Statistica Sinica **17**(2007), 241-264

THRESHOLD VARIABLE DETERMINATION AND THRESHOLD VARIABLE DRIVEN SWITCHING AUTOREGRESSIVE MODELS

Senlin Wu[1] and Rong Chen[1,2]

[1] *University of Illinois at Chicago and* [2] *Peking University*

Abstract: In this paper we propose a new class of nonlinear time series models, the threshold variable driven switching autoregressive models. It is a hierarchical model that combines two important nonlinear time series models, the threshold autoregressive (AR) models and the random switching AR models. The underlying time series process switches between two (or more) different linear models. The switching dynamics relies on an observable threshold variable (up to certain estimable parameters) as used in a threshold model, hence reveals the true nature of the switching mechanism. It also allows certain randomness in the switching procedure similar to that in a random switching model, hence provides some flexibility. Furthermore, we propose a model building procedure that concentrates on a fast determination of an appropriate threshold variable among a large set of candidates (and linear combinations of them). This procedure is applicable to the new models as well as the classical threshold models. A simulation study and two data examples are presented.

Key words and phrases: Model selection, posterior BIC, switching AR models, threshold AR models.

1. Introduction

Since the seminal paper of Tong and Lim (1980) on the threshold autoregressive (TAR) model, there have been a number of successful applications of the TAR model in various fields, such as economics, finance, biology, epidemiology, meteorology, and astronomy (Tong (1983), Chen (1995), Tong (1990), Watier and Richardson (1995), Montgomery, Zarnowitz, Tsay and Tiao (1998) and Tsay (1998)). TAR models provide a simple yet elegant approach to nonlinear time series (Tong (1990), Moeanaddin and Tong (1988) and Tsay (1989)). Various extensions of the TAR model have been proposed, including the threshold MA model (Gooijer (1998)), the threshold integrated MA model (Gonzalo and Martínez (2004)), the threshold ARMA model (Sáfadi and Morettin (2000)), the threshold ARCH model (Li and Lam (1995) and Li and Li (1996)), the threshold GARCH model (Zakoian (1994)), the threshold stochastic volatility model

(So, Li and Lam (2002)), and the smooth transition autoregressive model (Chan and Tong (1986), Teräsvirta (1994) and van Dijk, Teräsvirta and Franses (2002)). A more general class of models is proposed by Huerta, Jiang, and Tanner (2003), termed hierarchical mixture time series models.

Consider a (generalized) two-state TAR model

$$Y_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} Y_{t-1} + \cdots + \phi_p^{(1)} Y_{t-p} + \varepsilon_t^{(1)} & \text{if } Z_t \geq 0, \\ \phi_0^{(2)} + \phi_1^{(2)} Y_{t-1} + \cdots + \phi_p^{(2)} Y_{t-p} + \varepsilon_t^{(2)} & \text{if } Z_t < 0, \end{cases}$$

where Z_t is the threshold variable that determines the dynamic switching mechanism of the model. Many different choices of threshold variables have been used in applications. Particularly, in the standard self-exciting AR models, Z_t is assumed to be $Y_{t-d} - c$, where Y_{t-d} is a lag variable of the observed time series. In open-loop threshold models (Tong (1990)), Z_t takes the form $Z_t = X_{t-d} - c$. That is, the current mode of Y_t is determined by an exogenous time series X_t. Other choices include linear combinations of the lag variables or exogenous variables (Chen and So (2006) and Tsay (1998)) and nonlinear combinations of the form $Z_t = f(X_{t-d_1}, X_{t-d_2})$ (Chen (1995)).

A special case of the TAR model is the switching autoregressive (SAR) model, as first proposed in Tong and Lim (1980), subsequently formalized by Hamilton (1989), and used by McCulloch and Tsay (1994b). This model uses a random latent (unobservable) indicator as the threshold variable. Specifically, a two-state SAR model can be written as:

$$Y_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} Y_{t-1} + \cdots + \phi_p^{(1)} Y_{t-p} + \varepsilon_t^{(1)} & \text{if } I_t = 1, \\ \phi_0^{(2)} + \phi_1^{(2)} Y_{t-1} + \cdots + \phi_p^{(2)} Y_{t-p} + \varepsilon_t^{(2)} & \text{if } I_t = 2, \end{cases}$$

where I_t is the hidden state variable, or the "switching indicator". The switching mechanism can be either an independent process characterized as $P(I_t = 1) = p$ and I_i, I_j independent for $i \neq j$, or a Markovian dependent process characterized as $P(I_t = j \mid I_{t-1} = i) = p_{ij}$ for $i, j = 1, 2$. Hamilton (1989) and McCulloch and Tsay (1994a) used this model for analyzing the U.S. GNP series and identifying the "contraction" and "expansion" states of the economy.

TAR models assume that the state switching is deterministically controlled by an observable threshold variable. This observable threshold completely specifies the states of the process in the immediate future, hence produces a uni-modal predictive distribution. Also, the procedure to identify the threshold variable may reveal the possible relationship between the target series and other time series, which is helpful in understanding the underlying dynamics of the time series. However, this assumption is sometimes restrictive in cases. In practice

we often encounter situations that an TAR model works well except around the boundary between the two regimes. The smooth-transition TAR model (e.g., Chan and Tong (1986) and Teräsvirta (1994)) often helps. But when the states are mixed, there are no clear solutions. On the other hand, SAR models do not require an explicit observable threshold variable and enjoy a certain flexibility in the switching mechanism. However, it also has its limitations, especially when it is used for prediction. A two-state SAR model has a bi-modal predictive distribution, which results in a wide prediction interval. Further, in practice, it is often difficult to justify that the state switching mechanism is completely driven by a random process that does not depend on anything else.

To enjoy the strong information provided by an observable threshold variable, and to allow certain randomness in the switching mechanism at the same time, we introduce a combination of TAR and SAR models, termed as the *threshold variable driven switching AR models* (*TD-SAR*). Furthermore, we propose a model building procedure that concentrates on a fast determination of an appropriate threshold variable among a large set of candidates (and linear combination of them). It starts with classification (clustering) of observations into two (or more) classes. This preliminary classification forms the basis for fast searching. Once a small number of threshold variable candidates are identified, the full models are estimated, and model selection is carried out via certain criteria. This procedure is applicable to the new models as well as the threshold models.

The rest of the paper proceeds as follows. In Section 2 we formally introduce the TD-SAR model and propose a general strategy for the model building procedure, specially the procedure of threshold variable determination. This strategy can also be used for building standard TAR models. In Sections 3 to 5 we provide details on the three steps in the model building procedure. In Section 6 we study the empirical properties of this model and the modeling procedure through simulation. Section 7 contains two data examples and Section 8 presents a brief summary. Some technical details are contained in the Appendix.

2. Threshold Variable Driven Switching AR Models

2.1. The model

For a time series Y_t, $t = 1, \ldots, n$, a k-state TD-SAR(p) model can be expressed as

$$Y_t = \mathbf{Y}_{t-1}\boldsymbol{\phi}^{(I_t)} + \varepsilon_t^{(I_t)}, \qquad t = p+1, \ldots, n, \tag{1}$$

where $\mathbf{Y}_{t-1} = (1, Y_{t-1}, \ldots, Y_{t-p})$, and $I_t \in \{1, \ldots, k\}$. The general form of the switching mechanism is given by

$$P(I_t = i) = g_i(X_{1t}, \ldots, X_{mt}, \boldsymbol{\beta}_i), \quad i = 1, \ldots, k, \tag{2}$$

where $X_{1t}, \ldots, X_{mt}$ are observable variables (lag variables, exogenous variables, or their transformations) and $\boldsymbol{\beta}_i$ is a set of unknown parameters.

The link function $g_i(\cdot)$ in (2) is flexible. A natural choice would be the logistic link function. For a two-state switching model, we can use

$$P(I_t = 1) = \frac{e^{Z_t}}{1 + e^{Z_t}}, \tag{3}$$

where $Z_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_m X_{mt}$. For a three-state switching model, it may take the form of

$$P(I_t=1)=\frac{e^{Z_{1t}}}{e^{Z_{1t}}+e^{Z_{2t}}+1},\ P(I_t=2)=\frac{e^{Z_{2t}}}{e^{Z_{1t}}+e^{Z_{2t}}+1},\ P(I_t=3)=\frac{1}{e^{Z_{1t}}+e^{Z_{2t}}+1},$$

where

$$Z_{it} = \beta_0^{(i)} + \beta_1^{(i)} X_{1t} + \cdots + \beta_m^{(i)} X_{mt}, \quad i = 1, 2. \tag{4}$$

We call Z_t (or Z_{it}) the threshold variables. They are observable, given the parameters $\boldsymbol{\beta}$.

Remark 1. The standard TAR and SAR models are special cases of this general model. It simplifies to a TAR model if the function g has $g_1(Z_t) = I(Z_t - c \geq 0)$ and $g_2(Z_t) = 1 - g_1(Z_t)$. When $g_1(Z_t) = p$ and $g_2(Z_t) = 1 - p$, it becomes a two-state independent SAR model. When the threshold variable is taken as the latent variable $Z_t = I_{t-1}$, and the link function is set as $g_1(\cdot) = p_{i1}$, where p_{ij} ($i = 1, 2$ and $j = 1, 2$) is the transition probability $P(I_t = j \mid I_{t-1} = i)$, then we obtain the hidden Markovian SAR model.

Remark 2. The TD-SAR models have an extra layer of complexity compared to the standard TAR and SAR models. This additional complexity makes it possible to take the advantage of having an observable switching driving force enjoyed by the TAR model, and of having the flexibility of the switching mechanism enjoyed by the SAR model. Often the extra benefit outweighs the extra complexity of the model, as we demonstrate in the examples. This extra layer is specified by m extra parameters in a two-state model, and $(k-1)m$ extra parameters in a k-state model. In order to control the complexity and the tendency toward over-parametrization, a careful model determination procedure is required so that the number of extra parameters can be small.

Remark 3. The link function g_i in (2) is flexible. In this paper we choose the logistic function for its simplicity. It is commonly used to handle binary responses. In practice model assumption should always be checked, Note that

(4) includes a constant term that is to be estimated, There are no restrictions on the forms of the candidate variables.

Remark 4. The link function g_i in (2) can be multi-dimensional to accommodate multi-threshold situations. Multi-threshold models has been studied by Gonzalo and Pitarakis (2002) and Chong and Yan (2005). In this paper we focus on the single threshold case.

Remark 5. Note that a sufficient condition for the process (1) to be ergodic is that the AR model in each state be stationary, and that the Z_t process be a finite order stationary Markov chain. This can be easily proved using the tools in Tweedie (1975) and Tjostheim (1990). However, weaker conditions can be obtained in some cases. For example, if the threshold variable Z_t is independent of X_t and follows a discrete finite order stationary Markov chain taking values in Ω, then a sufficient condition for the ergodicity of the X_t process is

$$\max_{z\in\Omega}\{\sum_{i=1}^{k} p_i(z)||\mathbf{\Phi}^{(i)}||\} < 1,$$

where $p_i(z) = P(I_t = 1 \mid Z_t = z)$ in (2),

$$\mathbf{\Phi}^{(i)} = \begin{pmatrix} \phi_1^{(i)} & \phi_2^{(i)} & \cdots & \phi_{p-1}^{(i)} & \phi_p^{(i)} \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

and $||A||$ is the Euclidean norm. For example, for the AR(1) case, the process is ergodic if

$$\max\{p_1^*\phi^{(1)2} + (1-p_1^*)\phi^{(2)2}, (1-p_2^*)\phi^{(1)2} + p_2^*\phi^{(2)2}\} < 1,$$

where $p_1^* = \max_{z\in\Omega} p_1(z)$ and $p_2^* = \max_{z\in\Omega} p_2(z)$. More complicated cases are out of the scope of this paper. Tools used by Cline and Pu (1999) and Boucher and Cline (2007) can be used. For structural change between stationary and nonstationary time series, see Chong (2001).

2.2. Model building procedure

In practice it is always a difficult task to determine an appropriate threshold variable for a TAR model. TD-SAR model cannot avoid this difficulty either. To find the threshold variable, a commonly used method is to traverse all combinations of the possible threshold variables and threshold values, fit all

the corresponding models, and find the best one according to model selection criteria such as the Bayesian Information Criterion (BIC) or out-sample prediction performance (Tsay (1989)). This direct approach is only feasible for simple threshold variables such as lag variables or univariate exogenous variables. Recently researchers have started to consider linear combinations of several variables as the threshold variable (Chen (1995), Tsay (1998) and Chen and So (2006)) successfully. However, the traditional trial-and-error method is not sophisticated enough to handle even the linear combination of two variables. Here we propose a new approach to determine an appropriate threshold variable. It is fast, and has the capability to search among a large set of candidate variables and their linear combinations.

The approach is a reversed strategy. In TAR models, the threshold variable is used to determine the state I_t for each observation. Conversely, if the state is given, then an appropriate threshold variable should provide a close match to the state. It is much easier to check how well a variable agrees with the state indicator I_t than to use it directly in fitting the original model. Since I_t is usually unknown, we can estimate I_t first through a classification step. With the estimated I_t, we can then efficiently search for an appropriate threshold variable among a large set of candidates. Usually a small number of threshold variable candidates is retained and a full model is fit to each of them. The final model is selected by with model selection criteria.

We summarize our model building procedure as a three-stage algorithm.

1. **Classification.** Estimate the states I_t or the probabilities of the states $P(I_t = i)$. This step is essentially fitting a SAR model, with or without the Markovian structure in the switching mechanism. Following McCulloch and Tsay (1994a) and Chen (1995), we adopt a Bayesian approach in this step.
2. **Searching.** With the estimated $\hat{I}_t$ or $\hat{p}_{it} = \hat{P}(I_t = i)$, $i = 1, \ldots, k$, we search or construct the threshold variables that provide the best fit of $\hat{I}_t$ or $\hat{p}_{ti}$ under certain criteria. Here we propose to use CUSUM, minimum misclassification obtained via Support Vector Machine (SVM) algorithm, and the SVM-CUSUM (a combination of SVM and CUSUM) criterion.
3. **Full model estimation and model selection.** With several threshold variable candidates found in the searching step, a TD-SAR model fitting algorithm is used to estimate the full model. Again, this step is done with a Bayesian approach. A posterior BIC criterion is used for model selection.

This three-step algorithm can also be used to build a TAR model, with slight modification in the last step. Hence it also enhances our ability to use TAR models in applications.

In the next three sections we provide a more detailed implementation procedure.

3. Classification Algorithm

In Stage (I), we adopt the algorithm proposed by McCulloch and Tsay (1994b) and Chen (1995). For a time series Y_t, $t = 1, \ldots, n$, a k-state AR(p) model is

$$Y_t = \mathbf{Y}_{t-1}\boldsymbol{\phi}^{(I_t)} + \varepsilon_t^{(I_t)}, \qquad t = p+1, \ldots, n, \tag{5}$$

where $\mathbf{Y}_{t-1} = (1, Y_{t-1}, \ldots, Y_{t-p})$, $\boldsymbol{\phi}^{(i)} = (\phi_0^{(i)}, \phi_1^{(i)}, \ldots, \phi_p^{(i)})^T$, $i = 1, \ldots, k$, and I_t is the state indicator taking values from 1 to k. We assume the noises are independent, with possibly different variances for different states: $\varepsilon_t^{(I_t)} \sim N(0, \sigma_{I_t}^2)$.

Let $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_k)$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_k^2)$ and $\boldsymbol{I} = (I_{p+1}, \ldots, I_n)$. With $Y_1, \ldots, Y_p$ given and fixed, the conditional likelihood function of model (5) is

$$p(Y_{p+1}, \ldots, Y_n \mid \boldsymbol{\phi}, \boldsymbol{\sigma}^2, I_{p+1}, \ldots, I_n) \propto \prod_{t=p+1}^{n} \frac{1}{\sigma_{I_t}} \exp\left(-\frac{(Y_t - \mathbf{Y}_{t-1}\boldsymbol{\phi}^{(I_t)})^2}{2\sigma_{I_t}^2}\right). \tag{6}$$

Given independent priors on $\boldsymbol{\phi}$, $\boldsymbol{\sigma}^2$ and $\boldsymbol{I}$, the conditional posterior distribution becomes

$$\begin{aligned} p(\boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{I} \mid Y_{p+1}, \ldots, Y_n) &\propto p(Y_{p+1}, \ldots, Y_n \mid \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{I}) p(\boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{I}) \\ &\propto p(Y_{p+1}, \ldots, Y_n \mid \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{I}) p(\boldsymbol{\phi}) p(\boldsymbol{\sigma}^2) p(\boldsymbol{I}). \end{aligned}$$

Here the priors on $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}^2$ can be set as the standard conjugate priors, namely Gaussian and inverse χ^2. The indicator sequence I_t, $t = p+1, \ldots, n$ can either be assumed to be independent, with equal probability prior, or be assumed to follow a first order Markov chain, with unknown transition matrix. The Markov chain assumption is sometimes reasonable due to the possible autocorrelation in the underlying threshold variable series. If the Markovian model is used, one can specify a Dirichilet prior for the transition matrix.

We use the Gibbs sampler (Geman and Geman (1984), Gelfand and Smith (1990) and Robert and Casella (2004)) to draw samples from the posterior distribution, see the Appendix. Based on the samples drawn from the posterior distribution, we obtain the estimate of posterior probability $\hat{p}_{it} = P(I_t = i \mid X_{p+1}, \ldots, X_n)$, $i = 1, \ldots, k$, as well as the posterior mode estimator $\hat{I}_t = \arg\max_{i \in \{1,\ldots,k\}} \{\hat{p}_{it}\}$. The estimated states, or the probabilities, are then fed into the threshold variable determination procedure.

4. Searching Algorithm

Once the estimated indicators $\hat{I}_t$ or estimated posterior probability $\hat{p}_{it}$ are obtained, they can be matched with various candidate variables and their combinations or transformations to determine an appropriate threshold variable. Simple graphical methods can be used (Chen (1995)). However, graphical methods

have limitations when searching among a large candidate set. A more automatic procedure is needed.

Here we propose three quantitative criteria for the evaluation of the potential candidates as the threshold variables: CUSUM, misclassification via Support Vector Machine (SVM) and SVM-CUSUM. The CUSUM method is used to evaluate one-dimensional candidates; for evaluating linear combinations of multiple candidates as the threshold variables, we use SVM, which is commonly used as a supervised learning tool to find the best classification rule in high dimensional space; SVM-CUSUM is a combination of SVM and CUSUM.

4.1. CUSUM

The CUSUM method originated from control charts in production management. It has been used to find change points, e.g., Tong (1990) and Taylor (2000). Here we use it to measure the agreement between the preliminary classification $\hat{p}_{it} = \hat{P}(I_t = i)$ (or $\hat{I}_t$) and a threshold variable candidate. The idea is that, if a variable Z_t is indeed the correct threshold variable for a two-state TAR model, then there is a threshold value c such that for all $Z_t < c$, $\hat{p}_{1t} = \hat{P}(I_t = 1)$ are below (above) 0.5, and for all $Z_t > c$, $\hat{p}_{1t}$ are above (below) 0.5. Then if we cumulatively add $(p_{1t} - 0.5)$ in the ascending order of Z_t, the cumulative sum will reach its minimum (or its maximum) around $Z_t = c$. On the other hand, when Z_t is not the correct threshold variable, it does not provide any meaningful order for the $\hat{p}_{1t}$. Hence the resulting partial sum would be small because of the cancelation of the negative and positive deviations of $(\hat{p}_{1t} - 0.5)$. Specifically, in a k-state problem, we set $p_0 = 1/k$ and perform the following procedure.

1. For each state i, $(i = 1, \ldots, k)$, sort $\hat{p}_{it}$ according to the increasing order of Z_t. This forms a new series $\hat{p}_{it^*}$. Define CUSUM for state i for the variable Z_t as
$$\mathrm{CUSUM}_i(Z_t) = \max_{t_1} \sum_{s=1}^{t_1} (\hat{p}_{is^*} - p_0) - \min_{t_2} \sum_{s=1}^{t_2} (\hat{p}_{is^*} - p_0).$$
2. Define CUSUM for variable Z_t as
$$\mathrm{CUSUM}(Z_t) = \sum_{i=1}^{k-1} \mathrm{CUSUM}_i(Z_t). \tag{7}$$

Figure 1 demonstrates the CUSUM measure. It plots the estimated probability $P(I_t = 1) - 0.5$ against a wrong threshold variable X_1 (left) and the "true" threshold variable X_2 (right). The "true" threshold variable provides a clear separation between the two states. Figure 2 shows the partial cumulative sum, corresponding to $\mathrm{CUSUM}(X_1) = 21$, and $\mathrm{CUSUM}(X_2) = 112$.

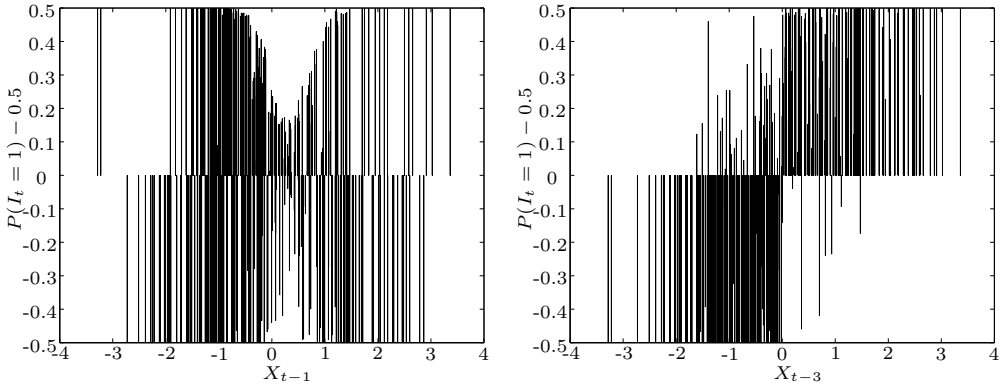

Figure 1. The effects of threshold variable. The classification probabilities are plotted against a wrong threshold variable and a correct threshold variable in the left and right panels, respectively.

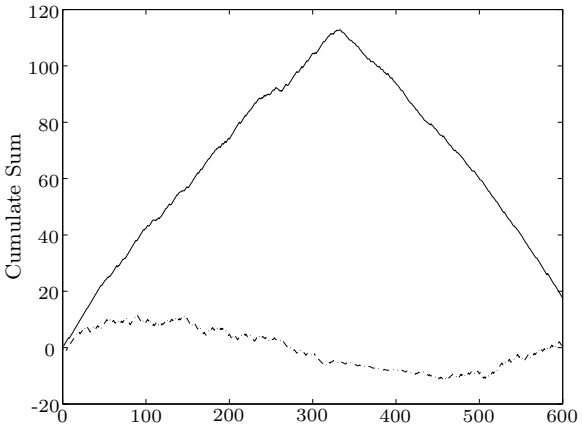

Figure 2. The CUSUM measure. The cumulative sums are plotted using a wrong threshold variable (dashed line) and a correct one (solid line).

4.2. SVM

The Support Vector Machine (SVM) (Vapnik (1998), Cristianini and Shawe-Taylor (2000) and Hastie, Tibshirani and Friedman (2001)) is a powerful tool for supervised classification. The original SVM finds the direction in the feature space that provides the largest separating margin between the data in the two classes. For our purpose, we concentrate on the inseparable case in relatively lower dimensional spaces, due to the concern of over-parametrization. The following is a simplified version from Hastie, Tibshirani and Friedman (2001), but tailored to our applications.

We first consider two-state cases. For a set of m variables $\{X_{1t}, \ldots, X_{mt}\}$, we consider a possible linear combination $Z_t = \sum_{j=1}^{m} \theta_j X_{jt}$ as our threshold variable. A valid threshold variable should group $\hat{I}_t$ according to the sign of $Z_t - c$, which is equivalent to using $\sum_{j=1}^{m} \theta_j X_{jt} - c$ to separate the I_t. The SVM is designed to find the optimal linear combination $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^T$. Based on this, the misclassification rate can be calculated and the optimal threshold variable can be identified.

Assume the classification sequence is $\hat{I}_t \in \{1, 2\}$, Let $\hat{I}_t^* = 2(\hat{I}_t - 1.5) \in \{-1, +1\}$. Given inseparable data (features) $\boldsymbol{X}_t = (X_{1t}, \ldots, X_{mt})$, $t = 1, \ldots, n$, and their classification (or state) indicator $\hat{I}_t^*$, the linear classification problem is expressed as:

$$\min_{\boldsymbol{\theta}, c, \zeta_t} \| \boldsymbol{\theta} \|^2 \text{ s.t. } \quad \zeta_t \geq 0, \ \sum_{t=1}^{n} \zeta_t \leq K, \ \hat{I}_t^* (\boldsymbol{X}_t \boldsymbol{\theta} - c) \geq 1 - \zeta_t \ \text{ for } t = 1, \ldots, n,$$

where $\boldsymbol{X}_t \boldsymbol{\theta} - c = 0$ is the hyperplane separating the two classes. Since there is no clear separation in the data, slack variables ζ_t are defined to tolerate the wrong classifications. The tuning parameter K sets the total budget for the error. The dual problem of this optimization problem can be solved efficiently with quadratic programming. Based on the optimal separating hyperplane, the classification is estimated by $\tilde{I}_t = \text{sign}(\boldsymbol{X}_t \boldsymbol{\theta} - c)$.

The misclassification rate is derived by comparing this estimate with the classification $\hat{I}_t^*, t = 1, \ldots, n$. For the cases with more than two states, one can obtain the optimal separation for each state (vs. all other states), hence obtaining $k - 1$ different separating hyperplanes. The overall performance of the candidate is the sum of the individual performances. It is possible, though more complicated, to require the separating hyperplanes to be parallel. Since we have an additional step for refinement, we choose to use the simpler procedure.

4.3. SVM-CUSUM

For a set of m variables $\{X_{1t}, \ldots, X_{mt}\}$, we can evaluate the selected optimal linear combination Z_t (from SVM) either by the misclassification rate r, or by a CUSUM measure CUSUM(Z_t) as defined in (7) using the estimated probability $\hat{P}(I_t = i)$. SVM uses the hard-decision estimate $\hat{I}_t$ and every sample is treated with equal weight. However, the samples with extreme probabilities, say 0.1 or 0.9, should be more informative than those with probabilities around 0.5 for a two-state problem. This can be solved by combining SVM with CUSUM measure. We call this combined method SVM-CUSUM.

These three methods can significantly reduce the size of potential candidate pool of threshold variables. The remaining few candidates are further examined in the third step.

5. TD-SAR Model Estimation and Model Selection

In the final step, TD-SAR model is fitted to a small number of threshold variable candidates selected from the searching step, and the final model is chosen based on a model selection criterion.

5.1. TD-SAR Model Estimation

Consider the two-state TD-SAR model in (1). Suppose the threshold variable candidate selected from the previous stage is in the form of $Z_t = \sum_{i=0}^{m} \beta_i X_{it}$, and assume the switching mechanism is the logistic model (3). Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$, $\boldsymbol{X} = (1, X_{1t}, \ldots, X_{mt})$, $\boldsymbol{\phi} = (\boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}^{(2)})$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2)^T$, $\boldsymbol{I} = (I_{p+1}, \ldots, I_n)^T$, and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_m)^T$. Then the joint posterior distribution is

$$p(\boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{I}, \boldsymbol{\beta} \mid \boldsymbol{Y}, \boldsymbol{X}) \propto p(\boldsymbol{Y} \mid \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{I}) p(\boldsymbol{I} \mid \boldsymbol{\beta}, \boldsymbol{X}) p(\boldsymbol{\phi}) p(\boldsymbol{\sigma}^2) p(\boldsymbol{\beta}),$$

where $p(\boldsymbol{Y} \mid \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{I})$ is the same as in (6), and

$$p(\boldsymbol{I} \mid \boldsymbol{\beta}, \boldsymbol{X}) = \prod_{t=p+1}^{n} \prod_{i=1}^{k} g_i(\boldsymbol{X}_t, \boldsymbol{\beta})^{I(I_t=i)},$$

where $g_i(\cdot)$ is defined in (3). Again, standard conjugate priors can be used for $p(\boldsymbol{\phi})$, $p(\boldsymbol{\sigma}^2)$ and $p(\boldsymbol{\beta})$. We use the MCMC algorithm to draw samples from the above posterior distribution. Detailed implementation is given in the Appendix.

5.2. Model Selection

Under likelihood-based inference, model selections for TAR models are usually done with information criteria such as AIC and BIC (Tong (1990)). Gonzalo and Pitarakis (2002) introduced inference-based model selection procedure for simple and multiple threshold models. For SAR models, Smith, Naik and Tsai (2006) proposed model selection criteria using Kullback-Leiber divergence.

Under the Bayesian framework we have adopted here, model selections are often based on the Bayes factor (Kass and Raftery (1995) and Berger and Pericchi (1998)) or the posterior probability of each possible model under consideration (e.g., Robert and Casella (2004) and the references therein). Due to the large number of candidate threshold variables and model choices under consideration., we use a Posterior BIC (PBIC) for model selection. PBIC is defined as the average BIC value under the posterior distribution of the parameters:

$$E(\mathrm{BIC}(\boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{\beta}) \mid \boldsymbol{Y}, \boldsymbol{X}) = \int \mathrm{BIC}(\boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{\beta}) p(\boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{\beta} \mid \boldsymbol{Y}, \boldsymbol{X}) d\boldsymbol{\phi}\, d\boldsymbol{\sigma}^2\, d\boldsymbol{\beta}.$$

For a two state model, we have

$$\mathrm{BIC}(\boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{\beta}) = -2 \sum_{t=p+1}^{n} \log\left(p_{1t} C_{t1} + p_{2t} C_{t2}\right) + k \log\left(n - p\right),$$

in which,

$$p_{1t} = \frac{\exp(\boldsymbol{X}_t \boldsymbol{\beta})}{1 + \exp(\boldsymbol{X}_t \boldsymbol{\beta})}, \qquad p_{2t} = \frac{1}{1 + \exp(\boldsymbol{X}_t \boldsymbol{\beta})},$$

$$C_{ti} = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(Y_t - \boldsymbol{Y}_{t-1}\boldsymbol{\phi}^{(i)})^2}{2\sigma_i^2}\right), \quad i = 1, 2,$$

and k is the number of the parameters in the model. This can be easily obtained by averaging the BIC values for all the samples of $(\boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{\beta})$ generated from the Gibbs sampler, i.e.,

$$\mathrm{PBIC} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{BIC}(\boldsymbol{\phi}^{(i)}, \boldsymbol{\sigma}^{2(i)}, \boldsymbol{\beta}^{(i)}).$$

Other model comparison procedures, such as out-sample forecasting comparison, can also be used. Such procedures are not automatic, hence might be used when the number of candidate models is greatly reduced. Bayesian model averaging (e.g., Hoeting et al. (1999)) can be used as well.

6. Simulation

In this section we present some simulation results to demonstrate the effectiveness of the proposed algorithms. The factors under consideration include the number of states, the number of threshold variable candidates, and the characteristics of the candidates.

6.1. Experimental Design

Following are the components used for the simulation.

1. AR model. Three models are considered.
   (a) M2W: Two-state model with $\varepsilon_{it} \sim N(0, \sigma_i^2)$, $\sigma_1^2 = 0.1, \sigma_2^2 = 0.05$.

$$Y_t = \begin{cases} -0.5 - 0.1Y_{t-1} + 0.4Y_{t-2} + \varepsilon_{1t} & I_t = 1, \\ 0.5 + 0.5Y_{t-1} - 0.5Y_{t-2} + \varepsilon_{2t} & I_t = 2. \end{cases}$$

   (b) M2N: Two-state model with $\varepsilon_{it} \sim N(0, \sigma_i^2)$, $\sigma_1^2 = 0.1, \sigma_2^2 = 0.05$.

$$Y_t = \begin{cases} -0.15 - 0.1Y_{t-1} + 0.4Y_{t-2} + \varepsilon_{1t} & I_t = 1, \\ 0.15 + 0.5Y_{t-1} - 0.5Y_{t-2} + \varepsilon_{2t} & I_t = 2. \end{cases}$$

Compared with the model M2W, the current one has a narrower margin on the constant items, and therefore it is harder to classify.

(c) M3: Three-state model with $\varepsilon_{it} \sim N(0, \sigma_i^2)$, $\sigma_1^2 = 0.1$, $\sigma_2^2 = 0.3$, $\sigma_3^2 = 0.05$.

$$Y_t = \begin{cases} -0.3 - 0.1Y_{t-1} + 0.4Y_{t-2} + \varepsilon_{1t} & I_t = 1, \\ -0.5Y_{t-1} + 0.7Y_{t-2} + \varepsilon_{2t} & I_t = 2, \\ 0.3 + 0.3Y_{t-1} - 0.3Y_{t-2} + \varepsilon_{3t} & I_t = 3. \end{cases}$$

2. Threshold variable candidates. Two models are used to generate the *true* threshold variable candidates:
   (a) T1: IID Standard Normal distribution $X_{1,t} \sim N(0, 1)$;
   (b) T2: AR(1) Model, which is $X_{2,t} = 0.6X_{2,t-1} + 0.8\varepsilon_j$.
   The threshold candidate set considered also includes 28 more variables listed in Table 1, with X_t being the true threshold variable used.

Table 1. Candidate variables used in the simulation.

(1) 5 U(0,1)	(2) N(0,1)	(3) 5 N(0,1)
(4) X_t + N(0,1)	(5) X_t + 5 N(0,1)	(6) $X_t^2 + X_t$
(7) X_t^2 + 0.5 X_t	(8) X_t^2 + 0.2 X_t + U(0,1)	(9) X_{t-1}
(10) X_{t-2}	(11) X_{t-3}	(12) X_{t-4}
(13) X_{t-5}	(14) X_{t-6}	(15) X_{t-1} + N(0,1)
(16) X_{t-2} + N(0,1)	(17) $X_t + X_{t-1}$	(18) X_t + 0.5 X_{t-1} + N(0,1)
(19) X_{t-1} + 0.5 X_{t-2}	(20) X_t - X_{t-2}	(21) X_{t-1} + 2 X_{t-2}
(22) X_{t-1}^2	(23) X_{t-2}^2	(24) X_{t-3}^2
(25) X_{t-1}^2 + 0.5 X_{t-1}	(26) $X_{t-2}^2 + X_{t-2}$	(27) X_{t-1}^2 + 0.5 X_{t-1}
(28) $X_{t-2}^2 + X_{t-1}$		

3. Switching mechanism. Three logistic models are considered.
   (a) L1: Binomial logistic model with a one-dimensional covariate; parameters (with intercept) are $\boldsymbol{\beta} = (-1, 4)^T$.
   (b) L2: Binomial logistic model with a two-dimensional covariate; parameters are $\boldsymbol{\beta} = (-1, 4, 2)^T$.
   (c) L3: Three-state logistic model with a one-dimensional covariate; parameters $\boldsymbol{\beta} = (\beta_1, \beta_2)$ are $\boldsymbol{\beta}_1 = (-1, 3)^T$ and $\boldsymbol{\beta}_2 = (-2, -4)^T$.

We consider eight different settings. In each simulation, we record the best five candidates according to the CUSUM values, the SVM classification rates, and the SVM-CUSUM values. If the true threshold variable is identified among the top five candidates, we mark it as a "success". In the third step, the TD-SAR model is applied on the best five candidates individually and their PBIC values are also calculated.

6.2. Results

Tables 2 and 3 summarize the results of the method for one-dimensional and two-dimensional settings, respectively.

Table 2. Simulation results for the one-dimensional settings.

Setting	Stage I (%)	CUSUM (%)	2nd best (%) (var)	PBIC (%)	2nd best (%) (var)	SVM (%)	2nd best (%) (var)	PBIC (%)	2nd best (%) (var)
M2W, T1, L1	88	100	90 (9)	100	60 (9)	100	60 (7)	100	60 (9)
M2W, T2, L1	89	100	50 (20)	100	80 (20)	100	60 (20)	100	90 (20)
M2N, T1, L1	76	100	40 (9)	100	20 (9)	100	30 (20)	100	40 (7)
M2N, T2, L1	76	100	40 (20)	100	30 (21)	100	30 (20)	100	30 (21)
M3, T1, L3	65	100	65 (10)	100	40 (9)	88	50 (9)	88	40 (10)
M3, T2, L3	70	100	55 (10)	100	60 (20)	83	35 (9)	83	45 (9)

Table 3. Simulation results for the two-dimensional settings.

Setting	Stage I (%)	SVM (%)	2nd best (%) (var)	PBIC (%)	2nd best (%) (var)	S-C (%)	2nd best (%) (var)	PBIC (%)	2nd best (%) (var)
M2N, (T1,T1), L2	88	88	35 (2,10)	88	35 (2,24)	95	30 (2,24)	95	35 (2,24)
M2N, (T1,T2), L2	89	93	25 (2,16)	93	30 (2,16)	95	15 (2,21)	95	20 (2,16)

Compared to the "true" state for each data point, the average classification rate in Stage (I) (classification stage) for the 40 data sets is shown in the second column of Tables 2 and 3. It is seen that the first stage provides a reasonable starting point for further analysis. In Table 2, column 3 shows the percentage of times the CUSUM algorithm includes the true threshold variable (or the two variables whose linear combination makes up the true threshold variables) among the top five variables. Column 4 shows the most frequently chosen variable (other than the true threshold variable) in the top five and the percentage of times it was chosen. The number in parentheses refers to the variable number listed in Table 1. This reveals what would happen if the true threshold variable were not included in the candidate set. From the results it seems that the search procedure tries to find a proxy variable that is highly correlated to the true threshold variable. But the choice is not persistent as the percentage is relatively low. Column 5 presents the percentage of times PBIC correctly identifies the true threshold variable in the final stage, based on the five candidates that CUSUM procedure identified in step 2. Column 6 shows the second best variable. Columns 7 to 10 are the same as columns 3-6, except the SVM is used as the search method. Table 3 reads the same way, where S-C denotes the SVM-CUSUM measure in Column 7.

Overall, the simulation shows that the CUSUM method works for checking one-dimensional variables. As expected, the SVM-CUSUM is slightly better than

the simple SVM. Another observation in our experiment is that the PBIC is a good model selection criterion.

7. Real Examples

We present two examples of U.S. economic indicators: nonfarm payroll numbers and the unemployment rate. The data are from the Bureau of Labor Statistics (*www.bls.gov*). The first series runs from January, 1939 to March, 2004, the second from January, 1948 to March, 2004. Both are seasonally adjusted.

7.1. U.S. nonfarm payroll numbers

Nonfarm payroll number is an important economic series. Its unexpected change is linked with large market volatility. First we transformed the original monthly data to quarterly differences:

$$Q_t = \frac{P_{3(t-1)+1} + P_{3(t-1)+2} + P_{3(t-1)+3}}{3}, \quad \text{and} \quad Y_t = \frac{Q_t - Q_{t-1}}{500,000},$$

for $t = 1, \ldots, 260$. Here, P_t is the monthly payroll number, Q_t represents the quarterly average, and Y_t represents the quarterly difference. We let the unit of Y_t be 500,000. The series Y_t is shown in Figure 3.

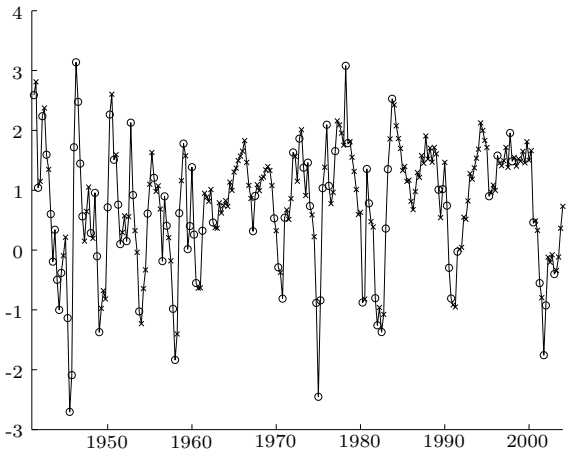

Figure 3. The states estimated for the U.S. nonfarm payroll numbers in the first stage ("o": State 1; "x": State 2).

For the sake of comparison, we fit the data with a linear ARMA model. Analysis suggests an AR(2) model $Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$. The MLE estimates and their standard errors (in parentheses) are $\hat{\phi}_0 = 0.2233(0.0396)$,

$\hat{\phi}_1 = 1.0054(0.0530)$, and $\hat{\phi}_2 = -0.3100(0.0537)$. It can be verified that the estimated model is stationary. The sum of the squared error (SSE) is 93.13.

After trying several model settings, we considered a two-state TD-SAR model with AR(2) in both states, and a logistic link function for the threshold driving mechanism. Specifically,

$$Y_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} Y_{t-1} + \phi_2^{(1)} Y_{t-2} + \varepsilon_1 & I_t = 1, \\ \phi_0^{(2)} + \phi_1^{(2)} Y_{t-1} + \phi_2^{(2)} Y_{t-2} + \varepsilon_2 & I_t = 2, \end{cases}$$

$$P(I_t = 1) = \frac{e^{Z_t}}{1 + e^{Z_t}},$$

where $Z_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_m X_{mt}$.

In this study, we limited the threshold variable candidate set to eight lag variables and their squares. We considered linear combination up to three variables in the candidate set, resulting totally 696 possible candidates. We also considered other economic indicators such as GDP, short-term Treasury rates and inflation rates, but did not find any of them to be good candidates.

Classification step: In the first classification step, 20,000 samples from the posterior distribution are obtained with the Gibbs sampler, with the first 10,000 samples discarded to reduce the effect of initial values. The posterior means of the AR coefficients are $\hat{\phi}_0^{(1)} = 0.1356$, $\hat{\phi}_1^{(1)} = 1.0535$, $\hat{\phi}_2^{(1)} = -0.1953$, $\hat{\phi}_0^{(2)} = 0.2873$, $\hat{\phi}_1^{(2)} = 1.9842$, and $\hat{\phi}_2^{(2)} = -0.4259$.

In Figure 3 we show the estimated states $\hat{I}_t$, labeled by 'o' and 'x'. It does not show a clear pattern, though most of the small Y_t's corresponds to State 1. We calculate two goodness-of-fit measures.

1. Hard SSE: $\sum_{t=p+1}^{n} (Y_t - \boldsymbol{Y}_{t-1} \hat{\boldsymbol{\phi}}^{(\hat{I}_t)})^2$, where $\hat{I}_t = 1$, if $\hat{p}_{t1} \geq 0.5$.
2. Soft SSE: $\sum_{t=p+1}^{n} \left(\hat{p}_{t1} (Y_t - \boldsymbol{Y}_{t-1} \hat{\boldsymbol{\phi}}^{(1)})^2 + \hat{p}_{t2} (Y_t - \boldsymbol{Y}_{t-1} \hat{\boldsymbol{\phi}}^{(2)})^2 \right)$.

For our first stage classification, the hard SSE is 80.77 and the soft SSE is 87.71.

Searching for an appropriate threshold variable: CUSUM is used to evaluate all one-dimensional candidates. Simple SVM and SVM-CUSUM are used to search and evaluate all linear combinations of up to three variables in the candidate set. Table 4 shows the best three candidates for each setting, with their corresponding CUSUM, SVM or SVM-CUSUM.

Full model estimation and model selection: Using each of the 16 different combinations in Table 4, we fit the corresponding TD-SAR models. Column 6 in Table 4 shows their PBIC values. The model using Y_{t-1}^2 as the threshold candidate was not stable, hence its PBIC is missing. The minimum PBIC is

reached with the linear combination of (Y_{t-2}, Y_{t-3}^2). The posterior means and the posterior standard deviations (in parentheses) of the parameters are: $\hat{\phi}_0^{(1)} = -0.31(0.15)$, $\hat{\phi}_1^{(1)} = 0.93(0.12)$, $\hat{\phi}_2^{(1)} = -0.95(0.19)$, $\hat{\sigma}_1^2 = 0.52(0.11)$, $\hat{\phi}_0^{(2)} = 0.32(0.10)$, $\hat{\phi}_1^{(2)} = 0.97(0.08)$, $\hat{\phi}_2^{(2)} = -0.29(0.08)$, and $\hat{\sigma}_2^2 = 0.24(0.04)$.

Table 4. Three selection methods applied to the U.S. payroll data (The TD-SAR model does not converge for Candidate 6).

No.	Dim.	Methods	Candidates	Target value	PBIC
1	1	CUSUM	(Y_{t-1})	8.86	498.55
2			(Y_{t-2})	8.80	478.04
3			(Y_{t-4})	8.49	490.17
4	1	Misclassification	(Y_{t-1})	0.3294	498.55
5			(Y_{t-2})	0.3254	478.04
6			(Y_{t-1}^2)	0.3294	–
7	2	Misclassification	(Y_{t-1}, Y_{t-2}^2)	0.3175	497.97
8			(Y_{t-2}, Y_{t-1}^2)	0.3175	491.86
9			(Y_{t-4}, Y_{t-1}^2)	0.3175	491.63
10	2	SVM-CUSUM	(Y_{t-2}, Y_{t-3}^2)	9.59	**476.42**
11			(Y_{t-2}, Y_{t-4}^2)	9.93	490.49
12			(Y_{t-4}, Y_{t-8})	9.90	493.82
13	3	Misclassification	$(Y_{t-2}, Y_{t-1}^2, Y_{t-8}^2)$	0.3016	478.90
14			$(Y_{t-1}, Y_{t-5}, Y_{t-4}^2)$	0.3016	504.28
15			$(Y_{t-3}, Y_{t-4}, Y_{t-1}^2)$	0.3056	530.95
16	3	SVM-CUSUM	$(Y_{t-1}, Y_{t-4}^2, Y_{t-6}^2)$	10.86	485.28
17			$(Y_{t-2}, Y_{t-5}, Y_{t-4}^2)$	11.44	478.77
18			$(Y_{t-2}, Y_{t-8}, Y_{t-4}^2)$	10.48	486.93

The threshold variable driven mechanism is

$$P(I_t = 1) = \frac{\exp(\beta_0 + \beta_1 Y_{t-2} + \beta_2 Y_{t-3}^2)}{1 + \exp(\beta_0 + \beta_1 Y_{t-2} + \beta_2 Y_{t-3}^2)}, \tag{8}$$

where the posterior mean of the coefficient is $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T = (2.6315, -4.4887, -3.3738)^T$. Hence the estimated threshold variable is $Z_t = 2.6315 - 4.4887 Y_{t-2} - 3.3738 Y_{t-3}^2$. We plot the probability $P(I_t = 1) = \exp(Z_t)/(1 + \exp(Z_t))$ vs Z_t in Figure 4. It can be seen that the switching is relatively sharp, with randomness in a very narrow range.

The left panel in Figure 5 shows the classification on the scatter plot of Y_{t-2} and Y_{t-3}, along with "optimal" (0.5 probability) linear combination that separates the classes. The right panel in Figure 5 shows the final estimated states for the payroll numbers. It can be seen that State 1 mainly includes those points

with small values (on the bottom) and the following upward curves, while State 2 mainly includes those points with large values (on the top) and the following downward curves. It is noted that the current classification is quite different from that in Figure 3, showing the effect of the threshold variable.

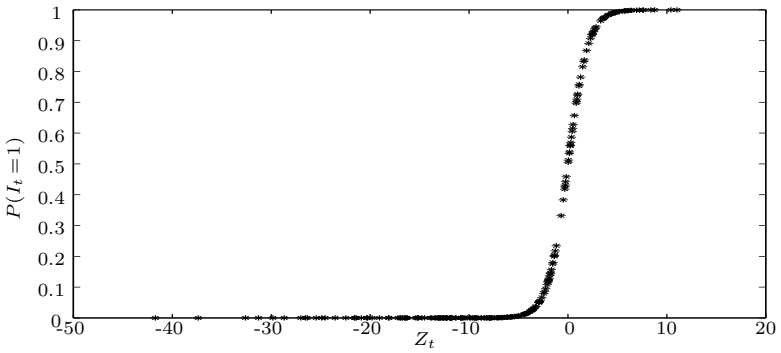

Figure 4. U.S. nonfarm payroll numbers: the switching driving mechanism as a logistic regression.

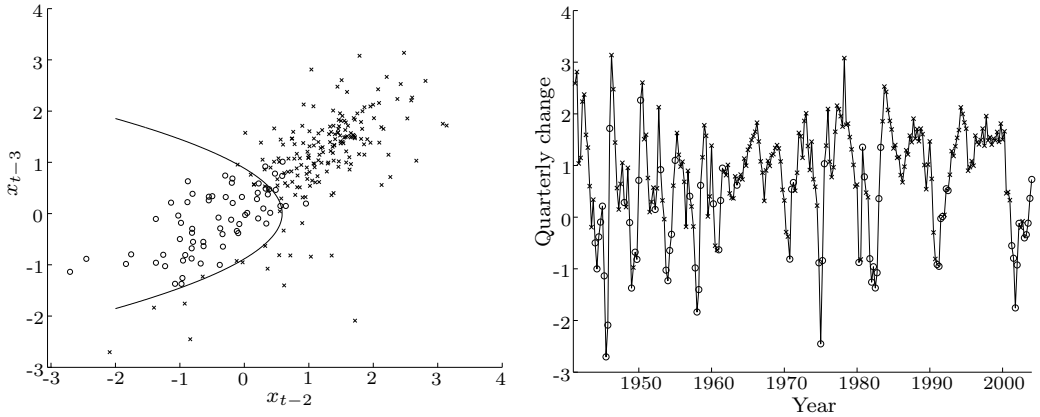

Figure 5. U.S. nonfarm payroll results. Left Panel: The estimated states and the threshold variable ("o": State 1; "x": State 2). The solid line is the 0.5 probability line. Right Panel: The estimated states in the original time series. ("o": State 1; "x": State 2).

Model comparison: Table 5 shows the hard and soft SSE for five different models, including the linear AR model, the standard TAR model, the TAR model with a threshold in the form of a linear combination, the standard SAR model, and the TD-SAR model. The smallest hard SSE for a TAR model comes from the one with the quadratic threshold. And it shows a large improvement over the

linear AR model. When using the SAR(2) model, the hard SSE is also reduced. Due to its bi-modal character, the soft SSE is calculated for the SAR model. The TD-SAR model further reduces the hard SSE and soft SSE, compared to the TAR(2) model and the SAR(2) model.

Table 5. Model comparison for nonfarm payroll numbers.

Model	Threshold Variable	Hard SSE	Soft SSE
AR(2)	-	93.13	-
TAR(2)	$Z_t = Y_{t-2} - c$	80.55	-
TAR(2)	$Z_t = \beta_0 + \beta_1 Y_{t-2} + \beta_2 Y_{t-3}^2$	77.05	-
SAR(2)	I_t IID	80.77	87.71
TD-SAR(2)	Equation (8)	64.37	75.08

7.2. U.S. unemployment rate

Montgomery et al. (1998) used a Markov Chain SAR model and a TAR model to analyze the U.S. unemployment rate series. Here we apply the proposed modeling procedure and the TD-SAR model to the same series (slightly longer). Following Montgomery et al. (1998), we obtain the quarterly differences (as with the nonfarm payroll series). Figure 6 shows the transformed series.

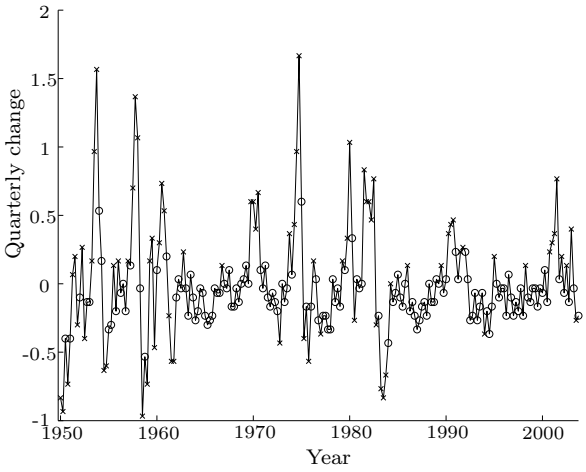

Figure 6. The states estimated for the U.S. unemployment rate in the first stage ("o": State 1; "x": State 2).

Again, following Montgomery et al. (1998), we use a two-state TD-SAR model with AR order $p = 2$ for both states. The threshold variable candidate set includes lag 1 to lag 8 of the observed series, their squares and linear combinations, resulting in a total of 136 candidates.

Classification step. We fit a two-state independent SAR(2) model. The posterior means of the AR coefficients are $\hat{\phi}_0^{(1)} = -0.0617$, $\hat{\phi}_1^{(1)} = 0.3471$, $\hat{\phi}_2^{(1)} = 0.0029$, $\hat{\phi}_0^{(2)} = 0.0567$, $\hat{\phi}_1^{(2)} = 1.0169$, and $\hat{\phi}_2^{(2)} = -0.2680$. The states of the samples (see Figure 6) and the probabilities of the states are also estimated. The hard SSE is 14.58, and soft SSE is 14.59.

Searching for an appropriate threshold variables. The searching results are shown in Table 6. There are 11 different candidates emerging in this stage.

Table 6. Three selection methods applied on the U.S. unemployment rate data.

No.	Dim.	Methods	Candidates	Target value	PBIC
1	1	CUSUM	(Y_{t-1})	10.19	111.78
2			(Y_{t-2})	10.63	**105.32**
3			(Y_{t-4}^2)	9.93	108.17
4	1	Misclassification	(Y_{t-4})	0.3472	112.13
5			(Y_{t-4}^2)	0.3241	108.17
6			(Y_{t-7}^2)	0.33.80	117.59
7	2	Misclassification	(Y_{t-1}, Y_{t-4}^2)	0.3148	**101.93**
8			(Y_{t-4}, Y_{t-4}^2)	0.3102	114.50
9			(Y_{t-4}, Y_{t-7}^2)	0.3148	113.66
10	2	SVM-CUSUM	(Y_{t-1}, Y_{t-2})	11.25	**104.98**
11			(Y_{t-1}, Y_{t-2}^2)	11.16	111.85
12			(Y_{t-2}, Y_{t-1}^2)	11.84	107.52

Table 7. Model comparison for U.S. unemployment rate data.

Model	Threshold Var.	Hard SSE	Soft SSE
TAR(2)	$Z_t = Y_{t-2} - c$	17.57	-
TAR(2)	$Z_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2}$	17.12	-
TAR(2)	$Z_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-4}^2$	18.27	-
SAR(2)	I_t IID	14.74	14.59
TD-SAR(2)	Logistic p_t (on Y_{t-2})	14.30	18.09
TD-SAR(2)	Logistic p_t (on Y_{t-1}, Y_{t-2})	14.98	17.22
TD-SAR(2)	Logistic p_t (on Y_{t-1}, Y_{t-4}^2)	17.28	17.99

Full model estimation and model selection. Using these different combinations, we fit the TD-SAR model. Column 6 in Table 6 shows their PBIC values. The attractive models are associated with (Y_{t-1}, Y_{t-4}^2), (Y_{t-1}, Y_{t-2}), and Y_{t-2} as threshold variables. The hard SSE and soft SSE for the related models are shown in Table 7. Although the variable pair (Y_{t-1}, Y_{t-4}^2) has the least PBIC, its SSEs are not satisfying. Instead, the variable pair (Y_{t-1}, Y_{t-2}) has the second least BIC, and both of its SSEs are better than the first candidate. Therefore we choose the

threshold variable as $Z_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2}$, and estimate the final model. The posterior means and the posterior standard deviation (in parentheses) of the parameters are $\hat{\phi}_0^{(1)} = -0.03(0.02)$, $\hat{\phi}_1^{(1)} = 0.51(0.11)$, $\hat{\phi}_2^{(1)} = 0.03(0.11)$, $\hat{\sigma}^{(1)} = 0.05(0.007)$, $\hat{\phi}_0^{(2)} = 0.32(0.17)$, $\hat{\phi}_1^{(2)} = 0.75(0.19)$, $\hat{\phi}_2^{(2)} = -0.68(0.24)$, and $\hat{\sigma}^{(2)} = 0.21(0.06)$. The logistic model driven mechanism is:

$$P(I_t = 1) = \frac{\exp(\beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2})}{1 + \exp(\beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2})},$$

and the estimated threshold variable $Z_t = 2.0604 - 4.0579 Y_{t-1} - 4.1071 Y_{t-2}$. We plot the probability $P(I_t = 1) = \exp(Z_t)/(1 + \exp(Z_t))$ vs Z_t in Figure 7. The classification by Y_{t-1}, Y_{t-2} is shown in the left panel of Figure 8. The hard SSE

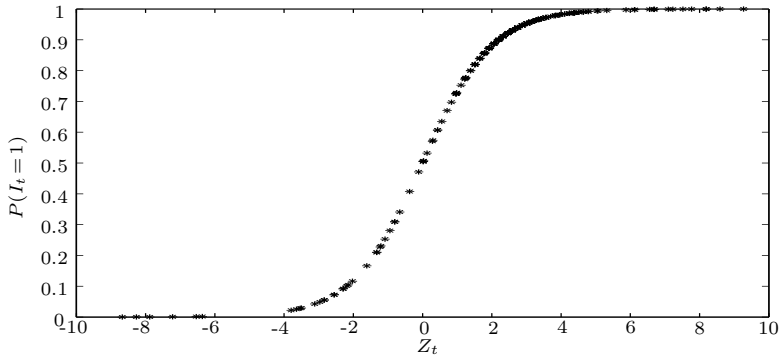

Figure 7. U.S. unemployment rate: the switching driving mechanism as a logistic regression.

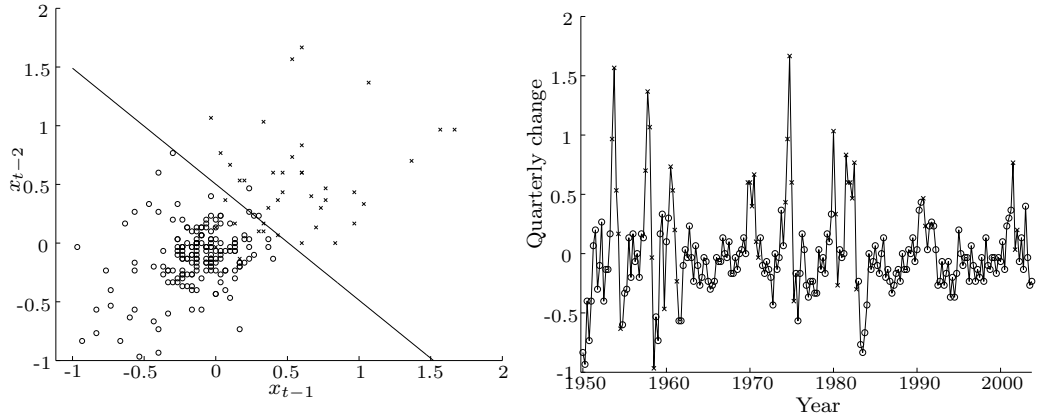

Figure 8. U.S. unemployment rate results. Left Panel: The estimated states and the threshold variable ("o": State 1; "x": State 2). The solid line is the 0.5 probability line. Right Panel: The estimated states in the original time series. ("o": State 1; "x": State 2).

is 14.98, and the soft SSE is 17.22. The right panel in Figure 8. shows the estimated states for the unemployment numbers. States 1 and 2 are labeled with "o" and "x". It can be seen that State 2 corresponds to those observations with value larger than 0.45, and the subsequent downward curves.

8. Summary

In this paper we proposed a new class of switching time series model, the threshold variable driven switching autoregressive models. It enjoys having an observable threshold variable (up to a set of unknown but estimable parameters) as the driving forces of model switching, as in TAR models, yet retains certain flexibility in the switching mechanism enjoyed by the SAR model. Stationary properties of the model are studied, estimation procedures are developed under a Bayesian framework, and a model building procedure is proposed so that the model is applicable in real practice. Specifically we consider the problem of threshold variable determination with a large set of candidate variables. A fast search algorithm is proposed to handle this difficult problem. Though heuristic, it is shown to perform reasonably well in examples. This search algorithm is useful for standard TAR model building as well.

Simulation and real examples have shown that the new model is useful in many cases. It is not clear if there is any 'observable signal' that can provide indication when a TD-SAR model should be used in any specific situation. We rely on model comparison procedures, as described in Section 5.2, to determine whether the new model, a standard TAR model, or a standard SAR model should be used. In fact, a standard TAR model or a standard SAR model should always be considered as a parsimonious alternative to the TD-SAR model. On the other hand, if a standard TAR or SAR model is considered for modeling a time series, a TD-SAR model should be considered as well.

In the above approach we have also assumed that the AR order in each state and the number of states are known. These restrictions can be easily relaxed; model selection criteria can be used to choose the AR order and the number of states.

Acknowledgement

This research is partially supported by NSF grant DMS 0244541 and NIH grant R01 Gm068958. The authors wish to thank the editors and three anonymous referees for their very helpful comments and suggestions that significantly improved the paper.

Appendix

The appendix is contained in a supplemental document available online at the *Statistica Sinica* website: http://www3.stat.sinica.edu.tw/statistica/

References

Berger, J. and Pericchi, L. (1998). Accurate and stable Bayesian model selection: the median intrinsic Bayes factor. *Sankhyā* **B16**, 1-18.

Boucher, T. and Cline, D. (2007). Stability of cylic threshold and threshold-like autoregressive time series models. *Statist. Sinica*. Accepted.

Chan, K. S. and Tong, H. (1986). On estimating thresholds in autoregressive models. *J. Time. Ser. Anal.* **7**, 179-190.

Chen, C. W. S. and So, M. K. P. (2006). On a threshold teteroscedastic model. *International Journal of Forecasting* **22**, 73-89.

Chen, R. (1995). Threshold variable selection in open-loop threshold autoregressive models. *J. Time. Ser. Anal.* **16**, 461-481.

Chen, R. and Liu, J. S. (1996). Predictive updating methods with application to Bayesian classification. *J. Roy. Statist. Soc. Ser. B* **58**, 397-415.

Chong, T. T.-L. (2001). Structural change in AR(1) models. *Econom. Theory* **17**, 87-155.

Chong, T. T.-L. and Yan, I. K.-M. (2005). Threshold model with multiple threshold variables and application to financial crises. Manuscript, The Chinese University of Hong Kong.

Cline, D. and Pu, H. (1999). Geometric ergodicity of nonlinear time series. *Statist. Sinica* **9**, 1103-1118.

Cristianini, N. and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines*. Cambridge University Press, Cambridge.

Gelfand, A. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398-409.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distribution and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721-741.

Gonzalo, J. and Martínez, O. (2004). Large shocks vs small shocks. Ph.D. Dissertation, U. Carlos III de Madrid.

Gonzalo, J. and Pitarakis, J. (2002). Estimation and model selection based inference in single and multiple threshold models. *J. Econometrics* **110**, 319-352.

Gooijer, J. G. D. (1998). On threshold moving-average models. *J. Time. Ser. Anal.* **19**, 1-18.

Hamilton, J. (1989). A new approach to the economics analysis of nonstationary time series and the business cycle. *Econometrica* **57**, 357-384.

Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, New York.

Hoeting, J. A., Madigan, A. D., Raftery, A. E. and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial (with discussion). *Statist. Sci.* **144**, 382-417.

Huerta, G., Jiang, W. and Tanner, M. A. (2003). Time series modeling via hierarchical mixtures. *Statist. Sinica* **13**, 1097-1118.

Kass, R. and Raftery, A. E. (1995). Bayes factors. *J. Amer. Statist. Assoc.* **90**, 773-795.

Li, C. W. and Li, W. (1996). On a double-threshold autoregressive heteroscedastic time series model. *J. Appl. Econ.* **11**, 253-274.

Li, W. K. and Lam, K. (1995). Modeling asymmetry in stock returns by a threshold autoregressive conditional heteroscedastic model. *The Statistician* **44**, 333-341.

Liu, J. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag, New York.

McCulloch, R. and Tsay, R. (1994a). Bayesian analysis of autoregressive time series via the Gibbs sampler. *J. Time. Ser. Anal.* **15**, 235-250.

McCulloch, R. and Tsay, R. (1994b). Statistical analysis of economic time series via Markov switching models. *J. Time. Ser. Anal.* **15**, 523-539.

Moeanaddin, R. and Tong, H. (1988). A comparison of likelihood ratio and CUSUM test for threshold autoregression. *The Statistician* **37**, 213-225.

Montgomery, A., Zarnowitz, V., Tsay, R. and Tiao, G. (1998). Forecasting the U.S. unemployment rate. *J. Amer. Statist. Assoc.* **93**, 478-493.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer-Verlag, New York.

Sáfadi, T. and Morettin, P. A. (2000). Bayesian analysis of threshold autoregressive moving average models. *Indian J. Statist.* **62**, 353-371.

Smith, A., Naik, P. and Tsai, C. (2006). Markov-switching model selection using Kullback-Leibler divergence. *J. Econometrics*. Accepted.

So, M. K. P., Li, W. K. and Lam, K. (2002). A Threshold stochastic volatility model. *J. Forecasting* **21**, 473-500.

Taylor, W. (2000). Change point analysis: a powerful new tool for detecting changes. Web: www.variation.com/cpa/tech/changepoint.html.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *J. Amer. Statist. Assoc.* **89**, 208-218.

Tjostheim, D. (1990). Nonlinear time series and Markov chains. *Adv. Appl. Probab.* **22**, 587-611.

Tong, H. (1983). Threshold Models in Non-linear Time Series Analysis. *Lecture Notes in Statistics*, 21.

Tong, H. (1990). *Non-linear Time Series: a Dynamical System Approach*. Oxford Univ. Press, New York.

Tong, H. and Lim, K. (1980). Threshold autoregression, limit cycles and cyclical data. *J. Roy. Statist. Soc. Ser. B* **42**, 245-292.

Tsay, R. (1989). Testing and modeling threshold autoregressive processes. *J. Amer. Statist. Assoc.* **84**, 231-240.

Tsay, R. (1998). Testing and modeling multivariate threshold models. *J. Amer. Statist. Assoc.* **93**, 1188-1202.

Tweedie, R. (1975). Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stochastic Process. Appl.* **3**, 385-403.

van Dijk, D., Teräsvirta, T. and Franses, P. H. (2002). Smooth transition autoregressive models - a survey of recent developments. *Econom. Rev.* **21**, 1-47.

Vapnik, V. (1998). *Statistical Learning Theory*. John Wiley, New York.

Watier, L. and Richardson, S. (1995). Modeling of an epidemiological time series by a threshold autoregressive model. *The Statistician* **44**, 353-364.

Zakoian, J. M. (1994). Threshold heteroshedastic models. *J. Econom. Dynam. Control* **18**, 931-955.

Department of Information and Decision Sciences, University of Illinois at Chicago, Chicago, IL 60607, U.S.A.

E-mail: wu_shenlin@yahoo.com

Department of Information and Decision Sciences, University of Illinois at Chicago, Chicago, IL 60607, U.S.A.

Department of Business Statistics and Econometrics, Peking University, Beijing, 100081, China.

E-mail: rongchen@uic.edu

(Received April 2005; accepted August 2006)

Statistica Sinica **17**(2007): Supplement, S36∼S38

THRESHOLD VARIABLE DETERMINATION AND THRESHOLD VARIABLE DRIVEN SWITCHING AUTOREGRESSIVE MODELS

Senlin Wu[1] and Rong Chen[1,2]

[1] *University of Illinois at Chicago and* [2] *Peking University*

Supplementary Material

Appendix

A.1. Implementation for the classification step

We illustrate the procedure with a two-state model. The general k-state models can be treated similarly.

We use the following prior distributions:

1. The coefficient vectors $\boldsymbol{\phi}^{(1)}$ and $\boldsymbol{\phi}^{(2)}$ follows truncated normal distribution, i.e., $(\boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}^{(2)}) \sim N(\boldsymbol{\mu}_0^{(1)}, \boldsymbol{\Sigma}_0^{(1)}) N(\boldsymbol{\mu}_0^{(2)}, \boldsymbol{\Sigma}_0^{(2)}) I(\phi_0^{(1)} < \phi_0^{(2)})$. The constraint on the coefficients $\phi_0^{(1)}$ and $\phi_0^{(2)}$ is used to avoid ambiguity. Note that the model with $\boldsymbol{\phi}_*^{(1)} = \boldsymbol{\phi}^{(2)}$, $\boldsymbol{\phi}_*^{(2)} = \boldsymbol{\phi}^{(1)}$ and $I_t^* = 3 - I_t$ is identical to the original model. Without the constraint, there will be two identical modes in the posterior distribution.
2. The variance σ_i^2 follows an inverse χ^2 distribution, i.e., $\sigma_i^2 \sim \chi^{-2}_{\sigma_0^2, \gamma_0}$, $i = 1, 2$.
3. The state I_t follows a Bernoulli distribution, i.e., $P(I_t = 1) = p_0$, $t = 1, \ldots, n$.

We use 'nearly' uninformative priors with the following parameters:

$$\boldsymbol{\phi}_0 = 0, \quad \boldsymbol{\Sigma}_0 \to \infty, \quad \gamma_0 = 1, \quad \sigma_0^2 \to 0 \quad p_0 = 0.5, \quad t = 1, \ldots, n.$$

Based on the likelihood function and the priors, standard calculation (e.g., Chen and Liu (1996)) yields the following conditional distributions:

1. For the coefficient $\boldsymbol{\phi}$:

$$\begin{aligned} p(\boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}^{(2)} \mid \boldsymbol{\sigma}^2, \boldsymbol{I}, \boldsymbol{Y}) &\propto p(\boldsymbol{Y} \mid \boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}^{(2)}, \boldsymbol{\sigma}^2, \boldsymbol{I}) p(\boldsymbol{\phi}), \\ &\propto N(\boldsymbol{\mu}_1^*, \Sigma_1^*) N(\boldsymbol{\mu}_2^*, \Sigma_2^*) I(\phi_0^{(1)} < \phi_0^{(2)}), \end{aligned} \tag{9}$$

where, for $i = 1, 2$,

$$\boldsymbol{\mu}_i^* = \left(\sum_{t: I_t = i} \boldsymbol{Y}_{t-1}^T \boldsymbol{Y}_{t-1} \right)^{-1} \sum_{t: I_t = i} \boldsymbol{Y}_{t-1}^T Y_t, \quad \Sigma_i^* = \sigma_i^2 \left(\sum_{t: I_t = i} \boldsymbol{Y}_{t-1}^T \boldsymbol{Y}_{t-1} \right)^{-1}.$$

It is still a truncated multivariate normal distribution.

2. For the variance $\sigma_i^2, i = 1, 2$:

$$p(\sigma_i^2 \mid \boldsymbol{\phi}, \boldsymbol{I}, \boldsymbol{Y}) \propto p(\boldsymbol{Y} \mid \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{I}) p(\sigma_i^2),$$
$$\propto \chi^{-2}\left(\frac{\gamma_0\sigma_0^2 + \sum_{t=p+1}^{n} I(I_t = i)(Y_t - \boldsymbol{Y}_{t-1}\boldsymbol{\phi}^{(i)})^2}{n_i + \gamma_0}, n_i + \gamma_0\right) \quad (10)$$

in which, n_i is the number of the observations in state i. This is still an inverse χ^2 distribution.

3. For the discrete $I_t, t = p+1, \ldots, n, i = 1, 2$:

$$p(I_t = i \mid \boldsymbol{\phi}, I_{[-t]}, \boldsymbol{Y}) \propto p(\boldsymbol{Y} \mid \boldsymbol{\phi}, \boldsymbol{\sigma}^2, I_t = i) p(I_t = i),$$
$$\propto \frac{1}{\sigma_i} \exp\left(-\frac{(Y_t - \boldsymbol{Y}_{t-1}\boldsymbol{\phi}^{(i)})^2}{2\sigma_i^2}\right).$$

We use the posterior means to estimate $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}^2$, use the posterior mode to estimate $\boldsymbol{I}$. We run the Gibbs sampler M+N cycles, discard the first M cycles and keep the results of the last N cycles.

A.2. Implementation for the model estimation step

The assumptions about the priors of $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}^2$ are the same as in the model in the first stage. Additionally, the prior of the coefficient of the logistic regression $\boldsymbol{\beta}$ is assumed to be uninformative normal distribution. The conditional posteriors are calculated here.

1. For the coefficient $\boldsymbol{\phi}$, $p(\boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}^{(2)} \mid \boldsymbol{\sigma}^2, \boldsymbol{I}, \boldsymbol{\beta}, \boldsymbol{Y}, \boldsymbol{X}) \propto p(\boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}^{(2)} \mid \boldsymbol{\sigma}^2, \boldsymbol{I}, \boldsymbol{Y})$, which is the same as (9).
2. For the variance σ_i^2, $p(\sigma_i^2 \mid \boldsymbol{\phi}, \boldsymbol{I}, \boldsymbol{\beta}, \boldsymbol{Y}, \boldsymbol{X}) \propto p(\sigma_i^2 \mid \boldsymbol{\phi}, \boldsymbol{I}, \boldsymbol{Y})$, which is the same as (10).
3. For the coefficients $\boldsymbol{\beta}$ and the state $\boldsymbol{I}$, we draw them jointly, using

$$p(\boldsymbol{\beta}, \boldsymbol{I} \mid \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{Y}, \boldsymbol{X}) \propto p(\boldsymbol{I} \mid \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{Y}, \boldsymbol{X}) p(\boldsymbol{\beta} \mid \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{Y}, \boldsymbol{X}).$$

That is, we draw $\boldsymbol{\beta}$ first from the marginal distribution, with $\boldsymbol{I}$ integrated out. Then $\boldsymbol{I}$ is drawn conditional on the sample of $\boldsymbol{\beta}$. Specifically,

$$p(\boldsymbol{\beta} \mid \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{Y}, \boldsymbol{X}) \propto p(\boldsymbol{Y} \mid \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2) p(\boldsymbol{\beta})$$
$$\propto \prod_{t=p+1}^{n} \left(\sum_{k=1}^{2} \left(p(Y_i \mid \boldsymbol{\phi}, \boldsymbol{\sigma}^2, I_t = k) p(I_t = k \mid \boldsymbol{X}_t, \boldsymbol{\beta})\right)\right) p(\boldsymbol{\beta})$$
$$\propto \prod_{t=p+1}^{n} \left(C_{t1} g(\boldsymbol{X}_t, \boldsymbol{\beta}) + C_{t2}(1 - g(\boldsymbol{X}_t, \boldsymbol{\beta}))\right) p(\boldsymbol{\beta}),$$

where $C_{ti} = \phi(Y_t; \boldsymbol{Y}_{t-1}\boldsymbol{\phi}^{(i)}, \sigma_i^2)$, the normal density evaluated at Y_t, and $g(\boldsymbol{X}_t, \boldsymbol{\beta})$ is the logistic link function (3). Then the independent hidden state variable $\boldsymbol{I}$ is drawn based on $\boldsymbol{\beta}$:

$$p(\boldsymbol{I} \mid \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{Y}, \boldsymbol{X}) \propto p(\boldsymbol{Y} \mid \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{I}) p(\boldsymbol{I} \mid \boldsymbol{X}, \boldsymbol{\beta}).$$

That is,

$$\begin{aligned} p(I_t = 1 \mid \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{Y}, \boldsymbol{X}) &\propto C_{t1} g(\boldsymbol{X}_t, \boldsymbol{\beta}), \\ p(I_t = 2 \mid \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{Y}, \boldsymbol{X}) &\propto C_{t2}(1 - g(\boldsymbol{X}_t, \boldsymbol{\beta})). \end{aligned}$$

Similar to the first stage, the distributions of $\boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{I}$ are easy to sample. We use random-walk Metropolis algorithm (e.g., Liu (2001)) to generate the samples of β.

References

Chen, R. and Liu, J. S. (1996). Predictive updating methods with application to Bayesian classification. *J. Roy. Statist. Soc. Ser. B* **58**, 397-415.

Liu, J. (2001). *Monte Carlo Strategies in Scientific Computing.* Springer-Verlag, New York.