

#2006-13

Discussion Paper #2006-13

Bayesian Analysis of Markov Switching

Vector Error Correction Model

by

Katsuhiro Sugita

Bayesian Analysis of Markov Switching Vector Error Correction Model

Katsuhiko Sugita*

Graduate School of Economics, Hitotsubashi University, 2-1, Naka,

Kunitachi, Tokyo, 186-8601, Japan

E-Mail: ksugita@econ.hit-u.ac.jp

Abstract

This paper introduces a Bayesian approach to a Markov switching vector error correction model that allows for regime shifts in the intercept terms, the lag terms, the adjustment terms and the variance-covariance matrix. The proposed Bayesian method allows for estimation of the cointegrating vector within a nonlinear framework through Gibbs sampling so that it generates more efficient estimation than classical approaches that require a multi-stage maximum likelihood procedure. The Bayes factors are applied to test for Markov switching and model specifications. We apply the proposed model to U.S. term structure of interest rates allowing the risk premium and other parameters in the model to change with regime.

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 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 intercept 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 in order 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. Other research using threshold cointegration model include Anderson (1997), Tsay (1998), Martens et al (1998), and Clements and Galvao (2002). Instead of threshold cointegration model, nonlinear cointegration model using Hamilton's (1989) Markov regime switching process is also developed (Krolzig, 1997). 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 a 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. Then, 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 a MS-VECM that allows the intercept, the lag terms, the adjustment coefficients, and the variance-covariance matrix to shift with Markov process. For a Bayesian approach to a MS-VECM, Paap and van Dijk (2003) propose a nonlinear VECM where the intercept terms are affected by Markov regime shift in order to investigate U.S. consumption and income. They employ a Bayesian cointegration analysis based on Kleibergen and Paap (2002) and Kleibergen and van Dijk (1998), which requires linear normalizing restrictions on the cointegrating vectors, that are criticized by Strachan (2003) as being likely to be invalid. Strachan and van Dijk (2003a) 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 et al (2006b) for details. In this paper, to estimate the cointegrating vectors, we apply Strachan and Inder's (2004) Bayesian method, that elicits a valid prior for the cointegrating space.

Our model in this paper is more general than Paap and van Dijk (2003), and is flexible to modify in order 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 constant in the cointegrating vectors to be affected by the

regime shifts by a slight modification. We present this modification in the application of the model which consider the regime dependent risk premium term to investigate U.S. term structure of interest rates.

The plan of the paper is as follows. Section 2 presents estimation method for the MS-VECM using a Gibbs sampler. We specify prior densities and likelihood functions, and then derive the posterior distributions. In Section 3, we describe testing for Markov switching and the number of cointegrating rank as model selection by Bayes factors. For testing for nonlinearity, the Schwarz BIC is used to approximate the Bayes factors. To determine the number of rank, the Savage-Dickey density ratio is applied to compute the Bayes factors. Then we show simulated experiments with artificially generated data to evaluate the performance of detecting an appropriate model by the proposed method. Section 4 illustrates application to U.S. term structure of interest rates, using the MS-VECM where the risk premium term is also affected by the regime shifts as well as the intercept terms, the lag terms, the variance-covariance matrix, and the adjustment coefficients. Section 5 contains concluding remarks. The computations reported in this paper were performed using code written in the Ox programming language (Doornik, 1998).

2 Markov Switching Vector Error Correction Model

This section introduces a MS-VECM and presents a Bayesian approach to estimate this model. Let X_t denote an $I(1)$ vector of n -dimensional time series with r linear cointegrating relations. A VAR(p) system with normally distributed Gaussian innovations $\varepsilon \sim iidN(0, \Sigma)$ can be written as a vector error correction model (VECM)

$$\Delta X_t = \mu + \alpha \beta' X_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta X_{t-i} + \varepsilon_t \quad (1)$$

where α ($n \times r$) is adjustment term; and β' ($n \times r$) is cointegrating vector. If we assume that the intercept term μ , the adjustment term α , the lag terms Ψ_i , and the variance-covariance matrix Σ in the VECM are subject to an unobservable discrete state variable

$s_t = i$ where $i = 1, 2, \dots, m$, then the VECM representation is written as

$$\Delta X_t = \mu(s_t) + \alpha(s_t)\beta'X_{t-1} + \sum_{i=1}^{p-1} \Psi_i(s_t)\Delta X_{t-i} + \varepsilon_t \quad (2)$$

where ε_t are assumed $N(0, \Sigma(s_t))$ and independent over time. Dimensions of matrices are μ and ε ($n \times 1$), Ψ_i and σ ($n \times n$). The state variable s_t evolves according to a m -state, first-order Markov switching process with the transition probabilities, $p(s_t = i | s_{t-1} = j) = p_{ij}$, $i, j = 1, \dots, m$.

Equation (2) can be rewritten in the matrix format as

$$Y = WB + E \quad (3)$$

where $Y = (\Delta X_p, \dots, \Delta X_T)'$, $W = (I_1 Z \beta, \dots, I_m Z \beta)$, $B = (\alpha, \Gamma)'$, $Z = (X_{p-1}, \dots, X_{T-1})'$,

$E = (\varepsilon_p, \dots, \varepsilon_T)'$, $\alpha = (\alpha_1, \dots, \alpha_m)$,

$\Gamma = (\mu_1, \dots, \mu_m, \Psi_{1,1}, \dots, \Psi_{p-1,1}, \dots, \Psi_{m,1}, \dots, \Psi_{m,p-1})'$,

$$X = \begin{pmatrix} s_{1,p} & \cdots & s_{m,p} & s_{1,p}\Delta X'_{p-1} & \cdots & s_{1,p}\Delta X'_1 & \cdots & s_{m,p}\Delta X'_{p-1} & \cdots & s_{m,p}\Delta X'_{p-1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{1,T} & \cdots & s_{m,T} & s_{1,T}\Delta X'_{T-1} & \cdots & s_{1,T}\Delta X'_{T-p+1} & \cdots & s_{m,T}\Delta X'_{T-1} & \cdots & s_{m,T}\Delta X'_{T-p+1} \end{pmatrix}.$$

Let τ be the number of rows of Y , so that $\tau = T - p + 1$, then X is $\tau \times m(1 + n(p - 1))$, Γ is $m(1 + n(p - 1)) \times n$, W is $\tau \times \kappa$ where $\kappa = m(1 + n(p - 1) + r)$, and B is $\kappa \times n$. $s_{i,j}$ in X is an indicate variable such that it equals to 1 if regime is i and 0 otherwise. I_i in W is an indicator matrix ($\tau \times \tau$) where the diagonal elements are 1 if t^{th} regime is i , otherwise 0, and the off-diagonal elements are all 0. Equation (3) represents the stacked form of (2).

2.1 Prior Distributions and Likelihood Functions

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 $\beta' = (I_r, \beta'_*)$ where β_* 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 . This prior with linear normalization, however, is criticized by Strachan (2003) and Strachan and Inder (2004) as invalid because this normalization restricts the estimable region of the cointegrating space. Instead of using this prior, they propose the Grassman approach that places a prior on the cointegrating space rather than the cointegrating vectors. There is a number of research based on the Grassman approach. Strachan and van Dijk (2003a), for example, analyze a vector error correction model using an uniform prior for the cointegrating space. Strachan and van Dijk (2003b) investigate an issue of model selection with a proper diffuse prior on the cointegrating vector. Koop et al (2006a) extend to a panel cointegration model, based on the Grassman prior. Koop et al. (2006b) discuss the prior elicitation for the cointegrating vector in detail. In this paper, we also adopt the Grassman approach to place a noninformative prior on the cointegrating space as

$$g(\beta) \propto \pi^{-(n-r)r} \prod_{j=1}^r \frac{\Gamma[(n+1-j)/2]}{\Gamma[(r+1-j)/2]} \quad (4)$$

where $\Gamma[q] = \int_0^\infty u^{q-1} e^{-u} du$, $q > 0$ with identification restrictions, $\beta'\beta = I_n$, that do not distort the weight on the cointegrating space.

For a prior for the transition probabilities p_{ij} , $i, j = 1, \dots, m$, we assign a beta distribution, assuming $m = 2$

$$p_{00} \sim \text{beta}(u_{00}, u_{01}) \quad (5)$$

$$p_{11} \sim \text{beta}(u_{11}, u_{10}) \quad (6)$$

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

With regard to priors for B , Ω_i , we assume prior independence between B and Ω_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 h

as

$$\Omega_i \sim IW(\Phi_i, h_i) \quad (7)$$

where $\Phi_i \in \mathbb{R}^{n \times n}$. As for a prior for B , we consider the vector form of B and assign a multivariate normal as

$$\text{vec}(B) \sim MN(\text{vec}(B_0), \Sigma_B) \quad (8)$$

where MN refers to a multivariate normal with mean $\text{vec}(B_0) \in \mathbb{R}^{kn \times 1}$, $k = m(1 + n(p-1) + r)$ and variance-covariance matrix $\Sigma_B \in \mathbb{R}^{kn \times kn}$. We assume that α_i , $i = 1, \dots, m$, is distributed independently, so that Σ_B is defined as

$$\Sigma_B = \begin{bmatrix} \Sigma_\alpha & 0 \\ 0 & \Sigma_\Gamma \end{bmatrix} \quad (9)$$

where Σ_α is $nrm \times nrm$ matrix such that $\Sigma_\alpha = \underline{V}_\alpha \otimes I_{rm}$, \underline{V}_α ($n \times n$) is prior variance-covariance matrix of $\alpha_i \sim MVN(\underline{\alpha}, \underline{V}_\alpha)$; Σ_Γ is $n(m + n(p-1)) \times n(m + n(p-1))$ matrix and is prior variance-covariance matrix of $\Gamma | \alpha \sim MVN(\underline{\Gamma}, \Sigma_\Gamma)$. This independence relation of distributions among $\alpha_1, \dots, \alpha_m$ is convenient for determining the cointegrating rank using the Savage-Dickey density ratio described in Section 3.

The likelihood function for $B, \Omega_1, \dots, \Omega_m, \beta$ and the state variables $\tilde{S}_T = \{s_1, s_2, \dots, s_T\}'$ is given by,

$$\begin{aligned} & \mathcal{L}(B, \beta, \Omega_1, \dots, \Omega_m, \tilde{S}_T | Y) \\ & \propto \left(\prod_{i=1}^m |\Omega_i|^{-t_i/2} \right) \exp \left(-\frac{1}{2} \text{tr} \left[\sum_{i=1}^m \{ \Omega_i^{-1} (Y_i - W_i B)' (Y_i - W_i B) \} \right] \right) \end{aligned} \quad (10)$$

$$= \left(\prod_{i=1}^m |\Omega_i|^{-t_i/2} \right) \exp \left(-\frac{1}{2} \sum_{i=1}^m \left[(\text{vec}(Y_i - W_i B))' (\Omega_i \otimes I_\tau)^{-1} (\text{vec}(Y_i - W_i B)) \right] \right) \quad (11)$$

where $Y_i = I_i Y$, $W_i = I_i W$ and t_i is the total number of observations when $s_t = i$, $i = 1, 2, \dots, m$.

The likelihood function for the transition probabilities p_{ij} , $i, j = 1, 2, \dots, m$, which are independent of the data but conditional on the set of the state variables, is given assuming $m = 2$:

$$\mathfrak{L}(p_{11}, p_{22} | \tilde{S}_T) = p_{11}^{m_{11}} (1 - p_{11})^{m_{12}} p_{22}^{m_{22}} (1 - p_{22})^{m_{21}} \quad (12)$$

where $m_{i,j}$, $i, j = 1$ or 2 , denotes the number of the transition from the regime i to j , that can be counted from given \tilde{S}_T .

2.2 Posterior Specifications

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

To sample the state variable \tilde{S}_τ 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 , $t = 1, 2, \dots, T$, as a block from the following conditional distribution:

$$p(\tilde{S}_\tau | \Theta, Y) = p(s_\tau | \Theta, Y) \prod_{t=p}^{\tau-1} p(s_t | s_{t+1}, \Theta, Y) \quad (13)$$

where $\Theta = \{B, \beta, \Omega_1, \Omega_2, p_{11}, p_{22}\}$. The first term of the right hand side of the above equation, $p(s_\tau | \Theta, Y)$, 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:

$$p(s_t | s_{t+1}, \Theta, Y) = \frac{p(s_{t+1} | s_t, \Theta, Y) p(s_t | \Theta, Y)}{p(s_{t+1} | \Theta, Y)} \propto p(s_{t+1} | s_t) p(s_t | \Theta, Y) \quad (14)$$

where $p(s_{t+1} | s_t)$ is the transition probability, and $p(s_t | \Theta, Y)$ can be obtained from

the Hamilton filter. Using Equation (14) we compute:

$$Pr(s_t = 0 | s_{t+1}, \Theta, Y) = \frac{p(s_{t+1} | s_t = 1) p(s_t = 1 | \Theta, Y)}{\sum_{j=0}^1 p(s_{t+1} | s_t = j) p(s_t = j | \Theta, Y)} \quad (15)$$

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 (15), we set $s_t = 1$, otherwise, set equal to 0.

After drawing \tilde{S}_τ by multi-move Gibbs sampling, we generate the transition probabilities, p_{11} and p_{22} , by multiplying (5) and (6) by the likelihood function (12)

$$p(p_{11}, p_{22} | \tilde{S}_\tau) \propto p_{22}^{u_{22} + m_{22} - 1} (1 - p_{22})^{u_{21} + m_{21} - 1} p_{11}^{u_{11} + m_{11} - 1} (1 - p_{11})^{u_{12} + m_{12} - 1} \quad (16)$$

Next, we can construct X and Z in (3) using the draw of \tilde{S}_τ , and then the joint posterior distribution can be obtained from the priors given in (7) and (8) and the likelihood function for B, β, Ω_i , and \tilde{S}_τ , that is,

$$\begin{aligned} & p(B, \beta, \Omega_1, \dots, \Omega_m, \tilde{S}_\tau | Y) \propto p(B, \beta, \Omega_1, \dots, \Omega_m, \tilde{S}_T) \mathcal{L}(B, \beta, \Omega_1, \dots, \Omega_m, \tilde{S}_T | Y) \\ & \propto g(\beta) \left[\prod_{i=1}^m (|J_i|^{h_i/2} |\Omega_i|^{-(t_i-1)/2}) \right] \exp \left\{ -\frac{1}{2} \left[\text{tr} \left(\sum_{i=1}^m \Omega_i^{-1} \right) \right. \right. \\ & \left. \left. + \sum_{i=1}^m \left[\text{vec}(Y_i - W_i B)' (\Omega_i \otimes I_\tau)^{-1} \text{vec}(Y_i - W_i B) \right] + [\text{vec}(B - B_0)' \Sigma_B^{-1} \text{vec}(B - B_0)] \right] \right\} \end{aligned} \quad (17)$$

where $g(\beta)$ refers to the prior for β given in (4). From the joint posterior (17), we have the following posterior distributions (see Appendix A.1 for derivation of these posteriors):

$$\Omega_i | \beta, B, \tilde{S}_\tau, Y \sim IW \left((Y_i - W_i B)' (Y_i - W_i B) + \Phi_i, t_i + h_i \right) \quad (18)$$

$$vec(B) | \Omega_i, \beta, \tilde{S}_\tau, Y \sim N(vec(B_\star), M_\star) \quad (19)$$

where

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

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

For the posterior for β , according to Strachan and Inder (2004), let $z_{1,t}$ be the $(t-1)^{th}$ row of Z and $z_{2,t}$ be the t^{th} row of X , then

$$p(\beta | \tilde{S}_\tau, Y) \propto g(\beta) |\beta' D_0 \beta|^{-\tau/2} |\beta' D_1 \beta|^{(\tau-n)/2} \quad (20)$$

where $D_0 = D_1 - D_2$, $D_1 = S_{11}$ and $D_2 = S_{10} S_{00}^{-1} S_{01}$, $S_{jk} = M_{jk} - M_{j2} M_{22}^{-1} M_{2k}$, $M_{jk} = h_{jk} + \sum z'_{j,t} z_{k,t}$, $h_{jk} = 0$ if $j \neq k$, $h_{jj} = \phi I$.

The posterior distributions for Ω_i in (18), B in (19), and β in (20) are not convenient analytical forms. Rather they are conditional on other parameters which must be estimated. Gibbs sampler can be employed to generate random draws from the conditional posteriors. While the conditional posterior densities for Σ_i and B are known form, the posterior for β in (20) is not a standard form and thus can be drawn by employing importance sampling, the Metropolis-Hastings algorithm (see Chib and Greenberg, 1995) or the Griddy-Gibbs sampling (see Ritter and Tanner, 1992). In this paper, we choose the Griddy-Gibbs sampling technique because the algorithm does not require the specification of function that approximate the distribution. Choosing the Griddy-Gibbs sampler, however, requires the appropriate choice of the grid of points and the computing cost is much higher than other algorithms. The algorithm is provided in Appendix A.2 for convenience.

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, $p_{ik}^{(0)}$, $B^{(0)}$, $\beta^{(0)}$ and $\Omega_i^{(0)}$.
- Step 2: Generate $\tilde{S}_\tau^{(j)} = \{s_1^{(j)}, s_2^{(j)}, \dots, s_\tau^{(j)}\}'$ from $p(\tilde{S}_\tau | \Theta^{(j-1)}, Y)$, where $\Theta = \{B, \Omega_1, \dots, \Omega_m, \beta, p_{ik}\}$ in (13), using multi-move Gibbs sampling algorithm.
- Step 3: Generate the transition probabilities $(p_{ik})^{(j)}$ from $p(p_{ik} | \tilde{S}_\tau^{(j)})$ in (16).
- Step 4: Generate $\beta^{(j)}$ from $p(\beta | \tilde{S}_\tau^{(j)}, Y)$ in (20) using the Griddy-Gibbs sampling algorithm.
- Step 5: Generate $B^{(j)}$ from $p(\text{vec}(B) | \beta^{(j)}, \Omega_i^{(j-1)}, \tilde{S}_\tau^{(j)}, Y)$ in (19).
- Step 6: Generate $\Omega_i^{(j)}$ from $p(\Omega_i | \beta^{(j)}, B^{(j)}, \tilde{S}_\tau^{(j)}, Y)$ in (18).
- 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.

3 Testing for Markov Switching, Cointegrating Rank, and Model Selection by Bayes Factors

In this paper, testing for Markov switching and the cointegrating rank is treated as a problem of model selection. In Bayesian framework, the posterior model probability $p(\mathcal{M}_j | Y)$ is used to assess the degree of support for each model, \mathcal{M}_j . From the Bayes rule, we have $p(\mathcal{M}_j | Y) = p(Y | \mathcal{M}_j)p(\mathcal{M}_j)/p(Y)$, where $p(Y | \mathcal{M}_j)$ is referred to as the marginal likelihood for \mathcal{M}_j ; and $p(\mathcal{M}_j)$ is the prior model probability for \mathcal{M}_j . Since $p(Y)$ is often hard to calculate, comparison of two models, j and 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 $PO_{ji} = p(\mathcal{M}_j | Y) / p(\mathcal{M}_i | Y) = \frac{p(Y|\mathcal{M}_j)p(\mathcal{M}_j)}{p(Y|\mathcal{M}_i)p(\mathcal{M}_i)}$, where the ratio of the marginal likelihoods $\frac{p(Y|\mathcal{M}_j)}{p(Y|\mathcal{M}_i)}$ is defined as the Bayes factor. With the posterior odds ratios, we can obtain the posterior model probability as $p(\mathcal{M}_j | Y) = PO_{ji} / \sum_{k=1}^M PO_{ki}$ 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). Among these, we choose the Schwarz BIC method to test for nonlinearity and select the most appropriate model since other methods are not possible or difficult to perform.¹ The Schwarz BIC can give a rough approximation to the Bayes factors. It is, however, easy to implement and does not require evaluation of the prior distribution, as Kass and Raftery (1995) note. The Schwarz BIC to approximate the Bayes factors is employed by Wang and Zivot (2000) for detecting the number of structural breaks. The Schwarz BIC for \mathcal{M}_j is calculated as

$$\text{BIC}_j = -2 \ln \mathcal{L}(\hat{\theta}_j | Y; \mathcal{M}_j) + q_j \ln(t) \quad (21)$$

where $\mathcal{L}(\hat{\theta}_j | Y; \mathcal{M}_j)$ denotes the likelihood function under the model j ; q_j denotes the total number of estimated parameters in \mathcal{M}_j . The likelihood function $\mathcal{L}(\hat{\theta}_j | Y; \mathcal{M}_j)$ is evaluated at $\hat{\theta}_j$, the posterior means of the parameters for \mathcal{M}_j .

With the Schwarz BICs for \mathcal{M}_j and \mathcal{M}_i , the Bayes factor for \mathcal{M}_j against \mathcal{M}_i can be approximated by

$$BF_{ji} = \exp[-0.5 (\text{BIC}_i - \text{BIC}_j)]. \quad (22)$$

¹ For example, a method by Chib (1995) requires to know the full form of prior, likelihood, and posterior. In the MS-VECM, the full forms of posterior for β in (20) is difficult to obtain. In this case, Chib and Jeliazkov (2001) show how the marginal likelihood can be calculated using the output from the Metropolis-Hastings algorithm.

The Savage-Dickey density ratio method is applicable only to nested models, while the Gelfand and Dey method is not suitable to multivariate models.

With the prior odds, defined as $p(\mathcal{M}_j)/p(\mathcal{M}_i)$, the posterior odds can be computed by multiplying the Bayes factor by the prior odds as $\text{PosteriorOdds}_{ji} = BF_{ji} \times \text{PriorOdds}_{ji}$. By using the Schwarz BIC to approximate to logarithm of the Bayes factor, it is easy to test Markov switching cointegration as a problem of model selection. In our case, we compute the Schwarz BIC such that

$$\begin{aligned} \text{BIC}_j &= -2 \ln \mathcal{L} \left(B, \Sigma_1, \dots, \Sigma_m, \beta, \tilde{S}_\tau, p_{ik} \mid Y; \mathcal{M}_j \right) + q_j \ln(t) \\ &= -2 \left\{ \ln \mathcal{L} \left(B, \Sigma_1, \dots, \Sigma_m, \beta, \tilde{S}_\tau \mid Y; \mathcal{M}_j \right) + \ln \mathcal{L} \left(p_{ik} \mid \tilde{S}_\tau; \mathcal{M}_j \right) \right\} + q_j \ln(t) \end{aligned}$$

To determine the number of cointegrating rank, Strachan and Inder (2004) employ the Laplace approximation method to calculate the Bayes factors. In this paper, instead of applying the Laplace approximation method, we choose the Savage-Dickey density ratio to calculate the Bayes factors to determine the number of cointegrating rank. The Bayes factor comparing zero rank $r = 0$ and non-zero rank $r = r^*$ is obtained using the Savage-Dickey density ratio as follows:

$$\begin{aligned} BF(r = 0 \mid r = r^*) &= BF(\alpha = 0 \mid \alpha \neq 0) \\ &= \frac{p(\alpha = 0 \mid Y, \mathcal{M}_{r^*})}{p(\alpha = 0 \mid \mathcal{M}_{r^*})} \end{aligned} \quad (23)$$

where $r^* > 0$ is the number of rank to test; \mathcal{M}_{r^*} denotes a model with rank r^* ; the denominator is the prior density evaluated at $\alpha = 0$; and the numerator is the posterior density evaluated at $\alpha = 0$. The prior for B , $\text{vec}(B) \sim MN(\text{vec}(B_0), \Sigma_B)$ with Σ_B defined in (9), implies $p(\alpha) = \prod_{i=1}^{m+1} p(\alpha_i)$, where $\alpha_i \sim MVN(\underline{\alpha}, \underline{V}_\alpha)$. The posterior for α_i is also independently distributed as $\alpha_i \mid \beta, \Omega_i, Y_i \sim MVN(\bar{\alpha}_i, \bar{V}_{\alpha,i})$ where $\bar{V}_{\alpha,i} = (\underline{V}_\alpha^{-1} + Z_i' Z_i \Omega_i^{-1})^{-1}$ and $\bar{\alpha}_i = \bar{V}_{\alpha,i} (\underline{V}_\alpha^{-1} \underline{\alpha} + Z_i' (Y_i - X_i \Gamma_i) \Omega_i)$. Since

$$\frac{1}{N - N_0} \sum_{j=N_0+1}^N p(\alpha = 0 \mid \beta^{(j)}, \Omega^{(j)}, Y, r^*) \rightarrow p(\alpha = 0 \mid Y, \mathcal{M}_{r^*}) \quad (24)$$

as N goes to infinity, the numerator of (23) can be easily calculated.

To evaluate the performance of detecting the Markov switching nonlinearity by the Schwarz BIC, we conducted Monte Carlo simulations. We consider three-variable MS-VECM for the experiments. The data generating processes (DGPs) are given as the following:

$$\mathcal{M}1: \Delta X_t = \mu_0 + \alpha_0 \beta' X_{t-1} + \Psi_0 \Delta X_{t-1} + \sigma_0 \varepsilon_t$$

$$\mathcal{M}2: \Delta X_t = \mu(s_t) + \alpha(s_t) \beta' X_{t-1} + \Psi(s_t) \Delta X_{t-1} + \sigma_0 \varepsilon_t$$

$$\mathcal{M}3: \Delta X_t = \mu_0 + \alpha(s_t) \beta' X_{t-1} + \Psi_0 \Delta X_{t-1} + \sigma(s_t) \varepsilon_t$$

$$\mathcal{M}4: \Delta X_t = \mu(s_t) + \alpha(s_t) \beta' X_{t-1} + \Psi(s_t) \Delta X_{t-1} + \sigma(s_t) \varepsilon_t$$

where $\mu(s_t = 0) = \mu_0 = (0.2, 0.2, 0.2)'$, $\mu(s_t = 1) = \mu_1 = (-0.2, -0.2, -0.2)'$, $\alpha(s_t = 0) = \alpha_0 = (-0.2, -0.2, 0.2)'$, $\alpha(s_t = 1) = \alpha_1 = (0, 0, 0)'$, $\beta' = (1, -1, 1)$, $\varepsilon_t \sim NID(0, I_3)$, $\sigma(s_t = 0) = \sigma_0 = 0.5I_3$, $\sigma(s_t = 1) = \sigma_1 = 0.1I_3$, $\Psi(s_t = 0) = \Psi_0 = 0.5I_3$, and $\Psi(s_t = 1) = \Psi_1 = 0.2I_3$ the sample size $T = \{100, 200, 500\}$. The transition probabilities are given as $(p_{00}, p_{11}) = (0.95, 0.95)$. These four DGPs represent:

$\mathcal{M}1$: Linear VECM model.

$\mathcal{M}2$: Homoskedastic MS-VECM with regime dependent mean

$\mathcal{M}3$: Heteroskedastic MS-VECM with constant mean

$\mathcal{M}4$: Heteroskedastic MS-VECM with regime dependent mean

It might be possible to compute Bayes factors for all model $\mathcal{M}1 - \mathcal{M}4$ to select the most appropriate model. However, if the true model is linear as $\mathcal{M}1$, computation of the Bayes factors for $\mathcal{M}2 - \mathcal{M}4$ might not be feasible because of the problem that the state variables and the transition probabilities are not identified through the Gibbs sampling, causing a convergence problem in the Gibbs sampler. Kim and Nelson (2001) overcome this problem by employing 'pseudo priors' (see Carlin and Chib, 1995). In this paper we restrict a priori that a certain percentage of the observations lies in each regime as Koop and Potter (1999) in order to avoid the problem. When the total number of either regime occurred in the generated state variables at j -th iteration

of Gibbs sampling, that is, $\sum_{i=1}^{\tau} s_i^{(j)}$ or $\sum_{i=1}^{\tau} (1 - s_i^{(j)})$, is less than given value, say, 5 percent of the sample size, then the previously drawn set of state variables, $\tilde{S}_{\tau}^{(j-1)}$, is used in the j -th iteration of the Gibbs sampler by setting $\tilde{S}_{\tau}^{(j)} = \tilde{S}_{\tau}^{(j-1)}$.

The Bayes factors were computed for all models ($\mathcal{M}1 - \mathcal{M}4$) to calculate the posterior model probability for each model. For prior hyperparameters, we set $\Phi_i = 0.01I_3$ and $h_i = 0.001$ for all i in (7), $\Sigma_B = 100I_{kn}$ and $B_0 = 0$ in (8) 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 (5) and (6). The number of cointegration rank and the lag length are assumed to be known. Each simulation is replicated 1000 times. For each replication of the simulations, the Griddy-Gibbs sampler is employed with 5,000 draws with the first 500 discarded to generate the cointegrating vector with the interval of integration (the deterministic Simpson's rule is used) for each element of β from -6.00 to 6.00 with the number of the grid at 1200 to avoid significant truncation of the posterior density.

Table 1 summarizes the results of Monte Carlo simulations for model selection. The value in each element of the Table represents the average posterior model probability. When the true model is the linear VECM, $\mathcal{M}1$, the average posterior model probability selects the correct model $\mathcal{M}1$ with more than 90% even when $T = 100$. When the true model is the MS-VECM, $\mathcal{M}2$, $\mathcal{M}3$ or $\mathcal{M}4$, the average posterior model probability selects the correct model with less than 90% for the three DGPs when $T = 100$. Increasing the sample size to 200 improves the performances as the correct model is selected with more than 90%.

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

We present an empirical study using the MS-VECM to analyze U.S. term structure of interest rates. It is possible to apply the MS-VECM of the form (2) to examine the expectations hypothesis of U.S. term structure of interest rates. However, with some

minor modifications, the model can be extended to allow the risk premium term in the cointegrating vector to shift with regime, rather than assuming that the risk premium is constant. With this model, we can investigate nonlinearity of U.S. term structure of interest rates by changes in monetary policy.

4.1 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:

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

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

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

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 $p - 1$:

$$\Delta X_t = \mu + \alpha(\beta' X_{t-1} - L_{f,t}) + \sum_{i=1}^{p-1} \Psi_i \Delta X_{t-i} + \varepsilon_t \quad (27)$$

where $X_t = (r_{f,t}, r_{1,t})'$; α (2×1) 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 $iidN(0, \Sigma)$.

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.

4.2 MS-VECM with Regime Switching Risk Premium

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, but with minor modification so that the risk premium term, $L_{f,t}$, is also subject to the regime shifts with Markov process:

$$\Delta X_t = \mu(s_t) + \alpha(s_t)(\beta' X_{t-1} - L(s_t)) + \sum_{i=1}^{p-1} \Psi_i(s_t) \Delta X_{t-i} + \varepsilon_t \quad (28)$$

where $\varepsilon_t \sim N(0, \Omega(s_t))$; $L(s_t)$ is the risk premium term depending upon the state variables. Compared with the MS-VECM considered in Section 2, there is an additional regime dependent parameter $L(s_t)$ to estimate. Since $\beta' X_t - L(s_t)$ is $I(0)$, we can estimate $L(s_t)$ as a parameter in a linear regression $\beta' X_t = L(s_t) + u_t = LS_t + u_t$ where $u_t \sim iidN(0, \sigma^2)$, $L = (L^0, L^1)$, and $S_t = (1, s_t)'$. With the natural conjugate priors for $L \mid \sigma^2 \sim N(\underline{L}, \sigma^2 \underline{V}_L)$ and $\sigma^2 \sim IG(\underline{\sigma}^2, \underline{\nu}_\sigma)$, where IG denotes an inverted Gamma distribution, the posterior for L is analytically obtained as a t -distribution with the mean $E(L \mid Y) = (\underline{V}_L^{-1} + S' S)^{-1} (\underline{V}_L^{-1} \underline{L} + S' \widehat{L})$ where each t^{th} row of S is $(1, s_t)$; \widehat{L} is the OLS estimator, $(S' S)^{-1} S' X \beta$. To estimate β , reexpress the model (28) as

$$y_t^* = \mu(s_t) + \alpha(s_t) \beta' X_{t-1} + \sum_{i=1}^{p-1} \Psi_i(s_t) \Delta X_{t-i} + \varepsilon_t \quad (29)$$

where $y_t^* = \Delta X_t + \alpha(s_t) L(s_t)$. The cointegrating vectors, β , in the model (29) can be

estimated by applying Strachan and Inder's (2004) method to the model.

Then, with regression $Y = WB + E$ where $W = (I_0(Z\beta - \mathbf{1}L_0), I_1(Z\beta - \mathbf{1}L_1), X)$; $\mathbf{1}$ is $\tau \times 1$ matrix of 1; $Y = (y_1^*, \dots, y_\tau^*)'$; B , X , and E are defined as (3) so that the posteriors for B and Ω_i are the same as (19) and (18). With these posteriors we implement the Gibbs sampler as follows (when $m = 2$):

- Step 1: Set $j = 1$. Specify starting values for the parameters of the model, $p_{11}^{(0)}$, $p_{22}^{(0)}$ $B^{(0)} = (\alpha_0^{(0)}, \alpha_1^{(0)}, \mu_0^{(0)}, \mu_1^{(0)}, \Psi_{1,0}^{(0)}, \dots, \Psi_{p-1,0}^{(0)}, \Psi_{1,1}^{(0)}, \dots, \Psi_{p-1,1}^{(0)})$, $L_i^{(0)}$, $\beta^{(0)}$ and $\Omega_i^{(0)}$.
- Step 2: Generate $\tilde{S}_\tau^{(j)} = \{s_1^{(j)}, s_2^{(j)}, \dots, s_\tau^{(j)}\}'$ from $p(\tilde{S}_\tau | \Theta^{(j-1)}, Y)$ in (13), where $\Theta = \{B, L, \Omega_1, \Omega_2, \beta, p_{ik}\}$, using multi-move Gibbs sampling algorithm.
- Step 3: Generate the transition probabilities, $p_{ik}^{(j)}$ from $p(p_{ik} | \tilde{S}_\tau^{(j)})$ in (16).
- Step 4: Add $\alpha(s_t)^{(j-1)}L(s_t)^{(j-1)}$ to the both sides of the model such that $\Delta X_t + \alpha(s_t)^{(j-1)}L(s_t)^{(j-1)} = \mu(s_t)^{(j-1)} + \alpha(s_t)^{(j-1)}\beta'X_{t-1} + \sum_{i=1}^{p-1}\Psi_i(s_t)^{(j-1)}\Delta X_{t-i} + \varepsilon_t$, then generate $\beta^{(j)}$ from $p(\beta | \tilde{S}_\tau^{(j)}, Y)$ in (20) using Strachan and Inder's method with the Griddy Gibbs sampling algorithm.
- Step 5: To generate $L(s_t)^{(j)}$, set up the regression $\beta^{(j)}X_t = LS_t^{(j)} + u_t$ where $L = (L^0, L^1)$, $S_t^{(j)} = (1, s_t^{(j)})$, and $u_t \sim iidN(0, \sigma^2)$. Estimate L as $L^{(j)} = (\underline{V}_L^{-1} + S^{(j)'}S^{(j)})^{-1}(\underline{V}_L^{-1}\underline{L} + S^{(j)'}S^{(j)}\hat{L})$, where \hat{L} is the OLS estimator. Note that the risk premium of regime 0 is $L(s_t = 0) = L_0 = L^0$, and that of regime 1 is $L(s_t = 1) = L_1 = L^0 + L^1$.
- Step 6: With the draws of $L^{(j)}$, $\beta^{(j)}$ and $\tilde{S}_\tau^{(j)}$, construct $I_0^{(j)}$ and $I_1^{(j)}$ to obtain $W^{(j)} = (I_0^{(j)}(Z\beta^{(j)} - \mathbf{1}L_0), I_1^{(j)}(Z\beta^{(j)} - \mathbf{1}L_1), X)$, and then generate $B^{(j)}$ from $p(vect(B) | \beta^{(j)}, L^{(j)}, \Omega_i^{(j-1)}, \tilde{S}_\tau^{(j)}, Y)$ in (19).
- Step 7: Generate $\Omega_i^{(j)}$ from $p(\Omega_i | \beta^{(j)}, L^{(j)}, B^{(j)}, \tilde{S}_\tau^{(j)}, Y)$ in (18).
- Step 8: Set $j = j + 1$, and go to Step 2.

4.3 Estimation Results

We analyze U.S. term structure of interest rates using the MS-VECM described above. The data set is monthly 3-month Treasury bill and 10-year Treasury bond covering the period 1960:1 to 2006:1 with 552 observations, obtained from the Federal Reserve Bank of St. Louis. Figure 1 plots the data set and its spread.

We consider the following four models:

$$\begin{aligned} \mathcal{M}1: \quad \Delta X_t &= \mu + \alpha(\beta' X_{t-1} - L) + \sum_{i=1}^{p-1} \Psi_i \Delta X_{t-i} + \varepsilon_t \\ \mathcal{M}2: \quad \Delta X_t &= \mu(s_t) + \alpha(s_t)(\beta' X_{t-1} - L(s_t)) + \sum_{i=1}^{p-1} \Psi_i(s_t) \Delta X_{t-i} + \varepsilon_t \\ \mathcal{M}3: \quad \Delta X_t &= \mu + \alpha(s_t)(\beta' X_{t-1} - L(s_t)) + \sum_{i=1}^{p-1} \Psi_i \Delta X_{t-i} + \sigma(s_t) \varepsilon_t \\ \mathcal{M}4: \quad \Delta X_t &= \mu(s_t) + \alpha(s_t)(\beta' X_{t-1} - L(s_t)) + \sum_{i=1}^{p-1} \Psi_i(s_t) \Delta X_{t-i} + \sigma(s_t) \varepsilon_t \end{aligned}$$

where $\varepsilon_t \sim iidN(0, \Sigma)$. $\mathcal{M}1$ represents a linear VECM. $\mathcal{M}2$ is a homoskedastic MS-VECM with regime dependent mean. $\mathcal{M}3$ is a heteroskedastic MS-VECM with constant mean. $\mathcal{M}4$ is a heteroskedastic MS-VECM with regime dependent mean.

To estimate four models, $\mathcal{M}1 - \mathcal{M}4$, we implement the Gibbs sampling algorithm described Section 4.2, with prior hyperparameters $\underline{L} = 0$, $\underline{V}_L = 1000$, $\underline{\sigma}^2 = 0.5$, $\underline{v}_\sigma = 0.001$ for ensuring a relatively noninformative for L . For other prior hyperparameters we set the same values as in the Monte Carlo simulation in Section 3. The Gibbs sampler is run with 10,000 times with the first 1,000 discarded.²

Testing for cointegration rank with the lag length $p = 4$ is conducted using the Savage-Dickey density ratio described in Section 3. The results are reported in Table 2. We find that there are very strong evidence of rank 1 for all four models $\mathcal{M}1 -$

² We did not check whether the draws from the Gibbs sampler converge by calculating, for example, Geweke's (1992) convergence diagnostics because there involve too many parameters to check. We believe that the number of iteration as 10,000 is generally sufficiently large number of draws to converge, and the first 1,000 discarded is enough to eliminate the effect of initial draw.

$\mathcal{M}4$ with the $p = 4$.³ Table 3 reports the posterior model probabilities calculated by using the Schwarz BIC for $\mathcal{M}1 - \mathcal{M}4$, varying the lag length $p = 2 - 5$. Clearly, nonlinearity by the Markov process is detected with almost 100 percent. The posterior model probability for $\mathcal{M}4$ is $Pr(\mathcal{M}4 | Y) = \sum_{p=2}^5 Pr(\mathcal{M}4, p | Y) \approx 0.934$, and thus there is strong evidence to support $\mathcal{M}4$. The highest posterior model probability is 44.1 percent given to $\mathcal{M}4$ with $p = 4$.

Table 4 reports the results of the posterior estimation of the parameters for $\mathcal{M}4$ with $p = 4$. From the results, the 95% HPDI 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 way⁴, 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 restrictions of $\beta = (1, -1)$, and the value is 278.14, which shows a very strong evidence to support the expectations hypothesis.⁵

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 1972 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 0. In regime 0, 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 1, the short rate fluctuates much more than the long rate. The risk premium in regime 1, L_1 , is lower than in regime 0, L_0 , that implies long run inflation expectations decrease during high inflation period due to the central bank's anti-inflationary monetary policy by rising the short term interest rate.

³ We have also tested for the number of cointegration rank of $\mathcal{M}1 - \mathcal{M}4$ with different lag length at $p = 2, 3$, and 5, and the results are all the same as when $p = 4$, that is, there is very strong evidence of rank 1.

⁴ 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."

⁵ 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 .

We find that the adjustment term for the long-term rate is positive as $\alpha_{i,0} < 0$ and for the short-term rate negative as $\alpha_{i,1} > 0$ for both regimes, which implies that the long-term rate tends to fall and the short-term rate tends to rise against the disequilibrium in either regime. The adjustment terms for both the long and the short-term rate in both regimes are significant as the 95% HPDIs for those values do not contain zero. Figure 3 plots the posterior densities for these adjustment terms and indicates that these densities do not contain zero. In regime 1 of higher volatility in the interest rates, the posterior mean of the adjustment speed for both the short ($\alpha_{1,0}$) and the long rates ($\alpha_{1,1}$) is $(-0.0800, 0.0625)$ which is much faster than those in the regime 0 ($\alpha_{0,0}, \alpha_{0,1}$) = $(-0.0115, 0.0135)$. This implies that interest rates adjust much faster in periods of high volatility with high inflation and anti-inflationary monetary policy.

5 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.⁶

Estimations are carried out entirely by a Bayesian method. The cointegrating vector is drawn using the method by Strachan and Inder (2004) 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.

To select the most appropriate model among linear, Markov switching, and other model specifications, we use the posterior model probabilities by approximating the Bayes factors by the Schwarz BIC. Although the Schwarz BIC does not generate the exact value of Bayes factor but just approximation, the Monte Carlo simulation show that it selects generally a correct model. To determine the number of cointegrating

⁶ 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.

rank, we employ the Savage-Dickey density ratio to calculate the Bayes factors for zero rank against non-zero rank.

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.

Appendix

A.1 Derivation of (18) and (19)

The joint prior of B , β , and Ω_i is given by multiplication of (4), (7) and (8) as follows:

$$\begin{aligned}
p\left(\text{vec}(B), \beta, \Omega_1, \dots, \Omega_m, \tilde{S}_T\right) &= g(\beta) p(\text{vec}(B)) p(\tilde{S}_T) \prod_{i=1}^m p(\Omega_i) \\
&\propto g(\beta) \left(\prod_{i=1}^m |\Phi_i|^{h_i/2} |\Omega_i|^{-(h_i+n+1)/2} \right) |\Sigma_B|^{-1/2} \exp \left\{ -\frac{1}{2} \left[\text{tr} \left(\sum_{i=1}^m \Omega_i^{-1} \Phi_i \right) \right. \right. \\
&\quad \left. \left. + \text{vec}(B - B_0)' \Sigma_B^{-1} \text{vec}(B - B_0) \right] \right\} \quad (30)
\end{aligned}$$

The likelihood function for $B, \Omega_1, \dots, \Omega_m, \beta$ and \tilde{S}_T is given by,

$$\mathcal{L}\left(B, \beta, \Omega_1, \dots, \Omega_m, \tilde{S}_T \mid Y\right)$$

$$\propto \left(\prod_{i=1}^m |\Omega_i|^{-t_i/2} \right) \exp \left(-\frac{1}{2} \text{tr} \left\{ \sum_{i=1}^m [\Omega_i^{-1} (Y_i - W_i B)' (Y_i - W_i B)] \right\} \right) \quad (31)$$

$$= \left(\prod_{i=1}^m |\Omega_i|^{-t_i/2} \right) \exp \left\{ -\frac{1}{2} \sum_{i=1}^m \left[\text{vec}(Y_i - W_i B)' (\Omega_i \otimes I_\tau)^{-1} \text{vec}(Y_i - W_i B) \right] \right\} \quad (32)$$

The joint posterior for deriving Ω_i is given as product of the joint prior (30) and the likelihood (31) as

$$\begin{aligned} & p(\text{vec}(B), \beta, \Omega_1, \dots, \Omega_m, \tilde{S}_T | Y) \\ & \propto p(\text{vec}(B), \beta, \Omega_1, \dots, \Omega_m, \tilde{S}_T) \mathcal{L}(B, \beta, \Omega_1, \dots, \Omega_m, \tilde{S}_T | Y) \\ & \propto g(\beta) \left(\prod_{i=1}^m |\Phi_i|^{h_i/2} |\Omega_i|^{-(t_i+h_i+n+1)/2} \right) |\Sigma_B|^{-1/2} \exp \left\{ -\frac{1}{2} [\text{vec}(B - B_0)' \Sigma_B^{-1} \text{vec}(B - B_0)] \right\} \\ & \times \exp \left(-\frac{1}{2} \left\{ \sum_{i=1}^m \Omega_i^{-1} [(Y_i - W_i B)' (Y_i - W_i B) + \Phi_i] \right\} \right) \end{aligned} \quad (33)$$

From the joint posterior (33), the conditional posterior density for Ω_i can be derived as

$$\begin{aligned} p(\Omega_i | B, \beta, \tilde{S}_T, Y) &= \frac{p(\text{vec}(B), \beta, \Omega_i, \tilde{S}_T | Y)}{p(\text{vec}(B), \beta, \tilde{S}_T | Y)} \propto p(\text{vec}(B), \beta, \Omega_i, \tilde{S}_T | Y) \\ &\propto |\Omega_i|^{-(t_i+h_i+n+1)/2} \exp \left(-\frac{1}{2} \text{tr} \left\{ \Omega_i^{-1} [(Y_i - W_i B)' (Y_i - W_i B) + \Phi_i] \right\} \right) \\ &= |\Omega_i|^{-(t_i+h_i+n+1)/2} \exp \left[-\frac{1}{2} \text{tr} (\Omega_i^{-1} \Phi_{*,i}) \right] \end{aligned} \quad (34)$$

where $\Phi_{*,i} = (Y_i - W_i B)' (Y_i - W_i B) + \Phi_i$. Thus, the conditional posterior of Ω_i is derived as an inverted Wishart distribution as

$$\Omega_i | \beta, B, \tilde{S}_T, Y \sim IW \left((Y_i - W_i B)' (Y_i - W_i B) + \Phi_i, t_i + h_i \right). \quad (35)$$

With regard to the conditional posterior density for $\text{vec}(B)$, we use the likelihood (32), instead of (31), to obtain the joint posterior as multiplying the joint prior in (30)

by (32), we have

$$\begin{aligned}
& p(\text{vec}(B), \beta, \Omega_1, \dots, \Omega_m, \tilde{S}_T | Y) \\
& \propto p(\text{vec}(B), \beta, \Omega_1, \dots, \Omega_m, \tilde{S}_T) \mathcal{L}(B, \beta, \Omega_1, \dots, \Omega_m, \tilde{S}_T | Y) \\
& \propto g(\beta) \left(\prod_{i=1}^m |\Phi_i|^{h_i/2} |\Omega_i|^{-(t_i+h_i+n+1)/2} \right) |\Sigma_B|^{-1/2} \exp \left\{ -\frac{1}{2} [\text{vec}(B - B_0)' \Sigma_B^{-1} \text{vec}(B - B_0)] \right\} \\
& \times \exp \left\{ -\frac{1}{2} \sum_{i=1}^m [\text{vec}(Y_i - W_i B)' (\Omega_i \otimes I_{t_i})^{-1} \text{vec}(Y_i - W_i B)] \right\} \tag{36}
\end{aligned}$$

From (36), we can write the key term in the last two lines as

$$\begin{aligned}
& \sum_{i=1}^m [\text{vec}(Y_i - W_i B)' (\Omega_i \otimes I_{t_i})^{-1} \text{vec}(Y_i - W_i B)] + \text{vec}(B - B_0)' \Sigma_B^{-1} \text{vec}(B - B_0) \\
& = \text{vec}(B - B_\star)' M_\star^{-1} \text{vec}(B - B_\star) + Q
\end{aligned}$$

where

$$Q = \sum_{i=1}^m [\text{vec}(Y_i)' (\Omega_i \otimes I_{t_i})^{-1} \text{vec}(Y_i)] + \text{vec}(B_0)' \Sigma_B^{-1} \text{vec}(B_0) - \text{vec}(B_\star)' M_\star^{-1} \text{vec}(B_\star)$$

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

$$\text{vec}(B_\star) = M_\star \left\{ \Sigma_B^{-1} \text{vec}(B_0) + \sum_{i=1}^m [(\Omega_i \otimes I_{t_i})^{-1} \text{vec}(W_i' Y_i)] \right\}.$$

For the proof of this derivation, see Appendix of Sugita (2006). Hence, the conditional posterior density for $\text{vec}(B)$ is derived as a multivariate normal density as follows:

$$p(\text{vec}(B) | \Omega_1, \dots, \Omega_m, \beta, \tilde{S}_T, Y) \propto |\Sigma_B|^{-1/2} \exp \left\{ -\frac{1}{2} [\text{vec}(B - B_\star)' M_\star^{-1} \text{vec}(B - B_\star)] \right\} \tag{37}$$

Thus, the conditional posterior distributions for Ω_i and B are given as (18) and (19) respectively.

A.2 The Griddy-Gibbs Sampler

The Griddy-Gibbs sampler is proposed by Ritter and Tanner (1992). This sampler can be implemented when the conditional posterior density is unknown to the researcher. The advantage of using this sampler over the importance sampler or the Metropolis-Hastings algorithm is that researcher does not have to provide an approximation of the function, that is not easy task in many cases. The disadvantage is that this sampler demands more computing time. The procedure for implementing the Griddy-Gibbs sampler is as following:

1. Before we begin the chain, we must choose the range of the grid and the number of the grid. The range should be chosen so that the generated numbers are not truncated.
2. Let $\text{vec}(\beta)' = (\beta_1, \beta_2, \dots, \beta_m)$. With an arbitrary starting value (within the upper and the lower bound of the grid), compute $f(\beta_1 | \beta_2^i, \beta_3^i, \dots, \beta_m^i, Y)$, where i denotes the i -th loop, over the grid $(\beta_{1,1}, \beta_{1,2}, \dots, \beta_{1,U})$, where $\beta_{1,1}$ is the lower bound of the grid of β_1 , and $\beta_{1,U}$ is the upper bound of the grid of β_1 .
3. Compute the values $G = (0, \Phi_2, \Phi_3, \dots, \Phi_U)$ where

$$\Phi_j = \int_{\beta_{1,1}}^{\beta_{1,j}} f(\beta_1 | \beta_2^i, \beta_3^i, \dots, \beta_m^i, Y) d\beta_1$$

$$j = 2, \dots, U$$

4. Compute the normalized pdf values $G_\zeta = G_j / \Phi_U$ of $\zeta(\beta_1 | \beta_2^i, \beta_3^i, \dots, \beta_m^i, Y)$.
5. Draw the random numbers from the uniform density with the lower bound as zeros and the upper bound as Φ_U and invert cdf G by numerical interpolation to obtain a draw β_1^i from $\zeta(\beta_1 | \beta_2^i, \beta_3^i, \dots, \beta_m^i, Y)$.

6. Repeat steps 2-5 for β_2, \dots, β_m .

7. Set $i = i + 1$ (increment i by 1) and go to step 2.

Note that integration at the step 3 can be done by the deterministic approximation such as the Simpson's rule or the Trapezoidal rule. The Simpson's rule approximates the integration of $f(\theta)$, based on an interpolation, such as

$$\int_0^1 f(\theta) d\theta \approx \frac{1}{6} [f(0) + 4f(0.5) + f(1)].$$

Or the extended Simpson rule can be applied as, with $2n$ intervals of equal length $d = \theta_i - \theta_{i-1} = 1/2n$ based on $2n + 1$ points $\theta_0 = 0, \theta_1, \dots, \theta_{2n} = 1$,

$$\int_0^1 f(\theta) d\theta \approx \frac{d}{3} \{f(\theta_0) + 4[f(\theta_1) + f(\theta_3) + \dots + f(\theta_{2n-1})] + 2[f(\theta_2) + f(\theta_4) + \dots + f(\theta_{2n})]\}.$$

See Bauwens et al (p68-71,1999) for details and other version of the Simpson rule.

References

- [1] 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.
- [2] Balke, N. S. and T. B. Fomby (1997). Threshold cointegration. *International Economic Review* 38, 627-645.
- [3] 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.
- [4] Bauwens, L. M. Lubrano and J-F, Richard (1999). *Bayesian Inference in Dynamic Econometric Models*, Oxford: Oxford University Press.
- [5] Campbell, J. Y. and R. J. Shiller (1987). Cointegration and tests of present value models. *Journal of Political Economy* 108, 205-51.
- [6] Carlin, B. P. and S. Chib (1995). Bayesian model choice via Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society Series B*, 57, 473-84.
- [7] Carter, C. K. and P. Kohn (1994). On Gibbs sampling for state space models. *Biometrika* 81, 541-553.
- [8] Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* 90, 432, Theory and Methods, 1313-1321.
- [9] Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician* 49, 327-335.
- [10] Chib, S. and I. Jeliazkov (2001). Marginal likelihood from the Metropolis Hastings output. *Journal of the American Statistical Association* 96, 270-81.

- [11] 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.
- [12] 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.
- [13] Doornik, J.A. (1998), *OX an Object-Oriented Matrix programming Language*. London: Timberlake Consultants Press.
- [14] Gelfand, A. and D. Dey (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society Series B* 56, 501-14.
- [15] 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*, 641-49. Oxford: Clarendon Press.
- [16] Goodfriend, M. (1998). Using the term structure of interest rates for monetary policy. *Economic Quarterly* 79, Federal Reserve Bank of Richmond, 1-23.
- [17] 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.
- [18] Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357-384.
- [19] Hansen, B. E. and B. Seo (2002). Testing for two-regime threshold cointegration in vector error-correction models. *Journal of Econometrics* 110, 293-318.
- [20] Johansen, S. (1988). Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control* 12, 231-54.
- [21] 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.

- [22] Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association* 90, 773-795.
- [23] 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.
- [24] Kim, C. J. and C. R. Nelson (2001). A Bayesian approach to testing for Markov switching in univariate and dynamic factor models. *International Economic Review* 42, 989-1013.
- [25] Kleibergen, F. and H. K. van Dijk (1998). Bayesian simultaneous equation analysis using reduced-rank regression. *Econometric Theory* 14, 701-43.
- [26] Kleibergen, F. and R. Paap (2002). Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration. *Journal of Econometrics* 111, 223-49.
- [27] Koop, G. (2004). *Bayesian Econometrics*. Wiley.
- [28] Koop, G. and S. M. Potter (1999). Bayes factors and nonlinearity: evidence from economic time series. *Journal of Econometrics* 88, 251-281.
- [29] Koop, G., R. Leon-Gonzalez and R. Strachan (2006a). Bayesian inference in a cointegrating panel data model. *Working Paper*, University of Leicester, UK.
- [30] Koop, G., R. Strachan and H. van Dijk (2006b). Bayesian approaches to cointegration. in *The Palgrave Handbook of Theoretical Econometrics* edited by K. Patterson and T. Mills. Palgrave Macmillan.
- [31] Krolzig, H. M. (1997). *Markov-switching vector autoregressions*. New York: Springer.
- [32] 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.

- [33] 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.
- [34] Psaradakis, Z., Sola, M., Spagnolo, F., (2004) forthcoming, On Markov Error-Correction Models, With an Application to Stock Prices and Dividends, *Journal of Applied Econometrics* 19, 69-88.
- [35] Ritter, C. and M. A. Tanner (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the griddy-Gibbs sampler. *Journal of the American Statistical Association* 87, 861-868.
- [36] 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, 345-76.
- [37] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6,461-64.
- [38] Shiller, R. J. (1990). The term structure of interest rate. in Friedman, B. and F. Hahn (eds). *Handbook of Monetary Economics I*, North-Holland: Amsterdam.
- [39] Strachan, R. W. (2003). Valid Bayesian estimation of the cointegrating error correction model. *Journal of Business and Economic Statistics* 21, 185-95.
- [40] Strachan, R. W. and H. K. van Dijk (2003a). Bayesian model selection with an uninformative prior. *Oxford Bulletin of Economics and Statistics* 65, 863-876.
- [41] Strachan, R. W. and H. K. van Dijk (2003b). The value of structural information in the VAR. *Econometric Institute Report EI 2003-17*. Erasmus University Rotterdam, Rotterdam.
- [42] Strachan, R. W. and B. Inder (2004). Bayesian analysis of the error correction model. *Journal of Econometrics* 123, 307-25.

- [43] Sugita, K. (2006). Bayesian Analysis of Dynamic Multivariate Models with Multiple Structural Breaks. *mimeo*.
- [44] Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* 93, 1188-1202.
- [45] 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.
- [46] Wang, J. and E. Zivot (2000). A Bayesian time series model of multiple structural changes in level, trend, and variance. *Journal of Business and Economic Statistics* 18, 374-386.

Table 1: Average posterior probabilities: Testing for Cointegration, Non-Cointegration and Markov Cointegration

True Model = $\mathcal{M}1$				True Model = $\mathcal{M}2$			
model	$T = 100$	$T = 200$	$T = 500$	model	$T = 100$	$T = 200$	$T = 500$
$\mathcal{M}1$	0.921	0.958	0.996	$\mathcal{M}1$	0.032	0.000	0.000
$\mathcal{M}2$	0.040	0.020	0.001	$\mathcal{M}2$	0.859	0.946	0.973
$\mathcal{M}3$	0.037	0.022	0.003	$\mathcal{M}3$	0.088	0.043	0.021
$\mathcal{M}4$	0.002	0.000	0.000	$\mathcal{M}4$	0.021	0.011	0.006

True model = $\mathcal{M}3$				True Model = $\mathcal{M}4$			
model	$T = 100$	$T = 200$	$T = 500$	model	$T = 100$	$T = 200$	$T = 500$
$\mathcal{M}1$	0.065	0.044	0.006	$\mathcal{M}1$	0.027	0.000	0.000
$\mathcal{M}2$	0.024	0.021	0.001	$\mathcal{M}2$	0.037	0.028	0.003
$\mathcal{M}3$	0.897	0.935	0.993	$\mathcal{M}3$	0.070	0.031	0.006
$\mathcal{M}4$	0.012	0.000	0.000	$\mathcal{M}4$	0.862	0.941	0.991

Note: Each element in Table shows the average posterior model probabilities calculated by using the Schwarz BIC.

Table 2: Cointegration Test for U.S. Term Structure of Interest Rates

	$\mathcal{M}1$	$\mathcal{M}2$	$\mathcal{M}3$	$\mathcal{M}4$
$r = 0$	0.023	0.001	0.000	0.000
$r = 1$	0.977	0.999	1.000	1.000
$r = 2$	0.000	0.000	0.000	0.000

Note: Each value in the Table shows the posterior model probability calculated by using the Savage-Dickey density ratio. Each model is with the lag length $p = 4$.

Table 3: Model Selection for U.S. Term Structure of Interest Rates with $r = 1$

	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$\Pr(\mathcal{M})$
$\mathcal{M}1$	0.000	0.000	0.000	0.000	0.000
$\mathcal{M}2$	0.026	0.002	0.000	0.000	0.028
$\mathcal{M}3$	0.000	0.005	0.025	0.008	0.038
$\mathcal{M}4$	0.255	0.188	0.441	0.050	0.934
$\Pr(p)$	0.281	0.195	0.466	0.058	1.000

Note: Each value in the Table shows the posterior model probability calculated by using the SBIC.

The bottom row of the Table shows the marginal probabilities for each lag length $\Pr(p) = \sum_{i=1}^4 \Pr(\mathcal{M}i, p)$.

The right column of the Table shows the marginal probabilities for each model $\Pr(\mathcal{M}) = \sum_{p=2}^5 \Pr(\mathcal{M}i, p)$.

Table 4: Posterior Results for $\mathcal{M}4$ with $p = 4$

() = standard deviation

parameter	mean	95% HPDI
β_2	-1.0331 (0.0618)	-1.1690, -0.9229
$\alpha_{0,0}$	-0.0115 (0.0031)	-0.0173, -0.0053
$\alpha_{0,1}$	0.0135 (0.0031)	0.0075, 0.0195
$\alpha_{1,0}$	-0.0800 (0.0111)	-0.0999, -0.0539
$\alpha_{1,1}$	0.0625 (0.0133)	0.0375, 0.0897
L_0	1.4310 (0.2953)	0.7755, 1.9582
L_1	0.6310 (0.5455)	-0.5606, 1.5853
p_{00}	0.9098 (0.0270)	0.8511, 0.9554
p_{11}	0.9696 (0.0097)	0.9482, 0.9860

$$\Sigma_0 = \begin{pmatrix} 0.0408 & 0.0169 \\ (0.0032) & (0.0017) \\ 0.0169 & 0.0270 \\ (0.0017) & (0.0026) \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 0.1670 & 0.2334 \\ (0.0240) & (0.0353) \\ 0.2334 & 0.6660 \\ (0.0353) & (0.0787) \end{pmatrix}.$$

Note: The subscript i in $\alpha_{i,j}$ denotes the regime, $j = 0$ for the long-term rate, and $j = 1$ for the short-term rate.

Figure 1: US 3-month bill rate, 10-year bond rate and the spread

Source: Federal Reserve Bank of St.Louis

Figure 2: Posterior expectation of the regime variable $E[S_t|Y]$ for th US Term
Structure of Interest rates

Figure 2: Histogram of posterior densities for α for US term structure of interest rates