Generalized Methods for Markov-Switching Models with Restricted Transition Matrices

Christopher Sims, Dan Waggoner, and Tao Zha

Princeton University and Federal Reserve Bank of Atlanta

9 September 2006

The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta nor the Federal Reserve System

Outline

Outline

- A generalized Markov-switching model with independent state variables and restricted transition matrices.

Outline

- A generalized Markov-switching model with independent state variables and restricted transition matrices.
- The block-wise algorithm versus the EM algorithm.

Outline

- A generalized Markov-switching model with independent state variables and restricted transition matrices.
- The block-wise algorithm versus the EM algorithm.
- The new implementation of the modified harmonic mean (MHM) procedure for computing posterior odds ratios.

Outline

- A generalized Markov-switching model with independent state variables and restricted transition matrices.
- The block-wise algorithm versus the EM algorithm.
- The new implementation of the modified harmonic mean (MHM) procedure for computing posterior odds ratios.
- Some *tentative* results with vector autoregession model.

Why Markov-switching models

Why Markov-switching models

- Allows one to begin to address questions concerning changes in agents behavior such as the current debate on whether monetary policy and the private sector's behavior have significantly changed in recent history. (Cogley and Sargent (2002, 2005), Canova and Gambetti (2004), Beyer and Farmer (2004), Primiceri (2005), Sims and Zha (2006)).

Why Markov-switching models

- Allows one to begin to address questions concerning changes in agents behavior such as the current debate on whether monetary policy and the private sector's behavior have significantly changed in recent history. (Cogley and Sargent (2002, 2005), Canova and Gambetti (2004), Beyer and Farmer (2004), Primiceri (2005), Sims and Zha (2006)).
- Does not require *that* much more work than the constant parameter case.

Markov-Switching Model

- y_t – a vector of endogenous variables;
- z_t – a vector of exogenous variables;
- s_t – a Markov state variable with finite state space H;
- $Q = (q_{i,j})$ – the transition matrix where $q_{i,j}$ is the probability of transition from state j to i;
- θ – all the other model parameters.
- The likelihood of Y_T is

$$p\left(Y_T \mid Z_T, \theta, Q\right) = \prod_{t=1}^{T} \left[\sum_{s_t \in H} p\left(y_t \mid Y_{t-1}, Z_t, \theta, s_t\right) p\left(s_t \mid Y_{t-1}, Z_{t-1}, \theta, Q\right) \right],$$

  which can be evaluated recursively.
- To evaluate the likelihood the *only* computation needed is

$$p\left(y_t \mid Y_{t-1}, Z_t, \theta, s_t\right)$$

The Gibbs Sampler

When the number of the model parameters is large, the MCMC simulation directly from $p(\theta, Q \mid Y_T, Z_T)$ can be inefficient and problematic. One can obtain the empirical joint posterior density $p(\theta, Q, S_T \mid Y_T, Z_T)$ by sampling alternatively from

The Gibbs Sampler

When the number of the model parameters is large, the MCMC simulation directly from $p(\theta, Q \mid Y_T, Z_T)$ can be inefficient and problematic. One can obtain the empirical joint posterior density $p(\theta, Q, S_T \mid Y_T, Z_T)$ by sampling alternatively from

- $p(S_T \mid Y_T, Z_T, \theta, Q)$, – can be evaluated recursively,

The Gibbs Sampler

When the number of the model parameters is large, the MCMC simulation directly from $p(\theta, Q \mid Y_T, Z_T)$ can be inefficient and problematic. One can obtain the empirical joint posterior density $p(\theta, Q, S_T \mid Y_T, Z_T)$ by sampling alternatively from

- $p(S_T \mid Y_T, Z_T, \theta, Q)$, – can be evaluated recursively,
- $p(Q \mid Y_T, Z_T, S_T, \theta)$, – Dirichlet with Dirichlet prior,

The Gibbs Sampler

When the number of the model parameters is large, the MCMC simulation directly from $p(\theta, Q \mid Y_T, Z_T)$ can be inefficient and problematic. One can obtain the empirical joint posterior density $p(\theta, Q, S_T \mid Y_T, Z_T)$ by sampling alternatively from

- $p(S_T \mid Y_T, Z_T, \theta, Q)$, – can be evaluated recursively,
- $p(Q \mid Y_T, Z_T, S_T, \theta)$, – Dirichlet with Dirichlet prior,
- $p(\theta \mid Y_T, Z_T, Q, S_T)$, – evaluation is model-dependent and no more difficult than drawing from the constant parameter model.

Restrictions on Q

Restrictions on q_j, the j^{th} column of Q, are of the form

$$\begin{bmatrix} q_1 \\ \vdots \\ q_h \end{bmatrix} = \underbrace{\begin{bmatrix} M_{1,1} & \cdots & M_{1,\nu} \\ \vdots & \ddots & \vdots \\ M_{h,1} & \cdots & M_{h,\nu} \end{bmatrix}}_{M} \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_\nu \end{bmatrix}$$

where $\omega_j = \left(\omega_{1,j}, \cdots, \omega_{r_j,j}\right)$ with $\omega_{i,j} \geq 0$ and $\sum_{i=1}^{r_j} \omega_{i,j} = 1$. Also, the following two conditions on M must hold.

- Each element of M is non-negative and each row of M has at most one positive element.
- The sum of any column of $M_{i,j}$ is equal to $\lambda_{i,j}$ and $\sum_{j=1}^{\nu} \lambda_{i,j} = 1$.

Restrictions on Q (Example)

Zero restrictions and asorbing state.

$$Q = \begin{bmatrix} q_{1,1} & q_{1,2} & 0 \\ q_{2,1} & q_{2,2} & 0 \\ 0 & q_{3,2} & 1 \end{bmatrix}.$$

Here $\nu = 3$ with $r_1 = 2$, $r_2 = 3$, and $r_3 = 1$. M is block diagonal with

$$M_{1,1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, M_{2,2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ and } M_{3,3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Symmetric jumping from the second state ($q_{1,2} = q_{3,2}$) can be obtained by letting $r_2 = 2$ and

$$M_{2,2} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \\ 1/2 & 0 \end{bmatrix}.$$

Block Optimization Algorithm

Block Optimization Algorithm

- It is essential to obtain the estimate of θ at the peak of the posterior distribution (e.g. starting point for MCMC and center for weighting distribution of MHM method).

Block Optimization Algorithm

- It is essential to obtain the estimate of θ at the peak of the posterior distribution (e.g. starting point for MCMC and center for weighting distribution of MHM method).
- EM algorithms – Hamilton (1994) and Chib (1996). Computationally expensive in VAR models.

Block Optimization Algorithm

- It is essential to obtain the estimate of θ at the peak of the posterior distribution (e.g. starting point for MCMC and center for weighting distribution of MHM method).
- EM algorithms – Hamilton (1994) and Chib (1996). Computationally expensive in VAR models.
- The ordinary BFGS minimization algorithm – Sims (2001). When the number of parameters is large, the Hessian update becomes inaccurate and the algorithm is likely to be unreliable.

Block Optimization Algorithm

- It is essential to obtain the estimate of θ at the peak of the posterior distribution (e.g. starting point for MCMC and center for weighting distribution of MHM method).
- EM algorithms – Hamilton (1994) and Chib (1996). Computationally expensive in VAR models.
- The ordinary BFGS minimization algorithm – Sims (2001). When the number of parameters is large, the Hessian update becomes inaccurate and the algorithm is likely to be unreliable.
- We use the Gibbs-sampling idea to break the parameters θ, Q into two blocks of parameters θ and Q. One can further break the block of parameters θ into sub-blocks. Apply the BFGS algorithm to each block alternately.

The MHM Theorem

The marginal data density (MDD) is defined as

$$\pi(Y_T) = \int \pi(Y_T \mid \theta)\pi(\theta)\, d\theta.$$

Let $h(\theta)$ be a proper density and define

$$m(\theta) = \frac{h(\theta)}{\pi(Y_T \mid \theta)\pi(\theta)}.$$

Since

$$\pi(Y_t)^{-1} = \int m(\theta)\pi(\theta \mid Y_T) d\theta,$$

if $\theta^{(i)}$ for $1 \leq i \leq N$ are draws from the posterior, the MDD can be approximated by

$$\hat{\pi}(Y_T)^{-1} = \frac{1}{N}\sum_{i=1}^{N} m(\theta^{(i)}),$$

The usual MHM method

- The proper density $h(\theta)$ is taken to be a truncated normal pdf (or some other pdf with bounded support that can be easily evaluated).

The usual MHM method

- The proper density $h(\theta)$ is taken to be a truncated normal pdf (or some other pdf with bounded support that can be easily evaluated).
- The distribution is centered at the estimated posterior mean and scaled by some multiple of the estimated posterior variance.

The usual MHM method

- The proper density $h(\theta)$ is taken to be a truncated normal pdf (or some other pdf with bounded support that can be easily evaluated).
- The distribution is centered at the estimated posterior mean and scaled by some multiple of the estimated posterior variance.
- The level of truncation and the scale multiple can be varied to get the best results.

The usual MHM method

- The proper density $h(\theta)$ is taken to be a truncated normal pdf (or some other pdf with bounded support that can be easily evaluated).
- The distribution is centered at the estimated posterior mean and scaled by some multiple of the estimated posterior variance.
- The level of truncation and the scale multiple can be varied to get the best results.
- First suggestion: when the posterior mean by be in a low density region, center at the posterior mode.

A different truncation

- When the posterior can be very small or zero at interior points, the usual MHM method can become inefficient. In high dimensions when the weighting density can be very peaked, a similar problem can occur.

A different truncation

- When the posterior can be very small or zero at interior points, the usual MHM method can become inefficient. In high dimensions when the weighting density can be very peaked, a similar problem can occur.
- Second suggestion: simply truncate these points away.

A different truncation

- When the posterior can be very small or zero at interior points, the usual MHM method can become inefficient. In high dimensions when the weighting density can be very peaked, a similar problem can occur.
- Second suggestion: simply truncate these points away.
- For $L_1, L_2 \in (0, \infty)$ define

$$\theta(L_1, L_2) = \{\theta \mid \pi(Y_T \mid \theta)\pi(\theta) > L_1 \text{ and } h(\theta) < L_2\}$$

A different truncation

- When the posterior can be very small or zero at interior points, the usual MHM method can become inefficient. In high dimensions when the weighting density can be very peaked, a similar problem can occur.

- Second suggestion: simply truncate these points away.

- For $L_1, L_2 \in (0, \infty)$ define

$$\theta(L_1, L_2) = \{\theta \mid \pi(Y_T \mid \theta)\pi(\theta) > L_1 \text{ and } h(\theta) < L_2\}$$

- Let $h^*(\theta) = c1_{\theta(L_1,L_2)}(\theta)h(\theta)$ where $1_{\theta(L_1,L_2)}(\theta)$ is the indicator function and c is the constant that makes h^* a proper density.

A different truncation (continued)

- How does one compute c? Since it is usually possible to obtain iid draws from h, these can be used to accurately compute c as long as it is not too small. Since the draws are iid, one can easily obtain estimates ot the accuracy of c.

A different truncation (continued)

- How does one compute c? Since it is usually possible to obtain iid draws from h, these can be used to accurately compute c as long as it is not too small. Since the draws are iid, one can easily obtain estimates ot the accuracy of c.
- Still want to be careful in the choice of h, as this can have a huge effect on the efficiency of the method.

Application

Consider

$$y_t = \begin{bmatrix} x_t & \pi_t & R_T \end{bmatrix}'.$$

and use the identification of Christiano, Eichenbaum, and Evans (2005).

Application

Consider

$$y_t = \begin{bmatrix} x_t & \pi_t & R_T \end{bmatrix}'.$$

and use the identification of Christiano, Eichenbaum, and Evans (2005).

- Time-varying variance of shocks only.

Application

Consider

$$y_t = \begin{bmatrix} x_t & \pi_t & R_T \end{bmatrix}'.$$

and use the identification of Christiano, Eichenbaum, and Evans (2005).

- Time-varying variance of shocks only.
- Only the interest rate equation can change (and variance of shocks).

Application

Consider

$$y_t = \begin{bmatrix} x_t & \pi_t & R_T \end{bmatrix}'.$$

and use the identification of Christiano, Eichenbaum, and Evans (2005).

- Time-varying variance of shocks only.
- Only the interest rate equation can change (and variance of shocks).
- All equations can change (and variance of shocks).

Application

Consider

$$y_t = \begin{bmatrix} x_t & \pi_t & R_T \end{bmatrix}'.$$

and use the identification of Christiano, Eichenbaum, and Evans (2005).

- Time-varying variance of shocks only.
- Only the interest rate equation can change (and variance of shocks).
- All equations can change (and variance of shocks).
- Independent state variables for equations and variances of shocks.

Application (continued)

	log MDD	% posterior	c
2v	1821.7	55%	0.14
2vm	1813.7	88%	0.000016
2vRm	1833.2	55%	0.0051
2v×2m	1857.7	55%	0.00024
2v×2Rm	1837.4	55%	0.00082
2v×3Rm	1839.5	88%	0.0000079
3v	1865.4	55%	0.00011
3v×2Rm	1863.2	55%	0.000023
4v	1868.8	55%	0.000028

Concluding Remarks

- We have developed a Bayesian method for estimating independent state variables and restricted transition matrices.

Concluding Remarks

- We have developed a Bayesian method for estimating independent state variables and restricted transition matrices.
- Our blockwise optimization method proves much more efficient than the existing EM algorithm for obtaining the MLEs or posterior estimates of the model parameters.

Concluding Remarks

- We have developed a Bayesian method for estimating independent state variables and restricted transition matrices.
- Our blockwise optimization method proves much more efficient than the existing EM algorithm for obtaining the MLEs or posterior estimates of the model parameters.
- Our new MHM method deals explicitly with the problem of zero likelihood at interior points and high dimensional weighting densities with extremely high peaks.