Economics Bulletin

Volume 36, Issue 3

Bayesian inference in Markov switching vector error correction model

Katsuhiro Sugita
Faculty of Law and Letters, University of the Ryukyus

Abstract

In this paper we consider a Bayesian approach to a Markov switching vector error correction model that allows for regime shifts in the number of cointegrating rank, the cointegrating vectors, the adjustment terms, the deterministic terms, the lag terms and the variance-covariance matrix. We use a valid prior for the cointegrating space, and sample the state variable by employing the multi-move Gibbs sampler, and estimate the cointegrating vectors by a collapsed Gibbs sampler. We also drive the posterior densities for the model where cointegrating vectors are regime-independent.

Citation: Katsuhiro Sugita, (2016) "Bayesian inference in Markov switching vector error correction model", *Economics Bulletin*, Volume 36, Issue 3, pages 1534-1546
Contact: Katsuhiro Sugita - ksugita@ll.u-ryukyu.ac.jp.
Submitted: February 24, 2016. **Published:** August 11, 2016.

1 Introduction

In this paper we show a method for Bayesian inference in a Markov switching vector error correction model (MS-VECM) that allows for regime shifts in the number of cointegrating rank, the cointegrating vectors, the adjustment terms, the deterministic terms, the lag terms and the variance-covariance matrix. Jochmann and Koop (2015) study Bayesian approach to an MS-VECM, based on Chib (1998), using a valid prior for the cointegrating space (Strachan and van Dijk (2003), Strachan and Inder (2004), and Villani (2005, 2006)) and a collapsed Gibbs sampling method to estimate the cointegrating vectors (see Koop et al. (2009)). In this paper we use different prior densities from those used by Jochmann and Koop (2015) and employ the multi-move Gibbs sampling method to sample the state variable (see Carter and Kohn (1994) and Shephard (1994) and Kim and Nelson (1998)), instead of the method by Chib (1998). The multi-move Gibbs sampler We derive the posterior densities for the MS-VECM allowing for regime shifts in the number of cointegrating rank as well as any parameter including the adjustment terms and the cointegrating vectors. We also consider more restrictive model where the cointegrating vectors are not regime-dependent since changing the long-run relationships with regime shifts might not be reasonable in many economic applications. We illustrates an empirical application to U.S. term structure of interest rates in Section 3. All results reported in this paper are generated using Ox version 7.00 for Linux (see Doornik, 2013).

2 Markov Switching Vector Error Correction Model

2.1 MS-VECM

This section introduces an MS-VECM and presents a Bayesian approach to estimate the 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

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

where β $(n \times r)$ contains the cointegrating vectors, α $(r \times n)$ contains the adjustment terms, μ $(1 \times n)$ is the vector of intercept terms, and Γ_l $(n \times n)$ is the lag term. If we assume that r, β, α, μ, Γ_l's, and Ω in the VECM in (1) are subject to an unobservable discrete state variable s_t that evolves according to an 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 an MS-VECM is written as

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

where $\beta(s_t)$ is $n \times r(s_t)$, α is $r(s_t) \times n$, and $\varepsilon_t \sim N(0,\Omega(s_t))$.

The MS-VECM in (2) can be rewritten as

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

where z_t is $1 \times m(1+np)$ and Φ is $m(1+np) \times n$, and defined as

$z_t = (\iota_t(1), \ldots, \iota_t(m), \iota_t(1)\Delta y_{t-1}, \ldots, \iota_t(1)\Delta y_{t-p}, \ldots, \iota_t(m)\Delta y_{t-1}, \ldots, \iota_t(m)\Delta y_{t-p}),$

$\Phi = (\mu(1)', \ldots, \mu(m)', \Gamma_1(1)', \ldots, \Gamma_p(1)', \ldots, \Gamma_1(m)', \ldots, \Gamma_p(m)')',$

and $\iota_t(i)$ in z_t is an indicator variable that equals to 1 if regime is i at t, and 0 otherwise. From equation (3), let define

the $T \times n$ matrices $Y^T = (\ \Delta y_1', \ \ldots, \ \Delta y_T'\ )'$ and $E = (\ \varepsilon_1', \ \ldots, \ \varepsilon_T'\ )'$,

the $T \times n$ matrix $X_i = (\ y_0'\iota_0(i), \ y_1'\iota_1(i), \ \ldots, \ y_{T-1}'\iota_{T-1}(i)\ )'$,

the $T \times m(1+np)$ matrix $Z^T = (\ z_1', \ z_2', \ \ldots, \ z_T'\ )'$,

the $T \times h$ (where $h = m(1+np) + \sum_{i=1}^{m} r(i)$) matrix $W = (\ X_1\beta(1), \ldots, \ X_m\beta(m), \ Z^T\ )$,

the $h \times n$ matrix $B = (\ \alpha(1)', \ldots, \ \alpha(m)', \ \Phi'\ )'$,

then we can simplify the model as follows:

$$\begin{aligned} Y^T &= \sum_{i=1}^{m} X_i\beta(i)\alpha(i) + Z^T\Phi + E && (4) \\ &= WB + E. && (5) \end{aligned}$$

Thus, to estimate the MS-VCEM in (2), we consider the simplified form given in (5).

2.2 Priors

In this paper we adopt the collapsed Gibbs sampling method to generate the cointegrating vector, developed by Koop et al. (2009).

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 a positive definite matrix and $a = \kappa^{-1}\alpha$ is semi-orthogonal. Thus, the transformation for the MS-VECM is

$$\begin{aligned} \beta(s_t)\alpha(s_t) &= [\beta(s_t)\kappa(s_t)]\left[\kappa(s_t)^{-1}\alpha(s_t)\right] \\ &= \left[\beta(s_t)\left(\alpha(s_t)\alpha(s_t)'\right)^{1/2}\right]\left[\left(\alpha(s_t)\alpha(s_t)'\right)^{-1/2}\alpha(s_t)\right] \equiv b(s_t)a(s_t), \end{aligned}$$

where $\kappa(s_t) \equiv [\alpha(s_t)\alpha(s_t)']^{1/2}$, and thus $b(s_t) = \beta(s_t)\kappa(s_t)$. We assign a matric-variate normal distribution to the prior for $b(s_t)$ as $b(s_t) \sim MVN\left(b_0(s_t), V_{b_0}(s_t)\right)$, that is, if $\widetilde{b}(s_t) = vec(b(s_t))$, then $\widetilde{b}(s_t)$ is a multivariate normal with mean $\widetilde{b}_0(s_t) = vec(b_0(s_t)) \in \mathbb{R}^{nr(s_t)\times 1}$ and variance-covariance matrix $V_{b_0}(s_t) \in \mathbb{R}^{nr(s_t)\times nr(s_t)}$:

$$\widetilde{b}(s_t) \sim MN(\widetilde{b}_0(s_t), V_{b_0}(s_t)). \tag{6}$$

Next, we consider a prior for the transition probabilities q_{ij}, $i, j = 1, \ldots, m$. We assign a beta distribution for this prior as

$$q_{ij} \sim beta\left(u_{ij}, \bar{u}_{ij}\right), \tag{7}$$

where u_{ij} and $\bar{u}_{ij}$ are the known hyperparameters of the priors, $\bar{u}_{ii} = Pr\left(s_t \neq i \mid s_{t-1} = i\right)$ and $\bar{u}_{ij} = Pr\left(s_t = j \mid s_{t-1} = i, s_t \neq i\right)$ for $i \neq j$, $beta$ refers to a beta distribution with density $\pi(q_{ij} \mid u_{ij}, \bar{u_{ij}}) = \frac{\Gamma(u_{ij}+\bar{u}_{ij})}{\Gamma(u_{ii j})\Gamma(\bar{u}_{ij})} q_{ij}^{u_{ij}-1}(1-q_{ij})^{\bar{u}_{ij}-1}$.

With regard to prior for B in (5), we use an inverted Wishart prior for $\Omega(s_t)$ with a positive definite matrix $\Omega_0(s_t) \in \mathbb{R}^{n\times n}$ and the degrees of freedom $\nu_0(s_t)$:

$$\Omega(s_t) \sim IW\left(\Omega_0(s_t), \nu_0(s_t)\right). \tag{8}$$

As for a prior for B, we consider the vector form of B and assign a multivariate normal prior for $vec(B)$ with mean $vec(B_0) \in \mathbb{R}^{nh\times 1}$ and variance-covariance matrix $V_{B_0} \in \mathbb{R}^{nh\times nh}$ where $h = m(1+np)+\sum_{i=1}^{m} r(i)$:

$$vec(B) \sim MN\left(vec(B_0), V_{B_0}\right). \tag{9}$$

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 $\widetilde{S}_T = \{s_1, s_2, \ldots, s_T\}'$. To sample the state variable $\widetilde{S}_T$ we employ the multi-move Gibbs sampling method proposed by Carter and Kohn (1994) and Shephard (1994) and 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:

$$p\left(\widetilde{S}_T \mid \Theta, Y^T\right) = p\left(s_T \mid \Theta, Y^T\right) \prod_{t=1}^{T-1} p\left(s_t \mid s_{t+1}, \Theta, Y^t\right), \tag{10}$$

where $\Theta = (B, \mathbf{b}, \Omega, \mathbf{q})$, $\mathbf{b} = (b(1), \ldots, b(m))$, $\Omega = (\Omega(1), \ldots, \Omega(m))$, and $\mathbf{q} = (q_{11}, q_{12}, \ldots, q_{mm})$. The first term of the right hand side of equation (10), $p\left(s_T \mid \Theta, Y^T\right)$, can be obtained from running the Hamilton filter (Hamilton, 1989). To draw s_t conditional on s_{t+1}, Θ and Y^t, we use the following results:

$$p\left(s_t \mid s_{t+1}, \Theta, Y^t\right) = \frac{p\left(s_{t+1} \mid s_t, \Theta, Y^t\right) p\left(s_t \mid \Theta, Y^t\right)}{p\left(s_{t+1} \mid \Theta, Y^t\right)} \propto p\left(s_{t+1} \mid s_t\right) p\left(s_t \mid \Theta, Y^t\right), \tag{11}$$

where $p\left(s_{t+1} \mid s_t\right)$ is the transition probability. The second mass function in (11), $p\left(s_t \mid \Theta, Y^t\right)$, can be obtained as follows. First, we determine $p(s_t|\Theta, Y^{t-1})$ by the prediction step:

$$p(s_t|\Theta, Y^{t-1}) = \sum_{i=1}^{m} p(s_t|s_{t-1} = i, \Theta) p(s_{t-1} = i|\Theta, Y^{t-1}), \tag{12}$$

then we determine $p(s_t|\Theta, Y^t)$ by the update step:

$$p(s_t|\Theta, Y^t) \propto p(s_t|\Theta, Y^{t-1}) f(y_t|\Theta, Y^{t-1}), \tag{13}$$

where $f(y_t|\Theta, Y^{t-1})$ is a density function of y_t conditional on Θ and Y^{t-1}. Repeat these steps for all $t = 1, \ldots, T$. Using equation (11) we compute:

$$p\left(s_t = 1 \mid s_{t+1}, \Theta, Y^t\right) = \frac{p\left(s_{t+1} \mid s_t = 1\right) p\left(s_t = 1 \mid \Theta, Y^t\right)}{\sum_{i=1}^{m} p\left(s_{t+1} \mid s_t = i\right) p\left(s_t = i \mid \Theta, Y^t\right)}. \tag{14}$$

Once above probability is computed, we draw a random number from a uniform distribution, and if the generated number is less than or equal to the value calculated by equation (14), we set $s_t = 1$. Otherwise, we generate another random number from the uniform distribution. Then, if the generated number is less than or equal to $p(s_t = 2|s_{t+1}, \Theta, Y^T, s_t \neq 1)$, then we set $s_t = 2$, and so on.

After drawing $\widetilde{S}_T$ by the multi-move Gibbs sampling, we follow Albert and Chib (1993) and Kim and Nelson (1998) in generating the transition probabilities q_{ij} by multiplying equation (7) by the likelihood function $q_{ij}^{m_{ij}}\left(1-\bar{q}_{ij}\right)^{\bar{m}_{ij}}$ where m_{ij}, $i,j=1,\ldots,m$, denotes the number of the transition from the regime i to j, that can be counted for given $\widetilde{S}_T$:

$$p\left(q_{ij}\mid\widetilde{S}_T\right)\propto q_{ij}^{u_{ij}+m_{ij}-1}\left(1-\bar{q}_{ij}\right)^{\bar{u}_{ij}+\bar{m}_{ij}-1}\quad . \tag{15}$$

Next, we consider deriving the conditional posterior densities for other parameters, B, Ω, and $\mathbf{b}$. From the joint posterior obtained from the joint prior multiplied by the likelihood, we have the following conditional posterior distributions (see Appendix for derivation of these posteriors):

$$\Omega(i)\mid b(i),B,\widetilde{S}_T,Y^T\sim IW\left((Y_i-W_iB)'(Y_i-W_iB)+\Omega_0(i),t_i+\nu_0(i)+n+1\right), \tag{16}$$

$$vec\,(B)\mid\mathbf{b},\Omega,\widetilde{S}_T,Y^T\sim MN\left(vec\,(B_1)\,,V_{B_1}\right), \tag{17}$$

where

$$V_{B_1}=\left\{V_{B_0}^{-1}+\sum_{i=1}^{m}\left[\Omega(i)^{-1}\otimes\left(W_i'W_i\right)\right]\right\}^{-1},$$

$$vec\,(B_1)=V_{B_1}\left\{V_{B_0}^{-1}vec\,(B_0)+\sum_{i=1}^{m}\left[(\Omega(i)\otimes I_h)^{-1}vec\left(W_i'Y_i\right)\right]\right\}.$$

To obtain the conditional posterior for $\tilde{b}(i)$, we rewrite the expression in equation (4)

$$\begin{aligned}Y_i-Z_i^T\Phi &= X_i\beta(i)\alpha(i)+E\\ &= X_ib(i)a(i)+E,\end{aligned} \tag{18}$$

where $Y_i=\mathscr{I}_iY\ (T\times n)$, $Z_i^T=\mathscr{I}_iZ^T\ (T\times m(1+np))$; $\mathscr{I}_i=\mathrm{diag}(\iota_1(i),\ldots,\iota_T(i))$ is the $T\times T$ diagonal matrix where $\iota_t(i)$ is an indicator variable that equals to 1 if regime is i at t and 0 otherwise; $a(i)$ and $b(i)$ are such that $a(i)=(\alpha(i)\alpha(i)')^{-\frac{1}{2}}\alpha(i)$ and $\beta(i)=b(i)\,[b(i)'b(i)]^{\frac{1}{2}}$. Then vectorize both side of equation (18) as

$$\begin{aligned}vec(Y_i-Z_i^T\Phi) &= vec\,[X_ib(i)a(i)]+vec(E)\\ &= \left(a(i)'\otimes X_i\right)vec(b(i))+vec(E),\end{aligned} \tag{19}$$

or

$$\widetilde{y}_i=A_i\widetilde{b}(i)+e, \tag{20}$$

where $\widetilde{y}_i=vec(Y_i-Z_i^T\Phi)$, $A_i=(a(i)'\otimes X_i)$, $\widetilde{b}(i)=vec(b(i))$, and $e=vec(E)$. With the prior for $\widetilde{b}(i)\sim MN(vec(b_0(i)),V_{\tilde{b}_0}(i))$, the conditional posterior distribution of $\widetilde{b}_i$ is obtained as

$$\widetilde{b}(i)\mid\Omega(i),vec(B),\widetilde{S}_T,Y^T\sim MN(\widetilde{b}_\star(i),V_{\tilde{b}_\star}(i)), \tag{21}$$

where

$$V_{\tilde{b}_\star}(i) = \left[V_{\tilde{b}_0}(i)^{-1} + \left\{\left(a(i)\Omega(i)^{-1}a(i)'\right) \otimes \left(X_i'X_i\right)\right\}\right]^{-1},$$

$$\widetilde{b}_\star(i) = V_{\tilde{b}\star}(i)\left[V_{\tilde{b}_0}(i)^{-1}vec(b_0(i)) + \left\{\left(a(i)\Omega(i)^{-1}\right) \otimes X_i'\right\}\widetilde{y}\right].$$

After drawing $\tilde{b}(i)$, we obtain $\beta(i)$ and $\alpha(i)$ through a collapsed Gibbs sampler.

So far, we have assumed that β is subject to regime shifts. However, changing the long-run relationships with regime shifts might not be reasonable in some cases. If we assume that β is not regime-dependent, then equation (4) can be written as $Y^T = \sum_{i=1}^m X_i\beta\alpha(i) + Z^T\Phi + E$, so that we can rewrite as

$$\begin{aligned} Y^T - Z^T\Phi &= \sum_{i=1}^m X_i\beta\alpha(i) + E \\ &= \sum_{i=1}^m X_i b a(i) + E. \end{aligned} \tag{22}$$

Then vectorize both side of equation (22) as

$$\begin{aligned} vec(Y^T - Z^T\Phi) &= \sum_{i=1}^m vec(X_i b a(i)) + vec(E) \\ &= \sum_{i=1}^m \left(a(i)' \otimes X_i\right) vec(b) + vec(E), \end{aligned} \tag{23}$$

or

$$\widetilde{y} = A\widetilde{b} + e, \tag{24}$$

where $\widetilde{y} = vec(Y^T - Z^T\Phi)$, $A = \sum_{i=1}^m \left(a(i)' \otimes X_i\right)$, $\widetilde{b} = vec(b)$, and $e = vec(E)$. With the prior for $\widetilde{b} \sim MN(vec(b_0), V_{\tilde{b}_0})$, the conditional posterior distribution of $\widetilde{b}_i$ is given by

$$\widetilde{b} \mid \Omega(1), \ldots, \Omega(m), vec(B), \widetilde{S}_T, Y^T \sim MN(\widetilde{b}_\star, V_{\tilde{b}_\star}), \tag{25}$$

where

$$V_{\tilde{b}_\star} = \left[V_{\tilde{b}_0}^{-1} + \sum_{i=1}^m \left\{\left(a(i)\Omega(i)^{-1}a(i)'\right) \otimes \left(X_i'X_i\right)\right\}\right]^{-1},$$

$$\widetilde{b}_\star = V_{\tilde{b}_\star}\left(V_{\tilde{b}_0}^{-1}vec(b_0) + \sum_{i=1}^m \left\{\left(a(i)\Omega(i)^{-1}\right) \otimes X_i'\right\}\widetilde{y}\right).$$

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

- Step 1: Set $g = 1$ where g denotes the number of iteration. Specify starting values for the parameters in the model, $\Theta^{(0)} = \left(B^{(0)}, \mathbf{b}^{(0)}, \Omega^{(0)}, \mathbf{q}^{(0)}\right)$.

- Step 2: Generate $\widetilde{S}_T^{(g)} = \left\{s_1^{(g)}, s_2^{(g)}, \ldots, s_T^{(g)}\right\}'$ from $p\left(\widetilde{S}_T \mid \Theta^{(g-1)}, Y^T\right)$ in (10), using the multi-move Gibbs sampling algorithm.

- Step 3: Generate the transition probabilities $(q_{ij})^{(g)}$ from $p\left(q_{ij} \mid \widetilde{S}_T^{(g)}\right)$ in (15) for $i, j = 1, \ldots, m$.
- Step 4: Generate $vec(B)^{(g)}$ from $p(vec(B) \mid \mathbf{b}^{(g-1)}, \Omega^{(g-1)}, \widetilde{S}_T^{(g)}, Y^T)$ in (17), then obtain $\alpha(i)^*$ and Φ from $vec(B)^{(g)}$. Compute $a(i)^* = (\alpha(i)^* \alpha(i)^{*\prime})^{-1/2} \alpha(i)^*$ for $i = 1, \ldots, m$.
- Step 5: Generate $b^*(i)$ from $p(\widetilde{b}(i) \mid \Omega(i)^{(g-1)}, vec(B)^{(g)}, \widetilde{S}_T^{(g)}, Y^T)$ in (21) for the model with regime-dependent β, then compute $\beta(i)^{(g)} = b(i)^*(b(i)^{*\prime} b(i)^*)^{-\frac{1}{2}}$ and $\alpha(i)^{(g)} = (b(i)^{*\prime} b(i)^*)^{\frac{1}{2}} a(i)^*$ for $i = 1, \ldots, m$. Or, for the model with constant β, generate b^* from $p(\widetilde{b} \mid \Omega^{(g-1)}, vec(B)^{(g)}, \widetilde{S}_T^{(g)}, Y^T)$ in (25), then compute $\beta^{(g)} = b^*(b^{*\prime} b^*)^{-\frac{1}{2}}$ and $\alpha(i)^{(g)} = (b^{*\prime} b^*)^{\frac{1}{2}} a(i)^*$ for $i = 1, \ldots, m$.
- Step 6: Generate $\Omega(i)^{(g)}$ from $p(\Omega(i) \mid b(i)^{(g)}, B^{(g)}, \widetilde{S}_T^{(g)}, Y^T)$ in (16) for $i = 1, \ldots, m$.
- Step 7: Set $g = g + 1$, and go to Step 2.

Step 2 through Step 7 can be iterated G times to obtain the posterior means or standard deviations. Note that the first G_0 times iterations are discarded in order to attenuate the effect of the initial values.

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. The expectations hypothesis of the term structure of interest rates implies a long-run relationship between long and short term interest rates. For an overview of the expectations hypothesis theory, see Shiller and McCulloch (1990). Let $R_t(f)$ be the yield to maturity for an f-period at time t, then the hypothesis implies that $R_t(f)$ and $R_t(1)$ are cointegrated with cointegrating vector $(1, -1)$, see Campbell and Shiller (1987). Thus, the expectations hypothesis implies the following vector error correction model with the lag length at p:

$$\Delta y_t = \mu + y_{t-1} \beta \alpha + \sum_{l=1}^{p} \Delta y_{t-l} \Gamma_l + \varepsilon_t, \tag{26}$$

where $y_t = (\ R_t(f), \quad R_t(1)\ )$ and $\varepsilon_t \sim iidN(0, \Omega)$.

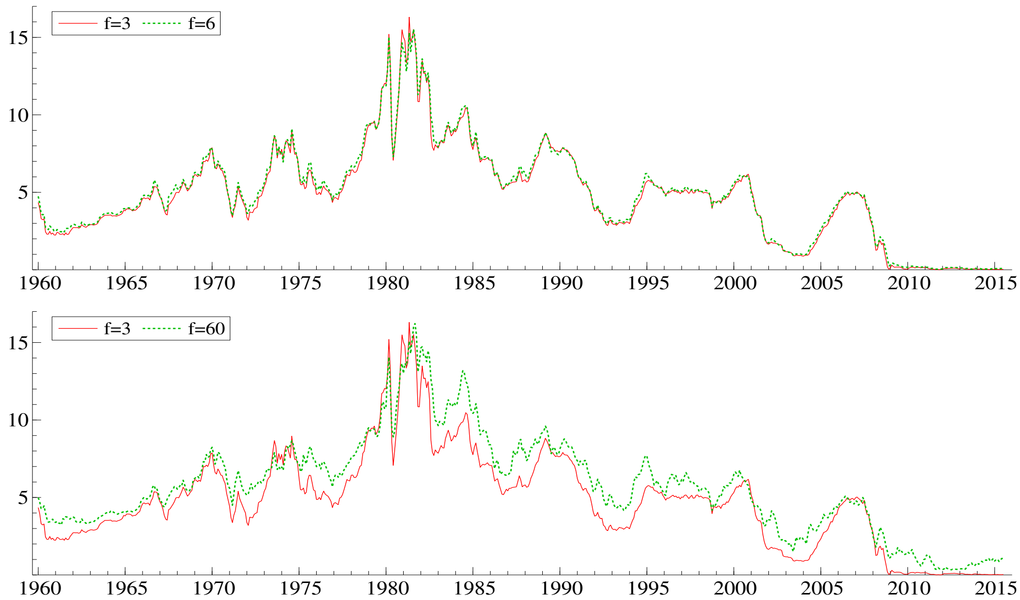
There is a number of research that confirms nonlinearity of U.S. term structure of interest rates due to changes in monetary policy. Tsay (1998), Hansen and Seo (2002), Clements and Galvao (2003) use a threshold cointegration model, while Clarida et al. (2006) and Tillmann (2007) employ a Markov switching 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.

We apply the MS-VECM to U.S. term structure of interest rate to account for the regime shifts as:

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

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

Figure 1: Interest rates (in % p.a.) on US bonds of maturity f (in months)



Source: Federal Reserve Bank of St.Louis

The data set is monthly three-month ($f = 3$), six-month ($f = 6$), and five-year ($f = 60$) US bonds covering the period 1960:1 to 2016:01 with 672 observations. These data are obtained from the Federal Reserve Bank of St. Louis. A pairwise plot of the series, using three-month ($f = 3$) rate as a short rate, is presented in Figure 1.

We consider bivariate two-regime MS-VECMs for the term structure because a parsimonious model is highly appreciated as Tillmann (2007) notes. Thus, we model bivariate MS-VECMs for two pairs of interest rates - one is for the $y_t = [R_t(6), R_t(3)]$ pair of interest rates and the other is for the $y_t = [R_t(60), R_t(3)]$ pair of interest rates. For each pair of the interest rates, we consider models with the cointegrating rank $r = 0$ or 1 for each regime, that is, $(r_1, r_2) =$ (0, 0), (0, 1), ..., (1, 1), where r_i denotes the number of rank in regime i. We also consider a model with $(r_1, r_2) = (1, 1)$ where β is unaffected by regime shifts as $\beta(s_t) = \beta$. For the lag length we consider the case $p = 1$ to 4. Thus, we consider a total of 20 bivariate models for each pair of the interest rates to select the most appropriate model among them. We choose Chib (1995)'s method for the marginal likelihood calculation to obtain the posterior model probabilities for model selection. Chib (1996), Chib (1998), and Kim and Nelson (1999) use Chib's method for a Markov switching model. To estimate these models and obtain the posterior model probabilities we employ the multi-move Gibbs sampling method to sample the state variable, and implement the collapsed Gibbs sampling algorithm to sample the cointegrating vectors described in Section 2.3. For prior hyperparameters, we set $\tilde{b}_0 = \tilde{b}_0(i) = (1, 0)'$, $V_{\tilde{b}_0} = V_{\tilde{b}_0}(i) = I_2$ for $i = 1$ or 2 in (6), $q_{11} \sim beta(u_{11}, u_{12}) = beta(9, 1)$ and $q_{22} \sim beta(u_{22}, u_{21}) = beta(9, 1)$ in (7), $\Omega_0(i) = I_2$ and $\nu_0(i) = 10$ for $i = 1$ or 2 in (8), $V_B = 10I_{\kappa n}$and $B_0 = 0$ in (9) favoring the absence of cointegration. These values are assigned to ensure fairly large variances for representing prior ignorance. The full Gibbs sampler is run with $G = 100,000$. In Table 1 we presents the posterior model probabilities obtained from the Bayes factors for all 20 models for each pair of the interest rates, the $[R_t(6), R_t(3)]$ and the $[R_t(60), R_t(3)]$. The highest posterior model probability is given to the model with $(r_1, r_2) = (0, 1)$ and $p = 1$ for the $[R_t(6), R_t(3)]$ pair, and $(r_1, r_2) = (1, 1)$ with constant β across regimes and $p = 3$ for the $[R_t(60), R_t(3)]$ pair.

Table 1: Posterior model probabilities: model selection

	$[R_t(6), R_t(3)]$					$[R_t(60), R_t(3)]$				
rank(1)	0	0	1	1	1	0	0	1	1	1
rank(2)	0	1	0	1	1	0	1	0	1	1
lag p					$(\beta(i) = \beta)$					$(\beta(i) = \beta)$
1	0.000	0.832	0.000	0.023	0.105	0.000	0.000	0.000	0.000	0.090
2	0.000	0.038	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.034
3	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.059	0.804
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 2: Posterior parameter estimates

	$[R_t(6), R_t(3)]$ $(r_1, r_2) = (0, 1)$	$[R_t(60), R_t(3)]$ $(r_1, r_2) = (1, 1)$ β is constant
β	0.7074 (0.0019)	0.7299 (0.0254)
	-0.7068 (0.0019)	-0.6835 (0.0273)
$\alpha(1)$	————	-0.0389 (0.0184)
	————	0.0177 (0.0285)
$\alpha(2)$	0.2272 (0.0941)	-0.0333 (0.0183)
	0.4747 (0.0842)	0.0369 (0.0130)
$\mu(1)$	-0.0417 (0.0576)	0.1083 (0.0781)
	-0.0519 (0.0616)	-0.0780 (0.1054)
$\mu(2)$	-0.0177 (0.0113)	0.0302 (0.0239)
	-0.0368 (0.0117)	-0.0324 (0.0161)
p_{11}	0.8808 (0.0346)	0.8507 (0.0380)
p_{22}	0.9597 (0.0129)	0.9537 (0.0125)
$\hat{\Omega}(1)$	0.5274 (0.0662)	0.2403 (0.0309)
	0.6045 (0.0727)	0.5930 (0.0625)
$\hat{\Omega}(2)$	0.0254 (0.0024)	0.0431 (0.0033)
	0.0213 (0.0017)	0.0201 (0.0017)

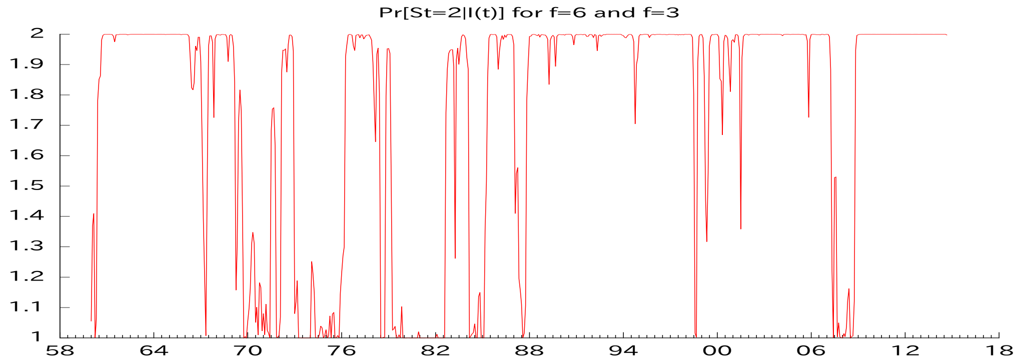
Note: Standard errors are in parentheses. The diagonal elements (the variances) of the regime-dependent variance-covariance matrices are given by $\hat{\Omega}(s_t)$.

We find that the equilibrium correction does not occur in regime 1 for the $[R_t(6), R_t(3)]$ pair, while the constant cointegrated relation can be found in both regimes for the $[R_t(60), R_t(3)]$ pair. These highest posterior model probabilities are dominant among other probabilities so that choosing one model for each pair is appropriate, otherwise we could consider a Bayesian model averaging as Peters et al. (2010).

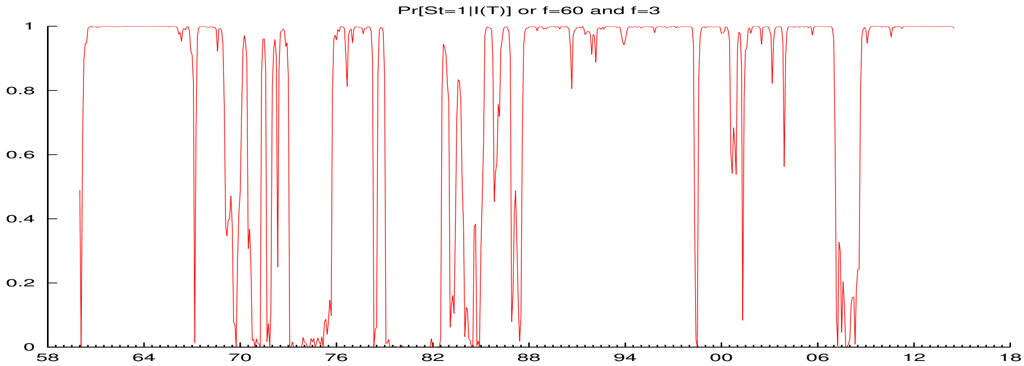
In Figure 2 we show the posterior expectation of the state variables for each interest rate pair, and find that both are almost identical. The *non-borrowed reserves operating procedure* between 1979 and 1982 is detected as the regime shift. Regime shift occurs also between 1973 and 1976, and between 1984 and 1985. 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 2.

The posterior means and standard deviations of parameters for each interest rate pair are reported in Table 2. We find that regime 1 is characterized by a much higher variance of both the long and the short rate than regime 2. For both pairs of the interest rates, the $\alpha(2)$

Figure 2: Posterior expectation of the regime variable



(a)



coefficients for short rate in regime 2 are significantly different from zero with positive sign. Thus, in the low volatility period the adjustment toward equilibrium occur mainly through the short rate $R_t(3)$.

4 Conclusion

In this paper we consider Bayesian inference in the Markov switching vector error correction model which allows any set of parameter in the model, including the number of cointegrating rank, to shift with the regime according to the first order unobservable Markov process. We employ the multi-move Gibbs sampling method to sample the state variable and derive the posterior densities by using different prior densities from those used by Jochmann and Koop (2015). We also consider the model where the cointegrating vectors are not regime-dependent.

Appendix

Derivation of the conditional posterior density for $\Omega(i)$ in (16) and $vec(B)$ in (17)

The joint prior of $vec(B)$, $\mathbf{b}$, and Ω is given by multiplication of equations (6), (8) and (9) as follows:

$$p\left(vec(B),\mathbf{b},\Omega\right) = p(vec(B))\prod_{i=1}^{m}\left(p(\tilde{b}(i))p(\Omega(i))\right)$$
$$\propto \left(\prod_{i=1}^{m} p(\tilde{b}(i))\left|\Omega_0(i)\right|^{\nu_0(i)/2}\left|\Omega(i)\right|^{-(\nu_0(i)+n+1)/2}\right)\left|\Sigma_B\right|^{-1/2}\exp\left\{-\frac{1}{2}\left[\mathrm{tr}\left(\sum_{i=1}^{m}\Omega(i)^{-1}\Omega_0(i)\right)\right.\right.$$
$$\left.\left.+vec\left(B-B_0\right)'\Sigma_B^{-1}vec\left(B-B_0\right)\right]\right\}. \tag{28}$$

The likelihood function is given by,

$$\mathfrak{L}\left(Y^T \mid vec(B),\mathbf{b},\Omega,\widetilde{S}_T\right) \tag{29}$$
$$\propto \left(\prod_{i=1}^{m}\left|\Omega(i)\right|^{-t_i/2}\right)\exp\left(-\frac{1}{2}\mathrm{tr}\left\{\sum_{i=1}^{m}\left[\Omega(i)^{-1}\left(Y_i-W_iB\right)'\left(Y_i-W_iB\right)\right]\right\}\right) \tag{30}$$
$$= \left(\prod_{i=1}^{m}\left|\Omega(i)\right|^{-t_i/2}\right)\exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\left[vec\left(Y_i-W_iB\right)'\left(\Omega(i)\otimes I_T\right)^{-1}vec\left(Y_i-W_iB\right)\right]\right\}, \tag{31}$$

where t_i is the total number of observations when $s_t = i$, $i = 1,\ldots,m$, $Y_i = \mathscr{I}_iY$ $(T\times n)$, $W_i = \mathscr{I}_iW$ $(T\times h)$, $\mathscr{I}_i = \mathrm{diag}(\iota_1(i),\ldots,\iota_T(i))$ is a $T\times T$ diagonal matrix where $\iota_t(i)$ is an indicator variable that equals to 1 if regime is i at t, and 0 otherwise.

The joint posterior for deriving $\Omega(i)$ is given as product of the joint prior in (28) and the likelihood (30) as

$$p(vec(B),\mathbf{b},\Omega \mid \widetilde{S}_T,Y^T) \propto p\left(vec(B),\mathbf{b},\Omega\right)\mathfrak{L}\left(Y^T \mid vec(B),\mathbf{b},\Omega,\widetilde{S}_T\right)$$
$$\propto \left(\prod_{i=1}^{m} p(\tilde{b}(i))\left|\Omega_0(i)\right|^{\nu_0(i)/2}\left|\Omega(i)\right|^{-(t_i+\nu_0(i)+n+1)/2}\right)\left|\Sigma_B\right|^{-1/2}\exp\left\{-\frac{1}{2}\left[vec(B-B_0)'V_B^{-1}vec(B-B_0)\right]\right\}$$
$$\times\exp\left(-\frac{1}{2}\left\{\sum_{i=1}^{m}\Omega(i)^{-1}\left[\left(Y_i-W_iB\right)'\left(Y_i-W_iB\right)+\Omega_\star(i)\right]\right\}\right). \tag{32}$$

From the joint posterior (32), the conditional posterior density for $\Omega(i)$ can be derived as

$$\begin{aligned}
p(\Omega(i) \mid B,b(i),\widetilde{S}_T,Y^T) &= \frac{p(B,b(i),\Omega(i),\widetilde{S}_T \mid Y^T)}{p(B,b(i),\widetilde{S}_T \mid Y^T)} \propto p(B,b(i),\Omega(i),\widetilde{S}_T \mid Y^T)\\
&\propto \left|\Omega(i)\right|^{-(t_i+\nu_0(i)+n+1)/2}\exp\left(-\frac{1}{2}\mathrm{tr}\left\{\Omega(i)^{-1}\left[\left(Y_i-W_iB\right)'\left(Y_i-W_iB\right)+\Omega_\star(i)\right]\right\}\right)\\
&= \left|\Omega(i)\right|^{-(t_i+\nu_0(i)+n+1)/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\Omega(i)^{-1}\Omega_\star(i)\right)\right], \qquad (33)
\end{aligned}$$

where $\Omega_\star(i) = (Y_i - W_iB)'(Y_i - W_iB) + \Omega_0(i)$. Thus, the conditional posterior of $\Omega(i)$ is derived as an inverted Wishart distribution as

$$\Omega(i) \mid B, b(i), \widetilde{S}_T, Y^T \sim IW\left((Y_i - W_iB)'(Y_i - W_iB) + \Omega_0(i), t_i + \nu_0(i) + n + 1\right). \quad (34)$$

With regard to the conditional posterior density for $vec(B)$, we use the likelihood (31) to obtain the joint posterior as multiplying the joint prior in (28) by (31), we have

$$\begin{aligned} & p(vec(B), \mathbf{b}, \Omega \mid \widetilde{S}_T, Y^T) \propto p\,(vec(B), \mathbf{b}, \Omega)\,\mathfrak{L}\left(Y^T \mid vec(B), \mathbf{b}, \Omega, \widetilde{S}_T\right) \\ & \propto \left(\prod_{i=1}^{m} p(\tilde{b}(i))\,|\Omega_0(i)|^{\nu_0(i)/2}\,|\Omega(i)|^{-(t_i+\nu_0(i)+n+1)/2}\right)|\Sigma_B|^{-1/2}\exp\left\{-\frac{1}{2}\left[vec(B-B_0)'V_B^{-1}vec(B-B_0)\right]\right\} \\ & \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\left[vec(Y_i - W_iB)'(\Omega(i)\otimes I_T)^{-1}vec(Y_i - W_iB)\right]\right\}. \end{aligned} \quad (35)$$

From equation (35), we can write the key term in the last two lines as

$$\begin{aligned} & \sum_{i=1}^{m}\left[vec\,(Y_i - W_iB)'\,(\Omega(i)\otimes I_T)^{-1}\,vec\,(Y_i - W_iB)\right] + vec\,(B-B_0)'\,V_B^{-1}vec\,(B-B_0) \\ & = vec\,(B-B_1)'\,V_{B_1}^{-1}vec\,(B-B_1) + Q, \end{aligned}$$

where

$$Q = \sum_{i=1}^{m}\left[vec\,(Y_i)'\,(\Omega(i)\otimes I_T)^{-1}\,vec\,(Y_i)\right] + vec\,(B_0)'\,V_{B_0}^{-1}vec\,(B_0) - vec\,(B_\star)'\,V_{B_1}^{-1}vec\,(B_\star)$$

$$V_{B_1}^{-1} = \left\{V_{B_0}^{-1} + \sum_{i=1}^{m}\left[\Omega(i)^{-1}\otimes\left(W_i'W_i\right)\right]\right\}^{-1},$$

$$vec\,(B_1) = V_{B_1}^{-1}\left\{V_{B_0}^{-1}vec\,(B_0) + \sum_{i=1}^{m}\left[(\Omega(i)\otimes I_h)^{-1}\,vec\left(W_i'Y_i\right)\right]\right\}.$$

For the proof of this derivation, see Appendix of Sugita (2008). Hence, the conditional posterior density for $vec(B)$ is derived as a multivariate normal density as follows:

$$p(vec\,(B) \mid \mathbf{b}, \Omega, \widetilde{S}_T, Y^T) \propto |V_B|^{-1/2}\exp\left\{-\frac{1}{2}\left[vec\,(B-B_1)'\,V_{B_1}^{-1}vec\,(B-B_1)\right]\right\}. \quad (36)$$

Thus, the conditional posterior distributions for $\Omega(i)$ and B are given as equations (16) and (17) respectively.

References

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

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

Carter, C. K. and Kohn, R. (1994). "On Gibbs sampling for state space models." *Biometrika*, 81(3): 541–553.

Chib, S. (1995). "Marginal likelihood from the Gibbs output." *Journal of the American Statistical Association*, 90(432): 1313–1321.

— (1996). "Calculating posterior distributions and modal estimates in Markov mixture models." *Journal of Econometrics*, 75(1): 79–97.

— (1998). "Estimation and comparison of multiple change-point models." *Journal of Econometrics*, 86(2): 221–241.

Clarida, R. H., Sarno, L., Taylor, M. P., and Valente, G. (2006). "The Role of Asymmetries and Regime Shifts in the Term Structure of Interest Rates." *The Journal of Business*, 79(3): 1193–1224.

Clements, M. P. and Galvao, A. B. C. (2003). "Testing the expectations theory of the term structure of interest rates in threshold models." *Macroeconomic Dynamics*, 7(04): 567–585.

Doornik, J. (2013). *Object-Oriented Matrix Programming using Ox*. London: Timberlake Consultants Press.

Goodfriend, M. (1998). "Using the term structure of interest rates for monetary policy." *FRB Richmond Economic Quarterly*, 84(3): 13–30.

Hamilton, J. D. (1989). "A new approach to the economic analysis of nonstationary time series and the business cycle." *Econometrica: Journal of the Econometric Society*, 357–384.

Hansen, B. E. and Seo, B. (2002). "Testing for two-regime threshold cointegration in vector error-correction models." *Journal of Econometrics*, 110(2): 293–318.

Jochmann, M. and Koop, G. (2015). "Regime-switching cointegration." *Studies in Nonlinear Dynamics & Econometrics*, 19(1): 35–48.

Kim, C.-J. and Nelson, C. R. (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*, 80(2): 188–201.

— (1999). "Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle." *Review of Economics and Statistics*, 81(4): 608–616.

Koop, G., León-González, R., and Strachan, R. W. (2009). "Efficient posterior simulation for cointegrated models with priors on the cointegration space." *Econometric Reviews*, 29(2): 224–242.

Peters, G. W., Kannan, B., Lasscock, B., and Mellen, C. (2010). "Model selection and adaptive Markov chain Monte Carlo for Bayesian cointegrated VAR models." *Bayesian Analysis*, 5(3): 465–491.

Shephard, N. (1994). "Partial non-Gaussian state space." *Biometrika*, 81(1): 115–131.

Shiller, R. J. and McCulloch, J. H. (1990). "The term structure of interest rates." *Handbook of Monetary Economics*, 1: 627–722.

Strachan, R. W. and Inder, B. (2004). "Bayesian analysis of the error correction model." *Journal of Econometrics*, 123(2): 307–325.

Strachan, R. W. and van Dijk, H. (2003). "Bayesian Model Selection with an Uninformative Prior." *Oxford Bulletin of Economics and Statistics*, 65(s1): 863–876.

Sugita, K. (2008). "Bayesian analysis of a vector autoregressive model with multiple structural breaks." *Economics Bulletin*, 3(22): 1–7.

Tillmann, P. (2007). "Inflation regimes in the US term structure of interest rates." *Economic Modelling*, 24(2): 203–223.

Tsay, R. S. (1998). "Testing and modeling multivariate threshold models." *Journal of the American Statistical Association*, 93(443): 1188–1202.

Villani, M. (2005). "Bayesian reference analysis of cointegration." *Econometric Theory*, 21(02): 326–357.

— (2006). "Bayesian point estimation of the cointegration space." *Journal of Econometrics*, 134(2): 645–664.