Statistical Inference in Markov Switching Vector Error Correction Model Using a Markov Chain Monte Carlo Method

Statistical Inference in Markov Switching Vector Error Correction Model Using a Markov Chain Monte Carlo Method

Katsuhiro Sugita*

Faculty of Law and Letters, University of the Ryukyus, Nishihara,

Senbaru, Okinawa, 903-0213, Japan

E-Mail: ksugita@ll.u-ryukyu.ac.jp

Abstract

This paper introduces statistical inference in a Markov switching vector error correction model using a Markov chain Monte Carlo method. The proposed model allows for regime shifts in the deterministic terms, the lag terms, the adjustment terms and the variance-covariance matrix. The proposed method allows for estimation of the cointegrating vector within a nonlinear framework through a collapsed Gibbs sampling. We apply the proposed model to U.S. term structure of interest rates.

Key words: Bayesian inference; Nonlinear cointegration; Markov switching model; Gibbs sampling; Bayes factor;

*The author would like to thank Mike Clements and the participants of seminar at Warwick, Hitotsubashi, and Kobe University for their useful comments.

1 Introduction

This paper proposes a Markov switching vector error correction model (MS-VECM) that allows for regime shifts in the deterministic terms, the lag terms, the adjustment terms and the variance-covariance matrix in a vector error correction model, using a Bayesian approach with a Markov chain Monte Carlo method.

A number of studies consider nonlinear cointegration models with regime switching. Balke and Fomby (1997) consider a threshold cointegration model to investigate the model in which there is discontinuous adjustment to a long-run equilibrium, based on the idea that only when the deviation from the equilibrium exceeds a critical threshold, do the benefits of adjustment exceed the costs and, hence economic agents act to move the system back toward the equilibrium. Anderson (1997), Tsay (1998), Martens et al (1998), and Clements and Galvao (2002) are examples of applying threshold cointegration model. Krolzig (1997) develops a regime switching cointegration model using Hamilton's (1989) Markov regime switching process instead of threshold cointegration model. Hall et al (1997) analyze the permanent income hypothesis using a single equation cointegration model with Markov regime switching. Psaradakis et al (2004) employ Markov switching to analyze an error correction model in a single equation. A vector error correction model with Markov regime switching is applied by Sarno and Valente (2005) for forecasting stock returns, and by Clarida et al (2006), who show regime switching in the term structure of interest rates.

Estimation for the MS-VECM by classical methods requires a multi-stage maximum likelihood procedure. The first stage consists of testing for the number of cointegrating relationships in the system and estimating the cointegrating vectors by implementing Johansen's (1988, 1991) maximum likelihood method. The second stage consists of estimating other parameters in the model by maximum likelihood method. Thus, the cointegrating vectors and other parameters in a nonlinear vector error correction model are estimated assuming the model is linear. The final stage consists of the implementation of an expectation-maximization (EM) algorithm for maximum likelihood estimation for unobserved Markov state variables conditional on estimated values of the cointegrating vectors and other parameters by maximum likelihood. Thus, to estimate the Markov state variables, the maximum likelihood estimates are treated as if they were the true values.

By applying a Bayesian approach, estimation of the MS-VECM is more efficient as inference on the state variable is based on a joint distribution, rather than a conditional distribution. The cointegrating vectors are estimated based on a joint distribution of other variables including the state variables, so that it allows estimation of the cointegrating vectors within a nonlinear framework, rather than assuming that the model is linear.

This paper proposes a Bayesian approach to the MS-VECM that allows any set of the parameters in the model to shift with Markov process. For a Bayesian approach to the MS-VECM, Paap and van Dijk (2003) propose a nonlinear VECM where only intercept terms are affected by Markov regime shift to investigate U.S. consumption and income. They employ a Bayesian approach based on Kleibergen and Paap (2002), which requires linear normalizing restrictions on the cointegrating vectors. This linear normalizing restrictions is criticized by Strachan (2003) as being likely to be invalid. Strachan and van Dijk (2003) and Strachan and Inder (2004) discuss the further problems associated with the use of linear normalizing restrictions, and propose the Grassman approach that places a valid prior on the cointegrating space. See Koop, Strachan and van Dijk (2006) for details. In this paper, we apply the method by Koop, Leon-Gonzalez and Strachan (forthcoming) who develop further prior elicitation for the cointegrating space. They propose an efficient posterior simulation algorithm for cointegrated models using a collapsed Gibbs sampler. Examples of application of this method include Koop, Leon-Gonzalez, and Strachan (2006) for a cointegrating panel data model and Strachan and van Dijk (2007) for model averaging in vector autoregressive models.

Our model in this paper is more general than Paap and van Dijk (2003), and is flexible to modify to consider the model in which other parameters are also subject to the regime shift. For example, in this paper we assume that the cointegrating vectors are unaffected by the regime shifts. It is, however, possible to consider the model where the cointegrating vectors are also dependent on the regime shifts. Also, it is possible to consider the trend or the drift in the cointegration relations to be affected by the regime shifts by a slight modification.

The plan of the paper is as follows. Section 2 presents estimation method for the MS-VECM using a collapsed Gibbs sampler. We specify prior densities and likelihood functions, and then derive the posterior distributions. Section 3 illustrates application to U.S. term structure of interest rates using the MS-VECM. Section 4 contains concluding remarks. All results reported in this paper are generated using Ox version 5.10 (see Doornik, 2007).

2 Markov Switching Vector Error Correction Model

This section introduces the MS-VECM and presents a Bayesian approach to estimate this model. Let y_t denote an $I(1)$ vector of $1 \times n$ with r linear cointegrating relations. A VAR system with normally distributed Gaussian innovations $\varepsilon_t \sim iidN(0, \Omega)$ can be written as a vector error correction model (VECM) with the number of lags p

$$\Delta y_t = y_{t-1}\beta\alpha + \delta\theta + \sum_{l=1}^{p} \Delta y_{t-l}\Gamma_l + \varepsilon_t \tag{1}$$

where α $(r \times n)$ is adjustment term; β $(n \times r)$ is cointegrating vector; Γ_l $(n \times n)$ is lag term. The deterministic terms $\delta\theta$ (δ is $1 \times d$, and θ is $d \times n$) are defined as follows. For example, if the process contains both a constant and linear time trend, then $\theta = (\mu', \gamma')'$ and $\delta = (1,t)$, or if the process contains a constant but no time trend, then $\theta = \mu$ and $\delta = 1$. If we assume that the deterministic term θ, the adjustment term α, the lag terms Γ_l, and the variance-covariance matrix Ω in the VECM are subject to an unobservable discrete state variable s_t that evolves according to a m-state, first-order Markov switching process with the transition probabilities, $p(s_t = i \mid s_{t-1} = j) = q_{ij}$, $i, j = 1, \ldots, m$, then the VECM representation is written as

$$\Delta y_t = y_{t-1}\beta\alpha(s_t) + \delta\theta(s_t) + \sum_{l=1}^{p} \Delta y_{t-l}\Gamma_l(s_t) + \varepsilon_t(s_t) \tag{2}$$

where $\varepsilon_t(s_t) \sim N(0, \Omega(s_t))$.

2.1 Likelihood

The MSVECM in (2) can be rewritten as

$$\Delta y_t = z_{1,t}\beta\alpha(s_t) + z_{2,t}\Phi + \varepsilon_t(s_t) \tag{3}$$

where $z_{2,t} = (\iota_t(1)\delta, \ldots, \iota_t(m)\delta, \iota_t(1)\Delta y_{t-1}, \ldots, \iota_t(1)\Delta y_{t-p}, \ldots, \iota_t(m)\Delta y_{t-1}, \ldots, \iota_t(m)\Delta y_{t-p})$, $\Phi = \left(\theta'(1), \ldots, \theta'(m), \Gamma_1'(1), \ldots, \Gamma_p'(1), \ldots, \Gamma_1'(m), \ldots, \Gamma_p'(m)\right)'$, $z_{1,t} = y_{t-1}$, and $\iota_t(i)$ in $z_{2,t}$ is an indicator variable that equals to 1 if regime is i at t, and 0 otherwise. From (3), let define the $T \times n$ matrices $Y = (\ \Delta y_1', \ \ldots, \ \Delta y_T'\ )'$ and $E = (\ \varepsilon_1'(s_1), \ \ldots, \ \varepsilon_T'(s_T)\ )'$, the $T \times n$ matrix $Z_i = (\ z_{1,1}'\iota_0(i), \ z_{1,2}'\iota_1(i), \ \ldots, \ z_{1,T}'\iota_{T-1}(i)\ )'$, the $T \times m(d+np)$ matrix $X = (\ z_{2,1}', \ z_{2,2}', \ \ldots, \ z_{2,T}'\ )'$, the $T \times h$ (where $h = m(r+d+np)$) matrix $W = (\ Z_1\beta, \ldots, \ Z_m\beta, \ X\ )$, the $h \times n$ matrix $B = (\ \alpha'(1), \ldots, \ \alpha'(m), \ \Phi'\ )'$, then we can simplify the model as follows:

$$Y = \sum_{i=1}^{m} Z_i \beta \alpha_i + X\Phi + E \quad (4)$$

$$= WB + E \quad (5)$$

The likelihood function for $B, \Omega(1), \ldots, \Omega(m)$, β and the state variables $\widetilde{S}_T = \{s_1, s_2, \ldots, s_T\}'$ is given by,

$$\mathfrak{L}\left(B, \beta, \Omega(0), \ldots, \Omega(m), \widetilde{S}_T \mid Y\right)$$

$$\propto \left(\prod_{i=1}^{m} |\Omega(j)|^{-t_i/2}\right) \exp\left(-\frac{1}{2}\mathrm{tr}\left[\sum_{i=1}^{m} \left\{\Omega(i)^{-1} (Y_i - W_i B)' (Y_i - W_i B)\right\}\right]\right) \quad (6)$$

$$= \left(\prod_{i=1}^{m} |\Omega(j)|^{-t_i/2}\right) \exp\left(-\frac{1}{2}\sum_{i=1}^{m} \left[(vec(Y_i - W_i B))' (\Omega(j) \otimes I_{t_i})^{-1} (vec(Y_i - W_i B))\right]\right) \quad (7)$$

$$(8)$$

where $Y_i = \mathscr{I}_i Y$, $W_i = \mathscr{I}_i W$, $\mathscr{I}_i$ is a vector consists of 1 if j-th row's regime is i and 0 otherwise, and t_i is the total number of observations when $s_t = i$, $i = 1, \ldots, m$.

The likelihood function for the transition probabilities q_{ij}, $i, j = 1, \ldots, m$, which are independent of the data set and the model's other parameters but conditional on the set of the state variables, is given:

$$\mathfrak{L}\left(q_{00}, q_{11} \mid \widetilde{S}_T\right) = q_{00}^{m_{00}} (1 - q_{00})^{m_{01}} q_{11}^{m_{11}} (1 - q_{11})^{m_{10}} \quad (9)$$

where $m_{i,j}$, $i, j = 0, \ldots, m$, denotes the number of the transition from the regime i to j, that can be counted from given $\widetilde{S}_T$. This likelihood for the transition probabilities is used by Albert and Chib (1993) and Kim and Nelson (1998).

2.2 Priors

In selecting a prior density for cointegrating vectors, one approach is to choose an informative prior such as a normal or a Student t distribution with r^2 linear normalization restrictions on β for identification such that $\beta' = (I_r, \beta_\star')$ where $\beta_\star$ is $(n-r) \times r$ unrestricted matrix. Bauwens and Lubrano (1996) and Kleibergen and Paap (2002) choose this type of prior with linear normalization on β.

Recently, several authors have argued that it is important to elicit a prior on the space spanned by the cointegrating vectors rather than to a particular identified choice for these vectors (see Strachan (2003), Strachan and Inder (2004), Strachan and van Dijk (2004 and 2006), Villani (2005 and 2006), Koop, Leon-Gonzalez, and Strachan (forthcoming)). Strachan (2003) and Strachan and Inder (2004) criticize the linear normalization on the cointegrating vector restricts the estimable region of the cointegrating space. Koop et al (forthcoming) develop efficient posterior simulation algorithms using a collapsed Gibbs sampler to estimate cointegrating space. The approach we use in this paper is based on the collapsed Gibbs sampling method proposed by Koop et al (forthcoming).

They propose the following transformation.

$$\beta\alpha = (\beta\kappa)(\kappa^{-1}\alpha) = \left[\beta\left(\alpha\alpha'\right)^{1/2}\right]\left[\left(\alpha\alpha'\right)^{-1/2}\alpha\right] \equiv ba$$

where $\kappa \equiv (\alpha\alpha')^{1/2}$ is positive definite matrix and $a = \kappa^{-1}\alpha$ is semi-orthogonal. Then, we assign the multivariate normal distribution to the prior for b as

$$b \sim N(b_0, V_{b_0}). \tag{10}$$

For a prior for the transition probabilities q_{ij}, $i, j = 1, \ldots, m$, we assign a beta distribution

$$q_{ii} \sim beta(u_{ii}, u_{ij}) \tag{11}$$

$$q_{jj} \sim beta(u_{jj}, u_{ji}) \tag{12}$$

where *beta* refers to a beta distribution with density $\pi(p_{ii} \mid u_{ii}, u_{ij}) = \frac{\Gamma(u_{ii}+u_{ij})}{\Gamma(u_{ii})\Gamma(u_{ij})} p_{ii}^{u_{ii}-1}(1-p_{ii})^{u_{ij}-1}$.

With regard to priors for B, $\Omega(i)$ in (5), we assume prior independence between B and $\Omega(i)$ such that $p(B,\Omega(1),\dots,\Omega(m)) = p(B)\prod_{i=1}^{m} p(\Omega(i))$. We assign prior for the variance-covariance matrix as an inverted Wishart distribution with the degrees of freedom $\nu_0(i)$ as

$$\Omega(i) \sim IW(\Omega_0(i), \nu_0(i)) \tag{13}$$

where $\Omega_0(i) \in \mathbb{R}^{n\times n}$. As for a prior for B, we consider the vector form of B unconditional on $\Omega(i)$ because if we consider that the prior for B is conditional on $\Omega(i)$ as is often used in regression models with the natural conjugate priors, it is not convenient to consider a case when the error covariance is subject to change with regime. We assign prior for B as a multivariate normal as

$$vec(B) \sim MN(vec(B_0), \Sigma_{B_0}) \tag{14}$$

where *MN* refers to a multivariate normal with mean $vec(B_0) \in \mathbb{R}^{nh\times 1}$, and variance-covariance matrix $\Sigma_{B_0} \in \mathbb{R}^{nh\times nh}$.

2.3 Posterior Specifications

In this subsection we derive the posterior densities from the priors and the likelihood functions. First, we derive the state variable $\widetilde{S}_\tau = \{s_1, s_2, \dots, s_\tau\}'$ by the multi-move Gibbs sampler, then derive the posterior distributions for other parameters.

To sample the state variable $\widetilde{S}_\tau$ we employ the multi-move Gibbs sampling method, which is originally proposed by Carter and Kohn (1994) and is applied to a Markov switching model by Kim and Nelson (1998). The multi-move Gibbs sampling refers to simulating $s_t, t = 1,2,\ldots,T$, as a block from the following conditional distribution:

$$p\left(\widetilde{S}_\tau \mid \Theta, Y\right) = p(s_\tau \mid \Theta, Y) \prod_{t=p}^{\tau-1} p(s_\tau \mid s_{\tau+1}, \Theta, Y) \tag{15}$$

where $\Theta = \{B, \beta, \Omega(1), \ldots, \Omega(m), q_{11}, \ldots, q_{mm}\}$. The first term of the right hand side of the equation (15), $p(s_\tau \mid \Theta, Y)$, can be obtained from running the Hamilton filter (Hamilton, 1989). To draw s_t conditional on s_{t+1}, Θ and Y, we use the following results:

$$p(s_t \mid s_{t+1}, \Theta, Y) = \frac{p(s_{t+1} \mid s_t, \Theta, Y)\, p(s_t \mid \Theta, Y)}{p(s_{t+1} \mid \Theta, Y)} \propto p(s_{t+1} \mid s_t)\, p(s_t \mid \Theta, Y) \tag{16}$$

where $p(s_{t+1} \mid s_t)$ is the transition probability, and $p(s_t \mid \Theta, Y)$ can be obtained from the Hamilton filter. Using Equation (16) we compute:

$$Pr(s_t = 0 \mid s_{t+1}, \Theta, Y) = \frac{p(s_{t+1} \mid s_t = 1)\, p(s_t = 1 \mid \Theta, Y)}{\sum_{j=1}^{m} p(s_{t+1} \mid s_t = j)\, p(s_t = j \mid \Theta, Y)} \tag{17}$$

Once above probability is computed, we draw a random number from a uniform distribution between 0 and 1, and if the generated number is less than or equal to the value calculated by (17), we set $s_t = 1$, otherwise, set equal to 0.

After drawing $\widetilde{S}_\tau$ by multi-move Gibbs sampling, we generate the transition probabilities, q_{00} and q_{11}, by multiplying (11) and (12) by the likelihood function (9)

$$p\left(q_{00}, q_{11} \mid \widetilde{S}_\tau\right) \propto q_{11}^{u_{11}+m_{11}-1}\left(1-q_{11}\right)^{u_{10}+m_{10}-1} q_{00}^{u_{00}+m_{00}-1}\left(1-q_{00}^{u_{01}+m_{01}-1}\right) \quad (18)$$

Next, we can construct X and Z in (4) using the draw of $\widetilde{S}_\tau$, and then the joint posterior distribution can be obtained from the priors given in (13) and (14) and the likelihood function for B, β,$\Omega(i)$, and $\widetilde{S}_\tau$, that is,

$$\begin{aligned} & p\left(B,\beta,\Omega(1),\ldots,\Omega(m),\widetilde{S}_\tau \mid Y\right) \propto p\left(B,\beta,\Omega(1),\ldots,\Omega(m),\widetilde{S}_T\right) \mathcal{L}\left(B,\beta,\Omega(1),\ldots,\Omega(m),\widetilde{S}_T \mid Y\right) \\ \propto \; & g(b)\left[\prod_{i=1}^{m}\left(|\Omega_0(i)|^{v_0(i)/2}|\Omega(i)|^{-(t_i+v_0(i)+n+1)/2}\right)\right]|\Sigma_{B_0}|^{-1/2}\exp\left\{-\frac{1}{2}\left[\mathrm{tr}\left(\sum_{i=0}^{1}\Omega(i)^{-1}\right)\right.\right. \\ & \left.\left.+\sum_{i=1}^{m}\left[vec(Y_i-W_iB)'\left(\Omega(i)\otimes I_{t_i}\right)^{-1}vec(Y_i-W_iB)\right]+\left[vec(B-B_0)'\Sigma_{B_0}^{-1}vec(B-B_0)\right]\right]\right\} \end{aligned}$$

where $g(b)$ refers to the prior for β given in (10). From the joint posterior, we have the following posterior distributions:

$$\Omega(i) \mid \beta, B, \widetilde{S}_\tau, Y \sim IW\left((Y_i-W_iB)'(Y_i-W_iB)+\Omega_0(i), t_i+v_0(i)+n+1\right) \quad (19)$$

$$vec(B) \mid \Omega(i), \beta, \widetilde{S}_\tau, Y \sim N\left(vec(B_\star), M_\star\right) \quad (20)$$

where

$$M_\star = \left\{\Sigma_{B_0}^{-1} + \sum_{i=1}^{m}\left[\Omega(i)^{-1}\otimes\left(W_i'W_i\right)\right]\right\}^{-1}$$

$$vec(B_\star) = M_\star\left\{\Sigma_{B_0}^{-1}vec(B_0) + \sum_{i=1}^{m}\left[\left(\Omega(i)\otimes I_h\right)^{-1}vec\left(W_i'Y_i\right)\right]\right\}$$

To obtain the conditional posterior for the cointegrating vectors, we rewrite the expression in (4) as

$$
\begin{aligned}
Y - X\Phi &= \sum_{i=1}^{m} Z_i \beta \alpha(i) + E \\
&= \sum_{i=1}^{m} Z_i b a(i) + E \qquad (21)
\end{aligned}
$$

where $a(i)$ and b are such that $a(i) = (\alpha(i)\alpha(i)')^{-\frac{1}{2}}\alpha(i)$ and $\beta = b(b'b)^{-\frac{1}{2}}$. Then vectorize both side of (21) as

$$
\begin{aligned}
vec(Y - X\Phi) &= \sum_{i=1}^{m} vec(Z_i b a(i)) + vec(E) \\
&= \sum_{i=1}^{m} \left(a(i)' \otimes Z_i\right) vec(b) + vec(E) \qquad (22)
\end{aligned}
$$

or

$$
\widetilde{y} = A\widetilde{b} + e \qquad (23)
$$

where $\widetilde{y} = vec(Y - X\Phi)$, $A = \sum_{i=1}^{m} a(i)' \otimes Z_i$, $\widetilde{b} = vec(b)$, and $e = vec(E)$. With the prior for $b \sim MVN(b_0, V_{b_0})$, that is, $\widetilde{b} \sim MN(vec(b_0), V_{b_0})$, the conditional posterior distribution of $\widetilde{b}_i$ is obtained as

$$
\widetilde{b} \mid k, \Omega(0), \Omega(1), B, Y \sim MN(\widetilde{b}_\star, V_{b_\star}) \qquad (24)
$$

where

$$
V_{b_\star} = \left[V_{b_0}^{-1} + \sum_{i=1}^{m} \left\{ \left(a(i)\Omega(i)^{-1}a(i)'\right) \otimes \left(Z_i'Z_i\right) \right\} \right]^{-1}
$$

$$
\widetilde{b}_\star = V_{b_\star} \left(V_{b_0}^{-1} vec(b_0) + \sum_{i=1}^{m} \left\{ \left(a(i)\Omega(i)^{-1}a(i)'\right)^{-1} \otimes \left(Z_i'Z_i\right)^{-1} \right\} \widehat{b} \right).
$$

$\widehat{b} = V_b \left[\sum_{i=1}^{m} \left\{ \left(a(i)\Omega(i)^{-1}\right) \otimes Z_i' \right\}\right] \widetilde{y}$, and $V_b = \sum_{i=1}^{m} \left[\left(a(i)\Omega(i)^{-1}a(i)'\right) \otimes (Z_i'Z_i)\right]$.

Given the conditional posterior distributions, we implement the Gibbs sampling to generate sample draws. The following steps can be replicated until convergence is achieved.

- Step 1: Set $j = 1$. Specify starting values for the parameters in the model, $\widetilde{S}_\tau^{(0)} = \left\{ s_1^{(0)}, s_2^{(0)}, \ldots, s_\tau^{(0)} \right\}'$, $B^{(0)}$, $\beta^{(0)}$ and $\Omega_i^{(0)}$.
- Step 2: Generate $\Omega(i)^{(j)}$ from $p(\Omega(i) \mid \widetilde{S}_\tau^{(j-1)}, \beta^{(j-1)}, B^{(j-1)}, Y)$ for $i = 0, 1$.
- Step 3: Generate $vec(B)^{(j)}$ from $p(vec(B) \mid \widetilde{S}_\tau^{(j-1)}, \beta^{(j-1)}, \Omega(0)^{(j)}, \Omega(1)^{(j)}, Y)$, then obtain $\alpha^*(0), \alpha^*(1)$, and Φ_i^*. Compute $a(i)^*$ using $a(i)^* = (\alpha(i)^*\alpha(i)^{*\prime})^{1/2}\alpha(i)^*$ for $i = 0, 1$.
- Step 4: Generate b^* from $p(\widetilde{b} \mid \widetilde{S}_\tau^{(j-1)}, B^{(j)}, \Omega(0)^{(j)}, \Omega(1)^{(j)}, Y)$. Then, compute $\beta_i^{(j)} = b_i^*(b_i^{*\prime}b_i^*)^{-\frac{1}{2}}$ and $\alpha(i)^{(j)} = a(i)^*(b_i^{*\prime}b_i^*)^{\frac{1}{2}}$ for $i = 0, 1$.
- Step 5: Generate the transition probabilities $(q_{ii})^{(j)}$ from $p\left(q_{ii} \mid \widetilde{S}_\tau^{(j-1)}\right)$ in (18) for $i = 0, 1$.
- Step 6: Generate $\widetilde{S}_\tau^{(j)} = \left\{ s_1^{(j)}, s_2^{(j)}, \ldots, s_\tau^{(j)} \right\}'$ from $p\left(\widetilde{S}_\tau \mid \Theta^{(j)}, Y\right)$, where $\Theta = \{B, \Omega(0), \Omega(1), \beta, q_{00}, q_{11}\}$ in (15), using multi-move Gibbs sampling algorithm.
- Step 7: Set $j = j + 1$, and go to Step 2.

Step 2 through Step 7 can be iterated N times to obtain the posterior means or standard deviations. Note that the first N_0 times iterations are discarded in order to attenuate the effect of the initial values. To check the Gibbs sampler to converge to a sequence of draws from the posteriors, Geweke (1992) suggests the MCMC diagnostic that tests whether the estimate based on the first set of the draws after N_0 burn-in replications is the same as the estimate based on the last set of the draws.

3 Application: U.S. Term Structure of Interest Rates

In this section, we present an empirical study using the MS-VECM to analyze U.S. term structure of interest rates.

3.1 Expectation Hypothesis

The expectations hypothesis of the term structure of interest rates implies an f-period interest rate is the weighted average of the expected future one-period interest rates plus risk premium. For an overview of the expectations hypothesis theory, see Shiller (1990). Let $r_{f,t}$ be the yield to maturity for an f-period at time t, $L_{f,t}$ be the risk premium for an f-period at time t, then the hypothesis implies:

$$r_{f,t} = f^{-1}\sum_{i=1}^{f} E_t r_{1,t+i-1} + L_{f,t} \tag{25}$$

By rewriting the above equation, the interest rate spread $S_{f,t}$ can be expressed as

$$S_{f,t} \equiv r_{f,t} - r_{1,t} = f^{-1}\sum_{i=1}^{f-1}\sum_{j=1}^{i} E_t \Delta r_{1,t+j} + L_{f,t}. \tag{26}$$

If $r_{1,t}$ is integrated of order one, then $r_{f,t}$ is also integrated of order one and thus $r_{f,t}$ and $r_{1,t}$ are cointegrated with cointegrating vector (1, -1) as analyzed by Campbell and Shiller (1987). The risk premium is assumed to be $I(0)$ so that the hypothesis states that $r_{f,t} - r_{1,t} - L_{f,t}$ is a stationary process.

The expectations hypothesis in (26) with constant risk premium implies the following vector error correction model with the lag length at p:

$$\Delta y_t = \eta + (y_{t-1}\beta - L_{f,t})\alpha + \sum_{l=1}^{p} \Delta y_{t-l}\Gamma_l + \varepsilon_t \tag{27}$$

where $y_t = (\ r_{f,t},\ \ r_{1,t}\ )'$; α (1×2) is the adjustment term; β (2×1) is the cointegrating vector; Ψ_i (2×2) is the lag coefficient; and ε_t (2×2) is $iidN(0, \Omega)$.

There is a number of research that confirms nonlinearity of U.S. term structure of interest rates due to changes in monetary policy. Tsay (1998), Hansen and Seo (2002), Clements and Galvao (2002) use a threshold cointegration model, while Clarida et al (2006) employ a Markov switching vector error correction model to detect regime switching. All these studies find nonlinearity due to the instability for interest rates between 1979 and 1982 as a potential source of shifts. This period between 1979 and 1982 is known as the *non-borrowed reserves operating procedure,* that the Federal Reserve moved from interest rate targeting to money growth targeting and allowed the interest rate to fluctuate freely.

3.2 MS-VECM

We apply the MS-VECM to U.S. term structure of interest rate based on (27) to account for the regime shifts. The MS-VECM considered in Section 2 is applied:

$$\begin{aligned}\Delta y_t &= \eta_{S_t} + (y_{t-1}\beta - L_{S_t})\alpha_{S_t} + \sum_{l=1}^{p} \Delta y_{t-i}\Gamma_{l,S_t} + \varepsilon_{t,S_t} &(28)\\ &= \mu_{S_t} + y_{t-1}\beta\alpha_{S_t} + \sum_{l=1}^{p} \Delta y_{t-i}\Gamma_{l,S_t} + \varepsilon_{t,S_t} &(29)\end{aligned}$$

where $\mu_{S_t} = \eta_{S_t} - L_{S_t}\alpha_{S_t}$, $\varepsilon_{t,S_t} \sim N(0, \Omega_{S_t})$; L_{S_t} is the risk premium term depending upon the state variables.

We analyze U.S. term structure of interest rates using the MS-VECM described above. The data set is monthly Federal fund rate, r_t^S, and 1-year Treasury bill rate, r_t^L, covering the period 1960:1 to 2008:8 with 584 observations, obtained from the Federal Reserve Bank of St. Louis. Figure 1 plots the data set and its spread.

Let $y_t = (r_t^L, r_t^S)$, then we consider the following seven models:

$$\mathscr{M}1: \quad \Delta y_t = \mu + y_{t-1}\beta\alpha + \sum_{l=1}^{p} \Delta y_{t-l}\Gamma_l + \varepsilon_t$$

$$\mathscr{M}2: \quad \Delta y_t = \mu(s_t) + y_{t-1}\beta\alpha + \sum_{l=1}^{p} \Delta y_{t-l}\Gamma_l + \varepsilon_t$$

$$\mathscr{M}3: \quad \Delta y_t = \mu(s_t) + y_{t-1}\beta\alpha(s_t) + \sum_{l=1}^{p} \Delta y_{t-l}\Gamma_l + \varepsilon_t$$

$$\mathscr{M}4: \quad \Delta y_t = \mu(s_t) + y_{t-1}\beta\alpha + \sum_{l=1}^{p} \Delta y_{t-l}\Gamma_l + \varepsilon_t(s_t)$$

$$\mathscr{M}5: \quad \Delta y_t = \mu(s_t) + y_{t-1}\beta\alpha(s_t) + \sum_{l=1}^{p} \Delta y_{t-l}\Gamma_l + \varepsilon_t(s_t)$$

$$\mathscr{M}6: \quad \Delta y_t = \mu(s_t) + y_{t-1}\beta\alpha + \sum_{l=1}^{p} \Delta y_{t-l}\Gamma_l(s_t) + \varepsilon_t(s_t)$$

$$\mathscr{M}7: \quad \Delta y_t = \mu(s_t) + y_{t-1}\beta\alpha(s_t) + \sum_{l=1}^{p} \Delta y_{t-l}\Gamma_l(s_t) + \varepsilon_t(s_t)$$

where $\varepsilon_t \sim iidN(0,\Omega)$ and $\varepsilon_t(s_t) \sim iidN(0,\Omega(s_t))$. $\mathscr{M}1$ represents a linear VECM, and other models, $\mathscr{M}2$ - $\mathscr{M}7$, are various specifications of the MSVECMs.

To estimate these models, we implement the collapsed Gibbs sampling algorithm described Section 2.2, For prior hyperparameters, we set $\tilde{b}_0 = (1,-1)'$, $V_{b_0} = \Omega_0(i) = 0.01I_2$ and $v_0(i) = 0.001$for $i=0$ or 1 in (13), $\Sigma_B = 100I_{\kappa n}$and $B_0 = 0$ in (14) favoring the absence of cointegration. These values are assigned to ensure fairly large variances for representing prior ignorance. For prior hyperparameters for the transition probabilities, we set $u_{00} = u_{11} = 9$, $u_{01} = u_{10} = 1$ in (11) and (12). The Gibbs sampler is run with 40,000 times with the first 5,000 discarded.

In this paper, selection of the number of the rank and lags is treated as a problem of model selection. In Bayesian framework, the posterior model probability $p(\mathscr{M}_j \mid Y)$ is used to assess the degree of support for each model, $\mathscr{M}_j$. From the Bayes rule, we have $p(\mathscr{M}_j \mid Y) = p(Y \mid \mathscr{M}_j)p(\mathscr{M}_j)/p(Y)$, where $p(Y \mid \mathscr{M}_j)$ is referred to as the marginal likelihood for $\mathscr{M}_j$; and $p(\mathscr{M}_j)$ is the prior model probability for $\mathscr{M}_j$. Since $p(Y)$ is often hard to calculate, comparison of two models, $\mathscr{M}_j$ and $\mathscr{M}_i$, by the posterior odds ratio, PO_{ji}, is often used to obtain the posterior model probability. The posterior odds ratio is defined as the ratio of their posterior model probabilities as $PO_{ji} = p(\mathscr{M}_j \mid Y)/p(\mathscr{M}_i \mid Y) = \frac{p(Y|\mathscr{M}_j)p(\mathscr{M}_j)}{p(Y|\mathscr{M}_i)p(\mathscr{M}_i)}$, where the ratio of the marginal likelihoods $\frac{p(Y|\mathscr{M}_j)}{p(Y|\mathscr{M}_i)}$ is defined as the Bayes factor. With the posterior odds ratios, we can obtain the posterior model probability as $p(\mathscr{M}_j \mid Y) = PO_{ji}/\sum_{k=1}^{M} PO_{ki}$ where M is the number of models under consideration. Thus, in order to obtain the posterior model probability by the posterior odds, we need to calculate the Bayes factor.

There are several methods to calculate the Bayes factor such as Chib (1995), Gelfand and Dey (1994), the Savage-Dickey density ratio (see Verdinelli and Wasserman, 1995), and the Schwarz Bayesian information criterion (BIC) approximation method (Schwarz, 1978). Chib (1995) provides a method of computing the marginal likelihood that utilizes the output of the Gibbs sampler. The marginal likelihood can be expressed from the Bayes rule as

$$p(y \mid \mathscr{M}_i) = \frac{p(y \mid \theta_i^\star)p(\theta_i^\star)}{p(\theta_i^\star \mid y)} \tag{30}$$

where $p(y \mid \theta_i^\star)$ is the likelihood for $\mathscr{M}_i$ evaluated at $\theta_i^\star$, which is the Gibbs output or the posterior mean of θ_i, $p(\theta_i^\star)$ is the prior density and $p(\theta_i^\star \mid y)$ is the posterior density. If the exact forms of the marginal posteriors are not known like our case, $p(\theta_i^\star \mid y)$ cannot be calculated. To estimate the marginal posterior density evaluated at $\theta_i^\star$ using the conditional posteriors, first block θ into l segments as $\theta = (\theta_1', \ldots, \theta_l')'$, and define $\varphi_{i-1} = (\theta_1', \ldots, \theta_{i-1}')$ and $\varphi^{i+1} = (\theta_{i+1}', \ldots, \theta_l')$. Since $p(\theta^\star \mid y) = \prod_{i=1}^{l} p(\theta_i^\star \mid y, \varphi_{i-1}^\star)$, we can draw $\theta_i^{(j)}$, $\varphi^{i+1,(j)}$, where j indicates the Gibbs output $j = 1, \ldots, N$, from $(\theta_i, \ldots, \theta_l) = (\theta_i, \varphi^{i+1}) \sim p(\theta_i, \varphi^{i+1} \mid y, \varphi_{i-1}^\star)$, and then estimate $\widehat{p}(\theta_i^\star \mid y, \varphi_{i-1}^\star)$ as

$$\widehat{p}(\theta_i^\star \mid y, \varphi_{i-1}^\star) = \frac{1}{N}\sum_{j=1}^{N} p(\theta_i^\star \mid y, \varphi_{i-1}^\star, \varphi^{i+1,(j)}).$$

Thus, the posterior $p(\theta_i^\star \mid Y)$ can be estimated as

$$\widehat{p}(\theta^\star \mid y) = \prod_{i=1}^{I} \left\{ \frac{1}{N}\sum_{j=1}^{N} p(\theta_i^\star \mid y, \varphi_{i-1}^\star, \varphi^{i+1,(j)}) \right\}. \tag{31}$$

Choosing a model among $\mathscr{M}1$ - $\mathscr{M}7$ and testing for cointegration rank with various lag length $p = 1$ to 3 is conducted using the Chib's method. There are three possible rank ($r = 0$, 1, and 2) for models with lag length $p = 1$ to 3 where the adjustment term is constant ($\mathscr{M}1$, $\mathscr{M}2$, $\mathscr{M}4$, and $\mathscr{M}6$). Thus, for models with constant α, we consider $3 \times 3 \times 4 = 36$ models. There are two possible rank ($r = 1$ and 2) for models with $p = 1$ to 3 where the adjustment term changes according to the regime ($\mathscr{M}3$, $\mathscr{M}4$, and $\mathscr{M}7$,). Thus, for models with regime dependent α, we consider $2 \times 3 \times 3 = 18$ models. Therefore, we consider total 54 models and select the most appropriate model among them. From the results of computing the Bayes factors for all 54 models shown in Table 1, the highest posterior model probability is 47.0 percent given to a model of $\mathscr{M}5$ with $p = 2$ and $r = 1$. Table 2 reports the results of the posterior estimation of the parameters for $\mathscr{M}5$ with $p = 2$ and $r = 1$[1]. From the results, the 95% of β (after normalizing) contains $\beta_2 = -1$, that is implied by the expectations hypothesis of the term structure. To examine whether the restriction of $\beta_2 = -1$ is appropriate in a more formal way [2], we calculate the Bayes factor as $BF \approx \exp[-0.5(\text{BIC}_R - \text{BIC}_{UR})]$, where BIC_{UR} is the unrestricted BIC, and BIC_R is the restricted BIC with the restriction of $\beta = (1,$-$1)$, and the value is 132.52, which shows a very strong evidence to support the

[1]Geweke's MCMC convergence diagnostic test (Geweke 1992) for one element of the cointegrating vector generates the statistic -0.17187 with probability 0.43177, that means a sufficiently large number of draws has been taken.

[2]As Koop (2004) note, "the justification for using the HPDIs to compare models is an informal one which, in contrast to posterior odds, is not rooted firmly in probability theory."

expectations hypothesis.[3]

The posterior expectation of the state variables is plotted in Figure 2. The *non-borrowed reserves operating procedure* between 1979 and 1982 is detected as the regime shift. Regime shift occurs also in 1973 and 1984. These regime shifts are corresponding to higher inflation regime (Goodfriend, 1998), and are characterized by a much higher variance of both the long and the short interest rate than those of regime 1. In regime 1, that is relatively stable period, the variance of the long rate is higher than that of the short rate; on the other hand, in regime 0, the short rate fluctuates much more than the long rate.

We find that the magnitudes of the adjustment terms for the short-term rate for both regimes are larger than those for the long-term rate, which implies that the short-term rate tends to adjust toward the equilibrium in either regime. The posterior mean of the adjustment term for the short rate in regime 0, α_0^S, is -0.2298, which is faster than that for the short rate in regime 1 of lower volatility. This implies that interest rates adjust much faster in periods of high volatility with high inflation and anti-inflationary monetary policy.

4 Conclusion

In this paper we consider a Markov switching vector error correction model where the adjustment terms, the lag terms, the intercept terms, and the variance-covariance matrix are subject to the regime shifts with the first order unobservable Markov process while the cointegrating vector is unaffected by the regime shifts.[4]

Estimations are carried out entirely by a Bayesian method. The cointegrat-

[3]See Kass and Raftery (1995) for a rule of thumb for evaluating Bayes factors. According to this rule of thumb, if BF_{ij} is between 20 and 150, there is a strong evidence against model j, and if BF_{ij} exceeds 150, there is a very strong evidence against model j.

[4]It is possible to allow the cointegrating vectors to change with Markov process by slight modification. However, we have not done this because changing the long-run relationship is not reasonable idea unless economic theory support this.

ing vector is drawn using the Markov Chain Monte Carlo method by Koop, Leon-Gonzalez, and Strachan (forthcoming) in a nonlinear framework so that the estimation of the cointegrating vector is more efficient than multi-step classical methods where the cointegrating vector is estimated assuming the model is linear.

As an application to illustrate the use of the MS-VECM, we illustrate U.S. term structure of interest rates using the MS-VECM with regime dependent risk premium. We find that regime with high volatility and high speed of adjustment captures the *non-borrowed reserves operating procedure* during the 1979-82 and other phases of inflation scare, while the stable regime with low volatility and low speed of adjustment prevails after the mid of 80's.

In this paper Markov switching is chosen as a switching behavior, assuming that one regime jumps to another regime suddenly at particular dates. It is of interest to consider alternative multivariate nonlinear models such as a smooth transition vector error correction models (ST-VECM) to analyze the nonlinear cointegration where the regime shifts occur not suddenly but smoothly, and compare the ST-VECM with the MS-VECM by the Bayes factors.

References

[1] Albert, J. H. and S. Chib (1993): "Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts," *Journal of Business and Economics Statistics*,11(1), 1-15.

[2] Anderson, H. M. (1997): "Transaction costs and non-linear adjustment towards equilibrium in the US treasury bill market", *Oxford Bulletin of Economics and Statistics*, 59, 465-483.

[3] Balke, N. S. and T. B. Fomby (1997): "Threshold cointegration", *International Economic Review,* vol.38, No.3, 627-645.

[4] Bauwens, L. and M. Lubrano (1996): "Identification restrictions and posterior densities in cointegrated Gaussian VAR systems", in T. B. Fomby (ed.), *Advances in Econometrics 11B*, Greenwich, CT: JAI press.

[5] Campbell, J. Y. and R. J. Shiller (1987): "Cointegration and tests of present value models", *Journal of Political Economy*, 108, 205-251.

[6] Carter, C. K. and P. Kohn (1994): "On Gibbs sampling for state space models", *Biometrika*, 81, 541-553.

[7] Chib, S. (1995): "Marginal likelihood from the Gibbs output", *Journal of the American Statistical Association*, 90, 432, Theory and Methods, 1313-1321.

[8] Clarida, R. H., L. Sarno, M. P. Taylor, and G. Valente (2006): "The role of asymmetries and regime shifts in the term structure of interest rates", *Journal of Business*, 79, 1193-1224.

[9] Clements, M. P. and A. B. Galvao (2002): "Testing the expectations theory of the term structure of interest rates in threshold models", *Macroeconomic Dynamics*, 1-19.

[10] Doornik, J.A. (2007): *Object-Oriented Matrix Programming Language Using Ox*, London: Timberlage Consultants Press.

[11] Gelfand, A. and D. Dey (1994): "Bayesian model choice: asymptotics and exact calculations", *Journal of the Royal Statistical Society Series*, B56, 501-514.

[12] Geweke, J. (1992): "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments", in Bernardo, J., J. Berger, A. Dawid, and A. Smith (eds.). *Bayesian Statistics*, 4, 641-649. Oxford: Clarendon Press.

[13] Goodfriend, M. (1998): "Using the term structure of interest rates for monetary policy", *Economic Quarterly*, 79, Federal Reserve Bank of Richmond, 1-23.

[14] Hall, S. G., Z. Psaradakis and M. Sola (1997): "Switching error-correction models of house prices in the United Kingdom", *Economic Modelling*, 14, 517-527.

[15] Hamilton, J. (1989): "A new approach to the economic analysis of nonstationary time series and the business cycle", *Econometrica*, 57, 357-384.

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

[17] Johansen, S. (1988): "Statistical analysis of cointegrating vectors", *Journal of Economic Dynamics and Control*, 12, 231-254.

[18] Johansen, S. (1991): "The power function for the likelihood ratio test for cointegration:, in J. Gruber, ed., *Econometric Decision Models: New Methods of Modelling and Applications*, New York: Springer Verlag, 323-335.

[19] Kass, R. E. and A. E. Raftery (1995): "Bayes factors", *Journal of the American Statistical Association*, 90, 773-795.

[20] Kim, C. J. and C. R. Nelson (1998): "Business cycle turning points, a new coincident index and tests of duration dependence based on a dynamic factor model with regime switching", *Review of Economics and Statistics*, 188-201.

[21] Kleibergen, F. and R. Paap (2002): "Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration", *Journal of Econometrics*, 111, 223-249.

[22] Koop, G. (2004): *Bayesian Econometrics*. Wiley.

[23] Koop, G., R. Leon-Gonzalez and R. Strachan (2006): "Bayesian inference in a cointegrating panel data model", *Woking Paper*, University of Leicester, UK.

[24] Koop, G., R. Leon-Gonzalez and R. Strachan (forthcoming): "Efficient posterior simulation for cointegrated models with priors on the cointegration space", *Econometric Reviews*. 29(2), 224-242

[25] Koop, G., R. Strachan and H. van Dijk (2006): "Bayesian approaches to cointegration", in *The Palgrave Handbook of Theoretical Econometrics* edited by K. Patterson and T.Mills. Palgrave Macmillan.

[26] Krolzig, H. M. (1997): *Markov-Switching Vector Autoregressions*. New York: Springer.

[27] Martens, M., P. Kofman and T. C. F. Vorst (1998): "A threshold error-correction model for intraday futures and index returns:, *Journal of Applied Econometrics*, 13, 245-263.

[28] Paap, R., and van Dijk, H.K., (2003): "Bayes Estimates of Markov Trends in Possibly Cointegrated Series: An Application to US Consumption and Income", *Journal of Business and Economic Statistics*, 21, 547-563.

[29] Psaradakis, Z., Sola, M., Spagnolo, F., (2004): "On Markov Error-Correction Models, With an Application to Stock Prices and Dividends", *Journal of Applied Econometrics*, 19, 69-88.

[30] Sarno, L., and G. Valente (2005): "Modelling and forecasting stock returns: Exploiting the futures market, regime shifts and international spillovers", *Journal of Applied Econometrics*, 20, 345-376.

[31] Schwarz, G. (1978): "Estimating the dimension of a model", *Annals of Statistics*, 6, 461-464.

[32] Shiller, R. J. (1990): "The term structure of interest rate", in Friedman, B. and F. Hahn (eds). *Handbook of Monetary Economics*, 1, North-Holland: Amsterdam.

[33] Strachan, R. W. (2003): "Valid Bayesian estimation of the cointegrating error correction model", *Journal of Business and Economic Statistics*, 21, 185-195.

[34] Strachan, R. W. and H. K. van Dijk (2003): "Bayesian model selection with an uninformative prior", *Oxford Bulletin of Economics and Statistics*, 65, 863-876.

[35] Strachan, R. W. and H. K. van Dijk (2004): "Valuing structure, model uncertainty and model averaging in vector autoregressive process", *Econometric Institute Report EI 2004-23*. Erasmus University Rotterdam, Rotterdam.

[36] Strachan, R. W. and H. K. van Dijk (2006): "Model uncertainty and Bayesian model averaging in vector autoregressive processes", *Discussion Papers in Economics No. 06/5*. Department of Economics, University of Leicester, UK.

[37] Strachan, R. W. and H. K. van Dijk (2007): "Bayesian averaging in vector autoregressive processes with an investigation of stability of the US great ratios and risk of a liquidity trap in the USA, UK and Japan", *Econometric Institute Report EI 2007-11*. Erasmus University Rotterdam. Rotterdam.

[38] Strachan, R. W. and B. Inder (2004): "Bayesian analysis of the error correction model", *Journal of Econometrics*, 123, 307-325.

[39] Sugita, K. (2009): "Bayesian Analysis of a Vector Autoregressive Model with Multiple Structural Breaks",*Economics Bulletin*, 3 (27), 1-7.

[40] Tsay, R. S. (1998): "Testing and modeling multivariate threshold models", *Journal of the American Statistical Association*, 93, 1188-1202.

[41] Verdinelli, I. and L. Wasserman (1995): "Computing Bayes factors using a generalization of the Savage-Dickey density ratio", *Journal of the American Statistical Association*, 90, 614-618.

[42] Villani, M. (2005): "Bayesian Reference Analysis of Cointegration", *Econometric Theory*, 21, 326-357.

[43] Villani, M. (2006): "Bayesian Point Estimation of the Cointegration Space", *Journal of Econometrics*, 134, 645-664.

Table 1: Model Selection for U.S. Term Structure of Interest Rate

		$\mathscr{M}1$	$\mathscr{M}2$	$\mathscr{M}3$	$\mathscr{M}4$	$\mathscr{M}5$	$\mathscr{M}6$	$\mathscr{M}7$
$p=1$	$r=0$	0.000	0.000	——	0.000	——	0.002	——
	$r=1$	0.000	0.000	0.003	0.012	0.152	0.011	0.032
	$r=2$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$p=2$	$r=0$	0.000	0.001	——	0.003	——	0.006	——
	$r=1$	0.000	0.000	0.027	0.073	0.470	0.052	0.102
	$r=2$	0.000	0.000	0.000	0.000	0.004	0.000	0.003
$p=3$	$r=0$	0.000	0.000	——	0.000	——	0.000	——
	$r=1$	0.000	0.000	0.000	0.001	0.025	0.003	0.008
	$r=2$	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: Each value in the Table shows the posterior model probability calculated by using the Chib's method.

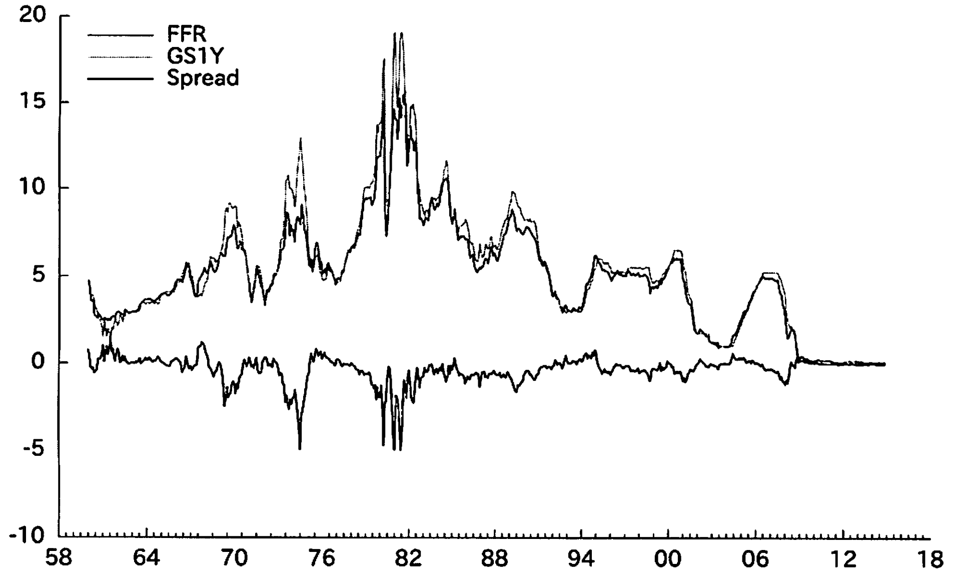
Table 2 Posterior Results for $\mathscr{M}5$ with $p = 2$

( ) = standard deviation

parameter	mean	95% HPDI
β_2	-1.0494 (0.0355)	-1.1212, -0.9819
$\alpha^L(0)$	0.0054 (0.0608)	-0.1143, 0.1242
$\alpha^S(0)$	-0.2298 (0.0752)	-0.3803, -0.0854
$\alpha^L(1)$	-0.0361 (0.0283)	-0.0917, 0.0186
$\alpha^S(1)$	-0.1210 (0.0223)	-0.1644, -0.0773
$\mu^L(0)$	-0.0162 (0.0885)	-0.1888, 0.1577
$\mu^S(0)$	-0.0144 (0.1383)	-0.2527, 0.2965
$\mu^L(1)$	-0.0127 (0.0174)	-0.0487, 0.0199
$\mu^S(1)$	-0.0465 (0.0246)	-0.0955, -0.0001
p_{00}	0.9814 (0.0067)	0.9663, 0.9921
p_{11}	0.9276 (0.0248)	0.8725, 0.9682

$$\Omega_0 = \begin{pmatrix} 0.7469 & 0.6008 \\ (0.0895) & (0.0936) \\ 0.6008 & 1.1828 \\ (0.0936) & (0.1809) \end{pmatrix}, \Omega_1 = \begin{pmatrix} 0.0580 & 0.0257 \\ (0.0035) & (0.0022) \\ 0.0257 & 0.0344 \\ (0.0022) & (0.0027) \end{pmatrix}.$$

Figure 1: US Federal fund rate, 1-year treasury bill rate a the spread



Source: Federal Reserve Bank of St.Louis

Figure 2: Posterior expectation of the regime variable for the US Term Structure of Interest rates