

## Subset Threshold Autoregression

MIKE K. P. SO<sup>1\*</sup> AND CATHY W. S. CHEN<sup>2</sup>

<sup>1</sup> *Hong Kong University of Science and Technology*

<sup>2</sup> *Feng-Chia University, Taiwan*

### ABSTRACT

We develop in this paper an efficient way to select the best subset threshold autoregressive model. The proposed method uses a stochastic search idea. Differing from most conventional approaches, our method does not require us to fix the delay or the threshold parameters in advance. By adopting the Markov chain Monte Carlo techniques, we can identify the best subset model from a very large of number of possible models, and at the same time estimate the unknown parameters. A simulation experiment shows that the method is very effective. In its application to the US unemployment rate, the stochastic search method successfully selects lag one as the time delay and five best models from more than 4000 choices. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS arranged autoregression; hierarchical Bayes; Markov chain Monte Carlo; stochastic search variable selection; Metropolis algorithm; unemployment rate

### INTRODUCTION

Non-linear time series modelling has attracted much attention in recent years. The threshold autoregressive (TAR) model proposed by Tong (1978, 1983) and Tong and Lim (1980) is one of the popular non-linear time series models that shows wide application in many areas. An important special case of TAR is the self-exciting threshold autoregressive (SETAR) model. Given the maximum autoregressive (AR) orders  $p_1$  and  $p_2$ , the SETAR(2 :  $p_1$ ;  $p_2$ ) model is stated as

$$y_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} y_{t-1} + \phi_2^{(1)} y_{t-2} + \cdots + \phi_{p_1}^{(1)} y_{t-p_1} + a_t^{(1)} & \text{if } y_{t-d} \leq r \\ \phi_0^{(2)} + \phi_1^{(2)} y_{t-1} + \phi_2^{(2)} y_{t-2} + \cdots + \phi_{p_2}^{(2)} y_{t-p_2} + a_t^{(2)} & \text{if } y_{t-d} > r \end{cases} \quad (1)$$

where  $r$  is the threshold parameter,  $d$  is commonly referred to the delay (or threshold lag) of the model,  $a_t^{(1)}$  and  $a_t^{(2)}$  are zero mean white noise with variances  $\sigma_j^2$ ,  $j = 1, 2$ . The SETAR model has

---

\* Correspondence to: Mike K. P. So, Department of Information and Systems Management, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong. E-mail: immkps@ust.hk  
Contract/grant sponsor: National Science Council of Taiwan; Contract/grant number: NSC90-2118-M-035-008.  
Contract/grant sponsor: Hong Kong RGC; Contract/grant numbers: 00/01.BM25; 01/02.BM38.

two different AR processes in the two regimes defined by the threshold variable  $y_{t-d}$  and  $r$ . One important issue in fitting the SETAR model is to select the best subset model. In other words, we need an efficient way to identify values of  $d$  and subsets of  $\phi_k^{(i)}$ ,  $i = 1, 2$  and  $k = 0, \dots, p_i$  that are 'important', given the maximum AR orders of  $p_1$  and  $p_2$  for the two regimes. The identification problem can be highly complicated because it involves a very large number of possible models. Taking a simple example of  $p_1 = p_2 = 3$  and the maximum delay of 4, the total possible models is  $4 \times 2^{3+3+2} = 1024$  which is already very large. In general, we have  $d_0 \times 2^{p_1+p_2+2}$  models to consider, where  $d_0$  is the maximum possible delay.

The orders of SETAR models are commonly identified by considering some information criteria such as the Akaike information criterion (AIC). For each possible delay parameter, Tong and Lim (1980) use the AIC to estimate  $r$  and find suitable autoregressive orders in both regimes of the threshold model. The delay parameter is then determined by minimizing the AIC. However, their method considers a selected number of  $r$  only and no best subset model is determined. Tsay (1989) proposes methods for building a SETAR model by using arranged autoregression. As in Tong and Lim (1980), the delay and threshold parameters have to be estimated prior to finding the autoregressive orders. Thanoon (1990) proposes a two-stage procedure which fits a full SETAR(2 :  $p_1; p_2$ ) model in the first stage to obtain preliminary estimates of  $r$  and  $d$ . Then, a method of finding the best subset autoregressive model is applied to the two regimes defined by the estimates of  $r$  and  $d$ . Overall, one major difficulty in identifying the best subset SETAR model is the specification of the threshold and delay parameters, especially the latter. When the threshold or delay parameters are changed, the best subset of SETAR derived from existing methods can be substantially different. Even worse is that in typical real applications, the number of models under consideration, i.e.  $d_0 \times 2^{p_1+p_2+2}$ , is large enough to prevent us from comparing all possible models directly using traditional methods, such as the AIC.

In this paper, we develop a selection method for SETAR models that can simulate the unknown parameters, including  $r$  and  $d$ , and identify the best subset SETAR model simultaneously. We formulate the identification problem in a Bayesian framework and adopt the Markov chain Monte Carlo (MCMC) methods. Previous applications of the MCMC methods to time series models can be found in Chen (1992, 1999), Chen and Lee (1995), Geweke and Terui (1993), McCulloch and Tsay (1993, 1994a, b). In particular, Chen and Lee (1995) adopt the idea of Carlin *et al.* (1992) to transform a SETAR model into a changepoint problem in linear regression via arranged autoregression and then make inference for all the parameters via a Bayesian approach. This paper generalizes the work of Chen and Lee (1995) and utilizes the stochastic search variable selection (SSVS) method proposed by George and McCulloch (1993) to select important subsets of SETAR models for further consideration. Two-component mixture of normal priors are assumed for  $\phi_k^{(i)}$ ,  $k = 0, \dots, p_i$ ,  $i = 1, 2$ . The binary indicator variables that indicate which component of the normal mixtures is used to determine whether the corresponding  $\phi_k^{(i)}$  are included in the SETAR models. By applying the MCMC methods to generate a posterior sample of the indicator variables, we can identify the best subset SETAR model by referring to those with the highest posterior probability. In this way, our method can select important subsets of SETAR models as well as making inference for the unknown parameters including  $r$  and  $d$ .

The structure of the remaining sections is as follows. The next section gives the Bayesian set-up for the threshold autoregressive time series models. The idea of the SSVS and the prior distributions for the unknown parameters are described. In the third section we show how the MCMC methods are implemented to generate posterior samples of the indicator variables for identifying the best subset SETAR model. In the fourth section we study the performance of our methodology using

simulated data. The results obtained from the simulation experiments tell us that our method can effectively pick up the true model as the best subset. In the fifth section we apply our method to the US unemployment rate where the time delay of lag one and five best models are successfully selected from more than four thousand choices. The final section presents some concluding remarks.

### THE BAYESIAN SET-UP FOR SETAR MODELS

#### The likelihood

Consider the SETAR(2 :  $p_1; p_2$ ) process in (1). Let  $p = \max\{p_1, p_2\}$ . We assume that the first  $p$  observations  $y_1, y_2, \dots, y_p$  are fixed. Let  $\pi_i$  be the time index of the  $i$ th smallest observation of  $\{y_{p+1-d}, y_{p+2-d}, \dots, y_{n-d}\}$ . By conditioning on the first  $p$  observations, we can write the likelihood function as

$$L(\Theta_1, \Theta_2, \sigma_1^2, \sigma_2^2, r, d | \mathbf{Y}) \propto \sigma_1^{-s} \sigma_2^{-(n-p-s)} \times \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^s (y_{\pi_i+d} - \phi_0^{(1)} - \sum_{k=1}^{p_1} \phi_k^{(1)} y_{\pi_i+d-k})^2 - \frac{1}{2\sigma_2^2} \sum_{i=s+1}^{n-p} (y_{\pi_i+d} - \phi_0^{(2)} - \sum_{k=1}^{p_2} \phi_k^{(2)} y_{\pi_i+d-k})^2 \right\}$$

where  $s$  satisfies  $y_{\pi_s} \leq r < y_{\pi_{s+1}}$ ,  $\mathbf{Y} = (y_{\pi_1+d}, y_{\pi_2+d}, \dots, y_{\pi_{n-p+d}})'$ ,  $\Theta_1 = (\phi_0^{(1)}, \phi_1^{(1)}, \dots, \phi_{p_1}^{(1)})'$ , and  $\Theta_2 = (\phi_0^{(2)}, \phi_1^{(2)}, \dots, \phi_{p_2}^{(2)})'$ . The parameters of the SETAR(2 :  $p_1; p_2$ ) model to be estimated are  $\Theta_1, \Theta_2, \sigma_1^2, \sigma_2^2, r$ , and  $d$ . Let  $Y_1^* = (y_{\pi_1+d}, y_{\pi_2+d}, \dots, y_{\pi_s+d})'$  be the observations in regime I and  $Y_2^* = (y_{\pi_{s+1}+d}, \dots, y_{\pi_{n-p}+d})'$  be those in regime II. So  $Y_1^*$  and  $Y_2^*$  are vectors of data arranged in the ascending order. Define  $X_1^* = (x_{1,\pi_1+d}, x_{1,\pi_2+d}, \dots, x_{1,\pi_s+d})'$ ,  $X_2^* = (x_{2,\pi_{s+1}+d}, \dots, x_{2,\pi_{n-p}+d})'$ , where  $x_{1,t} = (1, y_{\pi_t+d-1}, \dots, y_{\pi_t+d-p_1})'$  and  $x_{2,t} = (1, y_{\pi_t+d-1}, \dots, y_{\pi_t+d-p_2})'$ . An arranged autoregression is formed with the first  $s$  cases of  $\mathbf{Y}$  in the first regime and the last  $n - p - s$  cases in the second regime. Using the arranged autoregression, the likelihood function is simplified to

$$L(\Theta_1, \Theta_2, \sigma_1^2, \sigma_2^2, r, d | \mathbf{Y}) \propto \exp \left\{ -\sum_{i=1}^2 \frac{1}{2\sigma_i^2} (\mathbf{Y}_i^* - X_i^{*'} \Theta_i)' (\mathbf{Y}_i^* - X_i^{*'} \Theta_i) \right\} \sigma_1^{-s} \sigma_2^{-(n-p-s)}$$

#### A mixture specification

The primary objective of this paper is to identify the best subset SETAR model. Based on the idea of George and McCulloch (1993), we introduce the binary indicator variables  $\delta_{i,k}$ ,  $i = 1, 2$  and  $k = 0, \dots, p_i$ , which can take values of either 0 or 1. Each value of  $\delta_{i,k}$  determines the distribution of  $\phi_k^{(i)}$ . When  $\delta_{i,k} = 0$ ,  $\phi_k^{(i)}$  has the mean 0 and variance  $\tau_{i,k}^2$ . The variance of  $\phi_k^{(i)}$  will be scaled by  $c_{i,k}^2$  if  $\delta_{i,k} = 1$ . We represent our prior assumptions on a single  $\phi_k^{(i)}$  by the normal mixture distribution,

$$\phi_k^{(i)} | \delta_{i,k} \sim (1 - \delta_{i,k})N(0, \tau_{i,k}^2) + \delta_{i,k}N(0, c_{i,k}^2 \tau_{i,k}^2) \tag{2}$$

and

$$\delta_{i,k} = \begin{cases} 1 & \text{with probability } \gamma_{i,k} \\ 0 & \text{with probability } 1 - \gamma_{i,k} \end{cases} \tag{3}$$

where  $\gamma_{i,k}$  are the prior probabilities of having  $\delta_{i,k} = 1$ . In the general specification which allows correlations among  $\phi_k^{(i)}$ , the mixture in (2) can be stated as the following multivariate normal prior for the slope parameters  $\Theta_i$  in regime  $i$ ,

$$\Theta_i | \delta_i \sim N(\mathbf{0}, \mathbf{D}_{\delta_i} \mathbf{V}_i \mathbf{D}_{\delta_i})$$

where  $\delta_i = (\delta_{i,0}, \delta_{i,1}, \dots, \delta_{i,p_i})'$ ,  $\mathbf{V}_i$  is the prior correlation matrix and  $\mathbf{D}_{\delta_i}$  is the diagonal matrix  $\text{diag}\{a_{i,0}\tau_{i,0}, \dots, a_{i,p_i}\tau_{i,p_i}\}$  with  $a_{i,k} = 1$  if  $\delta_{i,k} = 0$  and  $a_{i,k} = c_{i,k}$  if  $\delta_{i,k} = 1$ . In particular when we do not have any prior information about the relationship among  $\phi_k^{(i)}$  that  $\mathbf{V}_i = \mathbf{I}$ , the covariance  $\mathbf{D}_{\delta_i} \mathbf{V}_i \mathbf{D}_{\delta_i}$  reduces to a diagonal matrix with elements  $a_{i,k}^2 \tau_{i,k}^2$ ,  $k = 0, \dots, p_i$ . In other words,

$$p(\Theta_i | \delta_i) = \prod_{k=0}^{p_i} p(\phi_k^{(i)} | \delta_{i,k})$$

We choose  $\tau_{i,0}, \dots, \tau_{i,p_i}$  to be small so that those  $\phi_k^{(i)}$  associated with  $\delta_{i,k} = 0$  are likely to be small in magnitude. We choose  $c_{i,0}, \dots, c_{i,p_i}$  to be large and greater than 1 to make  $c_{i,k}^2 \tau_{i,k}^2$  substantially greater than  $\tau_{i,k}^2$ . Then, those  $\phi_k^{(i)}$  associated with  $\delta_{i,k} = 1$  will have high variability. The above set-up fits well to our model selection framework. Variables associated with  $\delta_{i,k} = 1$  are regarded as 'useful' because the corresponding  $\phi_k^{(i)}$  are likely to be away from zero. In contrast, the variables having zero  $\delta_{i,k}$  are taken to be unimportant. Once we have created a posterior sample for  $\delta_i$ , we can identify a good model by referring to a particular combination of  $d$  and  $\delta_i$  which has a high posterior probability. For example for  $p_1 = p_2 = 2$ , the model

$$y_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} y_{t-1} + \phi_2^{(1)} y_{t-2} + a_t^{(1)} & \text{if } y_{t-1} \leq r \\ \phi_0^{(2)} + \phi_1^{(2)} y_{t-1} + a_t^{(2)} & \text{if } y_{t-1} > r \end{cases}$$

is selected as a good model if the posterior probability for the event  $\{d = 1, \delta_{1,0} = 1, \delta_{1,1} = 1, \delta_{1,2} = 1, \delta_{2,0} = 1, \delta_{2,1} = 1, \delta_{2,2} = 0\}$  is high. To implement the above procedures, suitable choices of  $\tau_{i,k}$  and  $c_{i,k}$  are required. We consider setting  $(\sigma_{\phi_k^{(i)}} / \tau_{i,k}, c_{i,k})$  as (0.5, 5), (0.5, 10), (1, 5) and (1, 10) where  $\sigma_{\phi_k^{(i)}}$  represents a measure of variation of  $\phi_k^{(i)}$ . Note that the last two sets of values were suggested by George and McCulloch (1993) while the first two sets are found to have superior performance from our simulation studies later in the paper. In our subset threshold autoregression, we can set  $\sigma_{\phi_k^{(i)}}$  as the standard error of the  $k$ th slope coefficients in fitting an ordinary AR process.

**Prior distributions**

To facilitate the sampling of  $\Theta_i, \sigma_i^2, \delta_i, r$  and  $d$  by using Monte Carlo Markov chain methods, we need to set their prior distributions. In this paper, we employ the inverse gamma conjugate prior for  $\sigma_i^2$ ,

$$\sigma_i^2 | \delta_i \sim IG(v_i/2, v_i \lambda_i / 2)$$

i.e.  $\frac{v_i \lambda_i}{\sigma_i^2} \sim \chi_{v_i}^2$

Note that in the above inverse gamma distribution,  $v_i$  and  $\lambda_i$  are constants. Nevertheless, if we want to allow prior dependence between  $\Theta_i$  and  $\sigma_i^2$ , we can also let the hyperparameters  $v_i$  and  $\lambda_i$

depend on  $\delta_i$ . For the threshold value  $r$ , we assume that it follows a uniform distribution on  $(a, b)$ . We further let the delay parameter  $d$  have a discrete uniform distribution on integers  $1, \dots, d_0$ , recalling that  $d_0$  is the maximum possible delay. Since it rarely happens that  $d$  is greater than 2, we set  $d_0 = 4$  in our study. Finally, we assume prior independence of  $\delta_{i,k}$ 's with the marginal distributions given in (3). Therefore,

$$p(\delta_i) = \prod_{k=0}^{p_i} \gamma_{i,k}^{\delta_{i,k}} (1 - \gamma_{i,k})^{1-\delta_{i,k}}$$

$i = 1, 2$ . In particular, we can have the flat prior  $p(\delta_i) = 2^{-(p_i+1)}$ , in which each lag  $k$  has an equal chance ( $\gamma_{i,k} = 1/2$ ) of being included.

### POSTERIOR DISTRIBUTIONS

To calculate the posterior probabilities of all  $d_0 \times 2^{p_1+p_2+2}$  subsets, SSVS uses MCMC to generate a sequence of

$$\delta_{i,0}, \delta_{i,1}, \dots, \delta_{i,p_i}$$

Under some regularity conditions (see for example, Tierney, 1994), the MCMC sequences of  $\delta_i$  converge in distribution to  $p(\delta_i|\mathbf{Y})$ . From the posterior sample of  $\delta_i$ , we can estimate the posterior probability of a particular model. Important subsets of SETAR models can then be identified via the MCMC sequences with the corresponding  $\delta_i$ 's occurring with high probabilities. In implementing the MCMC method,  $\Theta_i$  and  $\sigma_i^2$  are initialized at their least squares estimates when fitting an autoregressive process while  $\delta_i$  is initialized as  $(1, 1, \dots, 1)'$ . The MCMC iterates of  $\Theta_1, \Theta_2, \sigma_1^2, \sigma_2^2, r, d, \delta_1$ , and  $\delta_2$  are then generated from suitable full conditional distributions. Using standard Bayesian techniques (e.g. DeGroot, 1970 or Box and Tiao, 1973), we obtain the following full conditional distributions:

- (1) The conditional posterior distribution of  $\Theta_i$  is independent of  $\Theta_j$  for  $i \neq j$ ,

$$p(\Theta_i|\mathbf{Y}, \sigma_i^2, \delta_i, r, d) \sim N(\Theta_i^*, \mathbf{V}_i^*) \tag{4}$$

where

$$\Theta_i^* = \frac{\mathbf{V}_i^* X_i^{*'} X_i^* \hat{\Theta}_i}{\sigma_i^2} \quad \text{and}$$

$$\mathbf{V}_i^* = \left( \frac{X_i^{*'} X_i^*}{\sigma_i^2} + \mathbf{D}_{\delta_i}^{-1} \mathbf{V}_i^{-1} \mathbf{D}_{\delta_i}^{-1} \right)^{-1}$$

with  $\hat{\Theta}_i = (X_i^{*'} X_i^*)^{-1} X_i^{*'} Y_i^*$ .

- (2) The conditional posterior distribution of  $\sigma_i^2$  is independent of  $\sigma_j^2$ , for  $i \neq j$ ,

$$p(\sigma_i^2|\mathbf{Y}, \Theta_i, \delta_i, r, d) \sim IG \left( \frac{v_i + n_i}{2}, \frac{v_i \lambda_i + n_i s_i^2}{2} \right),$$

i.e.  $\frac{v_i \lambda_i + n_i s_i^2}{\sigma_i^2} \sim \chi_{v_i+n_i}^2, \quad i = 1, 2$  (5)

where

$$n_1 = \sum_{i=1}^{n-p} I_{\{y_{\pi_i} \leq r\}}, \quad n_2 = \sum_{i=1}^{n-p} I_{\{y_{\pi_i} > r\}}, \quad s_i^2 = n_i^{-1} (Y_i^* - \hat{Y}_i)' (Y_i^* - \hat{Y}_i), \quad \text{with } \hat{Y}_i = X_i^{*'} \Theta_i.$$

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

$$p(r|\mathbf{Y}, d, \Theta_i, \sigma_i^2, \delta_i, i = 1, 2) \propto \exp \left\{ - \sum_{i=1}^2 \frac{1}{2\sigma_i^2} (\mathbf{Y}_i^* - X_i^{*'} \Theta_i)' (\mathbf{Y}_i^* - X_i^{*'} \Theta_i) \right\} / \sigma_1^{n_1} \sigma_2^{n_2} \quad (6)$$

Note that  $n_1$  and  $n_2$  are functions of  $r$ .

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

$$p(d|\mathbf{Y}, r, \Theta_i, \sigma_i^2, \delta_i, i = 1, 2) = L(r, d, \Theta_i, \sigma_i^2, \delta_i, i = 1, 2|\mathbf{Y}) / \sum_{j=1}^{d_0} L(r, j, \Theta_i, \sigma_i^2, \delta_i, i = 1, 2|\mathbf{Y})$$

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

$$L(r, d, \Theta_i, \sigma_i^2, \delta_i, i = 1, 2|\mathbf{Y}) = \exp \left\{ - \sum_{i=1}^2 \frac{1}{2\sigma_i^2} (\mathbf{Y}_i^* - X_i^{*'} \Theta_i)' (\mathbf{Y}_i^* - X_i^{*'} \Theta_i) \right\} / \sigma_1^{n_1} \sigma_2^{n_2}$$

(5) Finally, the vector  $\delta_i$  is obtained by sampling  $\delta_{i,k}$  individually from the conditional distribution,

$$\delta_{i,k} \sim p(\delta_{i,k}|\mathbf{Y}, \Theta_i, \sigma_i^2, \delta_{i,(-k)}, r, d)$$

where  $\delta_{i,(-k)} = (\delta_{i,0}, \dots, \delta_{i,k-1}, \delta_{i,k+1}, \dots, \delta_{i,p_i})'$ . The conditional posterior probability function of  $\delta_{i,k}$  is a Bernoulli distribution with the probability

$$P(\delta_{i,k} = 1|\mathbf{Y}, \Theta_i, \sigma_i^2, \delta_{i,(-k)}, r, d) = \frac{\alpha_{i,k}}{\alpha_{i,k} + \beta_{i,k}} \quad (7)$$

where

$$\alpha_{i,k} = p(\phi_i^{(k)}|\delta_{i,(-k)}, \delta_{i,k} = 1)\gamma_{i,k} \quad \text{and} \\ \beta_{i,k} = p(\phi_i^{(k)}|\delta_{i,(-k)}, \delta_{i,k} = 0)(1 - \gamma_{i,k})$$

In particular if  $\mathbf{V}_i = \mathbf{I}$  and the uniform prior is specified for  $\delta_{i,k}$ , i.e.  $\gamma_{i,k} = 0.5$ , then  $\alpha_{i,k}$  and  $\beta_{i,k}$  are simplified to  $p(\phi_i^{(k)}|\delta_{i,k} = 1)$  and  $p(\phi_i^{(k)}|\delta_{i,k} = 0)$  respectively.

All conditional distributions except that of  $r$  are standard and easy to sample. For simulating  $r$ , we employ the Metropolis *et al.* (1953) algorithm. A symmetric random walk kernel is adopted to obtain a trial point at each iteration. Let  $f(r)$  be the conditional density in (6), suppressing the conditioning variables for notational simplicity. The resulting random walk Metropolis algorithm works as follows:

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

$$r^* = r^{(j-1)} + \varepsilon, \quad \varepsilon \sim N(0, a^2)$$

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

Step 2: Accept  $r^*$  as  $r^{(j)}$  with the probability  $p = \min\{1, f(r^*)/f(r^{(j-1)})\}$ . Otherwise, set  $r^{(j)} = r^{(j-1)}$ .

The positive scalar  $a$  controls the step size in the Metropolis move. In practice, we can tune the step size  $a$  to yield fast convergence of the MCMC. The larger the value of  $a$ , the smaller is the acceptance rate. A suitable value of  $a$  with good convergence properties can usually be achieved by having an acceptance rate of 25–50%. In summary, we use the following iterative sampling scheme to construct the desired posterior sample:

- (1) Draw  $\Theta_i$  from the multivariate normal distribution in (4).
- (2) Draw  $\sigma_i^2$  from the inverse Gamma distribution in (5).
- (3) Draw  $r$  using the random walk Metropolis algorithm.
- (4) Draw  $d$  from a multinomial distribution with probabilities proportional to the likelihood function.
- (5) Draw  $\delta_{i,k}$  from the Bernoulli distribution with the probability given in (7).

This completes one iteration. Of course, we can change the order in sampling the variables. Our experience tells us that fast convergence is attained irrespective of the order.

### SIMULATIONS

In this section, we perform simulation experiments to investigate the performance of the subset selection method for SETAR models. The following hyperparameters are set, namely  $\mathbf{V}_i = \mathbf{I}$ ,  $v_i = 3$  and  $\lambda_i = \tilde{\sigma}^2/3$  for  $i = 1, 2$ , where  $\tilde{\sigma}^2$  is the sample variance. The parameters  $\tau_{i,k}$  and  $c_{i,k}$  are constants independent of  $i$  and  $k$ . Uniform priors are adopted for  $\delta_i$ . In the first experiment, we simulate  $n = 1000$  observations from

$$\text{Model 1} \quad y_t = \begin{cases} -0.5 y_{t-1} + a_t^{(1)}, & a_t^{(1)} \sim N(0, 2) \quad \text{if } y_{t-1} \leq 0.4 \\ 0.5 y_{t-1} + a_t^{(2)}, & a_t^{(2)} \sim N(0, 1) \quad \text{if } y_{t-1} > 0.4 \end{cases}$$

This is the model used in the simulation study of Chen and Lee (1995). We run the MCMC for 10,000 iterations and discard the first 4000 burn-in iterates. To assess for the convergence to stationarity, we produce autocorrelation plots for the last 6000 MCMC iterates in Figures 1 and 2. The autocorrelations of the MCMC iterates decay very quickly, indicating fast convergence. The only variable that is of concern is  $r$  but it only exhibits moderate autocorrelations in the first 20 lags. In summary, we are very confident that the number of burn-in iterations are large enough to arrive at stationarity and so we keep this number together with the total iterations of 10,000 fixed throughout the simulation study.

As  $\tau_{i,k}$  and  $c_{i,k}$  are important in the subset selection, we focus on the choices of  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k})$  in the simulation. We set the maximum AR orders as 4 in both regimes, i.e.  $p_1 = p_2 = 4$ , and so there are altogether  $4 \times 2^{4+4+2} = 4096$  possible models. We choose  $\sigma_{\phi_k^{(i)}}$ 's to be the standard errors of the least square estimates of  $\phi_k^{(i)}$  obtained from fitting the AR(4) model to the simulated data.

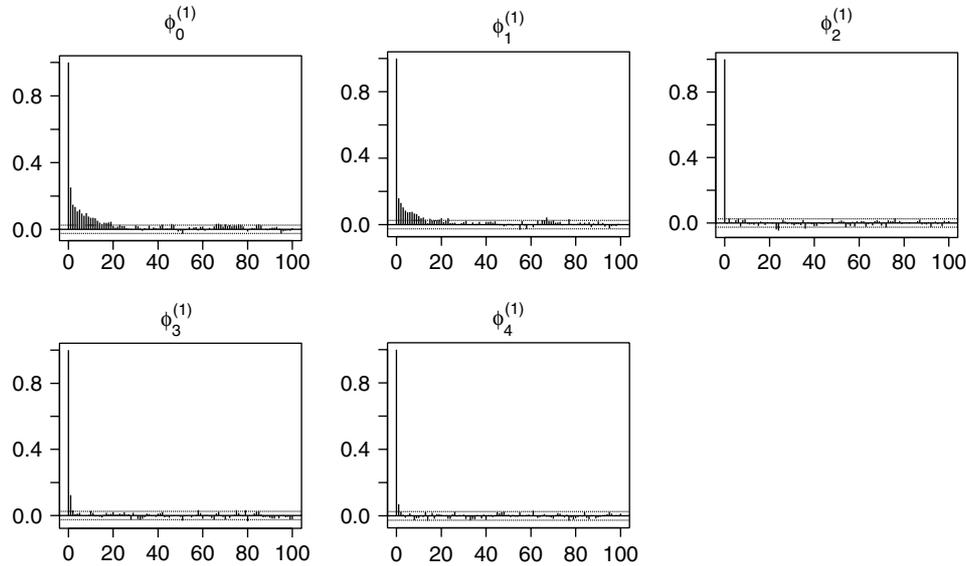


Figure 1. Autocorrelation plots of the MCMC iterates of  $\phi_i^{(1)}, i = 0, \dots, 4$ , in the case  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k}) = (0.5, 5)$  for the data simulated from Model 1

Four sets of  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k})$  are tested. They are  $(0.5, 5)$ ,  $(0.5, 10)$ ,  $(1, 5)$  and  $(1, 10)$ . The latter two are proposed by George and McCulloch (1993) and the former two are our suggestions. The true model can be expressed as

$$(\delta_1, \delta_2)' = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

The two ‘1’s mean having the coefficients  $\phi_1^{(1)}$  and  $\phi_1^{(2)}$  in the true model. We report in Table I the best three models selected based on the posterior probabilities  $P(d, (\delta_1, \delta_2)' | \mathbf{Y})$ . Note that all the twelve models are associated with  $d = 1$ . For  $(1, 5)$ , the best model is

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

with  $d = 1$  and the posterior probability of 0.013. Clear overestimation in the AR orders is observed. When we scan from the best model to the third best model, we find closer models to the true one. The same problem is also recorded in  $(1, 10)$ . Besides  $\phi_1^{(1)}$  and  $\phi_1^{(2)}$  which are in the true model, the best model also includes  $\phi_0^{(1)}$  and  $\phi_3^{(1)}$ . There is some improvement in  $(1, 10)$  compared with  $(1, 5)$ . Higher posterior probabilities are also recorded in  $(1, 10)$  but the overestimation still persists in this case. The problem is probably due to the use of a too small  $\tau_{i,k}$ . It is because setting  $\tau_{i,k}$  too small can include unimportant variables in the best model. So we suggest using  $(0.5, 5)$  and  $(0.5, 10)$  as supplements to  $(1, 5)$  and  $(1, 10)$  in this paper. The results from using the two priors are also presented in Table I. We see marked improvement over the two sets with  $\sigma_{\phi_k^{(i)}}/\tau_{i,k} = 1$  in the sense that the true model is correctly selected as the best. In addition, the highest posterior probabilities 0.018 and 0.046 are substantially larger than that of  $(1, 5)$  and  $(1, 10)$ , probably indicating that

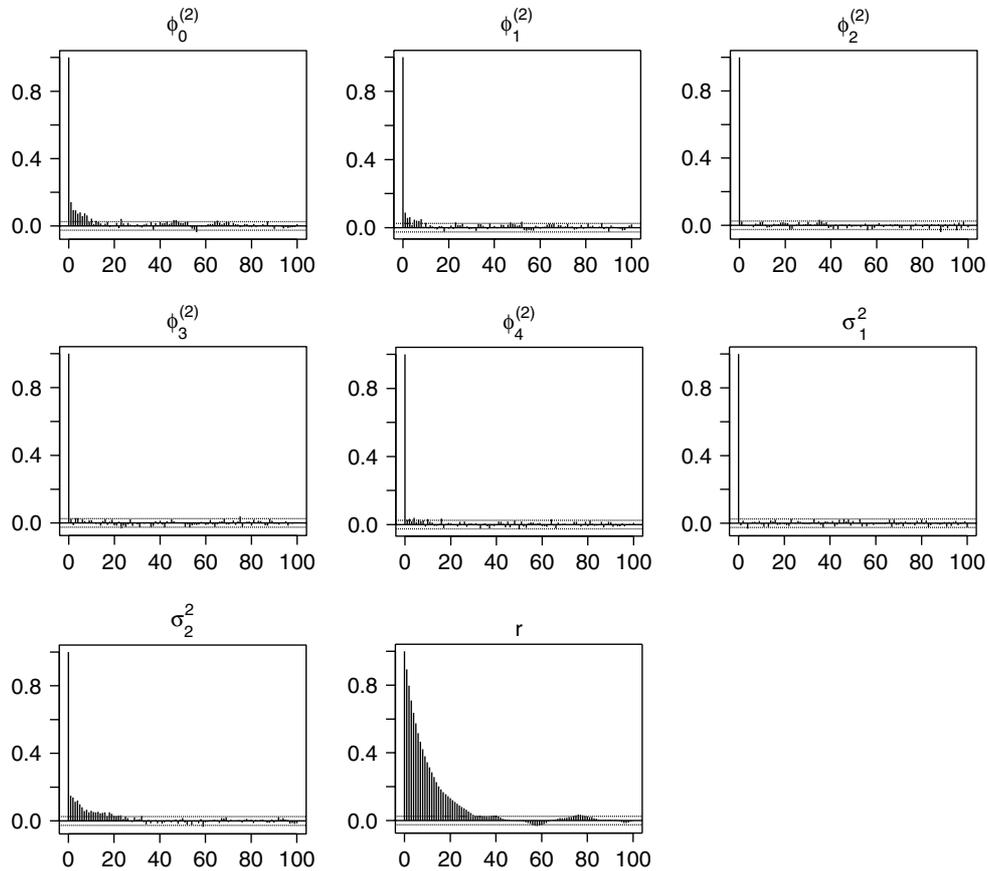


Figure 2. Autocorrelation plots of the MCMC iterates of  $\phi_i^{(2)}, i = 0, \dots, 4, \sigma_1^2, \sigma_2^2$  and  $r$  in the case  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k}) = (0.5, 5)$  for the data simulated from Model 1

good models are sharply identified. Indeed, the second and the third best models are very close to the true model in both cases.

To further investigate the differences in the performance among the four  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k})$ , we plot the posterior probabilities for all the 4096 models in Figure 3. We record the least variation in the probabilities in (1, 5), indicating that good models cannot be well-separated from others. We observe some improvements in (1, 10) by having more variation in the probabilities but the extra variation is still not enough to identify useful models. Higher dispersion exists in (0.5, 5) where the first two models are obviously outstanding. The one with (0.5, 10) is the best scenario that has the highest dispersion in the probabilities. More importantly, six models are easily sorted out with the probabilities  $>0.02$  while the others may not deserved of immediate attention. Overall, the above demonstration suggests that (0.5, 5) and (0.5, 10) are superior to (1, 5) and (1, 10) in the subset threshold autoregression. The one with (0.5, 10) is obviously the best.

To gain more evidence about the relative performance, we run another set of simulation on two more complicated models. The first one is Model 2 which consists of two AR(2) processes in the

Table I. The best three models expressed in terms of  $(\delta_1, \delta_2)'$  obtained from applying the four sets of  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k})$  to the simulated data generated from Model 1

	Best	Second best	Third best
(0.5, 5)	$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$ [0.018]	$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$ [0.016]	$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$ [0.014]
(0.5, 10)	$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$ [0.046]	$\begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$ [0.029]	$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$ [0.029]
(1, 5)	$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$ [0.013]	$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$ [0.012]	$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$ [0.012]
(1, 10)	$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$ [0.016]	$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$ [0.015]	$\begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$ [0.014]

Note:

The square brackets show the corresponding posterior probabilities. All the twelve models are associated with  $d = 1$ .

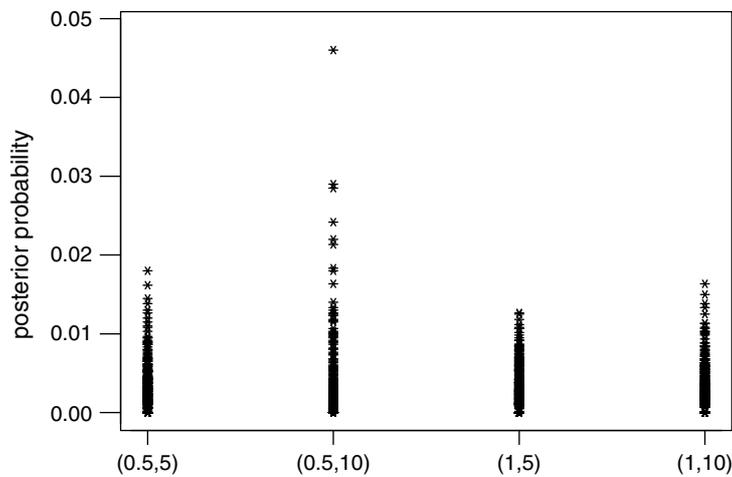


Figure 3. Plots of the posterior probability for all possible models in the four sets of  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k})$  when fitting the simulated data of Model 1

regimes.

$$\text{Model 2} \quad y_t = \begin{cases} 0.2 - 0.4 y_{t-1} + 0.3 y_{t-2} + a_t^{(1)}, & a_t^{(1)} \sim N(0, 0.64) \text{ if } y_{t-2} \leq 0.4 \\ 0.2 + 0.4 y_{t-1} + 0.3 y_{t-2} + a_t^{(2)}, & a_t^{(2)} \sim N(0, 0.25) \text{ if } y_{t-2} > 0.4 \end{cases}$$

The second one is Model 3 which is defined as

$$\text{Model 3} \quad y_t = \begin{cases} 0.3 + 0.4 y_{t-1} + 0.3 y_{t-4} + a_t^{(1)}, & a_t^{(1)} \sim N(0, 0.64) \text{ if } y_{t-1} \leq 0.5 \\ 0.6 - 0.4 y_{t-2} + a_t^{(2)}, & a_t^{(2)} \sim N(0, 0.25) \text{ if } y_{t-1} > 0.5 \end{cases}$$

The major difference between the two models is that some coefficients in the SETAR(2:4;2) specification are zero in Model 3 whereas Model 2 is a full SETAR(2:2;2). As advised in the first experiment, we fix 10,000 and 4000 as the total MCMC iterations and burn-in iterations. We consider two sample sizes;  $n = 500$  and  $n = 1000$  with 100 replications. In other words, 100 series from Model 2 and Model 3 are simulated and the SSVS method is then applied. The results are given in Tables II and III. We present the proportions (in percentages) of correctly selecting the true model. We also enumerate the proportions of times that the true model is declared as the best or the second best according to the posterior probabilities. Obviously, the proportions that include the first and the second best should be greater than that for the first best. The worst case is clearly (1, 5) where all the proportions for selecting the best are less than 25%. Even if we count the second best, the proportions are still less than 40%. Comparing with (1, 5), (1, 10) shows better results but the proportions for selecting the best are still small. In addition, increasing  $n$  does not seem to get any improvement. Obviously, (0.5, 5) and (0.5, 10) outperform (1, 5) and (1, 10). Using (0.5, 5) successfully selects the true model at least 62% of the trials while (0.5, 10) hits the best at least 87% of the trials except for  $n = 500$  in Model 2. Taking into account the second best model, the minimum proportions of hitting the true model increase to 84 and 82 respectively for (0.5, 5) and (0.5, 10). Generally speaking, increasing  $n$  tends to do better when  $\sigma_{\phi_k^{(i)}}/\tau_{i,k} = 0.5$ . This second simulation experiment confirms the remark in the first part that (0.5, 5) and (0.5, 10) are preferable.

Table II. The proportion of correctly selecting Model 2 out of 100 replications in different sets of  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k})$

	$n = 500$	$n = 1000$
(0.5, 5)	62 [84]	81 [93]
(0.5, 10)	48 [82]	87 [100]
(1, 5)	23 [35]	23 [34]
(1, 10)	47 [70]	46 [65]

Note:

The square brackets show the proportion of times that the true model is declared as either the best or the second best.

Table III. The proportion of correctly selecting Model 3 out of 100 replications in different sets of  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k})$

	$n = 500$	$n = 1000$
(0.5, 5)	76 [88]	77 [87]
(0.5, 10)	87 [96]	93 [98]
(1, 5)	12 [16]	8 [11]
(1, 10)	37 [56]	32 [53]

Note:

The square brackets show the proportion of times that the true model is declared as either the best or the second best.

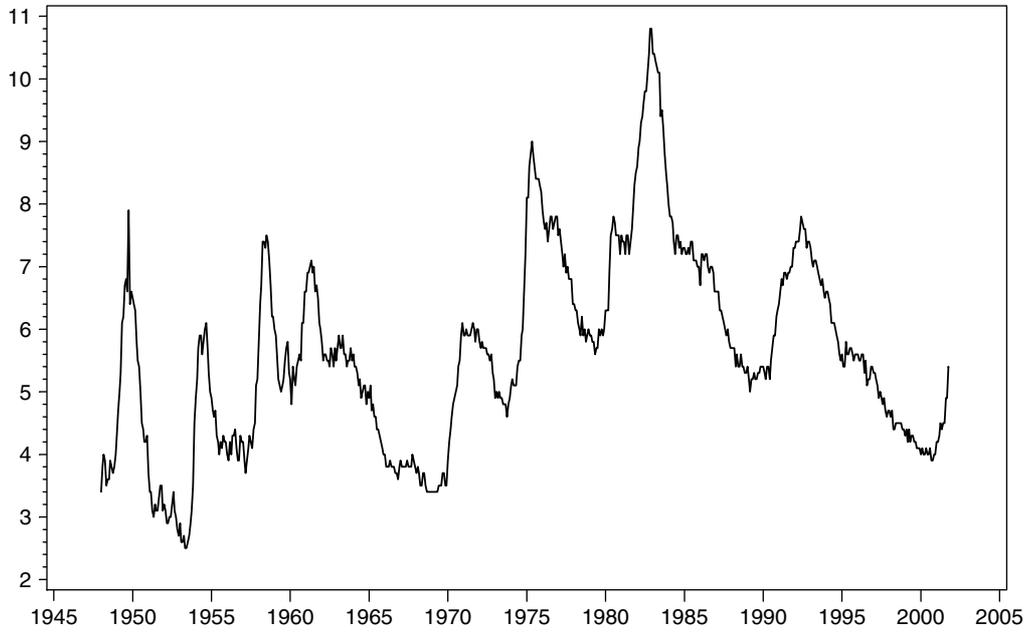


Figure 4. Monthly unemployment rate of the United States from January 1948 to October 2001

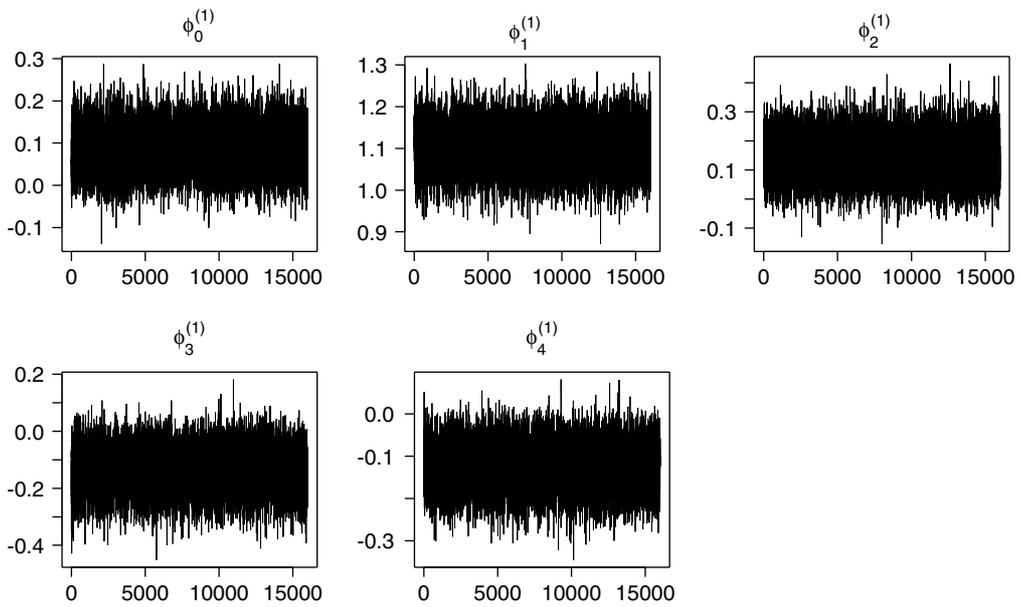


Figure 5. Time series plots of the last 16,000 MCMC iterates of  $\phi_k^{(1)}$ ,  $k = 0, \dots, 4$ , in the case  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k}) = (0.5, 5)$  for the unemployment data

We also recommend considering the second best model when doing model selection, as in some cases the second best is also a promising model choice.

ILLUSTRATIVE EXAMPLE

We apply the best subset selection method to the US unemployment rate from January 1948 to October 2001. The data set which consists of 646 monthly observations is shown in Figure 4. As suggested by the simulation studies, we use the two sets of prior parameters  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k}) = (0.5, 5)$  and  $(0.5, 10)$  for implementing the stochastic search algorithm where  $\sigma_{\phi_k^{(i)}}$ 's are the standard errors of the AR coefficients obtained from fitting an AR(4) model. The standard errors are found to be 0.031, 0.039, 0.056, 0.056 and 0.039 respectively for  $k = 0, 1, 2, 3,$  and  $4$ . The hyperparameters are chosen as  $\mathbf{V}_i = \mathbf{I}$ ,  $v_i = 3$ , and  $\lambda_i = \tilde{\sigma}^2/3$  for  $i = 1, 2$ , where  $\tilde{\sigma}^2$  is the sample variance of the data. Moreover, we choose 4 as the maximum possible delay, i.e.  $d_0 = 4$ . The maximum AR orders

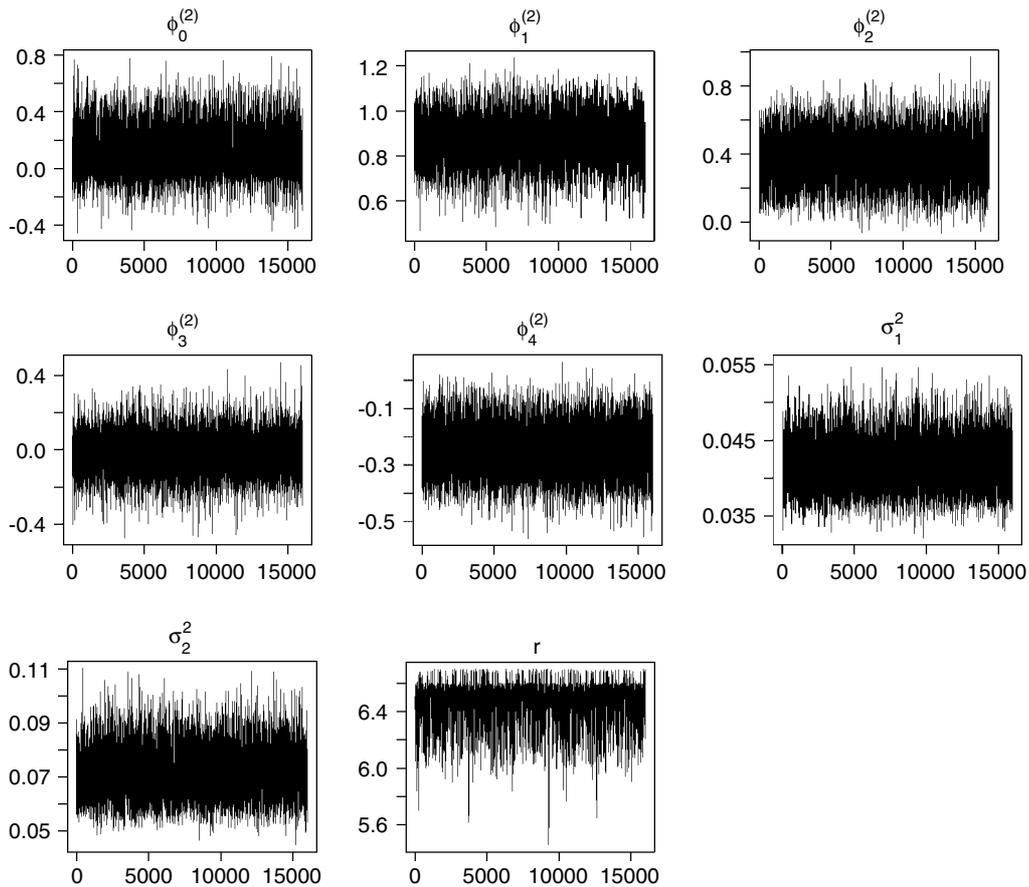


Figure 6. Time series plots of the last 16,000 MCMC iterates of  $\phi_k^{(2)}$ ,  $k = 0, \dots, 4$ ,  $\sigma_1^2$ ,  $\sigma_2^2$  and  $r$  in the case  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k}) = (0.5, 5)$  for the unemployment data

in both regimes are set at 4 and so we have altogether  $2^{4+4+2} = 1024$  possible combinations of  $\Theta_1$  and  $\Theta_2$ . Accounting also for 4 different delay parameters, we have altogether  $4 \times 1024 = 4096$  possible models to consider. Traditional methods such as AIC for selecting a good model would be very difficult as there are more than 4000 possible models and estimates of  $d$  and  $r$  have to be determined before doing the model comparison.

We iterate the MCMC 20,000 times and discard the first 4000 burn-in iterates. It takes less than 15 minutes in a 1GHz Pentium PC to complete the iterations. The time series and autocorrelation plots of all the parameters including  $\phi_k^{(i)}$ ,  $\sigma_i^2$  and  $r$ ,  $i = 1, 2$ ,  $k = 0, \dots, 4$ , are displayed in Figures 5 to 8. All the plots clearly suggest that convergence is attained after 4000 burn-in iterations. The time path and the correlograms of most parameters resemble that of a random sample, indicating that the convergence is fast. The only exception is  $r$  which has very mild autocorrelations in the first 20 lags. Overall, the MCMC sampling scheme adopted in this paper effectively constructs a sample from the joint posterior distribution of parameters for further statistical inference. The posterior means and variances of the parameters are given in Table IV. The means and variances for (0.5, 5) and (0.5, 10) are very similar. So the use of the two different  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k})$  only imposes minor changes in the posterior densities of the parameters. The threshold value  $r$  has means at around 6.5. The posterior probabilities  $P(d|Y)$  of  $d = 1, 2, 3$  and 4 are 0.596, 0.164, 0.239 and 0.001 for (0.5, 5) and 0.620, 0.181, 0.198 and 0.001 for (0.5, 10). The results from the two sets of  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k})$  are in good agreement and both indicate that the most appropriate choice of the delay parameter is 1.

Table V lists the best four models based on the posterior probabilities  $P(d, (\delta_1, \delta_2)'|Y)$ . The probabilities of the best four models are 0.025, 0.025, 0.021 and 0.017 for (0.5, 5) and 0.020, 0.019, 0.018 and 0.017 for (0.5, 10). All the eight best models have  $d = 1$  which are consistent

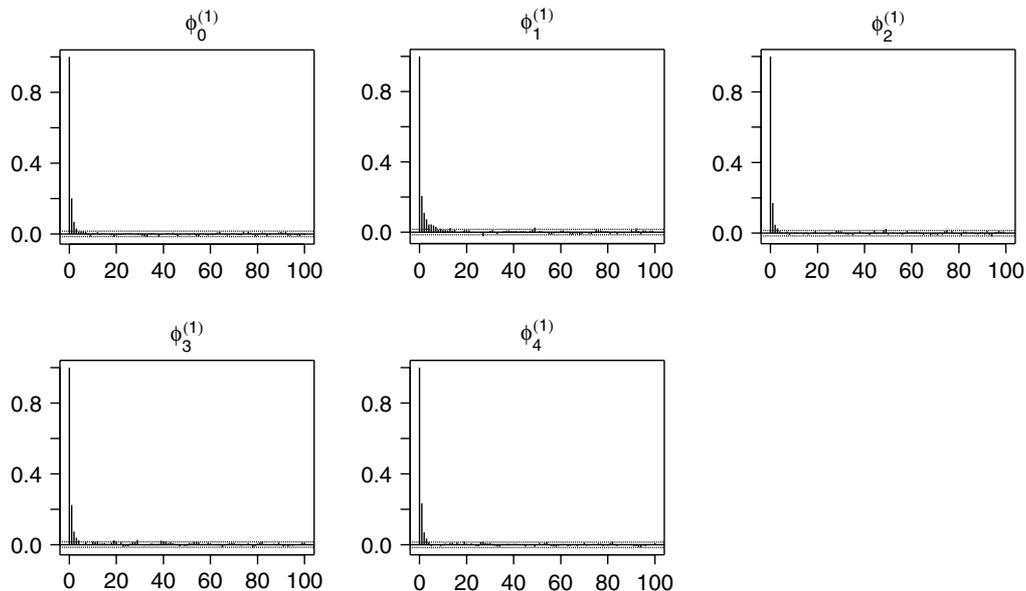


Figure 7. Autocorrelation plots of the MCMC iterates of  $\phi_i^{(1)}$ ,  $i = 0, \dots, 4$ , in the case  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k}) = (0.5, 5)$  for the unemployment data

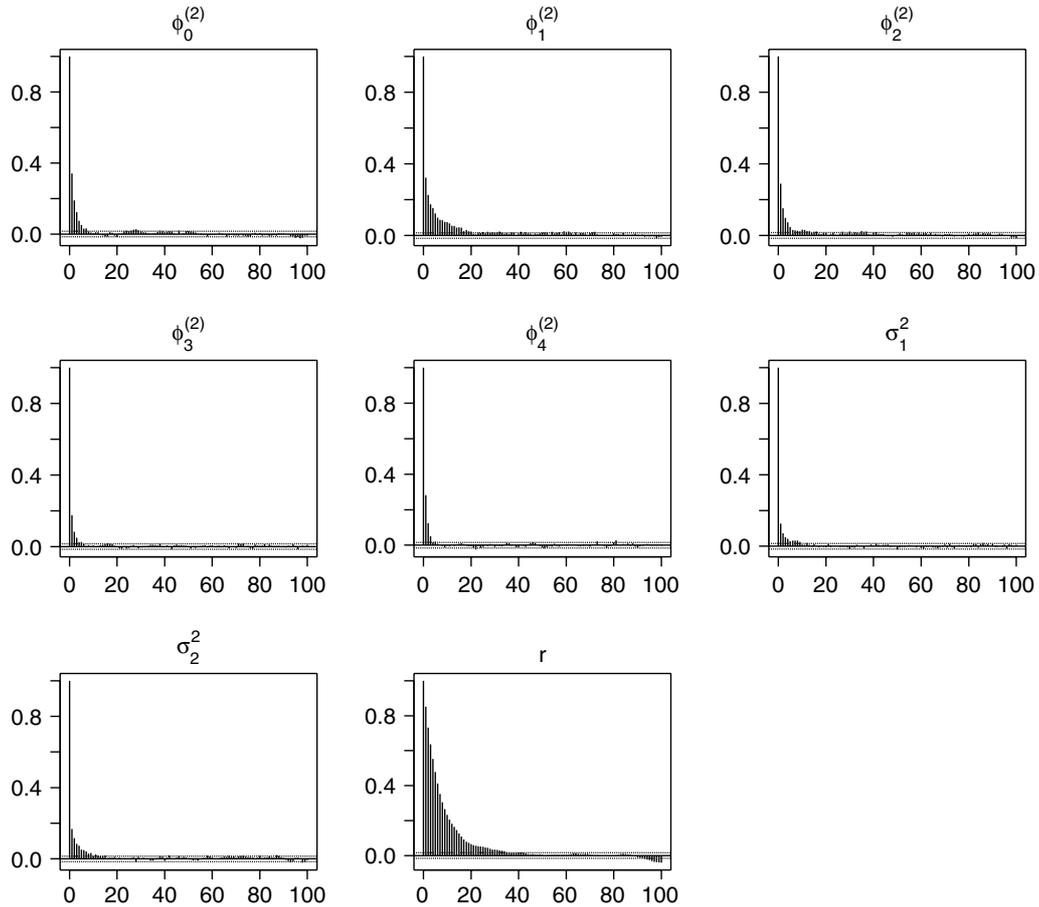


Figure 8. Autocorrelation plots of the MCMC iterates of  $\phi_i^{(2)}$ ,  $i = 0, \dots, 4$ ,  $\sigma_1^2$ ,  $\sigma_2^2$  and  $r$  in the case  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k}) = (0.5, 5)$  for the unemployment data

with the marginal posterior probabilities  $P(d|\mathbf{Y})$  that  $d = 1$  should be the most suitable choice. Using a smaller value of  $c_{i,k}$ , i.e. 5, the searching procedure tends to select more complicated models. The best model under  $(0.5, 5)$  is close to the full SETAR(2:4;4) model while the best choice under  $(0.5, 10)$ , i.e.  $M3$ , has only one parameter less. Since there are some overlaps in models between the two sets of  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k})$ , we end up with only five distinct models which are denoted by  $M1$  to  $M5$  in Table V. The five models,  $M1$  to  $M5$ , look very similar. Indeed, they have the same structure in the second regime and differ mostly in the two entries,  $\phi_0^{(1)}$  and  $\phi_2^{(1)}$ . The results indicate that the selection is robust to the two choices of  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k})$ . Our procedure based on MCMC sampling successfully reduces the models under consideration from 4096 to 5. More importantly, we do not need to fix  $r$  and  $d$  in advance as in other classical methods. Therefore, the uncertainty in the threshold value and the delay parameter can be taken into account while performing the stochastic search for the best model. Further comparisons among the five promising models identified by our method can be done using standard approaches such as AIC.

Table IV. Posterior means and standard deviations of the parameters for the unemployment data using  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k}) = (0.5, 5)$  and  $(0.5, 10)$

	(0.5, 5)		(0.5, 10)	
	Post. mean	Post. sd	Post. mean	Post. sd
$\phi_0^{(1)}$	0.08597	0.05018	0.08144	0.05046
$\phi_1^{(1)}$	1.08736	0.05194	1.10108	0.05237
$\phi_2^{(1)}$	0.15973	0.07095	0.14241	0.07042
$\phi_3^{(1)}$	-0.13514	0.07277	-0.13442	0.07174
$\phi_4^{(1)}$	-0.12722	0.05150	-0.12344	0.05154
$\phi_0^{(2)}$	0.09837	0.13913	0.10473	0.15030
$\phi_1^{(2)}$	0.83636	0.09737	0.86542	0.10011
$\phi_2^{(2)}$	0.40932	0.12659	0.38060	0.13319
$\phi_3^{(2)}$	-0.01179	0.09821	-0.01663	0.09805
$\phi_4^{(2)}$	-0.25093	0.07825	-0.24715	0.08064
$\sigma_1^2$	0.04193	0.00295	0.04190	0.00295
$\sigma_2^2$	0.07151	0.00818	0.07174	0.00827
$r$	6.46688	0.15404	6.45189	0.15743

Table V. The best four models expressed in terms of  $(\delta_1, \delta_2)'$  for the unemployment data using  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k}) = (0.5, 5)$  and  $(0.5, 10)$

	(0.5, 5)	(0.5, 10)
Best	$M1 \begin{pmatrix} 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix}$	$M3$
Second best	$M2 \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix}$	$M1$
Third best	$M3 \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix}$	$M5 \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix}$
Fourth best	$M4 \begin{pmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix}$	$M4$

Note:

All the eight best models are associated with  $d = 1$ .

Finding the best model under certain objectives becomes feasible as we can limit our focus to five models.

### CONCLUDING REMARKS

In this paper, we propose a subset selection method for the threshold autoregressive models. An advantage of our approach is that it can identify the best subset and estimate the model parameters simultaneously. Our method is feasible even though the number of possible models is very large. From the simulation experiments, we demonstrate that our two choices of  $(\sigma_{\phi_k^{(i)}}/\tau_{i,k}, c_{i,k})$ ,  $(0.5, 5)$  and  $(0.5, 10)$ , perform very well in picking up the correct model. The application to the US

unemployment rate data effectively limits our consideration to five best SETAR models for further exploration.

#### ACKNOWLEDGEMENTS

The authors would like to thank the Editor and the Associate Editor for their helpful suggestions and comments. C. W. S. Chen acknowledges research support from National Science Council of Taiwan grant NSC90-2118-M-035-008. The research of Mike So was supported by Hong Kong RGC Direct Allocation grants 00/01.BM25 and 01/02.BM38.

#### REFERENCES

- Box GEP, Tiao GC. 1973. *Bayesian Inference in Statistical Analysis*. Addison-Wesley: Reading, MA.
- Carlin BP, Gelfand AE, Smith AFM. 1992. Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics* **41**: 389–405.
- Chen CWS. 1992. Bayesian analysis of bilinear time series models: a Gibbs sampling approach. *Communications in Statistics—Theory and Methods* **21**: 3407–3425.
- Chen CWS. 1999. Subset selection of autoregressive time series models. *Journal of Forecasting* **18**: 505–516.
- Chen CWS, Lee JC. 1995. Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis* **16**: 483–492.
- DeGroot M. 1970. *Optimal Statistical Decisions*. McGraw-Hill: New York.
- George EI, McCulloch RE. 1993. Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**: 881–889.
- Geweke J, Terui N. 1993. Bayesian threshold autoregressive models for nonlinear time series. *Journal of Time Series Analysis* **14**: 441–454.
- McCulloch RE, Tsay RS. 1993. Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association* **88**: 968–978.
- McCulloch RE, Tsay RS. 1994a. Bayesian analysis of autoregressive time series via the Gibbs sampler. *Journal of Time Series Analysis* **15**: 235–250.
- McCulloch RE, Tsay RS. 1994b. Statistical inference of economic time series via Markov switching models. *Journal of Time Series Analysis* **15**: 523–539.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH. 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**: 1087–1091.
- Thanoon BY. 1990. Subset threshold autoregression with applications. *Journal of Time Series Analysis* **11**: 75–87.
- Tierney L. 1994. Markov chains for exploring posterior distributions *The Annals of Statistics* **22**: 1701–1762.
- Tong H. 1978. On a threshold model. In *Pattern Recognition and Signal Processing*, Chen CH (ed.). Sijthoff & Noordhoff: Amsterdam.
- Tong H. 1983. *Threshold Models in Nonlinear Time Series Analysis*. Springer-Verlag: New York.
- Tong H, Lim KS. 1980. Threshold autoregression, limit cycles and cyclical data (with discussion). *Journal of the Royal Statistical Society B* **42**: 245–292.
- Tsay RS. 1989. Testing and modeling threshold autoregressive process. *Journal of the American Statistical Association* **84**: 231–240.

#### Authors' biographies:

**Mike K. P. So** is an Assistant Professor of the Department of Information and Systems Management at the Hong Kong University of Science and Technology. He received his PhD in Statistics from The University of Hong Kong in 1996. His research interests include non-linear time series analysis and its applications, modelling of financial time series and market volatility study.

**Dr Cathy W. S. Chen** is Professor of Statistics, Feng Chia University, Taiwan. Her research interests include financial time series modelling, econometric methods, MCMC methods, and statistical approaches to epidemiology. She has published papers in *Journal of Time Series Analysis*, *Journal of Forecasting*, *Statistica Sinica*, *Statistics in Medicine*, and *Aids*. Her publication lists can be found at <http://www.stat.fcu.edu.tw/faculty/Chen.html>

*Authors' addresses:*

**Mike K. P. So**, Department of Information and Systems Management, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong.

**Cathy W. S. Chen**, Department of Statistics, Feng-Chia University, Taichung 40724, Taiwan.