Discuttion Paper #2006-14

Bayesian Analysis of Dynamic Multivariate Models with Multiple Structural Breaks

by

Katsuhiro Sugita

Bayesian Analysis of Dynamic Multivariate Models with Multiple Structural Breaks

Katsuhiro Sugita*

Graduate School of Economics, Hitotsubashi University, Tokyo 186-8601, Japan

Abstract

This paper considers a vector autoregressive model or a vector error correction model with multiple structural breaks in any subset of parameters, using a Bayesian approach with Markov chain Monte Carlo simulation technique. The number of structural breaks is determined as a sort of model selection by the posterior odds. For a cointegrated model, cointegrating rank is also allowed to change with breaks. Bayesian approach by Strachan (Journal of Business and Economic Statistics 21 (2003) 185) and Strachan and Inder (Journal of Econometrics 123 (2004) 307) are applied to estimate the cointegrating vectors. As empirical examples, we investigate structural changes in the predictive power of the yield curve and the US term structure of interest rates. We find strong evidence of three structural changes in both applications.

Key words: Bayesian inference; Structural break; Cointegration; Bayes factor;
JEL classification: C11; C12; C32

1 Introduction

The last decade has seen extensive study of the structural break in time series models. Papers such as Perron (1989) deals with this issue in the framework of a priori imposed break dates,

* Graduate School of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan. Tel.: +81-42-580-8525; fax: +81-42-580-8525. *E-mail Address*: ksugita@econ.hit-u.ac.jp

while others use methods where the break date is endogenized (Banerjee, Lumsdaine and Stock, 1992; Christiano, 1992; and Zivot and Andrews, 1992). Much of the subsequent research focus on testing for a structural break when the break date may not be known. Among these, the *supF* statistic of Andrews (1993) and the *expF* and *aveF* statistics of Andrews-Ploberger (1994) are most notable. Based on Andrews and Andrews-Ploberger's statistics, Hansen (2000) proposes a bootstrapping method for testing for a single structural break.

An extension of the literature on testing for a structural break involves allowing for more than one possible break date. For many macroeconomic or financial time series with the possibility of a structural break, the assumption of at most one break date is unrealistic and restrictive. Bai and Perron (1998) propose a test for multiple structural breaks at unknown dates using the double maximum test. Another testing method for detecting multiple changes is a likelihood ratio test with the null of l breaks against the alternative $l+1$ break points (Bai, 1999). While these methods only allow for structural breaks in mean, breaks in variance are often found in economic and financial data. Schwert (1990) finds that volatility of the stock-market is higher during and after the 1987 crash. Inclan (1993), Inclan and Tiao (1994), and Chen and Gupta (1997) detect multiple breaks in variance for several series of stock returns. Engel and Hakkio (1996) find that European Monetary System exchange rates have higher volatility during the periods of alignment, and Kim and Engel (1999) find multiple breaks in variance in real exchange rates associated with historically significant monetary events. Kim and Nelson (1999) combine a structural break with the Markov switching model to find evidence of variance breaks in postwar business cycles. For a Bayesian approach to multiple structural breaks, Wang and Zivot (2000) consider univariate models with multiple breaks in level, trend and variance. Another Bayesian approach to multiple structural breaks is provided by Chib (1998), who considers structural breaks as regime switching of discrete-state Markov process with restricted transition probabilities.

The above literatures consider structural break(s) in univariate models. The estimation of and testing for structural break in cointegrated models has been also received attention. Gregory and Hansen (1996a) study residual-based tests for cointegration with a single structural break in a single equation model. They proposed $ADF-$, $Z_\alpha-$, and Z_t-type tests designed to test the null of no cointegration against the alternative of cointegration in the presence of a possible regime shift.

Gregory and Hansen (1996b) extend this work, by permitting a trend shift as well as a regime shift and providing the critical values for testing cointegration with a single break. Seo (1998) derives the Lagrange multiplier test for structural breaks in cointegration relations and adjustment terms, using the framework of Andrews and Ploberger (1994). Hansen and Johansen (1999) test parameter instability in cointegrating vectors based on Nyblom's L statistic (1989). Hansen (2003) explores the multiple-break case in cointegrated systems, and allows changes in any subset of the parameters, where the time of the change points and the number of cointegration relations are treated as known. Inoue (1999) derives a rank test for cointegrated systems with a structural change in trend. Bai et al. (1998) develop methods for constructing confidence intervals for the date of a single break in multivariate time series, and show that the accuracy of the break point estimators can be improved with series that have common breaks. While these authors assume the constant volatility in VAR, Bai (2000) allows the variance-covariance matrices to be affected by the breaks, using the quasi-maximum likelihood method. He also considered multiple breaks instead of a single break.

The main contribution of this paper is the development of general multivariate structural break models. We consider multiple structural breaks in any subset of the parameters in VAR or co-integrated VAR models, using a Bayesian approach which extends Wang and Zivot's (2000) method for detecting multiple structural changes in univariate models. In cointegration analysis, as changes in volatility and other terms are likely to affect the strength of the adjustment toward the equilibrium, it is of interest to analyze a model where structural breaks also affect in the adjustment terms, cointegrating vectors, and/or cointegrating rank. Hansen (2003) considers similar general cointegration models with structural breaks in any subset of parameters, where the number of cointegration relations, the number of breaks and the location of the break points are known. This paper considers general multivariate cointegrated models with breaks in any subset of the parameters where the break points and the rank are unknown. This is possible by applying a valid Bayesian approach to cointegration proposed by Strachan (2003), which is based on the singular value decomposition method by Kleibergen and Paap (2002) and Kleibergen and van Dijk (1998). For a less general case where cointegrating rank is not allowed to change with breaks, a simpler method by Strachan and Inder (2004) can be applied.

The Bayesian approach has several advantages over the classical method in the context of structural break models as it is technically simpler, allows inferences that are optimal given the framework, and allows for nonnested model comparison by computing posterior odds (see Raftery, 1994). Additionally, inference from the Bayesian approach is based on the exact finite sample properties for all of the parameters of the model. Finally, unlike most classical methods for detecting structural breaks, the Bayesian approach provides information about uncertainty in all estimated parameters including the location of the break dates. When the posterior probability mass function of the change point exhibits a substantial range in dates, the structural break may occur smoothly, rather than suddenly at a particular date.

This paper is organized as follows. Section 2 presents a Bayesian approach to VAR model with multiple structural breaks, using a simple Gibbs sampler. In Section 3, we extend the approach of the VAR model with multiple breaks to vector error correction models with multiple breaks in deterministic terms, adjustment term, cointegrating vector, variance-covariance matrices, and cointegrating rank, using Metropolis-within-Gibbs sampling algorithm, based on the method by Strachan (2003). We also consider a case where cointegrating rank is not allowed to change with breaks. This case is treated by applying a simpler method by Strachan and Inder (2004) with the Griddy-Gibbs sampler to estimate the cointegrating vectors. Section 4 considers the issue of model selection for detecting multiple structural breaks using Bayes factors calculated by using Schwarz BIC method and Chib's (1995) method. In Section 5 determining the cointegrating rank is considered for the two cases - one for where the cointegrating rank is subject to change and the other is for where it is not subject to change with breaks. In Section 7, Monte Carlo simulations are presented using artificially generated data for VAR models and vector error correction models with multiple breaks in order to examine the performances of detecting the number of breaks using our methods. To illustrate an empirical study of the VAR model with multiple breaks, Section 8 presents the predictive power of the yield curve on output growth. For an application of the vector error correction model with multiple breaks, we apply the method to investigate US term structure of interest rates. Section 9 concludes. All computation in this paper are performed using code written by the author with Ox v3.30 for Linux (Doornik, 1998).

2 Bayesian Inference in Vector Autoregressive Model with Multiple Structural Breaks

2.1 Statistical Model for VAR with Multiple Structural Breaks

In this section we consider a Bayesian approach to VAR model with multiple structural breaks. Let y_t denote a vector of n-dimensional $(1 \times n)$ time series. If all parameters in a VAR are assumed to be subject to structural breaks, then the model is

$$y_t = \mu_t + t\delta_t + \sum_{i=1}^{p} y_{t-i}\Phi_{t,i} + \varepsilon_t \tag{1}$$

where $t = p, p+1, \ldots, T$; p is the number of lags; and ε_t are assumed $N(0, \Omega_t)$ and independent over time. Dimensions of matrices are μ_t, δ_t and ε_t $(1 \times n)$, $\Phi_{t,i}$ and Ω_t $(n \times n)$. The parameters μ_t, δ_t and Ω_t are assumed to be subject to m structural breaks ($m < t$) with break points $b_1, \ldots, b_m$, where $b_1 < b_2 < \cdots < b_m$, so that the observations can be separated into $m+1$ regimes.

Equation (1) can be rewritten in the matrix format as:

$$Y = XB + E \tag{2}$$

where $Y = \begin{bmatrix} y_p' & y_{p+1}', & \ldots, & y_T' \end{bmatrix}'$, $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$,

$$X_1 = \begin{bmatrix} s_{1,p} & \cdots & s_{m+1,p} & s_{1,p} & \cdots & s_{m+1,p} \\ s_{1,p+1} & \cdots & s_{m+1,p+1} & 2s_{1,p+1} & \cdots & 2s_{m+1,p+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \\ s_{1,T} & \cdots & s_{m+1,T} & (T-p+1)s_{1,T} & \cdots & (T-p+1)s_{m+1,T} \end{bmatrix},$$

$$X_2 = \begin{bmatrix} s_{1,p}y_{p-1}' & \cdots & s_{1,p}y_1' & \cdots\cdots & s_{m+1,p}y_{p-1}' & \cdots & s_{m+1,p}y_1' \\ s_{1,p+1}y_p' & \cdots & s_{1,p+1}y_2' & \cdots\cdots & s_{m+1,p+1}y_p' & \cdots & s_{m+1,p+1}y_2' \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ s_{1,T}y_{T-1}' & \cdots & s_{1,T}y_{T-p+1}' & \cdots\cdots & s_{m+1,T}y_{T-1}' & \cdots & s_{m+1,T}y_{T-p+1}' \end{bmatrix},$$

$$B = \begin{bmatrix} \mu_1', & \ldots, & \mu_{m+1}' & \delta_1', & \ldots, & \delta_{m+1}' & \Phi_{1,1}', & \ldots, & \Phi_{p,1}', & \ldots, & \Phi_{1,m+1}', & \ldots, & \Phi_{p,m+1}' \end{bmatrix}'$$

Let τ be the number of rows of Y $(\tau \times n)$, so that $\tau = T - p + 1$, then X is $\tau \times \kappa$ where $\kappa = (np+2)(m+1)$, and B is $\kappa \times n$. $s_{i,t}$ in X_1 and X_2 is an indicator variable which equals to 1 if regime is i and 0 otherwise.

2.2 Prior Distributions and Likelihood Functions for VAR with Multiple Structural Breaks

Let $b = (b_1, b_2, \ldots, b_m)'$ denote the vector of break dates. We specify priors for parameters, assuming prior independence between b, B and Ω_i, $i = 1, 2, \ldots, m$+1, such that $p(b, B, \Omega_1, \Omega_2, \ldots, \Omega_{m+1}) = p(b)\, p(B) \prod_{i=1}^{m+1} p(\Omega_i)$. This is because if we consider that the prior for B is conditional on Ω as is often used in regression models with the natural conjugate priors, it is not convenient to consider a case when the error covariance is also subject to structural breaks. Thus, the prior density for B is set as the marginal distribution and vectorized as $vec(B)$ unconditional on Ω_i for convenience. The prior for the covariance-variance matrix, Ω_i, is specified with an inverted Wishart density. For the prior for the location of the break dates b, we choose a diffuse prior such that the prior is discrete uniform over all ordered subsequences of $t = p+1, \ldots, T-1$. We consider that all priors for b, Ω_i, and $vec(B)$ are proper as:

$$p(b) \sim \mathcal{U}(p+1, T-1) \tag{3}$$

$$\Omega_i \sim IW(\psi_{0,i}, \nu_{0,i}) \tag{4}$$

$$vec(B) \sim MN(vec(B_0), V_0) \tag{5}$$

where $\mathcal{U}$ refers to a uniform distribution; IW refers to an inverted Wishart distribution with parameters $\psi_{0,i} \in \mathbb{R}^{n \times n}$ and degrees of freedom, $\nu_{0,i}$; MN refers to a multivariate normal with mean $vec(B_0) \in \mathbb{R}^{\kappa n \times 1}$, $\kappa = (np+2)(m+1)$ and covariance-variance matrix $V_0 \in \mathbb{R}^{\kappa n \times \kappa n}$.

The joint prior of b, B, and Ω_i is given by multiplication of (3) - (5) as follows:

$$
\begin{aligned}
& p\left(b, B, \Omega_1, \Omega_2, \ldots, \Omega_{m+1}\right) \\
& \propto \left(\prod_{i=1}^{m+1} \left|\psi_{0,i}\right|^{\nu_{0,i}/2} \left|\Omega_i\right|^{-(\nu_{0,i}+n+1)/2}\right) \left|V_0\right|^{-1/2} \\
& \times \exp\left[-\frac{1}{2}\left\{\operatorname{tr}\left[\sum_{i=1}^{m+1}\left(\Omega_i^{-1}\psi_{0,i}\right)\right] + vec\left(B-B_0\right)' V_0^{-1} vec\left(B-B_0\right)\right\}\right]
\end{aligned}
\tag{6}
$$

Using the definition of the matric-variate Normal density (see Bauwens, et al., 1999), the likelihood function for the structural break VAR model with the parameters, $b, B, \Omega_1, \ldots, \Omega_{m+1}$, is given by,

$$
\begin{aligned}
& \mathfrak{L}\left(b, B, \Omega_1, \ldots, \Omega_{m+1} \mid Y\right) \\
& \propto \left(\prod_{i=1}^{m+1} \left|\Omega_i\right|^{-t_i/2}\right) \exp\left(-\frac{1}{2}\operatorname{tr}\left[\sum_{i=1}^{m+1}\left\{\Omega_i^{-1}\left(Y_i - X_i B\right)'\left(Y_i - X_i B\right)\right\}\right]\right) \\
& = \left(\prod_{i=1}^{m+1} \left|\Omega_i\right|^{-t_i/2}\right) \exp\left(-\frac{1}{2}\sum_{i=1}^{m+1}\left[\left(vec\left(Y_i - X_i B\right)\right)'\left(\Omega_i \otimes I_\tau\right)^{-1}\left(vec\left(Y_i - X_i B\right)\right)\right]\right)
\end{aligned}
\tag{7}
$$

where t_i denotes the number of observations in regime i, $i = 1, 2, \ldots, m$+1; Y_i is the $t_i \times n$ partitioned matrix of Y values in regime i; and X_i is $t_i \times \kappa$ partitioned matrix of X values in regime i.

2.3 Posterior Specifications and Estimation for VAR with Multiple Structural Breaks

The joint posterior distribution can be obtained from the joint priors given in (6) multiplied by the likelihood function in (7), that is,

$$
\begin{aligned}
& p\left(b, B, \Omega_1, \ldots, \Omega_{m+1} \mid Y\right) \propto p\left(b, B, \Omega_1, \ldots, \Omega_{m+1}\right) \mathfrak{L}\left(b, B, \Omega_1, \ldots, \Omega_{m+1} \mid Y\right) \\
& \propto \left(\prod_{i=1}^{m+1} \left\{\left|\psi_{0,i}\right|^{\nu_{0,i}/2} \left|\Omega_i\right|^{-(t_i+\nu_{0,i}+n+1)/2}\right\}\right) \left|V_0\right|^{-1/2}
\end{aligned}
$$

$$\times \exp\left(-\frac{1}{2}\left[\text{tr}\left(\sum_{i=1}^{m+1}\Omega_i^{-1}\right) + \sum_{i=1}^{m+1}\left\{\left([vec\,(Y_i - X_iB)]'\,(\Omega_i \otimes I_\tau)^{-1}\,vec\,(Y_i - X_iB)\right)\right\}\right.$$
$$\left. + vec\,(B - B_0)'V_0^{-1}vec\,(B - B_0)\right]\Big) \tag{8}$$

Consider first the conditional posterior of b_i, $i = 1,2,\ldots,m$. Given that $p = b_0 < \cdots < b_{i-1} < b_i < b_{i+1} < \cdots < b_{m+1} = T$ and the form of the joint prior, the sample space of the conditional posterior of b_i only depends on the neighboring break dates b_{i-1} and b_{i+1}. It follows that, for $b_i \in [b_{i-1}, b_{i+1}]$,

$$p\,(b_i \mid [b - b_i], B, \Omega_1, \ldots, \Omega_{m+1}, Y) \propto p\,(b_i \mid b_{i-1}, b_{i+1}, B, \Omega_i, \Omega_{i+1}, Y_i) \tag{9}$$

for $i = 1,\ldots m$, which is proportional to the likelihood function evaluated with a break at b_i only using data between b_{i-1} and b_{i+1} and probabilities proportional to the likelihood function. Hence, b_i can be draw from multinomial distribution as

$$b_i \sim \mathcal{M}\,(b_{i+1} - b_{i-1}, p_{\mathcal{L}}) \tag{10}$$

where $p_{\mathcal{L}}$ is a vector of probabilities proportional to the likelihood functions.

Next, we consider the conditional posterior of Ω_i, and $vec\,(B)$. To derive these densities, the following theorem can be applied:

***Theorem**: In the linear multivariate regression model $Y = XB + E$, with the prior densities of $vec(B) \sim MN(vec(B_0), V_0)$ and $\Omega \sim IW(\Psi_0, \nu_0)$, the conditional posterior densities of $vec(B)$ and Ω are*

$$vec(B) \mid \Omega, Y \sim MN\,(vec(B_\star), V_B)$$

$$\Omega \mid B, Y \sim IW\,(\Psi_\star, \nu_\star)$$

where the hyperparameters are defined by

$$vec(B_\star) = \left[V_0^{-1} + \Omega^{-1} \otimes (X'X)\right]^{-1}\left[V_0^{-1}vec(B_0) + (\Omega \otimes I_\kappa)^{-1}\,vec(X'Y)\right]$$

$$V_B = \left[V_0^{-1} + \Omega^{-1} \otimes (X'X)\right]^{-1}$$

$$\Psi_\star = (Y - XB)'(Y - XB) + \Psi_0$$

$$\nu_\star = T + \nu_0$$

***Proof**: see Appendix A.*□

From (8), we can write two terms using the above theorem as:

$$\begin{aligned}
&\sum_{i=1}^{m+1} \left\{ [vec\,(Y_i - X_i B)]' (\Omega_i \otimes I_\tau)^{-1} vec\,(Y_i - X_i B) \right\} + [vec\,(B - B_0)]' V_0^{-1} vec\,(B - B_0) \\
&= [vec\,(B - B_\star)]' V_B^{-1} vec\,(B - B_\star) + Q
\end{aligned}$$

where

$$Q = \sum_{i=1}^{m+1} \left\{ [vec\,(Y_i)]' (\Omega_i \otimes I_\tau)^{-1} vec\,(Y_i) \right\} + [vec\,(B_0)]' V_0^{-1} vec\,(B_0) - [vec\,(B_\star)]' V_B^{-1} vec\,(B_\star)$$

Thus, the conditional posterior of Ω_i is derived as an inverted Wishart distribution as $\Omega_i \mid b, B, Y \sim IW(\Psi_{i,\star}, \nu_{\star,i})$ where $\Psi_{i,\star} = (Y_i - X_i B)'(Y_i - X_i B) + \psi_{0,i}$ and $\nu_{\star,i} = t_i + \nu_{0,i}$, thus:

$$p(\Omega_i \mid b, B, Y) = C_{IW}^{-1} |\Omega_i|^{-(t_i + \nu_i + n + 1)/2} \exp\left[-\frac{1}{2} \mathrm{tr}\left(\Omega_i^{-1} \Psi_{i,\star} \right) \right] \tag{11}$$

where $C_{IW} = 2^{n(t_i+\nu_{0,i})/2} \pi^{n(n-1)/4} \prod_{j=1}^{n} \Gamma\{(t_i + \nu_{0,i} + 1 - j)/2\} \, |\Psi_{i,\star}|^{-(t_i+\nu_{0,i})/2}$, $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x) dx$ for $x > 0$. The conditional posterior of $vec\,(B)$ is a multivariate normal density with covariance-variance matrix, V_B, that is,

$$p(vec\,(B) \mid b, \Omega_1, \ldots, \Omega_{m+1}, Y) = (2\pi)^{-\kappa n/2} \, |V_B|^{-1/2} \exp\left[-\frac{1}{2} \left\{ [vec\,(B - B_\star)]' V_B^{-1} vec\,(B - B_\star) \right\} \right] \tag{12}$$

where

$$vec\,(B_\star) = \left[V_0^{-1} + \sum_{i=1}^{m+1} \left\{\Omega_i^{-1} \otimes \left(X_i' X_i\right)\right\}\right]^{-1} \left[V_0^{-1} vec\,(B_0) + \sum_{i=1}^{m+1} \left\{\left(\Omega_i \otimes I_\kappa\right)^{-1} vec\left(X_i' Y_i\right)\right\}\right], \tag{13}$$

and

$$V_B = \left[V_0^{-1} + \sum_{i=1}^{m+1} \left\{\Omega_i^{-1} \otimes \left(X_i' X_i\right)\right\}\right]^{-1} \tag{14}$$

Given the full set of conditional posterior specifications above, we illustrate the Gibbs sampling algorithm for generating sample draws from the joint posterior. The following steps can be replicated:

- Step 1: Set $j = 1$. Specify starting values for the parameters of the model, $b^{(0)}$ $B^{(0)}$, and $\Omega_i^{(0)}$, where Ω_i is a covariance-variance matrix at regime i.

- Step 2a: Compute likelihood probabilities sequentially for each date at $b_1 = b_0^{(j-1)} + 1, \ldots, b_2^{(j-1)} - 1$ to construct a multinomial distribution. Weight these probabilities such that the sum of them equals 1.

- Step 2b: Generate a draw for the first break date b_1 on the sample space $(b_0^{(j-1)}, b_2^{(j-1)})$ from $p(b_1^{(j)} \mid b_0^{(j-1)}, b_2^{(j-1)}, B^{(j-1)}, \Omega_1^{(j-1)}, \Omega_2^{(j-1)}, Y)$.

- Step 3a: For $i = 3, \ldots, m+1$, compute likelihood probabilities sequentially for each date at $b_{i-1} = b_{i-2}^{(j-1)} + 1, \ldots, b_i^{(j-1)} - 1$ to construct a multinomial distribution. Weight these probabilities such that the sum of them equals 1.

- Step 3b: Generate a draw of the $(i-1)$th break date $b_{i-1}^{(j)}$ from the conditional posterior $p(b_{i-1}^{(j)} \mid b_{i-2}^{(j-1)}, b_i^{(j-1)}, B^{(j-1)}, \Omega_{i-1}^{(j-1)}, \Omega_i^{(j-1)}, Y)$. Go back to Step 3a to generate next break date, but with imposing previously generated break date. Iterate until all breaks are generated.

- Step 4: Generate $vec(B)^{(j)}$ from $p(vec\,(B) \mid b^{(j)}, \Omega_i^{(j-1)}, \ldots \Omega_{m+1}^{(j-1)}, Y)$ and convert to $B^{(j)}$.

- Step 5: Generate $\Omega_i^{(j)}$ from $p(\Omega_i \mid b^{(j)}, B^{(j)}, Y)$ for all $i = 1, \ldots, m+1$.

- Step 6: Set $j = j+1$, and go back to Step 2.

Step 2 through to Step 6 can be iterated N times to obtain the posterior densities. Note that the first L iterations are discarded in order to remove the effect of the initial values.

3 Bayesian Inference in Co-integrated VAR Model with Multiple Structural Breaks

3.1 VECM with Multiple Structural Breaks Where the Cointegrating Rank is Subject to Shift with Breaks

In this subsection, we consider a co-integrated multivariate model with multiple structural breaks where cointegrating rank and all parameters of the model are subject to shift with breaks. Let y_t denote an $I(1)$ vector of $1 \times n$ with r linear cointegrating relations. The long-run multiplier matrix Π is decomposed as $\beta\alpha$, where α is the adjustment term and β is the cointegrating vector, and both α' and β are $n \times r$ ($r \leq n$). Then, a vector error correction model (VECM) with p lags is expressed as

$$\begin{aligned} \Delta y_t &= \mu + t\delta + y_{t-1}\Pi + \sum_{l=1}^{p-1} \Delta y_{t-l}\Phi_l + \varepsilon_t \\ &= \mu + t\delta + y_{t-1}\beta\alpha + \sum_{l=1}^{p-1} \Delta y_{t-l}\Phi_l + \varepsilon_t \end{aligned} \tag{15}$$

where $\varepsilon \sim iid(0,\Omega)$; μ, δ and ε_t are $1 \times n$; Φ and Ω are $n \times n$.

If all parameters in the VECM (15) are subject to m structural breaks ($m < t$) with break points $b_1, \ldots, b_m$, where $b_1 < b_2 < \cdots < b_m$, so that the observations can be separated into $m+1$ regimes, then the VECM representation with p lag for observations $t = p, p+1, \ldots, T$, is:

$$\Delta y_t = \mu_t + t\delta_t + y_{t-1}\beta_t\alpha_t + \sum_{l=1}^{p-1} \Delta y_{t-l}\Phi_{l,t} + \varepsilon_t \tag{16}$$

where ε_t are assumed $N(0,\Omega_t)$.

Equation (**??**) can be rewritten in the matrix format as:

$$Y = Z\Pi + X\Gamma + E = WB + E \tag{17}$$

where

$$Y = \begin{bmatrix} \Delta y_p' & \cdots & \Delta y_T' \end{bmatrix}', W = \begin{bmatrix} Z & X \end{bmatrix}, B = \begin{bmatrix} \Pi' & \Gamma' \end{bmatrix}', E = \begin{bmatrix} \varepsilon'_p & \cdots & \varepsilon_T' \end{bmatrix}',$$

$$Z = \begin{bmatrix} Z_1 & \cdots & Z_{m+1} \end{bmatrix}, Z_i = \begin{bmatrix} s_{i,p-1}y_{p-1}' & \cdots & s_{i,T-1}y_{T-1}' \end{bmatrix}' \text{ for } i = 1,\ldots,m+1,$$

$$\Pi = \begin{bmatrix} \Pi_1' & \cdots & \Pi_{m+1}' \end{bmatrix}' \text{ where } \Pi_i = \beta_i\alpha_i,$$

$$\Gamma = \begin{bmatrix} \mu_1' & \cdots & \mu_{m+1}' & \delta_1' & \cdots \delta_{m+1}' & \Phi_{1,1}' & \cdots & \Phi_{1,p-1}' & \cdots\cdots & \Phi_{m+1,1}' & \cdots & \Phi_{m+1,p-1}' \end{bmatrix},$$

$$X = \begin{bmatrix} X_1 & X_2 \end{bmatrix},$$

$$X_1 = \begin{bmatrix} s_{1,p} & \cdots & s_{m+1,p} & s_{1,p} & \cdots & s_{m+1,p} \\ s_{1,p+1} & \cdots & s_{m+1,[+1} & 2s_{1,p+1} & \cdots & 2s_{m+1,p+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \\ s_{1,T} & \cdots & s_{m+1,T} & (T-p+1)s_{1,T} & \cdots & (T-p+1)s_{m+1,T} \end{bmatrix},$$

$$X_2 = \begin{bmatrix} s_{1,p}\Delta y_{p-1}' & \cdots & s_{1,p}\Delta y_1' & \cdots\cdots & s_{m+1,p}\Delta y_{p-1}' & \cdots & s_{m+1,p}\Delta y_1' \\ s_{1,p+1}\Delta y_p' & \cdots & s_{1,p+1}\Delta y_2' & \cdots\cdots & s_{m+1,p+1}\Delta y_p' & \cdots & s_{m+1,p+1}\Delta y_2' \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ s_{1,T}\Delta y_{T-1}' & \cdots & s_{1,T}\Delta y_{T-p+1}' & \cdots\cdots & s_{m+1,T}\Delta y_{T-1}' & \cdots & s_{m+1,T}\Delta y_{T-p+1}' \end{bmatrix},$$

Let τ be the number of rows of Y, so that $\tau = T - p + 1$, then X is $\tau \times (np+2)(m+1)$, Γ is $(np+2)(m+1) \times n$, W is $\tau \times \kappa$ where $\kappa = (np+n+2)(m+1)$, and B is $\kappa \times n$. $s_{i,t}$ in X is an indicator variable which equals to 1 if regime is i and 0 otherwise. Equation (17) represents the multivariate regression form of (**??**).

To estimate the VECM with multiple structural breaks, the method for a VAR model with breaks presented in the previous section, can be directly applied to estimate b, Ω_i, and $B = (\Pi', \Gamma')'$, and Strachan (2003)'s method is applied to decompose $\Pi_i = \beta_i\alpha_i$, which is based on the singular value decomposition (SVD) approach by Kleibergen and van Dijk (1998).

There are several Bayesian methods for estimating cointegrating vectors. The prior for the cointegrating vector β, might be chosen as a normal prior or Student t density with r^2 linear re-

strictions for identification and normalization on β such that $\beta' = (I_r, \beta'_\star)$, where $\beta_\star$ is $(n-r) \times r$ unrestricted matrix. This prior is used by Bauwens and Lubrano (1996) and Kleibergen and Paap (2002), but criticized by Strachan (2003) as this prior with the linear identifying restrictions on β is very likely to be invalid because this normalization restricts the estimable region of the cointegrating space, and the prior with this normalization is not invariant with respect to the ordering of the variables. Strachan (2003) proposes the method of identifying the space spanned by the cointegrating vectors. Strachan and van Dijk (2003) and Strachan and Inder (2004) discuss further problems associated with the use of linear identifying restrictions, and propose the Grassman approach which is valid prior on the cointegrating space. The identifying restrictions are $\beta'\beta = I_r$, that do not distort the weight on the cointegrating space, unlike the linear restrictions which entail several problems. Koop et al. (2004) provide general survey of Bayesian inference in the cointegrated model with a focus on the prior elicitation for the cointegrating space.

Strachan and Inder (2004) propose a simpler solution than Strachan (2003) to estimate the cointegrating vector and to detect the cointegration rank that uses a Laplace approximation. However, their method cannot be directly applied in our structural break model where the cointegration rank is subject to shift with breaks. Their transformation of the VECM in (**??**) is $Y = WB + E$ where $W = (X, Z\beta)$ and $B = (\Gamma', \alpha')'$, instead of (17), that is $Y = WB + E$ where $W = (X, Z)$ and $B = (\Gamma', \Pi')'$, so that the number of rank in each of the subsamples should be specified to generate draws of B within the Gibbs sampler. In order to use their method, we estimate total of $(n+1)^{(m+1)}$ models and calculate the Bayes factors for all these models to determine the number of cointegration relations in each of the regimes. By transforming the VECM to (17), generating draws of B does not depend on the number of rank, and thus we only need to estimate and calculate the Bayes factors of total of $(n+1)(m+1)$ models. However, their method can be used for models where the cointegrating rank is not subject to shift with breaks as shown in the next subsection.

In this paper Strachan's (2003) approach is used to identify the cointegrating vectors and the adjustment terms using the SVD of Π, and the number of rank is determined using approach based on the singular value decomposition method by Kleibergen and Paap (2002) and Kleibergen and van Dijk (1998) as Strachan (2003) applies this method.

Prior specifications for b, Ω_i, and $vec(B)$ are the same as those of a VAR model shown in the

previous section; $p(b) \sim \mathcal{U}(p+1, T-1)$, $\Omega_i \sim IW(\psi_{0,i}, \nu_{0,i})$, and $vec(B) \sim MN(vec(B_0), V_0)$. We assume that Π_i, $i = 1, \ldots, m+1$, is distributed independently, so that V_0 ($n\kappa \times n\kappa$) is defined as

$$V_0 = \begin{bmatrix} \Sigma_\Pi & 0 \\ 0 & \Sigma_\Gamma \end{bmatrix} \tag{18}$$

where Σ_Π is $n^2(m+1) \times n^2(m+1)$ matrix such that $\Sigma_\Pi = V_{\Pi_0} \otimes I_{n(m+1)}$, V_{Π_0} ($n \times n$) is prior covariance-variance matrix of $\Pi_i \sim MVN(\Pi_0, V_{\Pi_0})$; Σ_Γ is $n(np+2)(m+1) \times n(np+2)(m+1)$ matrix and is prior covariance-variance matrix of $\Gamma \mid \Pi \sim MVN(\Gamma_0, \Sigma_\Gamma)$.

With these priors, the posterior densities for b, Ω, and $vec(B)$ are given as:

$$p(b_i \mid B, \Omega_i, Y_i) \propto p(b_i \mid b_{i-1}, b_{i+1}, B, \Omega_i, \Omega_{i+1}, Y_i) \text{ for } \forall i \tag{19}$$

$$\Omega_i \mid b, B, Y \sim IW\left((Y_i - W_i B)'(Y_i - W_i B) + \psi_{0,i}, t_i + \nu_{0,i}\right) \text{ for } \forall i \tag{20}$$

$$vec(B) \mid b, \Omega_1, \ldots, \Omega_{m+1}, Y \sim N(vec(B_\star), V_B) \tag{21}$$

where $vec(B_\star)$ and V_B are given as

$$vec(B_\star) = \left[V_0^{-1} + \sum_{i=1}^{m+1} \left\{\Omega_i^{-1} \otimes \left(W_i' W_i\right)\right\}\right]^{-1} \left[V_0^{-1} vec(B_0) + \sum_{i=1}^{m+1} \left\{(\Omega_i \otimes I_\kappa)^{-1} vec\left(W_i' Y_i\right)\right\}\right],$$

and

$$V_B = \left[V_0^{-1} + \sum_{i=1}^{m+1} \left\{\Omega_i^{-1} \otimes \left(W_i' W_i\right)\right\}\right]^{-1}.$$

After drawing the posterior of $B = (\Gamma', \Pi')'$ from (21) in the Gibbs sampling, it is possible to identify the cointegrating vectors and the adjustment terms using Strachan's (2003) approach that is based on Kleibergen and van Dijk's (1998) SVD approach. Following Strachan (2003), we define the matrices $S_{jk,i}$ for $j,k = 0,1,2$ as $S_{jk,i} = M_{jk,i} - M_{j2,i} M_{22,i}^{-1} M_{2j,i}$ where $M_{jk,i} = (t_i + (p+3)n + 1)^{-1} \Sigma_{t=t_{i-1}+1}^{t_i} z_{j,t}' z_{k,t}$, $z_{0,t} = \Delta y_t$, $z_{1,t} = y_{t-1}$, and $z_{2,t} = x_t$. With the identifying restrictions

imposed on β_i in the normalizations $\beta_i' S_{11,i} \beta_i = I_r$ and $\beta_i' S_{10,i} S_{00,i}^{-1} S_{01,i} \beta_i = \Lambda_i = diag(\gamma_{1,i}, \ldots, \gamma_{r,i})$ $\gamma_{1,i} > \cdots > \gamma_{r,i}$, a total number of the restrictions is r^2, the transformation is given as:

$$\Pi_i = \beta_i \alpha_i + S_{11,i}^{-1} \beta_{\perp,i} \lambda_i \alpha_{\perp,i} \widetilde{\Sigma}_i \tag{22}$$

where $\widetilde{\Sigma}_i = S_{00,i} - S_{01,i} S_{11,i}^{-1} S_{10,i}$, $n \times (n-r)$ matrices $\beta_{\perp,i}'$ and $\alpha_{\perp,i}$ are orthogonal to β_i' and α_i such that $\beta_i' \beta_{\perp,i} = 0$ and $\alpha_{\perp,i} \alpha_i' = 0$. With this transformation, unrestricted (full rank) model is given by:

$$\Delta y_t = y_{t-1} \beta_t \alpha_t + y_{t-1} S_{11,t}^{-1} \beta_{\perp,t} \lambda_t \alpha_{\perp,t} \widetilde{\Sigma}_t + x_t \Phi + \varepsilon_t \tag{23}$$

If $\lambda_i = 0$, then Π_i is a reduced rank and thus the cointegration occurs. Thus, the posterior for $\zeta_i = (vec(\alpha_i), vec(\beta_i))$ is obtained as

$$p(\zeta_i \mid y) = p(\zeta_i, vec(\lambda_i) \mid y) \mid_{\lambda_i = 0} = p\left(vec(\Pi_i), \zeta_i, vec(\lambda_i)\right) \mid_{\lambda_i = 0} \left| J\left(vec(\Pi_i), \zeta_i, vec(\lambda_i)\right) \mid_{\lambda_i = 0} \right| \tag{24}$$

where $p(\zeta_i, vec(\lambda_i) \mid y)$ is posterior obtained from unrestricted (full rank) model in (23); $|J(\Pi_i, (\zeta_i, vec(\lambda_i)))|$ is the Jacobian for the transformation. See Appendix of Strachan (2003) for derivation of this Jacobian.

Define $\Pi_i^\star$ by $\Pi_i = S_{11,i}^{-1/2} \Pi_i^\star \widetilde{\Sigma}_i^{1/2}$, then the following transformation by the SVD is given as:

$$\begin{aligned} \Pi_i^\star &= U_i \underline{S}_i V_i' = \begin{bmatrix} U_{1,i} & U_{2,i} \end{bmatrix} \begin{bmatrix} \underline{S}_{1,i} & 0 \\ 0 & \underline{S}_{2,i} \end{bmatrix} \begin{bmatrix} V_{1,i}' \\ V_{2,i}' \end{bmatrix} \\ &= U_{1,i} \underline{S}_{1,i} V_{1,i}' + U_{2,i} \underline{S}_{2,i} V_{2,i}' \end{aligned} \tag{25}$$

where U_i are the eigenvectors of $\Pi_i^\star \Pi_i^{\star\prime}$, $U_{1,i}$ and $V_{1,i}'$ are $n \times r$, $U_{2,i}$ and $V_{2,i}$ are $n \times (n-r)$ and $\underline{S}_{1,i}$ and $\underline{S}_{2,i}$ are diagonal $r \times r$ and $(n-r) \times (n-r)$. Define the $r \times r$ orthogonal matrix Υ_i (thus, $\Upsilon_i \Upsilon_i' = \Upsilon_i' \Upsilon_i = I_r$) that contains the eigenvectors of the $U_{1,i}' S_{11,i}^{-1/2} S_{10,i} S_{00,i}^{-1} S_{01,i} S_{11,i}^{-1/2} U_{1,i}$, and $(n-r) \times (n-r)$ orthogonal matrices $\Upsilon_{1,i}$ and $\Upsilon_{2,i}$ that contain the eigenvector of the matrices $U_{2,i}' S_{11,i}^{-1} U_{2,i}$ and

$V_{2,i}'\widetilde{\Sigma}_i V_{2,i}$ respectively. Then the SVD in (25) is expressed as

$$\Pi_i^\star = U_{1,i}\Upsilon_i\Upsilon_i'\underline{S}_{1,i}V_{1,i}' + U_{2,i}\Upsilon_{1,i}\Upsilon_{1,i}\underline{S}_{2,i}\Upsilon_{2,i}\Upsilon_{2,i}'V_{2,i}' \tag{26}$$

and we obtain:

$$\beta_i = S_{11,i}^{-1/2}U_{1,i}\Upsilon_i \tag{27}$$

$$\alpha_i = \Upsilon_i'\underline{S}_{1.i}V_{1,i}'\widetilde{\Sigma}_i^{1/2} \tag{28}$$

$$\lambda_{r,i} = \Upsilon_{1,i}'\underline{S}_{2,i}\Upsilon_{2,i} \tag{29}$$

where the square root matrices $S_{11,i}^{-1/2}$ and $\widetilde{\Sigma}_i^{1/2}$ are defined as diagonal matrices with each of the diagonal elements replaced by its square root. From (27)-(29), the transformation (22) can be obtained. To draw ζ_i from (24), run the Metropolis-Hastings algorithm to draw from the posterior $p(\lambda_i,\alpha_i,\beta_i,\Omega_i \mid y) = g(\lambda_i \mid \alpha_i,\beta_i,\Omega_i,y)p(\alpha_i,\beta_i,\lambda_i,\Omega_i \mid y)$ where $g(\lambda_i \mid \alpha_i,\beta_i,\Omega_i,y)$ is the candidate-generating function that can be chosen by derivation from the conditional posterior density for $vec(B)$ in (21). With the assumption that Π_i, $i = 1,\ldots,m+1$, is distributed independently each other such that $\Pi_i \sim MVN(\Pi_0, V_{\Pi_0})$ where $V_{\Pi_0}(n\times n)$ is the first $m+1$ diagonal matrix of V_0 defined in (18), $p(B)$ is written as

$$\begin{aligned} p(B) &= p(\Pi)p(\Gamma \mid \Pi) \\ &= \left\{\prod_{i=1}^{m+1} p(\Pi_i)\right\} p(\Gamma \mid \Pi_1,\ldots,\Pi_{m+1}) \end{aligned} \tag{30}$$

Then, the conditional posterior density for Π_i can be written as $\Pi_i \mid b,\Omega_i,Y \sim MVN(\Pi_{\star,i}, V_{\Pi,i,\star})$ where $V_{\Pi,\star,i} = (V_{\Pi_0}^{-1} + Z_i'Z_i\Omega_i^{-1})^{-1}$ and $\Pi_{\star,i} = V_{\Pi,\star,i}(V_{\Pi_0}^{-1}\Pi_0 + Z_i'(Y_i - X_i\Gamma_i)\Omega_i)$ that is derived from (21). The decomposition of the trace in the posterior, as shown in Kleibergen and Paap (2002), gives a convenient choice for g is,

$$
\begin{aligned}
&g(\lambda_i \mid \alpha_i, \beta_i, \Omega_i, y) \\
&= (2\pi)^{-(n-r)^2/2} \left| \alpha_{\perp,i} \widetilde{\Sigma}_i \alpha'_{\perp,i} \right|^{(n-r)/2} \left| \beta'_{\perp,i} S_{11,i}^{-1} \left(V_{\Pi_0}^{-1} + Z_i' Z_i \Omega_i^{-1} \right)^{-1} \beta_{\perp,i} S_{11,i}^{-1} \right|^{(n-r)/2} \\
&\times \exp\left[-\frac{1}{2} \mathrm{tr} \left\{ \beta'_{\perp,i} S_{11,i}^{-1} \left(V_{\Pi_0}^{-1} + Z_i' Z_i \Omega_i^{-1} \right)^{-1} S_{11,i}^{-1} \beta_{\perp,i} (\lambda_i - \widetilde{\lambda}_i) \alpha_{\perp,i} \widetilde{\Sigma}_i \alpha'_{\perp,i} (\lambda_i - \widetilde{\lambda}_i)' \right\} \right]
\end{aligned} \tag{31}
$$

where

$$
\widetilde{\lambda}_i = \left\{ \beta'_{\perp,i} S_{11,i}^{-1} \left(V_{\Pi_0}^{-1} + Z_i' Z_i \Omega_i^{-1} \right)^{-1} S_{11,i}^{-1} \beta_{\perp,i} \right\}^{-1} \beta'_{\perp,i} S_{11,i}^{-1} \left(V_{\Pi_0}^{-1} + Z_i' Z_i \Omega_i^{-1} \right)^{-1} \widetilde{\Sigma}_i \alpha_{\perp,i} (\alpha_{\perp,i} \widetilde{\Sigma}_i \alpha'_{\perp,i})^{-1}.
$$

Appendix B provides the decomposition of the trace for the candidate-generating function g in (31). With the j-th draws, we can calculate the weight $w^{(j)}$ as follows;

$$
w_i^{(j)} = \frac{g\left(\lambda_i^{(j)} \mid \alpha_i, \beta_i, \Omega_i, y \right) p\left(\alpha_i^{(j)}, \beta_i^{(j)}, \lambda_i^{(j)}, \Omega_i^{(j)} \mid y \right) |_{\lambda_i = o}}{p(\alpha_i^{(j)}, \beta_i^{(j)}, \lambda_i^{(j)}, \Omega_i^{(j)} \mid y)}. \tag{32}
$$

Given the full set of conditional posterior specifications above, we illustrate the Metropolis-within-Gibbs sampling algorithm for generating sample draws from the joint posterior. The following steps are replicated N times to obtain the posterior densities with the first L iterations discarded:

- Step 1: Set $j = 1$. Specify starting values for the parameters of the model, $b^{(0)}$ $B^{(0)}$, and $\Omega_i^{(0)}$.
- Step 2 - 5: Generate $b^{(j)}$, $B^{(j)}$ and $\Omega_i^{(j)}$ as described in the previous section of the sampling scheme for the VAR model.
- Step 6a: Set $i = 1$. From the posterior of $B^{(j)} = (\Gamma^{(j)\prime} \Pi^{(j)\prime})'$ where $\Pi^{(j)} = (\Pi^{(j)\prime}_1 \cdots \Pi^{(j)\prime}_{m+1})$', perform the SVD of $\Pi_i^{(j)} = U_i^{(j)} S_i^{(j)} V_i^{(j)\prime}$, and then compute $\zeta_i^{(j)} = (\alpha_i^{(j)}, \beta_i^{(j)})$ given the number of rank r using (28) and (34).
- Step 6b: Generate $(\zeta_i^{(j)}, \lambda_i^{(j)})$ and from $p(\zeta_i, \lambda_i \mid y)$, and calculate $w_i^{(j)}$.

- Step 6c: Accept $(\zeta_i^{(j)}, \lambda_i^{(j)})$ with probability $\min(w_i^{(j)}/w_i^{(j-1)}, 1)$, otherwise $(\zeta_i^{(j)}, \lambda_i^{(j)}) = (\zeta_i^{(j-1)}, \lambda_i^{(j-1)})$.

- Step 6d: Set $i = i+1$, and go back to Step 6a for $i = 2, \ldots, m+1$.

- Step 7: Set $j = j+1$, and go back to Step 2

To determining the number of rank r in each regime $i = 1, \ldots, m+1$, we calculate the Bayes factors in Section 5.

3.2 VECM with Multiple Structural Breaks Where the Cointegrating Rank is Constant

The previous subsection dealt with general VECM with multiple structural breaks. If, however, the cointegrating rank is restricted to be constant over the whole sample, a simpler method by Strachan and Inder (2004) can be applied. The structural break VECM with $\varepsilon \sim N(0, \Omega_t)$

$$\Delta y_t = \mu_t + t\delta_t + y_{t-1}\beta_t\alpha_t' + \sum_{l=1}^{p-1} \Delta y_{t-l}\Phi_{l,t} + \varepsilon_t,$$

can be written as, instead of (17), $Y = WB + E$ where $Y = (\Delta y_p', \cdots, \Delta y_T')$, $W = (Z_1\beta_1, \cdots, Z_{m+1}\beta_{m+1}, X)$, $B = (\alpha', \Gamma')'$, $\alpha = (\alpha_1', \cdots, \alpha_{m+1}')'$, $E = (\ \varepsilon_p'\sigma_p' \quad \cdots \quad \varepsilon_T'\sigma_T'\ )'$, $Z_i = (s_{i,p-1}y_{p-1}', \ldots, s_{i,T-1}y_{T-1}')'$ for $i = 1, \ldots, m+1$. Γ and X are defined as in (17). Let τ be the number of rows of Y, then W is $\tau \times \kappa$ matrix and B is $\kappa \times n$ matrix, where $\kappa = (np+2+r)(m+1)$.

Prior specification for b, Ω_i, and $vec(B)$ are the same as those of previous subsection as $p(b) \sim \mathcal{U}(p+1, T-1)$, $\Omega_i \sim IW(\psi_{0,i}, \nu_{0,i})$, and $vec(B) \sim MN(vec(B_0), V_0)$. We assume that α_i, $i = 1, \ldots, m+1$, is distributed independently, so that V_0 ($n\kappa \times n\kappa$) is defined as

$$V_0 = \begin{bmatrix} \Sigma_\alpha & 0 \\ 0 & \Sigma_\Gamma \end{bmatrix} \tag{33}$$

where Σ_α is $nr(m+1) \times nr(m+1)$ matrix such that $\Sigma_\alpha = V_{\alpha_0} \otimes I_{r(m+1)}$, $V_{\alpha_0}(n \times n)$ is prior covariance-variance matrix of $\alpha_i \sim MVN(\alpha_0, V_{\alpha_0})$; Σ_Γ is $n(np+2)(m+1) \times n(np+2)(m+1)$ matrix and is prior covariance-variance matrix of $\Gamma \mid \alpha \sim MVN(\Gamma_0, \Sigma_\Gamma)$.

With these priors, the conditional posterior for b, Ω, and B are given as exactly same as (19), (20), and (21) respectively. However, we now have to specify a prior for β_i in W. Since the linear normalization, $\beta' = (I_r, \beta^{*\prime})$, is not valid as discussed by Strachan (2003) and Strachan and Inder (2004), we apply the Grassman approach by Strachan and Inder (2004) that specifies the prior on the cointegrating space rather than on the cointegrating vectors as $p(\beta) \propto \pi^{-(n-r)r} \prod_{j=1}^{r} \frac{\Gamma[(n+1-j)/2]}{\Gamma[(r+1-j)/2]}$ where $\Gamma[q] = \int_0^\infty u^{q-1}e^{-u}du$, $q > 0$ with identification restrictions, $\beta'\beta = I_n$. According to Strachan and Inder (2004), the resulting posterior for β is

$$p(\beta \mid y) \propto p(\beta) \left|\beta' D_0 \beta\right|^{-\tau/2} \left|\beta' D_1 \beta\right|^{(\tau-n)/2} \tag{34}$$

where $D_0 = D_1 - D_2$, $D_1 = S_{11}$ and $D_2 = S_{10}S_{00}^{-1}S_{01}$, $S_{jk} = M_{jk} - M_{j2}M_{22}^{-1}M_{2k}$, $M_{jk} = h_{jk} + \sum_{t=1}^{\tau} z'_{j,t}z_{k,t}$, $h_{jk} = 0$ if $j \neq k$, $h_{jj} = \varphi I$. To draw β from (34), the conditional posterior for β in (34) is not a known 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). Strachan and Inder (2004) use the Laplace approximation instead of the simulation methods. In this paper, we choose the Griddy-Gibbs sampling technique because the algorithm does not require the specification of the candidate-generating function that approximate the posterior. 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. Appendix C briefly explains the algorithm of the Griddy-Gibbs sampler for convenience.

4 Detecting for the Number of the Structural Breaks by Bayes Factors

In this section we consider detecting for the number of structural breaks as a problem of model selection. In Bayesian context, model selection for model i and j means computing the posterior odds ratio, that is the ratio of their posterior model probabilities, PO_{ij}:

$$PO_{ij} = \frac{p(\mathcal{M}_i \mid Y)}{p(\mathcal{M}_j \mid Y)} = \frac{p(Y \mid \mathcal{M}_i)}{p(Y \mid \mathcal{M}_j)} \cdot \frac{p(\mathcal{M}_i)}{p(\mathcal{M}_j)} = BF_{ij} \cdot \frac{p(\mathcal{M}_i)}{p(\mathcal{M}_j)} \tag{35}$$

where BF_{ij} denotes Bayes factor, defined as the ratio of marginal likelihood, $p(Y \mid \mathcal{M}_i)$ and $p(Y \mid \mathcal{M}_j)$. We compute the posterior odds for all possible models $i = 1, \ldots, J$ and then obtain the posterior probability for each model by computing

$$\Pr(\mathcal{M}_i \mid Y) = \frac{PO_{ij}}{\sum_{m=1}^{J} PO_{mj}} \tag{36}$$

where J is the number of models we consider.

There are several methods to compute the Bayes factor. Chib (1995) provides a method of computing the marginal likelihood that utilizes the output of the Gibbs sampler. The marginal likelihood can be expressed from the Bayes rule as

$$p(Y \mid \mathcal{M}_i) = \frac{p(Y \mid \theta_i^\star) p(\theta_i^\star)}{p(\theta_i^\star \mid Y)} \tag{37}$$

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

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

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

$$\widehat{p}(\theta^\star \mid Y) = \prod_{i=1}^{l} \left\{ \frac{1}{N} \sum_{j=1}^{N} p(\theta_i^\star \mid Y, \varphi_{i-1}^\star, \varphi^{i+1,(j)}) \right\}. \tag{38}$$

Note that $p(b_1, \ldots, b_m \mid B, \Omega_1, \ldots, \Omega_{m+1}, Y) = \prod_{i=1}^{m} p(b_i \mid b_{i-1}, b_{i+1}, B, \Omega_i, \Omega_{i+1}, Y_i)$ can be directly obtained from the Gibbs algorithm shown in Step 2 (a) in the section 2.3.

Chib's method can be used to determine the number of breaks for VAR models in Section 2 and cointegrated VAR models where the cointegrating rank is subject to change with breaks given

in Section 3.1, however, it cannot be used for cointegrated VAR models where the cointegrating rank is constant given in Section 3.2 due to non-standard form of the posterior for β.[1] In this case, we can adopt the Schwarz BIC method to approximate the Bayes factors as Yao (1988), Liu et al. (1997), and Wang and Zivot (2000) use the Schwarz BIC to determine the number of breaks. The Schwarz BIC is defined as

$$\text{BIC}_j\text{=-2}\ln \mathfrak{L}\left(\widehat{\theta}_j \mid Y;\text{M}_j\right)+q_j \ln(t) \tag{39}$$

where $\mathfrak{L}\left(\widehat{\theta}_j \mid Y;\text{M}_j\right)$ is the likelihood for model j evaluated at $\widehat{\theta}_j$, the posterior means of the parameters for model j; q_j denotes the total number of estimated parameters in the model j and M_j denotes the model indicator for model j. With the Schwarz BIC the Bayes factor for model i against model j can be approximated by $BF_{ij} \approx \exp\left[-0.5\left(\text{BIC}_i-\text{BIC}_j\right)\right]$

The BIC method described above gives a rough approximation to the Bayes factors, which is easy to use and does not require evaluation of the prior distribution, as Kass and Raftery (1995) note. However, it only provides an approximation, not an exact value of the Bayes factor. In this paper, the BIC method is only adopted for cointegrated VAR models where the rank is constant. For VAR models and cointegrated VAR models where the rank is allowed to change with breaks, we adopt Chib (1995)'s method to compute marginal likelihood $p(y \mid \mathcal{M}_i)$ to determine the number of structural breaks.[2]

5 Determining the Cointegrating Rank

To determine the cointegrating rank for the VECM in Subsection 3.1 where the cointegrating rank is also subject to change with breaks, the Bayes factor $BF(r \mid n)$ is calculated using the Savage-Dickey density ratio, that is the ratio of the marginal posterior density and the marginal prior density. With the approach of Chen (1994), the Bayes factor $BF(r \mid n)$ for all possible rank except $r=n$ (that is, full rank) can be obtained using draws of α_i, β_i, and λ_i from the posterior as follows;

[1] If the posterior is generated from non-standard form of density through the Metropolis-Hastings algorithm, one can estimate the marginal likelihood adopting a method by Chib and Jeliazkov (2001).

[2] An alternative approach for calculating the marginal likelihood is using the harmonic mean of the likelihood as $f(Y \mid \mathcal{M}_i)=N\left[\sum_{j=1}^{N} L(\theta^{(j)} \mid Y)\right]^{-1}$, where $\theta^{(j)}$, $j=1,\ldots,N$, are Gibbs output. Computing the harmonic mean of the likelihood is simple, however, as described in Kass and Raftery (1995), this method may exhibit unstable results.

$$\frac{1}{c_{r,i}}\left\{\frac{1}{N}\sum_{j=1}^{N} w_i^{(j)}\right\} \quad \rightarrow$$

$$BF_i(r \mid n) \quad = \quad \frac{\frac{1}{c_{r,i}} \int\int\int\int g(\lambda_i \mid \alpha_i, \beta_i, \Omega_i, y) p(\alpha_i, \beta_i, \lambda_i, \Omega_i \mid y) \mid_{\lambda_i=0} d\Omega_i d\lambda_i d\beta_i d\alpha_i}{\int\int\int\int p(\alpha_i, \beta_i, \lambda_i, \Omega_i \mid y) d\Omega_i d\lambda_i d\beta_i d\alpha_i} \tag{40}$$

where $c_{r,i}$ is a constant depending upon r and is calculated as:

$$c_{i,r} = \frac{\int\int\int\int p(\alpha_i, \beta_i, \lambda_i, \Omega_i) \mid_{\lambda_i=0} h(\lambda_i \mid \alpha_i, \beta_i, \Omega_i) d\Omega_i d\lambda_i d\beta_i d\alpha_i}{\int\int\int\int p(\alpha_i, \beta_i, \lambda_i, \Omega_i) d\Omega_i d\lambda_i d\beta_i d\alpha_i} \tag{41}$$

where $h(\lambda_i \mid \alpha_i, \beta_i, \Omega_i)$ is a proper conditional density. As shown in Kleibergen and Paap (2002), an appropriate density function h for the prior specification of $p(B)$ is a density function which is close to the conditional prior of λ, thus

$$\begin{aligned} h(\lambda_i \mid \alpha_i, \beta_i, \Omega_i) \quad &= \quad h(\lambda_i \mid \alpha_i, \beta_i) = (2\pi)^{-(n-r)^2/2} \left|\alpha_{\perp i} \widetilde{\Sigma}_i \alpha'_{\perp,i}\right|^{(n-r)/2} \left|\beta'_{\perp,i} V_{\Pi_0}^{-1} \beta_{\perp,i}\right|^{(n-r)/2} \\ &\quad \times \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\beta'_{\perp,i} V_{\Pi_0}^{-1} \beta_{\perp,i} (\lambda_i - \xi_i) \alpha_{\perp,i} \widetilde{\Sigma}_i \alpha'_{\perp,i} (\lambda_i - \xi_i)'\right\}\right] \end{aligned} \tag{42}$$

where $\xi_i = (\beta'_{\perp,i} V_{\Pi_0}^{-1} \beta_{\perp,i})^{-1} \beta'_{\perp,i} V_{\Pi_0}^{-1} (\Pi_0 - \beta_i \alpha_i) \widetilde{\Sigma}_i \alpha'_{\perp,i} (\alpha_{\perp,i} \widetilde{\Sigma}_i \alpha'_{\perp,i})^{-1}$. To obtain the value of (41), we simulate from the prior

$$p(\zeta_i, vec(\lambda_i), \Omega_i) \propto p(vec(\Pi_i), \Omega_i) \mid_{\Pi_i = \beta_i \alpha_i + S_{11,i}^{-1} \beta_{\perp,i} \lambda_i \alpha_{\perp,i} \widetilde{\Sigma}_i} |J(vec(\Pi_i), (\zeta_i, vec(\lambda_i)))|$$

to compute the ratio of the integrands of the numerator and denominator in (41), then take an average of these simulated ratios to estimate $c_{i,r}$. See Kleibergen and Paap (2002) for details.

For a model where the number of rank is not subject to change with breaks as shown in Subsection 3.2, the Bayes factors for all possible non-zero rank are obtained using the Savage-Dickey density ratio as follows:

$$\begin{aligned} BF(r = 0 \mid r \neq 0) \quad &= \quad BF(\alpha = 0 \mid \alpha \neq 0) \\ &= \quad \frac{p(\alpha = 0 \mid y)}{p(\alpha = 0)} \end{aligned} \tag{43}$$

where 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, $vec(B) \sim MN(vec(B_0), V_0)$ with V_0 defined in (33), implies $p(\alpha) = \prod_{i=1}^{m+1} p(\alpha_i)$, where $\alpha_i \sim MVN(\alpha_0, V_{\alpha_0})$. The posterior for α_i is also independently distributed as $\alpha_i \mid b, \Omega_i, Y_i \sim MVN(\alpha_{\star,i}, V_{\alpha,\star,i})$ where $V_{\alpha,\star,i} = (V_{\alpha_0}^{-1} + Z_i' Z_i \Omega_i^{-1})^{-1}$ and $\alpha_{\star,i} = V_{\alpha,\star,i}(V_{\alpha_0}^{-1}\alpha_0 + Z_i'(Y_i - X_i\Gamma_i)\Omega_i)$. Since

$$\frac{1}{N - N_0} \sum_{n=N_0+1}^{N} p(\alpha = 0 \mid b^{(n)}, \Omega^{(n)}, Y) \to p(\alpha = 0 \mid Y) \tag{44}$$

as N goes to infinity, the numerator of (43) can be easily calculated.

6 Simulation

In this section Monte Carlo simulation is conducted to examine the performance of the approach outlined in the previous sections. A simulation for VAR models with breaks is followed by another for VECM with breaks. Two structural breaks are given in artificially generated data for both simulations. We are interested in examining the performance in both detecting the number of breaks when the number of the breaks are unknown and the estimation of the location of the breaks when the number of breaks are correctly specified.

6.1 Monte Carlo Simulation: VAR with Structural Breaks

The first Monte Carlo simulation is for vector autoregressive models with multiple structural breaks. The following five data generation processes (DGPs) of two-variable VAR models with two structural breaks are considered:

$$\text{DGP 1: } y_t = \mu_1 + y_{t-1}\Phi_1 + \sigma_1\varepsilon_t$$

$$\text{DGP 2: } y_t = \mu_t + y_{t-1}\Phi_1 + \sigma_1\varepsilon_t$$

$$\text{DGP 3: } y_t = \mu_t + y_{t-1}\Phi_1 + \sigma_t\varepsilon_t$$

$$\text{DGP 4: } y_t = \mu_t + y_{t-1}\Phi_t + \sigma_1\varepsilon_t$$

$$\text{DGP 5: } y_t = \mu_t + y_{t-1}\Phi_t + \sigma_t \varepsilon_t$$

$$\text{for } t = 1, 2, \ldots, 300,$$

where $\varepsilon_t \sim iidN(0,1)$, $\mu_t = \mu_1 = (-0.1, -0.1)$, $\Phi_t = \Phi_1 = 0.2I_2$, $\sigma_t = \sigma_1 = 0.02I_2$ for $0 < t < 100$, $\mu_t = \mu_2 = (0,0)$, $\Phi_t = \Phi_2 = \begin{pmatrix} 0.3 & -0.2 \\ -0.2 & 0.5 \end{pmatrix}$, $\sigma_t = \sigma_2 = 0.1I_2$, for $100 \leq t < 200$, $\mu_t = \mu_3 = (0.1, 0.1)$, $\Phi_t = \Phi_3 = -0.2I_2$, $\sigma_t = \sigma_3 = 0.02I_2$, for $200 \leq t \leq 300$. DGP 1 contains no structural break while other models contain two structural breaks. In DGP 2, only the constant term changes with breaks. DGP 3 allows the constant terms and volatility to change with breaks. DGP 4 allows μ and Φ to change with breaks. DGP 5 is the most general model in which breaks affect all parameters of the model.

The Gibbs sampling algorithm presented in Subsection 2.3 is employed for the estimation of models for $m = 0, 1, \ldots, 4$ break points. For prior parameters, we set $\Psi_{0,i} = 0.1I_2$ and $\nu_{0,i} = 2.001$ for all i for the variance-covariance prior in (4), $B_0 = 0$ and $V_0 = 100 \times I_{nk}$ in (5) to ensure fairly large variance for representing prior ignorance. The number of lags in VAR is assumed to be known. Also, we assume that, except the number of breaks, correct model specifications are known for each model. We assign an equal prior probability to each model with i breaks, so that $\frac{Pr(m=i)}{Pr(m=0)} = 1$[3]. After running the Gibbs sampler for 500 iterations, we save the next 2,000 draws for inference. This procedure is replicated 500 times.

Table 1 summarizes the results of the Monte Carlo simulations. Each element in the Table shows the average posterior probability out of 500 replications for each number of breaks. We compute the posterior probability with Chib's method described in Section 4. For DGP 1, where there are no breaks, the average posterior probability when $m = 0$ is 94.2%. For DGP 2, 3, 4, and 5, the correct number of breaks, $m = 2$, is detected at about 94.5%, 99.5%, 96.7%, and 98.1% respectively. Thus, the DGP of the VAR models with breaks in volatility (DGP3 and 5) perform better than those of the homoskedastic VAR. Overall most of the iterations choose the correct number of breaks. Table 2 reports that the Monte Carlo mean of estimated break points that are the mode of the posterior when the correct number of breaks $m = 2$ is chosen. The estimates are

[3]Inclan (1993) and Wang and Zivot (2000) use the prior odds as an independent Bernoulli process with probability $p \in [0,1]$.

all close to the true values, $b = (100, 200)$.

6.2 Monte Carlo Simulation: VECM with Structural Breaks

The second experiment is for vector error correction models with multiple structural breaks. We consider the following five data generation processes (DGPs) of a two-variable co-integrated model:

$$\begin{aligned}
&\text{DGP 1: } \Delta y_t = \mu + y_{t-1}\beta\alpha + \sigma\varepsilon_t \\
&\text{DGP 2: } \Delta y_t = \mu_t + y_{t-1}\beta\alpha + \sigma\varepsilon_t \\
&\text{DGP 3: } \Delta y_t = \mu_t + y_{t-1}\beta\alpha + \sigma_t\varepsilon_t \\
&\text{DGP 4: } \Delta y_t = \mu_t + y_{t-1}\beta\alpha_t + \sigma\varepsilon_t \\
&\text{DGP 5: } \Delta y_t = \mu_t + y_{t-1}\beta\alpha_t + \sigma_t\varepsilon_t \\
&\text{for } t = 1, 2, \ldots, 300.
\end{aligned}$$

where $\varepsilon_t \sim iidN(0,1)$. DGP 1 represents a no structural break model. DGP 2 is a structural break model in μ only, and DGP 3 allows μ and σ to change with breaks. DGP 4 represents a structural break model in μ, α. DGP 5 allows μ, α and σ to change with breaks. In both DGP 4 and 5, the cointegrating rank is constant over the whole sample. The parameters given in each DGP 2-5 are shown in Table 3. For the DGP 1, the parameters are set as: $\mu = \mu_1$ of the DGP 2, and other parameters are the same as those of the DGP 2. These values are obtained by using Japanese short- and long-term interest rates.

The Gibbs sampling algorithm in Section 3.2 is implemented for the estimation of models for $m = 0, 1, \ldots, 4$ break points. For prior parameters, we set the same values for $\nu_{0,i}$, $\psi_{0,i}$, B_0, and V_0 as used in the previous simulation for the VAR models with breaks to ensure fairly large variance for representing prior ignorance. The cointegration rank and the number of the lags in VECM are assumed known. Also, we assume that correct model specifications are known for each model except the number of breaks. We assign an equal prior probability to each model with i breaks, so that $\frac{Pr(m=i)}{Pr(m=0)} = 1$. After running the Gibbs sampler for 500 iterations, we save the next 2,000 draws

for inference. This procedure is replicated 500 times.

Table 4 summarizes the results of the Monte Carlo simulations for model selection. Each element in the Table shows the average posterior probability out of 500 replications for each number of breaks. Unlike in the previous simulation for the VAR model, the Schwarz BIC method is adopted to calculate the marginal likelihood for the posterior probabilities. The Table shows that in most of the cases the correct number of break points, $m = 2$, is selected with dominantly high posterior probabilities. The heteroscedastic DGPs (DGP 3 and 5) perform better than the homoscedastic DGPs (DGP 2 and 4), as in the case of the simulation for VAR models in the previous subsection. DGP 5 shows the best performance with 94.4% of the time for $m = 2$.

Table 5 reports the Monte Carlo mode of the estimated break points. As in VAR models cases, these results show that in most of the cases the estimates are all closed to the true values, $b = (100, 200)$. The results of the homoscedastic DGPs, DGP 2 and DGP 4 show much higher standard deviations in estimating the break points.

7 Application 1: Predictive Power of the Yield Curve

In this section, we illustrate the instability of the predictive power of the yield curve on output growth in the United States as an empirical application of the VAR model with multiple structural breaks shown in Section 2.

7.1 Predictive Power of the Yield Curve on Output Growth

The predictive relationships between the slope of the yield curve and subsequent inflation or real output have been extensively studied. The consumption capital asset pricing model (CCAPM) with habit formation by Campbell and Cochrane (1999) shows that the term structure is related to the future economic activity - positive slopes of the real term structure precede economic expansion and negative slopes precede economic recession. Mishkin (1990a, 1990b), based on the Fisher decomposition, finds that the yield curve can predict inflation. Although Chen (1991), Estrella and Hardouvel (1991) and other studies find a positive correlation between the yield curve slopes and future real economic activities, Estrella et al (2003) suggest verifying the stability of the

relationship because the predictive power may depend on factors that may change over time such as monetary policy reaction function, real productivity, or monetary shocks.

Estrella et al (2003) investigate the instability of the predictive power based on the following model:

$$ip_{k,t} = \beta_0 + \beta_1 sp_t + \varepsilon_t \tag{45}$$

where sp_t is the spread between the two interest rates of bonds with different maturity; and $ip_{k,t}$ is the future growth rate of industrial production, IP_t, at a forecast horizon k and is defined as $ip_{k,t} \equiv (1200/k)\ln(IP_{t+k}/IP_t)$. We consider the forecast horizon of one year, that is, $k = 12$, as Estrella et al (2003) show that the predictive power of the spread on industrial production is maximum at $k = 12$.

7.2 Estimation Results

Instead of the linear single equation model given in (45), where future growth rate of industrial production is treated as the endogenous variable, we consider VAR models with $p = 3, 4$ and 5 lag terms as:

$$X_t = \mu_t + \sum_{i=1}^{p} X_{t-i}\Phi_i + \varepsilon_t \tag{46}$$

where $X_t = (sp_t, ip_{k,t})$ and $\varepsilon_t \sim iidN(0, \Omega_t)$. That is, we consider a VAR model with structural breaks in the intercept term μ and the volatility Ω.[4] The data for this model are, IP_t, the US industrial production, $r_{l,t}$, 10-year US treasury rate as a long-term interest rate, and $r_{s,t}$, the Federal fund rate as a short-term interest rate, based on monthly data obtained from the Saint Louis Federal Reserve Bank. The sample ranges from 1970:01 to 2005:11 with 430 observations. The two variables, $sp_t \equiv r_{l,t} - r_{s,t}$ and $ip_{12,t} \equiv 100\ln(IP_{t+12}/IP_t)$, are plotted in Figure 1. The prior parameters are the same as those used in the Monte Carlo simulation in Section 6.1. The Gibbs sampling is performed with 10,000 draws and the first 1,000 discarded for the VAR models with the number

[4] We also consider other models such that Φ_i also changes with breaks or the homoskedastic models where Ω does not change over time; however, the results prove to be insignificant as the Bayes factors are much lower than those in the model (46).

of structural breaks $m = 0, 1, \ldots, 4$ and the lags $p = 3, 4$ and 5.

Table 6 reports the Gibbs sampling results of model selection for the number of structural breaks, m, and the lag in the VAR, p. A VAR model with $m = 3$ and $p = 4$ results in the highest posterior model probability with 93.15%. Clearly, a VAR model with no break ($m = 0$) is rejected with nearly zero percent of the posterior model probability.

The estimates of the break points and other parameters of the VAR model with $m = 3$ and $p = 4$ are presented in Table 7. The posterior mass of each break date is plotted in Figure 2. The first break point is detected in the 95% HPDI (Highest Posterior Density Interval) between 1973:09 and 1975:07 with the posterior mode 1974:07. After the first break the variance of the interest rate spread decreased significantly and the productivity growth changed due to the first oil shock. The second break point is detected in the 95% HPDI between 1977:10 and 1979:10 with the posterior mode 1978:11. This second break date is associated with the advent of Fed Chairman Volcker in October 1979, initiating some fundamental changes until October 1982. However, the HPDI of the second date merely covers the assumed break date, October 1979, in the tail. The variance-covariance matrix of the regime between the second and third break dates, Ω_3, is much larger than that of the previous regime, Ω_2. The third estimated break date is found between 1982:09 and 1983:03 with the posterior mode 1983:01. This third break date is associated with the completion of the Volcker's monetary policies of the period with the *non-borrowed reserves operating procedure*, while the estimated mode of the third date is not exactly matched with the assumed date but the HPDI merely covers the assumed date in the tail. After the third break date the variance of both the spread and the industrial productivity growth was much reduced as shown in Ω_4.

8 Application 2: US Term Structure of Interest Rates

In this section, we analyze the US term structure of interest rates using the cointegration model with multiple structural breaks presented in Section 3.

8.1 The Expectations Hypothesis

The term structure of interest rates states that the expected future spot rate is equal to the future rate plus a time-invariant term premium. For an overview of the expectations hypothesis theory, see Shiller (1990). The continuously compounded yield to maturity for an f period bond is defined as $r_{f,t} = -(1/f)\, p_{f,t}$ where $p_{f,t}$ denote the log of the price of a unit-par-value discount bond at date t with f periods to maturity, and the one-period future rate of return, earned from period $t+f$ to $t+f+1$, is given by $1+F_{f,t} = P_{f,t}/P_{f+1,t}$. Let $r_{f,t}$ denote the yield to maturity f at t. Then the expectations hypothesis implies:

$$r_{f,t} - r_{1,t} = f^{-1} \sum_{j=1}^{f-1} \sum_{i=1}^{j} E_t \left(\Delta r_{1,t+i}\right) + L_f \tag{47}$$

where $L_f = f^{-1} \sum_{j=0}^{f-1} \Lambda_j$ and Λ_j is the term premium. If $r_{1,t}$ is integrated of order one, then $r_{f,t}$ must be integrated of order one and $y_{f,t}$ and $y_{1,t}$ are cointegrated with cointegration vector (1, -1), which is analyzed by Campbell and Shiller (1987). This cointegration relationship should be held in any pair of yield to maturity.

However, many studies find that the expectations hypothesis is rejected for US data. Hall et al (1992), and Engsted and Tanggaard (1994) consider this is due to the instability for interest rates between September 1979 and October 1982, known as the period with the *non-borrowed reserves operating procedure*. Taking this period into consideration, several studies such as Hansen and Johansen (1999), Bliss and Smith (1998), and Hansen (2003) show that the expectations hypothesis is held when structural breaks are imposed into the models.

8.2 Estimation Results

We analyze the US term structure of interest rates for detecting structural breaks in a vector error correction model applying the method outlined in Section 3. The data we use are the same as those of the previous application, that is, the Federal fund rate as a short-term interest rate and 10-year treasury bond yield as a long-term interest rate based on monthly data from the Saint Louis Federal Reserve Bank ranges from 1970:01 to 2006:01, with 432 observations. These series are plotted in Figure 3.

Let $y_t = (r_{l,t}, r_{s,t})$, where $r_{l,t}$ denotes the long-term interest rate at time t and $r_{s,t}$ denotes the short-term interest rate at time t, then the VECM with multiple structural breaks in the cointegrating rank, the adjustment term α, the cointegrating vector β, the risk premium δ and the covariance-variance matrix Ω can be expressed from the Granger representation theorem as:

$$\begin{aligned} \text{Model 1: } \Delta y_t &= (y_{t-1}\beta_t - \delta_t)\alpha_t + \sum_{i=1}^{p-1} \Delta y_{t-i}\Psi_i + \varepsilon_t \\ &= \mu_t + y_{t-1}\beta_t\alpha_t + \sum_{i=1}^{p-1} \Delta y_{t-i}\Psi_i + \varepsilon_t \end{aligned} \tag{48}$$

where $\varepsilon_t \sim N(0, \Omega_t)$ and $\mu_t = -\delta_t\alpha_t$. Thus, μ_t is restricted as its space spanned by α_t. However, we ignore this restriction as μ_t is independent upon the space spanned by α_t. Note that the risk premium δ is assumed to be constant in a given regime. This model is estimated using a method in Section 3.1, which is based on Strachan's (2003) approach with the Metropolis-within-Gibbs sampling algorithm. We also estimate a model where the cointegrating vectors, and thus the number of rank, are not subject to change with breaks as:

$$\text{Model 2: } \Delta y_t = \mu_t + y_{t-1}\beta\alpha_t + \sum_{i=1}^{p-1} \Delta y_{t-i}\Psi_i + \varepsilon_t \tag{49}$$

where $\varepsilon_t \sim N(0, \Omega_t)$. This model is estimated using a method in Section 3.2, based on Strachan and Inder's (2004) approach with the Griddy-Gibbs sampler. Model 1 allows μ, α, β and Ω to change with the breaks, while Model 2 restricts the cointegrating vector β to remain constant over the whole sample. We estimate both models and calculate the Bayes factors for $m = 0, 1, \ldots, 4$ break points. The number of lags in VAR, p, is varied with $p = 2, 3$ and 4. We set the same prior parameters used in the Monte Carlo simulation in Section 6.2. The MCMC sampling is performed with 10,000 draws with the first 1,000 discarded for both models.

Table 8 and Table 11 report the posterior model probabilities for the various number of break points as a model selection for Model 1 and Model 2 respectively. Note that the posterior model probabilities for Model 1 are calculated by Chib's method while those for Model 2 are by the Schwarz BIC method. Clearly, no-structural break, $m = 0$, is rejected by the data for both Model

1 and 2. Model with $m = 3$ and $p = 3$ is strongly favored with 99.7% for Model 1 and 97.2% for Model 2. Table 9 presents the posterior probabilities for different number of rank for Model 1 with $m = 3$ and $p = 3$, and it shows that $r = 1$ is favored for all regimes, although it is not clear in the regime 3. For Model 2 where the cointegrating rank does not change with break, the posterior probabilities for each rank is $\Pr(r = 0 \mid Y) = 0.0790$, $\Pr(r = 1 \mid Y) = 0.9210$, and $\Pr(r = 2 \mid Y) = 0.000$, so that the rank 1 is strongly supported. The estimates of the parameters excluding the lag terms in the vector error correction model with three structural breaks are given in Table 10 for Model 1 and in Table 12 for Model 2. The posterior mode for the first two breaks are detected around late 1979 and late 1982, which almost coincides with the period with the *non-borrowed reserves operating procedure*. The third break is detected in 1988:06 for both models. The posterior probability mass functions for the break dates are shown in Figure 4 for Model 1 and in Figure 6 for Model 2.

The results show that there are significant changes in μ, α , and Ω; however, no significant changes are shown in β. To compare Model 1 to Model 2, we calculate the Bayes factor using the Schwarz BIC as $BF_{21} \approx \exp[-0.5(\mathrm{BIC}_2 - \mathrm{BIC}_1)]$, that results in 33.91.[5] Thus, Model 2 is strongly favored over Model 1.

As for the adjustment term, $\alpha_i = (\alpha_{long,i}, \alpha_{short,i})'$ where $i = 1, \ldots, 4$, $\alpha_{long,i}$ is negative in all regimes, and $\alpha_{short,i}$ is positive in all regimes for both models. This suggests that positive deviation from the long-run equilibrium ($y_t\beta > 0$) would be corrected by rising in the short rate and/or by falling in the long rate. The adjustment terms for the short rate $\alpha_{short,i}$ (in absolute value) are much higher than those for the long term rate $\alpha_{long,i}$ for all regimes, which indicates that the short rate tends to have much more power to adjust toward the long-run equilibrium than the long-term rate. In regime 2, between late 1979 and late 1982, volatility of both the long- and short-term interest rates is quite high and the adjustment term in absolute value is also very high. After the third break (regime 4), the volatility of the interest rates and the adjustment term are quite small, especially α_{long} for both models seem not to be significant; that is, the long term rate does not respond by the deviations from the long-run equilibrium, while only short-term rate moves toward

[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} is more than 150, there is a very strong evidence against model j.

the equilibrium. The posterior densities for the adjustment terms are plotted in Figure 5 for Model 1 and in Figure 7 for Model 2.

The expectation hypothesis implies that $\beta_2 = -1$ and this value is included in the 95% HPDI of the posterior density. More formal testing for this over-identifying restrictions on the cointegrating vector can be done by computing Bayes factor with the null of $\beta_2 = -1$ against the alternative of $\beta_2 \neq -1$. The Bayes factor is computed using (39) as $BF \approx \exp[0.5(\text{BIC}_{UR} - \text{BIC}_R)]$, where BIC_{UR} denotes the unrestricted BIC and BIC_R denotes the restricted BIC with the restrictions of $\beta_2 = -1$. The Bayes factor is approximated to 338.98 for Model 2, which shows very strong evidence to support the expectation hypothesis.

9 Conclusion

We developed a Bayesian approach for analyzing a VAR model and co-integrated VAR model with multiple structural breaks based on the MCMC simulation methods. The number of structural breaks are chosen by the posterior model probability based on the estimation of the model. In the case of the cointegrated VAR models, the cointegration rank is also allowed to change with breaks, and the adjustment term and the cointegrating vectors are estimated using Strachan's (2003) method with the Metropolis-within-Gibbs sampling algorithm, a valid Bayesian method in terms of prior restrictions for the cointegrating vector. For a model where the cointegration rank is not allowed to change with breaks, we apply Strachan and Inder's (2004) method and use the Griddy-Gibbs sampling method to generate the cointegrating vectors. The Monte Carlo simulations demonstrated that our approach provides generally accurate estimation for the number of structural breaks as well as their locations. The Bayesian approach provides useful information such as uncertainty of models and all parameters including the location of the dates by the posterior mass function for each estimated break point.

The method is applied to two empirical studies, the predictive power of the yield curve and the US term structure of interest rates, in order to show that our Bayesian method is useful to analyze the case of multiple structural breaks. We found strong evidence of three structural breaks in both applications.

Appendix A. Proof of Theorem

For a linear regression model $Y = XB + E$, $E \sim iidN(0,\Omega)$, where Y and E are $T \times n$; X is $T \times \kappa$; B is $\kappa \times n$, given the prior density for $vec(B) \sim MN(vec(B_0), V_0)$ and $\Omega \sim IW(\Psi_0, \nu_0)$, the joint posterior is obtained by the joint prior

$$\begin{aligned} & p(vec(B),\Omega) = p(vec(B))p(\Omega) \\ & \propto |\Psi_0|^{\nu_0/2} |\Omega|^{-(\nu_0+n+1)/2} |V_0|^{-1/2} \exp\left[-\frac{1}{2}\left\{\mathrm{tr}\left(\Omega^{-1}\Psi_0\right) + vec(B-B_0)'V_0^{-1}vec(B-B_0)\right\}\right] \end{aligned} \tag{50}$$

with the likelihood

$$\mathfrak{L}(B,\Omega \mid Y) \propto |\Omega|^{-T/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(Y-XB)'(Y-XB)\right\}\right] \tag{51}$$

so that the joint posterior is

$$\begin{aligned} p(vec(B),\Omega \mid Y) \quad \propto \quad & p(vec(B),\Omega)\mathfrak{L}(B,\Omega \mid Y) \\ & |\Psi_0|^{\nu_0/2} |\Omega|^{-(T+\nu_0+n+1)/2} |V_0|^{-1/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left((Y-XB)'(Y-XB)+\Psi_0\right)\right\}\right] \\ & \times \exp\left[-\frac{1}{2}\left\{vec(B-B_0)'V_0^{-1}vec(B-B_0)\right\}\right]. \end{aligned} \tag{52}$$

From the joint posterior (52), it is easy to derive the conditional posterior density for Ω, which is the inverted Wishart density $IW(\Psi_\star, \nu_\star)$ as

$$\begin{aligned} p(\Omega \mid B,Y) \quad = \quad & \frac{p(B,\Omega \mid Y)}{p(B \mid Y)} \propto p(B,\Omega \mid Y) \\ \propto \quad & |\Omega|^{-(T+\nu_0+n+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left((Y-XB)'(Y-XB)+\Psi_0\right)\right\}\right] \\ = \quad & |\Omega|^{-(T+\nu_0+n+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Omega^{-1}\Psi_\star\right)\right] \end{aligned} \tag{53}$$

where $\Psi_\star = (Y - XB)'(Y - XB) + \Psi_0$ and $\nu_\star = T + \nu_0$.

As for the conditional posterior density for $vec(B)$, the likelihood

$$\begin{aligned}\mathfrak{L}(B, \Omega \mid Y) &\propto |\Omega|^{-T/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(Y - XB)'(Y - XB)\right\}\right] \\ &\propto |\Omega|^{-T/2} \exp\left[-\frac{1}{2}\left(vec(Y - XB)\right)'\left(\Omega \otimes I_T\right)^{-1}\left(vec(Y - XB)\right)\right] \qquad (54)\end{aligned}$$

can be used for obtaining the joint posterior density instead of (51) as:

$$\begin{aligned}p(vec(B), \Omega \mid Y) &\propto p(vec(B), \Omega)\mathfrak{L}(B, \Omega \mid Y) \\ &\propto |\Psi_0|^{\nu_0/2} |\Omega|^{-(T+\nu_0+n+1)/2} |V_0|^{-1/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Omega^{-1}\Psi_0\right)\right] \\ &\times \exp\left[-\frac{1}{2}\left\{\left(vec(Y - XB)\right)'\left(\Omega \otimes I_T\right)^{-1}\left(vec(Y - XB)\right) + \left(vec(B - B_0)\right)' V_0^{-1} vec(B - B_0)\right\}\right]. \qquad (55)\end{aligned}$$

The key term in the third line of the joint posterior density (55) can be written as:

$$\begin{aligned}&\left(vec(Y - XB)\right)'\left(\Omega \otimes I_T\right)^{-1}\left(vec(Y - XB)\right) + \left(vec(B - B_0)\right)' V_0^{-1} vec(B - B_0) \\ &= \left(vec(B - B_\star)\right)' V_B^{-1} vec(B - B_\star) + Q \qquad (56)\end{aligned}$$

where $Q = (vec(Y))'(\Omega \otimes I_T)^{-1} vec(Y) + (vec(B_0))' V_0^{-1} vec(B_0) - (vec(B_\star))' V_B^{-1} vec(B_\star)$, $V_B = \left[V_0^{-1} + (\Omega^{-1} \otimes (X'X)\right]^{-1}$, and $vec(B_\star) = V_B\left[V_0^{-1} vec(B_0) + (\Omega \otimes I_\kappa)^{-1} vec(X'Y)\right]$.

To prove equation (56), first rewrite the LHS of equation (56) as:

$$\begin{aligned}LHS &= \left(vec(Y - XB)\right)'\left(\Omega \otimes I_T\right)^{-1}\left(vec(Y - XB)\right) + \left(vec(B - B_0)\right)' V_0^{-1} vec(B - B_0) \\ &= (vec(Y))'(\Omega^{-1} \otimes I_T) vec(Y) + (vec(XB))'(\Omega^{-1} \otimes I_T) vec(XB) - 2\,(vec((Y))'(\Omega^{-1} \otimes I_T) vec(XB) \\ &+ (vec(B))' V_0^{-1} vec(B) + (vec(B_0))' V_0^{-1} vec(B_0) - 2\,(vec(B_0))' V_0^{-1} vec(B). \qquad (57)\end{aligned}$$

The RHS can be written as:

$$\begin{aligned} RHS &= (vec(B-B_\star))' V_B^{-1} vec(B-B_\star) + (vec(Y))' (\Omega \otimes I_T)^{-1} vec(Y) \\ &\quad + (vec(B_0))' V_0^{-1} vec(B_0) - (vec(B_\star))' V_B^{-1} vec(B_\star) \\ &= (vec(B))' V_B^{-1} vec(B) - 2\,(vec(B_\star))' V_B^{-1} vec(B) \\ &\quad + (vec(Y))' (\Omega^{-1} \otimes I_T) vec(Y) + (vec(B_0))' V_0^{-1} vec(B_0). \end{aligned} \tag{58}$$

So, from (57) and (58), $LHS - RHS$ is

$$\begin{aligned} LHS - RHS &= (vec(XB))' (\Omega^{-1} \otimes I_T) vec(XB) + (vec(B))' V_0^{-1} vec(B) - (vec(B))' V_B^{-1} vec(B) \\ &\quad -2 \left\{ (vec(Y))' (\Omega^{-1} \otimes I_T) vec(XB) + (vec(B_0))' V_0^{-1} vec(B) - (vec(B_\star))' V_B^{-1} vec(B) \right\} \\ &= \mathfrak{C} - 2\mathfrak{D} \end{aligned} \tag{59}$$

where $\mathfrak{C}$ and $\mathfrak{D}$ are defined as

$$\mathfrak{C} = (vec(XB))' (\Omega^{-1} \otimes I_T) vec(XB) + (vec(B))' V_0^{-1} vec(B) - (vec(B))' V_B^{-1} vec(B) \tag{60}$$

$$\mathfrak{D} = (vec(Y))' (\Omega^{-1} \otimes I_T) vec(XB) + (vec(B_0))' V_0^{-1} vec(B) - (vec(B_\star))' V_B^{-1} vec(B). \tag{61}$$

By substituting $V_B = \left[V_0^{-1} + \left\{ \Omega^{-1} \otimes (X'X) \right\} \right]^{-1}$, the third term of $\mathfrak{C}$ in (60) is

$$\begin{aligned} (vec(B))' V_B^{-1} vec(B) &= (vec(B))' \left[V_0^{-1} + \left\{ \Omega^{-1} \otimes (X'X) \right\} \right] vec(B) \\ &= (vec(B))' V_0^{-1} vec(B) + (vec(B))' \left[\Omega^{-1} \otimes (X'X) \right] vec(B) \\ &= (vec(B))' V_0^{-1} vec(B) + (vec(B))' vec\left[(X'X) B \Omega^{-1} \right]. \end{aligned} \tag{62}$$

Using (62) in (60), we have

$$\mathfrak{C} = (vec(XB))'(\Omega^{-1} \otimes I_T)vec(XB) - (vec(B))'vec(X'XB\Omega^{-1}).$$

Since $(vec(XB))'(\Omega^{-1} \otimes I_T)vec(XB) = ((I_n \otimes X)vec(B))'vec(XB\Omega^{-1}) = (vec(B))'(I_n \otimes X)'vec(XB\Omega^{-1})$, and $(vec(B))'vec(X'XB\Omega^{-1}) = (vec(B))'(I_n \otimes X)'vec(XB\Omega^{-1})$, so we have $\mathfrak{C} = 0$.

Next, we consider $\mathfrak{D}$. The first term of $\mathfrak{D}$ in (61) is

$$\begin{aligned} (vec(Y))'(\Omega^{-1} \otimes I_T)vec(XB) &= (vec(Y))'(\Omega^{-1} \otimes I_T)(I_n \otimes X)vec(B) \\ &= (vec(Y))'(\Omega^{-1} \otimes X)vec(B). \end{aligned} \tag{63}$$

Since $vec(B_\star) = V_B\left[V_0^{-1}vec(B_0) + (\Omega \otimes I_\kappa)^{-1}vec(X'Y)\right] = V_B\left[V_0^{-1}vec(B_0) + vec(X'Y\Omega^{-1})\right]$, the third term of $\mathfrak{D}$ is,

$$\begin{aligned} (vec(B_\star))'V_B^{-1}vec(B) &= \left[V_0^{-1}vec(B_0) + vec(X'Y\Omega^{-1})\right]'vec(B) \\ &= (vec(B_0))'V_0^{-1}vec(B) + \left[(\Omega^{-1} \otimes X')vec(Y)\right]'vec(B) \\ &= (vec(B_0))'V_0^{-1}vec(B) + (vec(Y))'(\Omega^{-1} \otimes X)vec(B). \end{aligned} \tag{64}$$

Thus, with (63) and (64), we have $\mathfrak{D}$ as:

$$\begin{aligned} \mathfrak{D} &= (vec(Y))'(\Omega^{-1} \otimes X)vec(B) + (vec(B_0))'V_0^{-1}vec(B) \\ &\quad - \left\{(vec(B_0))'V_0^{-1}vec(B) + (vec(Y))'(\Omega^{-1} \otimes X)vec(B)\right\} \\ &= 0. \end{aligned}$$

Therefore, with $\mathfrak{C} = \mathfrak{D} = 0$, we have $LHS - RHS = \mathfrak{C} - 2\mathfrak{D} = 0$, so that equation (56) is proved and thus the conditional posterior density for $vec(B)$ is

$$
\begin{aligned}
p(vec(B) \mid \Omega, Y) &= \frac{p(B, \Omega \mid Y)}{p(\Omega \mid Y)} \propto p(vec(B), \Omega \mid Y) \\
&\propto \exp\left[-\frac{1}{2}\left\{(vec(Y - XB))'(\Omega \otimes I_T)^{-1}(vec(Y - XB)) + (vec(B - B_0))' V_0^{-1} vec(B - B_0)\right\}\right] \\
&\propto \exp\left[-\frac{1}{2}\left\{(vec(B - B_\star))' V_B^{-1} vec(B - B_\star)\right\}\right]
\end{aligned}
$$

where

$$
V_B = \left[V_0^{-1} + (\Omega^{-1} \otimes (X'X))\right]^{-1}
$$

and

$$
vec(B_\star) = V_B\left[V_0^{-1} vec(B_0) + (\Omega \otimes I_\kappa)^{-1} vec(X'Y)\right],
$$

so that $vec(B) \mid \Omega, Y \sim MN(vec(B_\star), V_B)$.

Appendix B. Decomposition

The conditional posterior specification for $vec(B)$ in (21) is Normal. Since $B = (\Pi', \Gamma')' = (\Pi_1', \ldots, \Pi_{m+1}', \Gamma')'$, if we assume that $p(B) = \left\{\prod_{i=1}^{m+1} p(\Pi_i)\right\} p(\Gamma \mid \Pi_1, \ldots, \Pi_{m+1})$ as (30) where prior for Π_i is Normal such as $p(\Pi_i) \propto \exp\left[-\frac{1}{2}\mathrm{tr}\left\{(\Pi_i - \Pi_0)' V_{\Pi_0}^{-1}(\Pi_i - \Pi_0)\right\}\right]$, the conditional posterior for Π_i is also Normal such as $p(\Pi_i \mid \Omega_i, y) \propto \exp\left[-\frac{1}{2}\mathrm{tr}(\Pi_i - \Pi_{\star,i})' V_{\Pi,\star,i}^{-1}(\Pi_i - \Pi_{\star,i})\right]$ where $V_{\Pi,\star,i} = (V_{\Pi_0}^{-1} + Z_i' Z_i \Omega_i^{-1})^{-1}$ and $\Pi_{\star,i} = V_{\Pi,\star,i}(V_{\Pi_0}^{-1}\Pi_0 + Z_i'(Y_i - X_i\Gamma_i)\Omega_i)$. The trace in the posterior density can be decomposed as follows:

$$
\begin{aligned}
&\mathrm{tr}\left\{(\Pi_i - \Pi_{\star,i})' V_{\Pi,\star,i}^{-1}(\Pi_i - \Pi_{\star,i})\right\} \\
&= \mathrm{tr}\left\{(\beta_i\alpha_i + S_{11,i}^{-1}\beta_{\perp,i}\lambda_i\alpha_{\perp,i}\widetilde{\Sigma}_i - \Pi_{\star,i})' \left(V_{\Pi_0}^{-1} + Z_i' Z_i \Omega_i^{-1}\right)^{-1} (\beta_i\alpha_i + S_{11,i}^{-1}\beta_{\perp,i}\lambda_i\alpha_{\perp,i}\widetilde{\Sigma}_i - \Pi_{\star,i})\right\} \\
&+ \mathrm{tr}\left\{(\alpha_{\perp,i}\widetilde{\Sigma}_i\alpha_{\perp,i}')(\lambda_i - \widetilde{\lambda}_i)'\beta_{\perp,i}' S_{11,i}^{-1}\left(V_{\Pi_0}^{-1} + Z_i' Z_i \Omega_i^{-1}\right)^{-1} S_{11,i}^{-1}\beta_{\perp,i}(\lambda_i - \widetilde{\lambda}_i)\right\} \\
&- \mathrm{tr}\left\{\widetilde{\Sigma}_i\alpha_{\perp,i}'(\alpha_{\perp,i}\widetilde{\Sigma}_i\alpha_{\perp,i}')^{-1}\alpha_{\perp,i}\widetilde{\Sigma}_i(\beta_i\alpha_i - \Pi_{\star,i})\left(V_{\Pi_0}^{-1} + Z_i' Z_i \Omega_i^{-1}\right)^{-1} S_{11,i}^{-1}\beta_{\perp,i}\right\}
\end{aligned}
$$

$$\times \left[\left\{ \beta'_{\perp,i} S_{11,i}^{-1} \left(V_{\Pi_0}^{-1} + Z_i' Z_i \Omega_i^{-1} \right)^{-1} S_{11,i}^{-1} \beta_{\perp,i} \right\}^{-1} S_{11,i}^{-1} \beta_{\perp,i} \left(V_{\Pi_0}^{-1} + Z_i' Z_i \Omega_i^{-1} \right)^{-1} (\beta_i \alpha_i - \Pi_{\star,i}) \right]$$

where

$$\widetilde{\lambda}_i = \left\{ \beta'_{\perp,i} S_{11,i}^{-1} \left(V_{\Pi_0}^{-1} + Z_i' Z_i \Omega_i^{-1} \right)^{-1} S_{11,i}^{-1} \beta_{\perp,i} \right\}^{-1} \beta'_{\perp,i} S_{11,i}^{-1} \left(V_{\Pi_0}^{-1} + Z_i' Z_i \Omega_i^{-1} \right)^{-1} \widetilde{\Sigma}_i \alpha_{\perp,i} (\alpha_{\perp,i} \widetilde{\Sigma}_i \alpha'_{\perp,i})^{-1}.$$

So that equation (31) can be derived.

Appendix C. 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. 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 $\text{vec}(\beta)' = (\beta_1, \beta_2, \ldots, \beta_m)$. With an arbitrary starting value (within the upper and the lower bound of the grid), compute $f(\beta_1 | \beta_2^i, \beta_3^i, \ldots, \beta_m^i, Y)$, where i denotes the i-th loop, over the grid $(\beta_{1,1}, \beta_{1,2}, \ldots, \beta_{1,U})$, 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 $G = (0, \Phi_2, \Phi_3, \ldots, \Phi_U)$ where

$$\begin{aligned} \Phi_j &= \int_{\beta_{1,1}}^{\beta_{1,j}} f(\beta_1 | \beta_2^i, \beta_3^i, \ldots, \beta_m^i, Y) d\beta_1 \\ & j = 2, \ldots, U \end{aligned}$$

4. Compute the normalized pdf values $G_\zeta = G_j / \Phi_U$ of $\zeta(\beta_1 | \beta_2^i, \beta_3^i, \ldots, \beta_m^i, Y)$.

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 $\zeta(\beta_1|\beta_2^i,\beta_3^i,\ldots,\beta_m^i,Y)$.

6. Repeat steps 2-5 for $\beta_2,\ldots,\beta_m$.

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.

References

[1] Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821-856.

[2] Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383-1414.

[3] Bai, J., 1999. Likelihood ratio test for multiple structural changes. Journal of Econometrics 91, 299-323.

[4] Bai, J., 2000. Vector autoregressive models with structural changes in regression coefficients and in variance-covariance matrix. Annals of Economics and Finance 1, 303-339.

[5] Bai, J., Lumsdaine, R., Stock, J.H., 1998. Testing for and dating common breaks in multivariate time series. Review of Economic Studies 65, 395-432.

[6] Bai, J., Perron, P., 1998. Testing for and estimation of multiple structural changes. Econometrica 66, 817-858.

[7] Banerjee, A., Lumsdaine, R., Stock, J.H., 1992. Recursive and sequential tests of the unit-root and trend-break hypotheses: theory and international evidence. Journal of Business and Economic Statistics 10, 271-287.

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

[9] Bliss, R.R., Smith, D.C., 1998. The elasticity of interest rate volatility: Chan, Kaolyi, Longstaff and Sanders Revisited. Journal of Risk 1: 21-46.

[10] Campbell, J.Y., Cochrane, J.H., 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 108:2, 205-251.

[11] Campbell, J.Y., Shiller, R.J., 1987. Cointegration and tests of present value models. Journal of Political Economy 95, 1062-1088.

[12] Chen, M-H., 1994. Importance-weighted marginal Bayesian posterior density estimation. Journal of the American Statistical Association 89, 818-824.

[13] Chen, N-F., 1991. Financial investment opportunities and the macroeconomy. Journal of Finance 46, 529-554.

[14] Chen, J., Gupta, A.K., 1997. Testing and locating variance change-points with application to stock prices. Journal of the American Statistical Association 92, 739-747.

[15] Chib, S., 1995. Marginal likelihood from the Gibbs output. Journal of the American Statistical Association vol.90, 432, Theory and Methods, 1313-1321.

[16] Chib, S., 1998. Estimation and comparison of multiple change-point models. Journal of Econometrics 86, 221-241.

[17] Chib, S., Greenberg, E., 1995, Understanding the Metropolis-Hastings algorithm, The American Statistician 49, 327-335.

[18] Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the Metropolis Hastings output. Journal of the American Statistical Association 96, 270-281.

[19] Cristiano, L.J., 1992. Searching for a break in GNP. Journal of Business and Economic Statistics 10, 237-256.

[20] Doornik, J.A., 1998. OX An object-oriented matrix programming language. Timberlake Consultants Press, London.

[21] Engel, C., Hakkio, C.S., 1996. The distribution of exchange rates in the EMS. International Journal of Finance and Economics 1, 55-67.

[22] Engsted, T., Tanggaard, C., 1994. Cointegration and the US term structure. Journal of Banking and Finance 167-181.

[23] Estrella, A., Hardouvelis, G., 1991. The term structure as a predictor of real economic activity. Journal of Finance 46, 555-576.

[24] Estrella, A., Rodrigues, P., Schich, S., 2003. How stable is the predictive power of the yield curve? Evidence from Germany and the United States. The Review of Economics and Statistics 85(3), 629-644.

[25] Gregory, A.W., Hansen, B.E., 1996a. Residuals based tests for cointegration in models with regime shifts. Journal of Econometrics 70, 99-126.

[26] Gregory, A.W., Hansen, B.E., 1996b. Practitioners corner, testing for cointegration in models with regime and trend shifts. Oxford Bulletin of Economics and Statistics 58, 3, 555-560.

[27] Hall, A.D., Anderson, H.M., Granger, C.W.J., 1992. A cointegration analysis of treasure bill yield. The Review of Economic and Statistics 74, 116-126.

[28] Hansen, B.E., 2000, Testing for structural change in conditional vector autoregressive model. Journal of Econometrics 114, 93-115.

[29] Hansen, H., Johansen, S., 1999. Some tests for parameter constancy in cointegrated VAR-models. Econometrics Journal 2, 306-333.

[30] Hansen, P.R., 2003. Structural changes in the cointegrated vector autoregressive model. Journal of Econometrics 114, 261-295.

[31] Inclan, C., 1993. Detection of multiple structural changes of variance using posterior odds. Journal of Business and Economic Statistics 11, 289-300.

[32] Inclan, C., Tiao, G.C., 1994. Use of cumulative sums of squares for retrospective detection of changes of variance. Journal of the American Statistical Association 89, 913-923.

[33] Kass, R.E., Raftery, A. E., 1995. Bayes factors. Journal of the American Statistical Association 90, 773-795.

[34] Kim, C.J., Engel, C., 1999. The long run U.S/U.K real exchange rate. Journal of Money, Credit and Banking 31, 335-356.

[35] Kim, C-J., Nelson, C.R., 1999. Has the U.S.economy become more stable? A Bayesian approach based on a Markov switching model of business cycle. The Review of Economics and Statistics 81, 608-616.

[36] Kleibergen, F., Paap, R., 2002. Priors, Posteriors and Bayes factors for a Bayesian analysis of cointegration. Journal of Econometrics 111, 223-249.

[37] Kleibergen, F., van Dijk, H.K., 1998. Bayesian simultaneous equations analysis using reduced rank structures. Econometric Theory 14, 701-743.

[38] Koop, G., Strachan, R., van Dijk H., 2004. Bayesian approaches to cointegration. Working Paper No.04/27, University of Leicester.

[39] Liu, J., Wu, S., Zidek, J.V., 1997, On segmented multivariate regressions. Statistica Sinica 7, 497-525.

[40] Mishkin, F.S., 1990a. The information in the longer maturity term structure about future inflation. The Quarterly Journal of Economics CV:442, 815-828.

[41] Mishkin, F.S., 1990b. What does the term structure tell us about future inflation? Journal of Monetary Economics 25, 77-95.

[42] Nyblom, J., 1989. Testing for the constancy of parameters over time. Journal of the American Statistical Association 84, 223-230.

[43] Perron, P., 1989. The great crash, the oil price shock and the unit root hypothesis. Econometrica 57, 1361-1401.

[44] Raftery, A.E., 1994. Changepoint and change curve modeling in stochastic processes and spatial statistics. Journal of Applied Statistical Science 7, 403-424.

[45] Ritter. C., Tanner, M.A., 1992. Facilitating the Gibbs Sampler: The Gibbs stopper and the Griddy-Gibbs Sampler. Journal of the American Statistical Association 87, 861-868.

[46] Schwert, G.W., 1990. Stock volatility and the crash of '87. Review of Financial Studies 3, 77-102.

[47] Seo, B., 1998. Tests for structural change in cointegrated system. Econometric Theory 14, 222-259.

[48] Shiller, R.J., 1990. The term structure of interest rates. In: Friedman, B., Hahn, F. (eds), Handbook of monetary economics, vol. 1., North-Holland, Amsterdam, 627-722.

[49] Strachan, R., 2003. Valid Bayesian estimation of the cointegrating error correction model. Journal of Business and Economic Statistics 21, 185-195.

[50] Strachan, R., van Dijk, H.K., 2003. Bayesian model selection with an uninformative prior. Oxford Bulletin of Economics and Statistics 65, 863-876.

[51] Strachan, R., Inder, B., 2004. Bayesian analysis of the error correction model. Journal of Econometrics 123, 307-325.

[52] Yao, Y-C., 1988. Estimating the number of changepoints via Schwarz criterion. Statistics and Probability Letters 6, 181-189.

[53] Zivot, E., Andrews, D.W.K., 1992. Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. Journal of Business and Economic Statistics 10, 251-270.

[54] Wang, J., Zivot, E., 2000. A Bayesian time series model of multiple structural changes in level, trend, and variance. Journal of Business and Economic Statistics 18, 3, 374-386.

Table 1: Monte Carlo Results for VAR Models: Average Posterior Probabilities

DGP\#.of breaks	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
DGP 1	0.942	0.057	0.001	0.000	0.000
DGP 2	0.000	0.013	0.945	0.042	0.000
DGP 3	0.000	0.000	0.995	0.004	0.000
DGP 4	0.000	0.000	0.967	0.033	0.000
DGP 5	0.000	0.008	0.981	0.011	0.000

Table 2: Monte Carlo Mean of the Mode of the Posterior for the Break Points when $m=2$: VAR Models

()=Monte Carlo standard deviation

	DGP 2	DGP 3	DGP 4	DGP 5
1st break	99.571 (3.092)	100.06 (1.635)	99.987 (2.216)	100.03 (1.504)
2nd break	200.94 (2.237)	200.97 (1.403)	200.85 (3.093)	201.02 (1.883)

The true value of the first break is at $t = 100$, and the second is at $t = 200$.

Table 3: Parameters Given in DGPs for Monte Carlo Simulations of VECM with Two Structural Breaks

	DGP 2	DGP 3	DGP 4	DGP 5
μ_1	(-0.0551, -0.0370)	(-0.0072, 0.0250)	(-0.0389, -0.0862)	(-0.0342, 0.0355)
μ_2	(0.0746, 0.1578)	(0.0045, -0.1671)	(0.0378, 0.1123)	(-0.0017, -0.0535)
μ_3	(-0.0448, -0.2355)	(0.0421, -0.1282)	(-0.0558, -0.1406)	(-0.0227, -0.0997)
α_1	(0.0004, 0.1766)	(-0.0511, 0.0983)	(-0.0333, 0.1530)	(-0.0503, 0.1413)
α_2			(-0.2292, 0.1425)	(-0.0525, 0.0331)
α_3			(0.0118, 0.0884)	(0.0018, 0.0746)
β	(1, -0.9902)	(1, -0.9894)	(1, -0.9807)	(1, -0.9744)
σ_1	$\begin{pmatrix} 0.0764 & 0.0238 \\ 0.0238 & 0.0353 \end{pmatrix}$	$\begin{pmatrix} 0.1272 & 0.0499 \\ 0.0499 & 0.0787 \end{pmatrix}$	$\begin{pmatrix} 0.0828 & 0.0292 \\ 0.0292 & 0.1134 \end{pmatrix}$	$\begin{pmatrix} 0.3049 & 0.1010 \\ 0.1010 & 0.2131 \end{pmatrix}$
σ_2		$\begin{pmatrix} 0.0685 & 0.0196 \\ 0.0196 & 0.0252 \end{pmatrix}$		$\begin{pmatrix} 0.0865 & 0.0272 \\ 0.0272 & 0.0579 \end{pmatrix}$
σ_3		$\begin{pmatrix} 0.0112 & 0.0007 \\ 0.0007 & 0.0028 \end{pmatrix}$		$\begin{pmatrix} 0.0133 & 0.0007 \\ 0.0007 & 0.0027 \end{pmatrix}$

Table 4: Monte Carlo Results for VECM Models: Average Posterior Probabilities

DGP\ #.of breaks	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
DGP 1	0.946	0.054	0.000	0.000	0.000
DGP 2	0.084	0.030	0.875	0.030	0.000
DGP 3	0.000	0.000	0.902	0.098	0.000
DGP 4	0.085	0.108	0.807	0.000	0.000
DGP 5	0.000	0.000	0.944	0.056	0.000

Table 5: Monte Carlo Mean of the Mode of the Posterior for the Break Points when $m=2$: VECM

()=Monte Carlo standard deviation

	DGP 2	DGP 3	DGP 4	DGP 5
1st break	91.231 (21.24)	99.090 (3.335)	100.96 (17.43)	100.27 (0.793)
2nd break	188.21 (23.74)	200.33 (1.085)	203.01 (13.54)	200.82 (2.170)

The true value of the first break is at $t=100$, and the second is at $t=200$.

Table 6: Model Selection: Application 1

$p\backslash m$	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
$p=3$	0.0000	0.0000	0.0000	0.0002	0.0000
$p=4$	0.0000	0.0000	0.0000	0.9315	0.0120
$p=5$	0.0000	0.0000	0.0000	0.0344	0.0219

Note: Each element shows the posterior probability in () using Chib's (1995) method.

p: the number of the lag in a VAR

m: the number of the structural breaks

Table 7: Posterior Results of a VAR Model with $m = 3$ and $p = 4$ for Application 1

(): standard deviation

(a) Estimates of Break Points

	Posterior Mode	95% HPDR
1st break	1974:07 (0.5580)	1973:09, 1975:07
2nd break	1978:11 (0.5602)	1977:10, 1979:10
3rd break	1983:01 (0.1637)	1982:09, 1983:03

(b) Estimates of Other Parameters (Mean of the Posterior)

Parameters	sp	ip	Parameters	sp	ip
μ_1	-0.0123 (0.0170)	0.1283 (0.0573)	$sp(-2)$	-0.2612 (0.0228)	0.0212 (0.0217)
μ_2	0.0543 (0.0302)	0.2197 (0.0748)	$ip(-2)$	0.0590 (0.0081)	0.0762 (0.0221)
μ_3	-0.0488 (0.0952)	0.0823 (0.0442)	$sp(-3)$	0.0410 (0.0272)	-0.1705 (0.0213)
μ_4	0.0773 (0.0071)	0.1058 (0.l0133)	$ip(-3)$	0.0113 (0.0067)	0.0129 (0.0114)
$sp(-1)$	1.1967 (0.0131)	0.0675 (0.0119)	$sp(-4)$	-0.0292 (0.0157)	0.1290 (0.0133)
$ip(-1)$	0.0049 (0.0071)	1.0700 (0.0163)	$ip(-4)$	-0.0718 (0.0049)	-0.2303 (0.0075)

$$\Omega_1 = \begin{bmatrix} 0.1928 & 0.0322 \\ (0.0424) & (0.0221) \\ 0.0322 & 1.2628 \\ (0.0221) & (0.2661) \end{bmatrix}, \Omega_2 = \begin{bmatrix} 0.0916 & 0.0599 \\ (0.0375) & (0.0210) \\ 0.0599 & 0.9991 \\ (0.0210) & (0.2195) \end{bmatrix},$$

$$\Omega_3 = \begin{bmatrix} 2.0566 & 0.4891 \\ (0.5271) & (0.1277) \\ 0.4891 & 1.3336 \\ (0.1277) & (0.2728) \end{bmatrix}, \Omega_4 = \begin{bmatrix} 0.0822 & -0.0057 \\ (0.0073) & (0.0017) \\ -0.0057 & 0.4911 \\ (0.0017) & (0.0455) \end{bmatrix}$$

Table 8: Selection of the Number of the Breaks for Application 2 by Model 1

$p\backslash m$	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
$p=2$	0.0000	0.0000	0.0000	0.0000	0.0000
$p=3$	0.0000	0.0000	0.0000	0.9973	0.0018
$p=4$	0.0000	0.0000	0.0000	0.0007	0.0000

Note: Each element shows the posterior probability in ().

p: the number of the lag in a VAR

m: the number of the structural breaks

Table 9: Selection of the Number of Ranks of Model 1 with $m=3$ and $p=3$

rank$\backslash i$	$i=1$	$i=2$	$i=3$	$i=4$
$r=0$	0.0000	0.0311	0.0000	0.0742
$r=1$	1.0000	0.9699	0.6121	0.9258
$r=2$	0.0000	0.0000	0.3880	0.0000

Note: Each element shows the posterior probability in ().

r: the number of rank

i: regime divided by the structural breaks

Table 10: Posterior Results of Model 1 with $m = 3$ and $p = 3$, and $r = 1$ for all regimes for Application 2

(a) Estimates of Break Points

	Posterior Mode	95% HPDR
1st break	1979:10 (0.0548)	1979:09, 1979:10
2nd break	1983:01 (0.0601)	1982:10, 1983:02
3rd break	1988:06 (0.1226)	1988:04, 1988:08

(b) Estimates of Other Parameters (Mean of the Posterior)

Parameters	r_l	r_s	Parameters	r_l	r_s
μ_1	0.0219 (0.0053)	-0.0043 (0.0167)	α_3	-0.0265 (0.0049)	0.0755 (0.0093)
μ_2	-0.0005 (0.0763)	0.2087 (0.0819)	α_4	-0.0006 (0.0010)	0.0171 (0.0011)
μ_3	0.0238 (0.0114)	-0.1297 (0.0132)	β_1	1	-0.9911 (0.0636)
μ_4	-0.0194 (0.0017)	-0.0289 (0.0055)	β_2	1	-0.9915 (0.0793)
α_1	-0.0100 (0.0011)	0.0368 (0.0023)	β_3	1	-1.0023 (0.0193)
α_2	-0.0831 (0.0189)	0.1610 (0.0679)	β_4	1	-0.9935 (0.0473)

$$\Omega_1 = \begin{bmatrix} 0.0407 & 0.0378 \\ (0.0058) & (0.0075) \\ 0.0378 & 0.1796 \\ (0.0075) & (0.0247) \end{bmatrix}, \Omega_2 = \begin{bmatrix} 0.4822 & 0.5426 \\ (0.1193) & (0.1407) \\ 0.5426 & 2.8110 \\ (0.1407) & (0.5828) \end{bmatrix},$$

$$\Omega_3 = \begin{bmatrix} 0.1153 & 0.0423 \\ (0.0205) & (0.0077) \\ 0.0423 & 0.1140 \\ (0.0077) & (0.0181) \end{bmatrix}, \Omega_4 = \begin{bmatrix} 0.0471 & 0.0053 \\ (0.0047) & (0.0006) \\ 0.0053 & 0.0204 \\ (0.0006) & (0.0019) \end{bmatrix},$$

Table 11: Selection of the Number of the Breaks for Application 2 by Model 2

$p\backslash m$	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
$p=2$	0.0000	0.0000	0.0000	0.0000	0.0000
$p=3$	0.0000	0.0000	0.0000	0.9719	0.0263
$p=4$	0.0000	0.0000	0.0000	0.0009	0.0009

Note: Each element shows the posterior probability.

p: the number of the lag in a VAR

m: the number of the structural breaks

Table 12: Posterior Results of Model 2 with $m = 3$ and $p = 3$, and $r = 1$ for all regimes for Application 2

(a) Estimates of Break Points

	Posterior Mode	95% HPDR
1st break	1979:10 (0.0552)	1979:09, 1979:10
2nd break	1983:01 (0.0610)	1982:11, 1983:02
3rd break	1988:06 (0.1376)	1988:04, 1988:08

(b) Estimates of Other Parameters (Mean of the Posterior)

Parameters	r_l	r_s
μ_1	0.0225 (0.0053)	-0.0061 (0.0162)
μ_2	0.0000 (0.0644)	0.2117 (0.0762)
μ_3	0.0263 (0.0173)	-0.1228 (0.0165)
μ_4	-0.0189 (0.0024)	-0.0301 (0.0071)
α_1	-0.0102 (0.0012)	0.0371 (0.0022)
α_2	-0.0846 (0.0181)	0.1672 (0.0640)
α_3	-0.0257 (0.0051)	0.0709 (0.0216)
α_4	-0.0003 (0.0013)	0.0173 (0.0014)
β	1	-0.9844 (0.0610)

$$\Omega_1 = \begin{bmatrix} 0.0406 & 0.0376 \\ (0.0058) & (0.0075) \\ 0.0376 & 0.1789 \\ (0.0075) & (0.0249) \end{bmatrix}, \Omega_2 = \begin{bmatrix} 0.4829 & 0.5438 \\ (0.1179) & (0.1412) \\ 0.5438 & 2.7944 \\ (0.1412) & (0.5808) \end{bmatrix},$$

$$\Omega_3 = \begin{bmatrix} 0.1150 & 0.0420 \\ (0.0207) & (0.0078) \\ 0.0420 & 0.1147 \\ (0.0078) & (0.0182) \end{bmatrix}, \Omega_4 = \begin{bmatrix} 0.0472 & 0.0053 \\ (0.0047) & (0.0006) \\ 0.0053 & 0.0205 \\ (0.0006) & (0.0019) \end{bmatrix}$$

Figure 1: The Interest Rates Spread (sp) and the US Industrial Production Growth Rate (ip)

solid line - the interest rates spread, dotted line - the US industrial production growth rate

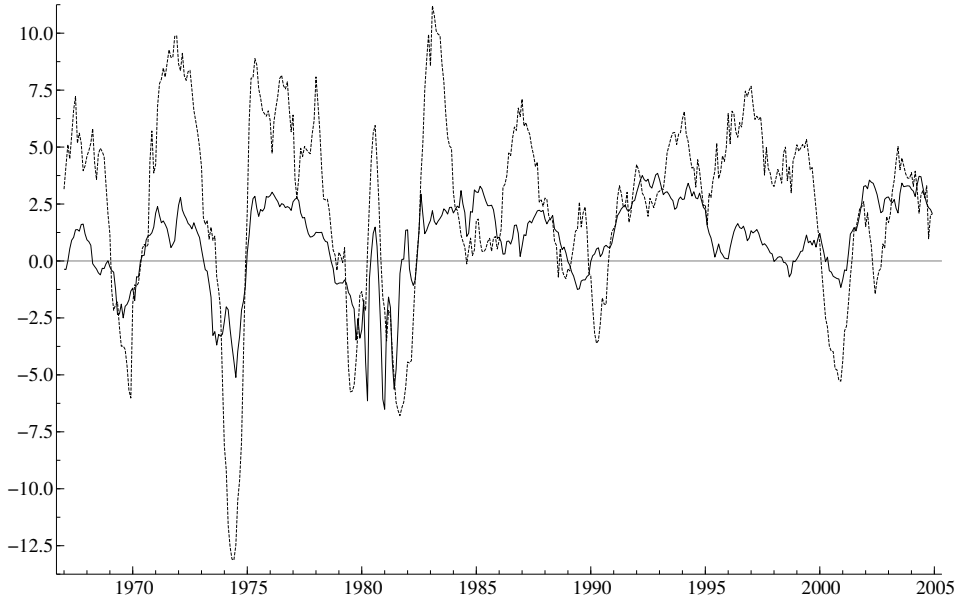

Figure 2: Posterior Probability Mass of the Break Dates for Application 1

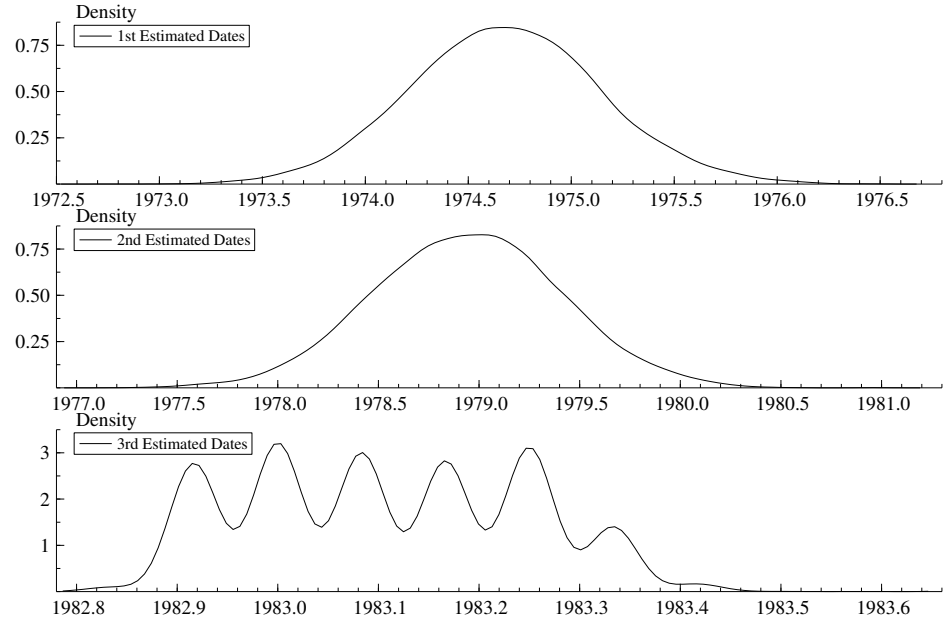

Figure 3: US Long- and Short-Term Interest Rates

solid line - 10-year TB rate, dotted line - 3-month TB rate

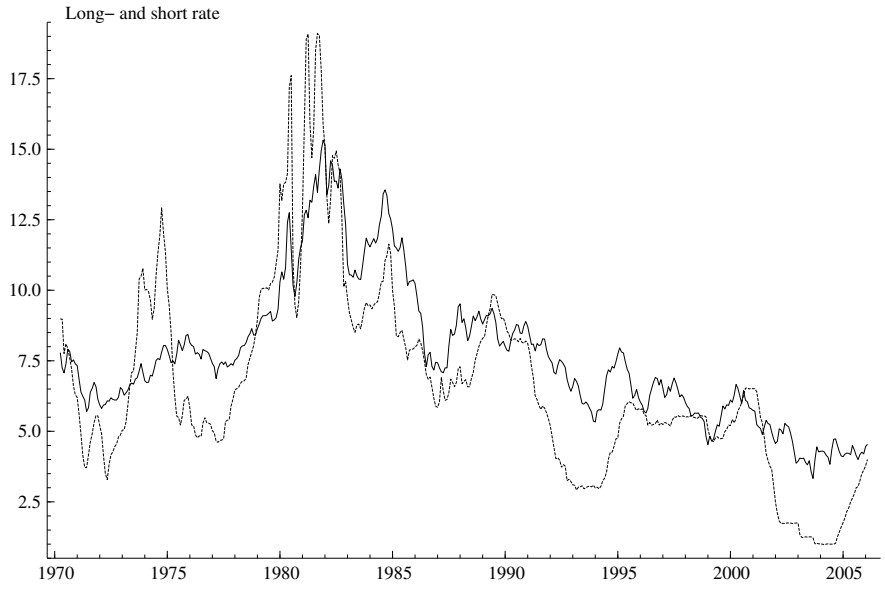

Figure 4: Posterior Probability Mass of the Break Dates for Application 2 - Model 1

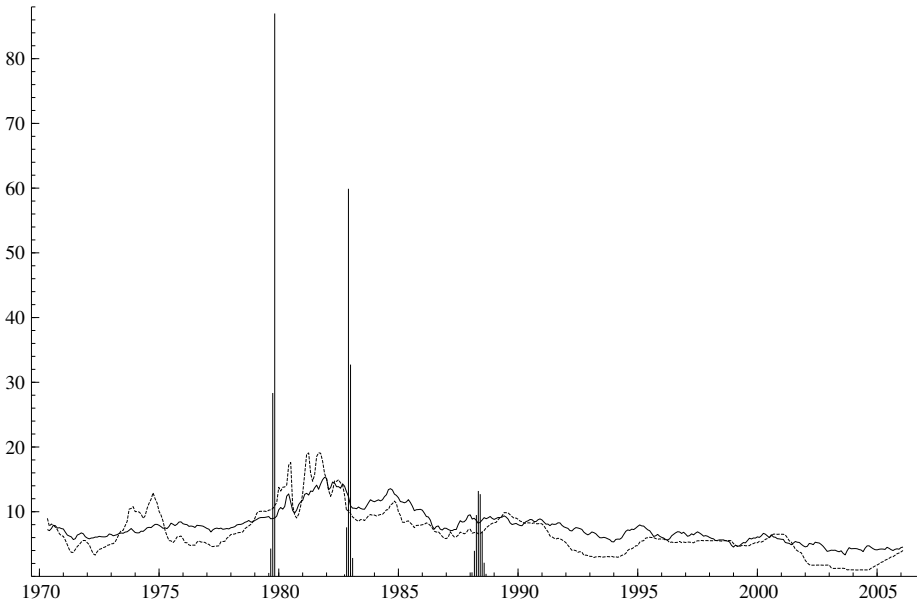

Figure 5: Posterior Density of α for Application 2 - Model 1

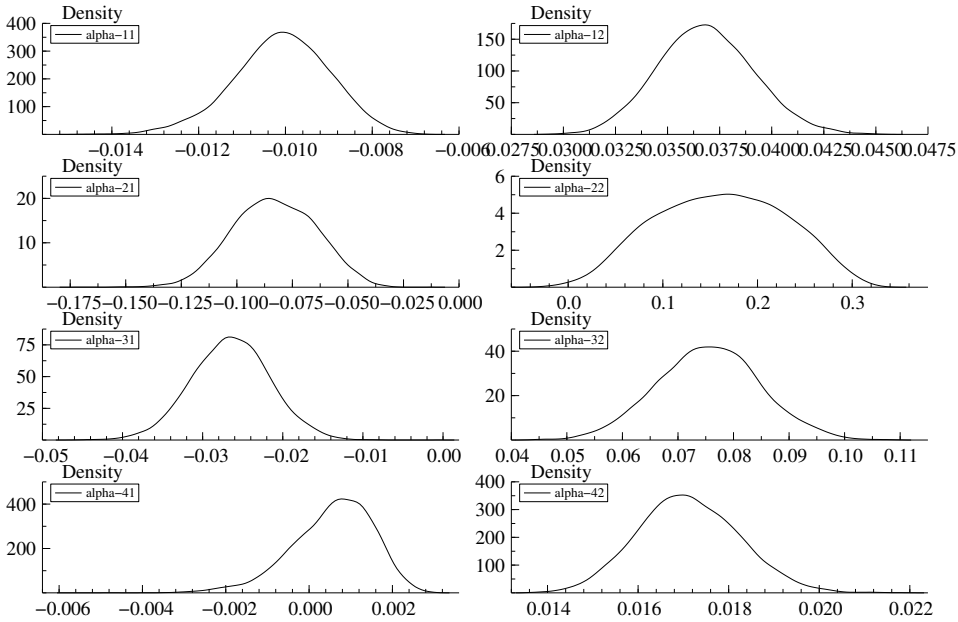

Figure 6: Posterior Probability Mass of the Break Dates for Application 2 - Model 2

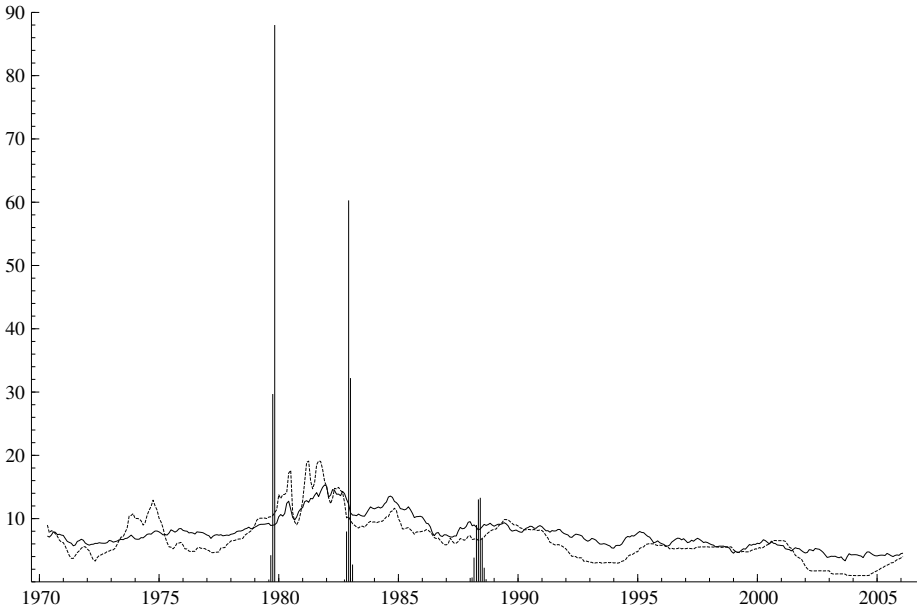

Figure 7: Posterior Density of α for Application 2 - Model 2