The Korean Communications
in Statistics Vol. 10 No. 3, 2003
pp. 981-996

Bayesian Inference for Switching Mean Models with ARMA Errors[1]

Young Sook Son[2], Seong W. Kim[3], Sinsup Cho[4]

Abstract

Bayesian inference is considered for switching mean models with the ARMA errors. We use noninformative improper priors or uniform priors. The fractional Bayes factor of O'Hagan (1995) is used as the Bayesian tool for detecting the existence of a single change or multiple changes and the usual Bayes factor is used for identifying the orders of the ARMA error. Once the model is fully identified, the Gibbs sampler with the Metropolis-Hastings subchains is constructed to estimate parameters. Finally, we perform a simulation study to support theoretical results.

Key Words : switching mean model; multiple change points; *ARMA* error; noninformative improper prior; fractional Bayes factor; Gibbs sampler; Metropolis-Hastings algorithm.

1. INTRODUCTION

Change point problems originally arisen in quality control have received interests in many fields. The bulk of studies on change point problems in frequentist perspective are found in Csörgő and Horváth (1997). Our interest in this paper is in the change point analysis of time series models with switching means using the Bayesian approach.

Ohtani (1982) presented a Bayesian procedure for estimating parameters of the switching regression model under noninformative priors when the subset of regression coefficients shifts and the error terms are generated by the $AR(1)$ process. Albert and Chip (1993) discussed Bayesian inference via Gibbs sampling for autoregressive time series models with Markov jumps in mean and variance. Garisch and Groenewald (1999) dealt with Bayesian change point analysis in the linear model with correlated errors. They assumed the multivariate nomal prior for a vector of regression parameters, the noninformative improper prior for the variance of white noises, and an uniform prior over (-1,1) for the correlation of errors. Two well-known

---

1)This work was supported by grant No. 2000-1-10400-003-2 from the Basic Research Program of the Korea Science Engineering Foundation.
2) Department of Statistics, Chonnam National University, Kwangju, 500-757, Korea.
E-mail: ysson@chonnam.ac.kr
3) Division of Applied Mathematics, Hanyang University, Ansan, Kyungkido, 426-791, Korea.
4) Department of Statistics, Seoul National University, Seoul, 151-742. Korea.

criteria are used for identifying the number of change points along with their positions. They are the arithmetic intrinsic Bayes factor (AIBF) of Berger and Pericchi (1996) and the fractional Bayes factor (FBF) of O'Hagan (1995).

Consider a simple regression model,

$$Y_t = \mu + \varepsilon_t . \tag{1}$$

We often call this a constant mean model. If a set of time series data, $\{y_t, t=1,2,\ldots,n\}$, is generated from the model in (1), the error term $\{\varepsilon_t\}$ will have the structure explaining autocorrelations of time series data. The model in (1) generally assumes that the mean is constant over all time periods. But a time series with a globally constant mean is practically very restrictive. There are rather many cases that the mean changes slowly or abruptly as the time passes.

In this paper, we consider the locally constant mean model, $M_{k,\ \boldsymbol{d}_k, p, q}$ with multiple mean changes at unknown time points $\boldsymbol{d}_k = (d_1, d_2, \ldots, d_k)$, assuming the *ARMA(p,q)* error. The proposed model is as follows:

$$M_{k,\ \boldsymbol{d}_k,p,q} : Y_t = \varepsilon_t + \begin{cases} \mu_0, & t=1,2,\ldots,d_1, \\ \mu_1, & t=d_1+1, d_1+2, \ldots, d_2, \\ \vdots & \\ \mu_{k-1}, & t=d_{k-1}+1, d_{k-1}+2, \ldots, d_k, \\ \mu_k, & t=d_k+1, d_k+2, \ldots, n, \end{cases} \tag{2}$$

where $\mu_{j-1} \neq \mu_j$ for $j=1,2,\ldots,k$ and $\{\varepsilon_t\}$ follows an *ARMA(p,q)* process, that is,

$$\Phi_p(B)\varepsilon_t = \Theta_q(B)a_t .$$

Here, $\Phi_p(B) = 1-\phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ and $\Theta_q(B) = 1-\theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$, where B is a backshift operator, and $\{a_t\}$ is a sequence of $N(0,\sigma^2)$ white noises. For this model, $\mu_0, \mu_1, \ldots, \mu_k, \sigma^2, k, \boldsymbol{d}_k, p, q, \underline{\phi}_p = (\phi_1, \phi_2, \ldots, \phi_p)$, and $\underline{\theta}_q = (\theta_1, \theta_2, \ldots, \theta_q)$ are all unknown parameters. For the stationarity and invertibility of $ARMA(p,q)$ error, $(\underline{\phi}_p, \underline{\theta}_q)$ must be in the region $C_p \times C_q$, where

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

We denote the model $M_{0,p,q}$ with $k=0$ in the model $M_{k,\ d_k,p,q}$ as the no-switching model with $\mu_0=\mu_1=\cdots=\mu_k$ in (2), which is known as a stationary and invertible *ARMA(p,q)* process. The model $M_{k,\ d_k,p,q}$ is a nonstationary process in the sense that the mean of process is not a globally constant.

We use the fractional Bayes factor (FBF) of O'Hagan (1995) as a Bayesian tool to determine the number of change points and the usual Bayes factor to identify the orders of *ARMA(p,q)* error. We propose a "binary segmentation" procedure. At the first level, we compare the models between no change point and a single change point using the FBF. If the test is in favor of the change point model, we locate the change point. Then we compute two FBF's similar to what we have done after dividing the data into two parts by the change point. We continue to conduct tests until no more change points are found in a subsegment. For the model being fully identified, we estimate parameters using the Gibbs sampler with the Metropolis-Hastings subchains.

When performing the Bayesian analysis for models including stationary and invertible *ARMA* structure, the most cumbersome problem is the specification of $C_p\times C_q$ for every p and q. We transform the region $C_p\times C_q$ into the region $(-1,1)^{p+q}$ to overcome this difficulty. This transformation is often used when integrating on $(\underline{\phi}_p,\ \underline{\theta}_q)$ for computing Bayes factors or randomly drawing $(\underline{\phi}_p,\ \underline{\theta}_q)$ in the Gibbs sampler (cf. Marriot, et al. (1992); Varshavsky (1995); Son (1999, 2001)).

The contents of this paper are as follows. In Section 2, we build a matrix form of the model and prior assumptions, Also, the exact and explicit likelihood functions are presented. In Section 3, the posterior probabilities of competing models are computed using the FBF for the identification of models. In Section 4, we construct the Gibbs sampler with the Metropolis-Hastings subchains for the estimation of parameters. In Section 5, some simulation results are provided. Finally, we finish this article with short concluding remarks in Section 6.

2. PRIOR ASSUMPTIONS AND LIKELIHOOD FUNCTION

Suppose that $Y_1, Y_2, \ldots, Y_n$ follow the process given in (2). Then, its matrix form is formulated as

$$M_{k,\ d_k,p,q}\ :\quad \boldsymbol{Y}=X_k\,\underline{\mu}_k+\underline{\varepsilon}, \tag{3}$$

where $\boldsymbol{Y}=(Y_1, Y_2,\ldots, Y_n)'$, $\underline{\mu}_k=(\mu_0,\mu_1,\ldots,\mu_k)'$, $\underline{\varepsilon}=(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)'$, and

$$X_k = \begin{bmatrix} \underline{1}_{d_1} & \underline{0}_{d_1} & \underline{0}_{d_1} & \underline{0}_{d_1} & \cdots & \underline{0}_{d_1} \\ & \underline{1}_{d_2-d_1} & \underline{0}_{d_2-d_1} & \underline{0}_{d_2-d_1} & \cdots & \underline{0}_{d_2-d_1} \\ & & \underline{1}_{d_3-d_2} & \underline{0}_{d_3-d_2} & \cdots & \underline{0}_{d_3-d_2} \\ & & & \ddots & \vdots & \vdots \\ & sym & & & & \underline{1}_{n-d_k} \end{bmatrix}$$

is an $n\times(k+1)$ matrix with $\underline{1}_a(\underline{0}_a)$ representing a $a\times 1$ column vector with ones(zeros) as its all elements. Since $\{\varepsilon_t\}$ follows a stationary and invertible *ARMA(p,q)* process, $E(Y)=X_k\,\underline{\mu}_k$ and $Cov(Y)=\sigma^2 V_{p,q}$, where $V_{p,q}$ is an $n\times n$ matrix composed of only $\underline{\phi}_p$ and $\underline{\theta}_q$. For the no-switching model, $M_{0,p,q}$, its matrix form is

$$M_{0,p,q} \; : \quad Y=\mu_0\,\underline{1}_n+\underline{\varepsilon} \tag{4}$$

We assume noninformative priors. Then the prior specifications are as follows:

$$\pi_k^N(\underline{\mu}_k,\sigma) \propto \sigma^{-s}, \quad \underline{\mu}_k \in R^{k+1}=(-\infty,\infty)^{k+1},\; 0<\sigma<\infty,\; s>0, \tag{5}$$

$$\pi_0^N(\mu_0,\sigma) \propto \sigma^{-s}, \quad \mu_0 \in R=(-\infty,\infty),\; 0<\sigma<\infty,\, s>0, \tag{6}$$

$$\pi(\underline{\phi}_p,\underline{\theta}_q|p,q)=I_{C_p\times C_q}(\underline{\phi}_p,\underline{\theta}_q)/\,Volume(C_p\times C_q),$$

where

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

Throughout this paper, the superscript N implies the use of noninformative improper prior or its result. The priors of discrete parameters, k, $\boldsymbol{d}_k$, p, and q, are assumed as uniform priors with each support, $K=\{k_1,k_2,\ldots\}$, $\boldsymbol{D}_k=\{\boldsymbol{d}_{k1},\boldsymbol{d}_{k2},\ldots\}$, $P=\{p_1,p_2,\ldots\}$, and $Q=\{q_1,q_2,\ldots\}$, respectively. Here, all elements of $\boldsymbol{D}_k$ are restricted so that all the parameters of each model generated by change points can be estimated and all elements of K,P, and Q are nonnegative integers. Finally, under the assumption of independence among sets of parameters, the prior of the switching model, $M_{k,\,d_k,p,q}$, for $k=1,2,\ldots$ is

$$\pi_k^N(\boldsymbol{d}_k,\underline{\mu}_k,\sigma,p,q,\underline{\phi}_p,\underline{\theta}_q)\propto\pi(\boldsymbol{d}_k)\cdot\pi_k^N(\underline{\mu}_k,\sigma)\cdot\pi(\underline{\phi}_p,\underline{\theta}_q|p,q) \tag{7}$$

and that of the no-switching model, $M_{0,p,q}$, is

$$\pi_0^N(\mu_0, \sigma, p, q, \underline{\phi}_p, \underline{\theta}_q) \propto \pi_0^N(\mu_0, \sigma) \cdot \pi(\underline{\phi}_p, \underline{\theta}_q | p, q).$$

Now, let an observed sequence of $\boldsymbol{Y}$ be $\boldsymbol{y}=\{y_1, y_2, \ldots, y_n\}'$, then from (3) and (4) the full likelihood functions under model $M_{k,\, d_k, p, q}$ and model $M_{0,p,q}$ can exactly and explicitly written as

$$l_k(\boldsymbol{d}_k, \underline{\mu}_k, \sigma, p, q, \underline{\phi}_p, \underline{\theta}_q | \boldsymbol{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} |V_{p,q}^{-1}|^{\frac{1}{2}}$$

$$\cdot \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-X_k\underline{\mu}_k)' V_{p,q}^{-1}(\boldsymbol{y}-X_k\underline{\mu}_k)\right\}, \tag{8}$$

and

$$l_0(\mu_0, \sigma, p, q, \underline{\phi}_p, \underline{\theta}_q | \boldsymbol{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} |V_{p,q}^{-1}|^{\frac{1}{2}}$$
$$\cdot \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\mu_0 \underline{1}_n)' V_{p,q}^{-1}(\boldsymbol{y}-\mu_0 \underline{1}_n)\right\},$$

where the specifications of $V_{p,q}^{-1}$ and $|V_{p,q}^{-1}|$ are shown in Leeuw (1994).

3. MODEL SELECTION BY THE FRACTIONAL BAYES FACTOR AND THE BAYES FACROR

Consider the problem of identifying a mean change model with multiple change points given a time series data, $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$. First, we are going to test the switching model M_1 with a single mean change against the no-switching model M_0. Now, We define the following function of data $\boldsymbol{y}$, a change point $\boldsymbol{d}_k$, and a constant b $(0 < b \le 1)$ for the model $M_{k,\, d_k, p, q}$ and $M_{0,p,q}$,

$$B_{k0}^N(\boldsymbol{d}_k, \boldsymbol{y} \mid b) = \frac{\sum_{p\in P, q\in Q} m^N{}_{(k,\, d_k, p, q)}(\boldsymbol{y} \mid b)}{\sum_{p\in P, q\in Q} m^N{}_{(0,p,q)}(\boldsymbol{y} \mid b)},$$

where

$$m^N_{(k,\, d_k, p, q)}(\boldsymbol{y} \mid b) = \int_{C_p\times C_q}\int_0^\infty \int_{R^{k+1}} \pi_k^N(\boldsymbol{d}_k, \underline{\mu}_k, \sigma, p, q, \underline{\phi}_p, \underline{\theta}_q)$$

$$\cdot \{l_k(\boldsymbol{d}_k, \underline{\mu}_k, \sigma, p, q, \underline{\phi}_p, \underline{\theta}_q \mid \boldsymbol{y})\}^b \, d\underline{\mu}_k \, d\sigma \, d(\underline{\phi}_p \times \underline{\theta}_q) \tag{9}$$

and

$$m^N_{(0,p,q)}(\mathbf{y} \mid b) = \int_{C_p \times C_q} \int_0^\infty \int_R \pi_0^N(\mu_0, \sigma, p, q, \underline{\phi}_p, \underline{\theta}_q)$$

$$\cdot \{l_0(\mu_0, \sigma, p, q, \underline{\phi}_p, \underline{\theta}_q \mid \mathbf{y})\}^b \, d\mu_0 \, d\sigma \, d(\underline{\phi}_p \times \underline{\theta}_q). \tag{10}$$

When computing the integrals in (9) and (10), the integrations for $\underline{\mu}_k$, μ_0, and σ are easily solved by using the kernels of the multivariate normal density for $\underline{\mu}_k$, the normal density for μ_0, and the inverse gamma density for σ. Specially, we use the following identity to integrate over $\underline{\mu}_k$,

$$\begin{aligned}
&(\mathbf{y} - X_k \underline{\mu}_k)' V_{p,q}^{-1} (\mathbf{y} - X_k \underline{\mu}_k) \\
&= \{(\mathbf{y} - X_k \hat{\underline{\mu}}_k) + X_k(\hat{\underline{\mu}}_k - \underline{\mu}_k)\}' V_{p,q}^{-1} \{(\mathbf{y} - X_k \hat{\underline{\mu}}_k) + X_k(\hat{\underline{\mu}}_k - \underline{\mu}_k)\} \\
&= S_{k,p,q} + (\underline{\mu}_k - \hat{\underline{\mu}}_k)'(X_k' V_{p,q}^{-1} X_k)(\underline{\mu}_k - \hat{\underline{\mu}}_k),
\end{aligned}$$

where

$$\begin{aligned}
S_{k,p,q} &= (\mathbf{y} - X_k \hat{\underline{\mu}}_k)' V_{p,q}^{-1} (\mathbf{y} - X_k \hat{\underline{\mu}}_k), \\
\hat{\underline{\mu}}_k &= (X_k' V_{p,q}^{-1} X_k)^{-1} X_k' V_{p,q}^{-1} \mathbf{y}.
\end{aligned}$$

But the region $C_p \times C_q$ with higher order of p and q than 2 is not explicit, and the integration over $(\underline{\phi}_p, \underline{\theta}_q)$ is very complicated. To circumvent the difficulty in identifying $C_p \times C_q$ with high order p and q, there is an useful reparameterization. Following Barndorff-Nielsen and Schou (1973), Monahan (1984), and Jones (1987), there is one to one transformation between $(\underline{\phi}_p, \underline{\theta}_q)$ and partial autocorrelations $(\underline{\gamma}_p, \underline{\gamma}_q)$, where $\underline{\gamma}_p = (\gamma_{p1}, \gamma_{p2}, \ldots, \gamma_{pp})$ and $\underline{\gamma}_q = (\gamma_{q1}, \gamma_{q2}, \ldots, \gamma_{qq})$, that maps $C_p \times C_q$ onto $(-1, 1)^{p+q}$. Let $\underline{z}^{(k)} = (z_1^{(k)}, z_2^{(k)}, \ldots, z_k^{(k)})$, $k = 1, 2, \ldots, p$. Then $z_i^{(k)}$ is calculated from the recursive relation, $z_i^{(k)} = z_i^{(k-1)} - r_k z_{k-i}^{(k-1)}$, $i = 1, 2, \ldots, k-1$, with $z_1^{(1)} = r_1$ as the initial setting and $z_k^{(k)} = r_k$ as the final setting. Finally, set $\underline{\phi}_p = \underline{z}^{(p)}$. For example of $p = 3$, $\phi_1 = r_1 - r_1 r_2 - r_2 r_3$, $\phi_2 = r_2 - r_1 r_3 + r_1 r_2 r_3$, and $\phi_3 = r_3$.

After integrating over $\underline{\mu}_k$ and σ in (9), we can let

$$S_{k,p,q} = |X_k' V_{p,q}^{-1} X_k|^{-1} \cdot |(X_k, \mathbf{y})' V_{p,q}^{-1} (X_k, \mathbf{y})|$$

using the fact given in Shilov (1961) that for some matricies, A and B,

$|A'A|^{\frac{1}{2}} \cdot \| (I-P_A)B \| = |(A,B)'(A,B)|^{\frac{1}{2}}$ with $P_A = A(A'A)^{-1}A'$ and $\| \underline{x} \| = (\underline{x}'\underline{x})^{1/2}$ for some column vector $\underline{x}$. Finally, after transforming from $(\underline{\phi}_p, \underline{\theta}_q)$ to $(\underline{\gamma}_p, \underline{\gamma}_q)$, the final form of (9) is obtained by

$$m^N_{(k,\ d_k, p, q)}(\ \mathbf{y} \mid b) = \frac{\Gamma\left\{\frac{1}{2}(bn+s-k-2)\right\} \cdot \pi(\ \mathbf{d}_k)}{b^{\frac{1}{2}(bn+s-1)} \cdot 2^{\frac{1}{2}(3-s)} \cdot \pi^{\frac{1}{2}(bn-k-1)}} \cdot g(s,\ \mathbf{d}_k, p, q, X_k,\ \mathbf{y} \mid b), \quad (11)$$

where

$$g(s,\ \mathbf{d}_k, p, q, X_k,\ \mathbf{y} \mid b) = \int_{[-1,1]^{p+q}} \frac{\left| V^*_{p,q}{}^{-1} \right|^{\frac{b}{2}} \left| X_k'\ V^*_{p,q}{}^{-1} X_k \right|^{\frac{1}{2}(bn+s-k-3)}}{\left| (X_k,\ \mathbf{y})'\ V^*_{p,q}{}^{-1} (X_k,\ \mathbf{y}) \right|^{\frac{1}{2}(bn+s-k-2)}} f(\ \underline{\gamma}_p,\ \underline{\gamma}_q) d(\ \underline{\gamma}_p \times\ \underline{\gamma}_q) \quad (12)$$

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

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

with $B_{\gamma_j}(\alpha_1, \alpha_2)$ denoting a rescaled beta probability density of a random variable γ_j defined on (-1,1) with two parameters, α_1 and α_2. Similarly, the final form, $m^N_{(0,p,q)}(\ \mathbf{y} \mid b)$, of (10) is obtained by replacing $k=0$, $X_k = \underline{1}_n$, and omitting the terms on the change point $\mathbf{d}_k$ in equation (11).

There is the fractional Bayes factor (FBF) of O'Hagan (1995) as a Bayes factor which can be used for Bayesian testing in spite of arbitrary constants in improper priors. The FBF is classified as a 'default' or an 'automatic' Bayes factor free from arbitrariness of noninformative improper priors. The default Bayes factors are simpler and more automatic to use since they don't need setting hyperparameters under conjugate priors or considering the imaginary constant as in Spiegelhalter and Smith (1982).

The FBF for testing the switching model M_1 with a fixed change point $\mathbf{d}_1$ against the no-switching model M_0 is defined as follows:

$$B^{FBF}_{10}(\ \mathbf{d}_1) = B^N_{10}(\ \mathbf{d}_1,\ \mathbf{y} \mid b=1) \cdot B^N_{01}(\ \mathbf{d}_1,\ \mathbf{y} \mid b),$$

where b=(the size of a minimal training sample)/ n is the common use of b in O'Hagan (1995) when robustness is not major concern. The minimal training sample implies the part of full sample with the minimal sample size to guarantee the finiteness of both $m^N_{(1,\ d_1, p, q)}(\ \mathbf{y} \mid b=1)$ and $m^N_{(0,p,q)}(\ \mathbf{y} \mid b=1)$. It is sufficient to check how the minimal training sample size for the model $m^N_{(1,\ d_1, p, q)}(\ \mathbf{y} \mid b=1)$ is, since the model

$m^N_{(1,\ d_1,p,q)}(\ \mathbf{y}\mid b=1)$ is a model including the model $m^N_{(0,p,q)}(\ \mathbf{y}\mid b=1)$. Four observations as a minimal training sample must be continuously sampled each two observations to estimate each μ_0, μ_1, and σ at both sides centering the change point, since all the priors except μ_0, μ_1, and σ have finite supports. For example, a minimal training sample of size 4 with a change point d_1 is $\{y_{d_1-1}, y_{d_1}, y_{d_1+1}, y_{d_1+2}\}$.

Finally, the posterior probability of the change model M_1 is given by

$$P(M_1\mid \mathbf{y})=\sum_{d_1\in D_1}\left\{\frac{p_0}{p_1}\{B^{FBF}_{10}(\ d_1)\}^{-1}+\{\pi(\ d_1\mid \mathbf{y})\}^{-1}\right\}^{-1}, \tag{13}$$

where $p_1(p_0)$ is the prior probability of the model $M_1(M_0)$ being true, and

$$\pi(\ d_1\mid \mathbf{y})=\frac{\sum_{p\in P, q\in Q} m^N_{(1,\ d_1,p,q)}(\ \mathbf{y}|b=1)}{\sum_{d_1\in D_1}\sum_{p\in P, q\in Q} m^N_{(1,\ d_1,p,q)}(\ \mathbf{y}|b=1))} \tag{14}$$

is the posterior probability of the change at each time point.

Theoretically, we can detect whether there is any change or not by the probability of (13), and find where the change occurs by the probability of (14). But the computation of the denominator in (14) for all $d_1\in D_1$ takes a lot of times. Setting $p=0$ and $q=0$ in the computation of (14) much more reduces the computation time and gives a reasonable result in the practical simulation of Section 5. We think that the mean change and the change point can be roughly detected under the assumption of random errors, since the $ARMA(p,q)$ error process is stationary.

For each group of data divided centering the change point with the maximum posterior probability of the change, the Bayesian procedure for detecting the existence of a single change is recursively repeated until any more changes are not detected. If the number of changes and the positions of change points are assumed to be determined as k and $\mathbf{d}_k$, respectively, in order to identify the orders p and q of $ARMA(p,q)$ errors, we use the usual Bayes factor,

$$B^N_{(k,\ \mathbf{d}_k,p,q)(k,\ \mathbf{d}_k,p',q')}(\ \mathbf{y}\mid b=1)\ =\ \frac{m^N{}_{(k,\ \mathbf{d}_k,p,q)}(\ \mathbf{y}\mid b=1)}{m^N{}_{(k,\ \mathbf{d}_k,p',q')}(\ \mathbf{y}\mid b=1)}, \tag{15}$$

for testing the model $M_{k,\ \mathbf{d}_k,p,q}$ with $ARMA(p,q)$ errors against the model $M_{k,\ \mathbf{d}_k,p',q'}$ ($p\neq p'$ or $q\neq q'$) with $ARMA(p',q')$ errors. Finally, the posterior probability of each model

$M_{k,\ d_k,p,q}$ with $ARMA(p,q)$ errors is given by

$$P(M_{k,\ d_k,p,q}\mid \mathbf{y})=\sum_{p'\in P,q'\in Q}\left\{\frac{p_{p',q'}}{p_{p,q}}\{B_{(k,\ d_k,p,q)(k,\ d_k,p',q')}(\mathbf{y}|b=1)\}^{-1}\right\}^{-1}, \qquad (16)$$

where $p_{p,q}(p_{p',q'})$ is the prior probability of the model $M_{k,\ d_k,p,q}(M_{k,\ d_k,p',q'})$ being true.

4. ESTIMATION BY GIBBS SAMPLING

When only the number of changes, k, and the orders of $ARMA(p,q)$ errors are known, we are going to estimate parameters, $(\mathbf{d}_k,\ \underline{\mu}_k,\ \sigma,\ \underline{\phi}_p,\ \underline{\theta}_p)$. After combining (7) and (8), and transforming parameters $(\underline{\phi}_p,\ \underline{\theta}_q)$ into $(\underline{\gamma}_p,\ \underline{\gamma}_q)$, the joint posterior distribution of $(\mathbf{d}_k,\ \underline{\mu}_k,\ \sigma,\ \underline{\gamma}_p,\ \underline{\gamma}_q)$ for fixed k,p, and q is given by

$$\pi(\mathbf{d}_k,\ \underline{\mu}_k,\ \sigma,\ \underline{\gamma}_p,\ \underline{\gamma}_q|\ \mathbf{y})\propto\sigma^{-(n+s)}|\ V^*_{p,q}{}^{-1}|^{\frac{1}{2}}\cdot\exp\left\{\frac{1}{2\sigma^2}(\mathbf{y}-X_k\underline{\mu}_k)'\ V^*_{p,q}{}^{-1}(\mathbf{y}-X_k\underline{\mu}_k)\right\}. (17)$$

The full conditional posterior densities for Gibbs sampling from equation (17) are as follows:

$$[\sigma^2|\ \mathbf{d}_k,\ \underline{\mu}_k,\ \underline{\gamma}_p,\ \underline{\gamma}_q]\sim IG\left\{\frac{1}{2}(n+s-1),2/Q(\mathbf{d}_k,\ \underline{\mu}_k,\ \underline{\gamma}_p,\ \underline{\gamma}_q)\right\},$$

where $IG\{\alpha,\beta\}$ denotes the inverse gamma distribution with parameters (α,β) which of density is given by $\pi(\sigma^2|\alpha,\beta)=\{\beta^\alpha\Gamma(\alpha)\}^{-1}(\sigma^2)^{-(\alpha+1)}e^{-1/(\beta\sigma^2)}$, and

$$Q(\mathbf{d}_k,\ \underline{\mu}_k,\ \underline{\gamma}_p,\ \underline{\gamma}_q)=(\mathbf{y}-X_k\underline{\mu}_k)'\ V^*_{p,q}{}^{-1}(\mathbf{y}-X_k\underline{\mu}_k).$$

$$[\underline{\mu}_k\mid\mathbf{d}_k,\sigma,\ \underline{\gamma}_p,\ \underline{\gamma}_q]\sim N_{k+1}\{\hat{\underline{\mu}}_k,\sigma^2(X'_k\ V^*_{p,q}{}^{-1}X_k)^{-1}\},$$

where

$$\hat{\underline{\mu}}_k=(X'_k\ V^*_{p,q}{}^{-1}X_k)^{-1}X'_k\ V^*_{p,q}{}^{-1}\ \mathbf{y}.$$

$$[\mathbf{d}_k|\ \underline{\mu}_k,\sigma,\ \underline{\gamma}_p,\ \underline{\gamma}_q]\sim g(\mathbf{d}_k)=\exp\{-Q(\mathbf{d}_k,\ \underline{\mu}_k,\ \underline{\gamma}_p,\ \underline{\gamma}_q)/(2\sigma^2)\}.$$

Finally,

$$[\underline{\gamma}_p,\ \underline{\gamma}_q|\ \mathbf{d}_k,\ \underline{\mu}_k,\sigma]\sim h(\underline{\gamma}_p,\ \underline{\gamma}_q)=|\ V^*_{p,q}{}^{-1}|^{\frac{1}{2}}\cdot\exp\left\{-\frac{1}{2\sigma^2}Q(\mathbf{d}_k,\ \underline{\mu}_k,\ \underline{\gamma}_p,\ \underline{\gamma}_q)\right\}\cdot f(\underline{\gamma}_p,\ \underline{\gamma}_q).$$

Since the conditional posterior densities of $\mathbf{d}_k$ and $(\underline{\gamma}_p,\ \underline{\gamma}_q)$ are not the standard form, we have to run the Metropolis-Hastings (MH) algorithm of Hasting (1970).

The MH algorithm for generating $\mathbf{d}_k$ is performed as follows:

STEP 0 : Set the initial value $\boldsymbol{d}_k^{(0)}$ as the value of $\boldsymbol{d}_k$ in the previous iteration of the Gibbs sampler and $j=0$.

STEP 1 : Generate $\boldsymbol{d}_k^* = (d_1, d_2, \ldots, d_k)$ from the discrete uniform distribution with a support $\boldsymbol{D}_k$.

STEP 2 : Compute $c = \min\{1, g(\boldsymbol{d}_k^*)/g(\boldsymbol{d}_k^{(j)})\}$.

STEP 3 : Generate U from Uniform(0,1) density.

STEP 4 : Set $\boldsymbol{d}_k^{(j+1)} = \begin{cases} \boldsymbol{d}_k^*, & \text{if } U \le c, \\ \boldsymbol{d}_k^{(j)}, & \text{if } U > c. \end{cases}$

STEP 5 : Set $j=j+1$, and go to **STEP 1**.

The MH algorithm for generating $(\underline{\gamma}_p, \underline{\gamma}_q)$ is performed as follows:

STEP 0 : Set the initial values, $\underline{\gamma}_p^{(0)} = (\gamma_1, \gamma_2, \ldots, \gamma_p)$ and $\underline{\gamma}_q^{(0)} = (\gamma_1', \gamma_2', \ldots, \gamma_q')$, as the values of $\underline{\gamma}_p$ and $\underline{\gamma}_q$ in the previous iteration step of the Gibbs sampler and $j=0$.

STEP 1: Generate $\gamma_i (i=1,2,\ldots,p)$ and $\gamma'_i (i=1,2,\ldots,q)$ independently from the uniform distribution with a space (-1,1). Then, set $\underline{\gamma}_p^* = (\gamma_1, \gamma_2, \ldots, \gamma_p)$ and $\underline{\gamma}_q^* = (\gamma_1', \gamma_2', \ldots, \gamma_q')$.

STEP 2 : Compute $d = \min\{1, h(\underline{\gamma}_p^*, \underline{\gamma}_q^*)/h(\underline{\gamma}_p^{(j)}, \underline{\gamma}_q^{(j)})\}$.

STEP 3 : Generate U from Uniform(0,1) density.

STEP 4 : Set $(\underline{\gamma}_p^{(j+1)}, \underline{\gamma}_q^{(j+1)}) = \begin{cases} (\underline{\gamma}_p^*, \underline{\gamma}_q^*), & \text{if } U \le d, \\ (\underline{\gamma}_p^{(j)}, \underline{\gamma}_q^{(j)}), & \text{if } U > d. \end{cases}$

STEP 5 : Set $j=j+1$, and go to **STEP 1**.

At each iteration of Gibbs sampler, $(\underline{\gamma}_p, \underline{\gamma}_q)$ is retransformed to $(\underline{\phi}_p, \underline{\theta}_q)$.

5. Simulation Study

We carry out a simulation study to check the Bayesian inference procedure for multiple switching mean models with ARMA errors discussed in the previous sections. All the computations are completed using the MATLAB (The MATH WORKS Inc., 1999).

Three time series data sets with each two change points are generated from the following models where $\sigma^2 = 1$, and shown in Figure 1.

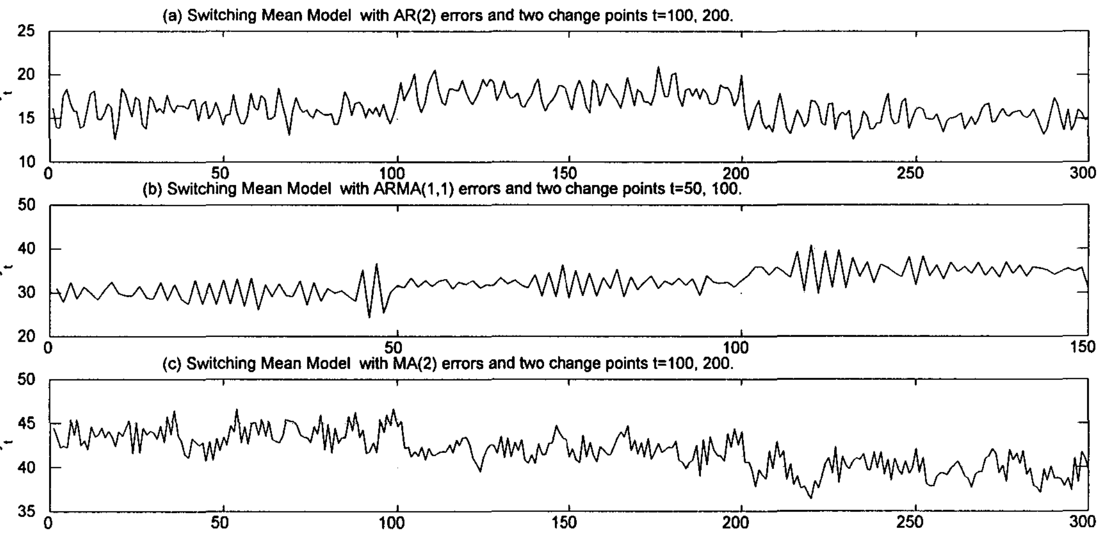

Figure 1. Plots of three simulated time series.

(i) A switching mean model with a sample size $n=300$, two change points, $\boldsymbol{d}_2=(100,200)$, $\boldsymbol{\mu}_2=(16,18,15)$, and the $AR(2)$ error with $\phi_1=0.3$ and $\phi_2=-0.5$.

(ii) A switching mean model with a sample size $n=150$, two change points, $\boldsymbol{d}_2=(50,100)$, $\boldsymbol{\mu}_2=(30,32,35)$, and the $ARMA(1,1)$ error with $\phi_1=-0.7$ and $\theta_1=0.6$.

(iii) A switching mean model with a sample size $n=300$,two change points, $\boldsymbol{d}_2=(100,200)$, $\boldsymbol{\mu}_2=(44,42,40)$, and the $MA(2)$ error with $\theta_1=-0.2$ and $\theta_2=-0.8$.

We set $s=1$ as the reference prior in the priors of (5) and (6). Also, we assume equal prior probabilities for each model, that is, $p_0=p_1$ in (13) and $p_{p',q'}=p_{p,q}$ in (16).

Table 1 shows the posterior probabilities of switching mean models computed using the FBF. Concerning the switching mean model with $AR(2)$ errors, at step 1 the posterior probability of the switching mean model for the data with a total of 300 observations is one and the change point(cp) is 200 with the maximum posterior probability(mpp), 0.4540.

At the next step the posterior probability of the switching mean model for the first 200 observations is also one and the change point(cp) is 99 with the maximum posterior probability(mpp), 0.4450. At the last step, each posterior probability of the switching mean models is 0.1408, 0.1388, and 0.1251, respectively, for three data groups with observation numbers, 1~99, 100~200, and 201~300, which implies that there is not any more change in each data group. Thus, two change points and their temporary positions are assumed as $(99, 200)$. Similarly, the positions of two change points for the rest of model are roughly put as $(49, 100)$ and $(101, 200)$.

Now, the posterior probabilities of $ARMA(p, q)$ errors computed using the usual Bayes factor (15) for each switching mean model with two temporary change points are shown in Table 2. But the computation of integral in (12) must be before solved. We estimate it by the Monte Carlo method through 200 importance sampling with a joint density of $p+q$ independent uniform variates distributed over $(-1, 1)$ as an importance density.

Table 3, 4, and 5 present the results of posterior distribution for parameters included in each model. When operating the Gibbs sampler, the initial values of $\boldsymbol{d}_k$, $\boldsymbol{\mu}_k$, $\boldsymbol{\gamma}_p$, and $\boldsymbol{\gamma}_q$ are required. Two temporary change points shown in Table 1 are used as initial values of $\boldsymbol{d}_k$, and the sample means of data groups divided centering each temporary change point are used as initial elements of $\boldsymbol{\mu}_k$. Initial elements of $\boldsymbol{\gamma}_p$ and $\boldsymbol{\gamma}_q$ are randomly generated from the uniform distribution over $(-1, 1)$. At step 1 of the MH algorithm for generating $\boldsymbol{d}_k$, each element of the support $\boldsymbol{D}_k$ of the discrete uniform distribution used as a transition probability distribution is put as (cp ± 10), where cp is obtained in Table 1.

In our simulation study, we estimate parameters from one sequence simulated for only one Gibbs sampler, and burn the first 30% after totally 130% iterations. The iterations of Gibbs sampler and the Metropolis-Hastings subchains are 100 and 50, respectively.

6. Concluding Remarks

We do not present all the simulation results due to space limit. However, we see that our methodologies presented in this article yield reasonable results in accordance with theoretical outcomes. In particular, they work out well for the data sets with larger sample sizes, larger differences in means, and more strictly stationary conditions. We also point out that larger sample sizes should be required as the first autocorrelation gets positively higher. When the positive first order autocorrelation is employed, the resulting data stay above or below means as time goes by. Meanwhile, the negative first order autocorrelation is employed, the resulting data fluctuate quite frequently between the mean. Hence, more data in case of the model with positive first order autocorrelation are required to capture overall pattern of time series than the case of negative first order autocorrelation. For example, in our simulation study, the first autocorrelations of the $AR(2)$, the $ARMA(1,1)$, and the $MA(2)$ error model are $0.2, -0.839$, and 0.214, respectively. Also, different $ARMA(p, q)$ errors can have similar values of the likelihood function, and the order (p, q) different from the true value of (p, q)

is can be selected. But, this is the problem similarly applied to the selection by the AIC (Akaike Information Criterion).

Table 1. The posterior probabilities of switching mean models computed using the FBF.

	Switching mean model with AR(2) error		Switching mean model with ARMA(1,1) error		Switching mean model with MA(2) error	
	Obs. no. cp(mpp)	Posterior probability	Obs. no. cp(mpp)	Posterior probability	Obs. no. cp(mpp)	Posterior probability
STEP 1	1~300 200(0.4540)	1	1~150 100(0.3090)	1	1~300 200(0.6633)	1
STEP 2	1~200 99(0.4450)	1	1~100 49(0.1879)	0.9969	1~200 101(0.5157)	1
STEP 3	1~99	0.1408	1~49	0.1564	1~101	0.4344
	100~200	0.1388	50~100	0.1772	102~200	0.2177
	201~300	0.1251	101~150	0.1662	201~300	0.1881

Table 2. The posterior probabilities of $ARMA(p,q)$ errors in switching mean models.

$ARMA(p,q)$	$AR(2)$ error	$ARMA(1,1)$ error	$MA(2)$ error
(0, 0)	0.0000	0.0000	0.0000
(0, 1)	0.0000	0.0000	0.0000
(0, 2)	0.0000	0.0000	1.0000
(1, 0)	0.0000	0.0002	0.0000
(1, 1)	0.0000	0.5035	0.0000
(1, 2)	0.0000	0.0010	0.0000
(2, 0)	0.7258	0.0008	0.0000
(2, 1)	0.2730	0.3004	0.0000
(2, 2)	0.0012	0.1942	0.0000

Table 3. A switching mean model with $AR(2)$ error.

True Parameter	Posterior Distribution				
	Mean	Std.	Median	Lower 95% limit	Upper 95% limit
$d_1 = 100$	100.0100	0.2245	100.0000	100.0000	100.0000
$d_2 = 200$	200.0000	0.0000	200.0000	200.0000	200.0000
$\mu_0 = 16$	15.9657	0.0843	15.9628	15.8241	16.1372
$\mu_1 = 18$	17.9057	0.0867	17.9042	17.7633	18.0442
$\mu_2 = 15$	15.1829	0.1011	15.1880	15.0262	15.3745
$\sigma^2 = 1$	0.9442	0.0585	0.9316	0.8646	1.0397
$\phi_1 = 0.3$	0.4136	0.0401	0.4172	0.3559	0.4589
$\phi_2 = -0.5$	-0.5023	0.0343	-0.5120	-0.5477	-0.4444

Table 4. A switching mean model with $ARMA(1,1)$ error.

True Parameter	Posterior Distribution				
	Mean	Std.	Median	Lower 95% limit	Upper 95% limit
$d_1 = 50$	50.0000	0	50	50	50
$d_2 = 100$	100.0000	0	100	100	100
$\mu_0 = 30$	29.9534	0.0373	29.9573	29.8863	30.0177
$\mu_1 = 32$	32.0613	0.0398	32.0640	31.9937	32.1254
$\mu_2 = 35$	35.0571	0.0417	35.0604	34.9917	35.1203
$\sigma^2 = 1$	1.1412	0.1380	1.1352	0.9161	1.3481
$\phi_1 = -0.7$	-0.7220	0.0380	-0.7301	-0.7796	-0.6505
$\theta_1 = 0.6$	0.5531	0.0413	0.5567	0.4887	0.6316

Table 5. A switching mean model with $MA(2)$ error.

True Parameter	Posterior Distribution				
	Mean	Std.	Median	Lower 95% limit	Upper 95% limit
$d_1 = 100$	100.77	0.7502	101.0000	100.0000	103.0000
$d_2 = 200$	200.00	0.0000	200.0000	200.0000	200.0000
$\mu_0 = 44$	43.6719	0.2073	43.6336	43.3854	44.0041
$\mu_1 = 42$	42.0689	0.1992	42.0803	41.6768	42.3593
$\mu_2 = 40$	39.7592	0.1969	39.7505	39.4381	40.1147
$\sigma^2 = 1$	0.9649	0.0801	0.9479	0.8456	1.0958
$\theta_1 = -0.2$	-0.1938	0.0292	-0.1886	-0.2280	-0.1592
$\theta_2 = -0.8$	-0.8888	0.0262	-0.9050	-0.9096	-0.8529

REFRERENCES

[1] Albert, J. H. and Chib, S. (1993), Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts, *Journal of Business and Economic Statistics*, Vol. 11. No.1. pp1-15.

[2] Barndorff-Nielsen, O. and Schou, G. (1973), On the Parametrization of Autoregressive Models by Parial Autocorrelations, *Journal of Multivariate Analysis* , Vol. 3,408-419.

[3] Berger, J. O. and Pericchi, L. R. (1996), The Intrinsic Bayes Factor for Model Selection and Prediction, *Journal of the American Statistical Association*, Vol.91, No.433, 109-122.

[4] Csörgő, M. and Horváth, L.(1997), *Limit Theorems in Change-Point Analysis*, John Wieley and Sons..

[5] Garisch, I. and Groenewald, P. C. N. (1999), The Nile revisted: Changepoint Analysis with Autocorrelation, *Bayesian Statistics, 6*, 753-760 .

[6] Hastings, W. K. (1970), Monte Carlo Sampling Methods using Markov Chains and their Applications, *Biometrika*, Vol.57, 97-109.

[7] Jones, M.C. (1987), Randomly Choosing Parameters from the Stationarity and Invertibility Region of Autoregressive-Moving Average Models, *Applied Statistics,* Vol. 36, No.2, 134-138.

[8] Leeuw, J. V. D. (1994), The Covariance Matrix of ARMA Errors in Closed Form, *Journal of Econometrics*, Vol.63, 397-405.

[9] Marriott, J. N., Ravishanker, A., Gelfand, A., and Pai, J. (1992), Baysian Analysis of ARMA Processes : Complete sampling based inference under full likelihoods, *Manuscript.*

[10] Monahan, J. F. (1984), A Note on Enforcing Stationarity in Autoregressive Moving Average Models, *Biometrika*, Vol.71, 403-404.

[11] O'Hagan, A. (1995), Fractional Bayes Factors for Model Comparison, *Journal of the Royal Statistical Society*, **B,** Vol.57, No.1, 99-138.

[12] Ohtani, K. (1982), Bayesian Estimation of the Switching Regression Model with Autocorrelated Errors, *Journal of Econometrics,* Vol. 18, pp251-261.

[13] Shilov, G. E. (1961), *An Introduction to the Theory of Linear Spaces.* Englewood Cliffs, NJ: Prentice-Hall.

[14] Son, Y. S. (1999), ARMA Model Indentification Using the Bayes Factor. *Journal of the Korean Statistical Society*, Vol.28, No.4, 503-513.

[15] Son, Y. S. (2001), Default Bayesian Inference of Regression Models with ARMA Errors Under Exact Full Likelihoods, *2001 Proceedings of the Spring Conference*, 163-168, Korean Statistical Society.

[16] Spiegelhalter, D. J. and Smith, A. F. M. (1982), Bayes Factors for Linear and Log-Linear Models with Vague Prior Information, *Journal of the Royal Statistical Society*, B, Vol. 44, 377-387.

[17] The MATH WORKS Inc. (1998), *MATLAB/Statistics Toolbox,* Version 5.2, Natick, MA 01760.

[18] Varshavsky, J. A. (1996), Intrinsic Bayes Factors for Model Selection with Autoregressive data, *Bayesian Statistics,* 5, J. M. Bernardo, et.al., London: Oxford University Press.

[Received April 2003, Accepted September 2003]