Ann. Inst. Statist. Math.
Vol. 47, No. 2, 211–224 (1995)

BAYESIAN MULTIPERIOD FORECASTS FOR ARX MODELS*

SHU-ING LIU

Graduate Institute of Statistics, National Central University, Chung-Li, Taiwan 320, R.O.C.

(Received February 21, 1994; revised January 9, 1995)

Abstract. Bayesian multiperiod forecasts for AR models with random independent exogenous variables under normal-gamma and normal-inverted Wishart prior assumptions are investigated. By suitably arranging the integration order of the model's parameters, a t-density mixture approximation is analytically derived to provide an estimator of the posterior predictive density for any future observation. In particular, a suitable t-density is proposed by a convenient closed form. The precision of the discussed methods is examined by using some simulated data and one set of real data up to lead-six-ahead forecasts. It is found that the numerical results of the discussed methods are rather close. In particular, when sample sizes are sufficiently large, it is encouraging to apply a convenient t-density in practical usage. In fact, this t-density estimator asymptotically converges to the true density.

Key words and phrases: ARX model, Bayesian forecast, t-density mixture, posterior predictive density, random regression.

1. Introduction

The purpose of this work is to develop approximations of the multiperiod predictive density of AR(p) models with random exogenous variables from the Bayesian point of view. Bayesian analysis of autoregressive moving-average (ARMA) models has been investigated since late 1960 and extensively applied in many areas. Referring to forecasting, in order to obtain a k-step posterior predictive density, theoretically at least, a $(k-1)$-multiple integration should be computed. This numerical difficulty limits the applications of the Bayesian procedure for multiperiod forecasting. In order to obtain the posterior predictive density, a suitable approximation is necessary. Chow (1974) presented formulations for estimating moments of predictive density for AR(1) models. Applying the usual numerical method, Monahan (1983) computed the percentiles of multiperiod (up to lead-five-ahead forecasts) predictive densities for AR(1) models. A Gaussian sum approximation, obtained by combining Laguerre and Gauss-Hermite

* This research was supported by the National Science Council, Republic of China under contract #NSC82-0208-M-008-086.

integration was proposed by Schnatter (1988) to estimate the predictive density for AR(p) models. Recently, Tanner (1993) summarized a variety of computational algorithms which provide an approximation of the desired density by a Monte Carlo simulation, such as the Gibbs sampling technique (Gelfand and Smith (1990)). These tools help to numerically compute the predictive density. A substitution method proposed by Liu (1994) was suggested in order to analytically approximate the predictive density for any future step by a suitable t-density for any ARMA model with strongly non-random exogenous variables under a normal-gamma prior assumption. In that paper, numerical results indicate that this approach is pretty accurate if shorter period forecasts are considered. Moreover, the approximate t-density will asymptotically converge to the true one.

Most Bayesian studies treated the so-called conditional case by considering the posterior predictive density for future values of the dependent variable, given future values of the independent variables. Zellner and Park (1987) investigated an unconditional case, Bayesian prediction for a random regression model. Theoretically, it becomes a non-linear problem and it is impossible to obtain a closed form for the predictive density even only one step ahead. The multiple integrations are very time consuming, especially when a longer future step is considered. To overcome this limitation, a reasonable approximation is needed. Here, a more extensive model, an AR(p) model with random independent exogenous variables is considered. Assuming conjugate priors, the posterior predictive density can be approximately treated as a mixture of some t-density functions. In particular, a suitable t-density is provided to approach the posterior predictive density.

The whole paper is organized as follows: in Section 2, a formal Bayesian analysis of the discussed model is presented. Most of the results are adapted from Zellner and Park (1987) and Broemeling and Shaarawy (1988). In Section 3, a detailed discussion of the proposed method is presented, including the development of a multivariate regression model adopted by Liu (1994). For examining the precision of the proposed methods, simulated data and one set real data of time series are analyzed numerically. The posterior predictive density is computed respectively by three methods, the t-density mixture method, the partial plug-in method and the path-sampling method. Results of some tail-percentiles and skewness of the estimated density are reported. All these evaluations are included in Section 4. Finally, Section 5 includes the conclusions.

2. The posterior predictive density

In this article, the Bayesian forecast for an AR(p) model with random independent exogenous variables is investigated. Let y_t denote the t-th observation of the dependent variable and let x_t be an $r \times 1$ random vector consisting of exogenous variables. The model is expressed as

$$\text{(2.1)} \qquad \begin{cases} y_t = \alpha + x_t^T \beta + y_{t,p}^T \phi + \epsilon_t, & \text{for } t = p+1, p+2, \ldots, n, \\ x_t = \eta + e_t, & \text{for } t = 1, 2, \ldots, n, \end{cases}$$

where $\beta = (\beta_1, \beta_2, \ldots, \beta_r)^T$, $\phi = (\phi_1, \phi_2, \ldots, \phi_p)^T$ and $y_{t,p} = (y_{t-1}, y_{t-2}, \ldots, y_{t-p})^T$. Here, $\{\epsilon_t\}$ is a set of independent normally distributed random variables

with mean zero and variance τ^{-1}. Similarily, $\{e_t\}$ is a set of independent normally distributed random vectors with mean zero and covariance matrix Σ_x. Also, $\{\epsilon_t\}$ and $\{e_t\}$ are independent.

Zellner and Park (1987) considered Bayesian analysis for a regression model without autoregressive components, that is $\phi = 0$. However, the posterior predictive density function even only one-step ahead can not be obtained in a closed form. As usual, a common problem in Bayesian analysis is met with, where the posterior predictive density involves multiple integrations. Therefore, how to obtain an explicit approximation of the predictive density is of interest. Before proposing a suitable approach, the expression of the exact posterior predictive density is first straightforwardly derived.

Suppose that $\{y_t\}_{t=1}^n$ and $\{x_t\}_{t=1}^n$ are observed, then for convenience a matrix form of formula (2.1) is written as

$$\begin{cases} Y_n = W_n\mu + \varepsilon_n, \\ X_n = J_n \otimes \eta^T + \upsilon_n, \end{cases}$$

where $Y_n = (y_{p+1}, y_{p+2}, \ldots, y_n)^T$, $W_n = (w_{p+1}, w_{p+2}, \ldots, w_n)^T$, $X_n = (x_1, x_2, \ldots, x_n)^T$, $w_j = (1, x_j^T, y_{j,p}^T)^T$, $\mu = (\alpha, \beta^T, \phi^T)^T$, $\varepsilon_n = (\epsilon_{p+1}, \epsilon_{p+2}, \ldots, \epsilon_n)^T$, $\upsilon_n = (e_1, e_2, \ldots, e_n)^T$ and J_n is an $n \times 1$ vector with all entries equal to 1. Moreover, the symbol $\otimes$ denotes the Kronecker product.

The likelihood function for (μ, τ) and (η, Σ_x) is

$$p(Y_n \mid X_n, \mu, \tau)p(X_n \mid \eta, \Sigma_x) \quad \text{with}$$

$$p(Y_n \mid X_n, \mu, \tau) \propto \tau^{m_1/2} \exp\left\{-\frac{\tau}{2}(Y_n - W_n\mu)^T(Y_n - W_n\mu)\right\} \quad \text{and}$$

$$p(X_n \mid \eta, \Sigma_x) \propto |\Sigma_x|^{-n/2} \exp\left\{-\frac{1}{2}\operatorname{tr}[(X_n - J_n \otimes \eta^T)^T(X_n - J_n \otimes \eta^T)\Sigma_x^{-1}]\right\},$$

where $m_1 = n - p$. Suppose the prior information between (μ, τ) and (η, Σ_x) is independent and the following prior pdf's for the parameters are set up, say

$$p(\mu, \tau, \eta, \Sigma_x) = p_1(\mu \mid \tau)p_2(\tau)p_3(\eta \mid \Sigma_x)p_4(\Sigma_x). \tag{2.2}$$

Here, each p_i is defined respectively as follows: p_1 is the pdf for a normal distribution, say $N(\mu_0, (\tau Q_0)^{-1})$, where Q_0 is a positive definite symmetric matrix; p_2 follows a gamma distribution with parameters a and b (Gamma(a, b)), say $p_2(\tau) \propto \tau^{a-1}\exp(-b\tau)$, where $a > 0$ and $b > 0$; p_3 is the pdf for a normal distribution, say $N(\eta_0, \Sigma_x)$; finally, p_4 follows an inverted Wishart distribution denoted by $IW(G_0, \nu_0, r)$, where G_0 is an $r \times r$ positive definite symmetric matrix and $\nu_0 > r$.

Adopting results from Broemeling and Shaarawy (1988), the posterior mean of μ, denoted by μ_n^*, is expressed as

$$\mu_n^* = \mathrm{E}(\mu \mid X_n, Y_n) = A_n^{-1}(Q_0\mu_0 + W_n^T Y_n), \quad \text{where} \tag{2.3}$$

$$A_n = W_n^T W_n + Q_0. \tag{2.4}$$

Moreover, the posterior distribution of τ and $\mu \mid \tau$ are respectively expressed as follows:

$$\tau \mid X_n, Y_n \sim \text{Gamma}\left(\frac{1}{2}(m_1 + 2a), \frac{1}{2}R_n\right) \quad \text{and} \tag{2.5}$$

$$\mu \mid X_n, Y_n, \tau \sim \text{Normal}\ (\mu_n^*, (\tau A_n)^{-1}), \tag{2.6}$$

where $R_n = Y_n^T Y_n + \mu_0^T Q_0 \mu_0 - \mu_n^{*T} A_n \mu_n^* + 2b$. Thus the posterior density of μ is a multivariate t-density written as:

$$p(\mu \mid X_n, Y_n) \propto \left\{1 + \frac{(\mu - \mu_n^*)^T A_n^{*-1} (\mu - \mu_n^*)}{m_1 + 2a}\right\}^{-(m_1+2a+p+r+1)/2}, \tag{2.7}$$

where $A_n^* = R_n A_n^{-1}/(m_1 + 2a)$. Hereafter, for convenience, the density (2.7) will be denoted by "$t(\mu_n^*, A_n^*, m_1 + 2a)$", and the conditional posterior predictive density of y_{n+1} given x_{n+1} is a univariate $t(w_{n+1}^T \mu_n^*, s_n, m_1 + 2a)$ distribution, where $s_n = (1 + w_{n+1}^T A_n^{-1} w_{n+1}) R_n/(m_1 + 2a)$.

On the other hand, the posterior predictive density for any future x_f is also a multivariate t-density with $n + \nu_0 + 1 - r$ degrees of freedom expressed as

$$p(x_f \mid X_n) \propto \left\{1 + \frac{(x_f - \eta_n)^T C_n^{-1} (x_f - \eta_n)}{n + \nu_0 + 1 - r}\right\}^{-(n+\nu_0+1)/2}, \tag{2.8}$$

where $C_n = (n + 2)B_n/\{(n + 1)(n + \nu_0 + 1 - r)\}$, $B_n = S_x + G_0 + n(\bar{x}_n - \eta_0)(\bar{x}_n - \eta_0)^T/(n + 1)$, $S_x = \sum_{t=1}^n (x_t - \bar{x}_n)(x_t - \bar{x}_n)^T$ and $\bar{x}_n = \sum_{t=1}^n x_t/n$, $\eta_n = (\eta_0 + n\bar{x}_n)/(n + 1)$. Thus the unconditional posterior predictive density of y_{n+1} is

$$p(y_{n+1} \mid X_n, Y_n) = \int p(y_{n+1} \mid x_{n+1}, X_n, Y_n) p(x_{n+1} \mid X_n) dx_{n+1}.$$

As mentioned before, there is no closed form for this integration, even if $\phi = 0$ is set. Actually, the integration will be more complicated if a further-step prediction is of interest, instead of only a one-step ahead prediction. In the following section, approximate methods are proposed to provide estimators for any k-step posterior predictive density, where k is any fixed positive integer.

3. Approximations for the posterior predictive density

In this section, some suitable approximations of the k-step posterior predictive density are considered. Instead of using the procedure demonstrated in Section 2, the k-step posterior predictive density can be represented as

$$p(y_{n+k} \mid X_n, Y_n) = \int p(y_{n+k} \mid X_n, Y_n, \theta) p(\theta \mid X_n, Y_n) d\theta,$$

where θ contains ϕ, β, α, η, τ and Σ_x. Since the integration can not be carried out completely for θ perhaps a partial integration is obtainable. Separating θ into two components, say θ_1 and θ_2, the posterior predictive density is then rewritten as

$$p(y_{n+k} \mid X_n, Y_n) = \int \left[\int p(y_{n+k} \mid X_n, Y_n, \theta_1, \theta_2) p(\theta_2 \mid X_n, Y_n, \theta_1) d\theta_2 \right] \cdot p(\theta_1 \mid X_n, Y_n) d\theta_1. \tag{3.1}$$

If the inner integration, denoted by $p(y_{n+k} \mid X_n, Y_n, \theta_1)$, can be carried out explicitly, the predictive density can be estimated by

$$\widehat{p(y_{n+k} \mid X_n, Y_n)} = \frac{1}{L} \sum_{l=1}^{L} p(y_{n+k} \mid X_n, Y_n, \theta_1^{(l)}),$$

where $\theta_1^{(l)}$ is sampled from $p(\theta_1 \mid X_n, Y_n)$. Therefore, our main goal is to find a suitable θ_1 set and an explicit expression of $p(y_{n+k} \mid X_n, Y_n, \theta_1)$ so that the above estimate can be obtained. The defining θ_1 set depends upon the nature of the conditional (given θ) predictive density function $p(y_{n+k} \mid X_n, Y_n, \theta)$. Formally, it is expressed as

$$p(y_{n+k} \mid X_n, Y_n, \theta) = \int \cdots \int p(y_{n+k}, x_{n+k}, y_{n+k-1}, x_{n+k-1}, \ldots, y_{n+1}, x_{n+1} \mid X_n, Y_n, \theta) dx_{n+k} dy_{n+k-1} dx_{n+k-1} \cdots dy_{n+1} dx_{n+1},$$

where $p(y_{n+k}, x_{n+k}, y_{n+k-1}, x_{n+k-1}, \ldots, y_{n+1}, x_{n+1} \mid X_n, Y_n, \theta) = \prod_{j=1}^{k} p(y_{n+j} \mid X_{n+j}, Y_{n+j-1}, \theta) p(x_{n+j} \mid \theta)$, the products of univariate normal density. To avoid the tedious multiple integrations, a recursive algorithm adopted from the developments of Chow (1974) and Liu (1994), by substituting $y_{n+1}, y_{n+2}, \ldots, y_{n+k-1}$ by past values $y_n, y_{n-1}, \ldots$ of the Y_n process is presented. The advantage of this is that conditional to θ the predictive density $p(y_{n+k} \mid X_n, Y_n, \theta)$ is a normal density function that can be written down explicitly and will help to conveniently select a suitable θ_1 set. The recursive representation of formula (2.1) is summarized as follows.

Define

$$y_{t+k}^* - x_{t+k}^* = a_k \alpha + \sum_{j=0}^{k-1} d_j \epsilon_{t+k-j},$$

where $y_{t+k}^* = y_{t+k} - \sum_{j=1}^{p} c_{k-1,j} y_{t+1-j}$, $x_{t+k}^* = \tilde{x}_{t+k}^T \beta$, $\tilde{x}_{t+k} = \sum_{j=0}^{k-1} d_j x_{t+k-j}$ and $a_k = \sum_{j=0}^{k-1} d_j$. Here, $c_{i,j}$'s and d_j's satisfy the following recursive formula: $d_j = c_{j-1,1}$, $c_{i,j} = c_{i-1,1}\phi_j + c_{i-1,j+1}$, for $i \geq 0$, $j \geq 1$ and $\phi_j = 0$, for $j \geq p+1$, with the initial condition

$$c_{-1,j} = \begin{cases} 1, & \text{if } j = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Moreover, let

$$z^*_{t+k} = \begin{pmatrix} y^*_{t+k} - x^*_{t+k} \\ \tilde{x}_{t+k} \end{pmatrix}, \quad \zeta = \begin{pmatrix} \alpha \\ \eta \end{pmatrix} \quad \text{and} \quad u_{t+k} = \begin{pmatrix} \sum_{j=0}^{k-1} d_j \epsilon_{t+k-j} \\ \sum_{j=0}^{k-1} d_j e_{t+k-j} \end{pmatrix},$$

then the coupled model expressed by formula (2.1) can be represented as

$$z^*_{t+k} = a_k \zeta + u_{t+k}. \tag{3.2}$$

Combining all the data together, the overall representation is

$$Z^*_n = a_k (J_{m_2} \otimes \zeta) + U_n, \tag{3.3}$$

where $Z^*_n = ({z^*_{p+k}}^T, {z^*_{p+k+1}}^T, \ldots, {z^*_n}^T)^T$, $U_n = (u^T_{p+k}, u^T_{p+k+1}, \ldots, u^T_n)^T$ and $m_2 = m_1 - k + 1$. In detail, U_n can be written as $U_n = F\Gamma$, where $F = D \otimes I_{r+1}$, $\Gamma = (\gamma^T_{p+1}, \gamma^T_{p+2}, \ldots, \gamma^T_n)^T$, $\gamma_t = (\epsilon_t, e^T_t)^T$ and D is an $m_2 \times m_1$ matrix defined as follows:

$$D = \begin{bmatrix} d_{k-1} & d_{k-2} & \cdots & d_1 & d_0 & 0 & 0 & \cdots & 0 \\ 0 & d_{k-1} & \cdots & d_2 & d_1 & d_0 & 0 & \cdots & 0 \\ & & & & \vdots & & & & \\ 0 & 0 & \cdots & 0 & 0 & 0 & d_{k-1} & \cdots & d_0 \end{bmatrix}.$$

Since γ_t's are independent and identical normally distributed, say $N(0, \Sigma)$, then Γ is also normally distributed with mean 0 and covariance matrix $I_{m_1} \otimes \Sigma$. Therefore, U_n is normally distributed with mean 0 and covariance matrix $\operatorname{cov}(U_n, U_n) = DD^T \otimes \Sigma$. Then the likelihood function of $Z_n = a_k^{-1} Z^*_n$ is expressed as

$$p(Z_n \mid \zeta, \Sigma) \propto |E^{-1} \otimes \Sigma|^{-m_2/2} \cdot \exp\left\{ -\frac{1}{2} (Z_n - J_{m_2} \otimes \zeta)^T (E^{-1} \otimes \Sigma)^{-1} (Z_n - J_{m_2} \otimes \zeta) \right\}, \tag{3.4}$$

where E is an $m_2 \times m_2$ matrix defined by $E = a_k^2 (DD^T)^{-1}$. Based on formula (3.3), Z_n is not only a function of $\{y_t\}$ and $\{x_t\}$, it is also a function of ϕ and β. Moreover, the covariance matrix of U_n involves both Σ and ϕ. To construct formula (3.1), θ_1 and θ_2 are defined respectively as $\theta_1 = (\phi^T, \beta^T)^T$ and θ_2 contains ζ and Σ. Furthermore, the conjugate prior density functions for ζ and Σ are assumed to be as follows:

(1) $p(\zeta \mid \Sigma)$ is normally distributed, say $N(\zeta_0, \Sigma)$, where $\zeta_0 = (\alpha_0, \eta_0^T)^T$.

(2) $p(\Sigma)$ has an inverted Wishart distribution denoted by $IW(G, \nu, r+1)$, where G is an $(r+1) \times (r+1)$ positive definite symmetric matrix and $\nu > r + 1$.

Then the posterior predictive density for y_{n+k}, given θ_1, can be obtained explicitly via the following assertion (the proof is given in the Appendix).

THEOREM 3.1. *Under the prior assumptions stated as above, the posterior predictive density for y_{n+k} given θ_1 is a multivariate t-density expressed as*:

$$p(y_{n+k} \mid X_n, Y_n, \theta_1) \propto \left\{ 1 + \frac{s(y_{n+k} - m_{n+k})^2}{m_2 + \nu - r} \right\}^{-(m_2+\nu-r+1)/2}, \tag{3.5}$$

where $\theta_1 = (\phi^T, \beta^T)^T$, $m_{n+k} = \sum_{j=1}^{p} c_{k-1,j} y_{n+1-j} + h^T \bar{z}_0$, $h = (1, \beta^T)^T$, $\bar{z}_0 = a_k(e\bar{z} + \zeta_0)/(e+1)$, $\bar{z} = \sum_{i=1}^{m_2}\sum_{j=1}^{m_2} e_{ij} z_{i+p+k-1}$, $e_{ij} = (E)_{ij}$, $e = J_{m_2}^T E J_{m_2}$, $s = \{h^T H h[b_k^2 + a_k^2/(e+1)]/(m_2+\nu-r)\}^{-1}$, $b_k^2 = \sum_{j=0}^{k-1} d_j^2$, $H = S_z + G + e(\bar{z} - \zeta_0)(\bar{z} - \zeta_0)^T/(e+1)$ *and* $S_z = \sum_{i=1}^{m_2}\sum_{j=1}^{m_2} e_{ij}(z_{i+p+k-1} - \bar{z})(z_{j+p+k-1} - \bar{z})^T$.

The remaining problem is to explicitly express the density function $p(\theta_1 \mid X_n, Y_n)$. By formula (2.7) the posterior density of θ_1 is

$$p(\theta_1 \mid X_n, Y_n) \propto \left\{1 + \frac{(\theta_1 - \theta_1^*)^T \tilde{A}_n^{-1}(\theta_1 - \theta_1^*)}{m_1 + 2a}\right\}^{-(m_1+2a+p+r)/2}, \tag{3.6}$$

θ_1^* is the last $(p+r)$ component of μ_n^* and $\tilde{A}_n$ is the lower $(p+r)\times(p+r)$ right-hand matrix of A_n^*. Therefore, the posterior predictive density of y_{n+k} can be estimated by

$$\widehat{p(y_{n+k} \mid X_n, Y_n)} = \frac{1}{L}\sum_{l=1}^{L} p(y_{n+k} \mid X_n, Y_n, \theta_1^{(l)}),$$

where $\{\theta_1^{(l)}\}_{l=1}^{L}$ are independently sampled from density (3.6) and $p(y_{n+k} \mid X_n, Y_n, \theta_1^{(l)})$ is given by formula (3.5).

In particular, for computational convenience we may set $L = 1$ and $\theta_1^{(1)} = \theta_1^*$, that is by plugging in the posterior mean of θ_1 and approximating the posterior predictive density using a multivariate t-density. Since the posterior mean θ_1^* is a consistent estimator of θ_1, the posterior predictive density provided by this plug-in method will converge to the true one when the sample size is sufficiently large. Adopting the aforementioned development, if ζ is moved from θ_2 to θ_1 and this new set denoted by $\theta_{(1)}$, then the posterior predictive density of y_{n+k} given $\theta_{(1)}$ is still a multivariate t-density expressed as follows.

LEMMA 3.1. *Under the inverted Wishart prior assumption of* Σ, *the posterior predictive density for* y_{n+k} *given* $\theta_{(1)}$ *is*

$$p(y_{n+k} \mid X_n, Y_n, \theta_{(1)}) \propto \left\{1 + \frac{s^*(y_{n+k} - m_{n+k}^*)^2}{m_2 + \nu - r}\right\}^{-(m_2+\nu-r+1)/2},$$

where $\theta_{(1)} = (\phi^T, \beta^T, \zeta^T)^T$, $m_{n+k}^* = m_{n+k} + h^T(a_k\zeta_1 - \bar{z}_0)$, $s^* = \{b_k^2 h^T H^* h/(m_2 + \nu - r)\}^{-1}$, $H^* = H + e\{(\bar{z} - \zeta_1)(\bar{z} - \zeta_1)^T - (\bar{z} - \zeta_0)(\bar{z} - \zeta_0)^T/(e+1)\}$ *and* ζ_1 *is the given value of* ζ. *Some notations such as* m_{n+k}, h, $\bar{z}_0$, b_k^2, e, H, $\bar{z}$ *and* ζ_0 *are stated in Theorem* 3.1.

Especially, if one sets $\zeta_1 = \zeta_0 = \bar{z}$, the following reductions are obtained: $m_{n+k} = m_{n+k}^*$, $H = H^*$ and $s^{-1} - (s^*)^{-1} = a_k^2 h^T H h/\{(e+1)(m_2+\nu-r)\}$, which is a positive value. Therefore, the two densities $p(y_{n+k} \mid X_n, Y_n, \theta_1)$ and $p(y_{n+k} \mid X_n, Y_n, \theta_{(1)})$ have the same mean and the same degrees of freedom, however the variation of the latter is smaller. More specifically, this is true if the

θ_1 set containing all the model's parameters, denoted by $\theta_{\{1\}}$, based on formula (3.2), z^*_{n+k} is a multivariate normal distribution with mean $a_k\zeta$ and covariance matrix $b_k^2\Sigma$. Therefore, the posterior predictive density for y_{n+k} is simplified by the following lemma.

LEMMA 3.2. *The posterior predictive density for y_{n+k} given $\theta_{\{1\}}$ is*

$$p(y_{n+k} \mid X_n, Y_n, \theta_{\{1\}}) \propto \xi_k^{-1} \exp\left\{-\frac{(y_{n+k} - m^*_{n+k})^2}{2\xi_k^2}\right\},$$

where $\theta_{\{1\}} = \{\phi, \beta, \zeta, \Sigma\}$ and $\xi_k^2 = b_k^2 h^T \Sigma h$.

Comparing the two densities introduced in Lemmas 3.1 and 3.2, we find that they have the same mean, however the former has a t-density which is heavier tailed than the latter, a normal density. If the given Σ matrix satisfies $h^T\Sigma h = h^T H^* h/(m_2 + \nu - r)$, then the two densities are almost the same if the degrees of freedom for the t-distribution are large enough. For example, restrict Σ as a diagonal matrix with τ^{-1} as the 1×1 upper left-hand matrix and Σ_x as the $r \times r$ lower right-hand matrix and define $\tau^{-1} = (h_{11} + 2h_{12}\beta)/(m_2 + \nu - r)$ and $\Sigma_x = h_{22}/(m_2 + \nu - r)$, where h_{11}, h_{12} and h_{22} are respectively, the 1×1 upper left-hand entry, the $1 \times (r-1)$ upper right-hand row and the $r \times r$ lower right-hand matrix of H^*. As expected, more elements are included in θ_1, the predictive density becomes less accurate. Though by substituting a consistent estimator of $\theta_{(1)}$ (or $\theta_{\{1\}}$), the conditional predictive density still converges to the true density, we would usually make the θ_1 set as small as possible. Thus in the following discussion $\theta_1 = (\phi^T, \beta^T)^T$ and this procedure is designated as the "partial plug-in method".

In the meantime, an extension of the so-called path-sampling method proposed by Thompson and Miller (1986) can also be used to estimate the desired predictive density. This is summarized by the following four steps:

Step 1. Sample k future values for x, say $x_{n+1}, x_{n+2}, \ldots, x_{n+k}$, independently from the multivariate t-distribution described by formula (2.8).

Step 2. Sample a value for τ from the gamma distribution described by formula (2.5).

Step 3. Sample a value for μ, conditional on the τ obtained from Step 2 from the normal distribution given by formula (2.6).

Step 4. Using the selected values, $\{x_{n+i}\}_{i=1}^k$, τ and μ respectively, in Steps 1, 2 and 3, a path k periods long, say $\{y_{n+1}, y_{n+2}, \ldots, y_{n+k}\}$ is simulated based on the first part of formula (2.1),

$$y_{n+i} = \alpha + x_{n+i}^T\beta + y_{n+i,p}^T\phi + \epsilon_{n+i}, \qquad i = 1, 2, \ldots, k,$$

where ϵ_{n+i} is i.i.d. $N(0, \tau^{-1})$ distributed, τ is the one sampled in Step 2 and again $\mu = (\alpha, \beta^T, \phi^T)^T$.

After a sufficiently large number of paths are generated, a bundle of simulated observations in each future step is obtained. Therefore, the posterior predictive

density for any future step can be estimated on the basis of these simulated observations. The main sampling scheme for the path-sampling method is that future observations should be simulated one by one sequentially. In the next section, numerical comparisons among the t-density mixture method, the partial plug-in method and the path-sampling method are performed with both simulated data and real data.

4. Numerical illustrations

In this section, numerical examinations are conducted to evaluate the performance of the three methods introduced in Section 3. The partial plug-in method is especially rather simple and interesting and can be used analytically. For simplicity, most of these hyper-parameters are substituted with ordinary least squares estimates stated as follows:

(1) $\mu_0 = (W_n^T W_n)^{-1} W_n^T Y_n$,
(2) α_0 is the first component of μ_0,
(3) $\eta_0 = \bar{x}_n$,
(4) $Q_0 = W_n^T W_n / q_0$, where q_0 is the 1×1 upper left-hand entry of $W_n^T W_n$,
(5) $a = (m_1 - p - r - 1)/2$,
(6) $b = \frac{1}{2} Y_n^T [I - W_n (W_n^T W_n)^{-1} W_n^T] Y_n$,
(7) $\nu = 2a + r$,
(8) $\nu_0 = \nu - 1$ and
(9) G is a diagonal matrix with $2b$ as the 1×1 upper left-hand matrix and G_0 as the $r \times r$ lower right-hand matrix such that $G_0 = \sum_{t=1}^{n} (x_t - \bar{x}_n)(x_t - \bar{x}_n)^T$.

For the simulated data, two types of model are examined including stationary and nonstationary components in the autoregressive part. For simplicity, the models are set up with $p = r = 1$,

$$\begin{cases} y_t = 0.3 + 0.5 x_t + \phi y_{t-1} + \epsilon_t, \\ x_t = 0.1 + e_t, \end{cases}$$

where $\phi = 0.5$ and 1. And the $\{(\epsilon_t, e_t^T)^T\}$ are assumed to be independent and identically normally distributed with mean 0 and a diagonal covariance matrix, with $\tau^{-1} = 1$ and $\Sigma_x = .5$. The influence of sample size for each method is investigated by designating n =50 and 100 for each model. In order to obtain a reliable impact on sample size, 350 observations are generated for each model, the last n observations are then used for analysis. Besides the simulated data, Lydia Pinkham's vegetable compound sales and advertising data, annual data from 1907 to 1960 (Vandaele (1983)), is used for investigation. A transfer function model is suggested in this book (p. 328) and expressed as

$$y_t = (\omega_0 - \omega_1 B) z_t + (1 - \phi B)^{-1} \epsilon_t,$$

where $y_t = \text{sales}_t - \text{sales}_{t-1}$, $z_t = \text{advertising}_t - \text{advertising}_{t-1}$ and $B z_t = z_{t-1}$. Also $\{\epsilon_t\}$ is a sequence of i.i.d. normally distributed random variables. To satisfy the assumptions of our discussed model, the following models are suggested.

$$\begin{cases} y_t = \alpha + \beta x_t + \phi y_{t-1} + \epsilon_t, \\ x_t = \eta + e_t, \end{cases}$$

Table 1. Inferences for the posterior predictive densities.

$$\begin{cases} y_t = 0.3 + 0.5x_{t-1} + \phi y_{t-1} + \epsilon_t, \\ x_t = 0.1 + e_t. \end{cases}$$

Percentiles

(a) $\phi = 0.5$, $n = 50$

step	5%					95%				
1	-0.33^a	-0.35^b	$(0.01)^c$	-0.35^d	$(0.02)^e$	3.22	3.23	(0.01)	3.21	(0.02)
2	−0.69	−0.71	(0.01)	−0.69	(0.02)	3.37	3.41	(0.01)	3.38	(0.03)
3	−0.83	−0.87	(0.01)	−0.80	(0.03)	3.33	3.39	(0.02)	3.41	(0.03)
4	−0.88	−0.90	(0.01)	−0.90	(0.03)	3.36	3.39	(0.02)	3.41	(0.03)

(b) $\phi = 0.5$, $n = 100$

step	5%					95%				
1	-0.42^a	-0.43^b	$(0.01)^c$	-0.42^d	$(0.02)^e$	3.17	3.18	(0.00)	3.17	(0.02)
2	−0.76	−0.77	(0.00)	−0.76	(0.02)	3.27	3.29	(0.01)	3.26	(0.03)
3	−0.89	−0.90	(0.01)	−0.88	(0.03)	3.26	3.28	(0.01)	3.27	(0.03)
4	−0.95	−0.97	(0.01)	−0.93	(0.03)	3.23	3.26	(0.02)	3.27	(0.03)
5	−0.96	−0.97	(0.01)	−0.95	(0.03)	3.24	3.28	(0.01)	3.24	(0.03)
6	−0.99	−1.03	(0.02)	−0.98	(0.03)	3.21	3.26	(0.02)	3.23	(0.03)

(c) $\phi = 1.0$, $n = 50$

step	5%					95%				
1	168.2^a	168.1^b	$(0.03)^c$	168.2^d	$(0.02)^e$	171.8	171.9	(0.04)	171.8	(0.02)
2	168.0	167.9	(0.06)	167.8	(0.04)	173.1	173.3	(0.07)	173.2	(0.03)
3	167.9	167.4	(0.12)	167.7	(0.04)	174.1	174.6	(0.12)	174.4	(0.04)
4	168.0	167.1	(0.13)	167.6	(0.05)	175.2	175.6	(0.15)	175.5	(0.05)

(d) $\phi = 1.0$, $n = 100$

step	5%					95%				
1	168.3^a	168.3^b	$(0.02)^c$	168.3^d	$(0.02)^e$	171.9	171.9	(0.02)	171.9	(0.02)
2	168.1	168.1	(0.04)	168.1	(0.03)	173.3	173.3	(0.04)	173.3	(0.03)
3	168.2	168.0	(0.06)	168.1	(0.04)	174.4	174.6	(0.07)	174.6	(0.04)
4	168.3	168.1	(0.07)	168.2	(0.05)	175.6	175.8	(0.08)	175.8	(0.05)
5	168.4	168.3	(0.10)	168.2	(0.06)	176.7	177.1	(0.10)	176.9	(0.06)
6	168.7	168.4	(0.11)	168.3	(0.07)	177.7	178.2	(0.14)	178.1	(0.07)

where $x_t = z_t - 0.091z_{t-1} + 0.411z_{t-2}$. Also, $\{e_t\}$ are i.i.d. normally distributed and independent of $\{\epsilon_t\}$.

As usual, it is impossible to expect precise longer period forecasts as the sample size is too small. Hence, the length of the forecast period is adjusted by the sample size, four periods ($k = 4$) as $n = 50$ and six periods ($k = 6$) as $n = 100$ are discussed. The estimated predictive density obtained by the t-density mixture method is obtained by setting $L = 100$ and the path-sampling technique is estimated by 10,000 bundles of paths. For each data set, some specified percentiles for each estimated predictive density are computed via the three methods. Since

Table 1. (continued).

Skewness

(a) $n = 50$

step	$\phi = 0.5$				$\phi = 1.0$			
1	0.00[b]	(0.00)[c]	−0.01[d]	(0.03)[e]	−0.02	(0.02)	0.00	(0.03)
2	0.01	(0.01)	0.01	(0.03)	0.02	(0.02)	0.01	(0.03)
3	0.01	(0.01)	0.01	(0.03)	0.06	(0.04)	0.03	(0.03)
4	0.01	(0.01)	0.01	(0.03)	−0.03	(0.05)	0.05	(0.03)

(b) $n = 100$

step	$\phi = 0.5$				$\phi = 1.0$			
1	0.00[b]	(0.00)[c]	0.00[d]	(0.03)[e]	0.00	(0.00)	0.00	(0.02)
2	0.00	(0.01)	0.01	(0.03)	0.00	(0.01)	0.01	(0.02)
3	0.01	(0.00)	0.01	(0.02)	0.01	(0.01)	0.01	(0.03)
4	0.01	(0.01)	0.00	(0.02)	0.03	(0.01)	0.02	(0.02)
5	0.00	(0.01)	0.00	(0.02)	0.01	(0.02)	0.02	(0.02)
6	0.00	(0.01)	0.00	(0.02)	0.06	(0.03)	0.04	(0.03)

a: partial plug-in method.
b: t-density mixture method.
c: standard error for t-density mixture method.
d: path-sampling method.
e: standard error for path-sampling method.

the standard errors for each estimate produced by the path-sampling method for simulated data of under 100 replications are rather small, presumbly 10,000 path bundles should be sufficient. On the other hand, for a stationary situation, the standard errors under 100 replications for the t-density mixture method indicate that $L = 100$ is sufficient. However, for models with nonstationary autoregressive components and small sample size, say $n = 50$, the standard errors for estimates are rather large. As values of L increase, up to 400, the magnitude of the standard error is almost the same as that for computing by the path-sampling method. Since the percentiles are rather close around to the center for all the three methods, only the tail-percentiles, the lower 5% and upper 95% are reported. Moreover, when sample sizes are large, say $n = 300$, the results of the three methods almost coincide, thus only $n = 50$ and $n = 100$ are reported in Table 1.

Intuitively, the partial plug-in method ignores the variation of the parameter vector θ_1, the predictive density should tend to be less variant. Also, the symmetric property of the t-density may limit the shape of the posterior predictive density. The impact of these two shortages could be investigated by comparing results with those obtained by the t-density mixture method and the path-sampling method. By examining the skewness as exhibited in Table 1, the results indicate that the symmetric restriction is rather reasonable. The 5% and 95% percentiles will roughly provide information about the dispersion of a density, while the results show that the lesser variation of the partial plug-in method is not too serious. Again, as sample sizes get larger, those limitations gradually disappear. For real data, most of the standard errors are rather large, especially those produced by

Table 2. Inferences for the posterior predictive densities.

Lydia pinkham vegetable data
Percentiles

step	5%		25%		50%		75%		95%	
1	−398[a]		−175		−20.7		133		358	
	−401[b]	(1.3)[c]	−176	(1.1)	−20.3	(0.9)	136	(1.3)	363	(2.9)
	−407[d]	(5.1)[e]	−181	(3.0)	−25.5	(2.9)	129	(3.0)	352	(5.0)
2	−392		−160		0.6		161		393	
	−401	(2.0)	−165	(1.1)	−2.0	(0.7)	160	(0.9)	396	(2.1)
	−399	(5.1)	−167	(3.6)	−0.7	(3.2)	159	(3.6)	399	(5.2)
3	−393		−158		3.5		165		399	
	−402	(1.7)	−162	(1.5)	1.9	(0.8)	166	(0.9)	405	(2.2)
	−396	(4.9)	−163	(3.5)	1.2	(2.9)	163	(3.3)	402	(4.1)
4	−397		−162		0.6		163		398	
	−405	(2.0)	−165	(1.6)	−0.1	(0.7)	164	(1.0)	405	(2.7)
	−400	(4.8)	−160	(2.9)	−1.5	(2.9)	162	(3.2)	405	(5.3)

step	Skewness			
1	0.01[b]	(0.00)[c]	−0.04[d]	(0.02)[e]
2	0.00	(0.00)	0.04	(0.03)
3	0.00	(0.00)	0.00	(0.03)
4	0.00	(0.00)	0.04	(0.02)

a: partial plug-in method.
b: t-density mixture method.
c: standard error for t-density mixture method.
d: path-sampling method.
e: standard error for path-sampling method.

the path-sampling method. The reason for this phenomenon may be that the sample size is too small ($n = 51$) or the fitted model deviates slightly from the true model. Though the results reported in Table 2 for the three methods are slightly different, they are almost indifferent statistically. All the discussed results were done on a VAX9320 computer at the National Central University in Taiwan. The data were generated from the DRNNOR, DRNGAM and DRNMVN subroutines in the IMSL package.

To summarize, the simulated experimental results show that under the normal-gamma and normal-inverted Wishart prior assumptions, the posterior predictive density for future steps so far considered can be almost simply approached by a suitable t-density. Though the t-density approximation has some limitations, the discrepancy is not too serious and disappears as the sample size increases. Moreover, since the posterior means of both β and ϕ are consistent estimators, the density produced by the partial plug-in method will converge to the true one as the sample size gets larger. In practical usage, it is recommended that one treats

approximately the predictive density as a suitable t-density. Moreover, for a more accurate approximation, a t-density mixture method will be another suggested approach.

5. Conclusions

This article investigates the Bayesian multiperiod forecasting problem for AR(p) models with random independent exogenous regressors under normal-gamma and normal-inverted Wishart prior assumptions. Methods are proposed to produce estimators for the posterior predictive density of any future observation. For any future step, the posterior predictive density can be estimated by a mixture of t-density functions or simply by a suitable t-density. The accuracy of the proposed methods has been compared with the path-sampling method by examining some simulated and some real data. For each series, some specified tail-percentiles and the skewness of the posterior predictive densities, up to six-step-ahead forecasts are computed respectively, by using the t-density mixture method, the partial plug-in method and the path-sampling method. Though the partial plug-in method provides a symmetric and less variant density, it seems these limitations do not cause too serious a discrepancy. The numerical results show that the three methods produce almost the same estimates. Furthermore, the partial plug-in method conveniently provides an explicitly estimated predictive density which asymptotically converges to the true one as the sample size gets larger.

Acknowledgements

The author wishes to thank the two anonymous referees for their valuable and constructive comments which led to a considerable improvement in this article.

Appendix

PROOF OF THEOREM 3.1. Combining the likelihood function (formula (3.4)) with the normal-inverted Wishart prior distributions together, we have

$$\begin{aligned} P(\zeta, \Sigma \mid Z_n) &\propto P(Z_n \mid \zeta, \Sigma) P(\zeta \mid \Sigma) P(\Sigma) \\ &\propto |\Sigma|^{-(m_2+\nu+r+3)/2} \exp\Big\{ -\frac{1}{2} \operatorname{tr}[S_z + G + e(\bar{z}-\zeta)(\bar{z}-\zeta)^T \\ &\qquad + (\zeta-\zeta_0)(\zeta-\zeta_0)^T]\Sigma^{-1} \Big\}, \end{aligned}$$

where S_z, e and $\bar{z}$ are stated in Theorem 3.1. Furthermore, the likelihood function for z_{n+k} is written as

$$P(z_{n+k} \mid Z_n, \zeta, \Sigma) \propto |\Sigma|^{-1/2} \exp\left\{ -\frac{a_k^2}{2b_k^2} (z_{n+k}-\zeta)^T \Sigma^{-1} (z_{n+k}-\zeta) \right\}.$$

After algebra, the posterior predictive density for z_{n+k} is

$$P(z_{n+k} \mid Z_n) \propto \left\{1 + \frac{(z_{n+k} - \tilde{z_0})^T Q_n^{-1} (z_{n+k} - \tilde{z_0})}{m_2 + \nu - r}\right\}^{-(m_2+\nu+1)/2},$$

where $Q_n = \{b_k^2/a_k^2 + 1/(e+1)\}H/(m_2 + \nu - r)$, $\tilde{z_0} = \bar{z_0}/a_k$ and b_k^2, $\bar{z_0}$ and H are stated in Theorem 3.1. Then z_{n+k} has a $t(\tilde{z_0}, Q_n, m_2 + \nu - r)$ distribution. Since $z_{n+k}^* = a_k z_{n+k}$ and $h^T z_{n+k}^* = y_{n+k}^*$, thus y_{n+k}^* has a $t(h^T \bar{z_0}, a_k^2(h^T Q_n h), m_2+\nu-r)$ distribution. Furthermore, after algebra $a_k^2(h^T Q_n h) = s^{-1}$, therefore the result for Theorem 3.1 is established.

REFERENCES

Broemeling, L. D. and Shaarawy, S. (1988). Time series: A Bayesian analysis in time domain, *Bayesian Analysis of Time Series and Dynamic Models* (ed. J. S. Spall), 1–21, Marcel Dekker, New York.

Chow, G. C. (1974). Multiperiod predictions from stochastic difference equations by Bayesian methods, *Studies in Bayesian Econometrics and Statistics* (eds. S. E. Fienberg and A. Zellner), 313–324, North-Holland, Amsterdam.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities, *J. Amer. Statist. Assoc.*, **85**, 398–409.

Liu, S. I. (1994). Multiperiod Bayesian forecasts for AR models, *Ann. Inst. Statist. Math.*, **46**, 429–452.

Monahan, J. F. (1983). Fully Bayesian analysis of ARMA time series models, *J. Econometrics*, **21**, 307–331.

Schnatter, S. (1988). Bayesian forecasting of time series by Gaussian sum approximation, *Bayesian Statistics* 3 (eds. J. M. Bernardo, M. H. Degroot, D. V. Lindley and A. F. M. Smith), 757–764, Clarendon Press, Oxford.

Tanner, M. A. (1993). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and the Likelihood Functions*, Springer, New York.

Thompson, P. A. and Miller, R. B. (1986). Sampling the future: A Bayesian approach to forecasting from univariate time series models, *Journal of Business and Economic Statistics*, **4**, 427–436.

Vandaele, W. (1983). *Applied Time Series and Box-Jenkins Models*, Academic Press, New York.

Zellner, A. and Park, S. (1987). Bayesian prediction with random regression, *Specification Analysis in the Linear Model* (eds. M. L. King and D. E. A. Giles), 234–251, Routledge and Kegan Paul, London.