Ann. Inst. Statist. Math.
Vol. 46, No. 3, 429–452 (1994)

MULTIPERIOD BAYESIAN FORECASTS FOR *AR* MODELS

SHU-ING LIU

Graduate Institute of Statistics, National Central University, Chung-Li, Taiwan 32054, R.O.C.

(Received March 1, 1993; revised September 6, 1993)

Abstract. The multiperiod Bayesian forecast under the normal-gamma prior assumption for univariate AR models with strongly exogenous variables is investigated. A two-stage approximate method is proposed to provide an estimator of the posterior predictive density for any future observation in a convenient closed form. Some properties of the proposed method are proven analytically for a one-step ahead forecast. The precision of the proposed method is examined by using some simulated data and two sets of real data up to lead-twelve-ahead forecasts by comparison with a path sampling method. It is found that most of the results for the two discussed methods are rather close for short period forecast. Especially when sample size is sufficiently large, the estimated predictive density provided by the two-stage method asymptotically converges to the true density. A heuristic proof of this asymptotic property is also presented.

Key words and phrases: AR model, Bayesian analysis, posterior mean, predictive density, regression analysis.

1. Introduction

Bayesian analysis of autoregressive-moving average ($ARMA$) models has been extensively applied in many fields. Estimation of the parameters in $ARMA$ models has been investigated since late 1960. However, there are few results for multiperiod forecasting via the Bayesian procedure. Theoretically, in order to obtain a k-th-step posterior predictive density, a $(k-1)$-multiple integration should be computed. This numerical difficulty limits the applications of the Bayesian procedure for multiperiod forecasting. Chow (1974) presented formulations for the estimation of future moments for $AR(1)$ models. Monahan (1983) gave a fully Bayesian treatment of $ARMA$ models. The percentiles of multiperiod (up to lead-five-ahead forecasts) predictive densities for $AR(1)$ models were computed with a numerical method. Geweke (1989) proposed a Monte Carlo integration with an importance sampling technique to compute the posterior expectation as a function of unknown vector parameters.

Using a normal-gamma prior assumption for the parameters of an $AR(p)$

model, Broemeling and Land (1984) obtained some theoretical results for the posterior predictive density. The joint predictive density of k future observations is the product of k univariate t-distributions. Therefore, to obtain the marginal predictive density, a multiple integration should be evaluated. This task is very difficult, especially where a longer future step is concerned. To overcome this limitation, a reasonable approximation is needed. Thompson and Miller (1986) suggested a simulation method to estimate the posterior predictive distribution for any $AR(p)$ model. The future observations are produced by generating the autoregressive parameters based on their corresponding posterior distributions. A bundle of 10,000 paths was created to compute the percentiles of the predictive density. This approach helps to explore the shape of the predictive density for any future observation in a unified manner. The difficulties of computing the integration are overcome through the help of the computer via a simulation.

In this paper, we consider $AR(p)$ models with strongly exogenous variables. From an analytical point of view, a reliable approximation in a closed form is valuable. It will be helpful to further understand the statistical property of the predictive density directly. Sometimes, a rough sketch of the density may be needed analytically, instead dealing with it case by case via simulation. Therefore, instead of performing a multiple integration or simulation, an explicit approximation of the desired predictive density is presented. The parameters in the model are separated into two parts, the first part arising from the $AR(p)$ model and the second part from the regression model. At first, parameters of the $AR(p)$ model are replaced by their posterior means. Then the problem is reduced to a regression model and the predictive result of a regression model can be adapted here. The proposed two-stage method creates a posterior predictive density for any future observation through a close formula. Comparisons with the path sampling method are performed to demonstrate the applicability of the proposed method. Though the approximate method places some restrictions on the shape and variation of the estimated predictive density, numerical results indicate that this limitation is negligible for a short period forecast.

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 Broemeling and Land (1984). In Section 3, a detailed discussion of the proposed method is presented, including the development of a transformed regression model and the two-stage procedure. Some analytical results for the two-stage method are proven for a one-step-ahead forecast. For further step forecasts, simulated data and two sets of time series real data are analyzed numerically. The posterior predictive density is computed respectively by two methods, the path sampling method and the two-stage method. The results show that the two methods are rather close to each other for short period forecasts. For a longer period forecast, the accuracy of the two-stage method increases as the sample size gets larger. All these evaluations are included in Section 4. Finally, a heuristic proof of the asymptotic convergence of the proposed method is also investigated.

2. *AR* models with exogenous variables

2.1 *The model*

In this paper, we consider autoregressive models with a regression component. Let y_t denote the t-th observation of the dependent variable in the time series and x_t denote the vector of strongly exogenous variables with r components. The model is represented by

$$y_t = x_t^T\beta + y_{t,p}^T\varphi + \epsilon_t, \qquad \text{for} \quad t = p+1, p+2, \ldots, n, \tag{2.1}$$

where $\beta^T = (\beta_1, \beta_2, \ldots, \beta_r)$, $\varphi^T = (\phi_1, \phi_2, \ldots, \phi_p)$ and $y_{t,p}^T = (y_{t-1}, y_{t-2}, \ldots, y_{t-p})$. Here, ϵ_t is a sequence of independent variables with normal distributions such that $\mathrm{E}(\epsilon_t) = 0$ and $\mathrm{Var}(\epsilon_t) = \tau^{-1}v_t$. Here v_t's are known constants. For simplicity, we treat x_t as known instead of as a random vector and β_1 is the parameter corresponding to the constant term.

Formula (2.1) can be rewritten in matrix form as

$$Y_n = Z_n\mu + \varepsilon_n, \tag{2.2}$$

where $Y_n^T = (y_{p+1}, y_{p+2}, \ldots, y_n)$, $Z_n^T = (z_{p+1}, z_{p+2}, \ldots, z_n)$, $z_j^T = (x_j^T, y_{j,p}^T)$, $\mu^T = (\beta^T, \varphi^T)$ and $\varepsilon_n^T = (\epsilon_{p+1}, \epsilon_{p+2}, \ldots, \epsilon_n)$. Then ε_n is multivariate normally distributed with mean zero and covariance matrix $(\tau V)^{-1}$, here V is an $(n-p)\times(n-p)$ diagonal matrix with v_t in the diagonal. The likelihood function of (2.2) is expressed as

$$p(Y_n \mid \mu, \tau) = \left(\frac{2\pi}{\tau}\right)^{-m/2} |V|^{1/2} \exp\left\{-\frac{\tau}{2}(Y_n - Z_n\mu)^T V (Y_n - Z_n\mu)\right\},$$

where $m = n - p$. Thus an appropriate prior density is

$$\xi(\mu, \tau) = \xi_1(\mu \mid \tau)\xi_2(\tau), \qquad \tau > 0, \quad \mu \in R^{p+r},$$

where ξ_1 is a normal density $N(\mu_0, (\tau Q)^{-1})$, Q is a positive definite matrix, and ξ_2 is a gamma density with parameters $a > 0$ and $b > 0$. Finally, the joint density of (Y_n, μ, τ) is

$$(2\pi)^{-(m+p+r)/2} |V|^{1/2} |Q|^{1/2} b^a \tau^{\delta-1} \exp\left\{-\frac{\tau}{2}(T_1 + T_2 + 2b)\right\},$$

where $\delta = (m + p + r + 2a)/2$, $T_1 = (Y_n - Z_n\mu)^T V (Y_n - Z_n\mu)$ and $T_2 = (\mu - \mu_0)^T Q (\mu - \mu_0)$.

If one is confident of one's prior information, one would specify the prior parameters μ_0, Q, a and b. On the contrary, one may use the flat prior density, $\xi(\mu, \tau) \propto \tau^{-1}$, $\tau > 0$ and $\mu \in R^{p+r}$. Then the joint density of (Y_n, μ, τ) is expressed as

$$p(Y_n, \mu, \tau) \propto \tau^{m/2-1} \exp\left\{-\frac{\tau}{2}(Y_n - Z_n\mu)^T V (Y_n - Z_n\mu)\right\}. \tag{2.3}$$

2.2 *Posterior distributions*

In this subsection, we consider the posterior distribution of μ, τ and the posterior predictive density of the observed data. Most of the developments follow from Broemeling and Lands' results (1984).

Let $F_n = Z_n^T V Z_n$, $SSR_n = Y_N^T(V - VZ_nF_n^{-1}Z_n^T V)Y_n$, $\hat{\mu}_n = F_n^{-1}Z_n^T V Y_n$, $A_n = F_n + Q$ and $\mu_n^* = A_n^{-1}(Q\mu_0 + Z_n^T V Y_n)$, then $T_1 + T_2 = SSR_n + (\mu - \mu_n^*)^T A_n(\mu - \mu_n^*) + \mu_0^T Q\mu_0 + \hat{\mu}_n^T F_n \hat{\mu}_n - \mu_n^{*T} A_n \mu_n^*$. After algebra, the following posterior distributions can be obtained directly.

Property 2.1. The posterior density of μ is a multivariate t-distribution with $(m + 2a)$ degrees of freedom expressed as

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

where $A_n^* = (m+2a)R_n^{-1}A_n$ and $R_n = SSR_n + \mu_0^T Q\mu_0 + \hat{\mu}_n^T F_n \hat{\mu}_n - \mu_n^{*T} A_n \mu_n^* + 2b$.

Property 2.2. The posterior density of τ is a gamma distribution with parameters $((m + 2a)/2, R_n/2)$ expressed as

$$p(\tau \mid Y_n) \propto \tau^{(m+2a)/2-1} \exp\left\{-\frac{\tau R_n}{2}\right\}.$$

Suppose we have observed a data set $\{y_1, y_2, \ldots, y_n\}$, then the predictive density of a future observation y_{n+1} is described as follows:

Property 2.3. The posterior predictive density of y_{n+1} is a univariate t-distribution with $m + 2a$ degrees of freedom expressed as

$$p(y_{n+1} \mid Y_n) \propto \left\{1 + \frac{s_n(y_{n+1} - \mu_n^{*T} z_{n+1})^2}{m + 2a}\right\}^{-(m+2a+1)/2}, \tag{2.4}$$

where $s_n = (m + 2a)\alpha_n R_n^{-1}$ and $\alpha_n = (1 + \upsilon_{n+1} z_{n+1}^T A_n^{-1} z_{n+1})^{-1}$.

If the flat prior density is used, that is $\xi(\mu, \tau) \propto \tau^{-1}$, then the above stated results are simply replaced by $2a \to -(p + r)$, $Q \to 0$ and $b \to 0$. Moreover, under these conditions, the following relationships hold:

$$R_n = SSR_n, \qquad A_n = F_n \qquad \text{and} \qquad \mu_n^* = \hat{\mu}_n.$$

3. Multiperiod forecasts

3.1 *The exact method*

Instead of considering a one-step-ahead forecast, multiperiod forecasts are investigated in this subsection. Suppose that the data have been observed up to time n, then the joint predictive density of future observations $y_{n+1}, y_{n+2}, \ldots, y_{n+k}$ is obtained with the following relationship:

$$p(y_{n+1}, y_{n+2}, \ldots, y_{n+k} \mid Y_n) = p(y_{n+1} \mid Y_n) \prod_{j=2}^{k} p(y_{n+j} \mid y_{n+j-1}, \ldots, y_{n+1}, Y_n).$$

Each density on the right-hand side is a univariate t-distribution described by Property 2.3. Thus the posterior predictive density of any future observation y_{n+k}, $k > 0$ is

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

It is very time consuming to evaluate this multiple integration, especially when further prediction is desired. In this paper, we focus our attention on how to analytically obtain an explicit and accurate approximation of the above marginal predictive density, instead of performing multiple integration.

3.2 *Path sampling methods*

A simulation method proposed by Thompson and Miller (1986) for $AR(p)$ models is used to estimate the posterior predictive densities for model (2.1). Under the normal-gamma prior assumption, the posterior distribution for the error precision τ is gamma and the distribution of μ conditional on Y_n and τ is normal. In detail,

$$\tau \mid Y_n \sim \text{Gamma}\left(\frac{1}{2}(m + 2a), \frac{1}{2}R_n\right) \tag{3.1}$$

and

$$\mu \mid Y_n, \tau \sim \text{Normal}(\mu_n^*, (\tau A_n)^{-1}), \tag{3.2}$$

where R_n, μ_n^* and A_n are defined in Section 2.

A simulation algorithm utilizing formulae (3.1) and (3.2) to generate joint k-step-ahead predictions is computed by performing the following three distinct steps as suggested by Thompson and Miller:

1. Choose a value of τ from the gamma distribution defined by equation (3.1).
2. From the conditional posterior of μ described by equation (3.2), choose a set of μ parameters, say $(\beta_1, \beta_2, \ldots, \beta_r, \phi_1, \phi_2, \ldots, \phi_p)$.
3. Using the selected values of (μ, τ) in steps 1 and 2, a path k periods long, say $\{y_{n+1}, y_{n+2}, \ldots, y_{n+k}\}$ is simulated based on model (2.1).

After a sufficient number of paths are generated, the posterior predictive density for any future observation is computed.

3.3 *A two-stage approximate method*

In this subsection, we propose a two-stage method to obtain an estimation of the posterior predictive density of any future observation y_{n+k}. In the following discussion, k is treated as a fixed integer. Now

$$y_{t+k} = \sum_{j=0}^{k-1} c_{j-1,1} x_{t+k-j}^T \beta + \sum_{j=1}^{p} c_{k-1,j} y_{t+1-j} + \sum_{j=0}^{k-1} c_{j-1,1} \epsilon_{t+k-j}, \quad \text{for} \quad t \geq p, \tag{3.3}$$

where $c_{i,j} = c_{i-1,1}\phi_j + c_{i-1,j+1}$, for $i \geq 0$ and $j \geq 1$, with the initial condition

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

Note that $\phi_j = 0$, for all $j \geq p+1$.

The purpose of this representation is to avoid the presence of $y_{t+1}, y_{t+2}, \ldots, y_{t+k-1}$ on the right-hand side of (3.4). The same technique was utilized by Chow (1974) in order to compute the moments of some multiperiod predictions. Furthermore, for $i \geq p+k$, define

$$y_i^* = y_i - \sum_{j=1}^{p} c_{k-1,j} y_{i-k+1-j}, \qquad x_i^* = \sum_{j=0}^{k-1} d_j x_{i-j} \qquad \text{and} \qquad \epsilon_i^* = \sum_{j=0}^{k-1} d_j \epsilon_{i-j},$$

where $d_j = c_{j-1,1}$, for $j = 0, 1, \ldots, k-1$.

Then, formula (3.3) can be rewritten as

$$y_{t+k}^* = x_{t+k}^{*T} \beta + \epsilon_{t+k}^*, \quad \text{for} \quad t \geq p. \tag{3.4}$$

And a matrix form is represented as

$$Y_n^* = X_n^* \beta + \varepsilon_n^*,$$

where $Y_n^* = \begin{pmatrix} y_{p+k}^* \\ y_{p+k+1}^* \\ \vdots \\ y_n^* \end{pmatrix}$, $X_n^* = \begin{pmatrix} x_{p+k}^* \\ x_{p+k+1}^* \\ \vdots \\ x_n^* \end{pmatrix}$ and $\varepsilon_n^* = \begin{pmatrix} \epsilon_{p+k}^* \\ \epsilon_{p+k+1}^* \\ \vdots \\ \epsilon_n^* \end{pmatrix}$.

Here, ε_n^* is multivariate normally distributed with mean zero and covariance matrix $\tau^{-1} D V^{-1} D^T$, where D is an $(m-k+1) \times m$ matrix defined as

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

Define

$$\tilde{Y}_n = U^{1/2}Y_n^*, \qquad \tilde{X}_n = U^{1/2}X_n^*, \qquad \tilde{\varepsilon}_n = U^{1/2}\varepsilon_n^* \quad \text{and} \quad U = (DV^{-1}D^T)^{-1},$$

then the following representation results:

$$\tilde{Y}_n = \tilde{X}_n\beta + \tilde{\varepsilon}_n, \tag{3.5}$$

here $\tilde{\varepsilon}_n$ is multivariate normally distributed with mean zero and covariance matrix $\tau^{-1}I_n$. That is the components of $\tilde{\varepsilon}_n$ are independent identically distributed. If we know values of $\tilde{Y}_n$, then model (3.5) is a standard multiple regression model. However, components of $\tilde{Y}_n$ involve the unknown parameters of ϕ_i's. To overcome this disadvantage, we suggest replacing ϕ_i by its posterior mean, this is the first stage. Now, model (3.5) is a multiple regression model and the posterior predictive density of y_{n+k} is derived explicitly as follows:

Along with model (3.5) the predictive model is written as

$$\tilde{y}_{n+k} = \tilde{x}_{n+k}^T\beta + \tilde{\epsilon}_{n+k},$$

where $\tilde{y}_{n+k} = dy_{n+k}^*$, $\tilde{x}_{n+k} = dx_{n+k}^*$, $\tilde{\epsilon}_{n+k} = d\epsilon_{n+k}^*$ and $d = (\sum_{j=0}^{k-1} d_j^2 v_{n+k-j})^{-1/2}$. Here $\tilde{\varepsilon}_n = U^{1/2}\varepsilon_n^* = U^{1/2}D\varepsilon_n$, where ε_n is defined by elements of $\{\epsilon_{p+1}, \epsilon_{p+2}, \ldots, \epsilon_n\}$ and $\tilde{\epsilon}_{n+k}$ is a linear combination of $\{\epsilon_{n+1}, \epsilon_{n+2}, \ldots, \epsilon_{n+k-1}\}$. Therefore, $\tilde{\epsilon}_{n+k}$ and $\tilde{\varepsilon}_n$ are mutually independent. Now, a Bayesian analysis of a multiple regression model can be performed and a posterior predictive density for any future observation y_{n+k} can be explicitly obtained with the following results. Most of the developments are adapted from Zellner (1971).

Let β_0 denote the first r components of μ_0 and Q_0^{-1} denote the right upper $r \times r$ matrix of Q^{-1}. Define

$$\tilde{R}_n = \tilde{Y}_n^T(I - \tilde{X}_n\tilde{F}_n^{-1}\tilde{X}_n^T)\tilde{Y}_n, \qquad \tilde{F}_n = \tilde{X}_n^T\tilde{X}_n,$$
$$\tilde{A}_n = Q_0 + \tilde{F}_n \quad \text{and} \quad \tilde{\beta} = \tilde{A}_n^{-1}(Q_0\beta_0 + \tilde{X}_n^T\tilde{Y}_n).$$

Under the same normal-gamma prior assumption as stated in Subsection 2.1, the following results are obtained:

Property 3.1. The posterior density of β is a multivariate t-distribution with $(m - k + 2a + 1)$ degrees of freedom expressed as

$$p(\beta \mid \tilde{Y}_n) \propto \left\{1 + (\beta - \tilde{\beta})^T \frac{\tilde{A}_n}{\tilde{R}_n}(\beta - \tilde{\beta})\right\}^{-(m-k+r+2a+1)/2}.$$

Property 3.2. The posterior predictive density of $\tilde{y}_{n+k}$ is a univariate t-distribution with $(m - k + 2a + 1)$ degrees of freedom which is expressed as

$$p(\tilde{y}_{n+k} \mid \tilde{Y}_n) \propto \left\{1 + \frac{w_{n+k}}{\eta}(\tilde{y}_{n+k} - \tilde{x}_{n+k}^T\tilde{\beta})^2\right\}^{-(\eta+1)/2},$$

where $w_{n+k} = \eta h_{n+k} \tilde{R}_n^{-1}$, $h_{n+k} = (1 + \tilde{x}_{n+k}^T \tilde{A}_n^{-1} \tilde{x}_{n+k})^{-1}$ and $\eta = m - k + 2a + 1$.

In order to make the degrees of freedom in Property 3.1 be positive, the condition $(m - k + 2a) \geq 0$ is required. In summary, the above proposed method involves two stages to obtain the posterior predictive density:

Stage 1. Obtain the posterior mean of φ (denoted by $\hat{\varphi}$) using the Bayesian method mentioned in Section 2.

Stage 2. Replace φ by $\hat{\varphi}$ and treat model (3.5) as a multiple regression model.

Applying Stages 1 and 2, a posterior predictive density for y_{n+k} is approximately represented as follows:

Property 3.3. The posterior predictive density of y_{n+k} is a univariate t-distribution with $(m - k + 2a + 1)$ degrees of freedom,

$$p(y_{n+k} \mid \tilde{Y}_n) \propto \left\{ 1 + \frac{w_{n+k}}{\eta} (y_{n+k} - m_{n+k})^2 \right\}^{-(\eta+1)/2},$$

where $m_{n+k} = \tilde{\beta}^T x_{n+k} + \hat{\varphi}^T y_{n+k,p}$ and η is defined in Property 3.2.

In the following discussion, we call this method the "two-stage method". And the Bayesian procedure discussed in Section 2 is referred to as "the regular Bayesian method". Moreover, for a pure $AR(p)$ model, the two-stage method can still be applied as long as the constant term is included, say

$$y_t = \beta_1 + y_{t,p}^T \varphi + \epsilon_t.$$

In the next section, some comparisons between the exact method and the two-stage method are investigated analytically for the one-step forecast. For further step ahead forecasts, numerical comparisons are performed between the two-stage method and the path sampling method. For simplicity, we restrict our discussion to the case where $V = I$ and the flat prior is applied.

4. Comparison between methods

4.1 *A special case*: $k = 1$

In this subsection, we analytically evaluate the performance of the two-stage method in a special situation: $k = 1$, $V = I$ and under the flat prior assumption. A posterior predictive density of y_{n+1} using the regular Bayesian method obtained from Property 2.3 is expressed as

$$p(y_{n+1} \mid Y_n) = \pi^{-1/2} (\alpha_n R_n^{-1})^{1/2} \Gamma\left(\frac{\zeta + 1}{2}\right) \Gamma^{-1}\left(\frac{\zeta}{2}\right) \cdot \{1 + \alpha_n R_n^{-1} (y_{n+1} - \mu_n^{*T} z_{n+1})^2\}^{-(\zeta+1)/2}, \tag{4.1}$$

where $\alpha_n = (1 + z_{n+1}^T A_n^{-1} z_{n+1})^{-1}$ and $\zeta = m - p - r$. On the other hand, from Property 3.3 a posterior density of y_{n+1} obtained by using the two-stage method is written as

$$p(y_{n+1} \mid \tilde{Y}_n) = \pi^{-1/2}(\gamma_n \tilde{R}_n^{-1})^{1/2} \Gamma\left(\frac{\zeta+1}{2}\right) \Gamma^{-1}\left(\frac{\zeta}{2}\right) \cdot \{1 + \gamma_n \tilde{R}_n^{-1}(y_{n+1} - m_{n+1})^2\}^{-(\zeta+1)/2}, \tag{4.2}$$

where $\gamma_n = (1 + x_{n+1}^T \tilde{A}_n^{-1} x_{n+1})^{-1}$ and $m_{n+1} = \tilde{\beta}^T x_{n+1} + \hat{\varphi}^T y_{n+1,p}$. Here, $\tilde{A}_n$ and $\tilde{R}_n$ are defined in Subsection 3.3.

We are interested in comparing the discrepancy between the two densities (4.1) and (4.2). Before investigating the differences, we summarize some results obtained from the two-stage method as follows. Their proofs are given in Appendix.

LEMMA 4.1. (1) $\tilde{\beta} = \hat{\beta}$, *here* $\tilde{\beta}$ *and* $\hat{\beta}$ *are the posterior means of* β *obtained respectively using the two-stage method and the regular method.*

(2) $\mu_n^{*T} z_{n+1} = m_{n+1}$.

(3) $\tilde{R}_n = R_n$.

LEMMA 4.2. $A_n^{-1} \geq B_n$, *where* B_n *is a* $(p+r) \times (p+r)$ *matrix defined by* $\begin{bmatrix} \tilde{A}_n^{-1} & 0 \\ 0 & 0 \end{bmatrix}$, A_n *is defined in Subsection* 2.2 *and* $\tilde{A}_n$ *is defined in Subsection* 3.3.

COROLLARY 4.1. (1) $\mathrm{Var}(\tilde{\beta}) \leq \mathrm{Var}(\hat{\beta})$,

(2) $\mathrm{Var}(y_{n+1} \mid \tilde{Y}_n) \leq \mathrm{Var}(y_{n+1} \mid Y_n)$,

that is the two-stage method produces small predictive variance.

The above results show that the posterior mean of β and y_{n+1} are respectively the same using both the regular Bayesian method and the two-stage method. Intuitively, the two-stage method treats φ as a constant, therefore the variation of the predictive density will be small. In fact, the above results prove that the two densities both follow the t-distribution with the same degrees of freedom, but expression (4.2) has smaller variance. However, some numerical examples discussed in the next section show that the shrinkages of variance are almost negligible when forecasting periods are not too far or sample sizes are moderate.

4.2 *Applications*

In this subsection, some numerical results are investigated to demonstrate the accuracy of the two-stage method, with both simulated data and real data. For the simulated data, $AR(1)$ models with one exogenous variable are applied and defined by

$$y_t = \alpha + \beta x_t + \phi y_{t-1} + \epsilon_t, \qquad t = 1, 2, \ldots, n,$$

where $\alpha = 0.5$, $\beta = 0.3$ and ϵ_t's are i.i.d. normal $(0, 1)$. The exogenous variable x_t's are independently generated from uniform $(0, 1)$. As the discussed model is not restricted to only a stationary model, values of ϕ are here set to be $\phi = 0.5$ and 1, representing a stationary and nonstationary model respectively. Moreover, to

investigate the influence of sample size, n's are set as 50, 100 and 300. For obtaining a reliable impact on sample size, the observations are drawn as follows: For each model 350 observations are generated, the last n observations then constituting a sample.

Besides the simulated data, two sets of real data are investigated. One set consists of the quarterly data of United States unemployment rates, from the first quarter of 1948 to the second quarter of 1991, a total of 174 observations. The other includes 179 observations of the United States real GNP (in 1982 dollars), from the first quarter of 1947 to the third quarter of 1991. These two economic time series are coded from the Citicrop Database (1991). The latter analyzed data has been translated by taking a natural logarithm. $AR(2)$ models are fitted to each series respectively described as:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t.$$

When sample size is small, it seems unreasonable to expect accurate longer period forecasts. Thus in the following discussion, the length of the forecast period is adjusted by sample size, ranging from six periods to twelve periods. For each data set, some specified percentiles for each estimated predictive density are computed via two methods, the path sampling method and the proposed two-stage method. The percentiles provided by the path sampling technique are estimated by 10,000 bundles of paths. Besides the comparison of percentiles, four statistics such as, mean, standard deviation, skewness and kurtosis, of each estimated density are computed. As discussed in Section 3, the two-stage method provides an estimated density explicitly, however, the path sampling technique obtains this by generating the unknown parameters. In order to further investigate the simulation variation of the latter, 200 replications are made. Then the estimation errors for each estimated percentile and the aforementioned four statistics via the path sampling method are computed. All these summary statistics could be used to measure the performance of the two-stage method.

As a result of Property 2.3 and Lemma 4.1, the one-step predictive densities are t-distributions, produced both by the regular method and the two-stage method, except that the degrees of freedom differ. Therefore, the estimated predictive densities created by these two methods should have the same mean, skewness and kurtosis. These phenomena are indicated by all the numerical results if the fluctuation errors are ignored. Moreover, since the two-stage method ignores the variation of the parameter vector φ, intuitively, the predictive density should tend to be less variant. Also, the symmetric property of the t-distribution may limit the shape of the predictive density. These two shortages could be examined by using those summary statistics.

The first phenomenon is examined by the standard deviation of each estimated density. In most case, the results show that the two-stage method tends to produce a less variant density than the one produced by the path sampling method. However, the results show that most of the differences between the two methods are quite small for short period forecasts. Referring to the symmetric restriction of the two-stage method, the experiments indicate that this limitation does not cause too serious a discrepancy. Except, for the nonstationary simulated

data with $n = 50$, the predictive density tends to skew to the right. In fact, the predictive medians for each density produced by the two methods are almost the same. This suggests that the bias of the true predictive mean and median is negligible. In general, these two disadvantages die out faster for stationary data than for nonstationary data as sample size gets larger.

The influence of sample size for the two-stage method can be investigated using results from simulated data. Based on the estimated percentiles, the performances of the two methods are quite close as long as the sample size is not too small and when $n = 300$, the percentage differences are almost negligible. This gives evidence that the accuracy of the two-stage method increases as sample size increases. In fact, the posterior mean substituting technique used in the two-stage method has some optimal property as pointed out by Chow (1974): Considering a one-step prediction for an $AR(p)$ model, the posterior means of autoregressive parameters will produce an optimal prediction of y_{n+1} under the criterion of minimum mean squared error. Moreover, a heuristic proof of the asymptotic convergence of the estimated predictive density approached by the two-stage method is presented in Appendix.

For the discussed real data sets, almost all the precentiles provided by the two methods are pretty close. Especially, for the GNP data, the two methods almost create the same estimated density. It is worthy to note that usually when analyzing GNP data, at least a first order differencing of the logarithm data is taken. However, without taking a suitable differencing, the existence of "nonstationarity" is allowed in the model. For this "nonstationary" data set, the result still shows that the two-stage approximation is quite accurate. All these discussed numerical results exhibited in Tables 1–8 were done on a VAX9320 computer at National Central University in Taiwan. The data were generated from the DRNNOR, DRNGAM and DRNMVN subroutines of the IMSL package.

For the sake of saving time, computations of the exact method are omitted, instead the path sampling method is performed for comparison. Presumably, the path sampling method will create a pretty accurate density if the number of paths are sufficiently large. Actually, the standard errors for specified percentiles and four statistics obtained from 200 replications are quite small. Thus 10,000 path bundles should reasonably be enough to estimate the true density. As the results of the two methods are close to each other, we can confidently say that the two-stage method is quite reliable. In fact, the closeness of the two methods is exhibited when sample sizes are not too small. However, when the sample size is small, the two-stage method is still applicable for short period forecasts.

According to the path sampling algorithm, once the posterior distributions described by (3.1) and (3.2) are obtained, then the computing time is not influenced by the sample size of the data and is decided by the number of path bundles. On the other hand, the two-stage method rewrites the model as a standard regression model. Therefore the computing time of the transformation and analysis of regression model increases as the sample size increases. In these simulation experiences, when $n = 50$, the two-stage method is far faster than the path sampling method. As $n = 100$, the former is still faster than the latter. However, when $n = 300$, the computing speed is reversed. Nevertheless, it is worthy to note that the path

Table 1. Inference for the predictive distributions. $y_t = 0.5 + 0.3x_t + 0.5y_{t-1} + \epsilon_t$, $n = 50$.

	Percentiles								
	.05	.25	.50	.75	.95	mean	standard deviation	skewness	kurtosis
$k = 1$	−0.44[a] (0.02)[c]	0.60 (0.01)	1.32 (0.01)	2.06 (0.01)	3.10 (0.02)	1.32[a] (0.01)[c]	1.08 (0.01)	0.02 (0.03)	3.12 (0.06)
	−0.45[b]	0.61	1.33	2.05	3.10	1.33[b]	1.08	0.00	3.14
$k = 2$	−0.50 (0.03)	0.69 (0.02)	1.53 (0.02)	2.37 (0.02)	3.61 (0.03)	1.53 (0.01)	1.25 (0.01)	0.03 (0.03)	3.14 (0.06)
	−0.54	0.69	1.53	2.37	3.60	1.53	1.26	0.00	3.15
$k = 3$	−0.90 (0.03)	0.37 (0.02)	1.23 (0.02)	2.09 (0.02)	3.39 (0.03)	1.23 (0.01)	1.30 (0.01)	0.01 (0.03)	3.22 (0.10)
	−0.96	0.31	1.17	2.04	3.31	1.17	1.30	0.00	3.15
$k = 4$	−1.16 (0.03)	0.16 (0.02)	1.08 (0.02)	1.97 (0.02)	3.30 (0.03)	1.07 (0.01)	1.36 (0.01)	0.00 (0.04)	3.29 (0.11)
	−1.21	0.09	0.97	1.85	3.15	0.97	1.33	0.00	3.15
$k = 5$	−0.97 (0.03)	0.35 (0.02)	1.21 (0.02)	2.10 (0.02)	3.46 (0.03)	1.22 (0.01)	1.36 (0.01)	0.00 (0.04)	3.42 (0.16)
	−1.03	0.28	1.16	2.05	3.35	1.16	1.33	0.00	3.16
$k = 6$	−0.98 (0.03)	0.33 (0.02)	1.21 (0.01)	2.10 (0.02)	3.50 (0.03)	1.23 (0.01)	1.36 (0.01)	0.05 (0.05)	3.49 (0.30)
	−0.97	0.34	1.22	2.11	3.41	1.22	1.33	0.00	3.16

[a]Path sampling method. [b]Two-stage method. [c]Standard error for path sampling.

Table 2. Inference for the predictive distributions. $y_t = 0.5 + 0.3x_t + 0.5y_{t-1} + \epsilon_t$, $n = 100$.

	Percentiles								
	.05	.25	.50	.75	.95	mean	standard deviation	skewness	kurtosis
$k = 1$	−0.53[a] (0.02)[c]	0.53 (0.01)	1.27 (0.01)	2.00 (0.02)	3.10 (0.02)	1.27[a] (0.01)[c]	1.10 (0.01)	0.04 (0.02)	3.07 (0.05)
	−0.52[b]	0.55	1.29	2.03	3.11	1.29[b]	1.11	0.00	3.07
$k = 2$	−0.78 (0.03)	0.41 (0.02)	1.22 (0.01)	2.05 (0.02)	3.23 (0.03)	1.22 (0.01)	1.22 (0.01)	0.01 (0.03)	3.12 (0.05)
	−0.80	0.40	1.23	2.05	3.11	1.23	1.23	0.00	3.07
$k = 3$	−1.00 (0.03)	0.22 (0.02)	1.07 (0.02)	1.91 (0.02)	3.16 (0.03)	1.07 (0.01)	1.26 (0.01)	0.02 (0.03)	3.06 (0.06)
	−0.96	0.25	1.09	1.92	3.13	1.09	1.25	0.00	3.07
$k = 4$	−1.08 (0.03)	0.15 (0.02)	1.03 (0.02)	1.88 (0.02)	3.11 (0.03)	1.02 (0.01)	1.28 (0.01)	0.01 (0.03)	2.98 (0.06)
	−1.10	0.13	0.98	1.84	3.07	0.98	1.27	0.00	3.07
$k = 5$	−1.06 (0.03)	0.20 (0.02)	1.04 (0.02)	1.87 (0.02)	3.10 (0.03)	1.04 (0.01)	1.26 (0.01)	−0.02 (0.03)	3.07 (0.06)
	−0.97	0.25	1.09	1.92	3.14	1.09	1.25	0.00	3.07
$k = 6$	−1.03 (0.03)	0.20 (0.02)	1.03 (0.02)	1.87 (0.02)	3.10 (0.03)	1.04 (0.01)	1.26 (0.01)	0.00 (0.03)	3.28 (0.06)
	−0.92	0.28	1.11	1.95	3.15	1.11	1.24	0.00	3.07
$k = 7$	−1.04 (0.03)	0.21 (0.02)	1.06 (0.02)	1.91 (0.02)	3.17 (0.03)	1.06 (0.01)	1.28 (0.01)	0.02 (0.03)	3.09 (0.06)
	−0.88	0.34	1.17	2.01	3.23	1.17	1.25	0.00	3.07
$k = 8$	−1.05 (0.03)	0.18 (0.02)	1.02 (0.02)	1.88 (0.02)	3.14 (0.03)	1.03 (0.01)	1.28 (0.01)	0.01 (0.03)	3.18 (0.06)
	−0.94	0.28	1.11	1.95	3.16	1.11	1.25	0.00	3.07

[a]Path sampling method. [b]Two-stage method. [c]Standard error for path sampling.

Table 3. Inference for the predictive distributions. $y_t = 0.5 + 0.3x_t + 0.5y_{t-1} + \epsilon_t$, $n = 300$.

	Percentiles								
	.05	.25	.50	.75	.95	mean	standard deviation	skewness	kurtosis
$k = 1$	−0.31[a] (0.02)[c]	0.66 (0.01)	1.36 (0.01)	2.04 (0.01)	3.03 (0.02)	1.36[a] (0.01)[c]	1.02 (0.01)	0.02 (0.02)	2.91 (0.05)
	−0.31[b]	0.67	1.35	2.03	3.02	1.35[b]	1.01	0.00	3.02
$k = 2$	−0.57 (0.02)	0.56 (0.02)	1.34 (0.01)	2.11 (0.02)	3.31 (0.02)	1.35 (0.01)	1.17 (0.01)	0.03 (0.02)	3.05 (0.05)
	−0.55	0.57	1.34	2.12	3.27	1.34	1.15	0.00	3.02
$k = 3$	−0.82 (0.02)	0.35 (0.02)	1.16 (0.01)	1.97 (0.02)	3.15 (0.02)	1.16 (0.01)	1.21 (0.01)	0.00 (0.03)	3.10 (0.05)
	−0.80	0.36	1.16	1.97	3.12	1.16	1.19	0.00	3.02
$k = 4$	−0.90 (0.02)	0.26 (0.02)	1.07 (0.01)	1.89 (0.02)	3.06 (0.03)	1.08 (0.01)	1.21 (0.01)	0.03 (0.02)	3.05 (0.05)
	−0.90	0.27	1.09	1.90	3.07	1.09	1.20	0.00	3.02
$k = 5$	−0.87 (0.03)	0.32 (0.02)	1.13 (0.02)	1.95 (0.02)	3.14 (0.03)	1.13 (0.01)	1.22 (0.01)	0.02 (0.03)	3.04 (0.05)
	−0.85	0.32	1.13	1.95	3.12	1.13	1.21	0.00	3.02
$k = 6$	−0.86 (0.03)	0.32 (0.02)	1.12 (0.01)	1.96 (0.02)	3.11 (0.03)	1.13 (0.01)	1.21 (0.01)	−0.01 (0.02)	2.96 (0.05)
	−0.85	0.32	1.13	1.95	3.12	1.13	1.21	0.00	3.02
$k = 7$	−0.85 (0.03)	0.33 (0.02)	1.15 (0.02)	1.95 (0.02)	3.13 (0.03)	1.14 (0.01)	1.21 (0.01)	0.00 (0.02)	2.99 (0.05)
	−0.83	0.35	1.16	1.98	3.15	1.16	1.21	0.00	3.02
$k = 8$	−0.95 (0.03)	0.28 (0.02)	1.07 (0.02)	1.91 (0.02)	3.07 (0.03)	1.08 (0.01)	1.22 (0.01)	0.02 (0.03)	3.09 (0.05)
	−0.89	0.29	1.11	1.92	3.10	1.11	1.21	0.00	3.02
$k = 9$	−0.93 (0.03)	0.28 (0.02)	1.12 (0.02)	1.93 (0.02)	3.07 (0.03)	1.10 (0.01)	1.22 (0.01)	−0.02 (0.03)	3.09 (0.05)
	−0.86	0.32	1.14	1.95	3.13	1.14	1.21	0.00	3.02
$k = 10$	−0.92 (0.03)	0.29 (0.02)	1.10 (0.02)	1.89 (0.02)	3.08 (0.03)	1.08 (0.01)	1.21 (0.01)	−0.04 (0.03)	3.04 (0.05)
	−0.86	0.32	1.14	1.96	3.13	1.14	1.21	0.00	3.02
$k = 11$	−0.95 (0.03)	0.27 (0.02)	1.10 (0.02)	1.92 (0.02)	3.04 (0.03)	1.09 (0.01)	1.21 (0.01)	−0.01 (0.02)	3.00 (0.05)
	−0.85	0.34	1.15	1.97	3.15	1.15	1.22	0.00	3.02
$k = 12$	−0.89 (0.03)	0.33 (0.02)	1.17 (0.02)	1.99 (0.02)	3.14 (0.02)	1.15 (0.01)	1.22 (0.01)	−0.03 (0.02)	3.00 (0.05)
	−0.82	0.37	1.18	2.00	3.19	1.18	1.22	0.00	3.02

[a]Path sampling method. [b]Two-stage method. [c]Standard error for path sampling.

Table 4. Inference for the predictive distributions. $y_t = 0.5 + 0.3x_t + 0.5y_{t-1} + \epsilon_t$, $n = 50$.

	Percentiles								
	.05	.25	.50	.75	.95	mean	standard deviation	skewness	kurtosis
$k = 1$	197.1[a] (0.03)[c]	198.1 (0.01)	198.9 (0.01)	199.6 (0.01)	200.7 (0.02)	198.9[a] (0.01)[c]	1.11 (0.01)	0.03 (0.03)	3.11 (0.06)
	197.1[b]	198.2	198.9	199.6	200.6	198.9[b]	1.07	0.00	3.14
$k = 2$	197.3 (0.04)	198.8 (0.02)	199.9 (0.02)	201.0 (0.02)	202.6 (0.04)	199.9 (0.02)	1.63 (0.01)	0.04 (0.02)	3.11 (0.06)
	197.4	198.9	199.9	201.0	202.5	199.9	1.55	0.00	3.15
$k = 3$	197.3 (0.04)	199.3 (0.03)	200.7 (0.03)	202.0 (0.03)	204.1 (0.05)	200.7 (0.02)	2.10 (0.01)	0.08 (0.03)	3.14 (0.06)
	197.5	199.4	200.6	201.9	203.8	200.6	1.92	0.00	3.15
$k = 4$	197.3 (0.04)	199.7 (0.03)	201.4 (0.03)	203.1 (0.03)	205.7 (0.06)	201.4 (0.02)	2.57 (0.02)	0.11 (0.02)	3.12 (0.07)
	197.7	199.9	201.3	202.8	204.9	201.3	2.20	0.00	3.15
$k = 5$	197.6 (0.06)	200.4 (0.04)	202.4 (0.04)	204.4 (0.04)	207.4 (0.06)	202.4 (0.03)	3.01 (0.02)	0.13 (0.03)	3.16 (0.07)
	198.1	200.7	202.4	204.1	206.6	202.4	2.60	0.00	3.16
$k = 6$	197.8 (0.07)	201.0 (0.03)	203.3 (0.04)	205.6 (0.04)	209.2 (0.08)	203.3 (0.03)	3.44 (0.03)	0.20 (0.03)	3.22 (0.07)
	198.5	201.3	203.3	205.2	208.1	203.3	2.92	0.00	3.16

[a]Path sampling method. [b]Two-stage method. [c]Standard error for path sampling.

Table 5. Inference for the predictive distributions. $y_t = 0.5 + 0.3x_t + 0.5y_{t-1} + \epsilon_t$, $n = 100$.

	Percentiles								
	.05	.25	.50	.75	.95	mean	standard deviation	skewness	kurtosis
$k = 1$	197.1[a] (0.02)[c]	198.2 (0.02)	198.9 (0.01)	199.7 (0.01)	200.8 (0.02)	198.9[a] (0.01)[c]	1.11 (0.01)	0.04 (0.03)	3.03 (0.05)
	197.1[b]	198.2	198.9	199.7	200.7	198.9[b]	1.09	0.00	3.07
$k = 2$	197.2 (0.04)	198.8 (0.02)	199.9 (0.02)	200.9 (0.02)	202.5 (0.04)	199.9 (0.02)	1.60 (0.01)	0.00 (0.02)	3.09 (0.05)
	197.3	198.8	199.9	200.9	202.4	199.9	1.56	0.00	3.07
$k = 3$	197.4 (0.04)	199.3 (0.03)	200.6 (0.02)	202.0 (0.03)	204.0 (0.05)	200.7 (0.02)	2.03 (0.01)	0.01 (0.03)	3.09 (0.05)
	197.5	199.4	200.7	202.0	203.9	200.7	1.94	0.00	3.07
$k = 4$	197.5 (0.05)	199.9 (0.03)	201.5 (0.03)	203.1 (0.03)	205.5 (0.06)	201.5 (0.02)	2.43 (0.02)	−0.01 (0.03)	3.07 (0.06)
	197.8	200.0	201.5	203.0	205.2	201.5	2.25	0.00	3.07
$k = 5$	197.8 (0.06)	200.5 (0.04)	202.4 (0.03)	204.2 (0.04)	207.0 (0.07)	202.4 (0.03)	2.79 (0.02)	0.00 (0.03)	3.08 (0.06)
	198.2	200.7	202.4	204.2	206.7	202.4	2.57	0.00	3.07
$k = 6$	198.1 (0.07)	201.2 (0.04)	203.3 (0.04)	205.4 (0.05)	208.4 (0.07)	203.3 (0.03)	3.14 (0.02)	0.00 (0.03)	3.09 (0.06)
	198.6	201.4	203.3	205.2	208.0	203.3	2.85	0.00	3.07
$k = 7$	198.5 (0.07)	201.9 (0.04)	204.2 (0.04)	206.6 (0.05)	210.0 (0.08)	204.2 (0.03)	3.49 (0.03)	0.02 (0.03)	3.04 (0.06)
	199.2	202.1	204.2	206.3	209.2	204.2	3.06	0.00	3.07
$k = 8$	198.9 (0.08)	202.5 (0.05)	205.1 (0.05)	207.6 (0.05)	211.5 (0.09)	205.1 (0.04)	3.84 (0.03)	0.05 (0.03)	3.07 (0.06)
	199.5	202.8	205.1	207.3	210.6	205.1	3.36	0.00	3.07

[a]Path sampling method. [b]Two-stage method. [c]Standard error for path sampling.

Table 6. Inference for the predictive distributions. $y_t = 0.5 + 0.3x_t + 0.5y_{t-1} + \epsilon_t$, $n = 300$.

	Percentiles								
	.05	.25	.50	.75	.95	mean	standard deviation	skewness	kurtosis
$k = 1$	196.9[a] (0.02)[c]	197.9 (0.01)	198.6 (0.01)	199.3 (0.01)	200.3 (0.02)	198.6[a] (0.01)[c]	1.02 (0.01)	0.03 (0.03)	2.94 (0.05)
	196.9[b]	197.9	198.6	199.3	200.3	198.6[b]	1.01	0.00	3.02
$k = 2$	196.8 (0.03)	198.2 (0.02)	199.2 (0.02)	200.2 (0.02)	201.7 (0.03)	199.2 (0.01)	1.47 (0.01)	0.03 (0.02)	2.99 (0.05)
	196.8	198.2	199.2	200.2	201.6	199.2	1.44	0.00	3.02
$k = 3$	196.7 (0.04)	198.4 (0.02)	199.6 (0.02)	200.9 (0.02)	202.6 (0.04)	199.6 (0.02)	1.81 (0.01)	0.02 (0.02)	3.03 (0.05)
	196.8	198.5	199.6	200.8	202.5	199.6	1.75	0.00	3.02
$k = 4$	196.7 (0.04)	198.7 (0.03)	200.1 (0.02)	201.5 (0.03)	203.5 (0.04)	200.1 (0.02)	2.09 (0.02)	0.03 (0.02)	3.05 (0.05)
	196.8	198.7	200.1	201.5	203.4	200.1	2.03	0.00	3.02
$k = 5$	196.7 (0.05)	199.1 (0.03)	200.6 (0.03)	202.2 (0.03)	204.5 (0.05)	200.6 (0.02)	2.35 (0.02)	0.00 (0.03)	2.99 (0.05)
	196.9	199.1	200.6	202.2	204.4	200.6	2.28	0.00	3.02
$k = 6$	196.9 (0.05)	199.4 (0.03)	201.2 (0.03)	202.9 (0.03)	205.4 (0.06)	201.1 (0.02)	2.59 (0.02)	−0.02 (0.03)	3.05 (0.05)
	197.0	199.4	201.1	202.8	205.2	201.1	2.49	0.00	3.02
$k = 7$	197.1 (0.06)	199.8 (0.04)	201.7 (0.03)	203.5 (0.04)	206.4 (0.06)	201.7 (0.03)	2.81 (0.02)	0.01 (0.03)	3.08 (0.05)
	197.2	199.9	201.7	203.5	206.1	201.7	2.71	0.00	3.02
$k = 8$	197.2 (0.06)	200.2 (0.04)	202.1 (0.04)	204.1 (0.04)	207.2 (0.06)	202.1 (0.03)	3.03 (0.02)	0.00 (0.03)	3.06 (0.04)
	197.4	200.2	202.1	204.1	206.9	202.1	2.90	0.00	3.02
$k = 9$	197.3 (0.07)	200.5 (0.04)	202.7 (0.04)	204.8 (0.04)	208.1 (0.07)	202.7 (0.03)	3.23 (0.03)	0.03 (0.03)	3.05 (0.05)
	197.6	200.6	202.7	204.7	207.7	202.7	3.06	0.00	3.02
$k = 10$	197.6 (0.07)	200.8 (0.05)	203.2 (0.04)	205.5 (0.04)	208.8 (0.08)	203.2 (0.03)	3.42 (0.03)	0.04 (0.03)	2.98 (0.05)
	197.9	201.0	203.2	205.4	208.5	203.2	3.23	0.00	3.02
$k = 11$	197.8 (0.08)	201.2 (0.05)	203.6 (0.05)	206.1 (0.05)	209.6 (0.08)	203.7 (0.04)	3.62 (0.03)	0.04 (0.03)	3.03 (0.05)
	198.1	201.4	203.7	206.0	209.3	203.7	3.41	0.00	3.02
$k = 12$	198.0 (0.08)	201.7 (0.05)	204.2 (0.05)	206.8 (0.05)	210.5 (0.09)	204.2 (0.04)	3.80 (0.03)	0.03 (0.03)	3.02 (0.05)
	198.4	201.9	204.3	206.6	210.1	204.3	3.55	0.00	3.02

[a]Path sampling method. [b]Two-stage method. [c]Standard error for path sampling.

Table 7. Inference for the predictive distributions. Quarterly Unemployment Rate, $n = 174$.

	Percentiles								
	.05	.25	.50	.75	.95	mean	standard deviation	skewness	kurtosis
$k = 1$	6.45[a] (0.01)[c]	6.76 (0.00)	6.99 (0.00)	7.21 (0.00)	7.53 (0.01)	6.99[a] (0.00)[c]	0.33 (0.00)	0.02 (0.02)	3.02 (0.05)
	6.45[b]	6.77	6.99	7.21	7.53	6.99[b]	0.33	0.00	3.04
$k = 2$	6.01 (0.01)	6.61 (0.01)	7.02 (0.01)	7.44 (0.01)	8.07 (0.01)	7.03 (0.01)	0.63 (0.00)	0.03 (0.02)	3.05 (0.05)
	6.00	6.61	7.03	7.45	8.06	7.03	0.62	0.00	3.04
$k = 3$	5.50 (0.02)	6.39 (0.01)	6.97 (0.01)	7.57 (0.01)	8.46 (0.02)	6.98 (0.01)	0.89 (0.01)	0.02 (0.02)	3.06 (0.06)
	5.51	6.38	6.98	7.58	8.44	6.98	0.89	0.00	3.04
$k = 4$	5.05 (0.02)	6.15 (0.01)	6.88 (0.01)	7.61 (0.01)	8.73 (0.02)	6.88 (0.01)	1.12 (0.01)	0.02 (0.03)	3.11 (0.06)
	5.04	6.12	6.87	7.62	8.71	6.87	1.12	0.00	3.04
$k = 5$	4.63 (0.03)	5.89 (0.02)	6.76 (0.02)	7.61 (0.02)	8.92 (0.03)	6.75 (0.01)	1.30 (0.01)	0.02 (0.03)	3.14 (0.06)
	4.61	5.87	6.74	7.61	8.88	6.74	1.30	0.00	3.04
$k = 6$	4.25 (0.03)	5.65 (0.02)	6.59 (0.02)	7.55 (0.02)	8.97 (0.03)	6.61 (0.01)	1.44 (0.01)	0.01 (0.03)	3.16 (0.06)
	4.23	5.63	6.59	7.56	8.96	6.59	1.44	0.00	3.04
$k = 7$	3.99 (0.03)	5.45 (0.02)	6.48 (0.02)	7.48 (0.02)	9.01 (0.03)	6.48 (0.01)	1.54 (0.01)	0.02 (0.03)	3.21 (0.07)
	3.94	5.42	6.45	7.48	8.97	6.45	1.53	0.00	3.04
$k = 8$	3.74 (0.04)	5.28 (0.02)	6.35 (0.02)	7.40 (0.02)	9.05 (0.04)	6.35 (0.02)	1.61 (0.01)	0.04 (0.03)	3.25 (0.07)
	3.70	5.24	6.30	7.37	8.91	6.30	1.59	0.00	3.04
$k = 9$	3.54 (0.04)	5.13 (0.02)	6.22 (0.02)	7.31 (0.02)	9.00 (0.04)	6.23 (0.02)	1.67 (0.01)	0.04 (0.03)	3.25 (0.08)
	3.49	5.09	6.20	7.31	8.92	6.20	1.66	0.00	3.04
$k = 10$	3.39 (0.04)	5.02 (0.02)	6.11 (0.02)	7.23 (0.02)	8.96 (0.04)	6.13 (0.02)	1.71 (0.01)	0.05 (0.03)	3.27 (0.08)
	3.37	4.96	6.06	7.16	8.75	6.06	1.64	0.00	3.04
$k = 11$	3.27 (0.04)	4.91 (0.02)	6.03 (0.02)	7.17 (0.02)	8.93 (0.04)	6.04 (0.02)	1.74 (0.01)	0.06 (0.03)	3.29 (0.09)
	3.22	4.82	5.93	7.05	8.66	5.93	1.66	0.00	3.04
$k = 12$	3.16 (0.04)	4.82 (0.02)	5.94 (0.02)	7.12 (0.02)	8.91 (0.04)	5.98 (0.02)	1.77 (0.01)	0.07 (0.03)	3.35 (0.09)
	3.10	4.73	5.85	6.98	8.60	5.85	1.67	0.00	3.04

[a]Path sampling method. [b]Two-stage method. [c]Standard error for path sampling.

Table 8. Inference for the predictive distributions. Quarterly GNP, $n = 179$.

	Percentiles								
	.05	.25	.50	.75	.95	mean	standard deviation	skewness	kurtosis
$k = 1$	8.32[a] (0.00)[c]	8.33 (0.00)	8.33 (0.00)	8.34 (0.00)	8.35 (0.00)	8.33[a] (0.00)[c]	0.01 (0.00)	0.01 (0.02)	3.05 (0.05)
	8.32[b]	8.33	8.33	8.34	8.35	8.33[b]	0.01	0.00	3.04
$k = 2$	8.31 (0.00)	8.33 (0.00)	8.34 (0.00)	8.35 (0.00)	8.37 (0.00)	8.34 (0.00)	0.02 (0.00)	−0.03 (0.03)	3.06 (0.05)
	8.31	8.33	8.34	8.35	8.37	8.34	0.02	0.00	3.04
$k = 3$	8.31 (0.00)	8.33 (0.00)	8.34 (0.00)	8.36 (0.00)	8.38 (0.00)	8.34 (0.00)	0.02 (0.00)	−0.01 (0.03)	3.11 (0.06)
	8.31	8.33	8.34	8.36	8.38	8.34	0.02	0.00	3.04
$k = 4$	8.30 (0.00)	8.33 (0.00)	8.35 (0.00)	8.37 (0.00)	8.40 (0.00)	8.35 (0.00)	0.03 (0.00)	0.00 (0.03)	3.14 (0.06)
	8.30	8.33	8.35	8.37	8.39	8.35	0.03	0.00	3.04
$k = 5$	8.30 (0.00)	8.33 (0.00)	8.35 (0.00)	8.38 (0.00)	8.41 (0.00)	8.35 (0.00)	0.03 (0.00)	−0.01 (0.03)	3.16 (0.06)
	8.30	8.33	8.35	8.38	8.41	8.35	0.03	0.00	3.04
$k = 6$	8.30 (0.00)	8.33 (0.00)	8.36 (0.00)	8.38 (0.00)	8.42 (0.00)	8.36 (0.00)	0.04 (0.00)	−0.02 (0.03)	3.16 (0.06)
	8.30	8.33	8.36	8.38	8.42	8.36	0.04	0.00	3.04
$k = 7$	8.30 (0.00)	8.34 (0.00)	8.36 (0.00)	8.39 (0.00)	8.43 (0.00)	8.36 (0.00)	0.04 (0.00)	−0.02 (0.03)	3.12 (0.06)
	8.30	8.34	8.36	8.39	8.43	8.36	0.04	0.00	3.04
$k = 8$	8.29 (0.00)	8.34 (0.00)	8.37 (0.00)	8.40 (0.00)	8.44 (0.00)	8.37 (0.00)	0.05 (0.00)	−0.02 (0.03)	3.13 (0.06)
	8.30	8.34	8.37	8.40	8.44	8.37	0.04	0.00	3.04
$k = 9$	8.29 (0.00)	8.34 (0.00)	8.37 (0.00)	8.41 (0.00)	8.45 (0.00)	8.37 (0.00)	0.05 (0.00)	−0.03 (0.03)	3.13 (0.06)
	8.30	8.34	8.37	8.40	8.45	8.37	0.05	0.00	3.04
$k = 10$	8.29 (0.00)	8.34 (0.00)	8.38 (0.00)	8.41 (0.00)	8.46 (0.00)	8.38 (0.00)	0.05 (0.00)	−0.04 (0.03)	3.11 (0.07)
	8.30	8.35	8.38	8.41	8.46	8.38	0.05	0.00	3.04
$k = 11$	8.29 (0.00)	8.35 (0.00)	8.38 (0.00)	8.42 (0.00)	8.47 (0.00)	8.38 (0.00)	0.06 (0.00)	−0.03 (0.03)	3.12 (0.07)
	8.30	8.35	8.38	8.42	8.47	8.38	0.05	0.00	3.04
$k = 12$	8.29 (0.00)	8.35 (0.00)	8.39 (0.00)	8.43 (0.00)	8.48 (0.00)	8.39 (0.00)	0.06 (0.00)	−0.01 (0.03)	3.15 (0.07)
	8.30	8.35	8.39	8.42	8.48	8.39	0.05	0.00	3.04

[a]Path sampling method. [b]Two-stage method. [c]Standard error for path sampling.

sampling method samples future observations sequentially one by one. In order to obtaining any future observations y_{n+k}, all the future observations before that, say y_{n+i}, $1 \leq i \leq k-1$, should meanwhile be simulated. However, the two-stage method provides a closed form for any future step. Thus the predictive density for each specified future step is computed independently.

The purpose of this paper is to explore the multiperiod predictive density analytically. Since an exact expression is impossible for more than a one-step forecast, then we try to approximate this predictive density by a suitable and commonly used explicit density. At least, we have a rough idea what this density looks like. Speaking overall, the results show that under the normal-gamma prior assumption, the multiperiod predictive density can almost be approached by a suitable t-distribution. Though the true density is slightly skewed and has a slight heavier tail when the sample size is small, this discrepancy is not too serious for short period forecasts and disappears as the sample size increases. In practical usage, it will be convenient and reliable to treat the predictive density as a t-distribution when statistical properties of this density are needed analytically.

5. Conclusions

This article investigates the multiperiod forecasting problem of $AR(p)$ models with strongly exogenous variables via the Bayesian approach using a normal-gamma prior assumption. A convenient and explicit method is proposed to produce an estimation of the posterior predictive density for any future observation. The accuracy of the proposed method has been examined by some simulated and real data. Some percentiles of the posterior predictive densities up to twelve-step-ahead forecasting are calculated for each series. For each forecasting step, percentiles and some statistics related to moments are computed using two methods respectively, the proposed two-stage method and the path sampling method.

Though the two-stage method provides a symmetrical and less variant density, it seems these restrictions do not cause too serious a discrepancy. The results show that the two methods almost produce the same estimate for short period forecast. When the sample size gets larger, a longer period forecast based on the two-stage method becomes more accurate. More precisely, when the sample size is large, the two-stage method provides an estimated density which converges asymptotically to the true one. A heuristic proof of this asymptotic property is shown analytically. In general, the forecasting experiments reveal that the proposed method is rather reliable and could be used explicitly to produce a posterior predictive density.

Meanwhile, the proposed two-stage method can be applied if the $AR(p)$ component is extended by an $ARMA(p,q)$ component. In this situation, just replace the exact likelihood function of the $AR(p)$ component by an approximate likelihood function which is proportional to

$$\tau^{m/2} \exp\left\{-\frac{\tau}{2}\left(y_t - x_t^T\beta - \sum_{i=1}^{p}\phi_i y_{t-i} - \sum_{j=1}^{q}\theta_j \epsilon_{t-j}\right)^2\right\}.$$

From the routine Bayesian approach for an $ARMA$ model (Broemeling and Shaarawy (1988)), the posterior means of ϕ_i's and θ_j's are substituted into the

model. Thus, an estimated predictive density is produced by a standard regression analysis as discussed in this paper.

Acknowledgements

This research was supported by the National Science Council, Republic of China under contract #NSC82-0208-M-008-086. Part of this work was done while the author was a visitor at the Graduate School of Business at the University of Chicago during the 1991–1992 academic year. The author wishes to thank the referees for their valuable comments and suggestions which led to considerable improvement of this article and to the GSB for their kind hospitality.

Appendix

1. Proofs of lemmas and corollary

In the following proofs, n is fixed and for simplicity the lower subscript n is sometimes appropriately dropped. Matrix Z_n is decomposed as $Z_n = (X_1 \mid X_2)$, where X_1 is an $m \times r$ matrix and X_2 is an $m \times p$ matrix. Again, the underlying assumptions are $k = 1$, $V = I$ and the flat prior is utilized.

A.1 $\tilde{\beta} = \hat{\beta}$

pf): Let $\hat{\beta}$ be the posterior mean of β using regular Bayesian analysis under the flat prior, then after algebra

$$\text{(A.1)} \qquad \hat{\beta} = (X_1^T W X_1)^{-1} X_1^T W Y, \qquad \text{where} \quad W = I - X_2 (X_2^T X_2)^{-1} X_2^T.$$

On the other hand, the posterior mean of β by the two-stage method under these assumptions is

$$\tilde{\beta} = (X_1^T X_1)^{-1} X_1^T [I + X_2 (X_2^T X_2)^{-1} X_2^T X_1 (X_1^T W X_1)^{-1} X_1^T] W Y.$$

Let $X_0 = (X_1^T X_1)^{-1/2} X_1^T$, then $X_1^T W X_1 = (X_1^T X_1)^{1/2} (X_0 W X_0^T)(X_1^T X_1)^{1/2}$ and $X_1 (X_1^T W X_1)^{-1} X_1^T = X_0 (X_0 W X_0^T)^{-1} X_0$. Therefore,

$$\begin{aligned} \tilde{\beta} &= (X_1^T X_1)^{-1/2} X_0 [I + (I - W) X_0^T (X_0 W X_0^T)^{-1} X_0] W Y \\ &= (X_1^T X_1)^{-1/2} (X_0 W X_0^T)^{-1} X_0 W Y \\ &= (X_1^T W X_1)^{-1} X_1^T W Y \\ &= \hat{\beta}. \end{aligned}$$

A.2 $\mu_n^{*T} z_{n+1} = m_n$

pf): By definition, $\mu_n^{*T} z_{n+1} = \hat{\beta}^T x_{n+1} + \hat{\varphi}^T y_{n+1,p}$ and $m_n = \tilde{\beta}^T x_{n+1} + \hat{\varphi}^T y_{n+1,p}$. Since $\tilde{\beta} = \hat{\beta}$, then $\mu_n^{*T} z_{n+1} = m_n$. That is the posterior means of the predictive density is the same both when using the regular Bayesian method and the two-stage method.

A.3 $\tilde{R}_n = R_n$

pf): Under the flat prior assumption, R_n is reduced to

$$R_n = SSR_n + Y^T[I - Z(Z^TZ)^{-1}Z^T]Y.$$

According to the partitioning of matrix Z, one has

$$R_n = Y^TWY - Y^TWX_1(X_1^TWX_1)^{-1}X_1^TWY,$$

where W is defined by (A.1).

On the other hand, with the two-stage method

$$\tilde{R}_n = \tilde{Y}^T[I - \tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T]\tilde{Y} = \tilde{Y}^T(\tilde{Y} - \tilde{X}_1\hat{\beta}) \quad \text{and} \quad \tilde{Y} = Y - X_2\hat{\varphi},$$

where $\hat{\varphi} = (X_2^TX_2)^{-1}X_2^T[I - X_1(X_1^TWX_1)^{-1}X_1^TW]Y$. After algebra,

$$\tilde{Y} - \tilde{X}_1\hat{\beta} = W[I - X_1(X_1^TWX_1)^{-1}X_1^TW]Y \quad \text{and} \quad \tilde{Y}^TW = Y^TW.$$

Therefore, $\tilde{R}_n = Y^TW[I - X_1(X_1^TWX_1)^{-1}X_1^TW]Y = R_n$.

A.4 $A_n^{-1} \geq B_n$

pf): By definition, $\tilde{A}_n^{-1} = (\tilde{X}_n^T\tilde{X}_n)^{-1} = (X_1^TX_1)^{-1}$ and

$$A_n^{-1} = (Z^TZ)^{-1} = \begin{bmatrix} X_1^TX_1 & X_1^TX_2 \\ X_2^TX_1 & X_2^TX_2 \end{bmatrix}^{-1}.$$

For simplicity, let C, B and $X_{i,j}$ denote A_n^{-1}, B_n and $X_i^TX_j$ respectively.

For matrices C and B, there exists matrix G (Basilevsky (1983), p. 235) such that

$$G^TCG = I \quad \text{and} \quad G^TBG = \Lambda = \begin{bmatrix} H & 0 \\ 0 & 0 \end{bmatrix},$$

where H is an $r \times r$ diagonal matrix. Then $C - B = (G^T)^{-1}(I - \Lambda)G^{-1}$, here $I - \Lambda = \begin{bmatrix} I - H & 0 \\ 0 & I_2 \end{bmatrix}$, and I_2 is a $p \times p$ identity matrix. Therefore, it is sufficient to show that $I - H$ is non-negative definite. By definition of B, it can be shown that

$$H = G_{11}^TX_{11}^{-1}G_{11},$$

where G_{11} is the upper left $r \times r$ matrix of G. Moreover, since H is idempotent and the non-negative property of matrix $I - H$ results.

A.5 $\mathrm{Var}(\tilde{\beta}) \leq \mathrm{Var}(\hat{\beta})$

pf): Under the assumptions and by using Property 3.1,

$$\mathrm{Var}(\tilde{\beta}) = \frac{\tilde{R}_n}{\nu}\tilde{A}_n^{-1} = \frac{\tilde{R}_n}{\nu}(X_1^TX_1)^{-1}, \quad \text{where} \quad \nu = m - p - r - 2,$$

and with Property 2.1,

$$\mathrm{Var}(\mu^*) = \frac{R_n}{\nu} A_n^{-1} \quad \text{and}$$
$$\mathrm{Var}(\hat{\beta}) = \frac{R_n}{\nu}[X_1^T X_1 - X_1^T X_2 (X_2^T X_2)^{-1} X_2^T X_1]^{-1}.$$

Since $\tilde{R}_n = R_n$, therefore $\mathrm{Var}(\tilde{\beta}) \leq \mathrm{Var}(\hat{\beta})$ is proved.

A.6 $\mathrm{Var}(y_{n+1} \mid \tilde{Y}_n) \leq \mathrm{Var}(y_{n+1} \mid Y_n)$

pf): By using Property 2.3, the predictive variance using the regular Bayesian method is

$$\mathrm{Var}(y_{n+1} \mid Y_n) = \frac{R_n}{\nu}(1 + z_{n+1}^T A_n^{-1} z_{n+1}), \qquad \text{where} \quad \nu = m - p - r - 2.$$

On the other hand, the predictive variance with the two-stage method is

$$\mathrm{Var}(y_{n+1} \mid \tilde{Y}_n) = \frac{\tilde{R}_n}{\nu}(1 + x_{n+1}^T (X_1^T X_1)^{-1} x_{n+1}).$$

Then resulting from the use of Lemma 4.2, the inequality is established.

2. Proof of asymptotical convergence of the two-stage method

Let $\Psi(B) = \sum_{j=0}^{k-1} d_j B^j$, then

$$x_{n+k}^* = \sum_{j=0}^{k-1} d_j x_{n+k-j} = \Psi(B) x_{n+k} \quad \text{and} \quad \epsilon_{n+k}^* = \sum_{j=0}^{k-1} d_j \epsilon_{n+k-j} = \Psi(B)\epsilon_{n+k}.$$

Formula (2.1) can be rewritten as

$$\text{(A.2)} \qquad \Phi(B) y_{n+k} = x_{n+k}^T \beta + \epsilon_{n+k}, \qquad \text{where} \quad \Phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j.$$

After algebra, it can be shown that

$$\Psi(B)\Phi(B) = 1 - \sum_{j=1}^{p} c_{k-1,j} B^{k+j-1}.$$

Hence, pre-multiplying both sides of formula (A.2) by the operator $\Psi(B)$, one has

$$\text{(A.3)} \qquad y_{n+k}^* = \Psi(B) x_{n+k}^T \beta + \Psi(B)\epsilon_{n+k} = x_{n+k}^{*T} \beta + \epsilon_{n+k}^*.$$

Therefore, formulae (A.2) and (A.3) are essentially equivalent. And they produce the same prediction for y_{n+k}, as long as the parameter vector φ is known.

Since the posterior mean of φ obtained in the first stage is consistent, therefore the posterior predictive density of y_{n+k} produced by model (A.3) will asymptotically converge to that produced by model (A.2).

REFERENCES

Basilevsky, A. (1983). *Applied Matrix Algebra in the Statistical Sciences*, North-Holland, New York.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, Holden-Day, Oakland.

Broemeling, L. and Land, M. (1984). On forecasting with univariate autoregressive process: A Bayesian approach, *Comm. Statist. Theory Methods*, **13**, 1305–1320.

Broemeling, L. and Shaarawy, S. (1988). *Bayesian Analysis of Time Series and Dynamic Models* (ed. J. C. Spall), 1–21, Marcel Dekker, New York.

Chow, G. C. (1974). *Multiperiod Predictions from Stochastic Difference Equations by Bayesian Methods* (eds. S. E. Fienberg and A. Zellner), 313–324, North-Holland, Amsterdam.

Geweke, J. (1989). Bayesian inference in econometrics models using Monte Carlo integration, *Econometrica*, **57**, 1317–1339.

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

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.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.