Statistics and Its Interface Volume 15 (2022) 51–61

Subset selection of double-threshold moving average models through the application of the Bayesian method*

Jinshan Liu, Jiazhu Pan, Qiang Xia†, and Ying Xiao

The Bayesian method is firstly applied for the selection of the best subset for the double-threshold moving average (DTMA) model. The Markov chain Monte Carlo (MCMC) techniques and the stochastic search variable selection (SSVS) method are used to identify the best subset model from a very large number of possible models. Simulation experiments show that the proposed method is feasible and efficient, despite the complexity being increased by the large number of subsets, and the uncertainty of the threshold and delay variables. Our method is illustrated by real data analysis on the Yen-Dollar exchange rate.

Keywords and phrases: Bayesian estimation, DTMA model, MCMC algorithm, SSVS, Exchange rate.

1. INTRODUCTION

Since Tong (1978) and Tong and Lim (1980) proposed the threshold autoregressive (TAR) models, nonlinear time series models have attracted more and more attention, and TAR model has become a hot topic in the analysis of nonlinear time series. There has been great interest for the theory of TAR models in the literature. See Chan and Tong (1985), Tong (1990) and Tsay (2005), and many others. Nevertheless some authors suggested the Bayesian approach, which can avoid complex analytical work and numerical multiplex integration for inference TAR models. For example, Mcculloch and Tsay (1993) proposed a Bayesian method for detecting the threshold value in the TAR model through the posterior probability map. 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. Pan et al. (2017) introduced the Bayesian stochastic search selection of TAR models to identify a threshold-dependent sequence with the highest probability. Meanwhile, in the literature, attention has been paid for threshold moving average (TMA) models using MCMC algorithm, because people realized TMA models are as important as TAR models in practice. See Liu and Susko (1992), Sáfadi and Morettin (2000), Ismail and Charif (2003), Ling and Tong (2005), Ling et al. (2007), Xia et al. (2010), Chen et al. (2010), Li et al. (2012) and Li (2012), among others.

However, the above papers only discussed the models with a single threshold variable. This greatly limits the scope of application of the model. In real data analysis, two and even more thresholds are required. For example, Leeper (1991) divided the policy parameter space into four disjointed regions depending on whether monetary and fiscal policies are active or passive. Tiao and Tsay (1994) proposed a four-region autoregressive model for the actual US quarterly GNP growth rate based on past growth rate levels and positive or negative signs. In terms of turnover and average price, Chen et al. (2012) and Ni et al. (2018) divided the daily return series of Hang Seng Index into three and four regimes, respectively. But it is found that the model selection discussion in these examples is on the double-threshold AR (DTAR) model. Therefore, as the new progress of study, a natural idea is to extend DTAR models to DTMA models in the relevant theory and application.

The aim of this paper is to study the Bayesian approach for the estimation problem for DTMA models. In literature, in order to avoid complex analytical work and numerical multiplex integration, Chen and Lee (1995), Ismail and Charif (2003), and Sáfadi and Morettin (2000) used the MCMC techniques and simultaneously estimated the threshold parameters and other parameters using permuted autoregressive methods. So and Chen (2003) developed a selection method of SETAR model, which can estimate unknown parameters, including threshold parameters and delay variables, and simultaneously identified the best subset SETAR model in Bayesian framework. In this paper, based on these work, we use the SSVS method proposed by George and McCulloch (1993) to select the best subset of DTMA models. Meanwhile, we apply the MCMC method which combines the Gibbs sampler and the M-H algorithm to generate the posterior samples of all parameters in the DTMA model. It is worth mentioning that we can identify the best subset of DTMA with the highest posteriori probability, and estimate the unknown parameters including the threshold parameters and delay variables well.

The contents of this paper is organized as follows: Section 2 describes the DTMA model and the selection of its

*The authors are arranged in alphabetical order.

†Corresponding author.

subset; Section 3 gives the details of Bayesian inference based on the MCMC algorithm; Some simulation studies are presented in Section 4; Section 5 shows an example of real data analysis; Section 6 contains our conclusion.

2. THE DTMA MODEL AND THE SELECTION OF ITS SUBSET

2.1 The DTMA model

We consider the following double-threshold moving average models, denoted by DTMA$(2; q_1, q_2, q_3, q_4, d_1, d_2)$:

(1)
$$y_t = \begin{cases} \theta_0^{(1)} + \varepsilon_t^{(1)} - \theta_1^{(1)}\varepsilon_{t-1}^{(1)} - \theta_2^{(1)}\varepsilon_{t-2}^{(1)} - \cdots - \theta_{q_1}^{(1)}\varepsilon_{t-q_1}^{(1)}, & z_{1,t-d_1} \le r_1, z_{2,t-d_2} \le r_2, \\ \theta_0^{(2)} + \varepsilon_t^{(2)} - \theta_1^{(2)}\varepsilon_{t-1}^{(2)} - \theta_2^{(2)}\varepsilon_{t-2}^{(2)} - \cdots - \theta_{q_2}^{(2)}\varepsilon_{t-q_2}^{(2)}, & z_{1,t-d_1} \le r_1, z_{2,t-d_2} > r_2, \\ \theta_0^{(3)} + \varepsilon_t^{(3)} - \theta_1^{(3)}\varepsilon_{t-1}^{(3)} - \theta_2^{(3)}\varepsilon_{t-2}^{(3)} - \cdots - \theta_{q_3}^{(3)}\varepsilon_{t-q_3}^{(3)}, & z_{1,t-d_1} > r_1, z_{2,t-d_2} \le r_2, \\ \theta_0^{(4)} + \varepsilon_t^{(4)} - \theta_1^{(4)}\varepsilon_{t-1}^{(4)} - \theta_2^{(4)}\varepsilon_{t-2}^{(4)} - \cdots - \theta_{q_4}^{(4)}\varepsilon_{t-q_4}^{(4)}, & z_{1,t-d_1} > r_1, z_{2,t-d_2} > r_2. \end{cases}$$

The above model (1) is a MA model with two threshold variables and four regimes (or four-mechanisms), where r_1, r_2 are the threshold parameters, d_1, d_2 are commonly referred to the delays (or threshold lags) of the model. For $j = 1, 2, 3, 4$, $\{\varepsilon_t^{(j)}\}$ is assumed to be a sequence of independent and identically distributed (i.i.d.) random variables with distribution $N(0, \sigma_j^2)$; the positive integer q_j is the order of the j-th regime; $z_{1,t-d_1}$ and $z_{2,t-d_2}$ are called the threshold variables, which are observable, exogenous or endogenous variables. If the threshold variables are endogenous, then they relate to y_t or functions of y_t. The parameters in model (1) are $\Theta^{(j)} = (\theta_0^{(j)}, \theta_1^{(j)}, \cdots, \theta_{q_j}^{(j)})'$ with $\Theta^{(i)} \ne \Theta^{(j)}$ for $i \ne j$.

Let $Y = \{y_1, y_2, \cdots, y_n\}$, $q = max\{q_1, q_2, q_3, q_4\} + 1$. Given the first $q-1$ observations, we can write the conditional likelihood function of the model (1) as follows:

(2)
$$\begin{aligned} & L(\Theta^{(j)}, \sigma_j^2, r_1, r_2, d_1, d_2 (j = 1,2,3,4)|Y) \\ & \propto \prod_{t=q}^{n}\prod_{j=1}^{4} \frac{1}{\sigma_j} \exp\{-\frac{1}{2\sigma_j^2}(y_t - \theta_0^{(j)} + \theta_1^{(j)}\varepsilon_{t-1}^{(j)} + \cdots \\ & + \theta_{q_j}^{(j)}\varepsilon_{t-q_j}^{(j)})^2 I^{(j)}\} \\ & \propto \sigma_1^{-n_1}\sigma_2^{-n_2}\sigma_3^{-n_3}\sigma_4^{-n_4} \exp\{ \\ & -\frac{1}{2\sigma_1^2}\sum_{t=q}^{n}(y_t - \theta_0^{(1)} + \theta_1^{(1)}\varepsilon_{t-1}^{(1)} + \cdots + \theta_{q_1}^{(1)}\varepsilon_{t-q_1}^{(1)})^2 \\ & \times I(z_{1,t-d_1} \le r_1, z_{2,t-d_2} \le r_2) \\ & -\frac{1}{2\sigma_2^2}\sum_{t=q}^{n}(y_t - \theta_0^{(2)} + \theta_1^{(2)}\varepsilon_{t-1}^{(2)} + \cdots + \theta_{q_2}^{(2)}\varepsilon_{t-q_2}^{(2)})^2 \\ & \times I(z_{1,t-d_1} \le r_1, z_{2,t-d_2} > r_2) \\ & -\frac{1}{2\sigma_3^2}\sum_{t=q}^{n}(y_t - \theta_0^{(3)} + \theta_1^{(3)}\varepsilon_{t-1}^{(3)} + \cdots + \theta_{q_3}^{(3)}\varepsilon_{t-q_3}^{(3)})^2 \\ & \times I(z_{1,t-d_1} > r_1, z_{2,t-d_2} \le r_2) \\ & -\frac{1}{2\sigma_4^2}\sum_{t=q}^{n}(y_t - \theta_0^{(4)} + \theta_1^{(4)}\varepsilon_{t-1}^{(4)} + \cdots + \theta_{q_4}^{(4)}\varepsilon_{t-q_4}^{(4)})^2 \\ & \times I(z_{1,t-d_1} > r_1, z_{2,t-d_2} > r_2)\}, \end{aligned}$$

where $I(\cdot)$ denotes the indication function, n_1 is the number of observations which satisfy the condition of $z_{1,t-d_1} \le r_1, z_{2,t-d_2} \le r_2$, and n_2, n_3, n_4 are defined accordingly. Further, $\varepsilon_t^{(j)}$ can be recursively computed by

$$\varepsilon_t^{(j)} = y_t - \theta_0^{(j)} + \theta_1^{(j)}\varepsilon_{t-1}^{(j)} + \cdots + \theta_{q_1}^{(j)}\varepsilon_{t-q_1}^{(j)}.$$

Denote $\Sigma^{(j)} = diag(\sigma_j^2, \sigma_j^2, \cdots, \sigma_j^2)$, which is an $(n-q+1) \times (n-q+1)$ matrix; $Y^{(j)} = (y_q, y_{q+1}, \cdots, y_n)$, $x_t^{(j)} = (1, -\varepsilon_{t-1}^{(j)}, \cdots, -\varepsilon_{t-q-j}^{(j)})'$, $X^{(j)} = (x_q^{(j)}, x_{q+1}^{(j)}, \cdots, x_n^{(j)})$; and

$$\begin{aligned} S^{(1)} = \; & diag(I(z_{1,q-d_1} \le r_1, z_{2,q-d_2} \le r_2), \ldots, \\ & I(z_{1,n-d_1} \le r_1, z_{2,n-d_2} \le r_2)), \\ S^{(2)} = \; & diag(I(z_{1,q-d_1} \le r_1, z_{2,q-d_2} > r_2), \ldots, \\ & I(z_{1,n-d_1} \le r_1, z_{2,n-d_2} > r_2)), \\ S^{(3)} = \; & diag(I(z_{1,q-d_1} > r_1, z_{2,q-d_2} \le r_2), \ldots, \\ & I(z_{1,n-d_1} > r_1, z_{2,n-d_2} \le r_2)), \\ S^{(4)} = \; & diag(I(z_{1,q-d_1} > r_1, z_{2,q-d_2} > r_2), \ldots, \\ & I(z_{1,n-d_1} > r_1, z_{2,n-d_2} > r_2)). \end{aligned}$$

Therefore, the conditional likelihood function (2) can be expressed as:

(3)
$$\begin{aligned} & L(\Theta^{(j)}, \sigma_j^2, r_1, r_2, d_1, d_2 (j = 1,2,3,4)|Y) \\ & \propto \prod_{j=1}^{4} \sigma_j^{-tr(S^{(j)})} \exp\{-\frac{1}{2}\sum_{j=1}^{4}(Y^{(j)} - X^{(j)'}\Theta^{(j)})' \\ & (\Sigma^{(j)})^{-1}(Y^{(j)} - X^{(j)'}\Theta^{(j)})\}, \end{aligned}$$

where $tr(S^{(j)})$ represent the trace of matrix $S^{(j)}$ $(j = 1, 2, 3, 4)$.

2.2 A mixture specification

The main objective of this paper is to identify the best subset of the DTMA model by SSVS method. George and McCulloch (1993) suggested the discrete indicators to identify good regression models using SSVS MCMC method, which was successfully employed to TAR, TARMA and DTAR models in So and Chen (2003), Chen et al. (2011) and Ni et al. (2018). Based on this idea, the binary indicator variables $\delta_{j,m}$, $m = 1, 2, \cdots, q_j$, $j = 1, 2, 3, 4$, are introduced which take a value of either 0 or 1. Each value of $\delta_{j,m}$ determines the distribution of $\Theta_m^{(j)}$. Then, prior assumptions are assumed on a single $\Theta_m^{(j)}$ by the normal mixture distribution, as follows:

$$\Theta_m^{(j)}|\delta_{j,m} \sim (1-\delta_{j,m})N(0,\tau_{j,m}^2) + \delta_{j,m}N(0,c_{j,m}^2\tau_{j,m}^2) \tag{4}$$

and

$$\delta_{j,m} = \begin{cases} 1, & \text{with probability} \quad \gamma_{j,m}, \\ 0, & \text{with probability} \quad 1-\gamma_{j,m}. \end{cases} \tag{5}$$

The general specification allows correlations among $\Theta_m^{(j)}$, and the mixture distribution in (4) can be stated as following multivariate normal prior for the slope parameters $\Theta^{(j)} = (\Theta_0^{(j)}, \Theta_1^{(j)}, \cdots, \Theta_{q_j}^{(j)})$:

$$\Theta^{(j)}|\delta_j \sim N(0, M_{\delta_j} D_j M_{\delta_j}) \tag{6}$$

where $\delta_j = (\delta_{j,0}, \delta_{j,1}, \cdots, \delta_{j,q_j})'$, D_j is the prior correlation matrix and $M_{\delta_j} = diag\{a_{j,0}\tau_{j,0},$
$\cdots, a_{j,q_j}\tau_{j,q_j}\}$ with $a_{j,m} = 1$ if $\delta_{j,m} = 0$ and $a_{j,m} = c_{j,m}$ if $\delta_{j,m} = 1$.

When no prior information about the relationship among $\Theta_m^{(j)}, j = 1,2,3,4$, are available, the prior distribution just simplified with $D_j = \Lambda$, where Λ is a diagonal matrix with elements $a_{j,m}^2\tau_{j,m}^2$, $m = 0,1,\cdots,q_j$. That is,

$$p(\Theta^{(j)}|\delta_j) = \prod_{m=0}^{q_j} p(\Theta_m^{(j)}|\delta_{j,m}). \tag{7}$$

If $\delta_{j,0}, \delta_{j,1}, \cdots, \delta_{j,q_j}$ are chosen to be small, those $\Theta_m^{(j)}$, which is associated with $\delta_{j,m} = 0$, are also likely to be small. On the other hand, $c_{j,0}, \cdots, c_{j,q_j}$ could be chosen greater than 1 to make $c_{j,m}^2\tau_{j,m}^2 \geq \tau_{j,m}^2$. Consequently, those $\Theta_m^{(j)}$ associated with $\delta_{j,m} = 1$ will have a high variability. In other words, variables associated with $\delta_{j,m} = 1$ are considered very useful for likely moving $\Theta_m^{(j)}$ away form zero. On contrary, the variables having $\delta_{j,m} = 0$ are taken to be unimportant.

In order to select the best subset, suitable choice of the hyper-parameters $\tau_{j,m}, c_{j,m}$ are very important. Note that the ratio combinations $(\sigma_{\Theta_m^{(j)}}/\tau_{j,m}, c_{j,m})$ were selected as $(1,5)$, $(1,10)$ by George and McCulloch (1993), and $(0.5,5)$, $(0.5,10)$ by So and Chen (2003). Both paper are found to have superior performance. Ni et al. (2018) set $(\Theta/\tau_{j,m}, c_{j,m})$ as $(0.1,5)$, $(0.2,5)$ and $(0.1,10)$, which also have achieved good results. In our models, we consider setting $(\Theta/\tau_{j,m}, c_{j,m})$ as $(0.1,5)$, $(0.1,10)$ and $(0.1,15)$, where Θ represents the maximum value of the absolute value of $\Theta_m^{(j)}$, without thinking about the value of $\sigma_{\Theta_m^{(j)}}$.

3. BAYESIAN INFERENCE

3.1 Conditional posterior distribution

In order to complete the selection of the best subset of the model by using MCMC methods, we need to infer the posterior distribution of each parameter in the DTMA model, so the priori information selection of the parameter is particularly important. Referring to Chen and Lee (1995), Xia et al. (2010) and Pan et al. (2017), the determination of the priori distribution can be completed.

(1) For $i \neq j$, assume $\Theta^{(j)}|\delta_j \sim N(0, M_{\delta_j} D_j M_{\delta_j})$, $(j = 1,2,3,4)$, which is independent of $\Theta^{(i)}$. Through the prior distribution, we conclude that the conditional posterior distribution of $\Theta^{(j)}$ is a pseudo-normal distribution as follows:

$$p(\Theta^{(j)}|Y,\delta_j,\sigma_j^2,r_1,r_2,d_1,d_2) \sim pN(U^{*(j)},\Sigma^{*(j)}), \tag{8}$$
$$U^{*(j)} = S^{(j)}\sum_{t=q}^{n}\{(\frac{x_t^{(j)}x_t^{(j)'}}{\sigma_j^2} + M_{\delta_j}^{-1}D_j^{-1}M_{\delta_j}^{-1})^{-1}\frac{x_t^{(j)}y_t}{\sigma_j^2}\},$$

and

$$\Sigma^{*(j)} = S^{(j)}\sum_{t=q}^{n}\{(\frac{x_t^{(j)}x_t^{(j)'}}{\sigma_j^2} + M_{\delta_j}^{-1}D_j^{-1}M_{\delta_j}^{-1})^{-1}\};$$

and $\Theta^{(j)}$ is independent of $\Theta^{(i)}$ for $i \neq j$.

(2) We employe the inverse gamma distribution for σ_j^2, which is independent of σ_i^2 for $i \neq j$, $\sigma_j^2 \sim IG(\frac{v_j}{2}, \frac{v_j\lambda_j}{2})$ $(j = 1,2,3,4)$. Then the posterior distribution of σ_j^2, independently with σ_i^2 for $i \neq j$, is also the inverse gamma distribution:

$$\begin{aligned} &p(\sigma_j^2|Y,\delta_j,\Theta^{(j)},r_1,r_2,d_1,d_2) \\ &\quad \sim IG(\frac{v_j+n_j}{2}, \frac{v_j\lambda_j + n_j s_j^2}{2}), \end{aligned} \tag{9}$$

where $j = 1,2,3,4$, $n_j = tr(S^{(j)})$, and

$$s_j^2 = \frac{1}{n_j}\sum_{t=q}^{n}(y_t - \Theta^{(j)'}x_t^{(j)}S^{(j)})^2, j = 1,2,3,4.$$

(3) When r_i follows the uniform distribution (a_i, b_i), $i = 1,2$, and r_1, r_2 are independent each other, the posterior density function of r_1 is derived from Bayes theorem:

$$\begin{aligned} &p(r_1|Y,\delta_j,\Theta^{(j)},\sigma_j^2,r_2,d_1,d_2) \\ &\propto \prod_{j=1}^{4}\sigma_j^{-n_j}\exp\{-\frac{1}{2}\sum_{j=1}^{4}(Y^{(j)} - X^{(j)'}\Theta^{(j)})' \\ &(\Sigma^{(j)})^{-1}(Y^{(j)} - X^{(j)'}\Theta^{(j)})\}, \end{aligned} \tag{10}$$

where $j = 1,2,3,4$, $n_j = tr(S^{(j)})$, and the posterior density function of r_2 is similar to r_1.

(4) Let d_1 follow the discrete uniform prior distribution over $0,1,2,\cdots,d_{1,0}$, and d_2 be independent of d_1. The conditional posterior distribution of d_1 is a multinomial

distribution:

(11)

$$p(d_1|Y,\delta_j,\Theta^{(j)},\sigma_j^2,r_1,r_2,d_2) = \frac{L(\Theta^{(j)},\sigma_j^2,r_1,r_2,d_1,d_2(j=1,2,3,4)|Y)}{\sum_{d_1=0}^{d_{1,0}} L(\Theta^{(j)},\sigma_j^2,r_1,r_2,d_1,d_2(j=1,2,3,4)|Y)},$$

where $j = 1,2,3,4$, $d_1 = 0,1,2,\cdots,d_{1,0}$, and $L(\Theta^{(j)},\sigma_j^2,r_1,r_2,d_1,d_2(j=1,2,3,4)|Y)$ is the likelihood function in (3). The conditional posterior distribution of d_2 is similar to that of d_1.

(5) The conditional posterior distribution of δ_j is a Bernoulli distribution with the probability

(12)

$$p(\delta_{j,m}|Y,\delta_{j,-m},\Theta^{(j)},\sigma_j^2,r_1,r_2,d_1,d_2(j=1,2,3,4)) = \frac{A_{j,m}}{A_{j,m}+B_{j,m}}.$$

where

$$A_{j,m} = p(\Theta_m^{(j)}|\delta_{j,-m},\delta_{j,m}=1)\gamma_{j,m},$$
$$B_{j,m} = p(\Theta_m^{(j)}|\delta_{j,-m},\delta_{j,m}=0)(1-\gamma_{j,m}).$$

In particular, if $D_j = \Lambda$ and the uniform prior is specified for $\delta_{j,m}$, for example $\gamma_{j,m} = \frac{1}{2}$, the $A_{j,m}$ and $B_{j,m}$ can be simplified to $p(\Theta_m^{(j)}|\delta_{j,m}=1)$ and $p(\Theta_m^{(j)}|\delta_{j,m}=0)$.

3.2 Sampling scheme

From the previous analysis, we can see that the full conditional posterior distributions of the parameters are identified except for $\Theta^{(j)}$, r_1 and r_2. Then, the Gibbs sampler should be used to the standard posterior distributions. Owing to the parameters $\Theta^{(j)}$, r_1 and r_2 don't have a specific distribution, we need to employ the M-H algorithm to draw it, which can be found in Chib and Greenberg (1995). For example, let $f(\cdot)$ be the conditional density in (10), the algorithm of drawing r_i is described below.

In order to enhance the convergence of the MCMC algorithm, we will modify the sampling plan of r_i in the following. First, we perform the M-H sampling step of random walk for M iterations. Then we use the previous M iterations to obtain the sample mean μ_{r_i} and the sample covariance Ω_{r_i}. Finally, using of the Gaussian proposal distribution with mean μ_{r_i} and covariance Ω_{r_i}, we apply the independent kernel M-H algorithm to r_i starting from the $M+1$ iteration, as follows:

Step 1: At iteration j, generate a point r_i^* from the independent kernel M-H algorithm,

$$r_i^* = \mu_{r_i} + \varepsilon_{r_i}, \quad \varepsilon_{r_i} \sim N(0,\Omega_{r_i}),$$

where r_i^{j-1} is the $(j-1)$th of r_i.

Step 2: Accept $r_i^* = r_i^j$, if the probability meets the condition $p = \min\{1, \frac{f(r_i^*)q(r_i)}{f(r_i^{j-1})q(r_i^*)}\}$. Otherwise, set $r_i^j = r_i^{j-1}$. Where, r_i^{j-1} is the $(j-1)$th of r_i, and $q(r_i^*) \propto exp\{-\frac{1}{2}(r_i - \mu_{r_i})'\Omega_{r_i}(r_i-\mu_{r_i})\}$. Similarly, the M-H sampling algorithm for other parameters can be shown as above. In summary, we use the following iterative sampling scheme to construct the desired posterior sample:

(1) Draw $\Theta^{(j)}$ using the random walk and the independent kernel M-H algorithm from the conditional posterior distribution in (8), $j = 1,2,3,4$;

(2) Draw σ_j^2 from the inverse Gamma distribution in (9), $j = 1,2,3,4$;

(3) Draw r_i using the random walk and the independent kernel M-H algorithm from the conditional posterior distribution in (10), $i = 1,2$;

(4) Draw d_i from the multinomial distribution in (11), $i = 1,2$;

(5) Draw $\delta_{j,m}$ from the Bernoulli distribution in (12), $j = 1,2,3,4$.

This completes one iteration. Of course, we can change the order to attain fast convergence in sampling the variables.

4. SIMULATION EXPERIMENTS

This section uses two different models for simulation experiments to verify the accuracy of the Bayesian optimal subset selection and the fitting effect of the sampling scheme. In order to improve the validity of the sampling results, the MCMC algorithm is performed a total of 8000 sampling iterations, the pre-iteration values of the previous 5000 transition periods are discarded, and the parameters are estimated based on sample values of back 3000 times. For each of the simulation models, 1500 observations are produced, and the last 1000 observations are selected as samples. 100 simulations are run twice to verify the accuracy of the Bayesian optimal subset selection. The values of the hyper-parameters for the prior distribution are taken as

$$v_j = 3, \quad \lambda_j = 1, \quad D_j = I,$$
$$d_{i,0} = 3, \quad a_i = p_{i,10}, \quad b_i = p_{i,90},$$

where $j = 1,2,3,4$, $i = 1,2$ and $p_{i,k}$ represents the k quantile of the data $\{z_{i,t}\}$.

4.1 Simulation I

The first model that we consider is

$$(13) \quad y_t = \begin{cases} 0.3 + 0.6\varepsilon_{t-1}^{(1)} + \varepsilon_t^{(1)}, & z_{1,t-1} \le 0, z_{2,t-1} \le 0, \\ 0.7 - 0.2\varepsilon_{t-1}^{(2)} + \varepsilon_t^{(2)}, & z_{1,t-1} \le 0, z_{2,t-1} > 0, \\ -0.2 - 0.7\varepsilon_{t-1}^{(3)} + \varepsilon_t^{(3)}, & z_{1,t-1} > 0, z_{2,t-1} \le 0, \\ 0.5 + 0.3\varepsilon_{t-1}^{(4)} + \varepsilon_t^{(4)}, & z_{1,t-1} > 0, z_{2,t-1} > 0, \end{cases}$$

Table 1. Estimated results of model (13) based on 100 replications

Parameter	True Value	Mean	S.D.
$\theta_0^{(1)}$	0.3	0.2856	0.0192
$\theta_1^{(1)}$	0.6	0.5913	0.0263
$\theta_0^{(2)}$	0.7	0.6935	0.0435
$\theta_1^{(2)}$	−0.2	−0.2290	0.0307
$\theta_0^{(3)}$	−0.2	−0.2263	0.0969
$\theta_1^{(3)}$	−0.7	−0.7211	0.0476
$\theta_0^{(4)}$	0.5	0.4889	0.0432
$\theta_1^{(4)}$	0.3	0.2822	0.0293
σ_1^2	1	1.0090	0.0932
σ_2^2	1	1.0062	0.0913
σ_3^2	1	0.9472	0.0835
σ_4^2	1	1.0480	0.0786
r_1	0	0.0037	0.0564
r_2	0	0.0013	0.0205
d_1	1	1	0
d_2	1	1	0

where $\varepsilon_t^{(j)} \sim N(0,1), j = 1,2,3,4,\ z_{i,t} \sim N(0,1), i = 1,2,$ and $y_1 = y_2 = y_3 = y_4 = 0$.

Table 1 records the posterior mean and posterior standard deviation (S.D.) of the estimated results of each parameter after the model (13) was repeated for 100 times. It can be seen from Table 1 that the posterior mean of each parameter is very close to the true value, and the posterior standard deviation of each parameter is very small, indicating that the sampling method has a very good fitting effect on the model (13).

Let $q_1 = q_2 = q_3 = q_4 = 1$, then the model (13) has $2^{4q+4} = 256$ possible subsets. A subset of the real model (13) can be represented by the following matrix:

$$(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4)' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Table 2 shows $(\Theta/\tau_{(j,m)}, c_{(j,m)})$ based on three different values, the optimal subset of $(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4)'$ selected from 100 repetitions and the posterior probability (Pos. Prob.). In each data set, the model with the highest a posteriori probability, or the model that is the most frequently selected during the MCMC operation, is selected as the best. In our simulation experiment, we use $(\Theta/\tau_{(j,k)}, c_{(j,k)})$ and select the three sets of values (0.1, 5), (0.1, 10) and (0.1, 15), where Θ represents the absolute maximum value of $\Theta_m^{(j)}$.

From the last 2000 MCMC iterations, the optimal subset of the model and the second best subset and posterior probabilities are recorded in Table 2, and three sets of $(\Theta/\tau_{(j,m)}, c_{(j,m)}) = (0.1, 5), (0.2, 5), (0.1, 10)$ are also recorded. For all sets, the true subset is successfully selected as the best subset of the model, and the posterior probabilities of the three groups are 0.2037, 0.4153 and 0.5447. In addition, the second best model found by the selection scheme is close to the real model, and the respective posterior probabilities are 0.0020, 0.0003, 0.0001. Compared with these results, the posterior probability of the best subset selection is much higher than the second, and $(\Theta/\tau_{(j,m)}, c_{(j,m)}) = (0.1, 15)$ is the preferred value in the selection scheme.

The selected results from the 100 replications of the simulations are shown in Table 3. We calculate the proportions of correct selection of the true model. Proportion 1 shows the accuracy of selecting the true model as the best one; proportion 2 indicates that the true model is selected either the best or the second best subsets. In this example, it seems that our Bayesian subset selection scheme performs well, especially when $(\Theta/\tau_{(j,m)}, c_{(j,m)}) = (0.1, 15)$.

4.2 Simulation II

The second model that we consider is:

$$y_t = \begin{cases} -0.5\varepsilon_{t-1}^{(1)} + 0.5\varepsilon_{t-2}^{(1)} + 0.5\varepsilon_{t-3}^{(1)} + \varepsilon_t^{(1)}, \\ \quad z_{1,t-1} \le 0, z_{2,t-2} \le -0.5, \\ 0.5 - 0.5\varepsilon_{t-2}^{(2)} + \varepsilon_t^{(2)}, \\ \quad z_{1,t-1} \le 0, z_{2,t-2} > -0.5, \\ 0.5\varepsilon_{t-1}^{(3)} - 0.5\varepsilon_{t-3}^{(3)} - 0.5\varepsilon_{t-4}^{(3)} + \varepsilon_t^{(3)}, \\ \quad z_{1,t-1} > 0, z_{2,t-2} \le -0.5, \\ 0.5 + 0.5\varepsilon_{t-4}^{(4)} + \varepsilon_t^{(4)}, \\ \quad z_{1,t-1} > 0, z_{2,t-2} > -0.5. \end{cases} \tag{14}$$

where $\varepsilon_t^{(j)} \sim N(0, 0.25),\ j = 1,2,3,4,\ z_{1,t-1} \sim N(0,1),$ $z_{2,t-2} \sim N(-0.5, 1)$, and $y_1 = y_2 = y_3 = y_4 = 0$.

Based on 100 samples, the estimated results are shown in Table 4, which lists the true values, posterior means, and posterior standard deviations for model (14). It can be seen from Table 4 that the posterior means are closed to the true values. Meanwhile, compared to the standard deviation of the posterior of some parameters, the average value is small. These results indicate an estimate of 0, such as $\Theta_0^{(1)}$ and $\Theta_4^{(1)}$ in the first mechanism. In addition, other posterior standard deviations are small. Therefore, we believe that the estimation of the parameters applies to the model (14).

Let $q_1 = q_2 = q_3 = q_4 = 4$, then model (14) has $2^{2q+4} = 4096$ possible subsets. A subset of the real model (14) can be expressed as follows:

$$(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4)' = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Like the simulation experiment for model (13), from the last 3000 MCMC iterations, the best model and the second best model with posterior probabilities for model (14) are recorded in Table 5. Three sets of $(\Theta/\tau_{(j,m)}, c_{(j,m)}) = (0.1, 5), (0.1, 10)$ and $(0.1, 15)$ are examined and recorded. For all sets, the true model was successfully chosen as the

Table 2. Subset selection results of model (13)

	(0.1, 5)	(0.1, 10)	(0.1, 15)
Best	$\begin{pmatrix}1&1&1&1\\1&1&1&1\end{pmatrix}$	$\begin{pmatrix}1&1&1&1\\1&1&1&1\end{pmatrix}$	$\begin{pmatrix}1&1&1&1\\1&1&1&1\end{pmatrix}$
Pos. Prob.	[0.2037]	[0.4153]	[0.5447]
Second best	$\begin{pmatrix}0&1&0&1\\1&0&1&0\end{pmatrix}$	$\begin{pmatrix}0&1&0&1\\1&0&1&0\end{pmatrix}$	$\begin{pmatrix}0&1&0&1\\1&0&1&0\end{pmatrix}$
Pos. Prob.	[0.0020]	[0.0003]	[0.0001]

Table 3. Proportion of correct selection of the true model (13) with 100 replications

$(\Theta/\tau_{(j,m)}, c_{(j,m)})$	Proportion 1	Proportion 2
(0.1, 5)	94	100
(0.1, 10)	92	100
(0.1, 15)	94	100

Table 4. Estimated results of model (14) based on 100 replications

Parameter	True Value	Mean	S.D.
$\theta_0^{(1)}$	0	−0.0002	0.0552
$\theta_1^{(1)}$	−0.5	−0.5087	0.0664
$\theta_2^{(1)}$	0.5	0.5087	0.0704
$\theta_3^{(1)}$	0.5	0.5166	0.0825
$\theta_4^{(1)}$	0	−0.0135	0.0694
$\theta_0^{(2)}$	0.5	0.5037	0.0479
$\theta_1^{(2)}$	0	0.0108	0.0696
$\theta_2^{(2)}$	−0.5	−0.4997	0.0668
$\theta_3^{(2)}$	0	−0.0042	0.0720
$\theta_4^{(2)}$	0	−0.0108	0.0758
$\theta_0^{(3)}$	0	0.0021	0.0509
$\theta_1^{(3)}$	0.5	0.5027	0.0738
$\theta_2^{(3)}$	0	0.0118	0.0713
$\theta_3^{(3)}$	−0.5	−0.4932	0.0680
$\theta_4^{(3)}$	−0.5	−0.5026	0.0773
$\theta_0^{(4)}$	0.5	0.4994	0.0439
$\theta_1^{(4)}$	0	−0.0166	0.0666
$\theta_2^{(4)}$	0	0.0056	0.0744
$\theta_3^{(4)}$	0	−0.0061	0.0583
$\theta_4^{(4)}$	0.5	0.5141	0.0729
σ_1^2	0.25	0.2458	0.0218
σ_2^2	0.25	0.2454	0.0220
σ_3^2	0.25	0.2439	0.0215
σ_4^2	0.25	0.2464	0.0223
r_1	0	−0.0011	0.0367
r_2	−0.5	−0.4843	0.0358
d_1	1	1	0
d_2	2	2	0

best subset selection for model (14), and the posterior probabilities of the three sets are 0.1414, 0.3549 and 0.4910, respectively. In addition, the second best model found by the selection scheme is close to the real model, and the respective posterior probabilities are 0.0004, 0.0001, 0.00002. Clearly, $(\Theta/\tau_{(j,m)}, c_{(j,m)}) = (0.1, 15)$ provides the best results.

In addition, the selected results of the 100 replicates of the model (14) simulation are recorded in Table 6. We calculate the proportion of the correct selection of the true model: the ratio 1 indicates the accuracy of selecting the real model as the best model, while ratio 2 means that true model is selected as either the best or the second best subset. These results show that our Bayesian subset selection scheme works well and $(\Theta/\tau_{(j,m)}, c_{(j,m)}) = (0.1, 15)$ is a good choice for the scheme.

From these two simulation experiments, we believe that the Bayesian estimation of unknown parameters in the DTMA model is satisfactory, the Bayesian subset selection method can separate the real model from other models, and $(\Theta/\tau_{(j,m)}, c_{(j,m)}) = (0.1, 15)$ is a good choice for this scheme.

5. ILLUSTRATIVE EXAMPLE

Interventions on foreign exchange market are widespread, and dynamics of exchange rate data has the characteristics of structural thresholds. We use the double-threshold variable moving average model to investigate the nonlinear characteristics of exchange rate of the Japanese Yen against the US dollar. We analyze this actual exchange rate using monthly data from January 1986 to December 2018, with a total of 396 observations. Ling and Tong (2005) and Xia et al. (2010) had analyzed such data sets, but their analysis were based on a single-threshold model. We break through the assumption of a single threshold variable, examine more comprehensively the trading information of the foreign exchange market, and establish a double-threshold moving average model. Liu (2007) used different linear and nonlinear time series models for the real historical data of the exchange rate of the RMB against the US dollar, and the nonlinear self-exciting threshold autoregressive model showed a good fitting effect. This paper will also use the self-excited threshold moving average model to analyze the historical data of the Yen against the US dollar, giving two threshold variables

Table 5. Subset selection results of model (14)

	(0.1, 5)	(0.1, 10)	(0.1, 15)
Best	$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$
Pos. Prob.	[0.1414]	[0.3549]	[0.4910]
Second Best	$\begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}$
Pos. Prob.	[0.0004]	[0.0001]	[0.00002]

Table 6. Proportion of correctly selecting the true model (14) with 100 replications

$(\Theta/\tau_{(j,m)}, c_{(j,m)})$	Proportion 1	Proportion 2
(0.1, 5)	96	100
(0.1, 10)	95	100
(0.1, 15)	96	100

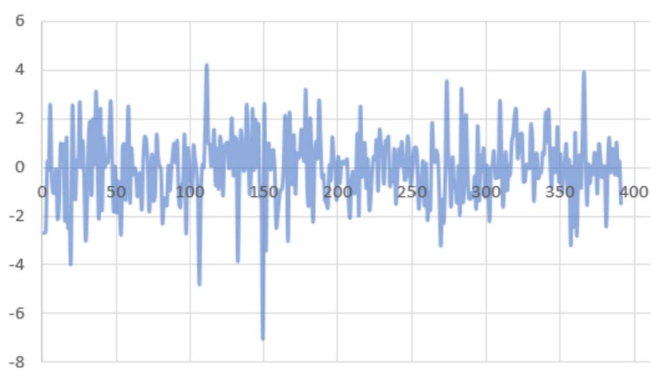

Figure 1. The y_t series from January 1986 to December 2018.

$z_{1,t-d_1}$ and $z_{2,t-d_2}$:

$$x_t = 100[\log(P_t) - \log(P_{(t-1)})],$$

$$y_t = x_t - \sum_{i=2}^{396} x_i/395 \ (t \geq 2),$$

$$z_{1,t-d_1} = y_{t-1}, z_{2,t-d_2} = y_{t-2},$$

where P_t represents the exchange rate of the Japanese dollar against the US dollar for the t-th month. Figure 1, Figure 2, Figure 3 show the logarithmic difference sequence of the yen-to-dollar monthly data from January 1986 to December 2018, and the time series diagram of two threshold variables, respectively.

For the empirical analysis, our hyperparameter selection is similar to that of Chen and Lee (1995) and Xia et al.

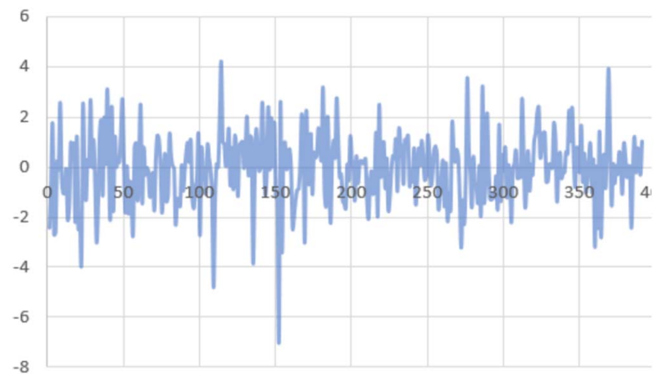

Figure 2. The $z_{1,t-d_1}$ series from January 1986 to December 2018.

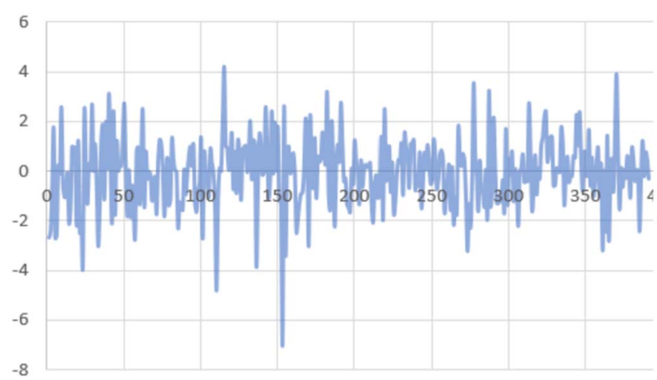

Figure 3. The $z_{2,t-d_2}$ series from January 1986 to December 2018.

(2010). We use the values of hyperparameters:

$$v_j = 3, \quad \lambda_j = 1, \quad \mathbf{V}_i = \mathbf{I},$$
$$d_{10} = d_{20} = 3, \quad a = p_5, \quad b = p_{95},$$

where p_k represents the k quantile of the sample data.

To construct the DTMA model with four mechanisms, we performed the MCMC iterations according to the sampling plan in Section 4. A total of 20,000 iterations were

Table 7. Coefficient estimation results of four-regime DTMA model

jth regime	$\theta_1^{(j)}$	$\theta_2^{(j)}$	$\theta_3^{(j)}$	$\theta_4^{(j)}$
1	−0.0748	0.0112	−0.0004	−0.0599
s.d.	0.0533	0.0554	0.0526	0.0634
c.i.	[−00977, −0.0518]	[−0.0122, 0.0335]	[−0.0232, 0.0224]	[−0.0850, −0.0349]
2	0.2352	−0.0844	0.0654	−0.0468
s.d.	0.0921	0.1046	0.0985	0.0797
c.i.	[0.2050, 0.653]	[−0.1166, −0.0523]	[0.0342, 0.0966]	−[0.0187, 0.0748]
3	0.2493	0.3640	0.1201	0.0204
s.d.	0.1009	0.1003	0.1072	0.0978
c.i.	[0.2177, 0.2809]	[0.3325, 0.3955]	[0.0876, 0.1527]	[−0.0107, 0.0515]
4	0.4737	0.6697	−0.0116	0.6704
s.d.	0.1661	0.1925	0.1809	0.2702
c.i.	[0.4332, 0.5142]	[0.6260, 0.7133]	[−0.0539, 0.0307]	[0.6187, 0.7221]

Table 8. Coefficient estimation results of four-regime DTMA model

Parameters	Posterior Mean	Posterior Standard Deviation	Confidence Interval
r_1	0.8466	0.0441	[0.8526, 0.8675]
r_2	0.5900	0.0684	[0.5641, 0.6160]

Table 9. AIC and BIC values for two models

Value	TMA	DTMA
AIC	644.81	639.80
BIC	735.05	732.44

performed and the first 15,000 iterations were discarded. The last 5000 iterations were retained to make the parameter estimation and optimal subset selection for the double-threshold moving average model. The estimation results and confidence intervals for each coefficient are recorded in Table 7. Table 8 shows the estimated values of the threshold parameters as well as the standard deviation and confidence intervals. Table 7 shows that the posterior standard deviation of the DTMA model of the Yen against the US dollar is relatively small, and the estimated values of parameters fall within their corresponding confidence intervals. According to the standard deviation of parameters and confidence intervals, some parameters in each mechanism are significant (the values with $\star$ in model (5.1) below mean that the parameters are significantly different from zero). Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for the DTMA model and the TMA model are calculated and compared with the TMA model in Ling and Tong (2005), and the results are recorded in Table 9. It can be seen that the AIC of the four-mechanism DTMA model is smaller than that of the TMA model. Moreover, the BIC of the four-mechanism DTMA model is also small. This indicates that the double-threshold moving average model with four mechanisms is more suitable for the Yen-Dollar exchange rate data.

Then the fitted four-regime DTMA model can be expressed as:

(15)

$$y_t = \begin{cases} -0.0748\varepsilon_{t-1}^{(1)} + 0.0112^*\varepsilon_{t-2}^{(1)} - 0.0004\varepsilon_{t-3}^{(1)} - 0.0599\varepsilon_{t-4}^{(1)} \\ +\varepsilon_t^{(1)}, \quad y_{t-1} \le 0.846^*, y_{t-2} \le 0.59^*, \\ 0.2352^*\varepsilon_{t-1}^{(2)} - 0.0844^*\varepsilon_{t-2}^{(2)} + 0.0654\varepsilon_{t-3}^{(2)} - 0.0468\varepsilon_{t-4}^{(2)} \\ +\varepsilon_t^{(2)}, \quad y_{t-1} \le 0.846^*, y_{t-2} > 0.59^*, \\ 0.2493^*\varepsilon_{t-1}^{(3)} + 0.3640\varepsilon_{t-2}^{(3)} + 0.1201\varepsilon_{t-3}^{(3)} + 0.0204\varepsilon_{t-4}^{(3)} \\ +\varepsilon_t^{(3)}, \quad y_{t-1} > 0.846^*, y_{t-2} \le 0.59^*, \\ 0.4737^*\varepsilon_{t-1}^{(4)} + 0.6697^*\varepsilon_{t-2}^{(4)} - 0.0116\varepsilon_{t-3}^{(4)} + 0.6704^*\varepsilon_{t-4}^{(4)} \\ +\varepsilon_t^{(4)}, \quad y_{t-1} > 0.846^*, y_{t-2} > 0.59^*. \end{cases}$$

Finally, the stationary convergence of Bayesian estimation based on the iterative traces of each parameter was tested. Figure 4 and Figure 5 show the iterative trajectory of the last 5000 MCMC sampling results. The frequency histograms of the posterior means of parameters are shown in Figure 6 and Figure 7.

Through the trace plot of each parameter sampling process, it is found that the trace of each parameter fluctuates and is stable above and below the estimated value, indicating that the sampling process is convergent. The posterior mean frequency histograms in Figure 6 and Figure 7 show that the parameter distribution of the four mechanisms is almost symmetrical.

Regarding the selection of subsets, Table 10 shows the best subset model and the second best subset model with the values of (0.1, 5), (0.1, 10) and (0.1, 15). Although dif-

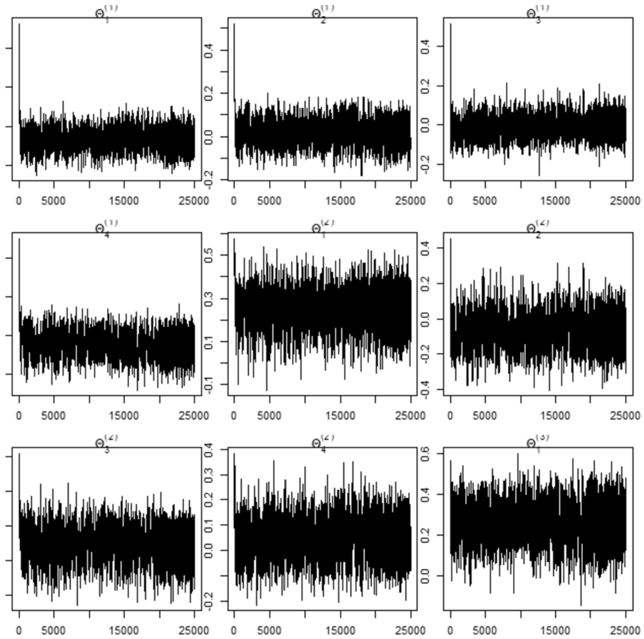

Figure 4. Trace plots of the last 5000 MCMC iterations of all estimate parameters.

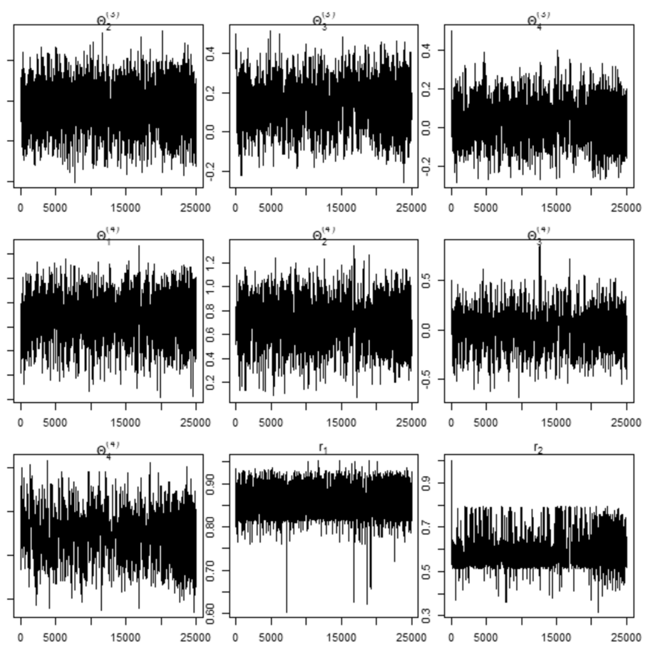

Figure 5. Trace plots of the last 5000 MCMC iterations of all estimate parameters.

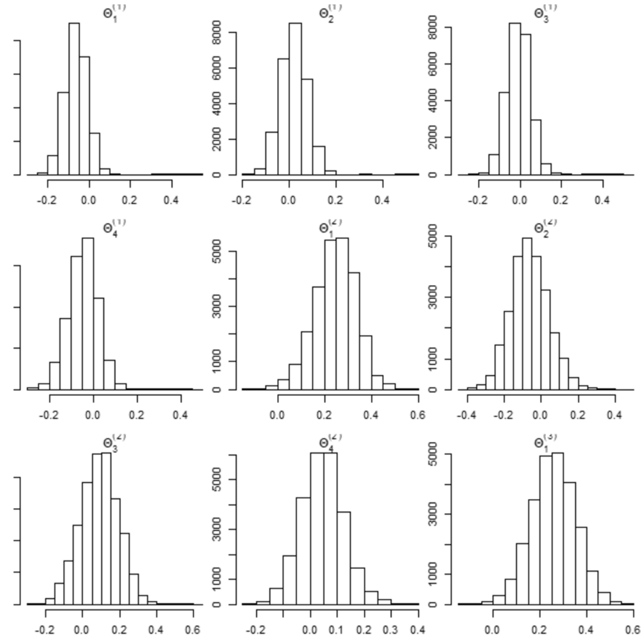

Figure 6. A posteriori mean frequency histogram of each parameter of the model (15).

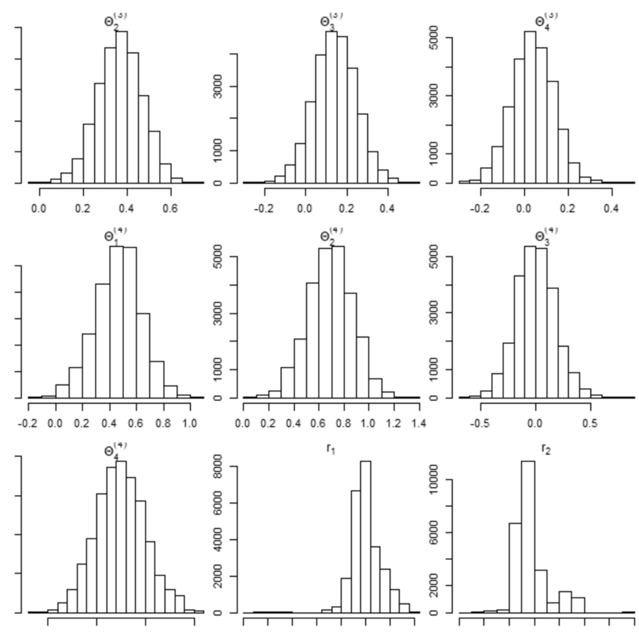

Figure 7. A posteriori mean frequency histogram of each parameter in the model (15).

ferent $(\Theta/\tau_{(j,m)}, c_{(j,m)})$ lead to the same selection of the best subset model, they have different posterior probabilities, $(\Theta/\tau_{(j,m)}, c_{(j,m)}) = (0.1, 15)$ has the highest posterior probability. For the second best subset, (0.1, 10) and (0.1, 15) select the same subset, and their difference is only reflected in the higher posterior probability of (0.1, 15). However, (0.1, 5) select more parameters in the third mechanism than the other two groups. Therefore, under the result of comprehensive comparison of the three sets of subsets, we tend to choose the best subset of the three groups, and

Table 10. Optimal subset results under different values of $(\Theta/\tau_{(j,m)}, c_{(j,m)})$ in empirical analysis

	(0.1, 5)	(0.1, 10)	(0.1, 15)
Best	$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$
Pos. Prob.	[0.0658]	[0.1281]	[0.1351]
Second Best	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$
Pos. Prob.	[0.00049]	[0.00051]	[0.00052]

by comparing the three posterior probabilities, we finally choose the highest probability of posterior (0.1, 15). In summary, the optimal subset model considered in this section is:

$$(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4)' = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}.$$

6. CONCLUSION

With the advent of the big data age, more and more time series in practical applications require complex nonlinear models, and the single-threshold MA model is not applicable. Therefore, this paper proposes a two-threshold moving average model, which uses the SSVS method to select the optimal subset. Due to the uncertainty of the two threshold variables and the two delay parameters in the DTMA model, we have conducted a more in-depth study on this model from the perspective of Bayesian estimation, and extended the optimal subset selection method to the two-threshold moving average case. Two simulation experiments verify the accuracy and effectiveness of the proposed method for parameter estimation and optimal subset selection. Finally, the empirical analysis of the Yen-Dollar exchange rate data shows the best subset selection method of the DTMA model is ideal and better than the TMA model.

ACKNOWLEDGEMENTS

This work was partially supported by the Major Research Plan of the National Natural Science Foundation of China (91746102), Ministry of Education in China Project of Humanities and Social Sciences (No. 17YJA910002), and the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science of East China Normal University, Ministry of Education.

Received 29 December 2019

REFERENCES

[1] Brooks, C. and Garrett, I. (2002). Can we explain the dynamics of the UK FTSE 100 stock and stock index futures markets. *Appl. Fin. Econ.* **12**, 25–31.

[2] Chan, K. S., Petruccelli, J. D., Tong, H. and Woolford, S. W. (1985). A multiple-threshold AR(1) model. *J. Appl. Probab.* **22**, 267–279. MR0789351

[3] Chan, K. S. and Tong, H. (1985). On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. Appl. Probab.* **17**, 666–678. MR0798881

[4] Chen, H. Q., Chong, T. T. L., Bai, J. S. (2012). Theory and applications of TAR model with two threshold variables. *Econometric Reviews* **31**(2), 142–170. MR2855216

[5] Chen, C. W. S. and Lee, J. C. (1995). Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis* **16**, 483–492. MR1365643

[6] Chen, C. W. S., Liu, F. C., Gerlach, R. (2011). Bayesian subset selection for threshold autoregressive moving-average models. *Computat. Statist.* **26**, 1–30. MR2773798

[7] Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician* **49**, 327–335.

[8] George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association* **88**, 881–889.

[9] 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.

[10] Hastings, W. K. (1970). Monte-Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109. MR3363437

[11] Ismail, M. A. and Charif, H. A. (2003). Bayesian inference for threshold moving average models. *Metron* **1**, 119–132. MR1994114

[12] Leeper, E. M. (1991). Equilibria under active and passive monetary and fiscal policies. *Journal of Monetery Economics* **27**, 129–147.

[13] Li, D. (2012). A note on moving-average models with feedback. *Journal of Time Series Analysis* **33**, 873–879. MR2991905

[14] Li, D., Ling, S. and Tong, H. (2012). On moving-average models with feedback. *Bernoulli* **18**, 735–745. MR2922468

[15] Ling, S. and Tong, H. (2005). Testing a linear moving-average model against threshold moving-average models. *The Annals of Statistics* **33**, 2529–2552.

[16] Ling, S., Tong, H., Li, D. (2007). Ergodicity and invertibility of threshold moving-average models. *Bernoulli* **13**, 161–168. MR2307400

[17] Liu, J. and Susko, E. (1992). On strict stationarity and ergodicity of a nonlinear ARMA model. *J. Appl. Probab.* **29**, 363–373. MR1165221

[18] Liu, T. (2007). Study on nonlinear characteristics of RMB's real exchange Rrates. *The Journal of Quantitative & Technical Economics* **2**, 11–18.

[19] Mcculloch, R. E. and Tsay, R. S. (1993). 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, pp. 253–262.

[20] 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.

[21] Ni, S., Xia, Q., Liu, J. (2018). Bayesian subset selection for two-threshold variable autoregressive models. *Studies in Nonlinear Dynamics and Econometrics* **22**(4),1–16. MR3858511

[22] Pan, J., Xia, Q., Liu, J. (2017). Bayesian analysis of multiple thresholds autoregressive model. *Computational Statistics* **32**, 219–237. MR3604215

[23] Sáfadi, T. and Morettin, P. A. (2000). Bayesian analysis of threshold autoregressive moving average models. *Sankýa* **62**, 353–371. MR1834161

[24] So, M. K. P. and Chen, C. W. S. (2003). Subset threshold autoregression. *Journal of Forecasting* **22**, 49–66.

[25] Tsay, R. S. (2005). *Analysis of Financial Time Series*, 2nd ed. New York: John Wiley and Sons. MR2162112

[26] Tiao, G. C. and Tsay, R. (1994). Some advances in nolinear and adaptive modeling in time-series. *Journal of Forecasting* **13**, 109–131.

[27] Tong, H. (1978). On a threshold model. In *Pattern Recognition and Signal Processing*, (ed. C. H. Chen). Amsterdam: Sijthoff and Noordhoff, pp. 101–141.

[28] Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press. MR1079320

[29] Tong, H. and Lim, K. S. (1980). Threshold autoregressions, limit cycles, and data. *Journal of the Royal Statistical Society, Series B* **42**, 245–292.

[30] Tong, H. (1990). *Non-linear Time Series. A Dynamical System Approach*. Oxford: Clarendon Press. MR1079320

[31] Xia, Q., Pan, J. Z., Zhang, Z. Q. and Liu, J. S. (2010). A Bayesian nonlinearity test for threshold moving average models. *J. Time Ser. Anal.* **31**, 329–336. MR2724513

Jinshan Liu
School of Financial Mathematics & Statistics
Guangdong University of Finance
Guangzhou, 510521
China
E-mail address: liujs58@163.com

Jiazhu Pan
Department of Mathematics & Statistics
University of Strathclyde
Glasgow G1 1XH
UK
E-mail address: jiazhu.pan@strath.ac.uk

Qiang Xia
Department of Applied Mathematics
South China Agricultural University
Guangzhou, 510642
China
E-mail address: xiaqiang@scau.edu.cn

Ying Xiao
Department of Applied Mathematics
South China Agricultural University
Guangzhou, 510642
China
E-mail address: 1511790083@qq.com