The Korean Communications
in Statistics Vol. 8, No. 3, 2001
pp. 777-786

Bayesian Approach for Determining the Order p in Autoregressive Models

Chansoo Kim[1] and Younshik Chung[2]

Abstract

The autoregressive models have been used to describe a wide variety of time series. Then the problem of determining the order in the times series model is very important in data analysis. We consider the Bayesian approach for finding the order of autoregressive(AR) error models using the latent variable which is motivated by Tanner and Wong(1987). The latent variables are combined with the coefficient parameters and the sequential steps are proposed to set up the prior of the latent variables. Markov chain Monte Carlo method(Gibbs sampler and Metropolis-Hasting algorithm) is used in order to overcome the difficulties of Bayesian computations. Three examples including $AR(3)$ error model are presented to illustrate our proposed methodology.

Keywords : Autoregressive model; $AR(p)$; Gibbs sampler; Latent variable; Sequential step.

1. Introduction

The autoregressive models have been used to describe a wide variety of time series. A general specification of the autoregressive process of order p, $AR(p)$, is given by

$$(x_t - m_t) = \phi_0 + \phi_1(x_{t-1} - m_{t-1}) + \phi_2(x_{t-2} - m_{t-2}) + \cdots$$
$$+ \phi_p(x_{t-p} - m_{t-p}) + \varepsilon_t, \tag{1.1}$$

where ε_t is a Gaussian white noise process with variance σ^2.

The process is stationary if in addition to the polynomial root condition, m_t is not depend on t. It is clear that different formulation of the $AR(p)$ model can obtained by different choices of the m_t .

The AR means model is obtained by taking m_t to be an arbitrary constant. Also, the

1) Research Institute of Computer Information and Communication, Pusan National University, Pusan, 609-735, Korea
2) Professor, Department of Statistics, Pusan National University, Pusan, 609-735, Korea
E-mail : yschung@hyowon.cc.pusan.ac.kr

AR error model obtained by taking $m_t = \mu$, a process mean parameter to be assumed constant for all t, together with $\phi_0 = 0$ to avoid over-parameterisation, is given by

$$(x_t - \mu) = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \cdots + \phi_p(x_{t-p} - \mu) + \varepsilon_t \tag{1.2}$$

and the error ε_t is white noise Gaussian with mean zero and variance $\sigma^2 > 0$.

For general class of time series models, a Bayesian analysis of the problem of consistent model selection is big issue. A Bayesian approach to model selection has attracted attention in the work of Atkinson(1978) and others. Pokitt and Tremayne(1983) consider a Bayesian analysis of the problem of consistent model selection for general class of time series models using the posterior odds. McCulloch and Tsay(1994) and Chib and Greenberg(1994) have utilized the Gibbs sampler for Bayesian analysis of AR and $ARMA$ processes, respectively. Marriot et al.(1994) have presented model selection not be based upon a single number, but that comparison of predictive performance of competing models be made at each time point in $ARMA(p,q)$. Recently Naylor and Marriott(1996) have presented analyses of the stationary and non-stationary models.

In this paper, we consider the Bayesian approach for finding the order of autoregressive error models. The assessment of the order of an autoregressive model is model selection problem encountered in many applications. Our methodology is oriented to the "sequential step" which is based on the idea of the data augmentation by Tanner and Wong(1987). This data augmentation method is used by George and McCulloch (1993) and Kuo and Mallick(1998) for linear and generalized linear regression using binary indicators for each predictor in the model. Using Gibbs sampling, or similar Markov chain Monte Carlo techniques, these variable selection procedures are set up to compute posterior probabilities for all the 2^p possible models having p different covariates. In the autoregressive model in (1.2), we consider, by introducing the latent variables into the prior of the coefficient parameter, the sequential step for the model selection which means that the model can take the ϕ_j only after $\phi_1, \cdots, \phi_{j-1}$ are reached. Therefore, we do not have to search over all the possible models since the number of autoregressive models is substantially smaller than the number of all possible models. For example, if there are $p=4$, then there are only four autoregressive models to consider; the first, the second, the third and forth order. Then, the marginal posterior probabilities of the latent variables will be computed to decide the order of $AR(p)$. In order to overcome the computational difficulties of the marginal posterior density, we use the Gibbs sampler(Gelfand and Smith, 1990). Also, we propose a simple approach by reparameterising parameters in terms of partial autocorrelation functions in order to run the Gibbs sampler easily in $AR(p)$ model.

The rest of the paper is organized as follows. Section 2 introduces the Bayesian formulation of $AR(p)$ model. Also, the "sequential step" is proposed in order to assign the probability distribution to the latent variables. In Section 3, to overcome the computational difficulty, we use Markov chain Monte Carlo method including the Gibbs sampler and Metropolis algorithm. Therefore, we derive the full conditional densities and a simple method is proposed for obtaining parameters randomly from the stationarity region of an $AR(p)$ process. Finally, in Section 4, our proposed methodology is applied to three examples.

2. Bayesian formulation for AR(p)

We consider the model in which an observation $\boldsymbol{y}$, at time t, is generated by autoregressive process of order p. Let $y_t=(x_t-\mu)$ in the model (1.2). Then we can rewrite the model (1.2) as follows;

$$y_t=\phi_1 y_{t-1}+\phi_2 y_{t-2}+\cdots+\phi_p y_{t-p}+\varepsilon_t, \quad t=1,\cdots,n. \tag{2.1}$$

This process is said to be stationary if the roots of

$$1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p=0$$

lie outside the unit circle and B is a back shift operator.

2.1. Likelihood

Motivated by the work of George and McCulloch(1993) and Kuo and Mallick(1998), we explore a simple method of subset selection. Instead of building a hierarchical model, we embed indicator variables in the autoregressive error model in (2.1). In other words, the order of autoregressive models is completely determined by the binary vector $\gamma=(\gamma_1,\ldots,\gamma_p)$. Let γ_j be an indicator variable supported at two points 1 and 0. We write the autoregressive error model for the ith subject, $i=1,\ldots,n$, by

$$y_t=\sum_{j=1}^{p}\gamma_j\phi_j y_{t-j}+\varepsilon_t, \qquad t=1,\cdots,n, \tag{2.2}$$

and

$$\varepsilon_t \sim N(0,\sigma^2)$$

where $\phi=(\phi_1,\ldots,\phi_p)$ is the usual unknown column vector of coefficients. When $\gamma_j=1$ if the order is added and $\gamma_j=0$ if the order is dropped. We assume that our model has at most the order p. The model (2.2) can be rewritten as the distribution of y_t given H_{t-1} and θ, where H_{t-1} is the past history, is

$$f(y_t|H_{t-1},\ \theta)\sim N(\widehat{y_{t|t-1,\ \theta}},\sigma^2) \tag{2.3}$$

where $\widehat{y_{t|t-1,\ \theta}}=\gamma_1\phi_1 y_{t-1}+\cdots+\gamma_p\phi_p y_{t-p}$, and $\theta=(\gamma_1,\cdots,\gamma_p,\phi_1,\cdots,\phi_p,\sigma^2)$. By the law of total probability, the joint of the observations, conditioned on the first p, is

$$f(y_{p+1},\cdots,y_n|y_p,\cdots,y_1,\ \theta)=\prod_{t=p+1}^{n} f(y_t|H_{t-1},\ \theta)$$

$$\propto\sigma^{-(n-p)}\exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{n}(y_t-\gamma_1\phi_1 y_{t-1}-\cdots-\gamma_p\phi_p y_{t-p})^2\right). \tag{2.4}$$

2.2. Prior distributions

Our interest is to combine the data $y=(y_1,\cdots,y_n)$, with the prior information, if available, to obtain the posterior of the unknown parameters $\theta=(\Phi,\sigma^2,r)$, where $\mathbf{r}=(\gamma_1,\cdots,\gamma_p)$ and $\Phi=(\phi_1,\cdots,\phi_p)$. Therefore, we suppose that the prior distribution of (Φ,r,σ^2) is given by

$$\pi[\Phi,\sigma^2,r]=\prod_{i=1}^{p}\pi[\phi_i]\pi[\sigma^2]P[\gamma_1,\cdots,\gamma_p]. \tag{2.5}$$

And let,

$$\pi[\Phi]=\pi[(\phi_1,\cdots,\phi_p)]\sim N_p(\mu_\phi,\sigma_\phi^2 I)I(\Phi\in C_p)$$

where I is an identity matrix and $I(\Phi\in C_p)$ denotes the indicator function and C_p is p -dimensional hypercube to satisfy the stationarity. Let

$$\pi[\sigma^2]\propto\sigma^{-2}.$$

In the autoregressive model $AR(p)$ with the order p, we consider the sequential step for our model selection which means that the model can take the ϕ_j only after $\phi_1,\ldots,\phi_{j-1}$ are reached. In other words, to get the coefficient parameter ϕ_j, one must pass through the coefficients $\phi_1,\ldots,\phi_{j-1}$. Let $p_j=\Pr(\gamma_j|\gamma_1,\ldots,\gamma_{j-1})$. Then, p_j can be determined when $\gamma_1=\cdots=\gamma_{j-1}=1$, but if at least one of the values of $\gamma_1,\ldots,\gamma_{j-1}$ is zero, p_j must be zero. Therefore, the joint prior density $\pi(r)$ of gamma may be defined as follows; let $P(\gamma_1=1)=p_1$ and $P(\gamma_1=0)=1-p_1$. We assume that ϕ_1 exists always, i.e. $P(\gamma_1=1)=p_1=1$. Since the value of ϕ_2 is considered only when ϕ_1 exits, let $P(\gamma_2=1|\gamma_1=1)=p_2$ and $P(\gamma_2=0|\gamma_1=1)=1-p_2$ Thus $P(\gamma_2=0|\gamma_1=0)=1$, $P(\gamma_2=1|\gamma_1=0)=0$. Also, let $P(\gamma_3=1|\gamma_2=1,\gamma_1=1)=p_3$,

$P(\gamma_3=0|\gamma_2=1, \gamma_1=1)=1-p_3$. Then $P(\gamma_3=0|\gamma_2=0, \gamma_1=1)=1$ and

$$P(\gamma_3=1|\gamma_2=0, \gamma_1=1)=P(\gamma_3=1|\gamma_2=1, \gamma_1=0)=$$
$$P(\gamma_3=1|\gamma_2=0, \gamma_1=0)=P(\gamma_3=0|\gamma_2=1, \gamma_1=0)=0.$$

Therefore, the joint prior density $P(\gamma_1,\cdots,\gamma_p)$ of $\gamma_1,\cdots,\gamma_p$ is of the general form

$$P(\gamma_1,\cdots,\gamma_p)=\left(\prod_{i=1}^{p}p_i\right)^{I(\gamma_p=\cdots=\gamma_1=1)}\times\left((1-p_p)\prod_{i=1}^{p-1}p_i\right)^{I(\gamma_p=0,\gamma_{p-1}=\cdots=\gamma_1=1)}$$
$$\times\cdots\times(p_1(1-p_2))^{I(\gamma_p=\cdots=\gamma_2=0,\gamma_1=1)}. \qquad (2.6)$$

For a given prior density function, $\pi(\boldsymbol{\Phi},\sigma^2,\boldsymbol{r})$, the posterior of interest is given by

$$\pi(\boldsymbol{\Phi},\sigma^2,\boldsymbol{\gamma}|\boldsymbol{y})\propto f(y_{p+1},\cdots,y_n|y_p,\cdots,y_1,\boldsymbol{\Phi},\sigma^2,\boldsymbol{r})\pi(\boldsymbol{\Phi},\sigma^2,\boldsymbol{r}). \qquad (2.7)$$

3. Full conditional distributions

Densities are denoted generically by brackets, so joint, conditional and marginal forms, for example, appear as, $[X,Y],[X|Y]$ and $[X]$, respectively. Given a prior distribution on $\pi(\boldsymbol{\Phi},\sigma^2,\boldsymbol{r})$, the joint posterior distribution for $(\boldsymbol{\Phi},\sigma^2,\boldsymbol{r})$ given $\boldsymbol{y}$ is

$$[\boldsymbol{\Phi},\sigma^2,\boldsymbol{r}|\boldsymbol{y}]\propto\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n-p}{2}}\exp\left(-\frac{\sum_{t=p+1}^{n}(y_t-\sum_{j=1}^{p}\gamma_j\phi_j y_{t-j})^2}{2\sigma^2}\right)$$
$$\times\left(\frac{1}{2\pi\sigma^2_\phi}\right)^{\frac{p}{2}}\exp\left(-\frac{\sum_{k=1}^{p}(\phi_k-\mu_\phi)^2}{2\sigma^2_\phi}\right)$$
$$\times\left(\frac{1}{\sigma^2}\right)^{-1}P[\gamma_1,\cdots,\gamma_p]. \qquad (3.1)$$

Bayesian inference proceeds by obtaining marginal posterior distribution of the components of $\boldsymbol{\theta}=(\boldsymbol{\Phi},\sigma^2,\boldsymbol{r})$ as well as features of these distributions. The Gibbs sampler introduced by Gelfand and Smith(1990) as a tool for carrying out Bayesian calculations, provided samples from the posterior in (3.1). This requires sampling from the complete conditional distributions associated with $\boldsymbol{\theta}$, each of which is proportional to the right side of (3.1). In order to obtain the joint conditional density of $(\gamma_1,\cdots,\gamma_p)$, let $S=\{S_1=(1,0,0,\cdots,0),S_2=(1,1,0,\cdots,0),\cdots,S_p=(1,1,\cdots,1)\}$ be a set which has p elements. Therefore, the full conditional densities are needed as follows;

$$[\sigma^2 | \boldsymbol{\Phi}, \boldsymbol{r}, \boldsymbol{y}] \sim \mathrm{IG}\left(\frac{n-p}{2}, \frac{\sum_{t=p+1}^{n}\left(y_t - \sum_{j=1}^{p} \gamma_j \phi_j y_{t-j} \right)^2}{2} \right). \tag{3.2}$$

$$[\phi_1, \cdots, \phi_p | \sigma^2, \boldsymbol{r}, \boldsymbol{y}] \propto \exp\left\{ -\frac{1}{2\sigma^2} \sum_{t=p+1}^{n}\left(y_t - \sum_{j=1}^{p} \gamma_j \phi_j y_{t-j} \right)^2 - \frac{1}{2\sigma_\phi^2} \sum_{k=1}^{p} (\phi_k - \mu_\phi)^2 \right\} \tag{3.3}$$

subject to the restriction to p-dimensional hypercube, C_p.

$$[(\gamma_1, \cdots, \gamma_p) \in S_i | \sigma^2, \boldsymbol{\Phi}, \boldsymbol{y}] = \frac{[\boldsymbol{y}, \sigma^2, \boldsymbol{\Phi} | (\gamma_1, \cdots, \gamma_p) \in S_i] P[(\gamma_1, \cdots, \gamma_p) \in S_i]}{\sum_{S_i \in S} [\boldsymbol{y}, \sigma^2, \boldsymbol{\Phi} | (\gamma_1, \cdots, \gamma_p) \in S_i] P[(\gamma_1, \cdots, \gamma_p) \in S_i]}. \tag{3.4}$$

In sampling scheme, the conditional distribution of σ^2 is straightforward. But, in equation (3.3), the joint conditional distribution of $(\phi_1, \cdots, \phi_p)$ has a restriction to satisfy the stationarity condition. This restriction should be incorporated in estimation. Therefore, we make use of some well-known time series results which are usefully summarized in notes of Monahan(1984) and Jones(1987).

Let $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_p)$ be the first p partial autocorrelations function of process. By reparameterizing $(\phi_1, \cdots, \phi_p)$ in terms of $(\rho_1, \cdots, \rho_p)$ as in Jones(1987) the stationarity constraints become $|\rho_i| < 1$ for $i = 1, \cdots, p$. The detail descriptions are as followings:
Set $\phi_1^{(1)} = \rho_1$ and

$$\phi_i^{(k)} = \phi_i^{(k-1)} - \rho_k \phi_{k-i}^{(k-1)}, \quad i = 1, \cdots, k-1 \tag{3.5}$$

with $\phi_k^{(k)} = \rho_k$ for $k = 2, \cdots, p$. Finally, set $\boldsymbol{\phi} = (\phi_1^{(p)}, \cdots, \phi_p^{(p)})$. The Jacobian of the transformation is

$$J = \prod_{k=2}^{p} (1-\rho_k)^{[k/2]} (1+\rho_k)^{[(k-1)/2]} \tag{3.6}$$

for $p \geq 2$ and is one when $p = 1$ and $[x]$ denotes the integer part of x .

The advantage of working with partial autocorrelations is that the stationarity conditions translate to the simple constraints that the partial autocorrelations range freely in the hypercube $C_1 \times \cdots \times C_1$, i.e., $|\rho_i| < 1$, $i = 1, \cdots, p$. For example, let $p = 3$, the transformation is simply obtained as follows;

$$\phi_1 = \rho_1 - \rho_1\rho_2 - \rho_2\rho_3,\ \ \phi_2 = \rho_2 - \rho_1\rho_3 - \rho_1\rho_2\rho_3,\ \ \phi_3 = \rho_3 \tag{3.7}$$

where $|\rho_i| < 1$, $i = 1, 2, 3$ and the Jacobian of the transformation is

$$J = (1-\rho_2)(1-\rho_3^2).$$

Therefore, the full conditional distribution of (3.3) in $AR(3)$ model is transformed into the following forms;

$$[\rho_1 | \rho_2, \rho_3, \sigma^2, \boldsymbol{r}, \boldsymbol{y}] \sim N^+\left(\frac{B}{A}, A^{-1}\right),$$

$$[\rho_2 | \rho_2, \rho_3, \sigma^2, \boldsymbol{r}, \boldsymbol{y}] \propto (1-\rho_2) N^+\left(\frac{D}{C}, C^{-1}\right),$$

$$[\rho_3 | \rho_1, \rho_2, \sigma^2, \boldsymbol{r}, \boldsymbol{y}] \propto (1-\rho_3^2) N^+\left(\frac{F}{E}, E^{-1}\right),$$

where

$$A = \frac{\sum_{t=4}^{n}[\rho_3(1-\rho_2)\gamma_2 y_{t-2} - \gamma_1 y_{t-1}(1-\rho_2)]^2}{\sigma^2} + \frac{[(1-\rho_2)^2 + (1-\rho_2)^2\rho_3^2]}{\sigma^2_\phi},$$

$$B = -\sum_{t=4}^{n}\left(\frac{\rho_3(1-\rho_2)\gamma_2 y_{t-2} - \gamma_1 y_{t-1}(1-\rho_2)}{\sigma^2}\right)$$

$$\times\left(\frac{y_t + \gamma_1 y_{t-1}\rho_2\rho_3 - \gamma_2 y_{t-2}\rho_2 - \gamma_3\rho_3 y_{t-3}}{\sigma^2}\right)$$

$$+\frac{(1-\rho_2)(\rho_2\rho_3 + \mu_\phi) + \rho_3(1-\rho_2)(\rho_2 - \mu_\phi)}{\sigma^2_\phi},$$

$$C = \frac{\sum_{t=4}^{n}[(\rho_1+\rho_3)\gamma_1 y_{t-1} - \gamma_2 y_{t-2}(1+\rho_1\rho_3)]^2}{\sigma^2} + \frac{(1+\rho_3)^2 + (1+\rho_1\rho_3)^2}{\sigma^2_\phi},$$

$$D = -\sum_{t=4}^{n}\left(\frac{(\rho_1+\rho_3)\gamma_1 y_{t-1} - \gamma_2 y_{t-2}(1+\rho_1\rho_3))}{\sigma^2}\right)$$

$$\times\left(\frac{y_t + \gamma_1 y_{t-1}\rho_1 + \gamma_2 y_{t-2}\rho_1\rho_3 - \gamma_3\rho_3 y_{t-3}}{\sigma^2}\right)$$

$$+\frac{(1+\rho_3)(\rho_1 - \mu_\phi) + (1+\rho_1\rho_3)(\rho_1\rho_3 + \mu_\phi)}{\sigma^2_\phi},$$

$$E=\frac{\sum_{t=4}^{n}[\rho_2\gamma_1 y_{t-1}+\gamma_2 y_{t-2}\rho_1(1-\rho_2)-\gamma_3 y_{t-3}]^2}{\sigma^2}+\frac{\rho_2^2+\rho_1^2(1-\rho_2)^2+1}{\sigma_\phi^2},$$

$$F=-\sum_{t=4}^{n}\left(\frac{\gamma_1 y_{t-1}\rho_2+\gamma_2 y_{t-2}\rho_1(1-\rho_2)-\gamma_3 y_{t-3}}{\sigma^2}\right)$$

$$\times\left(\frac{y_t-\gamma_1 y_{t-1}\rho_1(1-\rho_2)-\gamma_2 y_{t-2}\rho_2}{\sigma^2}\right)$$

$$+\frac{\rho_2(\rho_1(1-\rho_2)-\mu_\phi)+\mu_\phi+\rho_1(1-\rho_2)(\rho_2-\mu_\phi)}{\sigma_\phi^2},$$

and $N^+(a,b)$ denotes the truncated normal distribution with mean a and variance b. Hence, we have a simple method for obtaining parameters randomly from the stationarity region of an $AR(3)$ process. Using the one-for-one method and Metropolis-Hasting algorithm, we can generate three independent variates (ρ_1,ρ_2,ρ_3) and apply to the transformation (3.7) and then obtain (ϕ_1,ϕ_2,ϕ_3). Extension of the method to $AR(p)$ is immediate. Finally, the joint conditional probability of (3.4) is obtained by using Bayes theorem and this sampling is straightforward. The scheme goes through the sampling steps in (3.2)-(3.4) until the convergence is achieved.

4. Examples

We apply our proposed methodology to three examples involving autoregressive processes with different orders. The white noise ε_t was generated from normal distribution with mean 0 and variance 1. We deal with simulated data with the sample size $n=50$. Their true models are as follows;

Data 1 ($AR(1)$ model) : Let $\phi_1=0.7$.

Data 2 ($AR(2)$ model) : Let (ϕ_1,ϕ_2) = (0.2,0.5).

Data 3 ($AR(3)$ model) : Let (ϕ_1,ϕ_2,ϕ_3) = (0.7,-0.5,0.4).

These values of $\boldsymbol{\Phi}=(\phi_1,\phi_2,\phi_3)$ in each model ensure the condition of the stationarity in their models.

In the implementation of the Gibbs sampler, the first 200 draws are discarded and the algorithm is run to obtain 800 draws from the posterior. The results for Data 1, 2 and 3 are presented in Table 4.1 which shows the frequency of the orders over 800 runs. In Table 4.1, the values within parenthesis denote the percentage of each order which is the estimate of its posterior probability.

Table 4.1. The frequency of order in AR error model

True	Model Selection			Total
	order 1	order 2	order 3	
AR(1)	640(0.8)	112(0.14)	48(0.06)	800
AR(2)	26(0.0325)	575(0.71875)	199(0.24875)	800
AR(3)	3(0.00375)	1(0.00125)	796(0.995)	800

Data 1 is generated from $AR(1)$ model. Table 4.1 says that the posterior probability of supporting $AR(1)$ is 0.8, the posterior probability of supporting $AR(2)$ is 0.14 and that of supporting $AR(3)$ is 0.06. Therefore we conclude that $AR(1)$ model is supported. For Data 2 and 3, we can get the similar results from Table 4.1. Also, our methodology can be easily extended and applied to moving average process and to $ARMA(p,q)$ model.

References

[1] Atkinson, A. C. (1978) Posterior probabilities for choosing a regression model, *Biometrika,* 65, 39-48.

[2] Chib, S. and Greenberg, E. (1964) Bayes inference in regression models with ARMA(p,q) errors, *Journal of Econometrics*, 64, 183-206.

[3] Gelfand, A. E. and Smith A. F. M. (1990) Sampling based approaches to calculating marginal densities, *Journal of the American Statistical Association*, 85,398-409.

[4] George, E. I. and McCulloch, R. E. (1993) Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association*, 88, 881-889.

[5] Jones, M. C. (1987) Randomly choosing parameters from the stationarity and invertibility region of autoregressive-moving average models, *Applied Statistics*, 36, 134-138.

[6] Kuo, L. and Mallick, B. (1998) Variable selection for regression models. *Sankhya* B, 60, 65-81.

[7] Marriott, J., Ravishanker, N., Gelfand, A. and Pai, J. (1994) Bayesian analysis of ARMA process : Complete sampling-based inference under exact likelihoods, *Bayesian Analysis in Statistical and Econometrics*, edited by D. Berry, K. Chaloner and J, Geweke, 243-256.

[8] McColloch, R. and Tsay, R. (1994) Bayesian analysis of autoregressive time series via the Gibbs Sampler, *Journal of Time series Analysis*, 15, 235-250.

[9] Monahan, J. F. (1984) A note on enforcing stationarity in autoregressive-moving average process, *Biometrika*, 71, 403-404.

[10] Naylor, J. and Marriott, J. (1996) A Bayesian analysis of non-stationary AR series, *Bayesian Statistics* 5, 705-712.

[11] Pokitt, D. S. and Tremayne, A. R. (1983) On the posterior odds of time series models, *Biometrika*, 70, 157-162.

[12] Tanner, M. A. and Wong, W. H. (1987) The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association,* 82, 528-550.