

# Bayesian Estimation and Inference for Markov-Switching VARs in R

Matthieu Droumaguet<sup>a</sup>, Anders Warne<sup>b</sup>, Tomasz Woźniak<sup>c</sup>

<sup>a</sup>Department of Economics, European University Institute

<sup>b</sup>Directorate General Research, European Central Bank

<sup>c</sup>Department of Economics, University of Melbourne

email: [tomasz.wozniak@unimelb.edu.au](mailto:tomasz.wozniak@unimelb.edu.au)

© 2016 Matthieu Droumaguet, Anders Warne, & Tomasz Woźniak

---

## Abstract

A block Metropolis-Hastings algorithm for the Bayesian estimation of the Markov-switching Vector Autoregressive models with restrictions for Granger noncausality is provided, as well as an appropriate estimator for the marginal data density. The codes are available under the GNU General Public License v3.0. To refer to the codes in publications, please, cite the following paper:

Droumaguet, M., Warne, A., Woźniak, T. (2017) Granger Causality and Regime Inference in Markov-Switching VAR Models with Bayesian Methods, *Journal of Applied Econometrics*, 32(4), pp. 802–818.

**Keywords:** R, Markov-switching VARs, Block Metropolis-Hastings Sampler, Marginal Data Density

---

## 1. The Markov-switching VAR Model

The model under consideration is the Markov-switching Vector Autoregressive process with the restrictions for Granger noncausality. The model presented in this manuscript was used in [Droumaguet, Warne & Woźniak \(in press.a\)](#) to make inference about Granger causality between money and income data for the U.S. The exposition in this note is based on [Droumaguet, Warne & Woźniak \(in press.b\)](#).

*Notation and model.* Let  $\mathbf{y}_T = (y_1, \dots, y_T)'$  denote a time series of  $T$  observations, where each  $y_t$  is a  $N$ -variate real-valued vector for  $t \in \{1, \dots, T\}$ . We consider a class of parametric Markov mixture distribution models in which the stochastic process  $y_t$  depends on the realizations of a hidden discrete stochastic process  $s_t$  with finite state space  $\{1, \dots, M\}$  (see [Hamilton, 1989](#)). Conditioned on the state,  $s_t$ , and time series up to time  $t - 1$ , denoted by  $\mathbf{y}_{t-1}$ ,  $y_t$  follows an independent identical normal distribution. Specifically, we let:

$$y_t = \mu_{s_t} + A_{s_t}^{(1)} y_{t-1} + \dots + A_{s_t}^{(p)} y_{t-p} + \epsilon_t, \quad (1)$$

$$\epsilon_t | s_t \sim i.i.d. \mathcal{N}(\mathbf{0}, \Sigma_{s_t}), \quad (2)$$

for  $t \in \{1, \dots, T\}$ . We set the vector of initial values  $\mathbf{y}_0 = (y_{p-1}, \dots, y_0)'$  to the first  $p$  observations of the available data.

The variable  $s_t$  is assumed to be an irreducible aperiodic Markov chain with  $\Pr(S_0 = i | \mathbf{P}) = \pi_i$ , where  $\pi = (\pi_1, \dots, \pi_M)$  is the ergodic distribution of the Markov Switching (MS) process. Its properties are sufficiently described by the  $(M \times M)$  transition probabilities matrix:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1M} \\ p_{21} & p_{22} & \dots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \dots & p_{MM} \end{bmatrix}, \quad (3)$$

in which an element,  $p_{ij}$ , denotes the probability of transition from state  $i$  to state  $j$ ,  $p_{ij} = \Pr(s_{t+1} = j | s_t = i)$ . The elements of each row of matrix  $\mathbf{P}$  sum to one,  $\sum_{j=1}^M p_{ij} = 1$ . Equations (1)–(3) represent a MS-VAR model with  $M$  states and  $p$  lags.

*Likelihood function.* Let  $\theta$  be a vector of size  $k$ , collecting parameters of the transition probabilities matrix  $\mathbf{P}$  and all the state-dependent parameters of the VAR process,  $\theta_{s_t}: \mu_{s_t}, A_{s_t}^{(i)}, \Sigma_{s_t}$ , for  $s_t \in \{1, \dots, M\}$  and  $i \in \{1, \dots, p\}$ . The complete-data likelihood function is equal to the joint sampling distribution  $p(\mathbf{S}_T, \mathbf{y}_T | \theta)$  for the complete data  $(\mathbf{S}_T, \mathbf{y}_T)$  given the parameters, where  $\mathbf{S}_T = (s_0, s_1, \dots, s_T)$ ; see, e.g., [Frühwirth-Schnatter \(2006\)](#). This likelihood is further decomposed into:

$$p(\mathbf{S}_T, \mathbf{y}_T | \theta) = p(\mathbf{y}_T | \mathbf{S}_T, \theta) \Pr(\mathbf{S}_T | \mathbf{P}), \quad (4)$$

where the parameters  $\theta = (\theta_1, \dots, \theta_M, \mathbf{P})$ . The two components on the right hand side of equation (4) are the same as the standard MS-VAR model of [Frühwirth-Schnatter \(2006, Section 11.3.1\)](#).

*Prior distribution.* We assume that the prior distribution of the state-specific parameters for each state and the transition probabilities matrix are independent:

$$p(\theta) = \prod_{i=1}^M p(\theta_i) p(\mathbf{P}). \quad (5)$$

For the unrestricted MS-VAR model, we assume the following prior specification. Each row of the transition probabilities matrix,  $\mathbf{P}$ , *a priori* follows an  $M$ -variate Dirichlet distribution, with the  $M \times 1$  parameter vectors  $e_m$ , for  $m \in \{1, \dots, M\}$ . Finally, collect all of the vectors of parameters into a  $M^2 \times 1$  vector  $e = (e'_1, \dots, e'_M)'$ .

Furthermore, the state-dependent parameters of the VAR process are collected in vectors

$$\beta_{s_t} = \left( \mu'_{s_t}, \text{vec}(A_{s_t}^{(1)})', \dots, \text{vec}(A_{s_t}^{(p)})' \right)',$$

for  $s_t = 1, \dots, M$ . For each  $s_t \in \{1, \dots, M\}$   $\beta_{s_t}$  follows *a priori* a  $(N + pN^2)$ -variate normal distribution, with mean equal to  $\mu_\beta$  and a covariance matrix  $V_\beta$ . Furthermore, let  $B_{s_t} = (\mu_{s_t}, A_{s_t}^{(1)}, \dots, A_{s_t}^{(p)})'$  be a  $(1 + pN) \times N$  matrix of state-specific autoregressive parameters.

Let  $\sigma_{s_t}$  be an two-dimensional vector of standard deviations and  $\mathbf{R}_{s_t}$  a  $2 \times 2$  correlation matrix. The covariance matrix of  $\epsilon_t | s_t$  can now be decomposed as:

$$\Sigma_{s_t} = \text{diag}(\sigma_{s_t}) \mathbf{R}_{s_t} \text{diag}(\sigma_{s_t}), \quad (6)$$

Modeling covariance matrices using a decomposition into standard deviations and a correlation matrix, as in equation (6), was proposed for Bayesian inference by [Barnard, McCulloch & Meng \(2000\)](#). Each standard deviation  $\sigma_{j,s_t}$  for  $s_t = 1, \dots, M$  and  $j = 1, 2$ , follows a log-normal distribution, with the location parameter equal to  $\mu_\sigma$  and the standard deviation parameter set to  $\lambda_\sigma > 0$ .

Finally, we assume that the marginal prior distribution for the below-diagonal element of the (symmetric) correlation matrix  $\mathbf{R}_{s_t}$  is a uniform distribution on the interval  $(-1, 1)$ . We collect all the standard deviations in one vector,  $\sigma = (\sigma'_1, \dots, \sigma'_M)'$ , and all the unknown correlation coefficients into a vector

$$\mathbf{R} = (\text{vecl}(\mathbf{R}_1)', \dots, \text{vecl}(\mathbf{R}_M)')',$$

where the  $\text{vecl}$  operator stacks all the lower-diagonal elements of the correlation matrix into a vector.

To summarize, the prior specification (5) now takes the detailed form of:

$$p(\theta) = \prod_{i=1}^M p(\mathbf{P}_i) p(\beta_i) p(\mathbf{R}_i) \left( \prod_{j=1}^N p(\sigma_{i,j}) \right), \quad (7)$$

where each of the prior distributions is as assumed:

$$\begin{aligned} \mathbf{P}_i &\sim \mathcal{D}_M(e_i) \\ \beta &\sim \mathcal{N}(\mu_\beta, V_\beta) \\ \sigma_{i,j} &\sim \log \mathcal{N}(\mu_\sigma, \lambda_\sigma) \\ \mathbf{R}_{i,jk} &\sim \mathcal{U}(-1, 1) \end{aligned}$$

for  $i = 1, \dots, M$  and  $j, k = 1, \dots, N$ , and  $j \neq k$ .

## 2. Block Metropolis-Hastings algorithm

This section describes all the constituting blocks that form the MCMC sampler.

### 2.1. Simulating Hidden Markov Process

The first drawn parameter is the vector representing the states of the economy,  $\mathbf{S}_T$ . We first use a filter and smoother (see Section 11.2 of [Frühwirth-Schnatter, 2006](#), and references therein) and obtain the probabilities  $\Pr(s_t = i | \mathbf{y}_T, \theta^{(l-1)})$ , for  $t \in \{1, \dots, T\}$  and  $i \in \{1, \dots, M\}$ , and then draw  $\mathbf{S}_T^{(l)}$ , for  $l^{\text{th}}$  iteration of the algorithm. For the full description of the algorithm used in this work the reader is referred to [Droumaguet & Woźniak \(2012\)](#).

### 2.2. Sampling Transition Probabilities

In this step of the MCMC sampler, we draw from the posterior distribution of the transition probabilities matrix, conditioning on the states drawn in the previous step of the current iteration,  $\mathbf{P}^{(l)} \sim p(\mathbf{P} | \mathbf{S}_T^{(l)})$ . [Sims, Waggoner & Zha \(2008\)](#) provide a flexible analytical framework for working with restricted transition probabilities, and the reader is invited to consult Section 3 of their paper for an exhaustive description of the possibilities provided by the framework. We however limit the latitude given by the reparametrization in order to ensure the stationarity of Markov chain  $\mathbf{S}_T$ .

*Reparametrization.* The transition probabilities matrix  $\mathbf{P}$  is modeled with  $Q$  vectors  $w_j$ ,  $j = 1, \dots, Q$  and each of size  $d_j \leq M$ . Let all the elements of  $w_j$  belong to the  $(0, 1)$  interval and sum up to one, and stack all of them into the column vector  $\mathbf{w} = (w'_1, \dots, w'_Q)'$  of dimension  $d = \sum_{j=1}^Q d_j$ . Writing  $\mathbf{p} = \text{vec}(\mathbf{P}')$  as a  $M^2$  dimensional column vector, and introducing the  $(M^2 \times d)$  matrix  $\mathbf{M}$ , the transition matrix is decomposed as:

$$\mathbf{p} = \mathbf{M}\mathbf{w}, \quad (8)$$

where the  $\mathbf{M}$  matrix is composed of the  $M_{ij}$  sub-matrices of dimension  $(M \times d_j)$ , where  $i \in \{1, \dots, M\}$ , and  $j \in \{1, \dots, Q\}$ :

$$\mathbf{M} = \begin{bmatrix} M_{11} & \dots & M_{1Q} \\ \vdots & \ddots & \vdots \\ M_{M1} & \dots & M_{MQ} \end{bmatrix},$$

where each  $M_{ij}$  satisfies the following conditions:

1. For each  $(i, j)$ , all elements of  $M_{ij}$  are non-negative.
2.  $\sum_M M_{ij} = \Lambda_{ij} \mathbf{1}'_{d_j}$ , where  $\Lambda_{ij}$  is the sum of the elements in any column of  $M_{ij}$ .

3. Each row of  $\mathbf{M}$  has, at most, one non-zero element.
4.  $\mathbf{M}$  is such that  $\mathbf{P}$  is irreducible: for all  $j, d_j \geq 2$ .

The first three conditions are inherited from [Sims et al. \(2008\)](#), whereas the last condition assures that  $\mathbf{P}$  is irreducible, forbidding the presence of an absorbing state that would render the Markov chain  $\mathbf{S}_T$  non-stationary. The lack of independence of the rows of  $\mathbf{P}$  is described in [Frühwirth-Schnatter \(2006, Section 11.5.5\)](#). Once the initial state  $s_0$  is drawn from the ergodic distribution  $\pi$  of  $\mathbf{P}$ , direct MCMC sampling from the conditional posterior distribution becomes impossible. However, a Metropolis-Hastings step can be set up to circumvent this issue, since a kernel of joint posterior density of all rows is known:  $p(\mathbf{P}|\mathbf{S}_T) \propto \pi \prod_{j=1}^Q \mathcal{D}_{d_j}(w_j)$ . Hence, the proposal for transition probabilities is obtained by sampling each  $w_j$  from the convenient Dirichlet distribution. We then transform the column vector  $\mathbf{w}$  into our candidate matrix of transitions probabilities using equation (8). Finally, we compute the acceptance rate before retaining or discarding the draw.

**Algorithm 1.** *Metropolis-Hastings step for the restricted transition matrix.*

1.  $s_0 \sim \pi$ . The initial state is drawn from the ergodic distribution of  $\mathbf{P}$ .
2.  $w_j \sim \mathcal{D}_{d_j}(n_{1,j} + e_{1,j}, \dots, n_{d_j,j} + e_{d_j,j})$  for  $j \in \{1, \dots, Q\}$ .  $n_{i,j}$  corresponds to the number of transitions from state  $i$  to state  $j$ , counted from  $\mathbf{S}_T$ . The candidate transition probabilities matrix – in the transformed notation – are sampled from a Dirichlet distribution.
3.  $\mathbf{P}^{new} = \mathbf{M}\mathbf{w}$ . The proposal for the transitions probabilities matrix is reconstructed.
4. Accept  $\mathbf{P}^{new}$  if  $u \leq (\pi^{new}/\pi^{l-1})$ , where  $u \sim \mathcal{U}[0, 1]$ .  $\pi^{new}$  and  $\pi^{l-1}$  are the ergodic probabilities of  $s_0^{(l)}$  that are computed from  $\mathbf{P}^{new}$  and  $\mathbf{P}^{l-1}$  respectively.

### 2.3. Sampling Correlation Coefficients and Standard Deviations

Adapting the approach proposed by [Barnard et al. \(2000\)](#) to Markov-switching models, we sample from the full conditional distribution of unrestricted and restricted covariance matrices.

**Algorithm 2.** *Griddy-Gibbs for the standard deviations.* The algorithm iterates on all the standard deviation parameters  $\sigma_{s_t, j}$  for  $i \in \{1, \dots, N\}$  and  $s_t \in \{1, \dots, M\}$ . The grid is centered on the residuals' sample standard deviation  $\hat{\sigma}_{s_t, i}$  and divides the interval  $(\hat{\sigma}_{s_t, i} - 3\hat{\sigma}_{\hat{\sigma}_{s_t, i}}, \hat{\sigma}_{s_t, i} + 3\hat{\sigma}_{\hat{\sigma}_{s_t, i}})$  into  $G$  grid points.  $\hat{\sigma}_{\hat{\sigma}_{s_t, i}}$  is an estimator of the standard error of the estimator of the sample standard deviation.

1. Regime-invariant standard deviations: Draw from the unknown univariate density

$$p(\sigma_i | \mathbf{y}_T, \mathbf{S}_T, \mathbf{P}, \beta, \sigma_{-i}, \mathbf{R}).$$

This is done by evaluating a kernel on a grid of points, using the proportionality relation, with the likelihood function times the prior:  $\sigma_i | \mathbf{y}_T, \mathbf{S}_T, \mathbf{P}, \beta, \sigma_{-i}, \mathbf{R} \propto p(\mathbf{y}_T | \mathbf{S}_T, \theta) \cdot p(\sigma_i)$ . Reconstruct the c.d.f. from the grid through deterministic integration and sample from it.

2. Regime-varying standard deviations: For all regimes  $s_t \in \{1, \dots, M\}$ , draw from the univariate density

$$p(\sigma_{s_t, i} | \mathbf{y}_T, \mathbf{S}_T, \mathbf{P}, \beta, \sigma_{s_t, -i}, \mathbf{R}),$$

evaluating a kernel thanks to the proportionality relation, with the likelihood function times the prior:  $\sigma_{s_t, i} | \mathbf{y}_T, \mathbf{S}_T, \mathbf{P}, \beta, \sigma_{s_t, -i}, \mathbf{R} \propto p(\mathbf{y}_T | \mathbf{S}_T, \theta) \cdot p(\sigma_{s_t, i})$ .

**Algorithm 3.** *Griddy-Gibbs for the correlations* The algorithm iterates on all the correlation parameters  $\mathbf{R}_{s_t, i}$  for  $i \in \{1, \dots, (N-1)N/2\}$  and  $s_t \in \{1, \dots, M\}$ . The marginal posterior density of each correlation coefficient is bounded by  $a$  from below and  $b$  from above.  $a$  and  $b$  are computed so that although correlation coefficients are sampled from the marginal posterior distributions the resulting correlation matrix is positive definite. See [Barnard et al. \(2000\)](#) for the algorithm determining  $a$  and  $b$  for each  $\mathbf{R}_{s_t, i}$ . The grid divides interval  $(a, b)$  into  $G$  grid points.

1. Depending on the restriction scheme, set correlation parameters to 0.
2. Regime-invariant correlations: Draw from the univariate density

$$p(\mathbf{R}_i | \mathbf{y}_T, \mathbf{S}_T, \mathbf{P}, \beta, \sigma, \mathbf{R}_{-i}),$$

evaluating a kernel thanks to the proportionality relation, with the likelihood function times the prior:  
 $\mathbf{R}_i | \mathbf{y}_T, \mathbf{S}_T, \mathbf{P}, \beta, \sigma, \mathbf{R}_{-i} \propto p(\mathbf{y}_T | \mathbf{S}_T, \theta) \cdot p(\mathbf{R}_i).$

3. Regime-varying correlations: For all regimes  $s_t \in \{1, \dots, M\}$ , draw from the univariate density

$$p(\mathbf{R}_{s_t, i} | \mathbf{y}_T, \mathbf{S}_T, \mathbf{P}, \beta, \sigma, \mathbf{R}_{s_t, -i}),$$

evaluating a kernel thanks to the proportionality relation, with the likelihood function times the prior:  
 $\mathbf{R}_{s_t, -i} | \mathbf{y}_T, \mathbf{S}_T, \mathbf{P}, \beta, \sigma, \mathbf{R}_{s_t, -i} \propto p(\mathbf{y}_T | \mathbf{S}_T, \theta) \cdot p(\mathbf{R}_{s_t, i}).$

In their empirical application [Droumaguet et al. \(in press.a\)](#) reported the results for fifty grid points,  $G = 50$ , for each of the standard deviation and the correlation coefficients, however, the results were practically the same also for  $G = 30$ .

#### 2.4. Sampling Vector Autoregressive Parameters

Finally, we draw the state-dependent autoregressive parameters,  $\beta_{s_t}$  for  $s_t \in \{1, \dots, M\}$ . The Bayesian parameter estimation of finite mixtures of regression models when the realizations of states is known has been precisely covered in [Frühwirth-Schnatter \(2006, Section 8.4.3\)](#). The procedure consists of estimating all the regression coefficients simultaneously by stacking them into  $\beta = (\beta_0, \beta_1, \dots, \beta_M)$ , where  $\beta_0$  is a common regression parameter for each regime, and hence is useful for the imposing of restrictions of state invariance for the autoregressive parameters. The regression model becomes:

$$y_t = Z_t \beta_0 + Z_t D_{i,1} \beta_1 + \dots + Z_t D_{i,M} \beta_M + \epsilon_t, \quad (9)$$

$$\epsilon_t \sim i.i.d. \mathcal{N}(\mathbf{0}, \Sigma_{s_t}). \quad (10)$$

We have here introduced the  $D_{i,s_t}$ , which are dummy variables taking the value 1 when the regime occurs and set to 0 otherwise. A transformation of the regressors  $Z_T$  also has to be performed in order to allow for different coefficients on the dependent variables, for instance to impose zero restrictions on parameters. In the context of VARs, [Koop & Korobilis \(2010, Section 2.2.3\)](#) detail a convenient notation that stacks all the regression coefficients on a diagonal matrix for every equation. We adapt this notation by stacking all the regression coefficients for all the states on diagonal matrix. If  $z_{n,s_t,t}$  corresponds to the row vector of  $1 + Np$  independent variables for equation  $n$ , state  $s_t$  (starting at 0 for regime-invariant parameters), and at time  $t$ , the stacked regressor  $Z_t$  will be of the following form:

$$Z_t = \text{diag}(z_{1,0,t}, \dots, z_{N,0,t}, z_{1,1,t}, \dots, z_{N,1,t}, \dots, z_{1,M,t}, \dots, z_{N,M,t}).$$

This notation enables the restriction of each parameter, by simply setting  $z_{n,s_t,t}$  to 0 where desired.

**Algorithm 4.** *Sampling the autoregressive parameters.* We assume normal prior for  $\beta$ , i.e.  $\beta \sim \mathcal{N}(\mathbf{0}, V_\beta)$ .

1. For all  $Z_t$ s, impose restrictions by setting  $z_{n,s_t,t}$  to zero accordingly.
2.  $\beta | \mathbf{y}_T, \mathbf{S}_T, \mathbf{P}, \sigma, \mathbf{R} \sim \mathcal{N}(\bar{\beta}, \bar{V}_\beta)$ . Sample  $\beta$  from the conditional normal posterior distribution, with the following parameters:

$$\bar{V}_\beta = \left( V_\beta^{-1} + \sum_{t=1}^T Z_t' \Sigma_{s_t}^{-1} Z_t \right)^{-1}$$

and

$$\bar{\beta} = \bar{V}_\beta \left( \sum_{t=1}^T Z_t' \Sigma_{s_t}^{-1} y_t \right).$$

### 3. Estimation of marginal data densities

An appropriate marginal data density estimator used for the MS-VAR model with restrictions described above is the Modified Harmonic Mean estimator proposed by Geweke (1999, 2005). See Droumaguet & Woźniak (2012) for some more details regarding the computational details for the estimator for the MS-VAR models.

### References

- Barnard, J., McCulloch, R., & Meng, X.-L. (2000). Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage. *Statistica Sinica*, 10, 1281–1311.
- Droumaguet, M., Warne, A., & Woźniak, T. (in press.a). Granger Causality and Regime Inference in Markov-Switching VAR Models with Bayesian Methods. *Journal of Applied Econometrics*, .
- Droumaguet, M., Warne, A., & Woźniak, T. (in press.b). Online Appendix: Granger Causality and Regime Inference in Markov-Switching VAR Models with Bayesian Methods. *Journal of Applied Econometrics*, .
- Droumaguet, M., & Woźniak, T. (2012). *Bayesian Testing of Granger Causality in Markov-Switching VARs*. Working Paper Series 2012/06 European University Institute Florence, Italy.
- Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer.
- Geweke, J. (1999). Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication. *Econometric Reviews*, 18, 1–73.
- Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, Inc.
- Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57, 357–384. DOI: 10.2307/1912559.
- Koop, G., & Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3, 267–358.
- Sims, C. A., Waggoner, D. F., & Zha, T. (2008). Methods for inference in large multiple-equation markov-switching models. *Journal of Econometrics*, 146, 255–274. DOI: 10.1016/j.jeconom.2008.08.023.

**Description**

A block Metropolis-Hasting sampler for MS( $M$ )-VAR( $p$ ) models.

**Usage**

```
G.MSBVAR.griddy(N, h, priors, restrictions, starting.values,
                 debug=FALSE, print.iterations=50)
```

**Arguments**

N	A positive integer specifying the number of iterations of the sampling algorithm.
h	A positive integer specifying the forecast horizon.
priors	A list of objects specifying the prior distributions. It contains the following objects: <code>Beta.bar</code> – a $N + pN^2 \times 1$ matrix of the normal prior mean $\mu_\beta$ ; <code>V.Beta.bar.m1</code> – a $N + pN^2 \times N + pN^2$ covariance matrix of the normal prior $V_\beta$ ; <code>S.mean</code> – a scalar specifying the location parameter of the log-normal prior $\mu_\sigma$ ; <code>S.sd</code> – a scalar specifying the standard deviation parameter of the log-normal prior $\lambda_\sigma$ ; <code>w</code> – a $M^2 \times 1$ matrix of the parameters of the Dirichlet distribution $e$ .
restrictions	A list of objects specifying the restrictions on the parameters. It contains the following objects: <code>Beta.0</code> , <code>Beta.St</code> , <code>Sigma.0</code> , <code>Sigma.St</code> , <code>M</code> , <code>dj</code> that are specified below. <code>Beta.0</code> – a $(1 + pN) \times N$ matrix specifying which regime-invariant autoregressive parameters of matrix $B_{s_t}$ should be estimated, where value 1 is assigned to parameters to be estimated and value 0 to parameters not to be estimated. <code>Beta.St</code> – a list of $M$ $(1 + pN) \times N$ matrices specifying which regime-dependent autoregressive parameters of matrix $B_{s_t}$ should be estimated, where value 1 is assigned to parameters to be estimated and value 0 to parameters not to be estimated. <code>Sigma.0</code> – a $N \times N$ matrix specifying which regime-invariant elements of the covariance matrix $\Sigma_{s_t}$ should be estimated, where value 1 is assigned to parameters to be estimated and value 0 to parameters not to be estimated. <code>Sigma.St</code> – a list of $MN \times N$ matrices specifying which regime-dependent elements of the covariance matrices $\Sigma_{s_t}$ should be estimated, where value 1 is assigned to parameters to be estimated and value 0 to parameters not to be estimated. <code>M</code> – a $M^2 \times d$ matrix <code>M</code> setting the restrictions on the vectorized version of the transition probability matrix <code>p</code> . <code>dj</code> – a vector of length $Q$ specifying the values of $d_j$ for $j \in \{1, \dots, Q\}$ .

starting.values	A list of objects specifying the starting values of the parameters of the model from which the algorithm is initiated. It consists of the following objects: $Y$ , $U$ , $xi$ , $Beta$ , $Beta.0$ , $Beta.St$ , $Sigma$ , $PR.TR$ , $w$ , $S$ , $R$ , $G$ that are specified below. $Y$ – a $(T + p) \times N$ matrix of data. $U$ – a $T \times N \times M$ array of state-specific error terms evaluated at the starting values of the parameters. $xi$ – a $M \times T$ matrix specifying the realizations of the regimes taking a value of 1 when the regime occurs and 0 if it does not occur in a particular period. This matrix may be filled with values NA for the initialization of the algorithm. $Beta$ – a $(1 + pN) \times N \times M$ array collecting the regime-dependent and regime-invariant starting values for parameters $B_{s_t}$ . $Beta.0$ – a $(1 + pN) \times N$ matrix of the starting values for the regime-invariant parameters $B_{s_t}$ . $Beta.St$ – a $(1 + pN) \times N \times M$ array collecting the starting values of the regime-dependent parameters $B_{s_t}$ . $Sigma$ – a $N \times N \times M$ array collecting the regime-dependent and regime-invariant starting values for parameters $\Sigma_{s_t}$ . $S$ – a $N \times M$ matrix of the starting values for the regime-dependent standard deviations $\sigma_{s_t}$ . $R$ – a $N \times N \times M$ array collecting the starting values of the regime-dependent correlation matrices $R_{s_t}$ . $PR.TR$ – a $M \times M$ matrix of the starting values for the transition probabilities matrix $P$ . $w$ – a $d \times 1$ matrix of the starting values for the unrestricted elements of the transition probabilities matrix $w$ . $G$ – a positive integer number specifying the size of the grid, $G$ , for the Griddy Gibbs sampler for parameters $\sigma_{s_t}$ and $R_{s_t}$ .
debug	A logical value set by default to FALSE. If the argument is set to TRUE the algorithm displays information regarding the subsequent steps within each of the iteration corresponding to the groups of parameters being sampled. If the argument is set to FALSE the algorithm does not display such information.
print.iterations	An integer specifying how often should the algorithm's iteration count be displayed.

## Value

A list with the following arguments:

posteriors	A list containing draws from the posterior distribution of the parameters. It consists of the following elements $Beta$ , $Sigma$ , $S$ , $R$ , $PR.TR$ , $w$ , $S.t$ , $F.S.t$ , $F.Y$ that are described below. $Beta$ – a $(N + pN^2) \times M \times N$ array of the posterior draws of parameters $\beta_{s_t}$ . $Sigma$ – a $N^2 \times M \times N$ array of the posterior draws of parameters $vec(\Sigma_{s_t})$ . $S$ – a $N \times M \times N$ array of the posterior draws of parameters $\sigma_{s_t}$ . $R$ – a $(N - 1)N/2 \times M \times N$ array of the posterior draws of parameters $vecl(R_{s_t})$ . $PR.TR$ – a $M^2 \times N$ matrix of the posterior draws of parameters $p$ . $w$ – a $d \times N$ matrix of the posterior draws of parameters $w$ . $S.t$ – a $T \times N$ matrix of the posterior realizations of the Markov process $S$ . $F.S.t$ – a $h \times N$ matrix of draws of the forecasted states. $F.Y$ – a $h \times N \times N$ matrix of draws of the forecasted values of $y_t$ .
last.draws	A list containing the draws from the last iteration of the algorithm. This object can be delivered as a starting value for a new run of the MCMC algorithm in order to continue the chain. It consists of the following elements $Y$ , $U$ , $xi$ , $Beta$ , $Beta.0$ , $Beta.St$ , $Sigma$ , $PR.TR$ , $w$ , $S$ , $R$ , $G$ which definitions are identical to the elements of object starting.values.



## Authors

Matthieu Droumaguet and Tomasz Woźniak

## References

Droumaguet, M., Warne, A., Woźniak, T. (in press) Granger Causality and Regime Inference in Markov-Switching VAR Models with Bayesian Methods, *Journal of Applied Econometrics*.

Droumaguet, M., Woźniak, T. (2012). Bayesian Testing of Granger Causality in Markov-Switching VARs, *European University Institute Florence Working Paper Series*, **2012/06**, Italy.

---

G.MHM.MSVAR

*Marginal Data Density for MS-VAR models*

---

## Description

Estimates the marginal data density for MS(M)VAR( $p$ ) models with the estimator of [Geweke \(1999, 2005\)](#).

## Usage

```
G.MHM.MSVAR(Gibbs.output, priors, restrictions, alpha, debug=FALSE, print.iterations=100)
```

## Arguments

Gibbs.output	A list. An output of the estimation procedure in G.MSBVAR.griddy.
priors	A list. The same as argument priors from G.MSBVAR.griddy.
restrictions	A list. The same as argument restrictions from G.MSBVAR.griddy.
alpha	A scalar of value in (0,1). Truncation probability of the normal distribution from the Modified Harmonic Mean estimator.
debug	A logical value set by default to FALSE. If the argument is set to TRUE the algorithm displays information regarding the subsequent steps within each of the iteration corresponding to the groups of parameters being sampled. If the argument is set to FALSE the algorithm does not display such information.
print.iterations	An integer specifying how often should the algorithm's iteration count be displayed.

## Value

MHM	A scalar with the estimate of the natural logarithm of the marginal data density.
acceptance.rate	A scalar with the acceptance rate of the draws from the posterior distribution that are used for the marginal data density estimation. The rejections are due to the truncation rate specific for the Geweke (1999,2005) estimator (with tail probability alpha) as well as due to the intractable values of the likelihood function evaluated at the draws.
alpha	A scalar of value in (0, 1) reporting the value of the argument alpha of the function.
log.likelihood	A N-vector containing the values of the ordinates of the log-likelihood function at the posterior draws.
log.prior	A N-vector containing the values of the ordinates of the natural logarithm of the prior distribution at the posterior draws.
log.truncated.normal	A N-vector containing the values of the ordinates of the natural logarithm of the truncated normal distribution (with the mean set to the posterior mean of the parameters and the covariance matrix set to the posterior covariance matrix of the parameter vector and truncated symmetrically from both sides with probability N/2) evaluated at the posterior draws.
posterior.draws	A $k \times N$ matrix with the posterior draws of the parameters. $k$ denotes the overall number of the parameters of the model.

## Authors

Matthieu Droumaguet and Tomasz Woźniak

## References

- Droumaguet, M., Warne, A., Woźniak, T. (in press) Granger Causality and Regime Inference in Markov-Switching VAR Models with Bayesian Methods, *Journal of Applied Econometrics*.
- Droumaguet, M., Woźniak, T. (2012). Bayesian Testing of Granger Causality in Markov-Switching VARs, *European University Institute Florence Working Paper Series*, **2012/06**, Italy.
- Geweke, J. (1999). Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication. *Econometric Reviews*, **18**, 1-73.
- Geweke, J. (2005). Contemporary Bayesian Econometrics and Statistics. John Wiley & Sons, Inc.