E C O N O M I C S   B U L L E T I N

Bayesian analysis of a vector autoregressive model with multiple structural breaks

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

Abstract

This paper develops a Bayesian approach for analyzing a vector autoregressive model with multiple structural breaks based on MCMC simulation methods, extending a method developed for the univariate case by Wang and Zivot (2000). It derives the conditional posterior densities using an independent Normal-Wishart prior. The number of structural breaks is chosen by the posterior model probability based on the marginal likelihood, calculated here by the method of Chib (1995) rather than the Gelfand-Dey (1994) method used by Wang and Zivot. Monte Carlo simulations demonstrate that the approach provides generally accurate estimation for the number of structural breaks as well as their locations.

Citation: Sugita, Katsuhiro, (2008) "Bayesian analysis of a vector autoregressive model with multiple structural breaks." *Economics Bulletin,* Vol. 3, No. 22 pp. 1-7
Submitted: August 9, 2007. **Accepted:** April 14, 2008.
URL: http://economicsbulletin.vanderbilt.edu/2008/volume3/EB-07C10006A.pdf

1 Introduction

This paper considers a vector autoregressive (VAR) model with multiple structural breaks, using a Bayesian approach with Markov chain Monte Carlo simulation technique. In Bayesian analysis of VAR-models without structural break, the Minnesota prior advocated by Litterman (1980) is often used. Other priors that are often used include the diffuse (or Jeffreys'), the Normal-Wishart, and the Normal-diffuse priors, see Kadiyala and Karlsson (1997) for these priors in details. In considering a VAR model with structural breaks, both the diffuse and the Normal-diffuse priors cannot be used for detecting the number of breaks since all priors must be proper in computing the marginal likelihood using a method proposed by Chib (1995). The Normal-Wishart prior can be used only if all parameters are subject to change with breaks. Therefore, in order to detect the number of breaks in a VAR model, we consider independent Normal-Wishart prior in this paper.

The paper is structured as follows. Bayesian inference for a VAR model using independent Normal-Wishart prior is described in Section 2. Section 3 considers the issue of model selection for detecting multiple structural breaks using Bayes factors calculated by using Chib's (1995) method. In Section 4, Monte Carlo simulations are presented using artificially generated data for VAR models with multiple breaks in order to examine the performances of detecting the number of breaks using our method. Section 5 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 a Vector Autoregressive Model with Multiple Structural Breaks

2.1 Statistical Model for a VAR with Multiple Structural Breaks

In this section we consider a Bayesian approach to a VAR model with multiple structural breaks. Let y_t denote a vector of n-dimensional $(1\times n)$ time series. If all parameters 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 as:

$$y_t = x_t B + \varepsilon_t \tag{2}$$

where $x_t = (x_{1,t}, x_{2,t})$, $x_{1,t} = (s_{1,t},\ldots,s_{m+1,t}, ts_{1,t},\ldots,ts_{m+1,t})$, $x_{2,t} = (s_{1,t}y'_{t-1},\ldots,s_{1,t}y'_{t-p+1},\ldots,s_{m+1,t}y'_{t-1},\ldots,s_{m+1,t}y'_{t-p+1})$, $B = (\;\mu'_1 \;\; ,\ldots, \;\; \mu'_{m+1} \;\; \delta'_1 \;\; ,\ldots, \;\; \delta'_{m+1} \;\; \Phi'_{1,1} \;\; ,\ldots, \;\; \Phi'_{p,1} \;\; ,\ldots, \;\; \Phi'_{1,m+1} \;\; ,\ldots, \;\; \Phi'_{p,m+1}\;)'$, and $s_{i,t}$ in $x_{1,t}$ and $x_{2,t}$ is an indicator variable which equals to 1 if regime is i and 0 otherwise.

From equation (2), let define the $(T-p+1)\times$n matrices $Y = (\;y'_p \;\; ,\ldots, \;\; y'_T\;)'$ and $E = (\;\varepsilon'_p \;\; ,\ldots, \;\; \varepsilon'_T)'$, and $X = (\;x'_p \;\; ,\ldots, \;\; x'_T\;)'$, then we can simplify the model as follows:

$$Y = XB + E \tag{3}$$

2.2 Prior Distributions and Likelihood Functions

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 $p(b, B, \Omega_1, \Omega_2, \ldots, \Omega_{m+1}) = p(b)\, p(vec(B)) \prod_{i=1}^{m+1} p(\Omega_i)$. 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 but proper 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 $p(b) \sim \mathcal{U}(p+1, T-1)$, $\Omega_i \sim IW(\psi_{0,i}, \nu_{0,i})$, $vec(B) \sim MN(vec(B_0), V_0)$ 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 $V_0 \in \mathbb{R}^{\kappa n \times \kappa n}$.

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,

$$
\begin{aligned}
&\mathfrak{L}(b, B, \Omega_1, \ldots, \Omega_{m+1} \mid Y) \\
&\propto \left(\prod_{i=1}^{m+1} |\Omega_i|^{-t_i/2} \right) \exp\left(-\frac{1}{2} \mathrm{tr} \left[\sum_{i=1}^{m+1} \left\{ \Omega_i^{-1} (Y_i - X_i B)' (Y_i - X_i B) \right\} \right] \right) \\
&= \left(\prod_{i=1}^{m+1} |\Omega_i|^{-t_i/2} \right) \exp\left(-\frac{1}{2} \sum_{i=1}^{m+1} \left[(vec(Y_i - X_i B))' (\Omega_i \otimes I_{t_i})^{-1} (vec(Y_i - X_i B)) \right] \right) \qquad (4)
\end{aligned}
$$

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

The joint posterior distribution can be obtained from the joint priors multiplied by the likelihood function in (4), that is,

$$
\begin{aligned}
&p(b, B, \Omega_1, \ldots, \Omega_{m+1} \mid Y) \propto p(b, B, \Omega_1, \ldots, \Omega_{m+1})\, \mathfrak{L}(b, B, \Omega_1, \ldots, \Omega_{m+1} \mid Y) \\
&\propto \left(\prod_{i=1}^{m+1} \left\{ |\psi_{0,i}|^{\nu_{0,i}/2} |\Omega_i|^{-(t_i + \nu_{0,i} + n + 1)/2} \right\} \right) |V_0|^{-1/2} \\
&\times \exp\left(-\frac{1}{2} \left[\mathrm{tr}\left(\sum_{i=1}^{m+1} \Omega_i^{-1} \right) + \sum_{i=1}^{m+1} \left\{ \left([vec(Y_i - X_i B)]' (\Omega_i \otimes I_{t_i})^{-1} vec(Y_i - X_i B) \right) \right\} \right.\right. \\
&\left.\left. + vec(B - B_0)' V_0^{-1} vec(B - B_0) \right] \right) \qquad (5)
\end{aligned}
$$

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 $b_i \in [b_{i-1}, b_{i+1}]$,

$$
p(b_i \mid [b - b_i], B, \Omega_1, \ldots, \Omega_{m+1}, Y) \propto p(b_i \mid b_{i-1}, b_{i+1}, B, \Omega_i, \Omega_{i+1}, Y_i) \qquad (6)
$$

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 drawn from multinomial distribution as

$$b_i \sim \mathcal{M}(b_{i+1} - b_{i-1}, p_{\mathcal{L}}) \tag{7}$$

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.*□

From (5), we can write two terms using the above theorem as:

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

where

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

Thus, with the above theorem, 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:

$$p(\Omega_i \mid b, B, Y) = C_{IW}^{-1} |\Omega_i|^{-(t_i + \nu_i + n + 1)/2} \exp\left[-\frac{1}{2} \mathrm{tr}\left(\Omega_i^{-1} \Psi_{i,\star}\right) \right] \tag{8}$$

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,

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

where

$$vec(B_\star) = \left[V_0^{-1} + \sum_{i=1}^{m+1} \left\{ \Omega_i^{-1} \otimes (X_i' X_i) \right\} \right]^{-1} \left[V_0^{-1} vec(B_0) + \sum_{i=1}^{m+1} \left\{ (\Omega_i \otimes I_\kappa)^{-1} vec(X_i' Y_i) \right\} \right], \tag{9}$$

and

$$V_B = \left[V_0^{-1} + \sum_{i=1}^{m+1} \left\{ \Omega_i^{-1} \otimes \left(X_i' X_i \right) \right\} \right]^{-1} \tag{10}$$

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 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}:

$$PO_{ij} = \frac{p(\mathcal{M}_i \mid Y)}{p(\mathcal{M}_j \mid Y)} = \frac{p(Y \mid \mathcal{M}_i)}{p(Y \mid \mathcal{M}_j)} \cdot \frac{p(\mathcal{M}_i)}{p(\mathcal{M}_j)} = BF_{ij} \cdot \frac{p(\mathcal{M}_i)}{p(\mathcal{M}_j)} \tag{11}$$

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

$$\Pr(\mathcal{M}_i \mid Y) = \frac{PO_{ij}}{\sum_{m=1}^{J} PO_{mj}} \tag{12}$$

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

$$p(Y \mid \mathcal{M}_i) = \frac{p(Y \mid \theta_i)p(\theta_i)}{p(\theta_i \mid Y)}. \tag{13}$$

where $p(Y \mid \theta_i)$ is the likelihood for Model i evaluated at θ_i, $p(\theta_i)$ is the prior density and $p(\theta_i \mid Y)$ is the posterior density. Thus, for any value of $\theta_i^\star$ which is the Gibbs output or the posterior mean of θ_i, the marginal likelihood can be estimated using (13). 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

$$\widehat{p}(\theta^\star \mid Y) = \prod_{i=1}^{l} \left\{ \frac{1}{N} \sum_{j=1}^{N} p(\theta_i^\star \mid Y, \varphi_{i-1}^\star, \varphi^{i+1,(j)}) \right\}. \tag{14}$$

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.

For a VAR model with multiple structural breaks, we adopt Chib's (1995) method to compute marginal likelihood $p(y \mid \mathcal{M}_i)$ to determine the number of structural breaks.

4 Simulation

In this section Monte Carlo simulation is conducted to examine the performance of the approach outlined in the previous sections. 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 is unknown and the estimation of the location of the breaks when the number of breaks is correctly specified. The following five data generation processes (DGPs) of two-variable VAR models with two structural breaks are considered:

DGP 1: $y_t = \mu_1 + y_{t-1}\Phi_1 + \sigma_1\varepsilon_t$
DGP 2: $y_t = \mu_t + y_{t-1}\Phi_1 + \sigma_1\varepsilon_t$
DGP 3: $y_t = \mu_t + y_{t-1}\Phi_1 + \sigma_t\varepsilon_t$
DGP 4: $y_t = \mu_t + y_{t-1}\Phi_t + \sigma_1\varepsilon_t$
DGP 5: $y_t = \mu_t + y_{t-1}\Phi_t + \sigma_t\varepsilon_t$
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 $100 \le t < 200$, $\mu_t = \mu_3 = (0.1, 0.1)$, $\Phi_t = \Phi_3 = -0.2I_2$, $\sigma_t = \sigma_3 = 0.02I_2$, for $200 \le t \le 300$. 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

Table 1: Monte Carlo results: average posterior probabilities

DGP\no.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 $m=2$

()=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 $t=200$.

the variance-covariance prior, $B_0=0$ and $V_0=100\times I_{n\kappa}$ 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 $\frac{Pr(m=i)}{Pr(m=0)}=1$[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 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 all close to the true values, $b=(100,200)$.

5 Conclusion

This paper considers a vector autoregressive model with multiple structural breaks, using a Bayesian approach with Markov chain Monte Carlo simulation technique. We derive the conditional posterior densities for multivariate system using independent normal-Wishart prior. The number of structural breaks is determined as a sort of model selection by the posterior odds from the values of the marginal likelihood. The Monte Carlo experiments were conducted.

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

References

[1] Bauwens, L., Lubrano, M., and J-F. Richard (1999) *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press: Oxford.

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

[3] Doornik, J.A. (1998) *OX An Object-Oriented Matrix programming Language,* Timberlake Consultants Press: London.

[4] Gelfand, A.E., and D.K.Dey (1994) "Bayesian Model Choice: Asymptotics and Exacrt Calculations" *Journal of the Royal Statistical Society* Ser. B, 56, 501-514.

[5] Inclan, C. (1993) "Detection of Multiple Structural Changes of Variance Using Posterior Odds" *Journal of Business and Economic Statistics* 11, 289-300.

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

[7] Litterman, R.B. (1980) "A Bayesian Procedure for Forecasting with Vector Autoregressions", mimeo, Massachusetts Institute of Technology.

[8] Wang, J., and E. Zivot (2000) "A Bayesian Time Series Model of Multiple Structural Changes in Level" Trend, and Variance, *Journal of Business and Economic Statistics*18 3, 374-386.

Appendix. 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

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

with the likelihood

$$
\mathfrak{L}(B,\Omega \mid Y) \propto |\Omega|^{-T/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(Y-XB)'(Y-XB)\right\}\right] \tag{15}
$$

so that the joint posterior is

$$
\begin{aligned}
p(vec(B),\Omega \mid Y) \quad \propto \quad & p(vec(B),\Omega)\mathfrak{L}(B,\Omega \mid Y) \\
& |\Psi_0|^{\nu_0/2} |\Omega|^{-(T+\nu_0+n+1)/2} |V_0|^{-1/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left((Y-XB)'(Y-XB)+\Psi_0\right)\right\}\right] \\
& \times \exp\left[-\frac{1}{2}\left\{vec(B-B_0)' V_0^{-1} vec(B-B_0)\right\}\right].
\end{aligned} \tag{16}
$$

From the joint posterior (16), it is easy to derive the conditional posterior density for Ω, which is the inverted Wishart density $IW(\Psi_\star, \nu_\star)$ as

$$
\begin{aligned}
p(\Omega \mid B,Y) \quad &= \quad \frac{p(B,\Omega \mid Y)}{p(B \mid Y)} \propto p(B,\Omega \mid Y) \\
&\propto \quad |\Omega|^{-(T+\nu_0+n+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left((Y-XB)'(Y-XB)+\Psi_0\right)\right\}\right] \\
&= \quad |\Omega|^{-(T+\nu_0+n+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Omega^{-1}\Psi_\star\right)\right]
\end{aligned} \tag{17}
$$

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

$$
\begin{aligned}
\mathfrak{L}(B,\Omega \mid Y) \quad &\propto \quad |\Omega|^{-T/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(Y-XB)'(Y-XB)\right\}\right] \\
&\propto \quad |\Omega|^{-T/2} \exp\left[-\frac{1}{2}\left(vec(Y-XB)\right)' \left(\Omega \otimes I_T\right)^{-1} \left(vec(Y-XB)\right)\right]
\end{aligned} \tag{18}
$$

can be used for obtaining the joint posterior density instead of (15) as:

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

The key term in the third line of the joint posterior density in the above equation can be written as:

$$(vec(Y-XB))'(\Omega\otimes I_T)^{-1}(vec(Y-XB))+(vec(B-B_0))'V_0^{-1}vec(B-B_0)$$
$$=(vec(B-B_\star))'V_B^{-1}vec(B-B_\star)+Q \quad (19)$$

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 (19), first rewrite the LHS of equation (19) as:

$$\begin{aligned} LHS &= (vec(Y-XB))'(\Omega\otimes I_T)^{-1}(vec(Y-XB))+(vec(B-B_0))'V_0^{-1}vec(B-B_0) \\ &= (vec(Y))'(\Omega^{-1}\otimes I_T)vec(Y)+(vec(XB))'(\Omega^{-1}\otimes I_T)vec(XB)-2(vec((Y))'(\Omega^{-1}\otimes I_T)vec(XB) \\ &\quad +(vec(B))'V_0^{-1}vec(B)+(vec(B_0))'V_0^{-1}vec(B_0)-2(vec(B_0))'V_0^{-1}vec(B). \end{aligned} \quad (20)$$

The RHS can be written as:

$$\begin{aligned} RHS &= (vec(B-B_\star))'V_B^{-1}vec(B-B_\star)+(vec(Y))'(\Omega\otimes I_T)^{-1}vec(Y) \\ &\quad +(vec(B_0))'V_0^{-1}vec(B_0)-(vec(B_\star))'V_B^{-1}vec(B_\star) \\ &= (vec(B))'V_B^{-1}vec(B)-2(vec(B_\star))'V_B^{-1}vec(B) \\ &\quad +(vec(Y))'(\Omega^{-1}\otimes I_T)vec(Y)+(vec(B_0))'V_0^{-1}vec(B_0). \end{aligned} \quad (21)$$

So, from (20) and (21), $LHS-RHS$ is

$$\begin{aligned} LHS-RHS &= (vec(XB))'(\Omega^{-1}\otimes I_T)vec(XB)+(vec(B))'V_0^{-1}vec(B)-(vec(B))'V_B^{-1}vec(B) \\ &\quad -2\left\{(vec(Y))'(\Omega^{-1}\otimes I_T)vec(XB)+(vec(B_0))'V_0^{-1}vec(B)-(vec(B_\star))'V_B^{-1}vec(B)\right\} \\ &= \mathfrak{C}-2\mathfrak{D} \end{aligned} \quad (22)$$

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

$$\mathfrak{C}=(vec(XB))'(\Omega^{-1}\otimes I_T)vec(XB)+(vec(B))'V_0^{-1}vec(B)-(vec(B))'V_B^{-1}vec(B) \quad (23)$$

$$\mathfrak{D}=(vec(Y))'(\Omega^{-1}\otimes I_T)vec(XB)+(vec(B_0))'V_0^{-1}vec(B)-(vec(B_\star))'V_B^{-1}vec(B). \quad (24)$$

By substituting $V_B=\left[V_0^{-1}+\left\{\Omega^{-1}\otimes(X'X)\right\}\right]^{-1}$, the third term of $\mathfrak{C}$ in (23) is

$$\begin{aligned} (vec(B))'V_B^{-1}vec(B) &= (vec(B))'\left[V_0^{-1}+\left\{\Omega^{-1}\otimes(X'X)\right\}\right]vec(B) \\ &= (vec(B))'V_0^{-1}vec(B)+(vec(B))'\left[\Omega^{-1}\otimes(X'X)\right]vec(B) \\ &= (vec(B))'V_0^{-1}vec(B)+(vec(B))'vec\left[(X'X)B\Omega^{-1}\right]. \end{aligned} \quad (25)$$

Using (25) in (23), 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 (24) is

$$\begin{aligned} (vec(Y))'(\Omega^{-1}\otimes I_T)vec(XB) &= (vec(Y))'(\Omega^{-1}\otimes I_T)(I_n\otimes X)vec(B) \\ &= (vec(Y))'(\Omega^{-1}\otimes X)vec(B). \end{aligned} \quad (26)$$

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,

$$\begin{aligned}
(vec(B_\star))' V_B^{-1} vec(B) &= \left[V_0^{-1}vec(B_0) + vec(X'Y\Omega^{-1})\right]' vec(B) \\
&= (vec(B_0))' V_0^{-1} vec(B) + \left[(\Omega^{-1} \otimes X')vec(Y)\right]' vec(B) \\
&= (vec(B_0))' V_0^{-1} vec(B) + (vec(Y))' (\Omega^{-1} \otimes X) vec(B). \qquad (27)
\end{aligned}$$

Thus, with (26) and (27), 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 (19) 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 $vec(B) \mid \Omega, Y \sim MN(vec(B_\star), V_B)$.