A Bayesian Nonlinearity Test for Threshold Moving Average Models

QIANG XIA^{a*} JIAZHU PAN^b ZHIQIANG ZHANG^c JINSHAN LIU^a

^a *Department of Mathematics, South China Agricultural University, Guangzhou 510642, P.R. China*

^b *Department of Mathematics and Statistics, University of Strathclyde, Glasgow G1 1XH, UK*

^c *Department of Statistics, East China Normal University, Shanghai, P.R. China*

Abstract. We propose a Bayesian test for nonlinearity of threshold moving average (TMA) models. First of all, we obtain the marginal posterior densities of all parameters, including the threshold and delay, of TMA model using Gibbs sampler with Metropolis-Hastings algorithm. And then, we adopt reversible-jump Markov chain Monte Carlo (RJMCMC) methods to calculate the posterior probabilities for MA and TMA models. Posterior evidence in favor of the TMA model indicates threshold nonlinearity. Simulation experiments and a real example show that our method works very well in distinguishing MA and TMA models.

Key words: Bayesian inference; MA models; Gibbs sampler; Metropolis-Hastings algorithm; RJMCMC methods; TMA models.

*Corresponding author. E-mail address: xiaqiang@scau.edu.cn

1 Introduction

Since Tong (1978), threshold autoregressive (TAR) models have become a standard class of nonlinear time series models. There is a huge literature on theoretical property, estimation and test of TAR models, see Tong and Lim (1980), Tong (1990) and Tasy (2005) among others. To avoid complicated analytical works and numerical multiple integration in statistical inference of TAR models, some authors have applied the Bayesian method to simultaneous estimation of all parameters in TAR. Mcculloch and Tsay(1993a, 1993b) proposed a Bayesian procedure for detecting threshold values in the TAR model via posterior probability plots. Chen and Lee(1995) applied the Gibbs sampler of Geman and Geman (1984), and the Metropolis-Hastings (M-H) algorithm of Metropolis et al.(1953) and Hastings (1970), for inference of TAR models.

In recent years, attention has been paid for threshold moving average (TMA) models in the literature, because people realized TMA models are as important as TAR models in practice. For instance, Ismail and Charif (2003) introduced a Bayesian inference for TMA models, and Ling and Tong (2005) proposed a likelihood ratio test for linear MA model against TMA models. But fundamental theory about TMA models, such as identifying the threshold and delay values, estimating the parameters, and testing the threshold nonlinearity, needs to be developed further. The main objective of this paper is to propose a Bayesian method for testing the threshold nonlinearity of TMA models.

Firstly, we investigate Bayesian estimation of the threshold and other parameters of TMA models. Chen and Lee(1995), Ismail and Charif(2003) and Sáfadi and Morettin(2000) adopt MCMC techniques and use the arranged autoregression approach to estimate the threshold parameter and other parameters simultaneously. Basing upon their work, we combine Gibbs sampler and Metropolis-Hastings algorithm to give a Bayesian analysis of two-regime TMA models without employing the arranged autoregression. Secondly, we test the significance of threshold nonlinearity by comparing a MA and its threshold MA counterpart. We transfer the testing problem to a Bayesian model-selection problem. In order to do the model compari-

son, we choose the reversible-jump MCMC method of Green (1995) to compute the posterior probabilities for MA and TMA models. We then select the model with a higher posterior probability to determine whether the threshold nonlinearity is significant. It is demonstrated that the reversible-jump method is easy to implement and fits quite well within our framework of Bayesian modeling.

This paper is arranged as follows. Section 2 presents the TMA models and the methodology of Bayesian inference. Section 3 gives details of a Bayesian model selection procedure by RJMCMC method. Some simulation results and a real example are presented in Section 4. Section 5 is our conclusion.

Throughout the paper, we denote the transpose of a matrix A by A' .

2 Threshold MA Model and Bayesian Inference

2.1 TMA Model

A time series $\{y_t, t = 1, 2, \dots\}$ is said to follow a $TMA(2, q_1, q_2)$ model with two regimes if it satisfies the following equation

$$y_t = (\theta_0^{(1)} - \sum_{i=1}^{q_1} \theta_i^{(1)} \varepsilon_{t-i}) I(y_{t-d} > r) + (\theta_0^{(2)} - \sum_{i=1}^{q_2} \theta_i^{(2)} \varepsilon_{t-i}) I(y_{t-d} \leq r) + \varepsilon_t \quad (2.1.1)$$

where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with common distribution $N(0, \sigma^2)$, q_1, q_2, d are positive integers, and d is called the delay (or threshold lag) parameter of the model. I is the indicator function and $r \in R$ is called the threshold parameter.

2.2 Bayesian Inference

Suppose we have a sample $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$. Let $q = \max\{q_1, q_2\}$. Then the threshold variable y_{t-d} assumes values $\{y_{h-d}, \dots, y_{n-d}\}$, where $h = \max\{d+1, q+1\}$. Given the first $h-1$ observations and $\varepsilon_1 = \dots = \varepsilon_{h-1} = 0$, we can easily obtain the conditional likelihood

function of $TMA(2, q_1, q_2)$ model as follows

$$\begin{aligned}
L(\Theta_1, \Theta_2, \sigma^2, r, d \mid \mathbf{y}) &\propto \sigma^{-n+h-1} \\
&\times \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{t=h}^n \left(y_t - \theta_0^{(1)} + \theta_1^{(1)} \varepsilon_{t-1} + \dots + \theta_{q_1}^{(1)} \varepsilon_{t-q_1} \right)^2 I(y_{t-d} > r) \right. \right. \\
&\quad \left. \left. + \sum_{t=h}^n \left(y_t - \theta_0^{(2)} + \theta_1^{(2)} \varepsilon_{t-1} + \dots + \theta_{q_2}^{(2)} \varepsilon_{t-q_2} \right)^2 I(y_{t-d} \leq r) \right] \right\}
\end{aligned} \tag{2.2.1}$$

Denote

$$\begin{aligned}
Y &= (y_h, y_{h+1}, \dots, y_n)', \quad \Upsilon = (\varepsilon_h, \varepsilon_{h+1}, \dots, \varepsilon_n)', \\
\Theta_1 &= (\theta_0^{(1)}, \theta_1^{(1)}, \dots, \theta_{q_1}^{(1)})', \quad \Theta_2 = (\theta_0^{(2)}, \theta_1^{(2)}, \dots, \theta_{q_2}^{(2)})', \\
I_1 &= \text{diag}(I(y_{h-d} > r), I(y_{h+1-d} > r), \dots, I(y_{n-d} > r)), \\
I_2 &= \text{diag}(I(y_{h-d} \leq r), I(y_{h+1-d} \leq r), \dots, I(y_{n-d} \leq r)).
\end{aligned}$$

and

$$X_1 = \begin{bmatrix} 1 & -\varepsilon_{h-1} & \dots & -\varepsilon_{h-q_1} \\ 1 & -\varepsilon_h & \dots & -\varepsilon_{h+1-q_1} \\ \dots & \dots & \dots & \dots \\ 1 & -\varepsilon_{n-1} & \dots & -\varepsilon_{n-q_1} \end{bmatrix}, \quad X_2 = \begin{bmatrix} 1 & -\varepsilon_{h-1} & \dots & -\varepsilon_{h-q_2} \\ 1 & -\varepsilon_h & \dots & -\varepsilon_{h+1-q_2} \\ \dots & \dots & \dots & \dots \\ 1 & -\varepsilon_{n-1} & \dots & -\varepsilon_{n-q_2} \end{bmatrix}.$$

Then

$$\Upsilon = I_1(Y - X_1\Theta_1) + I_2(Y - X_2\Theta_2),$$

and (2.2.1) can be rewritten as

$$L(\Theta_1, \Theta_2, \sigma^2, r, d \mid \mathbf{y}) \propto |\Sigma|^{-\frac{1}{2}} \cdot \exp \left(-\frac{1}{2} \Upsilon' \Sigma^{-1} \Upsilon \right) \tag{2.2.2}$$

where $\Sigma = \text{diag}(\sigma^2, \dots, \sigma^2)$ and $\dim(\Sigma) = n - h + 1$. Assume $\{\varepsilon_t = 0, t < h\}$. Then $\{\varepsilon_t, t \geq h\}$ can be computed recursively: $\varepsilon_t = y_t - x_{1t}\Theta_1$ if $y_{t-d} > r$; $\varepsilon_t = y_t - x_{2t}\Theta_2$ if not, where $x_{1t} = (1, -\varepsilon_{t-1}, \dots, -\varepsilon_{t-q_1})$, $x_{2t} = (1, -\varepsilon_{t-1}, \dots, -\varepsilon_{t-q_2})$.

To implement the Bayesian inference about the parameters $\Theta_1, \Theta_2, \sigma^2, r, d$ in the $TMA(2, q_1, q_2)$ model, we need the joint posterior distribution $P(\Theta_1, \Theta_2, \sigma^2, r, d)$, which can be obtained by using conditional posterior distributions in a MCMC process. Therefore, we need to choose priors to derive the conditional posterior distribution for the unknown parameters.

Referring to Chen and Lee (1995), and Perreault(2000), we adopt that Θ_i follows $N(\Theta_{0i}, V_i^{-1})$, where V_i denotes the precision, $i = 1, 2$, and Θ_1 and Θ_2 are independent. Let σ^2 follow the inverse gamma distribution $IG(\alpha, \beta)$, where the hyper-parameters are known. Similar to Geweke and Terui(1993), we assume that r follows a uniform distribution on an interval (a, b) , and d follows a discrete uniform distribution on a set of integers $\{1, 2, \dots, d_0\}$, where d_0 is a prescribed positive integer.

Using Bayesian techniques, we derive the conditional posterior distributions of $\Theta_1, \Theta_2, \sigma^2, r, d$ based on the above priors as follows.

- (1) The conditional posterior probability function of Θ_i is

$$p(\Theta_i | \mathbf{y}, \Theta_j, \sigma^2, r, d) \propto \exp \left\{ -\frac{1}{2} \left[\Upsilon' \Sigma^{-1} \Upsilon + (\Theta_i - \Theta_{0i})' V_i (\Theta_i - \Theta_{0i}) \right] \right\} \quad (2.2.3)$$

for $i \neq j$ and $i = 1, 2$, where $\Upsilon = I_1(Y - X_1\Theta_1) + I_2(Y - X_2\Theta_2)$.

- (2) The conditional posterior distribution of σ^2 is

$$p(\sigma^2 | \mathbf{y}, \Theta_1, \Theta_2, r, d) \sim IG\left(\alpha + \frac{n - h + 1}{2}, \beta + \frac{s^2}{2}\right), \quad (2.2.4)$$

where $s^2 = \Upsilon' \Upsilon$.

- (3) The conditional posterior probability function of r is

$$p(r | \mathbf{y}, \Theta_1, \Theta_2, \sigma^2, d) \propto \exp \left(-\frac{1}{2\sigma^2} \cdot s^2 \right) \cdot I(a < r < b). \quad (2.2.5)$$

Note that s^2 is a function of r .

- (4) The conditional posterior probability function of d is a multinomial distribution with probability

$$p(d | \mathbf{y}, \Theta_1, \Theta_2, \sigma^2, r) = \frac{L(\Theta_1, \Theta_2, \sigma^2, r, d | \mathbf{y})}{\sum_{d=1}^{d_0} L(\Theta_1, \Theta_2, \sigma^2, r, d | \mathbf{y})} \quad (2.2.6)$$

where $d = 1, 2, \dots, d_0$.

2.3 Sampling Scheme

From the previous section, we can identify the conditional densities for σ^2 and d . Then the Gibbs sampler can be used. But we don't have closed forms of the conditional distributions for r and Θ_i ($i = 1, 2$). We will apply the random walk Metropolis-Hastings (M-H) algorithm to draw Θ_i , $i = 1, 2$, and r . Details of the Gibbs sampler and Metropolis-Hastings algorithm can be found in Casella and George (1992) and Chib and Greenberg (1995), respectively. Denote the target density in (2.2.3) by $f_1(\cdot)$. The algorithm of drawing Θ_i is described below.

Step 1: At iteration j , generate a point Θ_i^* from the random walk kernel

$$\Theta_i^* = \Theta_i^{[j-1]} + \varepsilon_{\Theta_i}, \quad \varepsilon_{\Theta_i} \sim N(0, \Sigma_{\Theta_i})$$

where $\Theta_i^{[j-1]}$ is the $(j-1)$ th iterate for Θ_i .

Step 2: Accept Θ_i^* as $\Theta_i^{[j]}$ with probability $p = \min\{1, f_1(\Theta_i^*)/f_1(\Theta_i^{[j-1]})\}$.

Otherwise, set $\Theta_i^{[j]} = \Theta_i^{[j-1]}$.

Σ_{Θ_i} is usually selected to be a diagonal matrix. The elements of Σ_{Θ_i} are turned by monitoring the acceptance rate between 0.25 and 0.50. Denote the target density in (2.2.5) by $f_2(\cdot)$, the algorithm of drawing r is described as follows.

- At iteration j , generate a point r^* from the random walk kernel

$$r_i^* = r^{[j-1]} + \varepsilon_{r^*}, \quad \varepsilon_{r^*} \sim N(0, \sigma_r^2)$$

where $r^{[j-1]}$ is the $(j-1)$ th iterate of r .

- Accept r^* as $r^{[j]}$ with probability $p = \min\{1, f_2(r^*)/f_2(r^{[j-1]})\}$. Otherwise, set $r^{[j]} = r^{[j-1]}$.

Remark 1. The Gibbs sampler combining with random walk M-H algorithm is carried out for first M iterates, and the sample mean $\mu_{\Theta_i}, \mu_r, \mu_{\sigma^2}, \mu_d$ and sample covariance matrix (or variance) $\Sigma_{\Theta_i}, \sigma_r^2, \sigma_{\sigma^2}^2, \sigma_d^2$ can be obtained. In the next stage, we will substitute μ_{σ^2} and μ_d for σ^2

and d in the model, and $\mu_{\Theta_i}, \Sigma_{\Theta_i}, \mu_r, \sigma_r^2$ are used to form the Gaussian kernels $N(\mu_{\Theta_i}, \Sigma_{\Theta_i})$ and $N(\mu_r, \sigma_r^2)$, which are used in RJMCMC scheme to test the significance of threshold nonlinearity by comparing a MA and its TMA counterpart.

3 Selecting Model by RJMCMC

The main goal of our study is to test for the threshold nonlinearity of TMA models. In other words, we want to select a model between MA and TMA models. To do a Bayesian model selection, we need to calculate the posterior probabilities $p(M_j|\mathbf{y})$, where M_1 and M_2 represent MA and TMA models respectively. We can adopt the RJMCMC method of Green (1995), which is a very useful mechanism to allow jumps between spaces of different dimensions while maintaining the detailed balance condition ensuring convergence of the Markov chain. This method was employed to choose between pairs of GARCH models by Vrontos, Dellaportas and Politis (2000), So, Chen and Chen (2005) and Chen, Gerlach and So (2008).

We consider the jump from M_1 with parameter $\Theta^{(1)}$ to M_2 with parameter $\Theta^{(2)}$. Here $\Theta^{(1)}$ consists of a set of MA parameters, say Θ , and $\Theta^{(2)}$ consists of Θ_1, Θ_2 and r . In general, $\Theta^{(1)}$ and $\Theta^{(2)}$ have different dimensions. To jump from M_1 to M_2 , we construct two variables $u^{(1)}$ and $u^{(2)}$ to form a bijection between $(\Theta^{(1)}, u^{(1)})$ and $(\Theta^{(2)}, u^{(2)})$, that is, $(\Theta^{(2)}, u^{(2)})$ is linked with $(\Theta^{(1)}, u^{(1)})$ by a one-to-one bijective transformation, which ensures the necessary condition that the dimensions of $(\Theta^{(1)}, u^{(1)})$ and $(\Theta^{(2)}, u^{(2)})$ are the same. i.e. $\dim(\Theta^{(1)}) + \dim(u^{(1)}) = \dim(\Theta^{(2)}) + \dim(u^{(2)})$. In other words, $(\Theta^{(2)}, u^{(2)}) = h(\Theta^{(1)}, u^{(1)})$ or $(\Theta^{(1)}, u^{(1)}) = h^{-1}(\Theta^{(2)}, u^{(2)})$ for some function h .

To jump from M_1 to M_2 , we simulate $u^{(1)}$ from a kernel $Q_1(u^{(1)}|\Theta^{(1)})$ and determine $\Theta^{(2)}$ from $h(\Theta^{(1)}, u^{(1)})$. The jump is then accepted with probability $\min\{1, p\}$, where

$$p = \frac{L(\mathbf{y}|M_2, \Theta^{(2)})\Pi(\Theta^{(2)}|M_2)p(M_2)J(M_1, M_2)Q_2(u^{(2)}|\Theta^{(2)})}{L(\mathbf{y}|M_1, \Theta^{(1)})\Pi(\Theta^{(1)}|M_1)p(M_1)J(M_2, M_1)Q_1(u^{(1)}|\Theta^{(1)})} \cdot \left| \frac{\partial(\Theta^{(2)}, u^{(2)})}{\partial(\Theta^{(1)}, u^{(1)})} \right| \quad (3.1)$$

The term $L(\mathbf{y}|M_i, \Theta^{(i)})$ is the likelihood for model M_i , $\Pi(\Theta^{(i)}|M_i)$ is the prior distribution and $p(M_i)$ is the prior probability, $i = 1, 2$. Denote the probability of the jump from M_i to M_j by $J(M_i, M_j)$. The last part in (3.1) is the Jacobian of the transformation. The jump from M_2 to M_1 can be implemented in the reversed way by simulating $u^{(2)}$ from a kernel $Q_2(u^{(2)}|\Theta^{(2)})$ and determining $\Theta^{(1)}$ from $h^{-1}(\Theta^{(2)}, u^{(2)})$ to calculate the acceptance probability $\min\{1, p^{-1}\}$.

To introduce the bijection for RJMCMC, we follow Vrontos et al. (2000), So, Chen and Chen (2005), and define $u^{(1)} = \Theta^{(2)}$, $u^{(2)} = \Theta^{(1)}$, which implies a Jacobian $\left| \frac{\partial(\Theta^{(2)}, u^{(2)})}{\partial(\Theta^{(1)}, u^{(1)})} \right| = 1$. In

addition, we set $J(M_i, M_j) = 1$ to allow a jump in each MCMC iteration, and $p(M_1) = p(M_2) = 0.5$ to reflect prior model ignorance. In this case, the acceptance probability of reversible jump in (3.1) is simplified to

$$p = \frac{L(\mathbf{y}|M_2, \Theta^{(2)})\Pi(\Theta^{(2)}|M_2)Q_2(u^{(2)})}{L(\mathbf{y}|M_1, \Theta^{(1)})\Pi(\Theta^{(1)}|M_1)Q_1(u^{(1)})} \quad (3.2)$$

with the kernels Q_1 and Q_2 being independent of $\Theta^{(1)}$ and $\Theta^{(2)}$, respectively.

It is important to choose appropriate kernels Q_1 and Q_2 to apply the RJMCMC successfully. Sampling Θ_i from $N(\mu_{\Theta_i}, \Sigma_{\Theta_i})$ and r from $N(\mu_r, \sigma_r^2)$ by the M-H steps, we select $Q_1(u^{(1)})$ to be the product of the three normals, i.e. $Q_1(u^{(1)}) \sim N(\mu_{\Theta_1}, \Sigma_{\Theta_1})N(\mu_{\Theta_2}, \Sigma_{\Theta_2})N(\mu_r, \sigma_r^2)$, as the kernel of drawing $\Theta^{(2)} = (\Theta'_1, \Theta'_2, r)'$. For the simulation of $u^{(2)}$, which is the parameter Θ of MA model, we use the same method as described in the previous section to construct $N(\mu_\Theta, \Sigma_\Theta)$ from the first M iterates of Θ . We then choose $Q_2(u^{(2)}) \sim N(\mu_\Theta, \Sigma_\Theta)$ as the kernel of drawing $\Theta^{(1)} = \Theta$. In summary, the jumping scheme is as follows.

- From MA(M_1) to TMA (M_2):
 - (1) Draw $\Theta_i \sim N(\mu_{\Theta_i}, \Sigma_{\Theta_i}), i = 1, 2$ and $r \sim N(\mu_r, \sigma_r^2)$ and accept the jump with probability $\min\{1, p\}$.
 - (2) If accepted, update $\Theta^{(2)}$. Otherwise, update $\Theta^{(1)}$.
- From TMA(M_2) to MA (M_1):
 - (1) Draw $\Theta \sim N(\mu_\Theta, \Sigma_\Theta)$, and accept the jump with probability $\min\{1, p^{-1}\}$.
 - (2) If accepted, update $\Theta^{(1)}$. Otherwise, update $\Theta^{(2)}$.

4 Simulation Experiments and a Real Example

In this section, we first present simulation results to show the effectiveness of our MCMC sampling scheme and model selection method, and then apply our method to a real data set of the exchange rate of Japanese Yen v.s. USA dollar.

4.1 Simulation Experiments

We set $M = 10000$ in all experiments, and apply the sampler scheme to draw all parameters and to form the means and the normal kernels by discarding the first 5000 iterations. We then perform 10000 iterations for posterior inference and model selection. The hyper-parameters are $\Theta_{0i} = 0, V_i = 0.1$ for $i = 1, 2$, and furthermore we choose

$$\alpha = 2.5, \quad \beta = 1.6, \quad d_0 = 3, \quad a = p_5, \quad b = p_{95},$$

where p_k denotes the k th percentile of the data.

Consider the following two models.

- Model 1: $MA(1)$

$$y_t = \varepsilon_t - 0.5\varepsilon_{t-1} \quad (4.1)$$

- Model 2: $TMA(2, 1, 1)$

$$y_t = \varepsilon_t + 0.5\varepsilon_{t-1}I(y_{t-1} > 0.25) - 0.5\varepsilon_{t-1}I(y_{t-1} \leq 0.25) \quad (4.2)$$

We carry out 20000 MCMC iterations for two sets of simulated data of sample size $n = 300$. Table I lists the posterior means, posterior standard deviations in columns 3 and 4 for model 1 and columns 6 and 7 for model 2. It is clear that the posterior means are closed to the true values, although $\theta_1^{(1)}$ has slightly positive bias. For the data simulated from Model 1, the estimated $\Theta^{(2)}$ parameters for model 2 in the two regimes are quite close to the true parameters as well. Using the RJMCMC method, the posterior probabilities of identifying the true models are 0.9371 and 0.9996, respectively for Model 1 and Model 2.

For the data simulated from Model 2, Figure 1 shows the trace plots and histograms of all parameters in the first iteration. Note that the Bayesian estimate and the trace plot of d just indicate $d = 1$. The histograms and trace plots show that the results are quite satisfactory.

Table I. Simulation Results for MA(1)and TMA(2,1,1)

parameter	true value 1	means	s. d.	true value 2	means	s. d.
$\theta_1^{(1)}$	—	0.4243	0.0753	-0.5	-0.3283	0.0748
$\theta_1^{(2)}$	—	0.4640	0.0740	0.5	0.5098	0.0665
σ^2	1	1.0590	0.0857	1	0.8671	0.0722
r	—	-0.1650	1.8168	0.25	0.2006	0.1535
d	—	—	—	1	1	0
θ	0.5	0.4498	0.0446	—	-0.0218	0.0566

Figure 1. The trace plots and histograms of all parameters in the first stage

Remark 2. We have also applied the likelihood ratio test of Ling and Tong to the data simulated from model 1 and model 2. For model 1, the maximum likelihood estimate of θ is $\hat{\theta} = 0.5039$ and $LR_n = 1.6408$. For model 2, $\hat{\theta}_1^{(1)} = -0.4529$, $\hat{\theta}_1^{(2)} = 0.5240$, $\hat{r} = 0.2204$ and $LR_n = 107.64$. Where LR_n is the test statistic of Ling and Tong and calculated with “ $\beta_1 = 0.1$ and $\beta_2 = 0.9$ ”. The corresponding critical values of the null limiting distribution of LR_n are 6.995, 7.483 and 10.831 at significant levels 0.10, 0.05 and 0.01, respectively, which were given by Ling and Tong with “ $p = d = 1$ ”. This shows that the estimated models are accepted at all three significant levels.

Remark 3. The initial values of the sampling scheme above are $\theta_1^{(1)} = -0.3$, $\theta_1^{(2)} = -0.7$, $r = -0.4$, $d = 1$, $\sigma^2 = 1$. If the initial values are changed to $\theta_1^{(1)} = 0.3$, $\theta_1^{(2)} = 0.7$, $r = 0.4$, $d = 3$, $\sigma^2 = 2$, then from Figure 2, we find the parallel trace plots of all parameters are almost stationary in the first iteration. There is also a indication that the trace plots are convergent in some sense, e.g. the similarity between the trace plots for the former one and for the latter one.

Figure 2. The parallel chains of all parameters for the first stage

To verify the usefulness of our method a bit further, we conduct an investigation about MA(2) and TMA(2,2,2) models.

- Model 3: $MA(2)$

$$y_t = \varepsilon_t - 0.5\varepsilon_{t-1} - 0.5\varepsilon_{t-2} \quad (4.3)$$

- Model 4: $TMA(2, 2, 2)$

$$y_t = \varepsilon_t - (0.5\varepsilon_{t-1} + 0.5\varepsilon_{t-2})I(y_{t-2} > 0.4) + (0.5\varepsilon_{t-1} + 0.5\varepsilon_{t-2})I(y_{t-2} \leq 0.4) \quad (4.4)$$

We carry out 20000 MCMC iterations for two sets of simulated data of sample size $n = 300$. The hyper-parameters of $\Theta_{0i} = (0, 0)'$, $V_i = \text{diag}(0.1, 0.1)$, $i = 1, 2$, other parameters are the same as the previous example. The simulation results are shown in Table II.

From Table II, we see that, for the second order MA and TMA models, the posterior means are also closed to true values. The posterior probabilities of identifying the true model are 0.8431 and 0.9998, respectively for Model 3 and Model 4 by the RJMCMC scheme.

The above simulation analysis shows that our MCMC sampling methods perform efficiently in providing posterior samplers for statistical inference of TMA models.

Table II. Simulation Results for MA(2) and TMA(2,2,2)

parameter	true value 3	means	s. d.	true value 4	means	s. d.
$\theta_1^{(1)}$	—	0.5055	0.0448	0.5	0.4282	0.0783
$\theta_2^{(1)}$	—	0.4736	0.0898	0.5	0.5064	0.0572
$\theta_1^{(2)}$	—	0.4744	0.0910	-0.5	-0.5110	0.0457
$\theta_2^{(2)}$	—	0.4644	0.0751	-0.5	-0.6057	0.0613
σ^2	1	1.0603	0.0863	1	1.0169	0.0825
r	—	0.6969	2.3570	0.4	0.3627	0.0015
d	—	3	0.1037	2	2	0
θ_1	0.5	0.4467	0.0437	—	-0.2579	0.0544
θ_2	0.5	0.4708	0.0448	—	-0.0753	0.0617

4.2 A Real Example

Now we analyze the exchange rate of Japanese Yen v.s. USA dollar. The monthly data from Jan. 1971 to Dec. 2000 are used and there are 360 observations. This data set were analyzed recently by Ling and Tong (2005). P_t denotes the exchange rate at t th month. Let $x_t = 100[\log(P_t) - \log(P_{t-1})]$ and $y_t = x_t - \sum_{i=2}^{360} x_i/359$ for $t \geq 2$. We employ $MA(1)$ and $TMA(2,1,1)$ models to fit the data $\{y_2, y_3, \dots, y_{360}\}$, and take $d = 1$ to keep consistent with Ling and Tong(2005). Similar to Chen and Lee(1995) and Perreault (2000), the hyper-parameters used are $\Theta_{0i} = 0$ for $i = 1, 2$, $V = 0.1$, $\alpha = 2.5$, $\beta = 1.6$, and $a = p_5$ and $b = p_{95}$ are defined as in subsection 4.1. Set $M = 10000$ to run MCMC iterations. Burning the first 5000 times, the posterior means of σ^2 with $MA(1)$ and $TMA(2,1,1)$ models are 6.5849 and 6.8192 respectively, they are almost equal and accord with our hypotheses, the means of $\theta_1^{(1)}, \theta_1^{(2)}, r$ are -0.2820, -0.6761, -4.4631 respectively. Figure 2 displays the trace plots and histograms of $\theta_1^{(1)}, \theta_1^{(2)}, r$ for TMA model.

The posterior means of $\theta_1^{(1)}, \theta_1^{(2)}, r$ form the normal kernels, which are applied to RJMCMC iterations for 10000 times. The estimate of posterior probabilities identifying the TMA model is 0.9997. The corresponding results of posterior mean and posterior stand deviation shown in square brackets are recorded in Table II. From Table II, we can see that $\theta_1^{(2)}$ and threshold r are slightly different from those in Ling and Tong (2005). But our procedure avoids determining the threshold and the delay values via some complicated ways. This is the main advantage of our Bayesian approach. Furthermore, our Bayesian threshold nonlinearity test also performs satisfactorily.

Figure 3. The trace plots and histograms of $\theta_1^{(1)}, \theta_1^{(2)}, r$ for the first iteration

Table III. Parameter Estimate for Monthly Exchange Rate
of *JPY* against *USD* (Jan. 1971 to Dec. 2000)

parameter	$\theta_1^{(1)}$	$\theta_1^{(2)}$	σ^2	r	θ
L&T	-0.281	-0.726	—	-4.93	-0.402
MA(1)	—	—	6.8192 (0.5016)	—	-0.4001 (0.0500)
TMA(2,1,1)	-0.2796 (0.0621)	-0.6902 (0.0938)	6.5849 (0.4954)	-4.6423 (0.2740)	—

Remark 4. Chen and Lee (1995) believed that the results of Bayesian inference are generally not dependent on the priors selected. In our work, we perform two sets of sensitivity analysis, and obtain the following results.

(1) When $V = 0.2, \alpha = 6, \beta = 4$, we get $\theta_1^{(1)} = -0.2827, \theta_1^{(2)} = -0.6886, \sigma^2 = 6.4752, r = -4.6005$ and $p(M_2|\mathbf{y}) = 0.9998$.

(2) When $V = 0.3, \alpha = 10, \beta = 8$, we get $\theta_1^{(1)} = -0.2810, \theta_1^{(2)} = -0.6892, \sigma^2 = 6.3729, r = -4.5581$ and $p(M_2|\mathbf{y}) = 0.9997$.

These results provide strong evidence to support Chen and Lee (1995).

5 Conclusion

Combining Gibbs sampler and Metropolis-Hastings algorithm, we have proposed a Bayesian analysis of $TMA(2, q_1, q_2)$ model without employing the arranged autoregression approach. Our procedure is simple to implement and requires no subjective specification of threshold and delay values. Using the proposed procedure, we develop a Bayesian testing scheme for threshold non-linearity for two-regime TMA models. The main idea is to compute the posterior probabilities of competitive models using RJMCMC method. Simulation results and application to a real example lend further support to our method.

ACKNOWLEDGMENTS

The authors thank the associate editor and anonymous reviewer(s) for their constructive comments, which lead to a significant improvement of this paper. The work in this paper was supported by a grant from President Science Foundation of South China Agricultural University (Project No:2008K015).

References

- [1] Casella, G. and George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, **46**, 167-174.
- [2] Chen, C. W. S. and Lee, J. C. (1995). Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis*, **16**, 483-492.
- [3] Chen, C. W. S, Lin, E. M. H. and Liu, F. C. (2008). Bayesian estimation for parsimonious threshold autoregressive models in R. *Newsletter of R project*, **8**, 26-33.
- [4] Chen, C. W. S., Gerlach, R. and So, M. K. P. (2008). Bayesian model selection for heteroskedastic models. *Advances in Econometrics*, **25**, 567-594.
- [5] Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49**, 327-335.
- [6] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.

- [7] Geweke, J. and Terui, N. (1993). Bayesian threshold autoregressive models for nonlinear time series. *Journal of Time Series Analysis*, **14**, 441-454.
- [8] Green, P.J. (1995) Reversible jump MCMC computation and Bayesian model determination. *Biometrika*, **82**, 711-732.
- [9] Hastings, W. K. (1970). Monte-Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97-109.
- [10] Ling, S. and Tong, H. (2005). Testing a linear moving-average model against threshold moving-average models. *Ann. Statist.*, **33**, 2529-2552.
- [11] Ling, S., Tong, H. and Li D. (2007). The ergodicity and invertibility of threshold moving-average models. *Bernoulli*, **13**, 161-168.
- [12] Perreault, L., Bernier, J., Bobée, B. and Parenta, E. (2000). Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. *Journal of Hydrology*, **235**, 221-241.
- [13] Perreault, L., Bernier, J., Bobée, B. and Parenta, E. (2000). Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting. *Journal of Hydrology*, **235**, 242-263.
- [14] Ismail, M. A. and Charif, H. A. (2003). Bayesian Inference for threshold moving average models. *METRON*, **1**, 119-132.
- [15] McCulloch, R. E. and Tsay, R. S. (1993a). Bayesian Analysis of Threshold Autoregressive Processes with a Random Number of Regimes. *Computing Science and Statistics. Proceedings of the 25th Symposium on the Interface*, Fairfax Station, VA: Interface Foundation of North America, 253-262.
- [16] McCulloch, R. E. and Tsay, R. S. (1993b). Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series. *Journal of the American Statistical Association*, **88**, 968-978.
- [17] Metropolis, N., Rosenbluth, A. W, Rosenbluth, M. N. and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087-1091.
- [18] So, M. K. P., Chen, C. W. S. and Chen, M. T. (2005) A Bayesian threshold nonlinearity test in Financial time series, *Journal of Forecasting*, **24**, 61-75.

- [19] Sáfadi, T. and Morettin, P. A. (2000). Bayesian Analysis of threshold autoregressive moving average models. *Sankhyā*, **62**, 353-371.
- [20] Tong, H. (1978). On a Threshold Model, in C.H. Chen (eds). *Pattern Recognition and Signal Processing*, Amsterdam: Sijthoff and Noordhoff, 101-141.
- [21] Tong, H. and Lim, K. S. (1980). Threshold Autoregressions, Limit Cycles, and Data. *Journal of the Royal Statistical Society*, **42**, 245-292.
- [22] Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford.
- [23] Tsay, R. S. (1987). Conditional heteroscedastic time series models. *Journal of the American Statistical Association*, **82**, 590-604.
- [24] Tsay, R. S. (1989). Testing and modeling threshold autoregressive process. *Journal of the American Statistical Association*, **84**, 231-240.
- [25] Tsay, R. S. (2005). *Analysis of Financial Time Series*. 2nd Edition, John Wiley & Sons.
- [26] Vrontos, I. D., Dellaportas, P. and Politis, D. N. (2000). Full Bayesian inference for GARCH and EGARCH models. *Journal of Business & Economic Statistics*, **18**, 187-198.