Journal of the Korean Statistical Society (2004), **33**: 2, pp 169–189

DEFAULT BAYESIAN INFERENCE OF REGRESSION MODELS WITH ARMA ERRORS UNDER EXACT FULL LIKELIHOODS†

YOUNG SOOK SON[1]

ABSTRACT

Under the assumption of default priors, such as noninformative priors, Bayesian model determination and parameter estimation of regression models with stationary and invertible ARMA errors are developed under exact full likelihoods. The default Bayes factors, the fractional Bayes factor (FBF) of O'Hagan (1995) and the arithmetic intrinsic Bayes factors (AIBF) of Berger and Pericchi (1996a), are used as tools for the selection of the Bayesian model. Bayesian estimates are obtained by running the Metropolis-Hastings subchain in the Gibbs sampler. Finally, the results of numerical studies, designed to check the performance of the theoretical results discussed here, are presented.

AMS 2000 subject classifications. Primary 62F15; Secondary 62M10.
Keywords. Regression models with stationary and invertible ARMA errors, fractional Bayes factor, arithmetic intrinsic Bayes factor, Gibbs sampler, Metropolis-Hastings algorithm.

1. INTRODUCTION

Strong autocorrelations are often seen in most residual analyses, which are performed following regression analyses of time series data. These phenomena require model structures incorporating autocorrelations of time series data. Regression models with ARMA errors can be considered as one alternative. Classical analyses of ARMA models are mainly based on asymptotic results. However, the Bayesian approach does not place any theoretical restriction on the sample size.

Received April 2003; accepted December 2003.
†This work was supported by grant No. R05-2000-000-00024-0 from the Basic Research Program of the Korea Science Engineering Foundation.
[1]Department of Statistics, Chonnam National University, Kwangju 500-757, Korea (e-mail : ysson@chonnam.ac.kr)

The focus of this paper is confined to the Bayesian inference of the regression model, $M_{k,p,q}$, with the stationary and invertible ARMA(p,q) error as follows:

$$M_{k,p,q} \ : \ y_t = \beta_0 + \sum_{i=1}^{k} \beta_i x_{ti} + \varepsilon_t, \quad t = 1, 2, \ldots, n, \tag{1.1}$$

where $\Phi_p(B)\varepsilon_t = \Theta_q(B)a_t$, $\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, $\Theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$, B is a backshift operator and $\{a_t\}$ is a sequence of $N(0, \sigma^2)$ white noises. In this model, $\beta_0, \beta_1, \ldots, \beta_k$, σ^2, $\boldsymbol{\phi}_p = (\phi_1, \phi_2, \ldots, \phi_p)$ and $\boldsymbol{\theta}_q = (\theta_1, \theta_2, \ldots, \theta_q)$ are all unknown parameters. For the stationarity-invertibility of the ARMA(p,q) errors, $(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)$ should be restricted to the region, $C_p \times C_q$, where

$$C_p \times C_q = \left\{ (\boldsymbol{\phi}_p, \boldsymbol{\theta}_q) : \Phi_p(x) = 0, \ |x| > 1 \text{ and } \Theta_q(y) = 0, \ |y| > 1 \right\} .$$

In particular, the model (1.1) with $\beta_i = 0$, $i = 1, 2, \ldots, k$, is refereed to as a stationary and invertible ARMA process.

Monahan (1983) gave a fully Bayesian analysis of the stationary and invertible ARMA models under the uniform prior over $C_p \times C_q$ for $(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)$ and the standard normal-inverse gamma conjugate prior for β_0 and σ^2. However, the numerical integration required in the inference was done using a fixed quadrature rule only for $p + q \leq 2$.

To carry out a fully Bayesian analysis of stationary and invertible ARMA models, Marriott *et al.* (1995) developed the concept of sampling based inference, based on Markov Chain Monte Carlo (MCMC) methods, such as Gibbs sampling and the Metropolis-Hastings algorithm under full likelihood, by introducing the unobserved history of time series data as latent variables. The priors that were used consisted of a multivariate normal prior for the latent variables, noninformative improper priors for β_0 and σ^2, and the uniform prior over Euclidean $(p+q)$ space for partial autocorrelations transformed from $(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)$ over $C_p \times C_q$. Because of the obscurity in the interpretation of the Bayes factor, due to the use of improper priors, the Bayes factor was not used for the determination of the models, but rather this was done by comparing the predictive performance of competing models. In addition, the problems of missing observations, outlier detection, and prediction are dealt with.

Chib and Greenberg (1994) developed the Bayesian estimation of the same regression models with stationary and invertible ARMA errors, as defined in (1.1). They expressed the conditional likelihood in terms of a number of pre-sample variables that was smaller than the number of latent variables used in the study by Marriott *et al.* (1995). They assumed multivariate normal priors for a vector of regression parameters and a vector of pre-sample variables, an inverse gamma prior for σ^2, and nonstandard proper priors for $\boldsymbol{\phi}_p$ and $\boldsymbol{\theta}_q$. The estimation technique that was used consisted of the Gibbs sampling and the Metropolis-Hastings algorithms. They also showed that the proposed Gibbs sampler converges to the true density.

Varshavsky (1996) used the arithmetic intrinsic Bayes factor (AIBF) of Berger and Pericchi (1996a) under exact full likelihoods, in order to identify regression models with AR errors. She assumed the uniform prior for $\boldsymbol{\phi}_p$ over its stationary region, C_p, noninformative improper priors for σ^2 and a vector of regression parameters. Also, the integral over C_p, which is required for the calculation of the AIBF, was computed by the Monte Carlo method after being transformed into partial autocorrelations over $(-1,1)^p$ from $\boldsymbol{\phi}_p$ over C_p.

Son (1999) discussed the identification of a stationary and invertible ARMA model using the usual Bayes factor under the exact full likelihood, and priors of the kind described in Marriott *et al.* (1995).

In this paper, we discuss the use of a fully Bayesian inference, in order to select the most appropriate model for a given set of data among competing models, regression models with stationary and invertible ARMA(p,q) errors, and to estimate all of the parameters included in the selected model. Our priors are based on the vagueness of information. Thus, we assume such default priors as noninformative improper priors for a vector of regression parameters and σ^2, and the uniform prior over $C_p \times C_q$ for $(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)$. The tools that we use for the Bayesian model determination are the default Bayes factors, the fractional Bayes factor (FBF) of O'Hagan (1995) and the arithmetic intrinsic Bayes factor (AIBF) of Berger and Pericchi (1996a). Bayesian estimation is performed using the Gibbs sampling and the Metropolis-Hastings subchain in the Gibbs sampler. Practically speaking, in computing the default Bayes factors for the identification of the models, or in applying the MCMC method for the estimation of the parameters, we used vectors of the partial autocorrelations over $(-1,1)^{p+q}$ reparameterized from $(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)$ over $C_p \times C_q$.

Here, we considerably extend the work of Varshavsky (1996) or Son (1999). Also, we apply a different methodology from theirs to the same regression models with stationary and invertible ARMA errors that Chib and Greenberg (1994) dealt with. Specifically, we use the fully exact likelihood function, expressed only in terms of real data, without introducing latent variables or pre-sample variables for the unobserved history of the time series data, as described in Marriott *et al.* (1995) or Chib and Greenberg (1994), which was rendered possible by Leeuw (1994), who presented the covariance matrix of ARMA errors in a closed form.

The contents of this paper are as follows. In Section 2, we describe the exact full likelihood function and our priors assumptions. In Section 3, the posterior probability is computed for each competing model, *via* the FBF and the AIBF. In Section 4, we provide the results of the full conditional posterior distributions, which are to be used when running the Gibbs sampler for the purpose of estimating the parameters. Finally, numerical studies designed to check the performance of the theoretical results discussed in this paper are provided.

2. Exact Likelihood Function and Priors Assumptions

For a sample of size n, the model described in (1.1) can be expressed in the form of a vector given by

$$M_{k,p,q} \; : \; \boldsymbol{Y} = \boldsymbol{X}_k\boldsymbol{\beta}_k + \boldsymbol{\varepsilon},$$

where $\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)'$, $\boldsymbol{\beta}_k = (\beta_0, \beta_1, \ldots, \beta_k)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ and $\boldsymbol{X}_k = (\boldsymbol{x}_{1,k}, \boldsymbol{x}_{2,k}, \ldots, \boldsymbol{x}_{n,k})'$ is an $n \times (k+1)$ matrix with $\boldsymbol{x}'_{t,k} = (1, x_{t1}, x_{t2}, \ldots, x_{tk})$ for $t = 1, 2, \ldots, n$. Since $\{y_t - \boldsymbol{x}'_{t,k}\boldsymbol{\beta}_k\} = \{\varepsilon_t\}$ follows a stationary and invertible ARMA(p,q) process, $E(\boldsymbol{Y}) = \boldsymbol{X}_k\boldsymbol{\beta}_k$ and $\text{Cov}(\boldsymbol{Y}) = \sigma^2\boldsymbol{V}_{p,q}$, where $\boldsymbol{V}_{p,q}$ is an $n \times n$ matrix composed of only $\boldsymbol{\phi}_p$ and $\boldsymbol{\theta}_q$.

Thus, the exact full likelihood function is explicitly given by

$$f\big(\boldsymbol{Y}|\boldsymbol{\beta}_k, \sigma, \; \boldsymbol{\phi}_p, \; \boldsymbol{\theta}_q\big) = (2\pi\sigma^2)^{-\frac{n}{2}} \; |\boldsymbol{V}_{p,q}^{-1}|^{\frac{1}{2}} \times \exp\Big\{ - \frac{1}{2\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}_k\boldsymbol{\beta}_k)'\boldsymbol{V}_{p,q}^{-1}(\boldsymbol{Y} - \boldsymbol{X}_k\boldsymbol{\beta}_k)\Big\}, \quad (2.1)$$

where the specification of $\boldsymbol{V}_{p,q}^{-1}$ and $|\boldsymbol{V}_{p,q}|$ in Leeuw (1994) is rewritten in Appendix.

Now, we make vague assumptions on priors as follows. For a real line, R,

$$\pi^N(\boldsymbol{\beta}_k, \sigma) \propto \sigma^{-(r+1)}, \; \boldsymbol{\beta}_k \in R^{k+1} \;, \; 0 < \sigma < \infty, \; r \geq 0. \quad (2.2)$$

The usual selection of r is $r = 0$ (the reference prior of Berger and Bernardo (1992)) or $r = k + 1$ (the Jeffreys' prior).

$$\pi(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q) = \frac{I_{C_p \times C_q}(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)}{\text{Volume}(C_p \times C_q)} \; .$$

where

$$I_{C_p \times C_q}(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q) = \begin{cases} 1, & \text{if } (\boldsymbol{\phi}_p, \boldsymbol{\theta}_q) \in C_p \times C_q \,, \\ 0, & \text{otherwise.} \end{cases}$$

Also, assuming that $(\boldsymbol{\beta}_k, \sigma)$ and $(\boldsymbol{\phi}_p,\ \boldsymbol{\theta}_q)$ are independent, we set

$$\pi^N(\boldsymbol{\beta}_k,\ \sigma,\ \boldsymbol{\phi}_p,\ \boldsymbol{\theta}_q) \ = \ \pi^N(\boldsymbol{\beta}_k,\ \sigma) \cdot \pi(\boldsymbol{\phi}_p,\ \boldsymbol{\theta}_q). \tag{2.3}$$

The superscript N, which appears in the form of notations throughout this paper, refers to the use of noninformative improper priors.

3. MODEL DETERMINATION USING THE FBF AND THE AIBF

Consider the problem of selecting one appropriate model, which generates a set of time series data, $\{y_1, y_2, y_3, \ldots, y_n\}$, among competing models,

$$M_{k,p,q} \ : \ y_t = \boldsymbol{x}'_{t,k}\boldsymbol{\beta}_k + \varepsilon_t \ , \qquad t = 1, 2, \ldots, n,$$

where $k \in I_{\text{Reg}} = \{k_i;\ i = 1, 2, \ldots\}$, $p \in I_{\text{AR}} = \{p_i;\ i = 1, 2, \ldots\}$, $q \in I_{\text{MA}} = \{q_i;\ i = 1, 2, \ldots\}$, and all the elements of each set, I_{Reg}, I_{AR} or I_{MA}, are nonnegative integers. In particular, we let $K = \max_i \{k_i\}$, $P = \max_i \{p_i\}$, $Q = \max_i \{q_i\}$.

At the first step of Bayesian inference, prior distributions are required for all of the parameters in the models. In the beginning of a Bayesian experiment, in the absence of any prior information on the parameters, default priors, such as noninformative priors, can be used. The Bayes factor or the posterior probability of the hypothesis or model can be used as tools for Bayesian testing or Bayesian model selection. The Bayes factor depends on prior distributions. However, because of the arbitrary constant incorporated into it, the usual Bayes factor cannot be used directly, in spite of their objectivity and simplicity, if default priors, most of which are typically improper, are assumed. On the other hand, the fractional Bayes factor (FBF) of O'Hagan (1995) and the intrinsic Bayes factor (IBF) of Berger and Perrichi (1996a), which are classified as 'default' or 'automatic' Bayes factors, are free from the arbitrariness of noninformative improper priors (Berger and Mortera, 1999).

Now, for use in the computation of the FBFs and the AIBFs later on, we define two functions of a given data, $\boldsymbol{Y}$ and a constant b, $0 < b \leq 1$ under the model $M_{k,p,q}$, as follows

$$m^N_{k,p,q}(\boldsymbol{Y}|b) = \int_{C_p \times C_q} \int_0^\infty \int_{R^{k+1}} \pi^N(\boldsymbol{\beta}_k, \sigma, \boldsymbol{\phi}_p, \boldsymbol{\theta}_q) \times \left\{ f(\boldsymbol{Y}|\boldsymbol{\beta}_k, \sigma, \boldsymbol{\phi}_p, \boldsymbol{\theta}_q) \right\}^b d\boldsymbol{\beta}_k \, d\sigma \, d(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)$$

and

$$B^N{}_{(k,p,q)(k',p',q')}(\boldsymbol{Y}|b) = \frac{m^N{}_{k,p,q}(\boldsymbol{Y}|b)}{m^N{}_{k',p',q'}(\boldsymbol{Y}|b)} .$$

To integrate over $\boldsymbol{\beta}_k$ and σ in the computation of $m^N_{k,p,q}(\boldsymbol{Y}|b)$ is straightforward, if the kernel of multivariate normal density and that of inverse gamma density are used, respectively. However, to integrate over $(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)$ is not explicit. For the Bayesian inference of a stationary and invertible ARMA process to be valid, $C_p \times C_q$ must be identified, since $(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)$ is restricted to the region, $C_p \times C_q$. However, $C_p \times C_q$ becomes very complicated for $p, q \geq 2$. To circumvent the difficulty in identifying $C_p \times C_q$ with high order p and q, there is a useful reparameterization, which is described in the literature. Following the work of Barndorff-Nielsen and Schou (1973), Monahan (1984) and Jones (1987), there is a one to one transformation between $(\boldsymbol{\phi}_p,\ \boldsymbol{\theta}_q)$ and the partial autocorrelations, $(\boldsymbol{\gamma}_p, \boldsymbol{\gamma}_q)$, that maps $C_p \times C_q$ to $(-1,1)^{(p+q)}$. Its applications are shown in Marriott *et al.* (1995), Varshavsky (1996), and Son (1999).

After integrating over $\boldsymbol{\beta}_k$ and σ, we can set

$$(\boldsymbol{Y} - \boldsymbol{X}_k\widehat{\boldsymbol{\beta}}_k)'\boldsymbol{V}^{-1}_{p,q}(\boldsymbol{Y} - \boldsymbol{X}_k\widehat{\boldsymbol{\beta}}_k) = \left|\boldsymbol{X}'_k\boldsymbol{V}^{-1}_{p,q}\boldsymbol{X}_k\right|^{-1} \cdot \left|(\boldsymbol{X}_k, \boldsymbol{Y})'\boldsymbol{V}^{-1}_{p,q}(\boldsymbol{X}_k, \boldsymbol{Y})\right|,$$

where $\widehat{\boldsymbol{\beta}}_{k,p,q} = (\boldsymbol{X}'_k\boldsymbol{V}^{-1}_{p,q}\boldsymbol{X}_k)^{-1}\boldsymbol{X}'_k\boldsymbol{V}^{-1}_{p,q}\boldsymbol{Y}$, using the fact (Shilov, 1961) that $|\boldsymbol{A}'\boldsymbol{A}|^{1/2} \cdot \|(\boldsymbol{I} - \boldsymbol{P}_A)\boldsymbol{B}\| = |(\boldsymbol{A}, \boldsymbol{B})'(\boldsymbol{A}, \boldsymbol{B})|^{1/2}$ with $\boldsymbol{P}_A = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$. Finally, transforming from $(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)$ to $(\boldsymbol{\gamma}_p, \boldsymbol{\gamma}_q)$ results in

$$\begin{aligned} & m^N{}_{k,p,q}(\boldsymbol{Y}|b) \\ &= \frac{2^{\frac{r}{2}-1}\Gamma\left\{(bn + r - k - 1)/2\right\}}{\pi^{\frac{1}{2}(bn-k-1)} b^{\frac{1}{2}(bn+r)}} \\ &\quad \times \int_{(-1,1)^{p+q}} \frac{|\boldsymbol{V}^{*-1}_{p,q}|^{\frac{b}{2}} |\boldsymbol{X}'_k\boldsymbol{V}^{*-1}_{p,q}\boldsymbol{X}_k|^{\frac{1}{2}(bn+r-k-2)}}{|(\boldsymbol{X}_k, \boldsymbol{Y})'\boldsymbol{V}^{*-1}_{p,q}(\boldsymbol{X}_k, \boldsymbol{Y})|^{\frac{1}{2}(bn+r-k-1)}} f(\boldsymbol{\gamma}_p, \boldsymbol{\gamma}_q)\, d(\boldsymbol{\gamma}_p, \boldsymbol{\gamma}_q), \end{aligned} \tag{3.1}$$

where $\boldsymbol{V}^*_{p,q}$ is an $n \times n$ matrix with $(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)$ in $\boldsymbol{V}_{p,q}$ replaced by $(\boldsymbol{\gamma}_p, \boldsymbol{\gamma}_q)$ and

$$f(\boldsymbol{\gamma}_p, \boldsymbol{\gamma}_q) = \prod_{u=1}^{p} B_{\gamma_u}\left(\left[\frac{u+1}{2}\right], \left[\frac{u}{2}\right] + 1\right) \prod_{v=1}^{q} B_{\gamma_v}\left(\left[\frac{v+1}{2}\right], \left[\frac{v}{2}\right] + 1\right)$$

with $B_{\gamma_j}(\alpha_1, \alpha_2)$ being a rescaled beta density of a random variable γ_j defined on $(-1, 1)$ with parameters α_1 and α_2.

The idea behind the IBF is to use minimal training samples to convert the improper prior to the proper posterior density. The minimal training sample refers to a part of the full sample, with the minimal sample size required to guarantee $0 < m^N_{k,p,q}(\boldsymbol{Y}|b = 1) < \infty$ for all (k, p, q). Since the prior of $(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)$ has a finite support, the minimal training sample is of size $K + 2$, as in the case of a regression model with independent random errors (Berger and Pericchi, 1996b). Our minimal training sampling scheme must also be planned to preserve the continuity of time. So, under the model $M_{k,p,q}$, one set of possible minimal training samples is $\{\boldsymbol{Y}(l), l = 1, 2, \ldots, n - K - 1\}$, where $\boldsymbol{Y}(l) = \{y_l, y_{l+1}, \ldots, y_{l+K+1}\}$ is a sample of size $K+2$ with corresponding design matrix $\boldsymbol{X}_k(l) = (\boldsymbol{x}_{l,k}, \boldsymbol{x}_{l+1,k}, \ldots, \boldsymbol{x}_{l+K+1,k})'$ such that all $(\boldsymbol{X}'_k(l)\boldsymbol{V}^{-1}_{p,q}(l)\boldsymbol{X}_k(l))$ are nonsingular, where $\mathrm{Cov}(\boldsymbol{Y}(l)) = \sigma^{-2}\boldsymbol{V}_{p,q}(l)$.

The idea behind the FBF is to use a fraction, b, of each likelihood function to change a noninformative improper prior into a proper prior. Thus, the definition of the FBF yields the FBF of $M_{k,p,q}$ to $M_{k',p',q'}$ as

$$B^{FBF}_{(k,p,q)(k',p',q')} = B^N_{(k,p,q)(k',p',q')}(\boldsymbol{Y}|b = 1) \cdot B^N_{(k',p',q')(k,p,q)}(\boldsymbol{Y}|b). \tag{3.2}$$

O'Hagan (1995) proposed a common and simple method of setting a fraction b as $b = m_0/n$, with m_0 being the minimal training sample size. See O'Hagan (1995) for the other use of b.

Berger and Perrichi (1996a) proposed an arithmetic IBF (AIBF), which uses an arithmetic mean of $B^N_{(k',p',q')(k,p,q)}(\boldsymbol{Y}(l)|b = 1)$, $l = 1, 2, \ldots, n - K - 1$ to prevent the IBF from depending on only one minimal training sample. Now, since the model $M_{K,P,Q}$ is an encompassing model, the AIBF, $B^{AIBF}_{(k,p,q)(k',p',q')}$ of $M_{k,p,q}$ to $M_{k',p',q'}$ can be defined as

$$B^{AIBF}_{(k,p,q)(k',p',q')} = B^N_{(k,p,q)(k',p',q')}(\boldsymbol{Y}|b = 1)\frac{\sum_{l=1}^{n-K-1} B^N_{(k',p',q')(K,P,Q)}(\boldsymbol{Y}(l)|b = 1)}{\sum_{l=1}^{n-K-1} B^N_{(k,p,q)(K,P,Q)}(\boldsymbol{Y}(l)|b = 1)},$$

where $m^N_{k,p,q}(\boldsymbol{Y}(l)|b = 1)$ included in the equation $B^N_{(\cdot)(\cdot)}(\boldsymbol{Y}(l)|b = 1)$ is given by replacing n, $\boldsymbol{X}_k$, $\boldsymbol{Y}$, $\boldsymbol{V}^*_{p,q}$ in $m^N_{k,p,q}(\boldsymbol{Y}|b = 1)$ of (3.1) with $K + 2$, $\boldsymbol{X}_k(l)$, $\boldsymbol{Y}(l)$, $\boldsymbol{V}^*_{p,q}(l)$, respectively, and $\boldsymbol{V}^*_{p,q}(l)$ is obtained from $\boldsymbol{V}_{p,q}(l)$.

Finally, the posterior probability of model $M_{k,p,q}$ *via* the default Bayes factor

$*$ is given by

$$P(M_{k,p,q}|\boldsymbol{Y}) = \left(\sum_{k' \in I_{\text{Reg}}} \sum_{p' \in I_{\text{AR}}} \sum_{q' \in I_{\text{MA}}} \frac{\delta_{k',p',q'}}{\delta_{k,p,q}} B^*_{(k',p',q')(k,p,q)} \right)^{-1}, \tag{3.3}$$

where $*$ denotes FBF or AIBF and $\delta_{k,p,q}$ is the prior probability of the model $M_{k,p,q}$ being true. The model to be selected for an observed time series data, $\boldsymbol{Y}$, is the model that gives the largest posterior probability, $P(M_{k,p,q}|\boldsymbol{Y})$ for all the combinations of k, p and q.

4. Parameter Estimation by Gibbs Sampling

From the joint posterior density which is obtained by combining the prior density (2.3) and the exact likelihood function (2.1), the full conditional distributions of $\boldsymbol{\beta}_k, \sigma^2, (\gamma_p, \gamma_q)$, which are used for constructing the Gibbs sampler, are obtained as follows:

$$\boldsymbol{\beta}_k \mid \sigma^2, \gamma_p, \gamma_q, \boldsymbol{X}_k, \boldsymbol{Y} \sim N_k\left(\widehat{\boldsymbol{\beta}}_{k,p,q}\,,\ \sigma^2(\boldsymbol{X}_k' \boldsymbol{V}^{*-1}_{p,q} \boldsymbol{X}_k)^{-1}\right)$$

where

$$\widehat{\boldsymbol{\beta}}_{k,p,q} = (\boldsymbol{X}_k' \boldsymbol{V}^{*-1}_{p,q} \boldsymbol{X}_k)^{-1} \boldsymbol{X}_k' \boldsymbol{V}^{*-1}_{p,q} \boldsymbol{Y}.$$

$$\sigma^2 \mid \boldsymbol{\beta}_k, \gamma_p, \gamma_q, \boldsymbol{X}_k, \boldsymbol{Y} \sim IG\left(\frac{(n+r)}{2},\ 2\left\{(\boldsymbol{Y} - \boldsymbol{X}_k\boldsymbol{\beta}_k)' \boldsymbol{V}^{*-1}_{p,q} (\boldsymbol{Y} - \boldsymbol{X}_k\boldsymbol{\beta}_k)\right\}^{-1}\right),$$

$$P(\gamma_p, \gamma_q \mid \boldsymbol{\beta}_k, \sigma^2, \boldsymbol{X}_k, \boldsymbol{Y}) \propto h(\gamma_p, \gamma_q),$$

where

$$h(\gamma_p, \gamma_q) = \left|\boldsymbol{V}^{*-1}_{p,q}\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}_k\boldsymbol{\beta}_k)' \boldsymbol{V}^{*-1}_{p,q} (\boldsymbol{Y} - \boldsymbol{X}_k\boldsymbol{\beta}_k)\right\} f(\gamma_p, \gamma_q).$$

Generating $\boldsymbol{\beta}_k$ or σ^2 from the conditional posterior distribution of $\boldsymbol{\beta}_k$ or σ^2, respectively, is straightforward. However, the conditional posterior distribution of (γ_p, γ_q) is not standard, so we use the following Metropolis-Hastings algorithm: Given the initial values $(\gamma_p^{(0)}, \gamma_q^{(0)})$, repeat from (i) to (iv) for $j = 1, 2, \ldots, J$.

(i) Generate γ_p^* and γ_q^* independently and randomly from $(-1,1)^p$ and $(-1,1)^q$, respectively.

(ii) Compute $c = \min\left\{1,\ \dfrac{h(\gamma_p^*, \gamma_q^*)}{h(\gamma_p^{(j-1)}, \gamma_q^{(j-1)})}\right\}$.

(iii) Generate U randomly from $(0, 1)$.

(iv) Set $(\gamma_p^{(j)}, \gamma_q^{(j)}) = \begin{cases} (\gamma_p^*, \gamma_q^*), & \text{if } U \le c, \\ (\gamma_p^{(j-1)}, \gamma_q^{(j-1)}), & \text{if } U > c. \end{cases}$

(γ_p, γ_q) which is finally estimated is again transformed to $(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)$.

5. NUMERICAL STUDY

To check the performance of the theoretical results discussed in this paper, we conducted a numerical study with three simulated data sets and one set of real data. The generated data and all the results of Bayesian inference were obtained using MATLAB (The Math Works Inc., 2000). Three time series data sets of sample size $n = 100$ were generated from the following models, and are shown in Figure 5.1. Normal random variates are generated by the NORMRND function of MATLAB.

(i) Linear trend model with AR(3) errors:
$y_t = 193.8 + 3.5t + \varepsilon_t,$
$\varepsilon_t = -0.6\varepsilon_{t-1} + 0.67\varepsilon_{t-2} + 0.36\varepsilon_{t-3} + a_t, \; a_t \sim iid\ N(0, 1^2).$

(ii) Curve trend model with MA(4) errors:
$y_t = 35.3 + 1.5t + 2.8t^2 + \varepsilon_t,$
$\varepsilon_t = a_t + 1.6a_{t-1} + 0.5a_{t-2} - 0.4a_{t-3} - 0.2a_{t-4}, \; a_t \sim iid\ N(0, 0.5^2).$

(iii) Linear trend model with ARMA(2, 2) errors:
$y_t = 10.6 + 0.5t + \varepsilon_t,$
$\varepsilon_t = -1.4\varepsilon_{t-1} - 0.5\varepsilon_{t-2} + a_t - 0.8a_{t-1} + 0.6a_{t-2}, \; a_t \sim iid\ N(0, 1.5^2).$

When applying the methodology introduced in this paper, it seems to take a lot of time to compute a $n \times n$ matrix, $\boldsymbol{V}_{p,q}^{-1}$ and its determinant, $|\boldsymbol{V}_{p,q}^{-1}|$. However, the core of the computations involves dealing with $p \times p$, $q \times q$, or at most $q + remainder(n/q)$ lower band matrices (Appendix).

We set $r = 0$ as a common choice of r in a prior setting of (2.2). In the computation of FBF of (3.2), we set $b = (K + 2)/n$. For the identification of the model, $M_{k,p,q}$, the computation of the integral in (3.1) must be evaluated in a simple way. We estimate the integral by means of the Monte Carlo method through 200 importance samples with a joint density of $p+q$ independent uniform variates distributed over $(-1, 1)$ as an importance density. To check the convergence of the Monte Carlo integration, we monitored the results of 100, 200, 300,

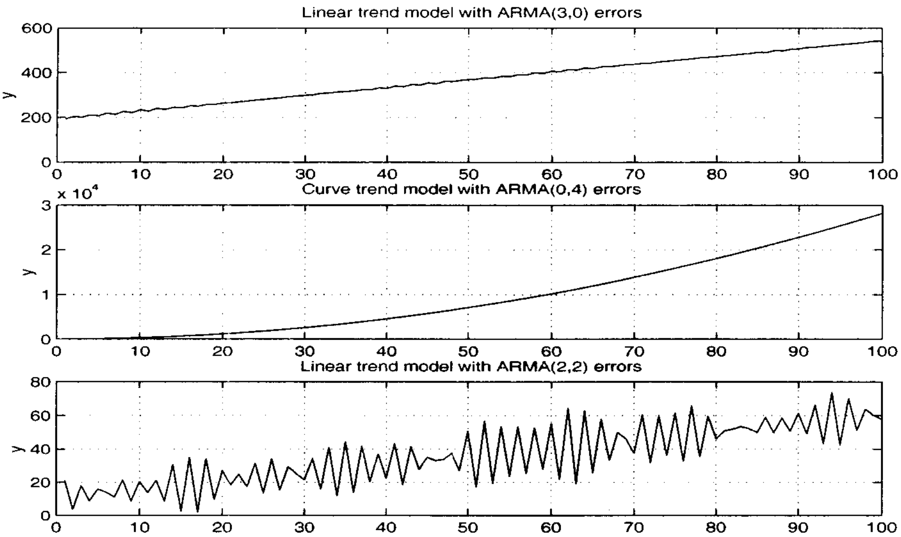

FIGURE 5.1 *Plots of three simulated time series data*

500, 1,000, 2,000, 3,000, 5,000 and 10,000 importance samplings. As a result, it was determined that 200 importance samplings were sufficient to get stability.

Table 5.1 presents the Akaike Information Criterion (AIC) for each model, $M_{k,p,q}$, obtained by means of the the maximum likelihood estimation (MLE), and posterior probabilities *via* the FBF and the AIBF. $*$ in the AIC, which was obtained using PROC ARIMA of SAS (SAS Institute Inc., 2000), denotes that the estimation algorithm may not converge and the results of the estimation are unstable, and $**$ denotes that the estimation algorithm did not converge. The competing models are $M_{k,p,q}$ for $k \in I_{\text{Reg}} = \{1,2,3\}$, $p \in I_{\text{AR}} = \{0,1,\ldots,5\}$, $q \in I_{\text{MA}} = \{0,1,\ldots,5\}$. All of the posterior probabilities are almost zero for the remainders in all of the competing models, except for the true trend model. Therefore, we don't present these results, because of space limitations. It takes about 9 minutes in the case where 200 importance samplings are used for the identification of the models for one data set. For the AR(3) error model, FBF, AIBF, and AIC all select a true model. For the MA(4) error model, FBF and AIBF select a true model, but many of the results obtained using the AIC are unstable. For the ARMA(2,2) error model, both AIBF and AIC select a true model. but FBF selects a more complex model, the ARMA(3,3) model. AIC imposes a penalty when more complex models are selected. Equal prior probabilities for all the models are assumed for the sake of the simplicity in computing

TABLE 5.1 *AIC and posterior probabilities via the FBF and the AIBF for the model* $M_{k,p,q}$

	linear trend (k=1) *AR(3) error*			*curve trend (k=2)* *MA(4) error*			*linear trend (k=1)* *ARMA(2,2) error*		
(p, q)	*FBF*	*AIBF*	*AIC*	*FBF*	*AIBF*	*AIC*	*FBF*	*AIBF*	*AIC*
(0,0)	0.0000	0.0000	457	0.0000	0.0000	269	0.0000	0.0000	774
(0,1)	0.0000	0.0000	399	0.0000	0.0000	189*	0.0000	0.0000	661*
(0,2)	0.0000	0.0000	338	0.0011	0.0008	150*	0.0000	0.0000	598*
(0,3)	0.0000	0.0000	326	0.2599	0.2258	142*	0.0000	0.0000	554*
(0,4)	0.0000	0.0000	310	0.7374	0.6464	252*	0.0000	0.0000	416*
(0,5)	0.0006	0.0001	299	0.0000	0.0000	354*	0.0000	0.0000	523*
(1,0)	0.0000	0.0000	302	0.0000	0.0000	254*	0.0000	0.0000	522
(1,1)	0.0000	0.0000	461*	0.0000	0.0000	186*	0.0000	0.0000	778*
(1,2)	0.0000	0.0000	282	0.0000	0.0000	188*	0.0000	0.0000	379
(1,3)	0.0000	0.0000	283	0.0007	0.0055	125*	0.0024	0.0005	622*
(1,4)	0.0000	0.0000	280	0.0000	0.0002	142*	0.0000	0.0000	365
(1,5)	0.0000	0.0000	282	0.0000	0.0000	129*	0.0000	0.0000	619*
(2,0)	0.0029	0.0027	286	0.0000	0.0000	196*	0.0000	0.0000	394
(2,1)	0.0103	0.0021	283	0.0000	0.0000	143	0.0000	0.0000	377
(2,2)	0.0000	0.0000	280	0.0000	0.0000	135*	0.1323	0.5492	361
(3,3)	0.0000	0.0000	282	0.0001	0.0181	**	0.0000	0.0000	363
(2,4)	0.0000	0.0000	282	0.0000	0.0001	125*	0.1548	0.0684	363
(2,5)	0.0000	0.0000	284*	0.0007	0.0697	126*	0.0000	0.0000	365
(3,0)	0.8488	0.9265	278	0.0000	0.0000	182	0.0000	0.0000	374
(3,1)	0.0000	0.0000	545*	0.0000	0.9000	143	0.0000	0.0000	375*
(3,2)	0.0023	0.0069	340*	0.0000	0.0020	132*	0.0000	0.0000	426*
(3,3)	0.0000	0.0000	282*	0.0000	0.0000	146*	0.7103	0.3818	363*
(3,4)	0.0000	0.0000	284	0.0000	0.0007	136*	0.0000	0.0000	366*
(3,5)	0.0000	0.0000	286	0.0000	0.0000	140*	0.0000	0.0000	427*
(4,0)	0.0024	0.0019	279	0.0000	0.0000	165	0.0000	0.0000	374
(4,1)	0.1325	0.0596	290*	0.0000	0.0000	139	0.0000	0.0000	583*
(4,2)	0.0011	0.0016	283	0.0000	0.0000	133*	0.0000	0.0000	403*
(4,3)	0.0000	0.0000	280*	0.0001	0.0305	213*	0.0000	0.0000	483*
(4,4)	0.0000	0.0000	281*	0.0000	0.0000	134*	0.0000	0.0000	367*
(4,5)	0.0000	0.0000	306*	0.0000	0.0000	130*	0.0000	0.0000	364*
(5,0)	0.0001	0.0000	281	0.0000	0.0000	165	0.0000	0.0000	372
(5,1)	0.0000	0.0000	279*	0.0000	0.0000	141	0.0001	0.0001	370
(5,2)	0.0000	0.0000	283*	0.0000	0.0000	171*	0.0000	0.0000	362
(5,3)	0.0000	0.0000	282*	0.0000	0.0001	136	0.0000	0.0000	408*
(5,4)	0.0002	0.0001	283*	0.0000	0.0000	125*	0.0000	0.0000	391*
(5,5)	0.0000	0.0000	491*	0.0000	0.0000	**	0.0000	0.0000	369*

the posterior probability of equation (3.3). However, the penalty imposed on the number of parameters can be incorporated into the prior probabilities, in order to select simpler models (Varshavsky, 1996).

Tables 5.2, 5.3 and 5.4 show the results of the MLE and posterior distribution for the parameters included in each model. The values in parentheses are approximate standard errors for the MLE and the numerical standard errors for the posterior distributions. The multivariate normal variates, inverse gamma variates, and uniform random variates used in the Gibbs sampler are generated by the functions of MATLAB, MVNRND, GAMRND, and UNIFRND, respectively. In our numerical study, we estimated the parameters from one sequence simulated from only one Gibbs sampler, and burned the first 10% after a total of 110% iterations. The number of iterations of the Gibb sampler and the Metropolis-Hastings subchain in the first two models is 100 in both cases, for which it takes almost 20 minutes, and those in the third ARMA$(2,2)$ error model are 200 and 100, respectively, for which it takes almost 40 minutes. Overall, the results of the posterior distribution are superior to those of the MLE from the viewpoint of the standard errors of the estimates. Specifically, in the results obtained by the MLE for the MA(4) error model, there is strong evidence that the estimation may not converge.

For the analysis of real data, Korea Population Projection data (in millions), with a sample size of 41, from 1960 to 2000 is used, and its plot is shown in Figure 5.2. In Figure 5.3, we present plots of the residuals obtained after fitting four regression trend models with ARMA$(0,0)$ errors as follows:

(i) Linear trend model: $y_t = \beta_1 + \beta_2 t + \varepsilon_t$.

(ii) Curve trend model: $y_t = \beta_1 + \beta_2 t + \beta_3 t^2 + \varepsilon_t$.

(iii) Third trend model: $y_t = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 t^3 + \varepsilon_t$.

(iv) Stepwise regression trend model:
$y_t = (\beta_1 + \beta_2 t + \beta_3 t^2) \cdot I_t + (\beta_4 + \beta_5 t) \cdot (1 - I_t) + \varepsilon_t$, where $I_t = 1$, if $t \leq 24$, $I_t = 0$, otherwise.

In Figure 5.3, the plot (a) of the residuals, obtained after fitting the linear trend model with *iid* errors, requires the use of a model incorporating a quadratic term. Plots, (b) or (c), obtained after fitting the curve trend model or the third trend model with *iid* errors, show strong positive autocorrelations and increasing variances. The stepwise regression trend model is considered, because of the

TABLE 5.2 *The linear trend model with AR(3) errors*

True Parameter	MLE	Posterior Distribution Mean	Std.	Median	Lower 95% limit	Upper 95% limit
$\beta_0 = 193.8$	193.8331	193.6754	0.3918	193.7773	192.9414	194.1381
	(0.2334)	(0.0391)				
$\beta_1 = 3.5$	3.4960	3.4965	0.0040	3.4962	3.4886	3.5028
	(0.0039)	(0.0004)				
$\phi_1 = -0.6$	−0.6734	−0.6106	0.0742	−0.6142	−0.7754	−0.5082
	(0.0963)	(0.0074)				
$\phi_2 = 0.67$	0.5809	0.6777	0.0994	0.6937	0.4642	0.8402
	(0.1012)	(0.0099)				
$\phi_3 = 0.36$	0.3262	0.3695	0.0699	0.3622	0.2769	0.4786
	(0.0995)	(0.0069)				
$\sigma^2 = 1.0$	0.8836	0.9234	0.1454	0.9017	0.7123	1.1629
		(0.0145)				

TABLE 5.3 *The curve trend model with MA(4) errors*

True Parameter	MLE	Posterior Distribution Mean	Std.	Median	Lower 95% limit	Upper 95% limit
$\beta_0 = 35.3$	35.0344	34.8099	0.6308	34.8878	33.3104	35.6158
	(0.2565)	(0.0630)				
$\beta_1 = 1.5$	1.5238	1.5242	0.0133	1.5236	1.5048	1.5469
	(0.0117)	(0.0013)				
$\beta_2 = 2.8$	2.7997	2.7998	0.0000	2.7998	2.7997	2.7999
	(0.0001)	(0.0000)				
$\theta_1 = -1.6$	−0.3830	−1.6626	0.1432	−1.6554	−1.8908	−1.4178
	(20.8508)	(0.0143)				
$\theta_2 = -0.5$	0.1628	−0.7014	0.3115	−0.7420	−1.2324	−0.4119
	(12.9306)	(0.0311)				
$\theta_3 = 0.4$	−0.1187	0.2199	0.3006	0.2195	−0.2072	0.5374
	(9.5223)	(0.0300)				
$\theta_4 = 0.2$	0.3353	0.1559	0.0943	0.1510	0.0198	0.2399
	(7.0306)	(0.0094)				
$\sigma^2 = 0.25$	0.6564	0.3263	0.1013	0.3112	0.1937	0.5144
		(0.0101)				

TABLE 5.4 *The linear trend model with ARMA(2,2) errors*

True Parameter	MLE	Posterior Distribution Mean	Std.	Median	Lower 95% limit	Upper 95% limit
$\beta_0 = 10.6$	10.5207	10.5144	0.1226	10.5362	10.2780	10.6507
	(0.0707)	(0.0086)				
$\beta_1 = 0.5$	0.5022	0.5021	0.0009	0.5021	0.5004	0.5035
	(0.0012)	(0.0000)				
$\phi_1 = -1.4$	−1.4576	−1.3021	0.2361	−1.1385	−1.6155	−1.0210
	(0.0991)	(0.0166)				
$\phi_2 = -0.5$	−0.5472	−0.4352	0.2089	−0.2908	−0.7027	−0.2096
	(0.0984)	(0.0147)				
$\theta_1 = 0.8$	0.8576	0.7520	0.0515	0.7653	0.6216	0.8022
	(0.0909)	(0.0036)				
$\theta_2 = -0.6$	−0.6410	−0.5452	0.2050	−0.7048	−0.7409	−0.2982
	(0.0902)	(0.0144)				
$\sigma^2 = 2.25$	1.9159	2.3967	0.4857	2.3310	1.7553	3.1801
		(0.0343)				

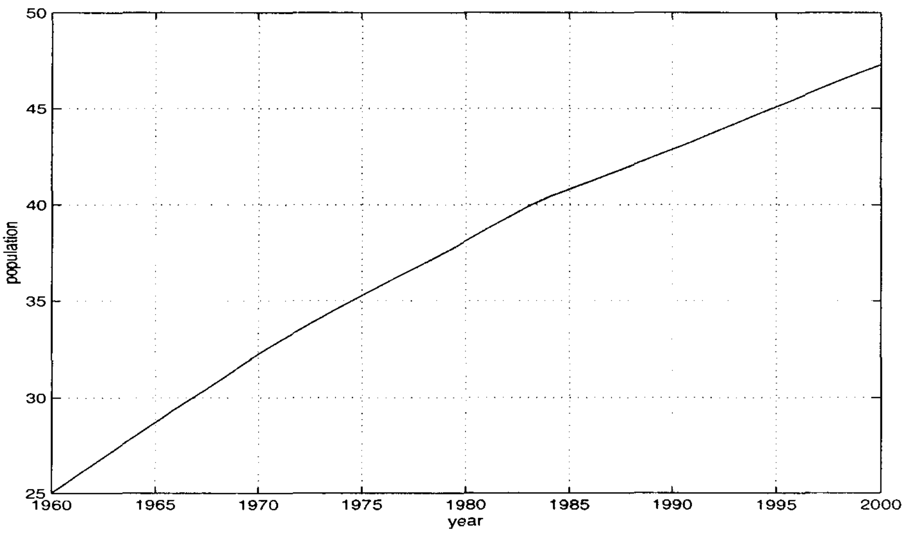

FIGURE 5.2 *Korea Population Projection Data (in millions) (Resource : Korea National Statistics Office, http://www.nso.go.kr)*

possibilities of the curve trend and the declining linear trend centered at $t = 24$ (year=1983) in the plot of (a) in Figure 5.3. As another possibility, fitting

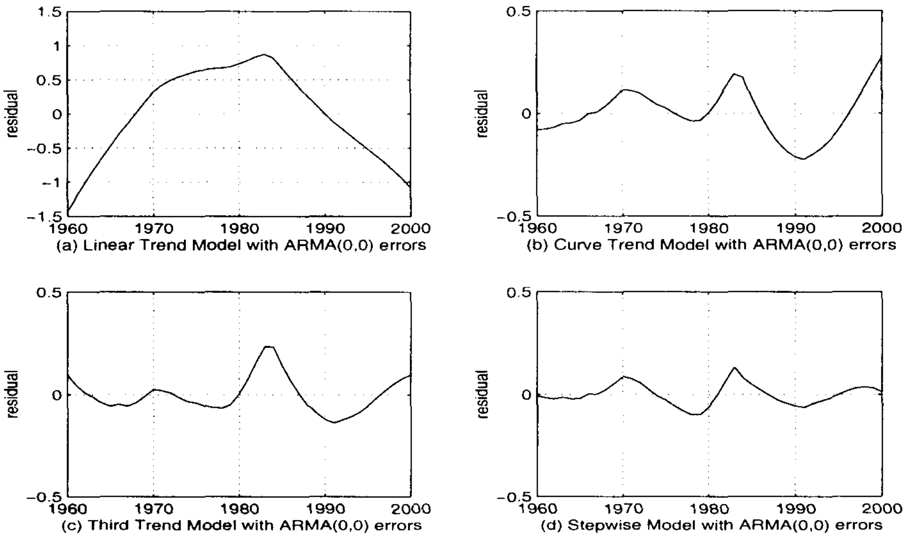

FIGURE 5.3 *Residual plots after fitting regression trend models with iid errors*

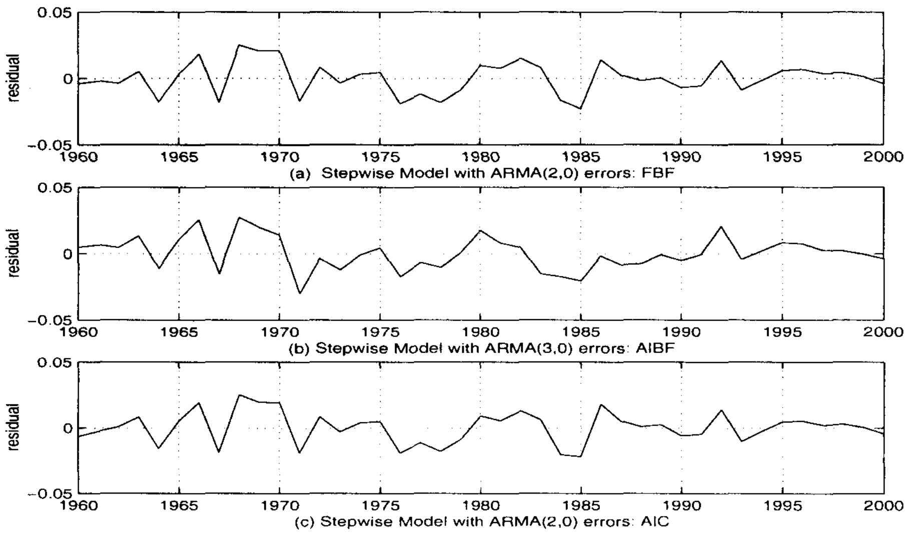

FIGURE 5.4 *Residual plots after fitting the stepwise regression trend model*

the third trend model before $t = 24$ doesn't give any significant improvement compared to the curve. The plot (d) of the residuals, obtained after fitting the stepwise model with *iid* errors, remains strongly autocorrelated, but the variance of the residuals becomes more stable.

The models selected by the FBF, AIBF and AIC, which are shown on the left side of Table 5.5, are the stepwise regression trend model with AR(2), AR(3), and AR(2) errors, respectively. The competing models are (i) the linear, (ii) the curve, (iii) the third, and (iv) the stepwise regression model with ARMA(p, q) errors, where $p, q = 0, 1, 2, 3$. For all of these models, except for the stepwise trend model, only the minimum AIC is presented for all p, q. It takes about 3 minutes to identify the model using 100 importance samplings.

The right side of Table 5.5 presents the results of the estimation for the selected models. The numbers of iterations in the Gibbs sampler and Metropolis-Hastings algorithm are both 100, for which it takes about 5 minutes. The computed test statistics are presented in the first line and the p-values are given in parentheses in the second line of the results of the t-test for zero residuals and Portmanteau χ^2 test to determine whether a sequence of residuals is white noise or not. The plots of the residuals in $(a), (b), (c)$ of Figure 5.4 are those for the models selected by the FBF, AIBF and AIC, respectively. The residuals are much more random than those of the stepwise regression trend model with *iid* errors. The entire analysis of the residuals refers to all of the well fitting models. However, the model selection made by the FBF or AIC is a little better than that of the AIBF from the viewpoint of the SSR (Sum of Squared Residuals).

6. CONCLUDING REMARKS

We present a fully Bayesian computation procedure for the model selection and parameter estimation of regression models with stationary and invertible ARMA errors. Under noninformative priors, the model identification is done by the posterior probability of each model computed *via* the AIBF or the FBF, and the model estimation is implemented by Gibbs sampling.

Our Bayesian procedure is meaningful in that the likelihood function is computed exactly using only observed data, without introducing any data augmentation techniques for unobserved data. We propose that the Bayesian approach can be used as an alternative, when the maximum likelihood estimation or the conditional least square estimation is unstable in terms of its convergence.

After performing many simulation experiments, we observed that both our model selection approach and the AIC often select the same true model, and that the Bayesian estimates obtained by Gibbs sampling are close to the maximum likelihood estimates. However, we can not but point out that the order (p, q), which is different from the true order of the ARMA error, can be selected, whether the Bayesian procedure or the AIC is used, since different ARMA(p, q) errors can be associated with similar values of the likelihood function.

TABLE 5.5 *Results of model selection and estimation for Korea Population Projection Data*

(p, q)	*FBF*	*AIBF*	*AIC*
	Linear trend model $(k = 2)$		
(p, q)	all zero	all zero	−53
	Curve trend model $(k = 3)$		
(p, q)	all zero	all zero	−198
	Third trend model $(k = 4)$		
(p, q)	all zero	all zero	−202
	Stepwise trend model $(k = 5)$		
(0,0)	0.0000	0.0000	−118
(0,1)	0.0000	0.0000	−167*
(0,2)	0.0000	0.0000	−198*
(0,3)	0.0010	0.0000	−205
(1,0)	0.0000	0.0000	−178
(1,1)	0.0000	0.0000	−206
(1,2)	0.0428	0.0008	−215
(1,3)	0.0000	0.0000	−218
(2,0)	0.7917	0.0395	−232
(2,1)	0.0164	0.0034	−230
(2,2)	0.0000	0.0000	−235*
(2,3)	0.0010	0.0004	−234*
(3,0)	0.1432	0.8360	−230*
(3,1)	0.0013	0.0023	−227*
(3,2)	0.0024	0.1174	−229*
(3,3)	0.0000	0.0000	−161*

	FBF	*AIBF*	*AIC*
β_1	24.2814	24.3083	24.2481
	(0.0011)	(0.0015)	(0.0118)
β_2	0.7705	0.7396	0.7760
	(0.0003)	(0.0009)	(0.0042)
β_3	−0.0051	−0.0035	−0.0053
	(0.0000)	(0.0000)	(0.0001)
β_4	29.4831	30.1419	29.4563
	(0.0047)	(0.0227)	(0.0935)
β_5	0.4333	0.4149	0.43405
	(0.0001)	(0.0006)	(0.0029)
ϕ_1	1.6092	1.7677	1.6613
	(0.0078)	(0.0047)	(0.0865)
ϕ_2	−0.8192	−0.8029	−0.8794
	(0.0079)	(0.0045)	(0.0870)
ϕ_3		−0.0226	
		(0.0013)	
σ^2	1.5792e-4	1.6438e-4	1.7402e-4
	(3.9716e-5)	(3.6662e-5)	(0.0132)
SSR	0.0058	0.0063	0.0059
	t-test for zero residuals		
t	0.0999	0.1992	0.1682
	(0.9209)	(0.8431)	(0.8673)
	Portmanteau χ^2 *test*		
lag=6	1.7162	1.3930	2.2226
	(0.7878)	(0.7072)	(0.6949)
lag=12	7.8260	8.7570	8.0684
	(0.6458)	(0.4600)	(0.6222)
lag=18	11.7533	11.6486	12.3483
	(0.7608)	(0.7054)	(0.7197)
lag=24	15.4693	16.6076	15.6963
	(0.8413)	(0.7346)	(0.8307)

APPENDIX

Denoting the covariance matrix of a stationary and invertible ARMA(p,q) process by $\sigma_\varepsilon^2 \boldsymbol{V}_{(p,q)}$, an expression of $\boldsymbol{V}_{(p,q)}^{-1}$ as only $\boldsymbol{\phi}_p$ and $\boldsymbol{\theta}_q$ follows the theorem by Leeuw (1994). We reconstruct its details for the algorithm in order to be easier programmed.

For the ARMA(p,q) model,

$$\boldsymbol{V}_{p,q}^{-1} = \boldsymbol{G}'(\boldsymbol{M}^{-1})'\Big\{\boldsymbol{I}_n - \boldsymbol{R}(\boldsymbol{R}'\boldsymbol{R} + \boldsymbol{G}^{(1)\prime}\boldsymbol{G}^{(1)} - \boldsymbol{G}^{(2)}\boldsymbol{G}^{(2)\prime})^{-1}\boldsymbol{R}'\Big\}\boldsymbol{M}^{-1}\boldsymbol{G}, \quad \text{(A.1)}$$

where $\boldsymbol{R}_{n\times p} = \boldsymbol{M}^{-1}\boldsymbol{G}\boldsymbol{N} - \boldsymbol{H}$,

$$\boldsymbol{M}_{n\times n} = \begin{pmatrix} \boldsymbol{M}^{(1)}_{q\times q} & \boldsymbol{0}_{q\times(n-q)} \\ \boldsymbol{M}^{(2)}_{q\times q} & \left(\boldsymbol{M}^{(1)}_{q\times q} \mid \boldsymbol{0}_{q\times(n-2q)}\right) \\ \boldsymbol{0}_{(n-2q)\times q} & \boldsymbol{M}^{(3)}_{(n-2q)\times(n-q)} \end{pmatrix} \quad \text{and} \quad \boldsymbol{N}_{n\times q} = \begin{pmatrix} \boldsymbol{M}^{(2)}_{q\times q} \\ \boldsymbol{0}_{(n-q)\times q} \end{pmatrix}$$

with

$$\boldsymbol{M}^{(1)}_{q\times q} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\theta_1 & 1 & 0 & \cdots & 0 & 0 \\ -\theta_2 & -\theta_1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -\theta_{q-2} & -\theta_{q-3} & -\theta_{q-4} & \cdots & 1 & 0 \\ -\theta_{q-1} & -\theta_{q-2} & -\theta_{q-3} & \cdots & -\theta_1 & 1 \end{pmatrix},$$

$$\boldsymbol{M}^{(2)}_{q\times q} = \begin{pmatrix} -\theta_q & -\theta_{q-1} & -\theta_{q-2} & \cdots & -\theta_2 & -\theta_1 \\ 0 & -\theta_q & -\theta_{q-1} & \cdots & -\theta_3 & -\theta_2 \\ 0 & 0 & -\theta_q & \cdots & -\theta_4 & -\theta_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -\theta_q \end{pmatrix},$$

$$\boldsymbol{M}^{(3)}_{(n-2q)\times(n-q)} = \begin{pmatrix} -\theta_q & -\theta_{q-1} & \cdots & -\theta_2 & -\theta_1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -\theta_q & \cdots & -\theta_3 & -\theta_2 & -\theta_1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & -\theta_4 & -\theta_3 & -\theta_2 & -\theta_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & -\theta_q & -\theta_{q-1} & \cdots & -\theta_2 & -\theta_1 & 1 \end{pmatrix},$$

and the matrices $\boldsymbol{G}$ and $\boldsymbol{H}$ have the same structure as $\boldsymbol{M}$ and $\boldsymbol{N}$ with θ_j being replaced by ϕ_j and order q by p, respectively.

The determinant of $\boldsymbol{V}_{p,q}$ is given by

$$|\boldsymbol{V}_{p,q}| = \Big|\boldsymbol{I}_{\max\{p,q\}} + (\boldsymbol{G}^{(1)'}\boldsymbol{G}^{(1)} - \boldsymbol{G}^{(2)}\boldsymbol{G}^{(2)'})^{-1}(\boldsymbol{G}^{(1)}\ \boldsymbol{M}^{(2)} - \boldsymbol{M}^{(1)}\ \boldsymbol{G}^{(2)})' \times \boldsymbol{M}_1{}'\boldsymbol{M}_1(\boldsymbol{G}^{(1)}\boldsymbol{M}^{(2)}\ -\ \boldsymbol{M}^{(1)}\boldsymbol{G}^{(2)})\Big|, \tag{A.2}$$

where $\boldsymbol{M}_1$ is an $n \times p$ matrix consisting of the first p colums of $\boldsymbol{M}^{-1}$.

When computing $\boldsymbol{V}_{p,q}^{-1}$ and $|\boldsymbol{V}_{p,q}|$ using (A.1) and (A.2), respectively, the orders of submatrices need to be adjusted; if $0 < p < q$ then make $q-p$ imaginary parameters, $\phi_{p+1} = \phi_{p+2} = \cdots = \phi_q = 0$, and use the order q instead of p in the specification of $\boldsymbol{G}^{(1)}$ and $\boldsymbol{G}^{(2)}$. Also, if $0 < q < p$ then make $p-q$ imaginary parameters, $\theta_{q+1} = \theta_{q+2} = \cdots = \theta_p = 0$, and use the order p instead of q in the specification of $\boldsymbol{M}^{(1)}$ and $\boldsymbol{M}^{(2)}$. To get $\boldsymbol{V}_{p,0}^{-1}$ and $|\boldsymbol{V}_{p,0}|$ for a AR(p) model, set $\boldsymbol{M} = \boldsymbol{I}_n$, $\boldsymbol{N} = 0_{n\times p}$, $\boldsymbol{M}^{(1)} = \boldsymbol{I}_{p\times p}$, and $\boldsymbol{M}^{(2)} = \boldsymbol{0}_{p\times p}$. To get $\boldsymbol{V}_{0,q}^{-1}$ and $|\boldsymbol{V}_{0,q}|$ for a MA(q) model, set $\boldsymbol{G} = \boldsymbol{I}_n$, $\boldsymbol{H} = \boldsymbol{0}_{n\times q}$, $\boldsymbol{G}^{(1)} = \boldsymbol{I}_{q\times q}$, and $\boldsymbol{G}^{(2)} = \boldsymbol{0}_{q\times q}$.

Most tedious job in the computation of $\boldsymbol{V}_{p,q}^{-1}$ and $|\boldsymbol{V}_{p,q}|$ is the computation of inverse matrix, $\boldsymbol{M}^{-1}$ of an $n \times n$ lower band matrix, $\boldsymbol{M}$. We use the following algorithm for the computation of $\boldsymbol{M}^{-1}$ in this paper. Given $\boldsymbol{M}$, $\boldsymbol{M}^{(1)}$, $\boldsymbol{M}^{(2)}$, n, p, q,

Step 1. (Initial step)

$m = [n/q]$, $d = remainder(n/q)$,
$\boldsymbol{B}_1 = \text{inv}(\boldsymbol{M}^{(1)})$,
$\boldsymbol{B}_2 = \boldsymbol{M}^{(2)} * \boldsymbol{B}_1$,
$\boldsymbol{B}_3 = \boldsymbol{M}(q*(m-1)+1 : n,\ q*(m-1)+1 : n)$,
$\boldsymbol{B}_4 = \text{inv}(\boldsymbol{B}_3)$,
$\boldsymbol{B}_5 = \boldsymbol{B}_4(1 : q+d,\ 1 : q)$,
$\boldsymbol{A}_{21} = -\boldsymbol{B}_5 * \boldsymbol{B}_2$,
$j = 1$,

Step 2. (Iterative step)

while ($j \le m-1$),
$\boldsymbol{A}_{11} = \boldsymbol{B}_1$,
$\boldsymbol{A}_{12} = \boldsymbol{0}_{q\times(q*j+d)}$,
$\boldsymbol{A}_{22} = \boldsymbol{B}_4$,
$\boldsymbol{B}_4 = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}$,

$$A_{21} = -\begin{pmatrix} B_1 \\ A_{21} \end{pmatrix} * B_2,$$

$j = j + 1,$

end,

Step 3. (Final step)

$M^{-1} = B_4.$

This algorithm requires the computation of the inverse of a matrix with at most the dimension $q + remainder(n/q)$ instead of the inverse of a $n \times n$ lower band matrix.

REFERENCES

BARNDORFF-NIELSEN, O. AND SCHOU, G. (1973). "On the parametrization of autoregressive models by partial autocorrelations", *Journal of Multivariate Analysis*, **3**, 408–419.

BERGER, J. O. AND BERNARDO, J. M. (1992). "On the development of reference priors", In *Bayesian Statistics IV* (J. M. Bernardo, M. H. DeGroat, D. V. Lindley and A. F. M. Smith, eds.), 61–77, Oxford University Press, London.

BERGER, J. O. AND MORTERA, J. (1999). "Default bayes factors for nonnested hypothesis testing", *Journal of the American Statistical Association*, **94**, 542–554.

BERGER, J. O. AND PERICCHI, L. R. (1996a). "The intrinsic Bayes factor for model selection and prediction", *Journal of the American Statistical Association*, **91**, 109–122.

BERGER, J. O. AND PERICCHI, L. R. (1996b). "The intrinsic Bayes factor for linear models", In *Bayesian Statistics V* (J. M. Bernardo, M. H. DeGroat, D. V. Lindley and A. F. M. Smith, eds.), 23–42, Oxford University Press, London.

CHIB, S. AND GREENBERG, E. (1994). "Bayes inference in regression models with ARMA(p, q) errors", *Journal of Econometrics*, **64**, 183–206.

JONES, M. C. (1987). "Randomly choosing parameters from the stationarity and invertibility region of autoregressive moving average models", *Applied Statistics*, **36**, 134–138.

LEEUW, J. (1994). "The covariance matrix of ARMA errors in closed form", *Journal of Econometrics*, **63**, 397–405.

MARRIOTT, J., RAVISHANKER, N., GELFAND, A. AND PAI, J. (1995). "Bayesian analysis of ARMA processes : Complete sampling based inference under exact likelihoods", In *Bayesian Statistics and Econometrics : Essays In Honor of Arnold Zellner* (D. Berry, K. Chaloner and J. Geweke, eds.), 241–256, John Wiley & Sons, New York.

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

MONAHAN, J. F. (1984). "A Note on enforcing stationarity in autoregressive-moving average models", *Biometrika*, **71**, 403–404.

O'HAGAN, A. (1995). "Fractional Bayes factors for model comparison", *Journal of the Royal Statistical Society*, **B57**, 99–138.

PICCOLO, D. (1982). "The size of the stationarity and invertibility region of an autoregressive-moving average process", *Journal of Time Series Analysis*, **3**, 245–247.

SAS INSTITUTE INC. (2000). *SAS/ETS User's Guide, version 6* (2nd ed.), SAS Institue Inc., Cary.

SHILOV, G. E. (1961). *An Introduction to the Theory of Linear Spaces*, Prentice-Hall, Englewood Cliffs.

SON, Y. S. (1999). "ARMA model identification using the Bayes factor", *Journal of the Korean Statistical Society*, **28**, 503–513.

THE MATH WORKS INC. (2000). *Statistics Toolbox : For Use with MATLAB, User's Guide, Version 3*, The Math Works Inc., Natick.

VARSHAVSKY, J. A. (1996). "Intrinsic Bayes Factors for Model Selection with Autoregressive data", In *Bayesian Statistics V* (J. M. Bernardo, M. H. DeGroat, D. V. Lindley and A. F. M. Smith, eds.), 757–763, Oxford University Press, London.