normally distributed errors. Thus, assuming that the errors have a normal distribution, when the true distribution is elliptical, will not lead to incorrect predictive inferences if the error variance structure is correctly specified. This extends the results of Zellner (1976). Finally, based on Monte Carlo numerical integration procedures, computations are provided in a model with multiplicative heteroscedasticity.

1. Introduction

Lately there has been much theoretical and applied interest in linear models with non-normal disturbances as several authors have explored the consequences of non-normality and heavy-tailed error distributions. In the context of one heavy-tailed error distribution, the multivariate-t distribution, Zellner (1976) provides a Bayes and classical analysis of a regression model, Ullah and Walsh (1984) consider the issue of testing restrictions, and Kelejian and Prucha (1985) point out the importance of distinguishing between independence and uncorrelatedness in non-normal situations. Models containing systems of equations are discussed in Prucha and Kelejian (1984) and Sutradhar and Ali (1986), and the issue of minimax estimation of the location parameters is taken up in Judge, Miyazaki and Yancey (1985). The problem of prediction

based on the assumption of spherical errors is considered in Dawid (1981) and Jammalamadaka, Tiwari and Chib (1987).

The purpose of this paper is to show that the assumption of elliptical errors in linear regression models with limited prior information on the parameters has no new consequences for prediction if the error variance structure is correctly specified. For example, correctly assuming that the errors are spherical but misspecifying the error distribution as normal, when the true distribution is a multivariate *t*, *does not* lead to incorrect predictive inferences. In Zellner (1976) it is shown that when prior information is of the non-informative type, the joint posterior of β, the regression coefficient vector, and τ^2, the error precision, differs depending on the error distribution assumed, although the marginal posterior of β is unaffected. The latter feature, which carries over in more general settings, is essentially the key as to why the predictive distribution is invariant to a wide class of error distributions. This means that in the study of robustness, perturbations from the assumed model can have very different consequences for estimation as opposed to prediction.

The plan of this paper is as follows. Section 2 contains the Bayes prediction densities for the linear regression model with elliptical errors. In section 3, we show that the resulting prediction density and the predictive moments can be computed numerically using procedures that are described in Geweke (1987a, b). We base our computational analysis on a model with multiplicative heteroscedasticity.

2. The elliptical error model

Consider the linear regression model with non-random regressors and elliptical errors,

$$y = X\beta + \psi(Z)u, \tag{2.1}$$

where $y \in R^n$, $\beta \in R^k$, $u \sim N(0, (\tau^2 \Lambda)^{-1})$, $\tau^2 \Lambda$ is the precision matrix, $\tau^2 > 0$, Z is a positive random variable with distribution G independent of u, and $\psi(\cdot)$ is a positive function. Model (2.1) implies that conditionally on Z,

$$y|Z, \beta, \tau^2, \Lambda \sim N\left(X\beta, (\tau^2 \Lambda)^{-1} \psi(Z)^2\right), \tag{2.2}$$

while the unconditional distribution of $y|\beta, \tau^2, \Lambda$ is

$$f(y|\beta, \tau^2, \Lambda) \propto \int_0^\infty |\tau^2 \Lambda|^{n/2} \left\{\psi(Z)^{-2}\right\}^{n/2}$$

$$\times \exp\left(-\frac{\tau^2}{2}\psi(Z)^{-1}(y - X\beta)'\Lambda(y - X\beta)\right) dG(Z). \tag{2.3}$$

From (2.3) several distributions including the ε-contaminated normal and multivariate exponential can be generated [cf. Muirhead (1983, pp. 32–34)]. 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 elliptical multivariate-t distribution with density given by [cf. Press (1982, p. 136)]

$$f(y|\beta, \tau^2, \Lambda) = c(\tau^2)^{n/2}|\Lambda|^{1/2}$$

$$\times \left[1 + \frac{\tau^2}{\nu}(y - X\beta)'\Lambda(y - X\beta)\right]^{-(n+\nu)/2}, \quad (2.4)$$

where

$$c = \Gamma\left(\frac{n+\nu}{2}\right)\bigg/\left(\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{n/2}\right).$$

The case $\nu = 1$ results in the multivariate Cauchy distribution. If $\Lambda = I_n$, the identity matrix of order n, we obtain the homoscedastic (or spherical) multivariate-t distribution.

The general Bayes prediction problem in the context of model (2.1) can be described as follows. Suppose that the vector of observations y is partitioned as $y = (y_1', y_2')'$, where $y_1 : n_1 \times 1$ is observed and interest centers on predicting $y_2 : n_2 \times 1$, an unobserved set of future observations, assuming that $X = (X_1', X_2')'$ is known. The Bayes prediction density is defined as

$$f^B(y_2|y_1) = \frac{\int\int f(y_1, y_2|\theta)\pi(\theta)\,d\theta}{\int f(y_1|\theta)\pi(\theta)\,d\theta}, \quad (2.5)$$

provided both the integrals are finite, θ is the parameter that indexes the pdf of y, and $\pi(\theta)$ is the pdf of θ prior to observing y_1. The Bayes prediction density above is an estimate of $f(y_2|y_1, \theta)$, the conditional pdf of y_2, given y_1 so that the resulting function is θ-free.

The definition in (2.5) is used to compute the Bayes prediction density for the elliptical regression error model. Observe that conditional on τ^2, the random vectors y_1 and y_2 are uncorrelated, however, they are not independent unless Z is a degenerate random variable.

We suppose that the matrix $\Lambda : n \times n$ is described by a fixed (i.e., independent of n) parameter vector η, $\eta \in R^p$. Also suppose that the non-informative prior of (β, τ^2, η) is

$$\pi(\beta, \tau^2, \eta) = (\tau^2)^{-1}\pi(\eta), \quad \beta \in R^k, \quad \tau^2 > 0, \quad \eta \in R^p. \quad (2.6)$$

First we note the key result that the marginal posterior density of β given y_1 is unaffected by a change in the error distribution from multivariate normal to elliptical, which extends the result of Zellner (1976).[1] If we let the prior pdf of (β, τ^2, η) be (2.6), then the marginal posterior pdf of β given $y_1 : n_1 \times 1$, for any Z, is

$$
\pi(\beta|y_1) \propto \int_{\eta \in R^p} \left[1 + \frac{1}{(n_1 - k)s_1^2(\eta)} (\beta - \hat{\beta}_1(\eta))' \right.
$$

$$
\left. \times X_1' \Lambda_1(\eta) X_1 (\beta - \hat{\beta}_1(\eta)) \right]^{-(k + n_1 - k)/2} \pi(\eta) \, d\eta,
$$

$$
(2.7)
$$

where

$$
\hat{\beta}_1(\eta) = \left(X_1' \Lambda_1(\eta) X_1 \right)^{-1} X_1' \Lambda_1(\eta) y_1,
$$

$$
s_1^2(\eta) = \mathrm{SSE}_1(\eta) / (n_1 - k),
$$

$$
\mathrm{SSE}_1(\eta) = \left(y_1 - X_1 \hat{\beta}_1(\eta) \right)' \Lambda_1(\eta) \left(y_1 - X_1 \hat{\beta}_1(\eta) \right),
$$

where $\Lambda_1 = \Lambda_1(\eta)$ is the $n_1 \times n_1$ submatrix corresponding to y_1. This is exactly the marginal posterior density of β obtained in the case where the errors are multivariate normal with precision matrix $\tau^2 \Lambda_1$ [cf. Leamer (1978)]. Consequently the posterior of β given y_1, y_2 and Z is equal to the posterior of β given y_1 and y_2, i.e.,

$$
\beta|y_1, y_2, Z \stackrel{d}{=} \beta|y_1, y_2 \quad \text{and} \quad (\beta, y_1, y_2)|Z \stackrel{d}{=} (\beta, y_1, y_2),
$$

where $\stackrel{d}{=}$ stands for equal in distribution. Hence, the posterior of y_2 given y_1 and Z is equal to the posterior of y_2 given y_1, i.e.,

$$
y_2|y_1, Z \stackrel{d}{=} y_2|y_1,
$$

which is stated in the next proposition.

Proposition 1. Let $\pi(\beta, \tau^2, \eta) \propto (\tau^2)^{-1}\pi(\eta)$. Then the Bayes prediction density of y_2, under model (2.1) is for any Z,

$$
f^B(y_2|y_1) = c_0^{-1} \int_{\eta \in R^p} |\Lambda(\eta)|^{1/2} |X'\Lambda(\eta)X|^{-1/2}
$$

$$
\times \mathrm{SSE}(\eta)^{-(n-k)/2} \pi(\eta) \, d\eta,
$$

$$
(2.8)
$$

---

[1] This point, that the invariance of the predictive distribution is connected to the invariance of the marginal posterior of β, was made to us by a referee. Hence the differential error distribution has an effect only on the posterior distribution of τ^2.

where

$$\text{SSE}(\eta) = \left(y - X\hat{\beta}(\eta) \right)' \Lambda(\eta) \left(y - X\hat{\beta}(\eta) \right),$$

$$\hat{\beta}(\eta) = \left(X'\Lambda(\eta)X \right)^{-1} X'\Lambda(\eta) y,$$

$$c_0 = \pi^{n_2/2} \Gamma\left(\frac{n_1 - k}{2} \right) \left(\Gamma\left(\frac{n - k}{2} \right) \right)^{-1}$$

$$\times \int_{-\infty}^{\infty} |\Lambda_1(\eta)|^{1/2} |X_1'\Lambda_1(\eta)X_1|^{-1/2}$$

$$\times \text{SSE}_1(\eta)^{-(n_1 - k)/2} \pi(\eta) \, d\eta.$$

Using the updating results in Chib, Jammalamadaka and Tiwari (1987), it is possible to also express this density as a mixture of multivariate-t densities, a representation that is useful under some circumstances. If we let

$$\Lambda^{-1}(\eta) = \begin{bmatrix} \Sigma_1(\eta) & C(\eta) \\ C'(\eta) & \Sigma_2(\eta) \end{bmatrix},$$

where

$$\Sigma_1(\eta) : n_1 \times n_1, \qquad C(\eta) : n_1 \times n_2, \qquad \Sigma_2(\eta) : n_2 \times n_2,$$

and define

$$\hat{\mu}_{2.1}(\eta) = X_2 \hat{\beta}_1(\eta) + C'(\eta)\Sigma_1(\eta)^{-1} \left(y_1 - X_1 \hat{\beta}_1(\eta) \right),$$

$$\Sigma_{2.1}(\eta) = \Sigma_2(\eta) - C'(\eta)\Sigma_1(\eta)^{-1} C(\eta),$$

$$Q(\eta) = X_2 - C'(\eta)\Sigma_1(\eta)^{-1} C(\eta),$$

$$\Omega(\eta) = \Sigma_{2.1} + Q(\eta) \left(X_1'\Sigma_1(\eta)^{-1}X_1 \right)^{-1} Q'(\eta),$$

then the Bayes prediction density of y_2 given in (2.8) can also be expressed as

$$f^B(y_2|y_1) = \int \psi_{n_2}(\hat{\mu}_{2.1}(\eta), P_{2.1}(\eta), n_1 - k) \pi(d\eta|y_1), \tag{2.9}$$

where the integrand ψ_{n_2} in (2.9) is the $(n_2 \times 1)$ multivariate-t density with mean vector $\hat{\mu}_{2.1}(\eta)$, precision matrix $P_{2.1}(\eta) = [\Omega(\eta)]^{-1}/s_1^2(\eta)$, and $n_1 - k$ degrees of freedom. The integration is performed with respect to the posterior

density of η, given y_1. This posterior density is

$$\pi(\eta|y_1) = m(y_1|\eta)\pi(\eta) \Big/ \int m(y_1|\eta)\pi(\eta)\,\mathrm{d}\eta,$$

where

$$m(y_1|\eta) = |\Lambda_1(\eta)|^{1/2}|X_1'\Lambda_1(\eta)X_1|^{-1/2}[\mathrm{SSE}_1(\eta)]^{-(n_1-k)/2}.$$

What is interesting about the MVt pdf in (2.8) is that the random-variable Z plays no role in the final answer, and that the prediction density is identical to that obtained under the assumption of multivariate normal errors. This shows that when there is limited prior information on the parameters, the assumption of normality is robust to deviation in the direction of elliptical distributions as far as prediction is concerned.

Based on this result, the prediction density with spherical errors (i.e., when $\Lambda = I_n$) can be obtained as a special case. Let the ordinary least squares estimates of β and $(\tau^2)^{-1}$, respectively, based on y_1, be given by

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y_1 \quad \text{and} \quad s_1^2 = \mathrm{SSE}_1/n_1 - k,$$

with

$$\mathrm{SSE}_1 = \|y_1 - X_1\hat{\beta}_1\|^2.$$

Then by setting $\Lambda = I_n$ in (2.9) we get the following result:

Corollary 1 [Jammalamadaka, Tiwari and Chib (1987)]. *Let the prior pdf of* (β, τ^2) *be* $\pi(\beta, \tau^2) \propto (\tau^2)^{-1}$, $\beta \in R^k$, $\tau^2 > 0$. *Then the Bayes prediction density of* y_2, *given* y_1, *for any* Z, *is* $\psi_{n_2}(X_2\hat{\beta}_1, [I_{n_2} + X_2(X_1'X_1)^{-1}X_2']^{-1}/s_1^2, n_1 - k)$, *a multivariate-t pdf with mean vector* $X_2\hat{\beta}_1$, *precision matrix* $[I_{n_2} + X_2(X_1'X_1)^{-1}X_2']^{-1}/s_1^2$, *and* $n_1 - k$ *degrees of freedom.*

This is the prediction density in Zellner (1971) under the assumption of multivariate normal errors with independent components.

3. An example

Consider a special case of the model in the previous section and suppose $(\tau^2\Lambda)^{-1}$ is a diagonal matrix representing heteroscedastic variances. In fact heteroscedastic linear models have been extensively studied, but almost exclusively from the viewpoint of parameter estimation and testing. Surekha and Griffiths (1984) is a useful recent work that compares some Bayes and sampling theory estimators in two heteroscedastic error models. We can apply the results of the previous section to examine the consequences of heteroscedasticity on predictive inferences.

Assume, along the lines of Geary (1966), Park (1966), Lancaster (1968) and Harvey (1976), that the variance of the error term is proportional to an unknown power of one of the covariates, say the ith, and write

$$\text{var}\left(\varepsilon_t | \left(\sigma^2, \gamma\right)\right) \equiv \sigma_t^2 = \left(\tau^2\right)^{-1} x_{ti}^\gamma, \qquad x_{ti} > 0, \quad \gamma \in R^1,$$

$$t = 1, 2, \ldots n_1, n_1 + 1, \ldots, n,$$

and let $i \geq 2$, since the first covariate in X is usually a constant.[2] Under this specification the distribution of the vector y in model (2.1) is elliptical with density

$$f\left(y | \beta, \tau^2, \gamma\right) \propto \int_0^\infty |\tau^2 P(\gamma)|^{n/2} \left\{\psi(Z)^{-2}\right\} \exp\left(-\frac{\tau^2}{2}\psi(Z)^{-1}\right.$$

$$\left. \times \left(y - X\beta\right)' P(\gamma)\left(y - X\beta\right)\right) dG(Z), \tag{3.1}$$

where $\tau^2 P(\gamma) = \tau^2 \text{diag}(x_{1i}^{-\gamma}, x_{2i}^{-\gamma}, \ldots, x_{ni}^{-\gamma})$ is the precision matrix of the error vector ε. Notice that $\gamma = 0$ reproduces the spherical case discussed in Corollary 1.

Suppose that *a priori* β, $\log \tau^2$ and γ are independent and uniformly distributed, which implies that the prior of (β, τ^2, γ) is given by

$$\pi\left(\beta, \tau^2, \gamma\right) \propto \left(\tau^2\right)^{-1}, \qquad \beta \in R^k, \quad \tau^2 > 0, \quad \gamma \in R^1. \tag{3.2}$$

When the error distribution is heteroscedastic multivariate normal, this is also Jeffreys' invariant prior of (β, τ^2, γ). See Surekha and Griffiths (1984, p. 91), where the above prior is adopted.

We define:

$$P_1(\gamma) = \text{diag}\left(x_{1i}^{-\gamma}, \ldots, x_{n_1 i}^{-\gamma}\right),$$

$$\hat{\beta}_1(\gamma) = \left(X_1' P_1(\gamma) X_1\right)^{-1} X_1 P_1(\gamma) y_1,$$

$$\text{SSE}_1(\gamma) = \left(y_1 - X_1 \hat{\beta}_1(\gamma)\right)' P_1(\gamma)\left(y_1 - X_1 \hat{\beta}_1(\gamma)\right),$$

---

[2] Notice that $x_{ti} > 0$ and thus this regression specification cannot be obtained from a joint normal distribution for the dependent and independent variables or from any other joint distribution involving doubly-infinite ranges. We owe this point to Arnold Zellner.

and

$$s_1^2(\gamma) = \mathrm{SSE}_1(\gamma)/(n_1 - k).$$

The next result follows from Proposition 1.

Corollary 2.   Let $\pi(\beta, \tau^2, \gamma) \propto (\tau^2)^{-1}$. *Let the observations* y *have density as in* (*3.1*). *Suppose* y_2 *is unobserved. Then the Bayes prediction density for* y_2 *is given by*

$$f^{\mathrm{B}}(y_2|y_1) = \frac{\displaystyle\int_{-\infty}^{\infty} f^{\mathrm{B}}(y_2|y_1, \gamma) m(y_1|\gamma)\, \mathrm{d}\gamma}{\displaystyle\int_{-\infty}^{\infty} m(y_1|\gamma)\, \mathrm{d}\gamma}, \tag{3.3}$$

where $f^{\mathrm{B}}(y_2|y_1, \gamma) = \psi_{n_2}(X_2\hat{\beta}_1(\gamma), [\Sigma_2(\gamma) + X_2(X_1'P_1(\gamma)X_1)^{-1}X_2']^{-1}/$ $s_1^2(\gamma), n_1 - k)$, *and* $m(y_1|\gamma)/\int m(y_1|\gamma)\, \mathrm{d}\gamma$ *is the marginal posterior density of* γ, *given* y_1, *with*

$$m(y_1|\gamma) = |P_1(\gamma)|^{1/2}|X_1'P_1(\gamma)X_1|^{-1/2}[\mathrm{SSE}_1(\gamma)]^{-(n_1-k)/2}.$$

This density can be analyzed numerically. Although the integral can be approximated by Laplace's method as discussed in Tierney and Kadane (1986), we employ a direct numerical integration procedure to produce the prediction density and predictive moments.

Suppose we let $y_2 : 1 \times 1$, and consider a model similar to the one in Surekha and Griffiths (1984), in which the data-generating process containing one explanatory variable is assumed to be

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t,$$

with $\beta_0 = 1$, $\beta_1 = 10$, $x_t \sim \mathrm{U}(1, 10)$, uniform on the interval $[1, 10]$. Since our results show that the prediction function with elliptical errors is identical to that with multivariate normal errors, the ε_t are random draws from a normal distribution with mean zero and variance $\sigma_t^2 = (\tau^2)^{-1}x_t^\gamma, (\tau^2)^{-1} = 4, \gamma = 1$. A sample size of $n_1 = 20$ is used.[3] The function $f^{\mathrm{B}}(y_2|y_1, \gamma)$ in (3.3) is numerically integrated with respect to (wrt) the posterior of γ given y_1, using Monte Carlo integration with importance sampling [see, e.g., Geweke (1987a)]. The choice of the importance function is critical in obtaining accurate estimates of the integral. In general, for $f^{\mathrm{B}}(y_2|y_1, \gamma)$ in (3.3), the ideal importance function is, of course, the posterior pdf of γ given y_1. Since it is difficult in our case to

---

[3] The actual data set is available from the authors. The analysis was also carried out with sample size $n_1 = 10$ and $n_1 = 30$. Since the results were similar, we decided to save space and not report those results.
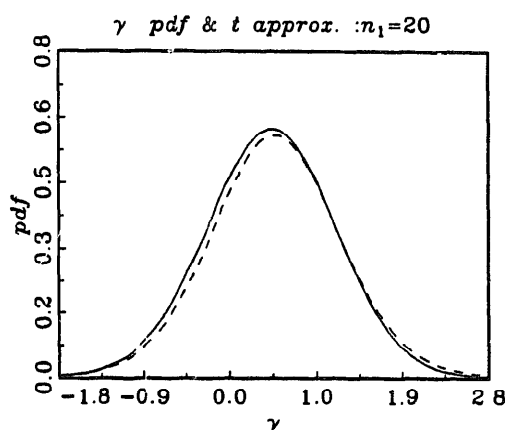
Fig. 1. $\pi(\gamma|y_1)$: ———, $\tilde{f}(\gamma)$: - - -, $\tilde{f}(\gamma) = \psi_1(C\ 524252, 2.302607, 18)$. [The normalizing constant of $\pi(\gamma|y_1)$, i.e., $\int m(y_1|\gamma)d\gamma$, is found by Monte Carlo integration. The importance function found adequate is the uniform density on $[-6,6]$ which was used to produce 1000 random draws of γ.]

sample from this posterior pdf, we have used an importance function, $\tilde{f}(\gamma)$, that mimics $\pi(\gamma|y_1)$. For large sample $\pi(\gamma|y_1)$ will tend to be approximately $N_1(\hat{\gamma}_{ML}, [\mathrm{var}(\hat{\gamma}_{ML})]^{-1})$, where $\hat{\gamma}_{ML}$ is the maximum likelihood estimate of γ and $\mathrm{var}(\hat{\gamma}_{ML})$ is the variance of the ML estimate (computed in the usual way from the observed Fisher information matrix). A possible choice for $\tilde{f}(\gamma)$ is a $\psi_1(\hat{\gamma}_{ML}, [\mathrm{var}(\hat{\gamma}_{ML})]^{-1}, \nu)$ pdf with d.f. ν chosen to ensure that the tails of $\tilde{f}(\gamma)$ are no sharper than those of $\pi(\gamma|y_1)$. The adequacy of this choice is illustrated in fig. 1.

Given that the Student-t density $\tilde{f}(\gamma) = \psi_1(0.524252, 2.302607, 18)$ is an adequate importance function, N values of γ, say $\gamma^{(i)}$, $i = 1, 2, \ldots, N$, may be generated randomly, where N is a suitably large number. The Monte Carlo estimate of (3.3) at the point y_2 is

$$f^{B}(y_2|y_1) \simeq \frac{\sum_{i=1}^{N} \left[f^{B}(y_2|y_1, \gamma^{(i)}) m(y_1|\gamma^{(i)}) / \tilde{f}(\gamma^{(i)}) \right]}{\sum_{i=1}^{N} \left[m(y_1|\gamma^{(i)}) / \tilde{f}(\gamma^{(i)}) \right]}. \tag{3.4}$$

Similarly, adapting the ideas in the interesting paper of Geweke (1987b), the predictive moments can be found as follows. It can be confirmed that the evaluation of the expectation $h(y_2)$ with respect to the predictive pdf is equivalent to finding a certain expectation with respect to the posterior of γ given y_1. In particular,

$$E^{f^{B}(y_2|y_1)}[h(y_2)] = E^{\pi(\gamma|y_1)}[g(\gamma)], \tag{3.5}$$

Table 1

Central moments, skewness and kurtosis of the predictive distribution.[a]

n_1	μ_1	μ_2	μ_3	μ_4	Sk	K
20	52.76888	24.61801	0.62324	2077.4587	0.00510	3.4279

[a] The moments are calculated using $S = 100$ antithetic replications of y_2, for each value of γ. Sk is a measure of skewness ($Sk = \mu_3/\sigma^3$) and K is a measure of kurtosis ($K = \mu_4/\sigma^4$).

where

$$g(\gamma) = \int h(y_2) f^B(y_2|y_1, \gamma) \, dy_2.$$

Clearly for a given value of γ, say $\gamma^{(i)}$, it is easy to generate random draws from $f^B(y_2|y_1, \gamma^{(i)})$. Suppose for each $\gamma^{(i)}$, S antithetic y_2 values, denoted $y_2^{(j|i)}$, $j = 1, \ldots, S$, are thus drawn. Then the Monte Carlo estimate of $g(\gamma^{(i)})$ is

$$\tilde{g}_i = \sum_{j=1}^{S} y_2^{(j|i)}/S,$$

and the Monte Carlo estimate of the LHS of (3.5) is

$$E[h(y_2)|y_1] \simeq \frac{\sum_{i=1}^{N} \left[\left(\tilde{g}_i m(y_1|\gamma^{(i)})/\tilde{f}(\gamma^{(i)}) \right) \right]}{\sum_{i=1}^{N} \left[m(y_1|\gamma^{(i)})/\tilde{f}(\gamma^{(i)}) \right]}, \tag{3.6}$$

which can be programmed. Notice that, since $E(y_2|y_1, \gamma)$ and $\text{var}(y_2|y_1, \gamma)$ are given by simple expressions, the unconditional mean and variance can be obtained by numerically integrating over $\pi(\gamma|y_1)$ for, e.g., $E(y_2|y_1) = E^{\pi(\gamma|y_1)}[E(y_2|y_1, \gamma)$. However, (3.6) is a more general procedure that can be used even when the latter approach is infeasible or complicated. Based on (3.4) and (3.6), the prediction density and predictive moments are provided below. In addition, we report the 'true' prediction density. The true prediction density is $f(y_2|y_1, \theta)$, which for our chosen parameter values and simulated $X_2 = (1, 5.21737)$ is the density of a $N_1(53.17366, 0.047917)$ distribution.

Examination of table 1 confirms the visual impression from fig. 2 that the prediction density is symmetric and, relative to a normal density with the same mean and variance, is slightly more peaked. It, therefore, appears that symmetric (around the mean) highest probability density prediction intervals are
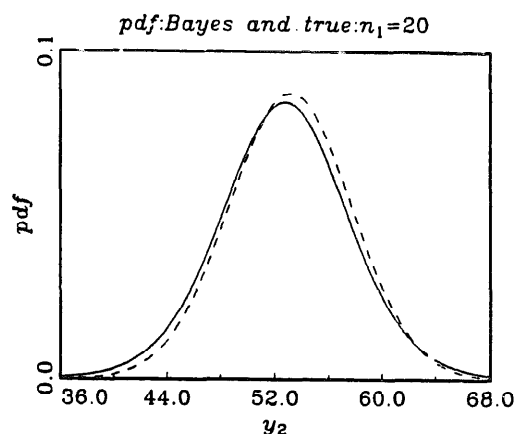
Fig. 2. $f^B(y_2|y_1)$: ——, true: ---. [The number of replications is $N = 2000$; antithetic sampling is used. Over the 2000 replications, the minimum value of $\gamma = -1.8775$ and the maximum is 2.9260.]

adequate for this problem. Whether the same phenomenon arises in other experiments with heteroscedasticity can be investigated along the lines developed in the paper. Of course, the ideas in this paper can be used quite generally in a variety of other contexts that are described by the elliptical error structure used here.

References

Chib, S., S. Rao Jammalamadaka and Ram C. Tiwari, 1987, Another look at the recursive estimation of the general linear model, American Statistician 41, 56–58.

Davison, A.C., 1986, Approximate predictive likelihood, Biometrika 73, 323–332.

Dawid, A.P., 1981, Some matrix-variate distribution theory: Notational considerations and a Bayesian application, Biometrika 68, 265–274.

Geary, R.C., 1966, A note on residual heterovariance and estimation efficiency in regression, American Statistician 20, 30–31.

Geweke, J., 1987a, Bayesian inference in econometric models using Monte Carlo integration, Mimeo.

Geweke, J., 1987b, Exact predictive densities in linear models with ARCH disturbances, Mimeo.

Harvey, A.C., 1976, Estimating regression models with multiplicative heteroscedasticity, Econometrica 44, 461–465.

Jammalamadaka, S. Rao, Ram C. Tiwari and S. Chib, 1987, Bayes prediction in the linear model with spherically symmetric errors, Economics Letters 24, 39–44.

Johnson, N.L. and S. Kotz, 1972, Continuous multivariate distributions (Wiley, New York).

Judge C., S. Miyazaki and T. Yancey, 1985, Mimimax estimates for the location vectors of spherically symmetric densities, Econometric Theory 1, 409–417.

Kelejian, Harry H. and I.R. Prucha, 1985, Independent or uncorrelated disturbances in linear regression: An illustration of the difference, Economics Letters 19, 35–38.

Lancaster, T., 1968, Grouping estimates on heteroscedastic data, Journal of the American Statistical Association 63, 182–191.

Leamer, E.E., 1978, Specification searches (Wiley, New York).

Muirhead, R.J., 1983, Aspects of multivariate statistical theory (Wiley, New York).

Park, R.E., 1966, Estimation with heteroscedastic error terms, Econometrica 34, 888.

Press, S.J., 1982, Applied multivariate analysis (Robert Krieger, Melbourne, FL).

Prucha, I.R. and H.H. Kelejian, 1984, The structure of simultaneous equation estimators: A generalization towards nonnormal disturbances, Econometrica 52, 721–736.

_urekha, K. and W.E. Griffiths, 1984, A Monte Carlo comparison of some Bayesian and sampling theory estimators in two heteroscedastic error models, Commun. Statist. Simula. Comp. 13, 85–105.

Sutradhar, B.C. and M.M. Ali, 1986, Estimation of the parameters of a regression model with a multivariate-t error variable, Communications Statistics Theory Method 15, 429–430.

Tierney, L. and J.B. Kadane, 1986, Accurate approximations for posterior moments and marginal densities, Journal of the American Statistical Association 81, 82–86.

Ullah, A. and V. Zande-Walsh, 1984, On the robustness of LM, LR, and W tests in regression models, Econometrica 52, 1055–1066.

West, M., 1984, Outlier models and prior distributions in Bayesian linear regression, Journal of the Royal Statistical Society 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 term, Journal of the American Statistical Association 71, 400–405.