8-2003

Bayesian Analysis of Threshold Autoregressive Models

Yongjae Kwon
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by Yongjae Kwon entitled "Bayesian Analysis of Threshold Autoregressive Models." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Business Administration.

Halima Bensmail, Major Professor

We have read this dissertation and recommend its acceptance:

Hamprasum Bozdogan, George C. Philippatos, John Barkoulas

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Yongjae Kwon entitled ”Bayesian Analysis of Threshold Autoregressive Models.” I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Business Administration.

Halima Bensmail
Major Professor

We have read this dissertation
and recommend its acceptance:

Hamprasum Bozdogan

George C. Philippatos

John Barkoulas

Acceptance for the Council:

Anne Mayhew
Vice Provost and
Dean of Graduate Studies

(Original signatures are on file with official student records.)

Bayesian Analysis of Threshold Autoregressive Models

A dissertation

Presented for the

Doctor of Philosophy Degree

The University of Tennessee, Knoxville

Yongjae Kwon

August 2003

Acknowledgments

First of all, I would like to thank my major professor, Halima Bensmail, for her guidance and support throughout this research. Also, I would like to thank the members of my committee, Hamprasum Bozdogan, George Philippatos, and John Barkoulas, for their patience and support.

I am also grateful to the faculties and colleagues of statistics department who have given me insights into the new world of statistics. Taking this opportunity, I would thank my parents, family, and friends. Without their suggestions and encouragements, this work would not be possible.

Abstract

Threshold Autoregression is a powerful statistical tool for modeling structural nonlinear relationships. This study presents a Bayesian modeling procedure for threshold autoregressions. To this end, the analytical framework of Bayesian analysis for a univariate SETAR and a threshold VAR were developed. For the estimation of parameters, a Markov-Chain Monte Carlo (MCMC) simulation and an importance/rejection sampling are used to obtain posterior samples.

In model determination, this study shows that Bayes factors are reliable testing procedures in model comparison, lag order selection, and threshold nonlinearity tests. However, it is difficult to get the exact figure of a Bayes factor because the analytical form of the marginal likelihood is occasionally unavailable. In this regard, a few approximation methods for the marginal likelihood as an element of Bayes factor are discussed and appropriate computational algorithms are investigated. Although the Laplace approximation method is a computationally convenient way of approximating marginal likelihood, the validity on small samples is doubtful. Together with Bayes factors, it provided a large scale simulation study on the performance of some information criteria such as SBC, AIC, ICOMP, CAICF_E, and BMS, and recommended they might be good alternatives in small samples or to avoid heavy computational burdens. As a model validation and sensitivity analysis on hyperparameter specifications, both a within-sample and an out-of-sample forecasting are recommended.

This study also provided empirical evidences of the proposed methodology through simulation studies and real data applications. The estimation algorithm of the delay and

threshold parameters is proved to be a stable process. In addition, the Laplace approximation method and Gelfand and Dey (1994) approximation method were used to obtain the marginal likelihoods as elements of Bayes factors. Also, the forecasting functions are approximated by a Monte Carlo simulation.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background

Motivation

For many statisticians, researchers, and practitioners handling time series data, forecasting is one of the most basic but difficult tasks to deal with. Although well-established diagnostic procedures are available for identification, estimation, and testing, there are serious limitations in linear time series modeling. Since all models are assumed to be symmetric, it is sometimes difficult to find models better than a random walk. Usually in time series modeling, the underlying assumption of a constant linear structure over time may be inappropriate when the external structural relationship substantially changes. These features include cycles, amplitude dependent frequencies, and jump phenomena. As a result, a large class of nonlinear models have been proposed in the literature. Granger and Tervrasvirta (1993) discussed a wide range of nonlinear time series models. Threshold regression models, Markov-switching models, bilinear models, and nonlinear moving average models are some of the examples. As a nonlinear time series model, threshold autoregression (TAR) developed by Tong (1978) has been proven to be a useful and relatively simple statistical methodology to model structural nonlinearities.

Usually for this type of model, however, conventional statistical techniques may not be applicable due to the intrinsic structure of the model itself. A most popular research question about threshold autoregressions, for example, is to test the threshold nonlinearity since some of the parameters are present only under the null hypothesis. Studies to overcome the

problems have been carried out along two main lines: the Classical and the Bayesian approaches.

In spite of many debates, the Bayesian approach has some advantages over the classical approach. More than anything else, parameter uncertainty is fully explained by the posterior simulation sampled from the entire parameter space. For this reason, the Bayesian analysis of threshold autoregression has received much attention recently, but the Bayesian modeling process in a unified manner has not been extensively discussed. Therefore, it is meaningful and desirable to bridge the gap in this endeavor, given the developed methodologies.

Research purposes

The main goal of this study is to propose a modeling procedure of threshold autoregressions in Bayesian framework. From this perspective, Bayesian parameter estimation, hypothesis testing, identification, and model validation process will be discussed extensively. To achieve this goal, the following three proposals will be accomplished.

First, the marginal posterior densities of parameters will be analyzed after examination and discussion of prior densities. Since direct simulation from the posterior densities may be occasionally impossible, numerical integration methods will be discussed and investigated to obtain posterior samples.

The second proposal is to investigate Bayesian testing and model identification techniques and to propose reliable modeling procedures. For this purpose, Bayes factor and information theoretic approach will be discussed and their properties will be examined empirically through simulation studies.

Finally, a Bayesian forecasting method will be discussed and the forecasting performance in a threshold model will be conducted for validation purposes by some empirical

applications.

Expected contribution to literature

This study makes several contributions to Bayesian threshold autoregression (TAR) modeling. First, it proposes a combined methodology of posterior analysis, model identification and validation. To this end, the posterior densities of threshold parameters will be derived analytically with conjugate priors for AR parameters and noninformative priors for the delay and threshold parameters.

Second, a Markov-chain Monte Carlo simulation (MCMC) algorithm is developed to obtain posterior samples when pseudo random numbers are not easily obtained from commercial statistical software packages. The Metropolis algorithm and importance sampling will be applied to estimate the delay and threshold parameters.

Third, this research investigates Bayesian testing and model determination methods in autoregression (AR)-lag order selection, threshold nonlinearity testing, and number of regime selection. To this end, the performance of Bayes factors (BF), Schwarz Bayesian criterion (SBC), Akaike's information criterion (AIC), and Bozdogan's informational complexity measure (ICOMP, CAICF_E, and BMS) in TAR models will be examined and applied to simulated data and real data. The AR-lag order selection in Bayesian TAR has not been extensively discussed yet, which this study investigates.

Fourth, model validation will be performed by examining one-step ahead forecasting (prediction) and multi-step ahead forecasting. Also, the sensitivity of hyperparameters will be studied based on the forecasting performances.

Finally, the proposed estimation, testing, and model identification procedure will be examined by applying the methodology to the simulated data and the real data.

Organization of dissertation

The layout of this thesis is as follows: The remainder of this chapter reviews literature relevant to this study. Chapter 2 presents the description of models considered and their analytical framework. Chapter 3 discusses Bayesian estimation and forecasting, mainly from the computational viewpoint. Chapter 4 explores methods of Bayesian model choice based on Bayes factors together with information criteria. Chapter 5 and 6 present simulation studies and applications to real data. Chapter 7 summarizes major findings and concludes.

1.2 Literature review

Since the threshold autoregression (TAR) was proposed by Tong (1978), many methodologies have been developed with two main lines: the Classical and the Bayesian approaches. Even though some methodologies might be used interchangeably, they have somewhat different backgrounds and inferential methods. Therefore, some important contribution in the literature will be reviewed about those two main approaches. Given the wide variations of TAR models, some empirical studies will be reviewed afterwards.

Classical approach

Some of the most important contributions in this area are by Tong (1983), Tsay (1989), Chan (1993), and Quian (1998). Tong (1983) developed self-exciting threshold autoregression (SETAR) modeling procedures and provided many empirical applications of the methodology. For modeling purposes, he used Akaike's information criterion (AIC) to estimate the delay and threshold parameters using grid search method and to select AR-lag orders. Many important properties of SETAR models such as ergodicity, stationary

distributions and moments, cyclical structure of models, and sampling properties of parameters are provided in the literature. On the contrary, Tsay (1989) proposed an arranged regression to obtain predictive residuals as a procedure for testing linear hypothesis as well as for locating the delay and threshold parameters.

Chan (1993) showed the consistency and the limiting distribution of the least squares estimators of a threshold autoregression under some regularity conditions for two-regime SETAR model. Quian (1998) obtained the consistency and limiting distribution of the maximum likelihood estimators.

The testing linear hypothesis is essential since it is a statistical procedure to test if a TAR model fits data significantly better than a linear autoregressive model does. However, it is nonstandard because some of the parameters are not identified under the null hypothesis. Therefore, the classical tests such as Lagrangian multiplier (LM), Wald, and likelihood ratio tests are not applicable. Chan (1990), Andrews and Ploberger (1994), Hansen (1996), and Hansen and Seo (2002) developed test statistics for this situation.

Chan (1990) developed a conditional likelihood-ratio test statistic under a normality assumption on noise terms. This is represented by the normalized reduction of sum of squares due to the piecewise linearity. He showed that this statistic is approximately a central Gaussian process under some regularity conditions. Andrews and Ploberger (1994) considered the optimal testing among the classical tests such as Lagrangian multiplier (LM) test, Wald test, and likelihood ratio test. They derived the optimal tests using a weighted average power criterion and obtained the asymptotics under some regularity conditions. Hansen (1996) proposed a test statistic based on a chi-square process and suggested a bootstrap method to obtain the critical values. As an extension, Hansen and Seo (2002) proposed *SupLM* test statistics for the two-regime threshold vector error correction (TVEC) models and suggested simulation method to obtain the figure of test statistics.

The Akaike information criterion (AIC) has been predominantly used in the identification of AR lag-order structures (Tong 1983, Tsay 1989, and many others). Kapetanios (2001) and De Gooijiter (2001) considered alternative methods in the strategy of AR lag-order selection. Kapetanios (2001) studied comparative analysis of such information criteria as Akaike information criterion (AIC), Schwarz Bayesian Criterion (SBC), Hannan-Quinn Information Criterion (HQ), Generalized information criterion (GIC), and Bozdogan's informational complexity measure (ICOMP). He compared the performances of AR-lag order selection of each criterion through simulation experiments of SETAR and Markov-switching models and concluded that Hanan-Quinn information criterion (HQ) and Schwarz Bayesian criterion (SBC) outperform other criteria. He also conducted simulation experiments of non-nested model comparison among SETAR, Markov-switching, and endogenous delay threshold autoregressive (EDTAR) models and reported that ICOMP performs relatively better than other criteria.

On the other hand, De Gooijiter (2001) focused on the performance of forecasting based on cross-validation criteria. Using some variations of cross-validation criteria such as original version of cross validation (CV) criterion, corrected version of cross validation criterion (CV_c), and unbiased version of cross validation criterion (CV_u), he compared the out-of-sample forecasting performance together with those of AIC and SBC. In his simulation experiment of SETAR, it was concluded that the unbiased version of cross validation criterion (C_u) outperforms other criteria.

The multivariate extension of threshold models have received much attention recently (See e.g., Balke and Fomby 1997, Tsay 1998, and Koop et al. 1996). Balke and Fomby (1997) applied the method of TAR to the system of cointegration in modeling discontinuous adjustment of long-run equilibrium, and also provided the testing procedures of threshold cointegration.

Tsay (1998) extended the methodology of univariate SETAR of Tsay (1989) to vector autoregressive system, including procedures of testing linear hypothesis and identification of the delay and threshold parameters. He provided an estimation procedure of the delay and threshold parameters using a grid search method based on AIC as well as conventional testing methodologies such as χ^2 test statistics.

The impulse response analysis for nonlinear models can be found in Koop, Pessaran, and Potter (1996). They proposed a *generalized impulse response function* to correct the bias and the history- and shock-dependence induced from nonlinearities and showed that the *generalized impulse response functions* could be used in both linear and nonlinear models. They also provided a method of computation for the generalized impulse response functions by a Monte Carlo simulation.

Bayesian approach

There are some limitations in the classical approach due to the unconventional behavior of likelihood function of a TAR. The Bayesian modeling is an attractive statistical method when we have limited data, contamination of measurement error, complicated structural relationship between underlying variables, and nonrandomized data structure (Litterman 1986). Therefore, the obvious alternative to the classical approach is a Bayesian modeling. Incorporating information from the entire parameter space, Bayesian methods capture the finite sample uncertainty about the true parameter space that the classical maximum likelihood (ML) approach does not. Furthermore, posterior densities may be used to combine dynamic features of different models so that the forecasts and impulse response functions will be more reflective of the underlying model and parameter uncertainty (Koop and Potter 1995).

Nevertheless, the literature on Bayesian treatment of TAR models are limited to a few. This is partly because Bayesian methods are computationally demanding. For most of the specifications, analytical forms of posterior distribution do not exist and Bayesian analysis requires simulation-based estimation such as Markov-Chain Monte Carlo (MCMC) and importance/rejection sampling. Another difficulty of Bayesian analysis lies with the appropriate choice of prior distribution and the corresponding hyperparameters.

Lubrano (1995), Broemeling and Cook (1992), Cook and Broemeling (1995), Chen and Lee (1995), and Forbes, Kalb, and Kofman (1999) are some of the references in this line of research. Lubrano (1995) considered a wide range of threshold models such as logistic smooth-transition autoregression (L-STAR), exponential smooth-transition autoregression (E-STAR), and SETAR in Bayesian framework. He provides an excellent description of their properties. He also pointed out that Bayesian TAR can effectively incorporate the properties of smooth transition between regimes that the classical approach does not.

Broemeling and Cook (1992) considered Bayesian SETAR modeling with noninformative priors and proposed a threshold nonlinearity test based on the highest posterior density (HPD) region. Cook and Broemeling (1996) proposed the extension of the previous work, and provided the analytical form of the posterior density of the threshold variable (r) given that the delay parameter (d) is known. In addition, the multi-step ahead forecasting function was approximated following the Monte Carlo simulation approach provided in Geweke and Terui (1993).

Geweke and Terui (1993) developed a methodology to incorporate the uncertainty of the delay and threshold parameters, deriving the exact posterior distribution of SETAR based on Bayes theorem. To compute the posterior probability of the delay and threshold parameters, the integrated density between intervals of the order statistics of the underlying series $\{y_t\}$ was computed and applied to Wolf's annual mean of sunspot numbers and the Canadian lynx

data. They also provided a Monte Carlo simulation to get the posterior samples of multi-step ahead forecasting functions.

Numerical integration methods such as such as Gibbs sampling and Metropolis algorithm were used in Chen and Lee (1995). For this purpose, marginal posterior densities of autoregressive parameters, the delay and threshold parameters were derived and Gibbs sampling algorithm was applied. Since pseudo-random numbers are not easily obtainable from the marginal posterior density of the threshold parameter (r), the Metropolis algorithm was used.

Forbes, et al. (1999) extended the methodology of Geweke and Terui (1993) to a threshold vector cointegrated model (TVECM) of financial arbitrage and obtained the exact posterior distribution of the delay and threshold parameters when Jeffrey's priors are chosen. Further, they developed a Bayesian estimation method by obtaining independent samples from the joint posterior density and the marginal densities of each parameter.

The problem of Bayesian model choice was addressed in Koop and Potter (1999). Given the wide range of variations of SETAR for the historical data of US unemployment rate, they applied Bayes factors to evaluate the non-nested set of models and threshold nonlinearity testing, as well. They showed that the posterior model probability based on marginal likelihoods is a reliable source of model comparison through simulation studies and further reported that Bayes factor contains a strong tendency toward a parsimonious model.

There are some other important literature relevant to Bayesian TAR modeling. Litterman (1986) showed that a Bayesian vector autoregression (BVAR) can be a powerful forecasting tool in economic time series. Kadiyala and Karlsson (1997) considered a wide range of prior densities for vector autoregressions (VAR): Minnesota prior, diffuse prior, normal-Wishart prior, normal-diffuse prior, and extended natural conjugate prior. They proposed several methods of Monte Carlo integration to estimate parameters and concluded that Gibbs

sampling performs better or at least as well as other methods.

The numerical integration of Bayes factor is also an important issue because analytical forms of integrated likelihood (or marginal likelihood) in most of model specifications are not easily available. This is caused by the advent of Markov chain Monte Carlo methods and the rapid development of computational technology. Newton and Raftery (1994), Gelfand and Dey (1994), and Verdinelli and Wasserman (1995) are some of the important contributions in this area. Newton and Raftery (1994) introduced the weighted likelihood bootstrap (WLB) to approximate posterior distributions and proved that the method is a simple and stable way of sampling from the posterior distribution.

Gelfand and Dey (1994) considered Bayes factor as a Bayesian testing procedure and provided asymptotics of the statistic by comparing it to the classical likelihood ratio test statistic. They developed an exact computation method based on posterior samples and the Laplace approximation based on the posterior modes. The finite sampling properties of estimates from the Laplace approximation is also reported in the literature. Although the Laplace approximation would be a convenient method of approximating Bayes factors and the marginal likelihood, they recommended the exact computation method, casting doubt on the accuracy of approximations for small to moderate sample sizes.

Verdinelli and Wasserman (1995) developed an alternative method of computing Bayes factor called the *generalized Savage-Dickey density ratio*, by multiplying the correction factor to the quantity known as the *Savage-Dickey density ratio*. They showed that the *generalized Savage-Dickey density ratio* could be evaluated based on posterior samples, which is easy to implement and interpret, even though this method has some disadvantages in computation of the correction factor and the limited applicability to non-nested testing problem.

Another method of sampling based estimation of Bayes factors can be found in Chib

(1995). Based on the *basic marginal likelihood identity* (*BMI*), he showed that the marginal likelihood could be computed by a posterior sample obtained from Gibbs sampling.

Some empirical studies

There have been many empirical studies applying threshold autoregressions in a wide range of variations. However, it is not desirable to review all the papers, nor is it the purpose of this dissertation. Rather, I would briefly review some selected papers to the extent that is relevant to the purposes of this study. Li and Lam (1995), Altissimo and Violante (2001), Materns et al. (1996), and Forbes et al. (1999) are some of the references of empirical study based on TAR model.

Li and Lam (1995) studied the asymmetric behavior of stock prices in growing and decreasing markets. With 2-regime threshold autoregression and conditional heteroscedasicity (TARCH) model, they analyzed Hong Kong Hang Seng index from 1970 to 1991 and reported strong evidence supporting for the asymmetries between the two markets.

Altissimo and Violante (2001) demonstrated the potential of applications of threshold VAR in their empirical study of macroeconomic model for the dynamics of output and unemployment. Using the data of US output and unemployment, they considered threshold VAR with the inclusion of threshold variable of the depth of the current recession and identified two lags for the variables of the output and the unemployment and four lags for the threshold variable based on the modified likelihood-ratio test proposed by Pesaran and Deaton (1978). They also found strong evidence of supporting the existence of nonlinearites in the dynamic system of output and unemployment based on the threshold nonlinearity test using the conditional likelihood ratio test statistic.

Materns et al. (1996) developed an empirical threshold error correction model (TECM)

to analyze the behavior of index-futures arbitrage relationship and applied their methodology to the data set of the S&P 500 index obtained from the New York Stock Exchange (NYSE) and matching index futures contracts obtained from the Chicago Mercantile Exchange (CME) during June and December 1993. Following Tsay (1989), they identified the two possible alternatives of 5-regime TECM. Forbes et al. (1999) analyzed the same data in the Bayesian threshold error correction system (BTECM) framework and identified the balanced three-regime Bayesian TECM.

The smooth transition autoregression (STAR) models have been heavily used in the efforts of finding the evidence of nonlinearites of some macroeconomic variables. Skalin and Terasvirta (1999), Ocal and Osborn (2000), Baum et al. (2001) are some of these examples. Skalin and Tersvirta (1999) applied exponentially smooth transition model (ESTAR) to the univariate time series of several Swedish macroeconomic variables such as industrial production, imports, exports, productivity, real wage, investment, consumption, and GDP. They identified the AR-lag order of each series based on AIC, and further conducted Granger causality test. They found strong interaction between model specification (linear vs. STAR) and the results of Granger causality test.

Okal and Osborn (2000) analyzed the univariate series of UK consumer consumption and industrial production from the 1st quarter/1955 to the 1st quarter/1995, with single- and two-transition ESTAR and LSTAR models. After identification of AR-lag orders based on AIC, they evaluated the fitted models of linear, two-regime, and three-regime ESTAR and LSTAR and found the evidence of supporting two-regime STAR for the consumption and three-regime STAR for the industrial production. With negligible differences between ESTAR and LSTAR, they also found that there is very little evidence of supporting a model among the linear, two-, and three-regime STAR models in forecasting performance standpoint.

Baum et al. (2001) studied the dynamic adjustment process of exchange rates of 7 industrialized countries to the long-run purchasing power parity (PPP) with three-regime balanced ESTAR models. After the Johansen's cointegration test, 7 countries among 17 countries were chosen for the further analysis. They found asymmetric behavior of mean reverting process based on the regimes in that deviations from the long-run PPP tend to persist for the time being in the middle regime, while in the outer regimes, they died out quickly in a nonlinear manner.

Chapter 2 Models and Bayesian analysis

2.1 SETAR

Model

The threshold autoregressive model can be described as a piecewise linear approximation to the general univariate autoregressive model of order p such that

$$y_t = g(y_{t-1}, ..., y_{t-p}) + \varepsilon_t,$$

where $g(\bullet)$ denotes a nonlinear function and ε_t is an *i.i.d*. random variable. More specifically, the model is given in the form,

$$y_t = \phi_0^{(j)} + \sum_{i=1}^{p} \phi_i^{(j)} y_{t-i} + \varepsilon_t^{(j)},\ if\ \ y_{t-d} \in R^j,\ j = 1, \cdots, M, \tag{1}$$

where R^i constitute a partition of the real line and d is a delay parameter.

The model is self-exciting because the regime (j) is the function of the past realizations of y_t sequence itself. If $y_{t-d} \in R^i$, the model is in regime i at time t. Note that the AR-lag order (p) may differ from regime to regime. If the noise variances $[var(\varepsilon_t^{(j)})]$ are different, then the TAR model reduces to a nonhomogenous linear AR model. If the coefficient vector $\phi^{(j)}$ are the same except the constant terms $\phi_0^{(j)}$, then the model becomes a random-level shift model (Lanne and Saikonen 2002).

Let us consider single-threshold model given in (2) since the extension to multi-regime models is straightforward.

$$y_t = \phi_0^{(1)} + \sum_{i=1}^{p} \phi_i^{(1)} y_{t-i} + \varepsilon_t^{(1)}, \ if \ y_{t-d} \le r,$$
$$\phi_0^{(2)} + \sum_{i=1}^{p} \phi_i^{(2)} y_{t-i} + \varepsilon_t^{(2)}, \ if \ y_{t-d} > r \tag{2}$$

where $0 < d \le D$ and r is the threshold parameter. Since the orders of autoregression (p) are not necessarily the same across the regime, some of the autocorrelation coefficients may be zero. Note that the likelihood function is a step function, r with breaks at the observed y_{t-d}.

For notational simplicity, let us write $y_t = (1, y_{t-1}, ..., y_{t-p})\phi + \varepsilon_t$, for $t = p+1, ..., n$, where $\phi \in R^{p+1}$ of coefficient vector and ε_t is the random noise process. Then, $(y_{t,} 1, y_{t-1}, ..., y_{t-p})$ is so-called a case of data for the $AR(p)$ model. This is the first step of constructing an arranged regression with cases rearranged based on the values of a particular regressor. For a given $SETAR(2; p_1, p_2)$ model with n observations, the threshold variable y_{t-d} may assume values $\{y_h, ..., y_{n-d}\}$ where $h = max\{1, p+1-d\}$. Then, we can rewrite the model as

$$y_{\pi_i+d} = \phi_0^{(1)} + \sum_{v=1}^{p} \phi_v^{(1)} y_{\pi_i+d-i} + \varepsilon_{\pi_i+d-i}^{(1)}, \ if \ i \le s$$
$$\phi_0^{(2)} + \sum_{v=1}^{p} \phi_v^{(2)} y_{\pi_i+d-i} + \varepsilon_{\pi_i+d-i}^{(2)}, \ if \ i > s \tag{3}$$

where s satisfies $y_{\pi_s} < r \le y_{\pi_{s+1}}$.

This is an arranged regression with the first s cases in the first regime and the rest in the second regime (see details in Tsay 1989). Let us assume that the first p observations $(y_1, \cdots, y_p)$ are fixed and the time index of ith smallest observations of $(y_{p+1-d}, ..., y_{n-d})$ is denoted by π_i. Then the likelihood function conditioning on the first p-observations are given by

$$L(\phi_{1,}\phi_2,\sigma_1^2,\sigma_2^2,r,d|y) \propto \sigma_1^{-s}\sigma_2^{-(n-p-s)}\exp\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{2}(y_{\pi_i+d}-\phi_0^{(1)}+\sum_{v=1}^{p}\phi_v^{(1)}y_{\pi_i+d-i})^2$$
$$-\frac{1}{2\sigma_2^2}\sum_{i=1}^{2}(y_{\pi_i+d}-\phi_0^{(2)}+\sum_{v=1}^{p}\phi_v^{(2)}y_{\pi_i+d-i})^2\}.$$

The parameters to be estimated for the model are $\phi^{(1)},\phi^{(2)},\sigma_1^2,\sigma_2^2,r,$ and $d,$ where $\phi^{(1)} = (\phi_0^{(1)},\phi_1^{(1)},\cdots,\phi_p^{(1)})^T$ and $\phi^{(2)} = (\phi_0^{(2)},\phi_1^{(2)},\cdots,\phi_p^{(2)})^T$.

Posterior analysis

For the time being, let us assume that the delay and threshold parameters $\varphi = (r,d)$ are known. Then, it is clear from equation (3) that the likelihood function conditioning on $\varphi = (r,d)$ is of the normal-inverse gamma form such that

$$L(\phi_{j,}\sigma_j^2,j = 1,2|\varphi,y) \propto \Pi_{j=1}^2\sigma_j^{-N_j/2}\exp\{-\frac{1}{2\sigma_j^2}(Y_j - X_j\phi_j)'(Y_j - X_j\phi_j)\}$$
$$= \Pi_{j=1}^2\sigma_j^{-(N_j-k-2)-1}\exp\{-\frac{1}{2\sigma_j^2}\sum_{i=1}^{2}(Y_j - X_j\hat{\phi}_j)'(Y_j - X_j\hat{\phi}_j)\}$$
$$\sigma_j^{-k/2}\exp\{-\frac{1}{2\sigma_j^2}\sum_{i=1}^{2}(\phi_j-\hat{\phi}_j)'X_j'X_j(\phi_j-\hat{\phi}_j)\} \tag{4}$$

where $x_t = (1,y_{t-1},\cdots,y_{t-p})$, $Y_1 = (y_{\pi_1+d},\cdots,y_{\pi_s+d})'$, $Y_2 = (y_{\pi_{s+1}+d},\cdots,y_{\pi_{n-k}+d})'$, $X_1 = (x_{\pi_1+d},$ $\cdots,x_{\pi_s+d})'$, and $X_2 = (x_{\pi_{s+1}+d},\cdots,x_{\pi_{n-k}+d})'$.

To proceed a Bayesian treatment, the specification of prior distribution for the unknown parameters should be included. Similar to Chen and Lee (1995), natural conjugate priors are chosen for the parameters in the following form

$$p(\phi_j|\sigma_j^2) \sim \ independent\ \ Normal(\phi_{0j},\sigma_j^2M_{0j}^{-1}), \tag{5}$$

$$p(\sigma_j^2) \sim independent\ \ Inverse\ \ Gamma(S_{0j},v_{0j}). \tag{6}$$

The delay parameter (d) is restricted to be a positive integer and the likelihood function is a step function with r breaking point at the observed y_{t-d}. Therefore, it may be helpful to define $\varphi = (d, r)$ so that the joint prior densities can be specified for Bayesian treatment. Having this in mind, r is assumed to follow a uniform distribution on (α, β) and d on the integers $1, 2, ..., D$ so that $p(d, r)$ could be of any form such that $\sum_{d=1}^{D} \int p(d, r) dr = 1$ (Geweke and Terui 1993).

Then, the posterior distribution conditioning on φ is given by

$$
\begin{aligned}
p(\phi_{j,} \sigma_j^2, j = 1, 2 | y, \varphi) \;\propto\; & \Pi_{j=1}^{2} (\sigma_j^2)^{N_j/2} \exp(-\frac{S_j}{2\sigma_j^2}) \exp\{-\frac{1}{2\sigma_j^2} (\phi_j - \hat{\phi}_j)' X_j' X_j (\phi_j - \hat{\phi}_j)\} \\
& (\sigma_j^2)^{-v_{0j}/2-1} \exp(-\frac{S_{0j}}{2\sigma_j^2}) \\
& (\sigma_j^2)^{-p/2} \exp\{-\frac{1}{2\sigma_j^2} (\phi_j - \phi_{0j})' M_{0j} (\phi_j - \phi_{0j})\} \\
= \; & \Pi_{j=1}^{2} (\sigma_j^2)^{-v_{1j}/2-1} \exp(-\frac{S_{1j}}{2\sigma_j^2}) \\
& (\sigma_j^2)^{-p/2} \exp\{-\frac{1}{2\sigma_j^2} (\phi_j - \tilde{\phi}_j)' M_{1j} (\phi_j - \tilde{\phi}_j)\}
\end{aligned}
\tag{7}
$$

where $\hat{\phi}_j = (X_j' X_j)^{-1} X_j' Y_j$, $S_j = (Y_j - X_j \hat{\phi}_j)'(Y_j - X_j \hat{\phi}_j)/(N_j - p - 1)$, $M_{1j} = M_{0j} + X_j' X_j$, $\tilde{\phi}_j = M_{1j}^{-1}(M_{0j}\phi_{0j} + X_j' X_j \hat{\phi}_j)$, $S_{1j} = S_{0j} + S_j + S_{\phi j}$, $S_{\phi j} = (\phi_{0j} - \hat{\phi}_j)'[M_{0j}^{-1} + (X_j' X_j)^{-1}]^{-1}(\phi_{0j} - \hat{\phi}_j)$, and $v_{1j} = v_{0j} + N_j$.

Since we assume that the delay and threshold parameters $\varphi = (d, r)$ are known, the conditional posterior distribution for the parameters can be easily obtained by conventional Bayesian technique such that

$$p(\phi_j | \sigma_j^2, \varphi, y) \sim independent\ Normal(\tilde{\phi}_j, \sigma^2 M_{1j}^{-1}) \;\; for \;\; j = 1, 2, \tag{8}$$

and

$$p(\sigma_j^2 | \varphi, y) \sim independent\ Inverse\ Gamma(S_{1j}, v_{1j}) \;\; for \;\; j = 1, 2. \tag{9}$$

Since $p(\phi,\sigma^2|\varphi,y) = p(\phi,\sigma^2,\varphi|y)/p(\varphi)$ and $\iint p(\phi,\sigma^2,\varphi|y)d\phi d\sigma^2 = p(\varphi|y)$, the marginal posterior density for the delay and threshold parameters $[\varphi = (d,r)]$ can be derived analytically by integrating out the joint posterior density with respect to (ϕ_j,σ_j^2) (see Geweke and Terui 1993). The marginal posterior distribution for ϕ_j conditioning on φ is multivariate t, which can be obtained by integrating out with respect to σ_j^2.

$$
\begin{aligned}
p(\phi_j|y,\varphi) &\propto \int_{\sigma_j^2\in R^+}(\sigma_j^2)^{-(v_{1j}+p)/2-1}\exp[-\frac{1}{2\sigma_j^2}\{S_{1j}+(\phi_j-\tilde{\phi}_j)'M_{1j}(\phi_j-\tilde{\phi}_j)\}]d(\sigma_j^2)\\
&= A^{-(v_{1j}+p)/2}2^{(v_{1j}+p)/2}\Gamma(\frac{v_{1j}+p}{2})\int_{\sigma_j^2\in R^+}2^{-(v_{1j}+p)/2}\Gamma^{-1}(\frac{v_{1j}+p}{2})\\
&\quad (\sigma_j^2/A)^{-(v_{1j}+p)/2-1}\exp(-A/2\sigma_j^2)d(\sigma_j^2/A)\\
&= 2^{(v_{1j}+p)/2}\Gamma(\frac{v_{1j}+p}{2})A^{-(v_{1j}+p)/2}
\end{aligned}
\tag{10}
$$

where $A = \{S_{1j}+(\phi_j-\tilde{\phi}_j)'M_{1j}(\phi_j-\tilde{\phi}_j)\}$.

And the marginal posterior distribution of $\varphi = (d,r)$ can be expressed as the product of the integrand of $p(\phi|y,\varphi)$ with respect to ϕ and $p(\varphi)$.

$$
\begin{aligned}
p(\varphi|y) &\propto p(\varphi)\Pi_{j=1}^2 2^{(v_{1j}+p)/2}\Gamma(\frac{v_{1j}+p}{2})\int_{R^k}\{S_{1j}+(\phi_j-\tilde{\phi}_j)'M_{1j}(\phi_j-\tilde{\phi}_j)\}^{-(v_{1j}+p)/2}d\phi_j\\
&= p(\varphi)\Pi_{j=1}^2 2^{(v_{1j}+p)/2}\Gamma(\frac{v_{1j}+p}{2})\Gamma(\frac{v_{1j}}{2})\pi^{v_{1j}/2}|M_{1j}|^{-1/2}S_{1j}^{-v_{1j}/2}\Gamma(\frac{v_{1j}+p}{2})^{-1}\\
&\quad \int_{R^k}c_j^{-1}\{S_{1j}+(\phi_j-\tilde{\phi}_j)'M_{1j}(\phi_j-\tilde{\phi}_j)\}^{-(v_{1j}+p)/2}d\phi_j\\
&= p(\varphi)\Pi_{j=1}^2 2^{(v_{1j}+p)/2}\Gamma(\frac{v_{1j}}{2})\pi^{v_{1j}/2}|M_{1j}|^{-1/2}S_{1j}^{-v_{1j}/2}
\end{aligned}
\tag{11}
$$

where $c_j = \Gamma(\frac{v_{1j}}{2})\pi^{v_{1j}/2}|M_{1j}|^{-1/2}S_{1j}^{-v_{1j}/2}\Gamma(\frac{v_{1j}+p}{2})^{-1}$.

Therefore, the marginal posterior distribution of the delay and threshold parameters is

$$
p(\varphi|y) \propto p(\varphi)\Pi_{j=1}^2\frac{2^{(v_{1j}+p)/2}\Gamma(v_{1j}/2)\pi^{v_{1j}/2}}{S_{1_j}^{v_{1j}/2}|M_{1j}|^{1/2}}. \tag{12}
$$

2.2 Threshold VAR

Model

Let us consider an arranged regression representation which consists of a k-vector $y_t = (y_{1t}, y_{2t}, \cdots, y_{kt})^T$ and v-dimensional exogenous variables $x_t = (x_{1t}, x_{2t}, \cdots, x_{vt})^T$ provided by Tsay (1998). Let $-\infty < r_0 < ... < r_M < \infty$. Then y_t follows a threshold VAR with threshold variable z_t if it satisfies

$$y_t = D^{(j)} + \sum_{i=1}^{p} B_i^{(j)} y_{t-i} + \sum_{i=1}^{q} \beta_i^{(j)} x_{t-i} + \varepsilon_t^{(j)}, \ \ if\, r_{j-1} < z_{t-d} < r_j, \tag{13}$$

where $j = 1, ..., M,$ and D_j are constant vectors and p and q are nonnegative intergers. The innovations satisfy $\varepsilon_t^{(j)} = \Sigma_j^{1/2} I,$ where $\Sigma_j^{1/2}$ are symmetric positive definite matrices. In addition, z_t is assumed to be stationary and continuous.

Given the observations $\{y_{t,}\ x_{t,}\ z_t\}$ for $t = 1, ..., n,$ we can rewrite the expression in the matrix form,

$$y'_t = X'_t B + \varepsilon'_t, \ \ for \ \ \ t = h+1, ..., n.$$

where $h = max(p, d, q),$ $X_t = (1, y'_{t-1}, y'_{t-2} ..., y'_{t-p}, x'_{t-1}, ..., x'_{t-q})'$ of which the dimension is $(pk + qv + 1)$.

Assuming that the threshold variable $z_{t-d} \in \Omega = \{z_{h+1-d}, ..., z_{n-d}\}$ and the ith smallest element of Ω is denoted by $z_{(i)}$ and the time index of $z_{(i)}$ as $t(i),$ we can rewrite above equation as follows

$$y'_{t(i)+d} = X'_{t(i)+d} B + \varepsilon'_{t(i)+d}, \ \ for \ \ i = 1, ..., n-h.$$

This is an arranged regression representation for a threshold VAR, exactly the same

structure to that of the univariate SETAR model (see details in Tsay 1989 and Tsay 1998). Let us consider two-regime threshold VAR models with exogenous variables. Let $Y_1 = (y_{\pi_1+d},...,y_{\pi_s+d})'$, $Y_2 = (y_{\pi_{s+1}+d},...,y_{\pi_{n-p}+d})'$, $X_1 = (X_{\pi_1+d},...,X_{\pi_s+d})'$, and $X_2 = (X_{\pi_{s+1}+d}, ...,X_{\pi_{n-p}+d})'$. Then, the model can be expressed in a compact matrix form

$$Y_j = X_j B_j + U_j, \;\; for \;\; j = 1,2. \tag{14}$$

The dimensions of each matrix for Y_j, X_j, B_j and U_j are $(N_j * k)$, $(N_j * \eta)$, $(\eta * k)$, and $(N_j * k)$ respectively, and where $\eta = (pk + qv + 1)$. By denoting y_j, b_j, and u_j the vectors obtained by stacking the columns of Y_j, B_j, and U_j, we can express the model as the vectorized form of VAR,

$$y_j = (I \otimes X_j) b_j + u_j, \; for \; j = 1,2.$$

In order to connect to the vectorized and the matrix forms of the model, it is useful to consider matricvariate generalizations of some distribution functions, which are found in Fernandez and Steel (1997).

A $(k \times \eta)$ matrix A is said to have a matricvariate normal distribution with the mean matrix $M \in R^{k\times\eta}$ and the covariance matrix of $vec(A)$ given by $\Sigma \otimes \Psi$, where $\Sigma \in R^k$ and $\Psi \in R^\eta$, if the density function of A is given by

$$\begin{aligned} f(A; M, \Sigma, \Psi) &= (2\pi)^{-k\eta/2} |\Sigma|^{-\eta/2} |\Psi|^{-k/2} \exp[-\frac{1}{2} tr\{\Sigma^{-1}(A-M)'\Psi^{-1}(A-M)\}] \\ &= (2\pi)^{-k\eta/2} |\Sigma \otimes \Psi|^{-1/2} \exp[-\frac{1}{2} tr\{(vec(A) - vec(M))'(\Sigma^{-1} \otimes \Psi^{-1}) \\ &\quad (vec(A) - vec(M))\}] \end{aligned}$$

Therefore, $vec(A) \sim$ Multivariate Normal $(vec(M),\, \Sigma \otimes \Psi)$.

A $(k \times k)$ positive semi-definite matrix Σ is said to have an inverted Wishart distribution if its density function is given by

$$f(\Sigma; S, \nu) = 2^{-\nu k/2}\pi^{k(k-1)/4}\prod_{i=1}^{k}\Gamma^{-1}(\frac{\nu+1-i}{2})|S|^{\nu/2}|\Sigma|^{-(\nu+k+1)/2}\exp[-\frac{1}{2}\Sigma^{-1}S]$$

where $S \in R^k$, and $\nu > k-1$.

A $(k \times \eta)$ matrix A is said to have a matricvariate Student-t distribution, if the density function of A is given by

$$\begin{aligned} f(A; M, \Sigma, \Psi, \nu) = \pi^{-k\eta/2}\prod_{i=1}^{k}\Gamma(\frac{\nu+\eta+1-i}{2})\Gamma^{-1}(\frac{\nu+1-i}{2})|\Sigma|^{\nu/2}|\Psi|^{\eta/2} \\ |\Sigma + (A-M)'\Psi(A-M)|^{-(\nu+k)/2} \end{aligned}$$

where $M \in R^{k\times\eta}$, $\Sigma \in R^k$, $\Psi \in R^\eta$, and $\nu > k-1$.

Posterior analysis

Following the equation (14), the likelihood function conditioning on the delay and threshold parameters is given by

$$\begin{aligned} L(B_j, \Sigma_j|\varphi, Y) &\propto |\Sigma_j|^{-N_j/2}\exp[-\frac{1}{2}tr\{\Sigma_j^{-1}(Y_j - X_jB_j)'(Y_j - X_jB_j)\}] \\ &= |\Sigma_j|^{-(N_j-\eta)/2}\exp[-\frac{1}{2}tr\{\Sigma_j^{-1}(Y_j - X_j\widehat{B}_j)'(Y_j - X_j\widehat{B}_j)\}] \\ &\quad |\Sigma_j|^{-\eta/2}\exp[-\frac{1}{2}tr\{\Sigma_j^{-1}(B_j - \widehat{B}_j)'(B_j - \widehat{B}_j)\}]. \end{aligned} \quad (15)$$

It is clear that the conjugate prior is of the matricvariate Normal-inverse Wishart form

$$p(B_j|\Sigma_j) \sim \textit{independent}\ \ MN_{k\eta}(B_{0j}, \Sigma_j, M_{0j}^{-1}) \quad (16)$$

$$p(\Sigma_j) \sim \textit{independent}\ \ \textit{Inverse}\ \ \textit{Wishart}_\eta(S_{0j}, \nu_{0j}). \quad (17)$$

Let us assume that the delay and threshold parameters $\varphi = (r, d)$ are known, the posterior distribution conditioning on the delay and threshold parameters can be obtained by the Bayes theorem and conventional Bayesian technique,

$$
\begin{aligned}
p(B_j, \Sigma_j|Y, \varphi) \propto\ & |\Sigma_j|^{-(N_j-\eta)/2} \exp[\{-\frac{1}{2} tr(\Sigma_j^{-1} S_j)\}] \\
& |\Sigma_j|^{-\eta/2} \exp[-\frac{1}{2} tr\{\Sigma_j^{-1} (B_j - \widehat{B}_j)' X_j' X_j (B_j - \widehat{B}_j)\}] \\
& |\Sigma_j|^{-(v_{0j}+k+1)/2} \exp[\{-\frac{1}{2} tr(\Sigma_j^{-1} S_{0j})\}] \\
& |\Sigma_j|^{-\eta/2} \exp[-\frac{1}{2} tr\{\Sigma_j^{-1} (B_j - B_{0j})' M_{0j} (B_j - B_{0j})\}] \\
=\ & |\Sigma_j|^{-(v_{1j}+k+1)/2} \exp[\{-\frac{1}{2} tr(\Sigma_j^{-1} S_{1j})\}] \\
& |\Sigma_j|^{-\eta/2} \exp[-\frac{1}{2} tr\{\Sigma_j^{-1} (B_j - B_{1j})' M_{1j} (B_j - B_{1j})\}] \qquad (18)
\end{aligned}
$$

where $M_{1j} = M_{0j} + X_j' X_j$, $B_{1j} = M_{1j}^{-1}(M_{0j} B_{0j} + X_j' X_j \widehat{B}_j)$, $S_{1j} = S_{0j} + S_j + S_{B_j}$, $S_{B_j} = (B_{0j} - \widehat{B}_j)'[M_{0j}^{-1} + (X_j' X_j)^{-1}]^{-1}(B_{0j} - \widehat{B}_j)$, $v_{1j} = v_{0j} + N_j$ and $\widehat{B}_j = (X_j' X_j)^{-1} X_j' Y_j$.

Note that

$$
B_j, |\Sigma_j, Y \sim \textit{Matricvariate Normal}_{\eta k}\ (B_{1j}, \Sigma_j, M_{1j}^{-1}) \qquad (19)
$$

and

$$
\Sigma_j|Y \sim \textit{ Inverse Wishart}_k\ (S_{1j}, v_{1j}). \qquad (20)
$$

The marginal posterior distribution can be obtained by integrating out the joint posterior distribution $p(B_j, \Sigma_j|Y, \varphi)$ with respect to B_j and Σ_j.

$$
\begin{aligned}
p(B_j, \varphi|y) \propto\ & p(\varphi) \int p(B_j, \Sigma_j|Y, \varphi) d\Sigma_j \\
\propto\ & p(\varphi) \int |\Sigma_j|^{-(v_{1j}+k+1)/2} \exp[\{-\frac{1}{2} tr(\Sigma_j^{-1} S_{1j})\}] \\
& |\Sigma_j|^{-\eta/2} \exp[-\frac{1}{2} tr\{\Sigma_j^{-1} (B_j - B_{1j})' M_{1j} (B_j - B_{1j})\}] d\Sigma_j \\
=\ & p(\varphi)\ 2^{k(v_{1j}+\eta)/2} \pi^{k(k-1)/4} \Pi_{i=1}^{k} \Gamma(\frac{v_{1j} + \eta + 1 - i}{2}) |A|^{-(v_{1j}+\eta)/2} \\
& \int (2^{k(v_{1j}+\eta)/2} \pi^{k(k-1)/4} \Pi_{i=1}^{k} \Gamma(\frac{v_{1j} + \eta + 1 - i}{2}))^{-1} |\Sigma_j|^{-(v_{1j}+k+\eta+1)/2} \\
& |A|^{(v_{1j}+\eta)/2} \exp[-\frac{1}{2} tr(\Sigma_j^{-1} A)] d\Sigma_j \\
\propto\ & p(\varphi)\ \zeta\ |A|^{-(v_{1j}+\eta)/2} \qquad (21)
\end{aligned}
$$

where $A = S_{1j} + (B_j - B_{1j})' M_{1j} (B_j - B_{1j})$ and $\zeta = 2^{k(v_{1j}+\eta)/2} \pi^{k(k-1)/4} \Pi_{i=1}^{k} \Gamma(\frac{v_{1j}+\eta+1-i}{2})$.

The marginal posterior density of the threshold parameters is given by

$$
\begin{aligned}
p(\varphi|y) &\propto p(\varphi)\Pi_{j=1}^{2}\zeta\int |S_{1j}+(B_j-B_{1j})'M_{1j}(B_j-B_{1j})|^{-(v_{1j}+k)/2}dB_j \\
&\propto p(\varphi)\Pi_{j=1}^{2}\zeta\xi\int^{-1}\xi^{-1}|S_{1j}+(B_j-B_{1j})'M_{1j}(B_j-B_{1j})|^{-(v_{1j}+k)/2}dB_j \\
&\propto p(\varphi)\Pi_{j=1}^{2}\{2^{\eta(v_{1j}+k)/2}\pi^{\eta(\eta-1)/4}\Pi_{i=1}^{k}\Gamma(\frac{v_{1j}+\eta+1-i}{2})\} \\
&\quad \{\pi^{\eta k}|S_{1j}|^{-v_{1j}/2}|M_{1j}|^{-k/2}\Pi_{i=1}^{k}\Gamma(\frac{v_{1j}+1-i}{2})\Gamma^{-1}(\frac{v_{1j}+\eta+1-i}{2})\}
\end{aligned}
$$

where $\xi = \pi^{\eta k}|S_{1j}|^{-v_{1j}/2}|M_{1j}|^{-k/2}\Pi_{i=1}^{k}\Gamma(\frac{v_{1j}+1-i}{2})\Gamma^{-1}(\frac{v_{1j}+\eta+1-i}{2})$.

Therefore,

$$p(\varphi|y) \propto p(\varphi)\Pi_{j=1}^{2}|S_{1j}|^{-v_{1j}/2}|M_{1j}|^{-k/2}\Pi_{i=1}^{k}\Gamma(\frac{v_{1j}+1-i}{2}). \tag{22}$$

Note that the posterior density of the delay and threshold parameters $\varphi = (r, d)$ is function of S_{1j}, M_{1j}, and v_{1j}.

Chapter 3 Bayesian estimation and forecasting

3.1 Bayesian estimation

Hyperparameters

One of the most difficult problems of Bayesian modeling is to find the appropriate hyperparameters. Unless Jefferey's or noninformative prior densities are chosen, we have to decide whether the hyperparameters should be estimated based on the underlying data or they should be specified with some prior beliefs. Unfortunately, the hierarchical modeling process does not look applicable to time series due to the lack of *exchangeability*.

The n values of y_i may be regarded as exchangeable if the joint density $(y_1, y_2, \cdots, y_n)$ is invariant to the permutation of the indexes (Gelman et al. 1997). The *exchangeability* can be interpreted as if the information contained in the data is distributed in a symmetric way.

Therefore, the hyperparameters should be specified in our models based on some prior belief. Assuming that no prior information is available except the data itself, Litterman (1980) suggested that the prior means of the regression parameters should be specified as a random walk process such that

$$y_{it} = y_{i,t-1} + u_{it} \tag{23}$$

for AR coefficients. With this setup, the prior mean of the parameter on the first own lag should be set to unity and the prior mean of the remaining parameters be set to zero.

The specification of covariance matrix is much more difficult. For this purpose, let's assume that the covariances of the parameters are negligible in that no other information is

available except the data itself. Kadiyala and Karlsson (1997) suggested a way of selecting values for the diagonal elements of the covariance matrix based on some reasonable assumptions. First, the prior parameter variances are set to decrease with lag length, considering that the importance of the lagged variables decreases as the lag length increases. Second, the relative tightness of the prior for the parameters on its own lags, foreign lags and exogenous variables should be based on the importance of the systems on interest such that

$$\begin{aligned} Var(B) &= \frac{k_1}{\lambda} \text{ for parameters on own lags} \\ &= \frac{k_2\sigma_i^2}{\lambda\sigma_j^2} \text{ for parameters on lags of variables } j \neq i \\ &= k_3\sigma_i^2 \text{ for parameters on exogenous/deterministic variables} \qquad (24) \end{aligned}$$

where λ denotes the lag length and σ_i^2 is a scale factor accounting for the differing variability of the variables, for which s_i^2 the residual standard error of a p-lag univariate autoregression for variables i may be a good candidate.

By changing the magnitude of $k_1, k_2,$ and $k_3,$ we can control over relative tightness of the hyperparameter space. Figure 1 shows prior distribution of the parameter of the first own lag (b_1) with varying magnitudes of variances when $S_0 = I$ and $(k_1, k_2) = \{(0.5, 0.05),$ $(0.05, 0.005),$ $(0.05^2, 0.005^2),$ $(0.005^2, 0.0005^2)\}$. The particular choice of those values is still of interest to this study. In this regard, Kadiyala and Karlsson (1997) proposed that the common values of the hyperparameters be 0.05 for k_1, 0.005 for k_2 and 10^5 for k_3 while Litterman (1980) suggested 0.05^2 for k_1, 0.005^2 for k_2. Throughout this study, (k_1, k_2) is specified to be $(0.05, 0.005)$ for parameter estimation and the sensitivity analysis will be conducted at different magnitudes of the constants.

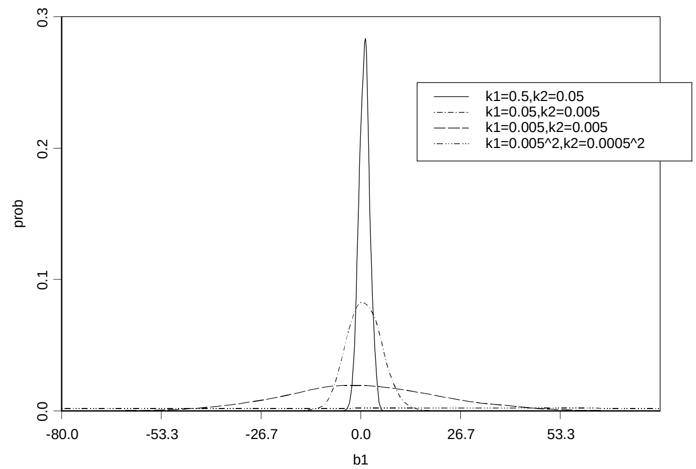

Figure 1 Prior density of AR coefficient $(beta_1)$ with varying variance

Posterior simulation

The difficulty of parameter estimation of TAR lies with estimation of the delay and threshold parameters $\varphi = (r, d)$. Given the marginal posterior density, a Markov-Chain Monte Carlo simulation is one of the possible options to obtain posterior samples. To sample from the posterior density, a Metropolis algorithm and an importance sampling have been used in this study. The Metropolis algorithm creates a sequence of Markov chain from a transition distribution and produces random variables $(\theta^{(1)}, \theta^{(2)}, \cdots \theta^{(l)}, \cdots)$ obtained by acceptance/rejection rule based on the ratio of the two densities evaluated at the previous and the current sample. Using the methodology of Gelman et al. (1997), the algorithm proceeds as follows.

i) Draw a starting point $\theta^{(0)}$ from the starting distribution $p(\theta^{(0)}|y)$.

ii) Sample a candidate point $(\theta^{(t)})$ from the jumping distribution $J_{t-1}(\theta^{(t)}|\theta^{(t-1)})$.

iii) Calculate the ratio of the densities

$$ratio = \frac{p(\theta^{(t)}|y)}{p(\theta^{(t-1)}|y)}. \tag{25}$$

iv) Set

$$\begin{aligned} \theta^{(t)} &= \theta^{(t)}, \text{ with probability } min(ratio, 1) \\ &\quad \theta^{(t-1)}, \text{ otherwise.} \end{aligned}$$

v) Repeat ii) to iv) to obtain desired number of sample points.

Importance sampling is also a useful numerical integration method to simulate random variables from an unnormalized density (Gelman et al. 1997). The problem of evaluating

$$E_{p*}(g(\theta|y)) = c\int g(\theta)p^*(\theta)d\theta = \frac{\int g(\theta)p^*(\theta)d\theta}{\int p^*(\theta)d\theta}$$

can be replaced by an equivalent problem using a Monte Carlo method. Suppose that $I(\theta)$ is proper density from which we can generate pseudo random numbers. Then the above equation can be replaced by

$$E_{p*}(g(\theta)) = \frac{\int g(\theta)w(\theta)I(\theta)d\theta}{\int w(\theta)I(\theta)d\theta} = \frac{E_I[g(\theta)w(\theta)]}{E_I[w(\theta)]} \tag{26}$$

where $w(\theta) = p^*(\theta)/I(\theta)$. By drawing pseudo-random numbers from $I(\theta)$, we can easily estimate $E_{p*}(g(\theta))$. If $I(\theta)$ is so chosen that the importance ratio is roughly constant, then precise estimate of the integral can be obtained.

To actually sample from the joint posterior density $p(r,d|y)$ in this study, the following algorithm has been used.

i) I used the Metropolis algorithm to simulate the threshold parameter (r) from the conditional density $p(r|d,y)$. More specifically, I simulated draws of r^l for each grid of $d = 1,2,\cdots,D$, with the starting value ($r^{(0)}$) obtained from $U(\alpha,\beta)$ and the jumping distribution $J_{t-1}(r^*|r^{(t-1)}) = N(r^{(t-1)},\tau^2)$ which is truncated at $(\varkappa_{15},\varkappa_{85})$, where $\varkappa_{15}$and $\varkappa_{85}$ are 15*th* and 85*th* percentile of the data. The accuracy of the algorithm depends on the choice of jumping distribution. Throughout this study, normal distribution has been used and the τ^2 is chosen such that the rejection rate falls between 45 to 55 %. To eliminate the abnormal effect of the starting point, some starting points at each grid were dropped for further analysis.

ii) To compute the normalizing factor of $p(r,d|y)$, I approximated the conditional density $p(r|d,y)$ by the normal distribution based on the simulated samples of r for each grid of d.

$$p_{approx}(r|d,y) \propto N(\hat{r}(d),\hat{\sigma}_r^2(d))$$

where $\hat{r}(d)$and $\hat{\sigma}_r^2(d)$ are estimated from the obtained samples at each grid.

iii) Since

$$p(d|y) = \frac{p(r,d|y)}{p(r|d,y)},$$

we can draw samples from

$$p_{approx}(d|y) = \frac{p(r,d|y)}{p_{approx}(r|d,y)} \tag{27}$$

using the idea of importance/rejection sampling.

The estimation of (ϕ, σ^2) for SETAR and (B, Σ) for threshold VAR is straightforward. Given the simulated values of $\varphi = (r,d)$, (ϕ, σ^2) and (B, Σ) can be drawn from the conditional posterior distribution of *Inverse Gamma*(S_{1j},v_{1j}) and *Normal*$(\tilde{\phi}_j,\sigma^2M_{1j}^{-1})$ for SETAR and inverse Wishart (S_{1j},v_{1j}) and Matricvariate Normal $(B_{1j},\Sigma_j,M_{1j}^{-1})$ for threshold

VAR, respectively (see chapter 2).

3.2 Forecasting

Comparing the posterior predictive distribution to the data that have actually realized is a good way of model validation. If a model fits well, then the replicated data generated under the model should look similar to observed data. The posterior predictive density of unobserved data $p(\tilde{y}|y)$ incorporates uncertainty affected by the parameter uncertainty in addition to the model variability represented by σ^2. Using the conventional technique, it can be easily shown that the predictive density for future observation in a usual autoregression such that

$$\tilde{y}_j\,|\hat{\phi}_j, \tilde{X} \sim multivariate\ t\ (\tilde{X}\hat{\phi}_j, S_{1j}, \sigma^2(I + \tilde{X}M_1^{-1}\tilde{X}'), v_1)$$

for the one dimensional y, and the predictive density for a k-vector Y is given by

$$\tilde{Y}_j\,|B_{1j}, \tilde{X} \sim matricvariate\ t\ (\tilde{X}_j B_{1j}, (I + \tilde{X}M_1^{-1}\tilde{X}')^{-1}, S_{1j}, v_1)$$

where $\sim$ denotes future observations.

However, Tong (1983) and Geweke and Terui (1993) showed that it is almost impossible to get the closed form solution of the Bayesian h-step ahead predictor. In non-linear autoregressive models, h-step ahead forecast $\hat{y}_{t+h}$ of y_{t+h} conditioning on the observations y_t is given by

$$\hat{y}_{t+h} = E[y_{t+h}|y_t, y_{t-1}, \cdots].$$

The evaluation of the conditional expectation is unrealistic because the joint distribution given in the following form should be marginalized and the closed form solution cannot be obtained easily,

$$p(y_{t+1}, y_{t+2}, \cdots, y_{t+h}|y) = \int p(y_{t+1}, y_{t+2}, \cdots, y_{t+h}|\phi, S, \varphi, Y)p(\phi, S, \varphi)d\phi dS d\varphi$$

where $\phi = (\phi_1, \cdots \phi_M)$, and $S = (S_1, \cdots S_M)$.

Therefore, a Monte Carlo integration method based on the decomposition of posterior densities was proposed (Geweke and Terui 1993).

(i) Draw a sample for $\varphi = (r, d)$ from the posterior density $p(\varphi|y)$.

(ii) Draw a sample for (σ^2, ϕ) conditional on φ from the posterior densities of $p(\sigma^2|\varphi, y)$ and $p(\phi|\sigma^2, \varphi, y)$. The sampling method from the posterior distribution is straightforward since $p(\sigma^2|\varphi, y) \sim \mathit{Inverse\ Gamma}(S_1, v_1)$ and $p(\phi|\sigma^2, \varphi, y) \sim \mathit{Multivariate\ normal}(\tilde{\phi}, \sigma^2 M_1^{-1})$.

(iii) Define $\hat{y}_t = y_t$ and create a predicted sample for $j = t+1, \cdots, t+h$

$$\hat{y}_{t+j} = \tilde{\phi}_0 + \sum_{i=1}^{p} \tilde{\phi}_i \, y_{t+j-i} + \varepsilon_{t+j} \tag{28}$$

where $\varepsilon_{t+j} \sim N(0, \sigma^2)$.

(iv) Repeat (i) - (iii) to get the desired number of samples.

(v) Get the average values from the sample

$$\hat{y}_{t+j} = \frac{1}{L} \sum_{l=1}^{L} \hat{y}_{t+j,l}. \tag{29}$$

The Bayesian h-step ahead predictor for the threshold VAR can be obtained in a similar manner.

(i) Draw a sample for $\varphi = (r, d)$ from the posterior density $p(\varphi|y)$.

(ii) Draw a sample for (Σ, B) conditional on φ from the posterior densities of $p(\Sigma|\varphi, y)$ and $p(B|\Sigma, \varphi, y)$. The samples can be easily obtained since

$$\Sigma|Y \sim \ Inverse \ \ Wishart_k \ (S_1, v_1)$$

$$B|\Sigma, Y \sim Matricvariate \ Normal_{\eta k} \ (B_1, \Sigma, M_1^{-1}).$$

(iii) Define $\widehat{Y}_t = Y_t$ and create a predicted sample for $j = t+1, \cdots, t+h$

$$\widehat{Y}_{t+j} = X_{t+j}B_1 + \widehat{\Sigma}^{1/2}, \ \ for \ \ j = 1, 2. \tag{30}$$

The sample $\widehat{\Sigma}^{1/2}$ can be easily obtained by the spectral decomposition of a sample from $\widehat{\Sigma}$.

(iv) Repeat (i) - (iii) to get the desired number of samples.

(v) Get the average values from the sample

$$\widehat{Y}_{t+j} = \frac{1}{L}\sum_{l=1}^{L}\widehat{Y}_{t+j,l}. \tag{31}$$

The forecasting performance is easily evaluated with the RMSE (Root Mean Squared Error) of the forecasted values and the realized values, given by

$$RMSE = [\frac{1}{H}\sum_{h=1}^{H}(y_{t+h} - \widehat{y}_{t+h})^2]^{1/2}. \tag{32}$$

Chapter 4 Bayesian model selection

4.1 Bayes factor

Bayes factor and the marginal likelihood

After describing the models of interest and the analytical forms of posterior densities in chapter 3, approximate Bayes factors will be used in this chapter to compare models, in the strategy of identification of number of regimes and AR lag-order structures as well as testing the evidence of nonlinearities.

In the classical approach, bootstrap methods are suggested to deal with the unidentification of some of the parameters under the null hypothesis (see, e.g., Chan 1993 and Hansen 1996). Tsay (1989 and 1998) proposed another approach to use predictive residuals obtained from recursive regressions.

In Bayesian approach, however, the problem of identifiability of some of the parameters does not arise due to the fact that Bayes factors automatically integrate out nuisance parameters (Koop and Potter 1999). Along the test of evidence of nonlinearity, Bayes factors may be used to assess the models of interest so that the most appropriate model will be identified given a data set and a possible model set. In this sense, Bayes factor is a useful tool for identification of models among possible alternatives, including the number of regimes and the AR-lag order selection. But the difficulty with Bayes factors lies in its computation since Bayes factors usually involve high-order multiple integration.

To choose a model M_2 against a model M_1, the approximate Bayes factor is given by

$$BF_{1,2} = p(\mathbf{y}|M_2)/p(\mathbf{y}|M_1) \tag{33}$$

where $p(\mathbf{y}|M_k) = \int p(\mathbf{y}|\theta_k)p(\theta_k|M_k)d\theta_k$ and θ_k is the vector of parameters of M_k, and $p(\theta_k|M_k)$ is its prior density; this is called the marginal likelihood of model M_k. Bayesian model selection is based on Bayes factor, whose key ingredient is the marginal likelihood of a model. Sometimes, Bayes factors is used for the ratio of the marginal likelihood under one model to another model. Although the competing models have a common set of parameters, this is not the necessary condition. For choosing the appropriate model, we calculate the Bayes factor for each pair of different combinations of AR order with a particular number of regimes varying from $1, \ldots, M$ for all models. By convention, $log(BF_{12}) < 2$ represents weak evidence, differences between 2 and 6 represent positive evidence, differences between 6 to 10 represent strong evidence, and differences > 10 represents very strong evidence (Jeffrey 1961).

$log(BF_{12})$	< 2	$2 - 6$	$6 - 10$	> 10
	weak evidence	positive evidence	strong evidence	very strong evidence

Laplace approximation

Often it is difficult to compute Bayes factors and therefore some approximation methods for the integrals have been proposed (Verdinelli and Wasserman 1995). One possibility is to approximate the integrals by Laplace's method using the normal approximation. This is simple to calculate and proven to give accurate estimates (Gelfand and Dey 1994, Lewis and Raftery 1994, Bensmail et al. 1997). Gelfand and Dey (1994) provide the large sample asymptotics of the estimator based on the central limit theorem. In this case, the Bayes factor is defined as

$$BF_{1,2} = \frac{p(\mathbf{y}|M_2)}{p(\mathbf{y}|M_1)} = \frac{|-H_2^{-1}(\tilde{\theta}^{(2)})|^{1/2} p(y|\tilde{\theta}^{(2)})p(\tilde{\theta}^{(2)})(2\pi)^{p_2}}{|-H_1^{-1}(\tilde{\theta}^{(2)})|^{1/2} p(\mathbf{y}|\tilde{\theta}^{(1)})p(\tilde{\theta}^{(1)})(2\pi)^{p_1}}, \tag{34}$$

where $\tilde{\theta}^{(i)}, (i = 1, 2)$ is the posterior mode of $\theta^{(i)}$, denoting the parameters $\phi, \sigma,$ and $\varphi = (r, d)$ of the model M_i, H_i is the Hessian of $h(\theta) = \log p(\mathbf{y}|\theta)p(\theta)$, evaluated at $\theta = \tilde{\theta}^{(i)}$, p_1 and p_2 are AR-lag orders for lower and upper regime, respectively. The problem of the Laplace approximation method is that we need the posterior mode $\tilde{\theta}$, and $|-H_2^{-1}(\tilde{\theta})|$ to be known (Bensmail et al. 1997). Since the estimator is based on the normal approximation asymptotics, some researchers cast doubt on the reliability of its accuracy in small samples (see, e.g., Gelfand and Dey 1994). Therefore, other alternatives should be investigated.

Posterior sample based estimator

Another sequence of approximation to Bayes factors is to use a sample from the posterior. Verdinelli and Wasserman (1995) discussed computation of the Bayes factors using *Savage-Dickey density ratio*. Dickey showed that

$$BF = p(\theta_0|y)/p(\theta_0), \text{ if } p(\psi|\theta_0) = p_0(\psi).$$

In this case, the computation of Bayes factors reduces to estimating the integral of $p(\theta_0|y)$. Verdinelli and Wasserman showed that this formula holds in the special case when the condition $p(\psi|\theta_0) = p_0(\psi)$ satisfies and proposed to add the correction factor in the following form

$$BF = p(\theta_0|y)/p(\theta_0)E[\frac{p_0(\psi)}{p(\psi, \theta_0)}] = p(\theta_0|y)/p(\theta_0)E[\frac{p_0(\psi)}{p(\psi|\theta_0)}] \quad (35)$$

which is called the *generalized Savage-Dickey density ratio*. Now, the evaluation of the generalized density ratio reduces to the computation of $p(\theta_0|y)$ and the correction factor $E[\frac{p_0(\psi)}{p(\psi|\theta_0)}]$. Assuming that a sample from the posterior $(\psi_1, \theta_1), (\psi_2, \theta_2), \cdots, (\psi_l, \theta_l), \cdots, (\psi_L, \theta_L)$ is available, we can estimate $p(\theta_0|y)$ by

$$\widehat{p}(\theta_0|y) = \frac{1}{L}\sum_{l=1}^{L} p(\theta_0|y,\psi_l) \tag{36}$$

and we can estimate the correction factor $CF = E[\frac{p_0(\psi)}{p(\psi|\theta_0)}]$ by

$$CF = \frac{1}{L}\sum_{l=1}^{L} \frac{p_0(\widetilde{\psi}_l)}{p(\theta_0,\widetilde{\psi}_l)} \tag{37}$$

where $(\widetilde{\psi}_1,\cdots,\widetilde{\psi}_L)$ represents a sample from the density $p(\psi|\theta_0,y)$. Further, the central limit theorem as a consequence of ergodicity of the chain allows us to approximate the standard error of estimated Bayes Factor

$$std.err(BF) = BF[\frac{s_1^2}{\widehat{p}^2(\theta_0|y)} + \frac{s_2^2}{\widehat{CF}^2}]^{1/2}. \tag{38}$$

where s_1 is an approximate standard error of $\widehat{p}(\theta_0|y)$ and s_2 is an approximate standard error of $\widehat{CF}$, assuming that those estimates are independent samples. Although this approach is an attractive method, there is serious limitations because it is only applicable to nested testing problems since ψ is assumed to be fixed at ψ_0. Therefore, more general approaches using posterior simulation method should be considered.

As the simplest form of the approximation using posterior samples, Newton and Raftery (1994) proposed the estimate

$$\int L(\theta|y)p(\theta)d\theta \approx [\frac{1}{L}\sum_{l=1}^{L}\frac{1}{L(\theta^l|y)}]^{-1} \tag{39}$$

where θ^l is a sample from the posterior. Since this estimate has often infinite variance, the modified version was proposed in the following form

$$\int L(\theta|y)p(\theta)d\theta \approx [\frac{1}{L}\sum_{l=1}^{L}\frac{q(\theta^l)}{L(\theta^l|y)p(\theta^l)}]^{-1} \tag{40}$$

where $q(\theta^l)$ is a density function which is supported by the posterior (Gelfand and Dey

1994). Of course, the accuracy of estimation depends on the choice of importance function q. Chen(1992) showed that a reasonable choice of q is usually a normal density whose mean and covariance is supported by the posterior sample. A different way of estimating Bayes factor is to use a sample from the prior given the same approximation functions above. Obviously, the problem of this method is that the prior will not sample intensively from the region where likelihood is not negligible. To deal with the problem, importance sampling is proposed (Verdinelli and Wasserman 1995).

In this study, the Laplace approximation method is used to compute the integrated likelihood for AR lag-order selection considering the computational burden of posterior sample based estimators. For the threshold nonlinearity testing and the number of regime selection, both estimation methods will be computed denoting by BF_{cond} for the Laplace approximation method and BF_{full} for the simulation based estimator given in equation (40).

4.2 Information theoretic approach

Another sequence of model evaluation can be conducted by a information theoretic approach which has become very popular since the seminal paper of Akaike (1973). The underlying idea of the approach is to select the best model based on some criterion value given a discrete set of models. The interesting feature of this approach is that a model is always preferred to other set of models. Although a wide variety of information criteria have been proposed, the basic structure consists of two components, the goodness of model fit and the penalty term, forming *penalized likelihood* (Bozdogan and Bearse 1999). Akaike's information criterion (AIC) compromises between the log-likelihood and the number of free parameters in a model, of which the basic form is given by

$$AIC_{post} = -2\log L(\widetilde{\theta}|y, M_k) + 2(\text{number of free parameters}) \qquad (41)$$

where $L(\tilde{\theta}|y, M)$ is the likelihood evaluated at the posterior mode of parameter (θ). More specifically, AIC of regime j in SETAR and threshold VAR reduce to

$$AIC_{post}^{(j)}(SETAR) = -2\log L(\tilde{\theta}^{(j)}|\varphi, y) + 2 \bullet (p_j + 1)$$

and

$$AIC_{post}^{(j)}(threshold\ VAR) = -2\log L(\tilde{\theta}^{(j)}|\varphi, y) + 2\kappa_j$$

respectively, where $L(\tilde{\theta}^{(j)}|\varphi, y)$ is the likelihood evaluated at the posterior mode of parameter ($\theta^{(j)}$), p_j is the AR-lag order, $\kappa_j = \{k(kp_j + vq + 1) \ +k(k+1)/2\}$, k is the dimension of dependent variable (y), v is the dimension of exogenous variable (x), and q is the lag order of x.

Among many modifications, Schwarz Bayesian criterion (SBC) given by the following form is the most popular in Bayesian model evaluation,

$$SBC = -2\log L(\tilde{\theta}|y, M_k) - 2\log(p(\tilde{\theta}|y, M_k)) \ + \log(n)(\text{number of free parameters}) \tag{42}$$

where n is the number of observations in the model. More specifically, SBC of regime j of each model is given by

$$SBC^{(j)}(SETAR) = -2\log L(\tilde{\theta}^{(j)}|\varphi, y) + (p_j + 1)\log(n_j)$$

and

$$SBC^{(j)}(threshold\ VAR) = -2\log L(\tilde{\theta}^{(j)}|\varphi, y) + \kappa_j \log(n_j).$$

Bozdogan's Information Complexity criterion (ICOMP) views the complexity of a model not just as the number of parameters or the sample size but as the degree of interdependence between the components in the model by which ICOMP can achieve automatic trade-off in a single criterion value (Bozdogan 1997). Following the utility maximization framework, the Bayesian version of ICOMP can be defined as

$$ICOMP_{post} = -2\log L(\tilde{\theta}|y, M_k) + (\text{number of free parameters}) + 2C_1(F^{-1}) \quad (43)$$

where $C_1(F^{-1})$ is the complexity measure of the inverse of Fisher information matrix. The complexity can be estimated by

$$C_1(F^{-1}) = \frac{1}{2}\{rank(\hat{F}^{-1}(\tilde{\theta}))\log[\frac{trace(\hat{F}^{-1}(\tilde{\theta}))}{rank(\hat{F}^{-1}(\tilde{\theta}))}] + \log\left|\hat{F}^{-1}(\tilde{\theta})\right|\}$$

where $\hat{F}^{-1}(\tilde{\theta})$ is the inverse of Fisher information matrix evaluated at the posterior mode $(\tilde{\theta})$. ICOMP of regime j in SETAR and threshold VAR is given by

$$ICOMP^{(j)}_{post}(SETAR) = -2\log L(\tilde{\theta}^{(j)}|\varphi, y) + (p_j + 1) + 2 \bullet C_1(F^{-1}(\tilde{\theta}))$$

and

$$ICOMP^{(j)}_{post}(SETAR) = -2\log L(\tilde{\theta}^{(j)}|\varphi, y) + \kappa_j + 2 \bullet C_1(F^{-1}(\tilde{\theta})).$$

The consistent Akaike's information criterion with Fisher information (CAICF$_E$) and Bayesian model selection criterion (BMS) are also considered in this study (Bozdogan 2000). The basic form of CAICF$_E$ and BMS are given by

$$\begin{aligned} CAICF_E = & -2\log L(\tilde{\theta}|y, M_k) + \log(n)(\text{number of free parameters}) + \log\left|F(\tilde{\theta})\right| \\ & + 2trace[F^{-1}(\tilde{\theta})R(\tilde{\theta})] \end{aligned}$$

and

$$\begin{aligned} BMS = & -2\log L(\tilde{\theta}|y, M_k) - 2\log(p(\tilde{\theta}|y, M_k)) + \log(n)(\text{number of free parameters}) \\ & + \log\left|F(\tilde{\theta})\right| + 2trace[F^{-1}(\tilde{\theta})R(\tilde{\theta})] \end{aligned}$$

where $R(\tilde{\theta})$ is the outer product form of the inverse of Fisher information matrix. Bozdogan (2000) showed that $trace[F^{-1}(\tilde{\theta})R(\tilde{\theta})]$ is approximated by a constant term $(\frac{nk}{n-k-2})$. Similarly, the two criteria of regime j can be easily computed by

$$CAICF_E^{(j)}(SETAR) = -2\log L(\widetilde{\theta}^{(j)}|\varphi,y) + (p_j+1)\log(n_j) + \log\left|F(\widetilde{\theta}^{(j)})\right| + 2[\frac{n_j(p_j+1)}{(n_j+p_j-1)}],$$

$$CAICF_E^{(j)}(threshod\ VAR) = -2\log L(\widetilde{\theta}^{(j)}|\varphi,y) + \kappa_j\log(n_j) + \log\left|F(\widetilde{\theta}^{(j)})\right| + 2[\frac{n_j\kappa_j}{(n_j+\kappa_j-2)}],$$

$$BMS^{(j)}(SETAR) = -2\log L(\widetilde{\theta}^{(j)}|\varphi,y) - 2*p(\widetilde{\theta}^{(j)}|\varphi,y) + \log(n_j)\bullet(p_j+1) + \log\left|F(\widetilde{\theta}^{(j)})\right| + 2[\frac{n_j(p_j+1)}{(n_j+p_j-1)}],$$

and

$$BMS^{(j)}(threshod\ VAR) = -2\log L(\widetilde{\theta}^{(j)}|\varphi,y) - 2p(\widetilde{\theta}^{(j)}|\varphi,y) + [\frac{n_j\kappa_j}{(n_j+\kappa_j-2)}]\log(n_j) + \log\left|F(\widetilde{\theta}^{(j)})\right| + 2[\frac{n_j\kappa_j}{(n_j+\kappa_j-2)}].$$

The only difference of $CAICF_E$ to BMS is the term of prior densities evaluated at the posterior mode $p(\widetilde{\theta}^{(j)})$.

Table 1 summarizes the penalty term of each information criterion. Note that the goodness of model fit term is the same for all of the criteria considered here. Finally, the criterion value for threshold nonlinearity testing and the number of regime selection can be easily obtained by summing up the criterion values of each regime, assuming that the delay and threshold parameters are fixed at the posterior mode.

Table 1 Information criteria for lag-order selection in regime *j*

		Penalty term
SETAR	AIC_{post}	$2(p_j + 1)$
	SBC	$(p_j + 1)\log(n_j)$
	$ICOMP_{post}$	$(p_j + 1) + 2C_1[F^{-1}(\widetilde{\theta}^{(j)})]$
	$CAICF_E$	$(p_j + 1)\log(n_j) + \log\left\|F(\widetilde{\theta}^{(j)})\right\| + 2[\frac{n_j(p_j+1)}{(n_j+p_j-1)}]$
	BMS	$(p_j + 1)\log(n_j) + \log\left\|F(\widetilde{\theta}^{(j)})\right\| + 2[\frac{n_j(p_j+1)}{(n_j+p_j-1)}]$
Threshold VAR	AIC_{post}	$2\kappa_j$
	SBC	$\kappa_j \log(n_j)$
	$ICOMP_{post}$	$\kappa_j + 2C_1[F^{-1}(\widetilde{\theta}^{(j)})]$
	$CAICF_E$	$\kappa_j \log(n_j) + \log\left\|F(\widetilde{\theta}^{(j)})\right\| + 2[\frac{n_j\kappa_j}{(n_j+\kappa_j-2)}]$
	BMS	$-2p(\widetilde{\theta}^{(j)}) + \kappa_j \log(n_j) + \log\left\|F(\widetilde{\theta}^{(j)})\right\| + 2[\frac{n_j\kappa_j}{(n_j+\kappa_j-2)}]$

Chapter 5 Simulation study

5.1 SETAR model

SETAR(2;1,1)

In this chapter, simulation study will be conducted. Let us first consider a data generated from the process of Chen and Lee (1995), in order to check the Bayesian estimation algorithm, the Metropolis algorithm in particular, such that

$$\begin{aligned} y_t = {} & \phi^{(1)} y_{t-i} + \varepsilon_t^{(1)},\ if\ y_{t-d} \le r, \\ & \phi^{(2)} y_{t-i} + \varepsilon_t^{(2)},\ if\ y_{t-d} > r \end{aligned}$$

where $\phi^{(1)} = -0.5$, $\phi^{(2)} = 0.5$, $\sigma_1^2 = 2$, $\sigma_2^2 = 1$, $r = 0.4$, $d = 1$, and $\varepsilon_t^{(j)}$ follows independent $N(0,\ \sigma_j^2)$. Figure 2 presents the time plot of simulated data of 200 realizations. Following the previous chapter, the hyperparameters are set to be $\phi_j = 0$ except $\phi_{1j} = 1$, $k = 0.05$ and $S_{0j} = s_j^2$ for the covariance matrix, and $v_{0j} = p$, where s_j^2 represents the sample variance of $AR\ (p)$ obtained by ordinary least squares (OLS) estimates where p is set to be 2.

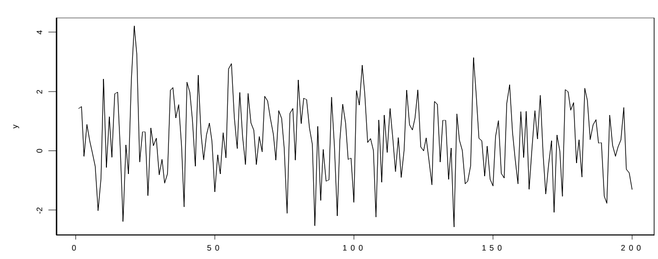

Figure 2 Time plot of simulated data of SETAR(2;1,1)

For the simulated data, linear AR(2) and SETAR(2;2,2) were considered when the delay parameter $d = 1$. For the parameter estimation, the Metropolis algorithm was used because pseudo random numbers are not available directly from the posterior density. Figure 3 presents the time plot of posterior sample of threshold parameter (r) along with the frequency histogram, obtained from 550 iterations, which shows that the algorithm converges quickly to the target density with less than 100 iterations around its true value of 0.4.

The frequency histogram of threshold value (r) for 500 iterations was obtained by eliminating the first 50 samples to burn-in. It is noticed that the empirical distribution is almost symmetric around the true value. This suggests the crude evidence of nlnlineairities existed when $d = 1$.

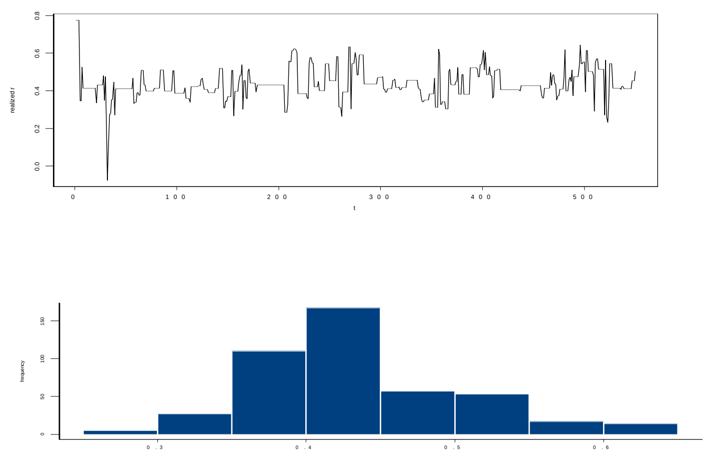

Figure 3 Time plot and frequency histogram of posterior r of SETAR(2;1,1)

Table 2 presents the summary of posterior simulation for SETAR(2;2,2). The estimated AR coefficients of the lower regime are slightly biased and skewed to the left while the estimated threshold value is almost symmetric around its true value. The final analysis for this simulation study is to test the threshold nonlinearity testing based on Bayes factor using Gelfand and Dey approximation method with posterior samples of size 500. Table 3 presents the Bayes factor of M_1 [linear AR(2)] against M_2 [SETAR(2;2,2)], which shows that there is a very strong evidence of supporting SETAR(2;2,2).

SETAR(2;2,2)

Here, a second data set was considered which is generated from the following,

$$\begin{aligned} y_t = {} & \phi_1^{(1)} y_{t-1} + \phi_2^{(1)} y_{t-1} + \varepsilon_t^{(1)},\ if\ y_{t-d} \le r, \\ & \phi_1^{(2)} y_{t-1} + \phi_2^{(2)} y_{t-1} + \varepsilon_t^{(2)},\ if\ y_{t-d} > r, \end{aligned}$$

where $\phi^{(1)} = \begin{bmatrix} \phi_1^{(1)} \\ \phi_2^{(1)} \end{bmatrix} = \begin{bmatrix} 0.7 \\ -0.3 \end{bmatrix},\ \phi^{(2)} = \begin{bmatrix} \phi_1^{(2)} \\ \phi_2^{(2)} \end{bmatrix} = \begin{bmatrix} -0.7 \\ 0.3 \end{bmatrix},\ \sigma_1^2 = 2,\ \sigma_2^2 = 1,$

$r = 0,\ d = 2,$ and $\varepsilon_t^{(j)}$ follows independent $N(0,\ \sigma_j^2)$. Figure 4 presents the time plot of simulated data of 200 observations. Compared with previous example, the nonlinear structure is more apparent as shown in the plot. For the simulated data, linear AR(p) and SETAR($2;p_1,p_2$) were considered with the maximum lag order set to be 5 for both models. To estimate the threshold parameter $\varphi = (r,d),$ the Metropolis algorithm was applied at each grid of the delay parameter (d) with the parameter space of $r \in \{-2.31, 1.58\}$ and $d \in \{1,2,3\}$.

The boundary of the threshold parameter space is chosen such that at least 15% of the data falls in either of the regime.

Table 2 Parameter estimation of SETAR(2;1,1)

Parameter	True value	Mean	Median	Standard deviation
$\phi_1^{(1)}$	–0.50	–0.7955	–0.8030	0.1470
$\phi_2^{(1)}$	0	–0.0318	–0.0291	0.1670
$\phi_1^{(2)}$	0.50	0.4665	0.4694	0.0547
$\phi_2^{(2)}$	0	–0.0611	–0.0634	0.0568
σ_1^2	2	2.0161	1.6443	1.4187
σ_2^2	1	0.9284	0.6475	0.9109
r	0.4	0.4004	0.3984	0.1023

Table 3 Bayes factor for threshold non-linearity testing of SETAR(2;1,1)

Model	Log integrated likelihood	$log(BF_{12})$
Linear AR(2)	–324.95	
SETAR(2;2,2)	–303.87	21.08

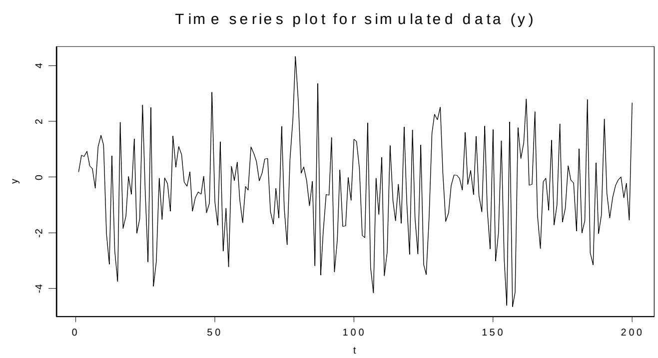

Figure 4 Time plot of simulated data of SETAR(2;2,2)

Figure 5 presents the time plot of simulated r of 1000 iterations at each grid of $d = 1, 2, 3$ and the corresponding frequency histogram of 500 realizations by eliminating the first 500 iterations to burn-in. It is interesting to note that the behavioral patterns of posterior samples of threshold parameter (r) at different grid of d are quite different. The 90% highest posterior density regions of r for each grid are $\{\hat{r}|d = 1\} \in (-0.583, 0.149)$, $\{\hat{r}|d = 2\} \in (-0.070, 0.058)$ and $\{\hat{r}|d = 3\} \in (0.916, 1.088)$. Using the posterior samples obtained from the Metropolis algorithm, the posterior probability of delay parameter $p(\widehat{d}|y)$ was approximated by applying the importance sampling technique explained in the previous chapter and finally the posterior sample of (r, d) was obtained by simple resampling method. Figure 6 shows bar plot of d and the frequency histogram of the posterior sample of r which approximates empirical distribution of posterior d and r, respectively. Note that the delay parameter is estimated to be the true value with more than 99% accuracy.

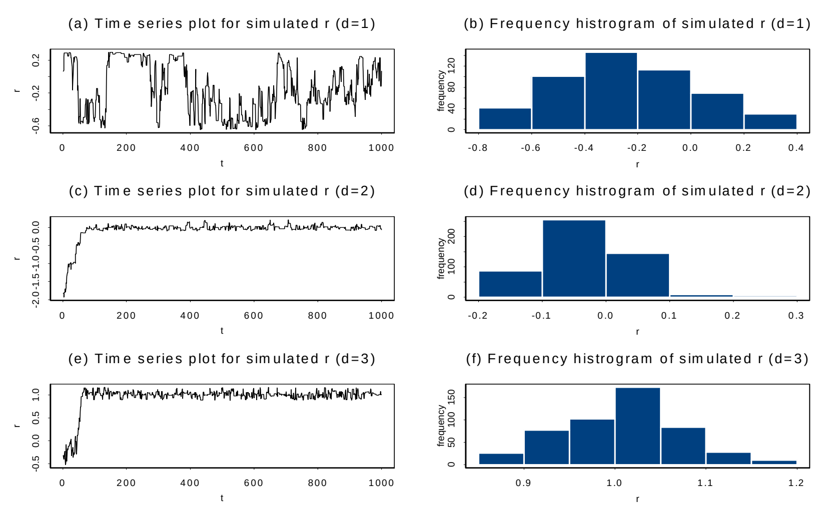

Figure 5 Time plot and the frequency histogram of posterior r of SETAR(2;2,2)

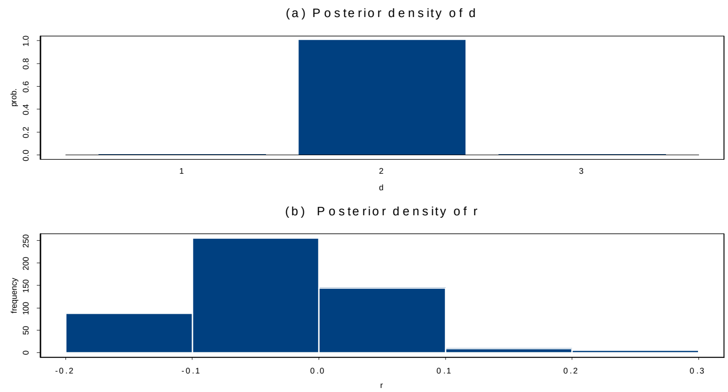

Figure 6 Posterior density of delay (d) and threshold parameter (r)

Next, the AR-lag order selection of SETAR$(2; p_1, p_2)$ was investigated. To examine the performance of the Bayes factors and information criteria such as SBC, ICOMP, AIC, CAICF$_E$, and BMS, 100 simulation experiments were conducted at each level of sample sizes of 100, 200, 500, and 1000 generated from the same data generating process. At each level of samples, the posterior samples of the delay and threshold parameter $\hat{\varphi} = (\hat{r}, \hat{d})$ were obtained. Conditioning on the estimation of the delay and threshold parameter, the log integrated likelihoods as elements of Bayes factors and the information criteria such as SBC, AIC, ICOMP, CAICF$_E$, and BMS were computed and the best models were selected based on the decision rule of each criterion. To check the effect of prior densities on performances of AIC, SBC, and ICOMP, both with- and without-prior densities evaluated at the posterior mode were computed. Special consideration should be given to the Bayes factors because they are considered to be testing procedures, not decision theoretic approach. For Bayes factors, the most preferred models based on the integrated likelihoods were regarded as the best choice among possible alternatives in this experiment. Table 4 summarizes the result of the experiment and more detailed results are found in the appendix. The key findings of the

experiment are summarized in the following:

(i) Positive correlations between sample size and the performance of each criterion have been noticed. For the data of sample size more than 200, Bayes factor, $SBC_{no\ prior}$, SBC_{prior}, AIC_{prior}, $ICOMP_{prior}$, and BMS selects the true lag orders of at least 98 % of the time.

(ii) The Bayes factor performs reasonably well in selecting the true lag orders with sample size of more than about 100. In case of smaller samples, the performances were doubtful, which is suspected to be caused by the approximation method based on normal density.

(iii) Schwarz Bayesian criterion (SBC) and Akaike's information criterion (AIC) without prior adjustments are relatively robust to small samples. However, the performances of SBC and AIC with prior adjustment is slightly better than those of without prior adjustment.

(iv) ICOMP_{prior} and BMS always perform better than $\mathrm{ICOMP}_{no\ prior}$ and CAICF_E.

(v) In small samples, $SBC_{no\ prior}$, $AIC_{no\ prior}$ perform the best.

Given the major findings of the experiment, the Bayes factor and the information criteria (AIC_{prior}, $\mathrm{SBC}_{no\ prior}$, and ICOMP_{prior}) were computed to estimate $\widehat{p}$ for linear AR(p) and $(\widehat{p_1}, \widehat{p_2})$ for SETAR$(2; p_1, p_2)$. The result is summarized in the appendix, which shows that all the criteria agree that $(\widehat{p_1}, \widehat{p_2}) = (2, 2)$ for SETAR$(2; p_1, p_2)$ and $\widehat{p} = 3$ for linear AR(p). It is interesting to note that there is a tendency toward overparameterization with misspecification of the model.

Table 4 Result of simulation experiment for lag-order selection of SETAR(2;2,2)

(a) Lower regime

Approximate sample size	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
65	0.47	0.73	0.43	0.61	0.55	0.42	0.53	0.61	0.56
120	0.96	0.94	0.94	0.76	0.97	0.59	0.98	0.70	0.99
300	0.99	0.98	1.00	0.79	0.99	0.69	0.99	0.78	0.99
600	1.00	0.98	1.00	0.72	0.99	0.63	0.99	0.67	0.99

(b) Upper regime

Approximate sample size	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
30	0.35	0.70	0.30	0.62	0.42	0.51	0.47	0.54	0.46
80	0.63	0.90	0.58	0.74	0.77	0.56	0.81	0.59	0.79
200	1.00	0.99	1.00	0.78	1.00	0.57	1.00	0.61	1.00
400	1.00	0.99	1.00	0.81	1.00	0.57	1.00	0.60	1.00

Finally, the threshold nonlinearity testing was conducted based on Bayes factor computed by Gelfand and Dey approximation method as reported in table 5. It is concluded that there is a very strong evidence of supporting SETAR(2;2,2) against the linear AR(3) given by the figure of log Bayes fator of 75.51. The parameter estimation for the identified SETAR(2;2,2) based on posterior sample of size 500 is summarized in table 6. It is noticed that the estimation result is consistent with the true data generating process.

5.2 Threshold VAR(2;2,2)

Next, bivariate-vector series generated from the following equation was considered where the threshold parameters are set to be $\varphi = (r, d) = (0, 1)$ and the threshold variable is defined to be $z_t = y_{1,t}$.

Table 5 Bayes factor for threshold nonlinearity testing of SETAR(2;2,2)

	Log integratedl ikelihood	$log(BF_{12}^{full})$
1 *regime model*$(\hat{p} = 3)$	–397.61	
2 *regime model*$(\widehat{p_1} = 2, \widehat{p_2} = 2)$	–322.10	75.51

Table 6 Parameter estimation of SETAR(2;2,2)

Parameter	True value	Mean	Median	Standard deviation
$\phi_1^{(1)}$	–0.70	–0.7473	–0.7443	0.0615
$\phi_2^{(1)}$	0.30	0.3438	0.3437	0.0534
$\phi_2^{(2)}$	0.70	0.6703	0.67152	0.0663
$\phi_2^{(2)}$	–0.30	–0.4234	–0.4263	0.0841
σ_1^2	2	2.2407	1.9524	1.4339
σ_2^2	1	1.0724	0.7679	1.0020
r	0	–0.0037	0.0028	0.0516
d	2	2.0000	2.0000	0.0000

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0.9 & 0.9 \\ 0 & 0.7 \end{bmatrix}\begin{bmatrix} y_{t-1,1} \\ y_{t-1,2} \end{bmatrix} + \begin{bmatrix} -0.33 & -0.33 \\ 0 & 0.35 \end{bmatrix}\begin{bmatrix} y_{t-2,1} \\ y_{t-2,2} \end{bmatrix} + \varepsilon_t, \text{ if } y_{1,t-1} \geq 0$$

$$= \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} -0.7 & -0.3 \\ 0 & 0.7 \end{bmatrix}\begin{bmatrix} y_{t-1,1} \\ y_{t-1,2} \end{bmatrix} + \begin{bmatrix} 0.3 & 0.1 \\ 0 & 0.3 \end{bmatrix}\begin{bmatrix} y_{t-2,1} \\ y_{t-2,2} \end{bmatrix} + \varepsilon_t, \text{ if } y_{1,t-1} < 0$$

$$\text{where } \Sigma_1 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}, \ \Sigma_2 = \begin{bmatrix} 1 & -0.3 \\ -0.3 & 1 \end{bmatrix}.$$

Figure 7 presents time plot of 500 realizations. we considered linear and threshold VAR$(2; p_1, p_2)$ of lag orders up to 5 with the boundary condition of the delay parameter $d \in \{1, 2, 3\}$ and the threshold parameter $r \in (-2.13, 0.41)$, assuming that at least 15% of the data falls in either of the regime, for the two-regime threshold VAR.

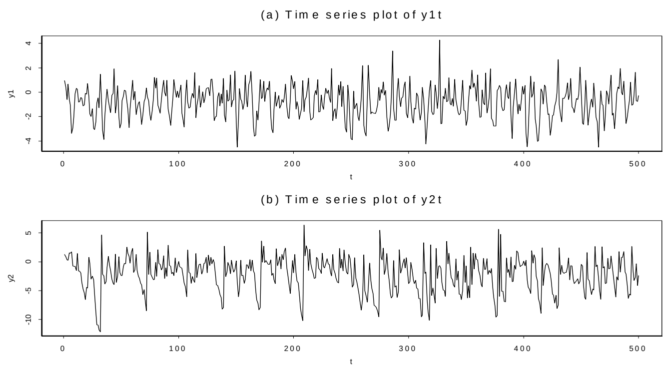

Figure 7 Time plot of threshold VAR(2;2,2)

The hyperparameters of AR coefficients are set to be 0 except the first own lag being 1, $S_{0j} = \begin{bmatrix} s_{1j}^2 & 0 \\ 0 & s_{2j}^2 \end{bmatrix}$, $k_1 = 0.05$ and $k_2 = 0.005$ for M_{0j}, following chapter 5 (see also Kadiyala and Karlsson 1999). The analysis starts with estimation of the delay and threshold parameters $\varphi = (r, d)$. Figure 8 presents the time plot of the posterior output created from the metropolis algorithm of 1000 iterations and the corresponding frequency histogram of the last 500 realization at each grid of $d = 1, 2, 3$. Note that the algorithm is unstable at the grid of $d = 2$, while it converges to the target density with less than 100 iterations at the grid of $d = 1$ and 3. This is consistent with the fact that the nonlinear relationship exists when $d = 1$. Based on the posterior output, the marginal posterior density of delay parameter $p(d|y)$ and the marginal posterior distribution of threshold value $p(r|y)$ are computed (see figure 9). Note that the posterior probabilities of other grid points except $p(\widehat{d} = 1|y)$ are negligible.

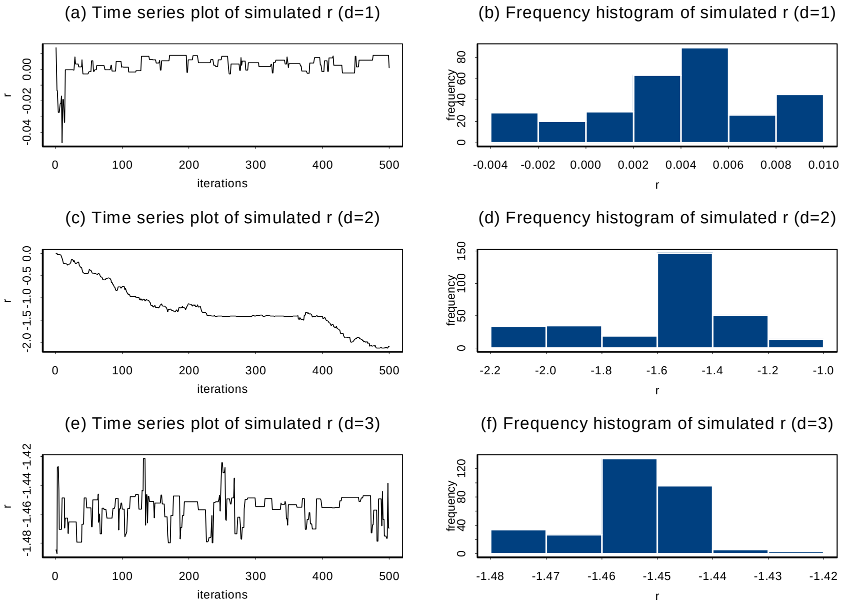

Figure 8 Time plot and frequency histogram of posterior r of TVAR(2;2,2)

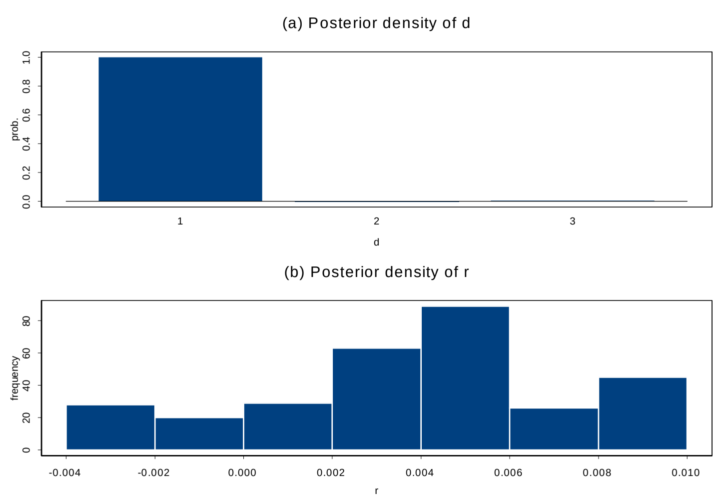

Figure 9 Posterior density of delay (d) and threshold parameter (r)

To examine the performance of the Bayes factors and other information criteria in a threshold VAR, a simulation experiment was conducted. The design of the experiment is similar to that of SETAR in the previous section. The data sets were generated from the same generating process but varying sizes of samples of 150, 250, 500, and 1000. To proceed the study, the posterior samples of threshold parameter $\widehat{\varphi} = (\hat{r}, \widehat{d})$ were obtained at each level of sample size, which were used to estimate parameters and to compute criteria values for 100 sets of data. Conditioning on the delay and the threshold parameter, the log integrated likelihoods as elements of Bayes factors and the information criteria such as SBC, AIC, ICOMP, $CAICF_E$, and BMS were computed and the best models were selected by the criterion values.

Table 7 summarizes the result of the experiment and the detailed information is reported in the appendix. Each cell of table 7 presents the percentages of selecting the true AR-lag orders by each criterion. The major findings are summarized in the following.

(i) There is a strong correlation between sample size and the performance of information criteria, as in SETAR. For samples of more than 100, BF, $SBC_{no\ prior}$, SBC_{prior}, AIC_{prior}, $ICOMP_{prior}$, and BMS select the true lag orders more than 98 % of the time. However, there is no such relationship for $AIC_{no\ prior}$ and $CAICF_E$.

(ii) The performance of Bayes factor was the poorest in very small samples, which might be caused by the normal approximation.

(iii) For small samples of less than 100, $SBC_{no\ prior}$, $ICOMP_{no\ prior}$ and $CAICF_E$ perform the better. Unlike in SETAR, AIC_{prior} always performs better than $AIC_{no\ prior}$ which significantly overestimates the true lag order.

Table 7 Result of simulation experiment for lag-order selection of TVAR(2;2,2)

(a) Lower regime

Approximate sample size	BF	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	$CAICF_E$	BMS
110	0.99	1.00	0.99	0.77	1.00	0.98	1.00	0.56	1.00
190	1.00	1.00	1.00	0.91	1.00	1.00	1.00	0.60	1.00
380	1.00	1.00	1.00	0.85	1.00	1.00	1.00	0.36	1.00
600	1.00	1.00	1.00	0.81	1.00	1.00	1.00	0.27	1.00

(b) Upper regime

Approximate sample size	BF	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	$CAICF_E$	BMS
35	0.41	0.93	0.62	0.80	0.79	0.85	0.72	0.91	0.71
55	0.84	0.96	0.90	0.81	0.96	0.98	0.94	0.87	0.95
120	1.00	1.00	1.00	0.85	1.00	1.00	1.00	0.68	1.00
400	1.00	1.00	1.00	0.84	1.00	1.00	1.00	0.54	1.00

(iv) There is no significant difference between $SBC_{no\ prior}$ SBC_{prior} except when sample size is 35.

Given the results of experiments, lag orders for linear VAR(p) and VAR$(2;p_1,p_2)$ are estimated to be $\hat{p} = 3$ and $(\widehat{p_1},\widehat{p_2}) = (2,2)$ (See appendix). The threshold nonlinearity testing based on Bayes factor is reported in table 8 and the parameter estimation for the identified VAR$(2;2,2)$ is summarized in table 9.

Table 8 Bayes factor for threshold nonlinearity testing of TVAR(2;2,2)

	Log integratedl ikelihood	$log(BF_{12})$
Linear VAR $(\hat{p} = 3)$	-3021.5	
2-regime Threshold VAR $(\widehat{p_1} = 2, \widehat{p_2} = 2)$	-2515.1	506.4

Table 9 Parameter estimation of TVAR(2;2,2)

	Lower regime		Upper regime	
	y_{1t}	y_{2t}	y_{1t}	y_{1t}
$y_{1,t-1}$	0.9140	0.8974	-0.5643	-0.2368
	(0.0064)	(0.0267)	(0.0181)	(0.1290)
$y_{2,t-1}$	0.0051	0.7042	-0.0495	-0.6688
	(0.0014)	(0.01254)	(0.0026)	(0.0127)
$y_{1,t-2.}$	-0.3114	-0.2698	0.4210	0.0699
	(0.0015)	(0.0104)	(0.0376)	(0.1244)
$y_{2,t-2}$	0.0002	0.3275	0.0471	0.2853
	(0.0015)	(0.0049)	(0.0025)	(0.0176)
$\hat{\Sigma}$	0.9988	-0.0205	0.8335	-0.1711
	-0.0205	1.6593	-0.1711	1.0809

* Fugures in parenthesis ( ) represent standard error of the estimated parameters

Chapter 6 Application

6.1 Sunspot numbers

In this chapter, two examples of Bayesian analysis for SETAR and threhold VAR will be provided. Figure 10 presents the time plot of annual mean of sunspot numbers ranging from 1700 to 1979. Because of cyclical behavior of the series, they have long been of interest to astrophysical theorists and statistical analysts as well. Following Yule in 1927, Box and Jenkins, and Akaike applied the linear autoregressive models to the data (See details in Tong 1983). Tong (1983) was the first to model nonlinearities applying TAR model to these figures: He identified the SETAR(2;4,12) based on Akaike's information criterion (AIC). Geweke and Terui (1993) obtained the exact posterior density of the delay and threshold parameter $\varphi = (r, d)$ and the empirical distribution of regime change in Bayesian TAR framework.

Given the previous findings, the analysis covers estimation of parameters, AR-lag order selection, threshold nonlinearity testing, forecasting and the sensitivity analysis.

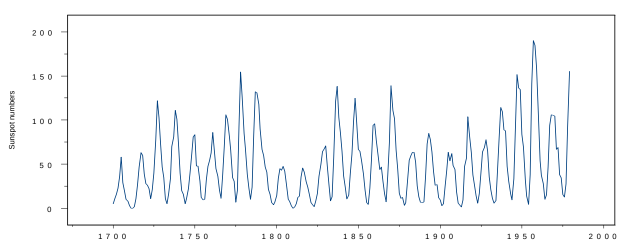

Figure 10 Time plot of annual sunspot numbers (1700-1979)

Here, linear $AR(p)$ and $SETAR(2; p_1, p_2)$ were considered with maximum lag-order of 18. Since the parameter estimation of linear $AR(p)$ is straightforward, let us focus on the estimation procedures of $SETAR(2; p_1, p_2)$ in detail. First, the posterior samples of $\varphi = (r, d)$ was obtained by the Metropolis algorithm and the importance sampling (see chapter 3). With the maximum of delay parameter $d = 4$, 1000 samples of r for each grid of d were obtained from the Metropolis algorithm. Figure 11 presents time plots of the Metropolis algorithm at each grid in the left side and the corresponding frequency histogram of 500 realizations after dropping the first 500 sample points to burn-in. While the algorithm converges to the target density at $d = 3$ and 4, they are not stable at $d = 1, 2$. This represents a crude evidence that the nonlinearities exists at the delay parameter $d = 3$ and 4.

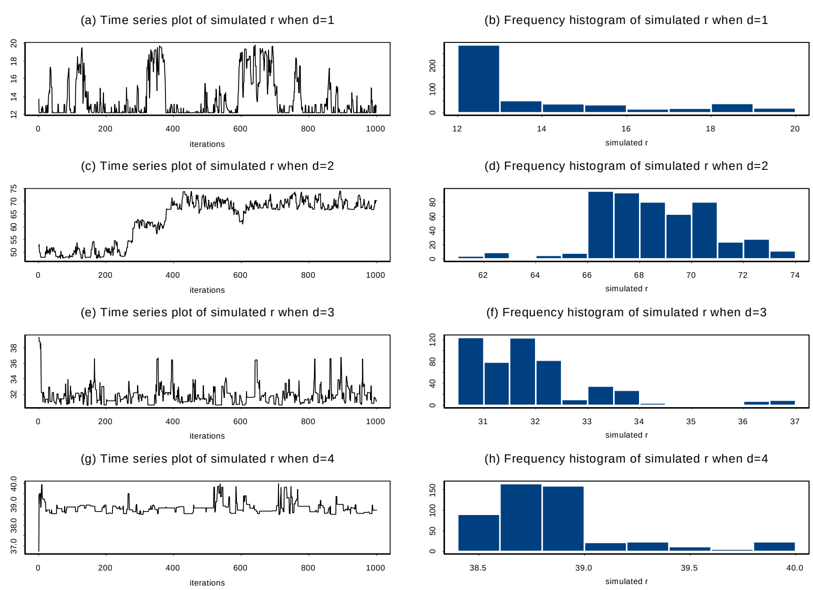

Figure 11 Time plot and frequency histogram of posterior r

From the samples, the marginal posterior densities of d and r were computed based on the technique of importance sampling (see figure 12). It appears that the empirical distribution of posterior density for the threshold value (r) is bi-modal with its mode around 32 and 37 at $d = 3$, supporting the result of Geweke and Terui (1993).

The following analysis is the choice of AR-lag orders for each model. With the maximum lag orders of 18 for each model, the Bayes factors and information criteria were computed and reported in the appendix. For the linear AR(p) model, the Schwarz Bayesian criterion is minimized at AR(2) and the Bayes factors also suggest strong evidence of supporting AR(2), whereas $\text{AIC}_{no\ prior}$ selects AR(11) model. Similarly for two-regime model, the SBC selects $(\hat{p}_1, \hat{p}_2) = (2, 3)$ which is supported by the Bayes factor while the $\text{AIC}_{no\ prior}$ overestimates $(\hat{p}_1, \hat{p}_2) = (6, 18)$. Given the limited validity of $\text{AIC}_{no\ prior}$ in Bayesian framework, the AR-lag orders estimated by the Bayes factor and the SBC were used for further analysis. Based on the estimated lag-orders, the Bayes factor of $SETAR(2; 2, 3)$ against $AR(2)$ was computed using Laplace approximation $[log(BF_{12}^{cond})]$ and Gelfand and Dey approximations $[log(BF_{12}^{full})]$ together with information criteria, as reported in table 10.

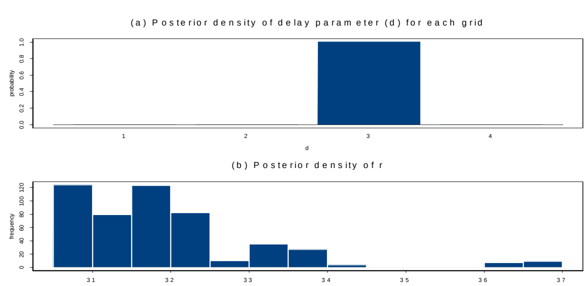

Figure 12 Posterior density of delay (d) and threshold parameter (r)

Table 10 Bayes factors and information criteria for threshold nonlinearity testing

	$\text{AIC}_{no\ prior}$	$\text{SBC}_{no\ prior}$	ICOMP_{prior}	$log(BF_{12}^{cond})$	$log(BF_{12}^{full})$
$LinearAR(2)$	2233.2	2357.2	2356.3		
$SETAR(2;2,3)$	2117.2	2318.5	2297.5	44.6	12.8

Table 11 Parameter estimation of Sunspot numbers

Parameter	Mean	Median	Standard deviation
$\phi_0^{(1)}$	17.6655	17.6991	2.5520
$\phi_1^{(1)}$	1.7056	1.7064	0.1008
$\phi_2^{(1)}$	–1.2678	–1.2712	0.1772
$\phi_0^{(2)}$	7.1153	7.1312	2.6768
$\phi_1^{(2)}$	0.9107	0.9127	0.0987
$\phi_2^{(2)}$	–0.0318	–0.0265	0.1519
$\phi_3^{(2)}$	–0.1778	–0.1773	0.7738
σ_1^2	248.5516	247.9273	15.6690
σ_2^2	113.2076	113.2507	11.0396
r	31.887	31.64	1.2112
d	3	3	0.0000

The logarithm of Bayes factor suggests that there is a very strong evidence of supporting $SETAR(2;2,3)$. Finally, table 11 presents the result of parameter estimation of the identified $SETAR(2;2,3)$ obtained from 500 posterior samples.To check the validity of the identified model, one-step ahead and multi-step ahead forecasting exercises were conducted by partitioning the data into a test set (1770-1929) and a validation set (1930-1979). The delay and threshold parameters were estimated that $\widehat{\varphi} = (\widehat{r}, \widehat{d}) = (40.68, 3)$ with the testing data set. The results of one-step ahead forecasting at each grid of λ are plotted against realized figures in figure 13. Also, the root mean squared errors (RMSE’s) of up to 50-ahead forecasting horizons are presented in table 12 and 13.

Table 12 RMSE of one-step ahead forecasting (sunspot numbers; 1930-1979)

		λ			
	Forecasting horizon (h)	0.5	0.05	0.005	0.005^2
RMSE	5	12.7	12.4	12.4	13.8
	10	11.3	10.8	10.8	11.7
	15	10.0	10.1	9.8	10.4
	20	9.8	10.0	9.7	10.0
	30	13.7	14.4	13.7	14.1
	50	18.0	17.3	19.9	22.0

Table 13 RMSE of h-step ahead forecasting (sunspot numbers; 1930-1979)

		λ			
	Forecasting horizon (h)	0.5	0.05	0.005	0.005^2
RMSE	5	7.3	8.0	8.8	4.5
	10	10.6	10.5	15.3	6.0
	15	21.3	18.6	25.0	16.0
	20	31.9	28.8	34.5	24.6
	30	45.3	43.7	50.5	40.8
	50	59.9	59.1	63.4	55.8

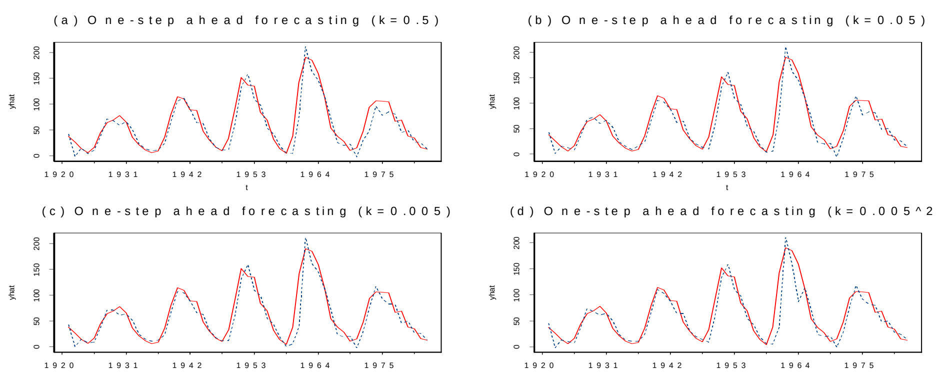

Figure 13 Time plot of one-step ahead forecasting (1930-1979)

Although the difference is not significant, the RMSE at $\lambda = 0.05$ gives us the minimum RMSE. Table 13 reports the root mean squared error (RMSE) of multi-step ahead forecasting exercise. The figures show that the RMSE's at $\lambda = 0.005^2$ are significantly lower than the other values of λ. The other three cases do not look significantly better than any others (see figures 13 through 16).

Tong (1983) reported the RMSE's of forecasting horizon 1921-1955 with the fitting period of 1721-1890. For the comparison purposes, the multi-step ahead forecasting of this period was conducted for the same period of test and forecasting data set. The RMSE's of maximum time horizon of 35 are presented in the appendix, which shows that the Bayesian multi-step ahead predictor has a significantly lower RMSE. Figure 17 shows the time series plot of multi-step ahead forecasting.

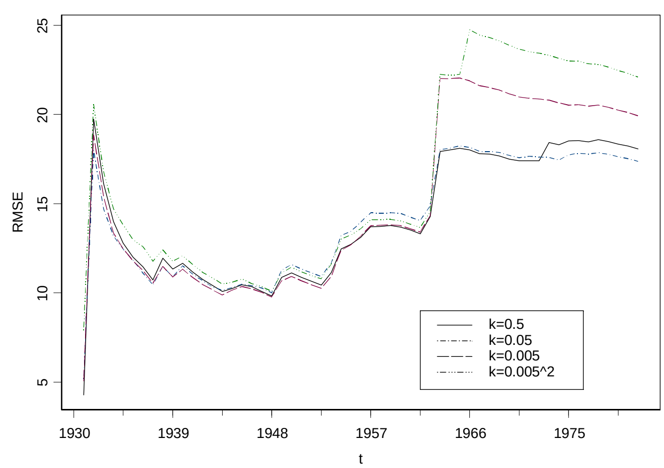

Figure 14 Time plot of RMSE of one-step ahead forecasting (1930-1979)

Figure 15 Time plot of h-step ahead forecasting (1930-1979)

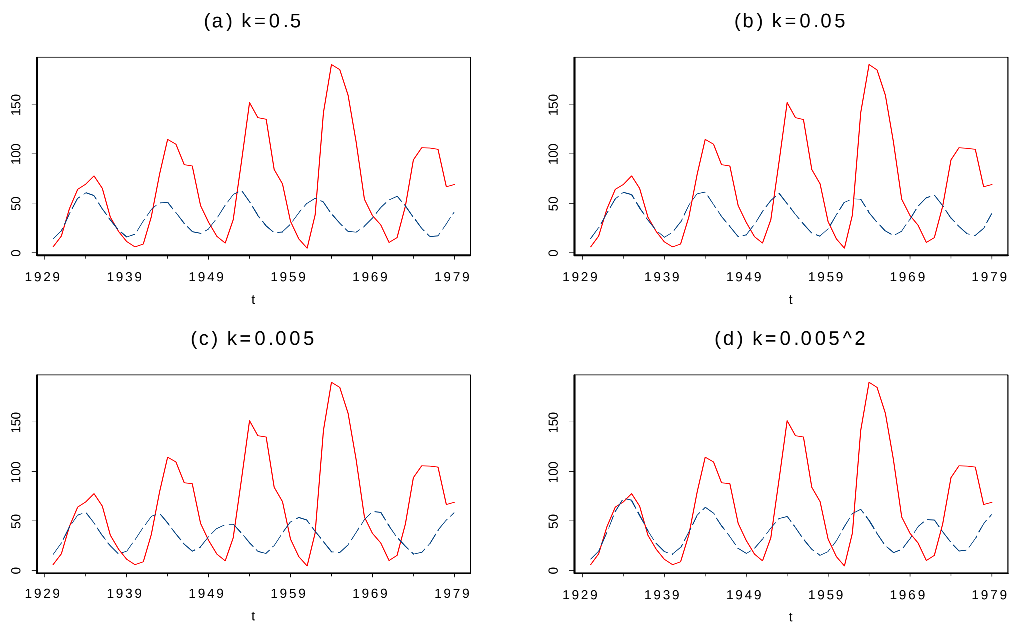

Figure 16 Time plot of RMSE of h-step ahead forecasting(1930-1979)

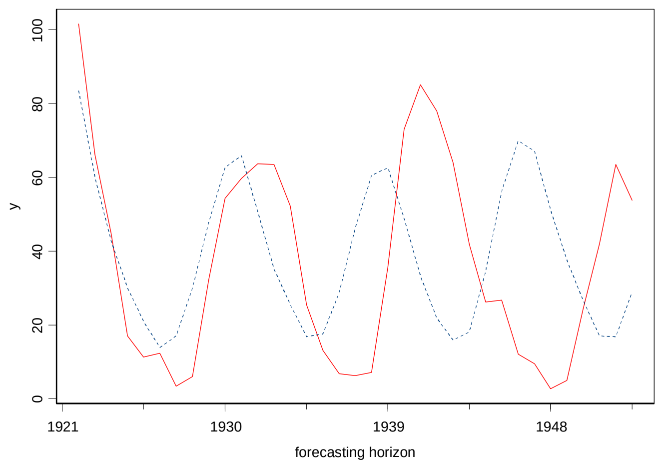

Figure 17 Time plot of h-step ahead forecasting (1921-1950)

6.2 US interest rate

Tsay (1998) analyzed bivariate series of monthly interest rates of 3-month treasury bills and 3-year treasury notes from January 1959 to February 1993, in the perspective of recursive least squares framework. In his analysis, three-regime threshold VAR model with $(\hat{p}, \hat{d}, \hat{s}) = (7, 4, 3)$ was identified based on AIC, where $p, d,$ and s denotes AR-lag orders, delay parameter, and number of regimes, respectively. Further, AR lag orders for each regime were estimated to be $(\widehat{p_1}, \widehat{p_2}, \widehat{p_3}) = (2, 6, 7)$.

The data was obtained from the website of Federal Reserve Base of St. Louis (http://research.stlouisfed.org/fred/data/rates.html) but the time horizon was extended from April 1953 to September 2002.

The two series are constructed such that $y_t = (y_{1t}, y_{2t})$ where y_{it} is the growth series of natural logarithm of original sequences $[y_{it} = \ln(Y_{it}) - \ln(Y_{i,t-1})]$, where Y_{1t} and Y_{2t} represent

the interest rate of 3-month treasury-bill and the yield to maturity of 3-year treasury-note. The interest rate spread of the logarithm of the two series was used as the threshold variable such that $z_t = \ln(Y_{1t}) - \ln(Y_{2t})$. In addition, 3-month moving average of the interest rate spread was constructed so that the threshold variable incorporates quarterly economic information (See details in Tsay 1998). Figure 18 presents the time series plot of the two series and the 3-month moving average of the interest rate spread.

For this data, linear VAR, two-regime, and three-regime threshold VAR models were considered for the possible alternatives. The analysis begins with the estimation of threshold parameters $\varphi = (r, d)$ with the restrictions of $r \in (z_{.15}, z_{.85})$ and $d \in (1, 4)$ where z_α reflects α percentile of the order statistics of $\{z_t\}$. Following the previous chapter, the Metropolis algorithm was used to generate posterior samples of r at each grid of d.

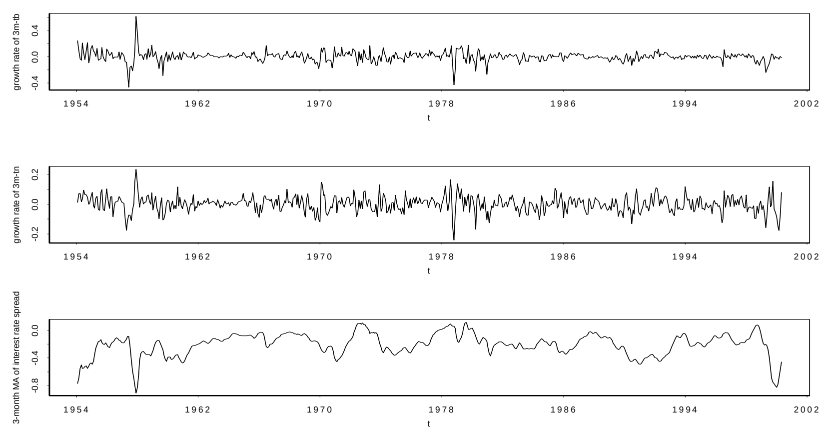

Figure 18 Time plot of 3-month treasury bill rate (a), 3-year treasury note yield to maturity (b), and 3-month moving average of the spread between the two series (c)

The following figures 19 and 20 present time plots of 2000 iterations of the algorithm, for a two-regime and a three-regime model, respectively. As shown in the plot, the simulation algorithm converges quickly to the target density about less than 100 iterations as in the SETAR case. Note that it is stable at the grid of d=1,2 in two-regime threshold VAR while it converges to the target density quickly at all the grid points in three-regime VAR model.

From the obtained samples, the approximate posterior probability of d was computed (See table 14 and figure 21). Unlike Tsay (1998) 's analysis, the delay parameter is estimated to be as $\widehat{d} = 1$ in either case of two-regime and three-regime model. Figure 22 and 23 present the frequency histogram of the posterior samples of $\hat{r}$, from which two posterior mode around (-0.34) is identified.

The identification of lag-order structures followed by the estimation of threshold variable (r) and delay parameter (d). For each regime, Bayes factors based on the Laplace approximation of integrated likelihood as well as Akaike information criterion (AIC) and Schwartz Bayesian criterion (SBC) were computed up to the maximum AR lag-order of 15 which is reported in the appendix. According to the decision rule of Bayes factors provided by Jefferey (1961), $\hat{p} = 3$ for the linear VAR, $(\hat{p}_1, \hat{p}_2) = (2, 3)$ for the two-regime threshold VAR, and $(\hat{p}_1, \hat{p}_2, \hat{p}_3) = (2, 3, 2)$ for the three-regime threshold VAR model were identified.

Table 14 Marginal posterior probability of delay parameter

	$d = 1$	$d = 2$	$d = 3$	$d = 4$
Two-regime model	0.9980	0.0000	0.0001	0.0019
Three-regime model	0.9969	0.0000	0.0031	0.0000

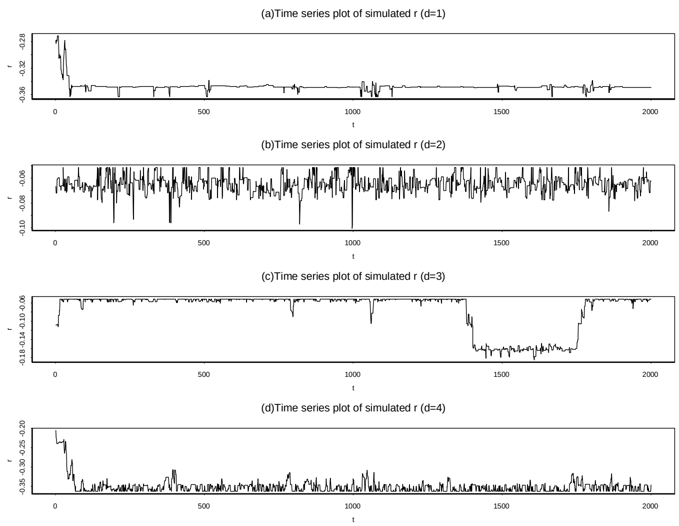

Figure 19 Time plot of posterior simulation of r (two-regime case)

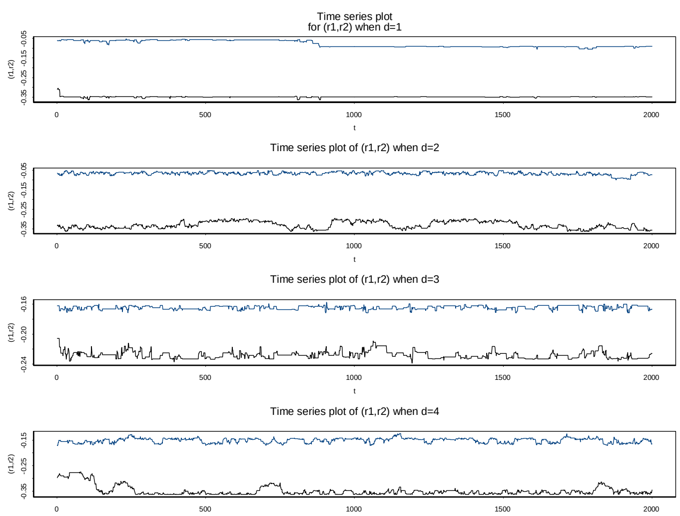

Figure 20 Time plot of posterior simulation of $r = (r_1, r_2)$ (three-regime case)

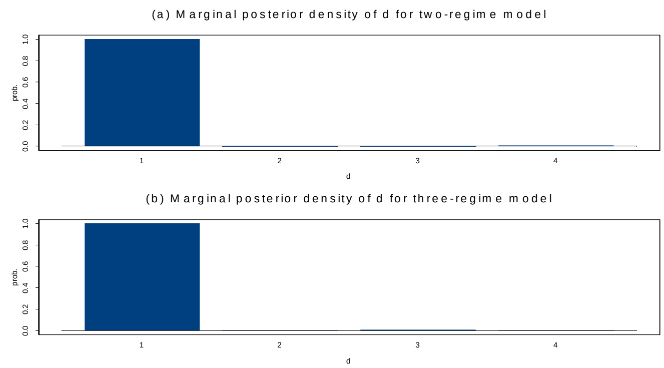

Figure 21 Approximate posterior density of the delay parameter (d) (two-and three-regime case)

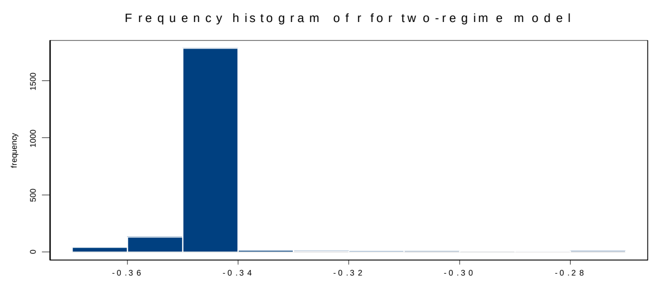

Figure 22 Posterior density of threshold parameter (r) in two-regime case

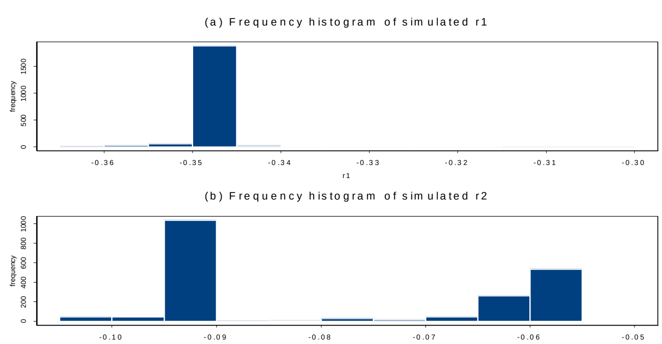

Figure 23 Posterior density of threshold parameter (r) in three-regime case

The Bayes factor and the $\mathrm{SBC}_{no\ prior}$ on the choice of AR-lag orders while there are some discrepancies with $\mathrm{AIC}_{no\ prior}$. $\mathrm{AIC}_{no\ prior}$ recommends $\hat{p} = 7$ for the single-regime VAR, $(\hat{p}_1, \hat{p}_2) = (2, 3)$ for the two-regime VAR, and $(\hat{p}_1, \hat{p}_2, \hat{p}_3) = (2, 3, 1)$ for the three-regime VAR. There are strong or very strong evidences of supporting VAR(2) for single-regime VAR and VAR(3) for the middle-regime of three-regime threshold VAR. However, the evidences of supporting any model between VAR(1) and VAR(2) for the lower-regime of two-regime model, VAR(1), VAR(2), and VAR(3) for the upper-regime of two-regime model, VAR(1) and VAR(2) for the lower-regime of the three-regime case, and VAR(1) and VAR(2) for the upper-regime of the three-regime model are not strong.

As a final step of the analysis, the Bayes factors of linear VAR, two-regime and three-regime threshold VAR were computed by the Gelfand and Dey approximation method using 500 posterior samples as well as BF based on Laplace approximation and information criteria, which is summarized in the table 15. As shown in the table, there is a strong evidence of supporting two-regime model and also a strong evidence of supporting three-regime threshold VAR against two-regime threshold VAR. Therefore, the threshold VAR(3;2,3,2) were finally identified for the data. The estimation of parameters is summarized in table 16.

Table 15 Bayes factors and information criteria for threshold linearity testing (US interest rates)

	$\mathrm{AIC}_{no\ prior}$	$\mathrm{SBC}_{no\ prior}$	ICOMP_{prior}	$\log(\mathrm{BF}_{cond})$	$\log(\mathrm{BF}_{full})$
Single-regime VAR $(\hat{p} = 3)$	−1523.6	−1524.4	−1504.2		
Two-regime threshold VAR $(\hat{p}_1 = 2, \hat{p}_2 = 3)$	−1521.1	−1576.6	−1548.3	$BF_{12} = 24.7$	$BF_{12} = 32.36$
Three-regime threshold VAR $(\hat{p}_1 = 2, \hat{p}_2 = 3, \hat{p}_3 = 2)$	−1617.1	−1660.9	−1546.3	$BF_{23} = 8.6$	$BF_{23} = 42.26$

Table 16 Parameter estimation (US interest rates)

	Lower regime		Middle regime		Upper regime	
	y_{1t}	y_{2t}	y_{1t}	y_{2t}	y_{1t}	y_{1t}
constant	0.0158	0.0775	–0.0011	0.0926	–0.0102	–0.0311
	(0.0092)	(0.4786)	(0.0043)	(0.2903)	(0.0121)	(0.4895)
$y_{1,t-1}$	0.4568	0.4226	0.4073	0.3020	0.3655	0.4104
	(0.2290)	(0.6550)	(0.2223)	(0.3552)	(0.2900)	(0.7039)
$y_{1,t-2}$	–0.1808	–0.0540	–0.1251	–0.1391	–0.0812	–0.1431
	(0.1751)	(0.4897)	(0.2611)	(0.3429)	(0.2085)	(0.5392)
$y_{1,t-3}$			0.0380	0.1299		
			(0.2156)	(0.3832)		
$y_{2,t-1}$	0.1239	–0.0475	0.1408	–0.0858	0.4004	–0.0133
	(0.3125)	(0.3509)	(0.2740)	(0.3395)	(0.3965)	(0.3282)
$y_{2,t-2}$	0.1674	0.0123	0.0278	–0.0424	–0.1560	0.0147
	(0.2485)	(0.0018)	(0.2704)	(0.2859)	(0.3147)	(0.0017)
$y_{2,t-3}$			0.0706	0.0068		
			(0.2903)	(0.0006)		
$\hat{\Sigma}$	0.0063	0.0021	0.0029	0.0016	0.0056	0.0028
	0.0021	0.0027	0.0016	0.0018	0.0028	0.0027

* Figures in parenthesis ( ) represent the standard deviations of parameters

As a validation process, one-step ahead and multi-step ahead forecasting analyses were conducted. The fitting period is from 06/1953 to 09/1999 and the forecasting period is from 10/1999 to 09/2002. To investigate the sensitivity of specification of the hyperparameters, forecasting exercises at different combinations of (k_1, k_2) were performed. Figure 24 and figure 25 presents the time plot of one-step ahead forecasting exercises at each combination of (k_1, k_2). Table 17 and 18 present the time plot of RMSE's of one-step ahead and h-step ahead forecasting, respectively. Although the realized RMSE's of each level are not significantly different, they are minimized at $(k_1, k_2) = (0.05^2, 0.005^2)$.

Table 17 RMSE of one-step ahead forecasting (US interest rates)

	Forecasting horizon (h)	(λ_1, λ_2)			
		$(0.5, 0.05)$	$(0.05, 0.005)$	$(0.05^2, 0.005^2)$	$(0.005^2, 0.0005^2)$
$RMSE(y_1)$	3	0.046	0.043	0.038	0.060
	6	0.052	0.053	0.048	0.055
	12	0.079	0.076	0.076	0.078
	24	0.085	0.084	0.083	0.086
$RMSE(y_2)$	3	0.060	0.056	0.067	0.059
	6	0.059	0.059	0.063	0.063
	12	0.062	0.060	0.063	0.066
	24	0.073	0.077	0.070	0.077

Table 18 RMSE of multi-step ahead forecasting (US interest rates)

	Forecasting horizon (h)	(λ_1, λ_2)			
		$(0.5, 0.05)$	$(0.05, 0.005)$	$(0.05^2, 0.005^2)$	$(0.005^2, 0.0005^2)$
$RMSE(y_1)$	3	0.044	0.063	0.061	0.068
	6	0.071	0.083	0.082	0.076
	12	0.087	0.094	0.101	0.096
	24	0.085	0.090	0.088	0.087
$RMSE(y_2)$	3	0.062	0.082	0.076	0.060
	6	0.068	0.064	0.072	0.061
	12	0.089	0.073	0.066	0.068
	24	0.089	0.088	0.081	0.091

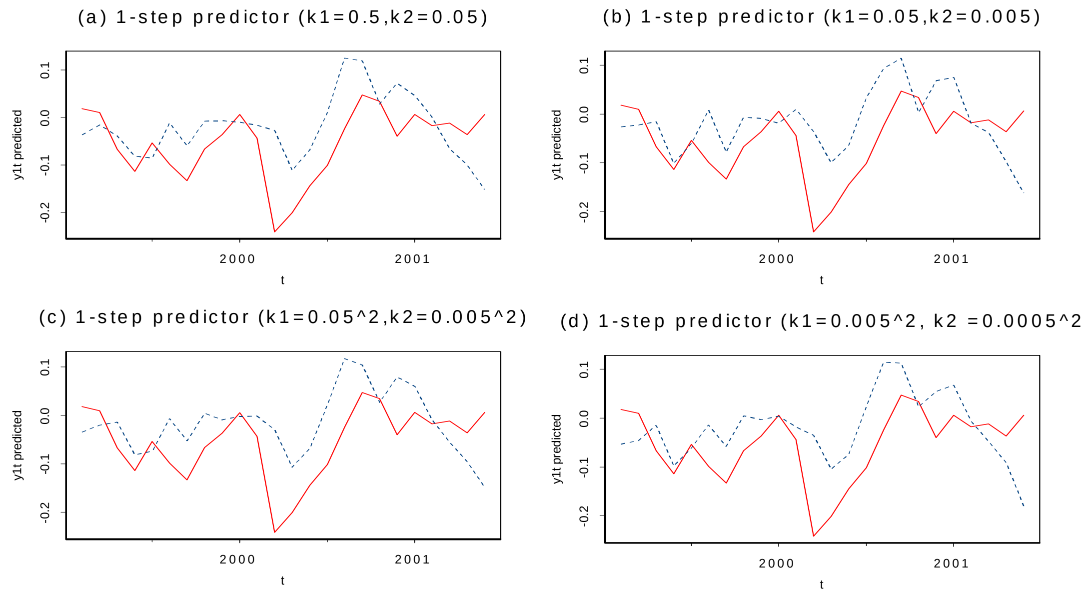

Figure 24 Time plot of one-step ahead forecasting (y_1)

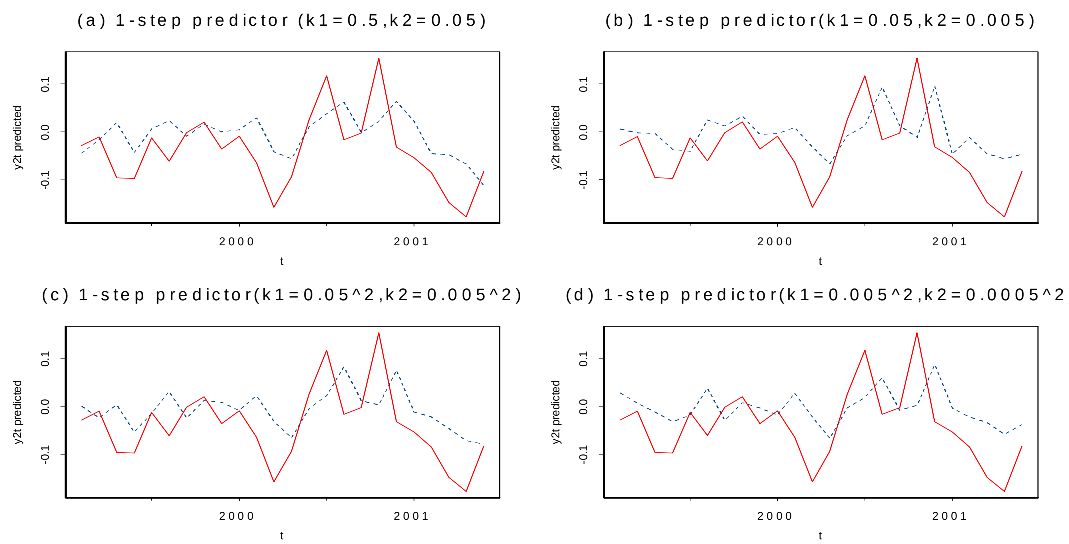

Figure 25 Time plot of one-step ahead forecasting (y_2)

Figure 26 and 27 present the time plot of multi-step ahead forecasting at $(k_1, k_2) \in \{(0.5, 0.05), (0.05,\ 0.005), (0.05^2, 0.005^2), (0.005^2, 0.0005^2)\}$ for y_1 and y_2, and figure 28 and 29 shows the RMSE's of one-step ahead and multi-step ahead forecasting. They show that there are negligible differences among the different choices of (k_1, k_2). In conclusion, threshold VAR analysis for the interest rates is robust to hyperparameter specification in the forecasting viewpoint.

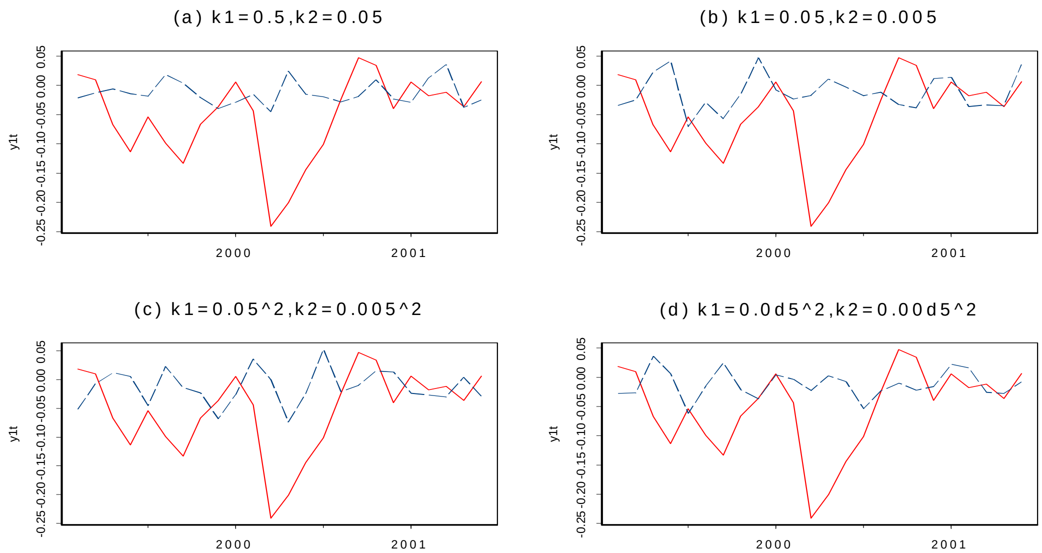

Figure 26 Time plot of multi-step ahead forecasting (y_1)

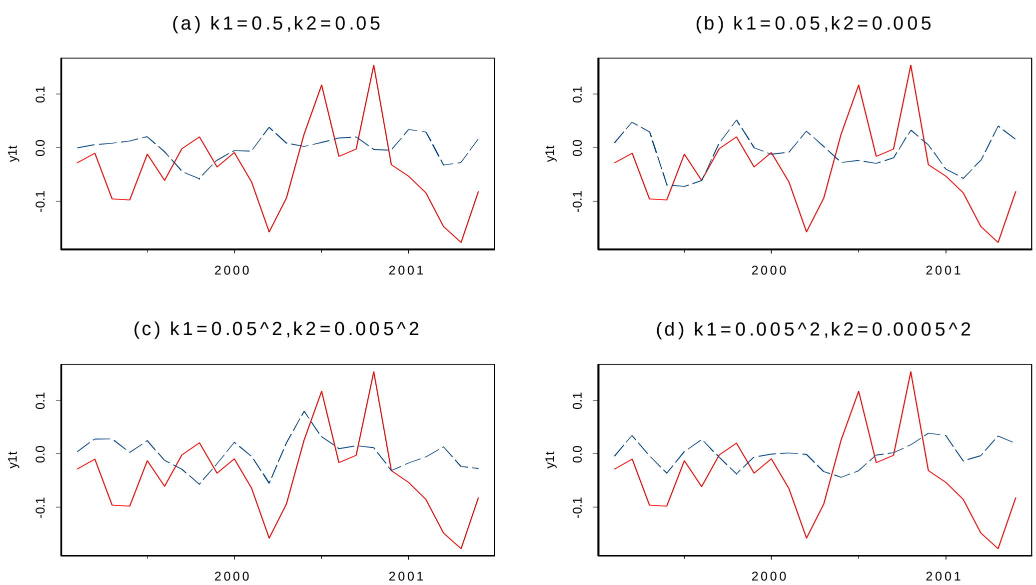

Figure 27 Time plot of multi-step ahead forecasting (y_2)

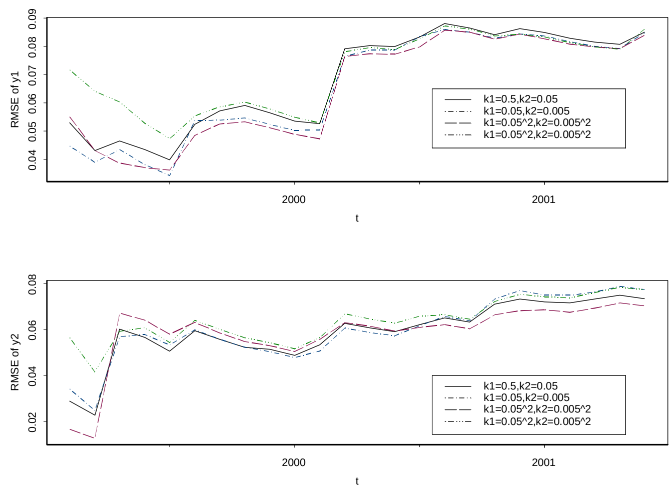

Figure 28 Time plot of RMSE of one-step ahead forecasting

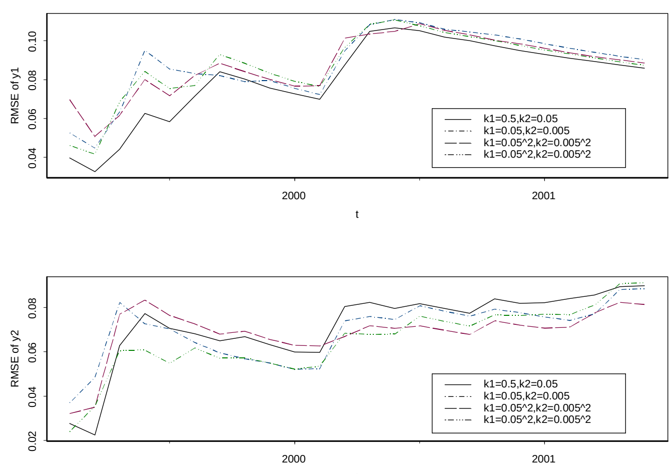

Figure 29 Time plot of RMSE of multi-step ahead forecasting

Chapter 7 Summary and conclusion

7.1 Summary

The main goal of this thesis is to propose a Bayesian modeling procedure for threshold autoregressions. To this end, three major purposes were presented at the outset; (1) to obtain the marginal posterior densities in closed form; (2) to propose Bayesian testing and model determination procedures; and (3) to propose a method of model validation and sensitivity analysis for the hyperparameters of prior densities.

To achieve this goal, the analytical framework of a Bayesian SETAR and a threshold VAR model were provided in the first place when conjugate prior densities are specified for autoregressive parameters and noninformative priors for the delay and threshold parameters. For the estimation of parameters, a Markov-Chain Monte Carlo (MCMC) simulation methods was developed. For model choice, this study showed that Bayes factors are reliable methods of testing in model comparison, AR-lag order selection, and threshold nonlinearity test. Together with Bayes factors, we studied performances of information criteria such as SBC, AIC, ICOMP, CAICF_E, and BMS for AR-lag order selection and found that they may be good alternatives in small samples. A few approximation methods of the integrated likelihood (or marginal likelihood) as a element of Bayes factor were discussed and appropriate computational algorithm was investigated. In addition, forecasting performance was recommended to use in model validation and sensitivity analysis on hyperparameters.

The proposed methodology was applied to simulated data and real data of univariate and vector time series. The algorithm for estimating the delay and threshold parameter was proven to be a stable process in the simulation study for both SETAR and threshold VAR.

Also, through the lag order selection experiments, the performances of Bayes factors and Schwarz Bayesian criterion (SBC) together with those of AIC, ICOMP, CAICF_E, and BMS were examined with varying sizes of samples.

For the threshold nonlinearity testing and the number of regime selection, the Bayes factors based on posterior samples were computed for simulation studies and real examples. Although full consideration of information criterion is not applicable, the quantity evaluated at the posterior mode of the delay and threshold parameters was computed as well as Bayes factor based on the Laplace approximation. In the simulation study, Bayes factors were proven to give stable results in selection of true models. The Bayes factors applied in annual mean of sunspot numbers and US interest rates supported the model identification results of previous studies. On the contrary, there were some disagreements on AR-lag order selection between the Bayes factors, SBC, AIC, and ICOMP and the previous research. A major finding is that Bayes factors and SBC tend to choose parsimonious models.

The forecasting functions were approximated by a Monte Carlo integration method and applied to real data. It showed that examination of forecasting performance would be a good way of model validation as well as sensitivity check on hyperparameters.

7.2 Conclusion and future study

In conclusion, we developed a Bayesian modeling procedure which has some advantages over the classical maximum likelihood (ML) approach. First, the posterior simulation contains full information due to the draws from joint posterior distribution of the entire parameter space. The normal approximation of some posterior densities may mislead the true densities which is displayed in some empirical studies. Second, simulation based estimation

method is easy to implement and interpret. Although some estimation process requires a little more CPU time and capacity than the classical ML approach, rapid development of information technology and the Markov chain Monte Carlo simulation technique make such routine calculations more handy.

There are several limitations in this study which motivate future study. First, I considered only the diagonal elements of covariance matrix for hyperparameter specifications. The main difficulty of full matrix lies with reasonable choice of the off-diagonal elements. One possibility is to use data adaptive prior specification. Second, some applications of threshold model with autoregressive conditional heteroscedasity (ARCH/GARCH) process and vector cointegrated systems will be welcomed considering that threshold models are flexible to other model settings. In addition, comparative studies of similar type of models such as Markov-switching models in Bayesian framework would be highly desirable. Third, more empirical studies in a varied disciplines are also desirable. Finally, it was noticed that the forecasting performance is very poor for the annual mean of sunspot numbers. Sometimes, periodically varying components in economics and hydrological time series such as monthly sales and river flow can be modeled by a nonparametric approach which is a useful technique in forecasting performance as shown in Lewis and Ray (2002).

References

Akaike,H. (1973). *Information Theory and an Extension of the Maximum Likelihood Principles*. in B.N.Petrov and F Csaski, eds., Second International Symposium of Information Theory, Academia Kiado, Budapest.

Albert, J.A. and S.Chib. (1993). *Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shift*. Journal of Business and Economic Statistics, v.11. N.1. 1-15.

Altissimo, F., and G.L.Violante. (2001). *The Nonlinear Dynamics of Output and Unemployment In the U.S*. Journal of Applied Econometrics 16, 461-486.

Andrews, D. W. and W. Proberger. (1994). *Optimal Tests When a Nuisance Parameter is Present only Under the alternative*. Econometrica, 62(6), 1383-1414.

Baum, C.F., J.T. Barkoulas, and M. Caglayan. (2001). *Nonlinear adjustment to purchasing Power Parity in the Post-Bretton Woods era*. Journal of International Money and Finance, 20, 379-399.

Balke, N.S. and T.B. Fomby. (1997). *Threshold Cointegration*. International Economic Review, 38(3). 627-645.

Bensmail, H., G. Celeux, A. E. Raftery, and C. Robert. (1997) *Inference in Model-based Cluster Analysis*. Computing and Statistics, 7, 1-10.

Bozdogan, H. (1997). *Akaike's Information Criterion and Recent Developments in Informational Complexity*. Working Paper, University of Tennessee.

Bozdogan, H. and P.M. Bearse. (1999). *Information Complexity Criteria for Detecting Influential Observations in Dynamic Multivariate Linear Models using the Genetic Algorithm*. Working paper, University of Tennessee.

Bozdogan, H. and M. Ueno. (2000). *A Unified Approach to Information-Theoretic and Bayesian Model Selection Criteria*. Paper presented in the Technical Session Track C, The

6th World Meeting of the International Society for Bayesian Analysis, Japan.

Bozdogan, H. and P. Bearse. (2003). *Information Complexity Criteria for Detecting Influential Observations in Dynamic Multivariate Linear Models using the Genetic algorithm*. Journal of Statistical Planning and Inference, 114, 31-44.

Broemeling, L.D. and P. Cook. (1993). *Bayesian Analysis of Threshold Autoregressions*. Communication in Statistics-Theory and Methods. 21 (9), 2459-2482.

Caner,M. and B.E. Hansen (2001). *Threshold Autoregression with a Unit Root*. Econometrica, 69, 1555-1596.

Chan, K.S. (1990). *Testing For Threshold Autoregression*. Annals of Statistics, 18(4), 1886-1894.

Chan, K.S. (1991). *Percentage Points of Likelihood Ratio Tests for Threshold Autoregression*. Journal of Royal Statistical Society. 53(3), 691-696.

Chan, K.S. (1993). *Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model*. Annals of Statistics, 21, 520-533.

Chen, C.W.S. and J.C. Lee. (1995). *Bayesian Inference of Threshold Autoregressive Models*. Journal of Time Series Analysis, 16(5), 483-492.

Chen, M-H. (1992). *Importance Weighted Marginal Bayesian Posterior Density Estimation*. Technical Report 92-22, Purdue University, Dept. of Statistics.

Chib, S. (1995). *Marginal Likelihood from the Gibbs Output*. Journal of the American Statistical Association, 90, 1313-1321.

Chib, S and I. Jeliazkov. (2002). *Accept-Reject Metropolis-Hastings Sampling and Marginal Likelihood Estimation*. Technical Report, Washington University, St. Louis.

Cook,P., and L.D. Broemeling. (1996). *Analyzing Threshold Autoregressions with a Bayesian Approach*. in T.B. Fomby (ed.), Advances in Econometrics: Bayesian Methods

Applied to Time series Data, Vol. 11. Part B. Greenwich: JAI Press. 89-107.

De Gooijter, J.G. (2001). *Cross-Validation Criteria for SETAR Model Selection*. Journal of Time Series Analysis, 22(3), 267-281.

Fernandez,C. and M.Steel. (1997). *Multivariate Student-t Regression Models*: *Pitfalls and Inference*. Center for Economic Research and Department of Econometrics, Tilburg University, The Netherlands.

Forbes, C.S., G.R.J. Kalb, and P. Kofman. (1999). *Bayesian Threshold Analysis*. Journal of Business and Economic Statistics, 17(3), 364-372.

Gelfand, A.E. and D.K.Dey. (1994). *Bayesian Model Choice* : *Asymptotics and Exact Calculations*. Journal of the Royal Statistical Society, 56(3), 501-514.

Gelfand, A. and A.F.M. Smith. (1990). *Sampling-Based Approaches to Calculating Marginal Densities*. Journal of the American Statistical Association, 85, 398-409.

Gelman,A., J.B.Carlin, H.S.Stern, and D.B.Rubin. (1995). *Bayesian Data Analysis*. Chapman and Hall/CRC, USA.

Geweke, J. (1995). *Monte Carlo simulation and numerical integration*. in H.Amman, D. Kendrick and J. Rust (eds), North-Holland, Amsterdam.

Geweke,J. and N. Terui. (1993). *Bayesian Threshold Autoregressive models for Nonlinear Time series*. Journal of Time Series Analysis 14(5). 441-454.

Granger,C.W. and T. Tervrasvirta. (1993). *Modelling nonlinear economic relationships*. Oxford University Press.

Hansen, B.E. (1996). *Inference When a Nuisance Parameter is not Identified Under the Null Hypothesis*. Econometrica, 64(2),413-430

Hansen, B.E. (1999). *Threshold Effects in Non-Dynamic Panels*: *Estimation*, *Testing*, *and Inference*. Journal of Econometrics, 93, 345-368.

Hansen, B.E. (2000). *Sample Spliting and Threshold Estimation*. Econometrica, 68(3), 575-603.

Hansen, B.E. and B. Seo. (2002). *Testing for two-regime threshold cointegration in vector error correction models*. Journal of Econometrics, 110, 293-318.

Jeffreys, H. (1961). *Theory of Probability*. London

Kadiyala, K. R. and S. Karlsson. (1997). *Numerical Methods for Estimation and Inference in Bayesian VAR-Models*. Journal of Applied Econometrics, 12, 99-132.

Karlsson, S. (2001). *Bayesian Methods in Econometrics Linear Regression*. http://web.hhs.se/stat/courses/bayes/regression.pdf.

Kapetanios, G. (2001). *Model Selection in Threshold Models*. Journal of Time Series Analysis, 22(6), 733-754.

Koop,G., M.H.Pessaran, and S.Potter. (1996). *Impulse Response Analysis in Nonlinear Multivariate Models*. Journal of Econometrics, 74, 119-147

Koop,G. and S. Potter. (1995). *Nonlinearity, structural breaks, or outliers in economic time series*. in (ed.) Barnett, W.A. et. al., Nonlinear econometric modelling in the time series analysis, Proceedings of the eleventh international symposium in economic theory. Cambridge University Press.

Koop,G. and S.M.Potter. (1999). *Dynamic Asymmetries in U.S. Unemployment*. Journal of Business and Economic Statistics, 17(3), 298-312.

Koop,G. and S.M.Potter. (1999). *Bayes Factors and Nonlinearity : Evidence from Economic Time Series*. Journal of Econometrics, 88, 251-281.

Lanne,M. and P. Saikkonen. (2002), "Threshold Autoregressions for Strongly Autocorrelated Time Series," Journal of Business and Economic Statistics. 20(2)

Lewis, S.M. and R.E. Raftery. (1994). *Estimating Bayes Factors via Posterior*

Simulation with the Laplace-Metropolis Estimator. Technical Report, University of Washington.

Lewis, P.A.W. and B.K.Ray. (2002). *Nonlinear Modelling of Periodic Threshold Autoregression Using TSMARS*. Journal of Time Series Analysis, 23(4).

Litterman, R.B. (1980). *A Bayesian Procedures for forecasting with Vector Autoregressions*. Mimeo, MIT.

Litterman, R.B. (1986). *Forecasting with Bayesian Vector Autoregressions - Five Years of Experience*. Journal of Business and Economic Statistics, 4(1).

Lubrano,M. (1995). *Bayesian analysis of nonlinear time series models with a threshold*. in (ed.) Barnett, W.A., et. al., Nonlinear econometric modelling in the time series analysis, Proceedings of the eleventh international symposium in economic theory. Cambridge University Press.

Meng, X.L. and W.H. Wong. (1996). *Simulating Ratios of Normalizing constants via A Simple Identity* : *A Theoretical Explanation*. Statistica Sinica 6, 831-860

Newton, M.A. and A.E. Raftery. (1994). *Approximate Bayesian Inference With the Weighted Likelihood Bootstrap*. Journal of the Royal Statistical Society, Ser. B. 56, 3-48

Okal, N. and D.R.Osborn. (2000). *Business Cycle Non-linearities in UK Consumption and Production*. Journal of Applied Econometrics, 15, 27-43.

Pesaran, M.H. and A. Deaton. (1978). *Testing non-nested nonlinear regression models*. Econometrica, 46, 677-694.

Raftery, A.E. (1994). *Bayes Factors and BIC*: *Comment on Weakliem*. University of Washington, Technical Report. N. 34.

Qian,L. (1998). *On Maximum Likelihood Estimators for Threshold Autoregression*. Journal of Statistical Planning and Inference, 75, 21-46.

Skalin, J. and T. Terasvirta. (1999). *Another Look at Swedish Business Cycles*, 1861-1988. Journal of Applied Econometrics, 14, 359-378.

Tong, H. (1978). *On a Threshold Model*. Pattern Recognition and Signal Processing, ed. C. H. Chen, The Netherlands: Sijthoff and Noordhoff.

Tong, H. (1983). *Threshold Models in Non-Linear Time Series Analysis*. Springer-Verlag New York Inc., USA.

Tong,H., and K.S. Lim. (1980). *Threshold Autoregression*, *Limit Cycles and Cyclical Data* (*with Discussion*). Journal of Royal Statistical Society, B. 42. 245-292

Tsay, R.S. (1989). *Testing and Modelling Threshold Autoregressive Processes*. Journal of the American Statistical Association, 84(405), 231-240

Tsay, R.S. (1998). *Testing and Modelling Multivariate Threshold Models*. Journal of the American Statistical Association, 93(443), 1188-1998

Verdinelli, I and L Wasserman. (1995). *Computing Bayes factors Using a Generalization of the Savage-Dickey Density Ratio*. Journal of American Statistical Association, v. 90. N. 430, 614-618

Appendix

I. Performance study on lag order selection

1. SETAR(2;2,2)

Table A1 SETAR ($n = 100,\ n_1 \approx 65, n_2 = 35$)

(a) Lower regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)	52	19	56	4	44	3	44	4	41
VAR(2)	47	73	43	61	55	42	53	61	56
VAR(3)	1	7	1	19	1	24	3	16	3
VAR(4)		1		10		18		12	
VAR(5)				6		13		7	

(b) Upper regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)	65	22	70	8	58	2	53	5	54
VAR(2)	35	70	30	62	42	51	47	54	46
VAR(3)		4		13		18		16	
VAR(4)		2		7		12		10	
VAR(5)		2		10		17		15	

Table A2 SETAR ($n = 200,\ n_1 \approx 120, n_2 = 80$)

(a) Lower regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)	4	1	6		3		2		1
VAR(2)	96	94	94	76	97	59	98	70	99
VAR(3)		2		10		11		11	
VAR(4)				4		9		5	
VAR(5)		3		10		21		14	

(b) Upper regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)	35	6			21		17		19
VAR(2)	65	90	41	74	77	56	81	59	79
VAR(3)		2	58	12	2	17	2	20	
VAR(4)		1	1	7		12		10	
VAR(5)		1		7		15		11	

Table A3 SETAR ($n = 500,\ n_1 \approx 300, n_2 = 200$)

(a) Lower regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)									
VAR(2)	99	98	100	79	99	69	99	78	99
VAR(3)	1	2		12	1	14	1	12	1
VAR(4)				5		8		6	
VAR(5)				4		9		4	

(b) Upper regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)									
VAR(2)	100	100	99	78	100	57	100	61	100
VAR(3)			1	13		21		17	
VAR(4)				5		11		9	
VAR(5)				4		11		13	

Table A4 SETAR ($n = 1000,\ n_1 \approx 300, n_2 = 200$)

(a) Lower regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)									
VAR(2)	100	98	100	72	99	63	99	67	99
VAR(3)		2		16	1	21	1	20	1
VAR(4)				5		7		6	
VAR(5)				7		9		7	

(b) Upper regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)									
VAR(2)	100	99	100	81	100	57	100	60	100
VAR(3)		1		11		19		21	
VAR(4)				5		10		7	
VAR(5)				3		14		12	

2. Threshold VAR(2;2,2)

Table A5 Threshold VAR ($n = 150,\ n_1 \approx 100, n_2 \approx 50$)

(a) Lower regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)	1		1						
VAR(2)	99	100	99	77	100	98	100	56	100
VAR(3)				16		2		24	
VAR(4)				4				11	
VAR(5)				3				9	

(b) Upper regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)	59	7	38	1	21	3	28	9	29
VAR(2)	41	93	62	80	79	85	72	91	79
VAR(3)				7		4			
VAR(4)				7		4			
VAR(5)				5		4			

Table A6 Threshold VAR ($n = 250,\ n_1 \approx 180, n_2 \approx 65$)

(a) Lower regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)									
VAR(2)	100	100	100	91	100	100	100	60	100
VAR(3)				5				17	
VAR(4)				3				9	
VAR(5)				1				14	

(b) Upper regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)	16	3	10		4	1	6		5
VAR(2)	84	96	90	81	96	98	94	87	95
VAR(3)				14				13	
VAR(4)				2					
VAR(5)				3					

Table A7 Threshold VAR ($n = 500,\ n_1 \approx 380, n_2 \approx 120$)

(a) Lower regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)									
VAR(2)	100	100	100	85	100	100	100	36	100
VAR(3)				8				19	
VAR(4)				4				22	
VAR(5)				3				23	

(b) Upper regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)									
VAR(2)	100	100	100	85	100	100	100	68	100
VAR(3)				9				19	
VAR(4)				4				4	
VAR(5)				2				9	

Table A8 Threshold VAR ($n = 1000,\ n_1 \approx 750, n_2 \approx 250$)

(a) Lower regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)									
VAR(2)	100	100	100	81	100	100	100	27	100
VAR(3)				14				21	
VAR(4)								19	
VAR(5)								33	

(b) Upper regime

Model	*BF*	$SBC_{no\ prior}$	SBC_{prior}	$AIC_{no\ prior}$	AIC_{prior}	$ICOMP_{no\ prior}$	$ICOMP_{prior}$	*CAICF*	*BMS*
VAR(1)									
VAR(2)	100	100	100	84	100	100	100	54	100
VAR(3)				13				18	
VAR(4)				2				10	
VAR(5)				1				18	

II. Bayes factor and information criteria for AR lag-order selection

1.SETAR(2;2,2)

Table A9 Log integrated likelihood and information criteria [Linear AR(p)]

Model	Log integrated likelihood	AIC	SBC	ICOMP
$AR(1)$	–423.66	844.89	830.08	841.39
$AR(2)$	–430.41	858.36	832.26	851.16
$AR(3)$	–404.70	805.27	769.53	793.84
$AR(4)$	–411.08	817.94	770.64	802.44
$AR(5)$	–417.70	831.12	772.07	811.52

Table A10 Log integrated likelihood and information criteria (2-regime SETAR: Lower regime)

Model	Log integrated likelihood	AIC	SBC	ICOMP
$AR(1)$	–1070.8	2122.4	2136.0	2131.1
$AR(2)$	–1031.2	2031.1	2055.3	2045.3
$AR(3)$	–1036.6	2030.4	2066.0	2050.8
$AR(4)$	–1040.2	2025.5	2072.7	2052.2
$AR(5)$	–1047.3	2027.3	2086.5	2060.7

Table A11 Log integrated likelihood and information criteria (2-regime SETAR: Upper regime)

Model	Log integrated likelihood	AIC	SBC	ICOMP
$AR(1)$	–620.02	1222.0	1234.2	1229.5
$AR(2)$	–580.51	1131.9	1153.3	1143.6
$AR(3)$	–586.68	1133.6	1164.9	1150.4
$AR(4)$	–592.73	1135.2	1176.7	1157.4
$AR(5)$	–599.05	1136.6	1188.5	1164.3

2. Annual mean of sunspot numbers

Table A12 Log integrated likelihood and information criteria (Linear AR(p) model)

Model	Log likelihood	Log integrated likelihood	AIC	ICOMP	SBC
$AR(1)$	–1197.3	–1252.1	2398.6	2517.0	2517.5
$AR(2)$	–1113.6	–1173.4	2233.2	2356.3	2357.2
$AR(3)$	–1111.3	–1176.6	2230.6	2363.3	2363.9
$AR(4)$	–1111.1	–1182.0	2232.2	2373.8	2374.9
$AR(5)$	–1110.9	–1187.4	2233.9	2384.7	2386.1
$AR(6)$	–1107.6	–1189.5	2229.1	2388.6	2390.4
$AR(7)$	–1102.1	–1189.5	2220.1	2387.9	2390.3
$AR(8)$	–1096.0	–1188.9	2210.0	2385.8	2388.9
$AR(9)$	–1088.9	–1187.2	2197.8	2381.3	2385.3
$AR(10)$	–1088.9	–1192.7	2199.8	2391.8	2396.6
$AR(11)$	–1088.5	–1197.8	2201.1	2401.5	2407.1
$AR(12)$	–1088.1	–1202.8	2202.2	2411.1	2417.4
$AR(13)$	–1088.0	–1208.2	2204.1	2421.4	2428.6
$AR(14)$	–1087.5	–1213.2	2205.1	2430.7	2438.8
$AR(15)$	–1086.8	–1218.0	2205.7	2439.4	2448.5
$AR(16)$	–1086.5	–1223.1	2207.1	2449.1	2459.1
$AR(17)$	–1082.6	–1224.6	2201.3	2450.7	2461.9
$AR(18)$	–1081.8	–1229.3	2201.6	2459.0	2471.4

Table A13 Log integrated likelihood and information criteria (2-regime SETAR: lower regime)

Model	Log likelihood	Log integrated likelihood	AIC	ICOMP	SBC
$AR(1)$	–464.79	–498.02	933.59	1011.09	1010.08
$AR(2)$	–442.39	–479.50	890.78	973.89	972.88
$AR(3)$	–439.18	–480.46	886.36	977.47	976.50
$AR(4)$	–438.04	–483.69	886.09	985.55	984.56
$AR(5)$	–436.35	–486.46	884.70	992.33	991.35
$AR(6)$	–436.15	–490.95	886.30	1002.55	1001.41
$AR(7)$	–434.54	–494.11	885.08	1009.54	1008.27
$AR(8)$	–434.36	–498.61	886.11	1019.22	1017.64
$AR(9)$	–434.02	–503.73	888.04	1030.05	1028.82
$AR(10)$	–433.99	–508.82	889.98	1040.81	1038.41
$AR(11)$	–433.97	–513.87	891.94	1051.53	1048.82
$AR(12)$	–433.96	–518.77	893.93	1062.06	1059.27
$AR(13)$	–433.96	–523.56	895.92	1072.45	1069.72
$AR(14)$	–433.94	–528.06	897.87	1082.47	1080.12
$AR(15)$	–433.36	–531.91	898.73	1090.98	1089.20
$AR(16)$	–433.36	–536.40	900.72	1100.95	1099.64
$AR(17)$	–429.41	–536.91	894.83	1100.67	1100.43
$AR(18)$	–428.64	–540.93	895.27	1108.75	1108.88

Table A14 Log integrated likelihood and information criteria (2-regime SETAR: upper regime)

Model	Log likelihood	Log integrated likelihood	AIC	ICOMP	SBC
$AR(1)$	–634.08	–673.53	1272.2	1355.5	1356.2
$AR(2)$	–623.87	–668.54	1253.7	1344.7	1345.2
$AR(3)$	–609.27	–659.12	1226.5	1324.0	1324.7
$AR(4)$	–608.12	–663.40	1226.2	1331.8	1332.4
$AR(5)$	–607.84	–668.57	1227.7	1341.3	1341.9
$AR(6)$	–606.47	–672.58	1226.9	1348.3	1349.1
$AR(7)$	–605.95	–677.41	1227.9	1357.0	1358.0
$AR(8)$	–605.94	–682.49	1229.9	1366.5	1368.1
$AR(9)$	–593.41	–674.43	1206.8	1347.7	1350.8
$AR(10)$	–591.01	–677.30	1204.0	1351.4	1355.4
$AR(11)$	–580.82	–671.96	1185.6	1337.2	1342.8
$AR(12)$	–580.44	–676.78	1186.9	1345.3	1351.7
$AR(13)$	–579.76	–681.34	1187.5	1352.7	1360.0
$AR(14)$	–577.34	–684.19	1184.7	1356.1	1364.3
$AR(15)$	–577.33	–689.54	1186.7	1365.1	1374.0
$AR(16)$	–577.27	–694.80	1188.5	1373.9	1383.6
$AR(17)$	–576.06	–698.89	1188.1	1380.0	1390.6
$AR(18)$	–573.13	–701.11	1184.3	1381.8	1393.6

3.US interest rate

Table A15 Log integrated likelihood and information criteria (Linear VAR)

Model	Log likelihood	Log integrated likelihood	AIC	ICOMP	SBC
VAR(1)	750.63	730.88	–1483.3	–1464.5	–1491.7
VAR(2)	774.79	746.90	–1523.6	–1504.2	–1524.4
VAR(3)	782.02	746.08	–1530.0	–1510.6	–1523.1
VAR(4)	783.52	739.56	–1525.0	–1505.7	–1510.1
VAR(5)	788.70	736.74	–1527.4	–1508.3	–1504.6
VAR(6)	799.64	739.69	–1541.3	–1522.5	–1511.1
VAR(7)	805.76	737.83	–1545.5	–1526.8	–1507.8
VAR(8)	809.04	733.16	–1544.1	–1525.8	–1498.7
VAR(9)	810.10	726.19	–1538.2	–1520.2	–1485.0
VAR(10)	811.58	719.80	–1533.2	–1515.6	–1472.2
VAR(11)	814.60	714.90	–1531.2	–1514.2	–1462.7
VAR(12)	824.84	717.24	–1543.7	–1526.9	–1468.4
VAR(13)	827.67	712.19	–1541.3	–1524.7	–1458.6
VAR(14)	829.00	705.67	–1536.0	–1519.5	–1445.6
VAR(15)	832.64	701.45	–1535.3	–1518.8	–1437.6

Table A16 Log integrated likelihood and information criteria (two-regime threshold VAR: lower regime)

Model	Log likelihood	Log integrated likelihood	AIC	ICOMP	SBC
VAR(1)	76.77	60.97	–135.55	–117.52	–153.54
VAR(2)	82.89	61.08	–139.79	–120.55	–155.50
VAR(3)	84.28	56.61	–134.57	–114.77	–147.47
VAR(4)	84.92	51.72	–127.84	–108.24	–137.87
VAR(5)	86.51	47.89	–123.03	–103.83	–130.54
VAR(6)	91.70	47.72	–125.42	–106.02	–131.67
VAR(7)	94.87	45.67	–123.76	–103.42	–128.41
VAR(8)	96.86	42.42	–119.73	–98.83	–122.58
VAR(9)	103.75	44.30	–125.50	–104.38	–129.00
VAR(10)	105.77	41.18	–121.54	–99.89	–123.82
VAR(11)	108.37	38.76	–118.75	–94.82	–120.24
VAR(12)	118.60	44.00	–131.20	–106.61	–136.56
VAR(13)	120.02	40.90	–126.06	–100.15	–130.63
VAR(14)	120.35	36.70	–118.72	–90.90	–121.89
VAR(15)	123.66	35.69	–117.33	–86.99	–121.17

Table A17 Log integrated likelihood and information criteria (two-regime threshold VAR: upper regime)

Model	Log likelihood	Log integrated likelihood	AIC	ICOMP	SBC
VAR(1)	716.92	698.18	–1415.8	–1394.0	–1428.2
VAR(2)	733.70	707.86	–1441.4	–1418.3	–1447.6
VAR(3)	742.56	709.67	–1451.1	–1427.8	–1451.1
VAR(4)	744.73	704.64	–1447.5	–1424.1	–1441.0
VAR(5)	748.32	701.01	–1446.6	–1423.2	–1433.9
VAR(6)	753.42	698.91	–1448.8	–1425.7	–1430.0
VAR(7)	759.21	697.50	–1452.4	–1429.6	–1427.6
VAR(8)	762.90	694.08	–1451.8	–1429.6	–1420.9
VAR(9)	764.77	688.76	–1447.5	–1425.7	–1410.5
VAR(10)	766.18	683.04	–1442.4	–1421.0	–1399.1
VAR(11)	771.82	681.62	–1445.6	–1425.0	–1396.7
VAR(12)	775.64	678.32	–1445.3	–1425.1	–1390.5
VAR(13)	777.71	673.17	–1441.4	–1421.3	–1380.6
VAR(14)	780.31	668.61	–1438.6	–1418.4	–1372.0
VAR(15)	781.47	662.52	–1432.9	–1412.8	–1360.2

Table A18 Log integrated likelihood and information criteria (three-regime VAR: lower regime)

Model	Log likelihood	Log integrated likelihood	AIC	ICOMP	SBC
VAR(1)	85.8	69.9	−153.7	−135.7	−171.5
VAR(2)	91.8	69.8	−157.7	−138.8	−173.6
VAR(3)	93.6	65.7	−153.3	−133.9	−165.7
VAR(4)	94.3	60.9	−146.7	−127.5	−156.1
VAR(5)	95.9	56.9	−141.8	−123.0	−148.5
VAR(6)	101.6	57.3	−145.3	−126.3	−150.6
VAR(7)	105.3	55.6	−144.6	−124.7	−148.3
VAR(8)	106.6	51.5	−139.2	−119.0	−140.7
VAR(9)	113.4	53.0	−144.8	−123.5	−146.5
VAR(10)	114.7	49.0	−139.4	−117.5	−139.3
VAR(11)	117.6	46.7	−137.2	−113.5	−136.2
VAR(12)	127.2	51.1	−148.4	−123.9	−150.2
VAR(13)	128.3	47.5	−142.6	−116.8	−143.1
VAR(14)	128.9	43.4	−135.8	−108.4	−134.7
VAR(15)	133.4	43.4	−136.9	−107.4	−137.0

Table A19 Log integrated likelihood and information criteria (three-regime threshold VAR: middle regime)

Model	Log likelihood	Log integrated likelihood	AIC	ICOMP	SBC
VAR(1)	563.4	545.9	–1108.9	–1085.4	–1126.9
VAR(2)	577.9	554.2	–1129.9	–1105.2	–1144.1
VAR(3)	584.6	554.7	–1135.3	–1110.3	–1145.6
VAR(4)	585.9	549.5	–1129.9	–1104.6	–1136.0
VAR(5)	588.9	545.9	–1127.8	–1102.3	–1130.0
VAR(6)	590.4	540.8	–1122.9	–1097.3	–1121.0
VAR(7)	594.0	537.8	–1122.0	–1096.4	–1116.4
VAR(8)	606.4	543.9	–1138.9	–1114.1	–1130.7
VAR(9)	609.8	540.9	–1137.6	–1113.3	–1125.9
VAR(10)	611.4	536.1	–1132.8	–1109.1	–1117.4
VAR(11)	618.8	537.2	–1139.6	–1116.7	–1121.5
VAR(12)	625.4	537.4	–1145.0	–1122.3	–1124.1
VAR(13)	626.4	531.8	–1138.8	–1116.2	–1114.1
VAR(14)	629.8	528.7	–1137.6	–1114.7	–1110.1
VAR(15)	631.2	523.5	–1132.4	–1109.6	–1101.5

Table A20 Log integrated likelihood and information criteria (three-regime threshold VAR: upper regime)

Model	Log likelihood	Log integrated likelihood	AIC	ICOMP	SBC
VAR(1)	169.7	154.4	–321.5	–296.6	–341.5
VAR(2)	175.0	154.8	–324.1	–297.2	–341.7
VAR(3)	177.6	152.8	–321.1	–293.3	–336.2
VAR(4)	179.3	150.1	–316.6	–288.4	–329.2
VAR(5)	181.2	147.8	–312.5	–284.5	–322.8
VAR(6)	186.4	148.8	–314.9	–287.4	–323.8
VAR(7)	193.5	151.6	–321.0	–294.6	–329.5
VAR(8)	193.7	147.8	–313.4	–286.7	–319.6
VAR(9)	198.6	148.8	–315.3	–287.1	–320.6
VAR(10)	201.5	147.8	–313.0	–283.6	–317.2
VAR(11)	206.6	149.2	–315.3	–284.9	–319.3
VAR(12)	208.5	147.3	–311.1	–279.0	–313.9
VAR(13)	210.1	145.4	–306.2	–271.5	–307.8
VAR(14)	213.3	145.1	–304.6	–268.2	–305.9
VAR(15)	214.6	142.9	–299.3	–260.7	–299.4

III. RMSE's of forecasting exercise

1. Sunspot numbers

Table A21 RMSE of multi-step ahead forecasting

	λ			
Lead time	0.5	0.05	0.005	0.005^2
1	7.8	8.2	10.2	5.1
2	6.5	8.3	10.3	4.1
3	6.0	7.2	8.4	4.7
4	6.9	7.9	8.4	4.7
5	7.3	8.0	8.8	4.5
6	10.3	10.7	14.6	5.0
7	12.5	12.4	17.5	5.9
8	11.7	11.7	16.8	5.7
9	11.0	11.0	15.9	5.8
10	10.6	10.5	15.3	6.0
11	10.8	11.0	16.4	6.6
12	12.2	12.3	18.6	7.5
13	11.9	12.3	18.6	7.3
14	13.9	13.0	18.8	9.5
15	21.3	18.6	25.0	16.0
16	26.9	23.6	30.3	20.1
17	29.8	26.2	33.0	22.2
18	32.9	29.3	35.3	25.0
19	32.7	29.4	35.4	25.0
20	31.9	28.8	34.5	24.6
21	31.4	28.2	34.2	24.0
22	31.7	28.4	34.2	23.9
23	31.5	28.1	33.6	23.5
24	32.4	28.1	34.8	24.4
25	36.7	34.4	42.1	30.8

Table A21 (continued.)

	λ			
Lead time	0.5	0.05	0.005	0.005^2
26	40.9	38.8	47.3	35.3
27	45.1	43.2	51.6	39.9
28	45.9	44.1	51.9	41.0
29	46.0	44.5	51.3	41.5
30	45.3	43.7	50.5	40.8
31	44.8	43.2	50.2	40.3
32	44.8	43.3	50.1	40.3
33	44.2	42.8	49.3	39.8
34	46.2	44.7	52.3	41.5
35	52.2	50.8	59.1	47.2
36	57.6	56.3	64.6	52.7
37	61.2	60.0	67.4	56.4
38	62.2	61.1	67.6	57.7
39	61.5	60.6	66.7	57.2
40	60.7	59.8	65.9	56.5
41	60.0	59.1	65.3	55.9
42	59.7	58.8	64.8	55.6
43	59.3	58.5	64.1	55.2
44	58.7	57.9	63.4	54.6
45	58.7	57.9	63.8	54.8
46	59.2	58.4	64.4	55.7
47	60.0	59.2	64.7	56.5
48	60.7	59.9	64.7	56.9
49	60.4	59.6	64.1	56.3
50	59.9	59.1	63.4	55.8

Table A22 RMSE of multi-step ahead forecasting

Lead time	SETAR(2;2,3)	SETAR(2;12,2)*	SETAR(2;9,2)*
1	18.1	10.6	9.6
2	13.5	15.7	19.6
3	11.1	20.0	22.8
4	11.6	21.2	23.8
5	10.3	20.9	24.8
6	10.8	20.1	25.1
7	13.2	19.5	25.1
8	13.5	19.2	25.5
9	13.1	19.3	26.9
10	13.5	19.6	26.9
11	12.6	19.9	26.8
12	12.6	21.1	27.0
13	14.4	22.4	27.8
14	15.6	23.1	28.4
15	15.3	23.0	28.6
16	14.8	23.0	28.4
17	15.3	24.7	28.5
18	17.6	26.9	29.0
19	21.1	32.3	29.7
20	21.4	32.7	34.0
21	21.5	33.3	33.9
22	23.8	34.1	33.7
23	26.0	36.1	33.8
24	27.3	37.8	34.1
25	27.2	38.6	34.1
26	26.7	38.5	33.7
27	26.8	38.3	33.5
28	28.5	38.9	33.6
29	30.0	38.5	34.2
30	30.7		34.5
31	30.8		
32	30.3		
33	30.2		
34	30.8		
35	30.6		

2. US interest rate (multi-step ahead forecasting)

Table A23 RMSE of multi-step ahead forecasting (y_1)

	(λ_1, λ_2)			
Lead time	$(0.5, 0.05)$	$(0.05, 0.005)$	$(0.05^2, 0.005^2)$	$(0.005^2, 0.0005^2)$
1	0.039	0.052	0.069	0.046
2	0.032	0.044	0.050	0.041
3	0.044	0.063	0.061	0.068
4	0.062	0.095	0.079	0.084
5	0.058	0.085	0.071	0.075
6	0.071	0.083	0.082	0.076
7	0.084	0.082	0.088	0.092
8	0.080	0.078	0.084	0.088
9	0.075	0.079	0.079	0.083
10	0.072	0.075	0.076	0.079
11	0.069	0.072	0.076	0.076
12	0.087	0.094	0.101	0.096
13	0.104	0.108	0.103	0.108
14	0.106	0.110	0.104	0.110
15	0.105	0.109	0.108	0.107
16	0.101	0.105	0.105	0.104
17	0.100	0.104	0.103	0.102
18	0.097	0.103	0.100	0.100
19	0.094	0.100	0.098	0.097
20	0.092	0.098	0.096	0.095
21	0.090	0.096	0.093	0.093
22	0.089	0.094	0.091	0.091
23	0.087	0.091	0.090	0.089
24	0.085	0.090	0.088	0.087

Table A24 RMSE of multi-step ahead forecasting (y_2)

	(λ_1, λ_2)			
Lead time	$(0.5, 0.05)$	$(0.05, 0.005)$	$(0.05^2, 0.005^2)$	$(0.005^2, 0.0005^2)$
1	0.027	0.036	0.032	0.023
2	0.022	0.048	0.034	0.035
3	0.062	0.082	0.076	0.060
4	0.077	0.072	0.083	0.060
5	0.070	0.070	0.076	0.054
6	0.068	0.064	0.072	0.061
7	0.065	0.059	0.067	0.057
8	0.066	0.056	0.069	0.057
9	0.063	0.054	0.065	0.054
10	0.059	0.052	0.062	0.052
11	0.059	0.052	0.062	0.053
12	0.080	0.073	0.066	0.068
13	0.082	0.075	0.071	0.067
14	0.079	0.074	0.070	0.067
15	0.081	0.080	0.071	0.076
16	0.079	0.078	0.069	0.073
17	0.077	0.075	0.067	0.071
18	0.083	0.079	0.073	0.076
19	0.081	0.077	0.071	0.076
20	0.082	0.075	0.070	0.076
21	0.083	0.074	0.071	0.076
22	0.085	0.077	0.077	0.080
23	0.089	0.087	0.082	0.090
24	0.089	0.088	0.081	0.091

IV. Computations in Splus

1. SETAR

```
# create data set (simulation)
SETAR(2,1,1)
y<-rnorm(1)
n<-200
for (i in 2:n)
{
if (y[i-1]<=0.4)
{y0<-y[i-1]*(-0.5)+rnorm(1,0,sqrt(2))}
else
{y0<-y[i-1]*(0.5)+rnorm(1,0,sqrt(1))}
y<-c(y,y0) }
SETAR(2;2,2)
y<-rnorm(2)
n<-200
for (i in 3:n)
{
if (y[i-2]<=0)
{y0<-y[i-1]*(-0.7)+y[i-2]*(0.3)+rnorm(1,0,sqrt(2))}
else
{y0<-y[i-1]*(0.7)+y[i-2]*(-0.3)+rnorm(1,0,sqrt(1))}
y<-c(y,y0) }
x<-tsmatrix(y, lag1=lag(y,-1), lag2=lag(y,-2), lag3=lag(y,-3), lag4=lag(y,-4),lag5=lag(y,-5))
# define delay parameter
d<-2
# case ordered data
x<-x[order(x[,(d+1)]),]
ys<-sort(y)
a<-ys[round(0.15*n)]
b<-ys[round(0.85*n)]
u<-runif(1,a,b)
r<-runif(1,a,b)
simu<-0
rej<-0
p1<-5
p2<-5

# Metropolis algorithm
for (it in 1:1000){
data<-pardata.unv20(x,d,u,p1,p2)
x1<-data$x1
y1<-data$y1
x2<-data$x2
y2<-data$y2
hyper<-hyper.unv0(x1,y1)
bhat<-hyper$bhat
ssols<-hyper$ssols
m0<-hyper$m0
b0<-hyper$b0
```

```
v0<-hyper$v0
s0<-hyper$s0
prob1<-prob.unv(x1, y1,bhat,ssols,m0,b0,v0,s0)
hyper<-hyper.unv0(x2,y2)
bhat<-hyper$bhat
ssols<-hyper$ssols
m0<-hyper$m0
b0<-hyper$b0
v0<-hyper$v0
s0<-hyper$s0
prob2<-prob.unv(x2, y2,bhat,ssols,m0,b0,v0,s0)
probu<-prob1+prob2
data<-pardata.unv20(x,d,r,p1,p2)
x1<-data$x1
y1<-data$y1
x2<-data$x2
y2<-data$y2
hyper<-hyper.unv0(x1,y1)
bhat<-hyper$bhat
ssols<-hyper$ssols
m0<-hyper$m0
b0<-hyper$b0
v0<-hyper$v0
s0<-hyper$s0
prob1<-prob.unv(x1, y1,bhat,ssols,m0,b0,v0,s0)
hyper<-hyper.unv0(x2,y2)
bhat<-hyper$bhat
ssols<-hyper$ssols
m0<-hyper$m0
b0<-hyper$b0
v0<-hyper$v0
s0<-hyper$s0
prob2<-prob.unv(x2, y2,bhat,ssols,m0,b0,v0,s0)
probr<-prob1+prob2
prob<-probr-probu
if(prob>log(runif(1)))
{u<-r
probu<-probr
rej<-rej+1}
r<-rnorm(1,u,0.075)
while(!(a<r & r<b))
{r<-rnorm(1,u,0.075)}
simu<-c(simu,u,probu)
}
simu<-simu[-1]
simu<-matrix(simu,ncol=2,byrow=T)

## function to partition data set
pardata.unv20
function(x, d, u, p1, p2)
{
x11 <- x[x[, (d + 1)] <= u, ]
x22 <- x[x[, (d + 1)] > u, ]
n1 <- nrow(x1)
n2 <- nrow(x2)
y1 <- x11[, 1]
y2 <- x22[, 1]
y1 <- matrix(y1, ncol = 1)
y2 <- matrix(y2, ncol = 1)
```

```
x1 <- x11[, 2:(p1 + 1)]
x2 <- x22[, 2:(p2 + 1)]
tmpfil <- tempfile(”pardata.unv20”)
on.exit(unlink(tmpfil))
list(x1 = x1, y1 = y1, x2 = x2, y2 = y2)
}

## function to specify hyper parameters
> hyper.unv0
function(x, y)
{
#ols estimates
p <- ncol(x)
bhat <- solve(t(x) %*% x) %*% t(x) %*% y
ssols <- (t(y - x %*% bhat) %*% (y - x %*% bhat))
#specification of hyper-parameters
k1 <- 0.05
m0 <- diag(c(k1/seq(1:p)))
b0 <- (c(1, rep(0, (p - 1))))
v0 <- p
s0 <- 3
tmpfil <- tempfile(hyper.unv0)
on.exit(unlink(tmpfil))
list(bhat = bhat, ssols = ssols, m0 = m0, b0 = b0,
v0 = v0, s0 = s0)
}

## function to score the joint prob. of delay and threshold parameters
> prob.unv
function(x, y, bhat, ssols, m0, b0, v0, s0)
{
k <- ncol(x)
m1 <- m0 + t(x) %*% x
bpost <- solve(m1) %*% ((m0 %*% b0) + (t(x) %*% x) %*%bhat)
ssb <- t(b0 - bhat) %*% solve((solve(m0) + solve(t(x) %*% x))) %*% (b0 - bhat)
v1 <- nrow(x) + v0
sspost <- (ssols + ssb + s0)
prob <- lgamma(v1/2) + v1/2 * log(pi) - v1/2 * log(sspost) - 1/2 * log(det(m1))
}

## compute posterior probability of delay parameter
probd1<-prob.d2(setard1[501:1000,])
probd2<-prob.d2(setard2[501:1000,])
probd3<-prob.d2(setard3[501:1000,])
all<-probd1+probd2+probd3
pr1<-probd1/all
pr2<-probd2/all
pr3<-probd3/all
proball<-c(pr1,pr2,pr3)
proball
probd1<-prob.d3(simu3d1[501:1000,])
probd2<-prob.d3(simu3d2[501:1000,])
probd3<-prob.d3(simu3d3[501:1000,])
probd4<-prob.d3(sim[501:1000,])
all<-probd1+probd2+probd3+probd4
pr1<-probd1/all
pr2<-probd2/all
pr3<-probd3/all
pr4<-probd4/all
```

```
proball3<-c(pr1,pr2,pr3,pr4)
proball3
```

2. Threshold VAR

```
# create data set (simulation)
set.seed(1000)
mu0<-c(0,0)
vmat0<-diag(2)
y<-matrix(mnorm(2,mu0,vmat0),ncol=2,byrow=T)
vmat1<-matrix(c(1,.2,.2,1),ncol=2)
vmat2<-matrix(c(1,-.3,-.3,1),ncol=2)
beta12<-matrix(c(-.33,0,-.33,.35),ncol=2)
beta11<-matrix(c(.9,0,.9,.7),ncol=2)
beta22<-matrix(c(.3,0,.1,.3),ncol=2)
beta21<-matrix(c(-.7,0,-.3,-.7),ncol=2)
y<-matrix(y,ncol=2)
n<-500
for (i in 3:n)
{
if (y[(i-1),1]<=0)
{y0<-y[(i-1),]%*%beta11+y[(i-2),]%*%beta12+mnorm(1,c(0,0),vmat1)}
else
{y0<-y[(i-1),]%*%beta21+y[(i-2),]%*%beta22+mnorm(1,c(0,0),vmat2)}
y<-rbind(y,y0) }
x<-tsmatrix(y, lag1=lag(y,-1), lag2=lag(y,-2), lag3=lag(y,-3), lag4=lag(y,-4), lag5=lag(y,-5))
k<-ncol(y)
# define delay parameter
d<-3

# create case ordered data
x<-x[order(x[,(k+d)]),]
a<-x[round(0.15*n),(k+d)]
b<-x[round(0.85*n), (k+d)]
u<-runif(1,a,b)
r<-runif(1,a,b)
simud<-0

# Metropolis algorithm
for (it in 1:500)
{
parmdata20(x,k,d, u)
hyperm0(x1, y1)
prob1<-parm(x1, y1, m0,b0,s0,v0)
hyperm0(x2,y2)
prob2<-parm(x2, y2, m0,b0,s0,v0)
probu<-prob1+prob2
parmdata20(x,k, d, r)
hyperm0(x1, y1)
prob1r<-parm(x1, y1, m0,b0,s0,v0)
hyperm0(x2,y2)
prob2r<-parm(x2, y2, m0,b0,s0,v0)
probr<-prob1r+prob2r
prob<-probr-probu
if (prob>=log(runif(1)))
{u<-r
rej<-rej+1
```

```
probu<-probr}
simud<-c(simud,d,u,probu)
r<-rnorm(1,u,sqrt(var(x[,1]))/40)
while(!(a<r & r<b))
{r<-rnorm(1,u,sqrt(var(x[,1]))/40)}
}
simud<-simud[-1]
simud<-matrix(simud,ncol=3,byrow=T)

## function for partition of data
> parmdata20
function(x, d, u, p1, p2)
{
x11 <- x[x[, (d + 2)] <= u, ]
x22 <- x[x[, (d + 2)] > u, ]
n1 <- nrow(x11)
n2 <- nrow(x22)
y1 <- x11[, 1:2]
y2 <- x22[, 1:2]
x1 <- x11[, 3:(2 + p1 * 2)]
x2 <- x22[, 3:(2 + p2 * 2)]
tmpfil <- tempfile(”parmdata20”)
on.exit(unlink(tmpfil))
list(x1 = x1, y1 = y1, x2 = x2, y2 = y2)
}

## function to hyperparameter specification
> hyperm0
function(x, y)
{
#ols estimates
p <- ncol(x)
pp <- p/2
k <- ncol(y)
n <- nrow(x)
bhat <<- solve(t(x) %*% x) %*% t(x) %*% y
ssols <<- (t(y - x %*% bhat) %*% (y - x %*% bhat))
#specification of hyper-parameters
k2 <- (0.005)^2
k1 <- (0.05)^2
v0 <- k + 1
hypervar0(x, y)
m0 <- kronecker(diag(1/seq(1:pp)), diag(c(sig1/n,sig2/n)))
hconst <- c(rep(c(k1, k2), pp), rev(rep(c(k1, k2),pp)))
m0 <- m0 * hconst
m0 <<- (m0)
bpri <- (c(1, rep(0, (p - 1))))
b0 <<- cbind(bpri, 0)
s0 <<- diag(c((sig1)/(n), (sig2)/(n)))
tmpfil <- tempfile(”hyperm1”)
on.exit(unlink(tmpfil))
}

## function to compute s^2
> hypervar0
function(x1, y1)
{
x10 <- x1[, -1]
p1 <- ncol(x10)
```

```
x11 <- x10[, seq(1, (p1 - 1), by = 2)]
x12 <- x10[, seq(2, p1, by = 2)]
y11 <- y1[, 1]
y12 <- y1[, 2]
y11 <- matrix(y11, ncol = 1)
y12 <- matrix(y12, ncol = 1)
x11 <- cbind(1, x11)
x12 <- cbind(1, x12)
bb11 <- solve(t(x11) %*% x11) %*% t(x11) %*% y11
bb12 <- solve(t(x12) %*% x12) %*% t(x12) %*% y12
sig1 <<- (t(y11 - x11 %*% bb11) %*% (y11 - x11 %*%bb11))
sig2 <<- (t(y12 - x12 %*% bb12) %*% (y12 - x12 %*%bb12))
}

## function to compute joint prob. of d and r
> parm
function(x, y, m0, b0, s0, v0)
{
n <- nrow(x)
k <- ncol(y)
p <- (ncol(x) - 1)/k
etha <- p * k + 1
# parameter estimation #
bols <- solve(t(x) %*% x) %*% t(x) %*% y
m1 <- m0 + (t(x) %*% x)
b1 <<- solve(m1) %*% (m0 %*% b0 + t(x) %*% x %*% bols)
sb <- t(b0 - bols) %*% solve(solve(m0) + solve(t(x) %*% x)) %*% (b0 - bols)
so1 <- t(y - x %*% bols) %*% (y - x %*% bols)
s1 <- so1 + s0 + sb
v1 <- v0 + n
g <- 0
for(j in 1:k) {
gg <- lgamma((v1 + 1 - j)/2)
g <- c(g, gg)
}
g <- sum(g)
prob <<- - (n * k) * log(2 * pi) - v1/2 * log(det(s1)) - k/2 * log(det(m1)) + g
tmpfil <- tempfile(”parm”)
on.exit(unlink(tmpfil))
prob
}
```

3. Lag order selection

```
## function to compute BF and Information criteria for lag-order selection (SETAR)
> lagsel.unv01
function(x, y, bf)
{
p <- ncol(x)
n <- length(y)
bf <- 0
xx <- x
for(j in 1:p) {
if(j == 1) {
x <- xx[, 1]
x <- matrix(x, ncol = 1)
## ols estimates
```

```
bhat <- solve(t(x) %*% x) %*% t(x) %*% y
ssols <- (t(y - x %*% bhat) %*% (y - x %*% bhat))
## specification of hyper-parameters
k1 <- 0.05
m0 <- k1
b0 <- 1
s0 <- 3
v0 <- j
m1 <- m0 + t(x) %*% x
bpost <- ((m0 * b0) + (t(x) %*% x) * bhat)/m1
ssb <- ((b0 - bhat) * 1)/(1/m0 + 1/(t(x) %*% x)) * (b0 - bhat)
v1 <- nrow(x) + v0
sspost <- (ssols + ssb + s0)
sigmasq <- sspost/(v1 - 1)
sigmasq <- c(sigmasq)
lik <- - n/2 * log(2 * pi) - n/2 * log(sigmasq) - (0.5 * (t(y - x %*% bpost) %*% (y - x %*% bpost)))/sigmasq
m00 <- sigmasq/(m0)
prbeta <- -1/2 * log(m00) - j/2 * log(2 * pi) - 1/2 * (((bpost - b0) * (bpost - b0))/m00)
prsigma <- - (v0/2 + 1) * log(sigmasq) - (0.5 * s0)/sigmasq
postprob <- lik + prbeta + prsigma
hessian1 <- sigmasq/m1
hessian2 <- sspost^2/((v1 - 1)^2 * (v1 - 2))
hessian <- 1/2 * (log(hessian1) + log(hessian2))
}
else {
x <- xx[, 1:j]
## ols estimates
bhat <- solve(t(x) %*% x) %*% t(x) %*% y
ssols <- (t(y - x %*% bhat) %*% (y - x %*% bhat))
## specification of hyper-parameters
k1 <- 0.05
m0 <- (diag(k1/seq(1:j)))
b0 <- (c(1, rep(0, (j - 1))))
s0 <- 3
v0 <- j
## estimation of posterior parameter
m1 <- m0 + t(x) %*% x
bpost <- solve(m1) %*% ((m0 %*% b0) + (t(x) %*% x) %*% bhat)
ssb <- t(b0 - bhat) %*% solve((solve(m0) + solve(t(x) %*% x))) %*% (b0 - bhat)
v1 <- nrow(x) + v0
sspost <- (ssols + ssb + s0)
sigmasq <- sspost/(v1 - 1)
sigmasq <- c(sigmasq)
lik <- - n/2 * log(2 * pi) - n/2 * log(sigmasq) - (0.5 * (t(y - x %*% bpost) %*% (y - x %*% bpost)))/sigmasq
m00 <- sigmasq * solve(m0)
prbeta <- -1/2 * log(det(m00)) - j/2 * log(2 * pi) - 1/2 * (t(bpost - b0) %*% solve(m00) %*% (bpost - b0))
prsigma <- - (v0/2 + 1) * log(sigmasq) - (0.5 * s0)/sigmasq
postprob <- lik + prbeta + prsigma
hessian1 <- sigmasq * solve(m1)
hessian2 <- sspost^2/((v1 - 1)^2 * (v1 - 2))
hessian <- 1/2 * (log(det(hessian1)) + log((hessian2)))
}
bbf <- postprob + hessian + j/2 * log(2 * pi)
prior <- prbeta + prsigma
s <- ncol(hessian1) + ncol(hessian2)
ifim <- s/2 * log((sum(diag(hessian1)) + sum(diag(hessian2)))/s)
aic <- -2 * lik + 2 * (j)
aic1 <- -2 * postprob + 2 * (j)
bic <- -2 * lik + log(n) * (j)
bic1 <- -2 * postprob + log(n) * (j)
```

```
icomp <- -2 * lik + 2 * (ifim - hessian) + j
icomp1 <- -2 * postprob + 2 * (ifim - hessian) + j
caicf <- -2 * lik + log(n) * j + 2 * hessian + (2 * n * j)/(n - j - 2)
bms <- bic1 + 2 * hessian + (2 * n * j)/(n - j - 2)
bff <- c(bbf, aic, aic1, bic, bic1, icomp, icomp1, caicf, bms)
bf <- c(bf, bff)
}
tmpfil <- tempfile("lagsel.unv01")
on.exit(unlink(tmpfil))
bf <- bf[-1]
}
>
## function to compute BF and Information criteria for lag-order selection (Threshold VAR)
> lagsel.m0
function(x, y, bf)
{
p <- ncol(x)/2
n <- nrow(x)
k <- ncol(y)
xx <- x
bf <- 0
for(j in (1:p)) {
x <- xx[, 1:(k * j)]
#ols estimates
bhat <- solve(t(x) %*% x) %*% t(x) %*% y
ssols <- (t(y - x %*% bhat) %*% (y - x %*% bhat))
#specification of hyper-parameters
hyperm0(x, y)
m1 <- m0 + t(x) %*% x
bpost <- solve(m1) %*% ((m0 %*% b0) + (t(x) %*% x) %*% bhat)
ssb <- t(b0 - bhat) %*% solve((solve(m0) + solve(t(x) %*% x))) %*% (b0 - bhat)
v1 <- n + v0
sspost <- (ssols + ssb + s0)
etha <- k * j
sigma <- sspost/(v1 - k)
lik <- - (n * k) * log(2 * pi) - n/2 * log(det(sigma)) - 0.5 * sum(diag(solve(sigma) %*% t(y - x %*% bpost) %*%
(y - x %*% bpost)))
sigprior <- - (v0 + k + 1)/2 * log(det(sigma)) + v0/2 * log(det(s0)) - 0.5 * sum(diag(solve(sigma) %*% s0))
betaprior <- - (etha * k) * log(2 * pi) - etha/2 * log(det(sigma)) - 0.5 * sum(diag(solve(sigma) %*% t(bpost - b0)
%*% m0 %*% (bpost - b0)))
prior <- sigprior + betaprior
postprob <- lik + sigprior + betaprior
hess1 <- kronecker(sigma, solve(m1))
hessian1 <- log((det(hess1)))
hess2 <- dd %*% kronecker(sigma, sigma) %*% t(dd)
hessian2 <- log(det(hess2))
hessian <- 1/2 * ((hessian1 + hessian2))
s <- ncol(hess1) + ncol(hess2)
ifim <- s/2 * log((sum(diag(hess1)) + sum(diag(hess2)))/s)
nparm <- (k * etha + (k * (k + 1))/2)
bbf <- postprob + hessian + j/2 * log(2 * pi)
aic <- -2 * lik + 2 * nparm
aic1 <- -2 * postprob + 2 * nparm
bic <- -2 * lik + log(n) * nparm
bic1 <- -2 * postprob + log(n) * nparm
icomp <- - lik * 2 + 2 * (ifim - hessian) + nparm
icomp1 <- - postprob * 2 + 2 * (ifim - hessian) + nparm
caicf <- -2 * lik + log(n) * nparm + 2 * hessian + (2 * n * nparm)/(n - nparm - 2)
bms <- bic1 + 2 * hessian + (2 * n * nparm)/(n - nparm - 2)
```

```
bff <- c(bbf, aic, aic1, bic, bic1, icomp, icomp1, caicf, bms)
bf <- c(bf, bff)
}
tmpfil <- tempfile(”lagsel.m0”)
on.exit(unlink(tmpfil))
bf <- bf[-1]
}
```

4. Forecasting

4.1 SETAR
One-step ahead forecasting

```
# Annual sunspot data
sun<-data.frame(sun)
yy<-sun
xx<-cbind(1,tsmatrix(yy, lag(yy,-1), lag(yy,-2), lag(yy,-3)))
n<-nrow(xx)
p<-ncol(xx)
x<-xx[1:220,]
xnew<-xx[(221-3):n,]
yt<-xx[218:220,2]
p1<-2
p2<-3
all<-rep(0,6)
for (j in 1:57)
{
ypred<-0
for (it in 1:100)
{
u<-sim[round(runif(1,0.5,500.5)),]
x<-x[order(x[,(u[1]+2)]),]
data<-pardata.unv2(x[,2:p],u[1],u[2],p1,p2)
x1<-data$x1
x2<-data$x2
y1<-data$y1
y2<-data$y2
if (yt[j]<=u[2])
{param<-par.unv(x1,y1)
sigma <- rgamma(1, param$sspost/param$v1)
beta <- mnorm(1, param$beta, (sigma * solve(param$m1)))
yhat<-beta%*%xnew[j,c(1,3:(p1+2))]+sqrt(sigma)*rnorm(1)}
else
{param<-par.unv(x2,y2)
sigma <- rgamma(1, param$sspost/param$v1)
beta <- mnorm(1, param$beta, (sigma * solve(param$m1)))
yhat<-beta%*%xnew[j,c(1,3:(p2+2))]+sqrt(sigma)*rnorm(1)}
ypred<-c(ypred,yhat)
}
yt<-c(yt,yhat)
ypred<-ypred[-1]
ypred<-sort(ypred)
yf20<-ypred[21]
yf80<-ypred[79]
yf05<-ypred[6]
yf95<-ypred[94]
yf<-mean(ypred)
```

```
yyf<-cbind(xnew[j,2],yf,yf05,yf20,yf80,yf95)
all<-rbind(all,yyf)
}
all<-all[-1,]
```

h-step ahead forecasting

```
# Annual sunspot data
sun<-data.frame(sun)
yy<-sun
xx<-cbind(1,tsmatrix(yy, lag(yy,-1), lag(yy,-2), lag(yy,-3)))
n<-nrow(xx)
p<-ncol(xx)
#x<-xx[1:220,]
#yt<-xx[221:n,2]
#xnew<-xx[218:220,2]
x<-xx[17:169,]
yt<-xx[170:234,2]
xnew<-xx[167:169,2]
p1<-2
p2<-3
all<-rep(0,6)
for (j in 1:35)
{
ypred<-0
for (it in 1:100)
{
u<-sim[round(runif(1,0.5,500.5)),]
x<-x[order(x[,(u[1]+2)]),]
data<-pardata.unv2(x[,2:p],u[1],u[2],p1,p2)
x1<-data$x1
x2<-data$x2
y1<-data$y1
y2<-data$y2
if (xnew[j]<=u[2])
{param<-par.unv(x1,y1)
sigma <- rgamma(1, param$sspost/param$v1)
beta <- mnorm(1, param$beta, (sigma * solve(param$m1)))
yhat<-c(1,xnew[(j+2):(j+3-p1)])%*%beta+sqrt(sigma)*rnorm(1)}
else
{param<-par.unv(x2,y2)
sigma <- rgamma(1, param$sspost/param$v1)
beta <- mnorm(1, param$beta, (sigma * solve(param$m1)))
yhat<-c(1,xnew[(j+2):(j+3-p2)])%*%beta+sqrt(sigma)*rnorm(1)}
ypred<-c(ypred,yhat)
}
ypred<-ypred[-1]
ypred<-sort(ypred)
yf20<-ypred[21]
yf80<-ypred[79]
yf05<-ypred[6]
yf95<-ypred[94]
yf<-mean(ypred)
xnew<-c(xnew,yf)
yyf<-cbind(yt[j],yf,yf05,yf20,yf80,yf95)
all<-rbind(all,yyf)
}
all<-all[-1,]
```

4.2 US interest rates

One-step ahead forecasting

```
dimnames(tb1)[[1]]<-timeSeq("04/01/1953","09/01/2002",by="months")
y<-data.frame(tb1)
y<-log(y)
yy<-y[,1]-y[,2]
yy1<-yy[1]
yy2<-(yy[2]+yy[1])/2
yyyy<-tsmatrix(yy,lag(yy,-1),lag(yy,-2))
yyy<-c(yy1,yy2,apply(yyyy,1,mean))
y<-diff(y,lag=1)
k<-ncol(y)
yy<-tsmatrix(yyy,lag(yyy,-1),lag(yyy,-2),lag(yyy,-3),lag(yyy,-4))
### Create case ordered data ###
x<-tsmatrix(y, lag1=lag(y,-1), lag2=lag(y,-2), lag3=lag(y,-3))
nyy<-nrow(yy)
x<-cbind(yy[,2],x)
xx<-x
n<-nrow(xx)
p<-ncol(xx)
x<-xx[1:566,]
xnew<-xx[567:n,2:ncol(xx)]
yt<-xx[567:n,1]
p1<-2
p2<-3
p3<-2
all<-rep(0,8)
for (j in 1:24)
{
ypred<-rep(0,2) for (it in 1:100)
{
u<-sim[round(runif(1,0.5,500.5)),c(1,3:4)]
x<-x[order(x[,u[1]]),]
data<-parmdata3(x,2,1,u[2:3],p1,p2,p3)
x1<-data$x1
x2<-data$x2
x3<-data$x3
y1<-data$y1
y2<-data$y2
y3<-data$y3
if (yt[j]<=u[2])
{param<-par.m(x1,y1)
sigma <- wishart(param$v1, zero, param$s1)
beta <- mnorm(1, param$beta, kronecker(sigma, solve(param$m1)))
beta<-matrix(beta,ncol=2,byrow=F)
v<-eigen(sigma)
vsigma<- v$vectors %*% diag(sqrt(v$values)) %*% t(v$vectors)
yhat<-c(1,xnew[(j),(3:(2*p1+2))])%*%beta+t(vsigma%*%mnorm(1,c(0,0),diag(2)))
}
if (yt[j]>u[2] & yt[j]<=u[3])
{param<-par.m(x2,y2)
sigma <- wishart(param$v1, zero, param$s1)
beta <- mnorm(1, param$beta, kronecker(sigma, solve(param$m1)))
```

```
beta<-matrix(beta,ncol=2,byrow=F)
v<-eigen(sigma)
vsigma<- v$vectors %*% diag(sqrt(v$values)) %*% t(v$vectors)
yhat<-c(1,xnew[(j),(3:(2*p2+2))])%*%beta+t(vsigma%*%mnorm(1,c(0,0),diag(2)))
}
if(yt[j]>u[3])
{param<-par.m(x3,y3)
sigma <- wishart(param$v1, zero, param$s1)
beta <- mnorm(1, param$beta, kronecker(sigma, solve(param$m1)))
beta<-matrix(beta,ncol=2,byrow=F)
v<-eigen(sigma)
vsigma<- v$vectors %*% diag(sqrt(v$values)) %*% t(v$vectors)
yhat<-c(1,xnew[(j),(3:(2*p3+2))])%*%beta+t(vsigma%*%mnorm(1,c(0,0),diag(2)))
}
ypred<-rbind(ypred,yhat)
}
ypred<-ypred[-1,]
yf<-apply(ypred,2,mean)
y1f90<-ypred[order(ypred[,1]),1][89]
y1f10<-ypred[order(ypred[,1]),1][11]
y2f90<-ypred[order(ypred[,2]),2][89]
y2f10<-ypred[order(ypred[,2]),2][11]
allm<-c(xnew[(j),1:2],yf,y1f10,y1f90,y2f10,y2f90)
all<-rbind(all,allm)
}
all<-all[-1,]
```

h-step ahead forecasting

```
dimnames(tb1)[[1]]<-timeSeq(”04/01/1953”,”09/01/2002”,by=”months”)
y<-data.frame(tb1)
y<-log(y)
yy<-y[,1]-y[,2]
yy1<-yy[1]
yy2<-(yy[2]+yy[1])/2
yyyy<-tsmatrix(yy,lag(yy,-1),lag(yy,-2))
yyy<-c(yy1,yy2,apply(yyyy,1,mean))
y<-diff(y,lag=1)
k<-ncol(y)
yy<-tsmatrix(yyy,lag(yyy,-1),lag(yyy,-2),lag(yyy,-3),lag(yyy,-4),lag(yyy,-5),lag(yyy,-6)
,lag(yyy,-7),lag(yyy,-8),lag(yyy,-9),lag(yyy,-10),lag(yyy,-11),lag(yyy,-12),lag(yyy,-13))
# Create case ordered data#
#x<-tsmatrix(y, lag1=lag(y,-1), lag2=lag(y,-2), lag3=lag(y,-3))
x<-tsmatrix(y, lag1=lag(y,-1), lag2=lag(y,-2), lag3=lag(y,-3), lag4=lag(y,-4), lag5=lag(y,-5),
lag6=lag(y,-6), lag7=lag(y,-7), lag8=lag(y,-8), lag9=lag(y,-9),lag10=lag(y,-10)
,lag11=lag(y,-11), lag12=lag(y,-12))
nyy<-nrow(yy)
x<-cbind(yy[,2],x)
xx<-x
n<-nrow(xx)
p<-ncol(xx)
x<-xx[1:566,]
xnew<-xx[554:566,2:3]
xp<-xx[567:n,2:3]
yt<-xx[567,1]
ydiff1<-xnew[1,1]-xnew[1,2]
ydiff2<-xnew[2,1]-xnew[2,2]
```

```
ydiff<-c(ydiff1,ydiff2)
p1<-2
p2<-12
p3<-2
all<-rep(0,8)
for (j in 1:24)
{
ypred<-rep(0,2) for (it in 1:100)
{
u<-sim[round(runif(1,0.5,500.5)),c(1,3:4)]
x<-x[order(x[,u[1]]),]
data<-parmdata3(x,2,1,u[2:3],p1,p2,p3)
x1<-data$x1
x2<-data$x2
x3<-data$x3
y1<-data$y1
y2<-data$y2
y3<-data$y3
xt<-c(t(xnew[(j+11):j,]))
if (yt[j]<=u[2])
{param<-par.m(x1,y1)
sigma <- wishart(param$v1, zero, param$s1)
beta <- mnorm(1, param$beta, kronecker(sigma, solve(param$m1)))
beta<-matrix(beta,ncol=2,byrow=F)
v<-eigen(sigma)
vsigma<- v$vectors %*% diag(sqrt(v$values)) %*% t(v$vectors)
yhat<-c(1,xt[1:(2*p1)])%*%beta+t(vsigma%*%mnorm(1,c(0,0),diag(2)))
}
if (yt[j]>u[2] & yt[j]<=u[3])
{param<-par.m(x2,y2)
sigma <- wishart(param$v1, zero, param$s1)
beta <- mnorm(1, param$beta, kronecker(sigma, solve(param$m1)))
beta<-matrix(beta,ncol=2,byrow=F)
v<-eigen(sigma)
vsigma<- v$vectors %*% diag(sqrt(v$values)) %*% t(v$vectors)
yhat<-c(1,xt[1:(2*p2)])%*%beta+t(vsigma%*%mnorm(1,c(0,0),diag(2)))
}
if(yt[j]>u[3])
{param<-par.m(x3,y3)
sigma <- wishart(param$v1, zero, param$s1)
beta <- mnorm(1, param$beta, kronecker(sigma, solve(param$m1)))
beta<-matrix(beta,ncol=2,byrow=F)
v<-eigen(sigma)
vsigma<- v$vectors %*% diag(sqrt(v$values)) %*% t(v$vectors)
yhat<-c(1,xt[1:(2*p3)])%*%beta+t(vsigma%*%mnorm(1,c(0,0),diag(2)))
}
ypred<-rbind(ypred,yhat)
}
ypred<-ypred[-1,]
yf<-apply(ypred,2,mean)
xnew<-rbind(xnew,yf)
y1f90<-ypred[order(ypred[,1]),1][89]
y1f10<-ypred[order(ypred[,1]),1][11]
y2f90<-ypred[order(ypred[,2]),2][89]
y2f10<-ypred[order(ypred[,2]),2][11]
allm<-c(xp[j,1:2],yf,y1f10,y1f90,y2f10,y2f90)
all<-rbind(all,allm)
ydiff<-c(ydiff,(yf[1]-yf[2]))
yd<-(ydiff[j]+ydiff[(j+1)]+ydiff[(j+2)])/3
```

```
yt<-c(yt,yd)
}
all<-all[-1,]
```

5. US interest rates data

```
y<-tb1
y<-log(y)
yy<-y[,1]-y[,2]
yy1<-yy[1]
yy2<-(yy[2]+yy[1])/2
yyyy<-tsmatrix(yy,lag(yy,-1),lag(yy,-2))
yyy<-c(yy1,yy2,apply(yyyy,1,mean))
y<-diff(y,lag=1)
k<-ncol(y)
y<-ts(y,start=(1953+4/12),end=(2002+9/12),frequency=12)
yy<-tsmatrix(lag(yyy,-1),lag(yyy,-2),lag(yyy,-3),lag(yyy,-4))
### Create case ordered data ###
x<-tsmatrix(y, lag1=lag(y,-1), lag2=lag(y,-2), lag3=lag(y,-3), lag4=lag(y,-4), lag5=lag(y,-5),
lag6=lag(y,-6), lag7=lag(y,-7), lag8=lag(y,-8), lag9=lag(y,-9),lag10=lag(y,-10)
,lag11=lag(y,-11), lag12=lag(y,-12),lag13=lag(y,-13), lag14=lag(y,-14),lag15=lag(y,-15))
nyy<-nrow(yy)
n<-nrow(x)
x<-cbind(yy[13:nyy,],x)
```

6. Computation of Bayes factor (Gelfand and Dey method)

```
{
dimnames(tb1)[[1]]<-timeSeq(”04/01/1953”,”09/01/2002”,by=”months”)
y<-data.frame(tb1)
y<-log(y)
yy<-y[,1]-y[,2]
##yy<-tsmatrix(yy,lag(yy,-1),lag(yy,-2),lag(yy,-3),lag(yy,-4))
yy1<-yy[1]
yy2<-(yy[2]+yy[1])/2
yyyy<-tsmatrix(yy,lag(yy,-1),lag(yy,-2))
yyy<-c(yy1,yy2,apply(yyyy,1,mean))
y<-diff(y,lag=1)
k<-ncol(y)
y<-ts(y,start=(1953+4/12),end=(2002+9/12),frequency=12)
yy<-tsmatrix(lag(yyy,-1),lag(yyy,-2),lag(yyy,-3),lag(yyy,-4))
### Create case ordered data ###
x<-tsmatrix(y, lag1=lag(y,-1), lag2=lag(y,-2), lag3=lag(y,-3), lag4=lag(y,-4), lag5=lag(y,-5),
lag6=lag(y,-6), lag7=lag(y,-7), lag8=lag(y,-8), lag9=lag(y,-9),lag10=lag(y,-10)
,lag11=lag(y,-11), lag12=lag(y,-12),lag13=lag(y,-13), lag14=lag(y,-14),lag15=lag(y,-15))
nyy<-nrow(yy)
x<-cbind(yy[12:(nyy-1),],x)
## define lag-orders for each regime ##
p1<-2
p2<-3
## Define threshold values ##
nn<-nrow(sim)
sb1<-0
sb2<-0
```

```
postp1<-0
postp2<-0
for (it in 1:nn)
parmdata2p(x,sim[it,1],sim[it,2],p1,p2)
bf.m2(x1,y1)
sb1<-c(sb1,b1)
postp1<-c(postp1,postprob)
bf.m2(x2,y2)
sb2<-c(sb2,b1)
postp2<-c(postp2,postprob)
}
postp1<-postp1[-1]
postp2<-postp2[-1]
sb1<-sb1[-1]
sb2<-sb2[-1]
sb1 <- matrix(sb1, ncol = (k*(k * p1+1) + k), byrow = T)
sb2 <- matrix(sb2, ncol = (k*(k * p2+1) + k), byrow = T)
mu1 <- apply(sb1, 2, mean)
mu2 <- apply(sb2, 2, mean)
vmat1 <- var(sb1)
vmat2 <- var(sb2)
dimb1 <- nrow(vmat1)
qq1 <- 0
qq2<-0
for(it in 1:nrow(sb1)) {
bnorm1 <- -1/2 * (t(mu1 - sb1[it, ]) %*% solve(vmat1) %*% (mu1 - sb1[it, ]))
bnorm2 <- -1/2 * (t(mu2 - sb2[it, ]) %*% solve(vmat2) %*% (mu2 - sb2[it, ]))
qq1 <- c(qq1, bnorm1)
qq2 <- c(qq2, bnorm2)
}
qq1 <- qq1[-1]
qq2<-qq2[-1]
postprob<-postp1+postp2
qq<-qq1+qq2
bff <- -qq+postprob
bf<-mean(bff)

## function to compute Bayes factor (Gelfand and Dey method)
> bf.m1
function(x, sn, k, p)
{
y <- x[, 1:k]
x <- cbind(1, x[, ((k + 1):(2 * p))])
n <- nrow(x)
postprob <- 0
b1 <- 0
hyperm1(x, y)
m1 <- m0 + t(x) %*% x
bpost <- solve(m1) %*% ((m0 %*% b0) + (t(x) %*% x) %*%
bhat)
ssb <- t(b0 - bhat) %*% solve((solve(m0) + solve(t(x) %*% x))) %*% (b0 - bhat)
v1 <- nrow(x) + v0
beta <- c(bpost)
sspost <- (ssols + ssb + s0)
sigma <- sspost/(v1 - k - 1)
lik <- - (n * k) * log(2 * pi) - n/2 * log(det(sigma)) - 0.5 * sum(diag(solve(sigma) %*% t(y - x %*%bpost) %*% (y
- x %*% bpost)))
sigprior <- - (v0 + k + 1)/2 * log(det(sigma)) + v0/2 * log(det(s0)) - 0.5 * sum(diag(solve(sigma)%*%s0))
betaprior <- - (etha * k) * log(2 * pi) - etha/2 *
```

```
log(det(sigma)) + k/2 * log(det(m0)) - 0.5 *
sum(diag(solve(sigma) %*% t(bpost - b0) %*% m0
%*%
(bpost - b0)))
postprob <- lik + sigprior + betaprior
for(i in 1:sn) {
ssigma <- wishart(v1, zero, sspost)
sbeta <- mnorm(1, beta, kronecker(ssigma,solve(m1)))
b1 <- c(b1, sbeta, diag(ssigma))
}
b1 <- b1[-1]
b1 <- matrix(b1, ncol = (2 * p + k), byrow = T)
mu <- apply(b1, 2, mean)
vmat1 <- var(b1)
dimb1 <- nrow(vmat1)
qq <- 0
for(it in 1:nrow(b1)) {
bnorm <- -1/2 * (t(mu - b1[it, ]) %*% solve(vmat1) %*% (mu - b1[it, ]))
qq <- c(qq, bnorm)
}
qq <- qq[-1]
bff <- - qq + postprob
bf <- mean(bff)
tmpfil <- tempfile("bf.m1")
on.exit(unlink(tmpfil))
list(bf = bf, b1 = b1)
}
>
## function to compute Bayes factor (Gelfand and Dey method)
> bf.m2
function(x, y)
{
n <- nrow(x)
hyperm1(x, y)
m1 <- m0 + t(x) %*% x
bpost <- solve(m1) %*% ((m0 %*% b0) + (t(x) %*% x) %*%bhat)
ssb <- t(b0 - bhat) %*% solve((solve(m0) + solve(t(x) %*% x))) %*% (b0 - bhat)
v1 <- nrow(x) + v0
beta <- c(bpost)
sspost <- (ssols + ssb + s0)
sigma <- sspost/(v1 - k - 1)
lik <- - (n * k) * log(2 * pi) - n/2 * log(det(sigma)) - 0.5 * sum(diag(solve(sigma) %*% t(y - x %*%bpost) %*% (y
- x %*% bpost)))
sigprior <- - (v0 + k + 1)/2 * log(det(sigma)) + v0/2 * log(det(s0)) - 0.5 * sum(diag(solve(sigma)%*%s0))
betaprior <- - (etha * k) * log(2 * pi) - etha/2 *log(det(sigma)) + k/2 * log(det(m0)) - 0.5 *sum(diag(solve(sigma)
%*% t(bpost - b0) %*% m0%*%(bpost - b0)))
postprob <<- lik + sigprior + betaprior
ssigma <- wishart(v1, zero, sspost)
sbeta <- mnorm(1, beta, kronecker(ssigma, solve(m1)))
b1 <<- c(sbeta, diag(ssigma))
}
## function to draw samples from inverse wishart dendity
> wishart
function(nu, zero, s0, tol = 1e-007)
{
p <- ncol(s0)
if(max(abs(s0 - t(s0))) > tol)
print("vmat is not symmetric")
zero <- rep(0, p)
simul <- mnorm(nu, zero, s0)
```

```
wish <- t(simul) %*% (simul)
wish <- solve(wish)
}

## function to draw samples from multivariate normal density
> mnorm
function(m, mu, vmat, tol = 1e-007)
{
p <- ncol(vmat)
if(length(mu) != p)
print(”mu vector is the wrong length”)
if(max(abs(vmat - t(vmat))) > tol)
print(”vmat is not symmetric”)
vs <- eigen(vmat)
vsqrt <- vs$vectors %*% diag(sqrt(vs$values)) %*% t(vs$vectors)
simul <- matrix(rnorm(m * p), nrow = m) %*% vsqrt
simul <- sweep(simul, 2, mu, ”+”)
drop(simul)
}
```

VITA

Yongjae Kwon was born in South Korea in 1966. He received his Bachelor of Business administration from Seoul National University in 1988 and his Master of Business Administration (concentration on Finance) from St. Louis University in 1997. Mr. Kwon expects to receive the Doctor of Philosophy in August 2003. From 1990 to 1999, he worked at the Ministry of Science and Technology as a assistant director and a deputy director at divisions of research and development management, international cooperation, nuclear cooperation, and R&D planning. He then pursued Doctor of Philosophy in Business Administration with concentration on Statistics from 1999.