

Bayesian Markov Regime-Switching Models for Cointegration

Kai Cui¹, Wenshan Cui²

¹Department of Statistical Science, Duke University, Durham, USA

²School of Science and Information, Qingdao Agricultural University, Qingdao, China

Email: kc52@stat.duke.edu, wshcui@qau.edu.cn

Received October 10, 2012; revised November 11, 2012; accepted November 18, 2012

ABSTRACT

This paper introduces a Bayesian Markov regime-switching model that allows the cointegration relationship between two time series to be switched on and off over time. Unlike classical approaches for testing and modeling cointegration, the Bayesian Markov switching method allows for estimation of the regime-specific model parameters via Markov Chain Monte Carlo and generates more reliable estimation. Inference of regime switching also provides important information for further analysis and decision making.

Keywords: Cointegration; Regime-Switching; Bayesian; MCMC

1. Introduction

Since the development of the concept of cointegration [1], there has been a rich literature on testing cointegration and applying cointegration approaches to real data analysis. One of the most illustrative examples in practice is the pair trading strategy [2]. The basic idea is that: find two securities whose prices have been historically moving together. So when the spread between them widens, we short the winner and buy the loser. And if we believe that the history would repeat itself, prices will converge again and the arbitrageur will profit. This moving-together relationship between two nonstationary time series is called cointegration. Mathematically, if two nonstationary time series U_t and V_t are cointegrated, then there exists a number δ called the cointegration ratio, such that $Y_t = U_t - \delta V_t$ is stationary.

Although there have been many statistical studies to find cointegrated time series, there are still many unsolved problems. First of all, it is often hard simply to find cointegration given a specific period of time. There are several statistical explanations for failing to reject the null of no cointegration including the span of the data set, structural breaks [3] and the choice of test model [4]. Secondly, there are few statistical decision-making rules after identifying candidate pairs. Taking pair trading as an example, typically, people simply use the decision rule that they open a long-short position when the pair prices have diverged by a certain amount (e.g. two standard deviations from the historical mean) and close the position when the prices have reverted [5].

This paper proposes the Bayesian Markov regime-switching model that allows the cointegration relationship between two time series to be switched on or off over time via a discrete-time Markov process. This is an improvement to the traditional cointegration tests considering that the model flexibly allows local non-cointegration rather than assuming global cointegration over the whole period of time. By using a fully Bayesian models, uncertainty about cointegration ratio is also incorporated into the model and inferred simultaneously with all other unknown quantities. Furthermore, inference of the hidden regime-switching is also critical to decision making and further generic analysis.

2. Markov Regime-Switching Models for Cointegration

Suppose we have two nonstationary time series U_t and V_t with integration order 1, and $Y_t = U_t - \delta V_t$ (δ is known, typically people propose a δ and then test the stationary property of Y_t). If Y_t is stationary, then we say time series U_t and V_t are cointegrated. To test for stationarity, the Engle-Grange method [6] tests the $\gamma = 0$ null hypothesis using the ADF unit root test [5] based on the Error Correction Model (EVM) with lag order K (as compared to $\gamma < 0$ in which case it is stationary):

$$\Delta Y_t = \mu + \gamma Y_{t-1} + \sum_{k=1}^K \beta_k \Delta Y_{t-k} + \varepsilon_t \quad (1)$$

where μ is a constant, β_k s are autoregression coeffi-

cients and $\varepsilon_t \sim N(0, \Sigma_\varepsilon)$.

In comparison, the Markov regime-switching model we proposed allows Y_t to switch between cointegrated or non-cointegrated regimes in a Markovian manner, by introducing the regime indicator variable X_t , regime specific parameters and the Markov transition matrix P_X . For the simplicity of exposition, we assume that $X_t \in \{0, 1\}$, with $X_t = 0$ denoting that that Y_t is stationary (*i.e.* U_t and V_t are cointegrated) at time t and $X_t = 1$ meaning non-cointegration. Then the model can be written as:

$$\begin{aligned} \Delta Y_t &= \mu^{(X_t)} + \gamma I_{\{X_t=0\}} Y_{t-1} + \sum_{k=1}^K \beta_k^{(X_t)} \Delta Y_{t-k} \\ &+ \varepsilon_t, \varepsilon_t \sim N(0, \Sigma_\varepsilon^{(X_t)}) \end{aligned} \tag{2}$$

$$P_X = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}$$

where $Y_t = U_t - \delta V_t$ and thus $\Delta Y_t = \Delta U_t - \delta \Delta V_t$. P_X is the Markov transition matrix of X_t , with

$$p_{ij} = \Pr(X_{t+1} = j | X_t = i)$$

and initial value X_0 .

Clearly, when $X_t = 0$ the model reduces to model (1) with negative γ , while $X_t = 1$ specifies unit root process for Y_t and thus no cointegration exists for time series U_t and V_t . By obtaining inference of the underlying regimes X_t , regime-specific parameters and segmentation of regime-specific data, the model provides much information for further generic analysis and decision making.

3. Bayesian Computation

We propose to use Bayesian analysis for the inference of parameters and latent regimes X_t , where posterior samples of all unknown quantities are drawn using Markov Chain Monte Carlo (MCMC).

Under this model (2), the likelihood function is:

$$\begin{aligned} L(X, Y | \Theta) &\propto P(X_0) P(\Delta Y_0 | X_0) \\ &\cdot \left[\prod_{t=1}^T P(\Delta Y_t | X_t) P(X_t | X_{t-1}) \right] \\ &\propto P(X_0) P(\Delta Y_0 | X_0) \\ &\left[\prod_{t: X_t=0} P(\Delta Y_t | X_t) P(X_t | X_{t-1}) \right] \\ &\cdot \left[\prod_{t: X_t=1} P(\Delta Y_t | X_t) P(X_t | X_{t-1}) \right] \end{aligned}$$

Conjugate prior distributions are placed on model parameters [7]. Specifically, conjugate Dirichlet priors are assigned to each row of the transition matrix P_X and

$q = (q_0, q_1)$, where $q_i = \Pr(X_0 = i)$. Conjugate Normal-Gamma priors are assigned for all the regression coefficients $\beta_k^{(X_t)}$ and the corresponding precisions $\phi_\varepsilon^{(X_t)}$

$$q \sim Dir(\alpha_1, \alpha_2); p_{ij} \sim Dir(\alpha_1^{(i)}, \alpha_2^{(i)}); i, j = 0, 1$$

$$\beta_k^{(X_t)} \sim N\left(0, \frac{\lambda}{\phi^{(X_t)}}\right); \phi^{(X_t)} \sim \Gamma\left(\frac{v_0}{2}, \frac{SS_0}{2}\right)$$

To obtain the posterior marginal distributions of the unknown parameters and the hidden regimes X_t , Gibbs sampler is constructed to iterate the following steps:

1) Sample q and P_X from full conditional distributions:

$$\begin{aligned} q | \dots &\sim Dir(\alpha_1 + I_{\{X_0=0\}}, \alpha_2 + I_{\{X_0=1\}}); (p_{00}, p_{01}) | \dots \\ &\sim Dir\left(\alpha_1^{(0)} + \sum_{t=1} I_{\{X_{t-1}=0, X_t=0\}}, \alpha_2^{(0)} + \sum_{t=1} I_{\{X_{t-1}=0, X_t=1\}}\right) \\ &\cdot (p_{10}, p_{11}) | \dots \\ &\sim Dir\left(\alpha_1^{(1)} + \sum_{t=1} I_{\{X_{t-1}=1, X_t=0\}}, \alpha_2^{(1)} + \sum_{t=1} I_{\{X_{t-1}=1, X_t=1\}}\right) \end{aligned}$$

2) Sample the regression coefficients β and variance from Normal-Inverse Gamma full conditional distributions given the conjugacy of the priors.

3) Sample the whole path of X_t .

Since X_t s are highly correlated, Gibbs sampler constructed via regular full conditional distribution would be extremely inefficient [7]. To overcome this, Forward Filtering and Backward Sampling algorithm is applied to draw block samples of X_t . To achieve this, define $\alpha_n(i) = P(\Delta Y_0, \dots, \Delta Y_n, X_n = i | \Theta)$, then by recursion:

$$\alpha_0(i) = q_i P(\Delta Y_0 | X_0 = i);$$

and

$$\alpha_n(i) = \left[\sum_j \alpha_{n-1}(j) p_{ji} \right] P(\Delta Y_n | X_n = i)$$

With this, the results follow that:

$$\begin{aligned} P(X_t = i | \dots) &\propto \alpha_t(i) \\ P(X_{t-1} = j | X_t = i, \dots) &\propto \alpha_{t-1}(j) p_{ji} \end{aligned}$$

By using this algorithm, a sample of X_T is first drawn from a bernoulli (multinomial if X_t takes more than two values) distribution, and $\forall 0 \leq t \leq T-1$, samples of X_t are drawn sequentially and backward from the conditional bernoulli distribution, with

$$P(X_t = j | X_{t+1} = i) = \frac{\alpha_t(j) p_{ji}}{\sum_j \alpha_t(j) p_{ji}},$$

until the whole X_t time series are sampled.

4. Simulated Time Series Analysis

4.1. Model Assessment

To testify the performance of the proposed framework, we simulated a Markov regime-switching times series of length $T = 500$, which switches between one stationary AR(2) process (State 0) and one non-stationary AR(2) process (State 1). The two AR(2) models and the corresponding Error Correction Models (ECM) are shown as follows: (the (non-)stationary property can be easily tested by the Unit Root Test)

$$\begin{aligned}
 \text{State 0: } & y_t = 3 + 0.6y_{t-1} - 0.28y_{t-2} + \varepsilon_t, \\
 \text{ECM: } & \Delta y_t = 3 - 0.68y_{t-1} + 0.28\Delta y_{t-1} + \varepsilon_t; \\
 \text{State1: } & y_t = 4 + 0.7y_{t-1} + 0.3y_{t-2} + \varepsilon_t^*, \\
 \text{ECM: } & \Delta y_t = 4 - 0.3\Delta y_{t-1} + \varepsilon_t^*.
 \end{aligned}
 \tag{3}$$

where $\varepsilon_t \sim N(0,1)$ and $\varepsilon_t^* \sim N(0,1)$. The transition matrix is specified as

$$P_X = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 \\ 0.8 & 0.2 \end{bmatrix}.$$

A simulated data was shown in **Figure 1**.

The proposed model was applied to the time series to find regime switching, with the priors specified as follows:

$$\begin{aligned}
 q & \sim \text{Dir}(1,1), (p_{00}, p_{01}) \sim \text{Dir}(1,1), (p_{10}, p_{11}) \sim \text{Dir}(1,1), \\
 \beta_k^{(x_t)} & \sim N\left(0, \frac{1}{\phi^{(x_t)}}\right); \phi^{(x_t)} \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)
 \end{aligned}$$

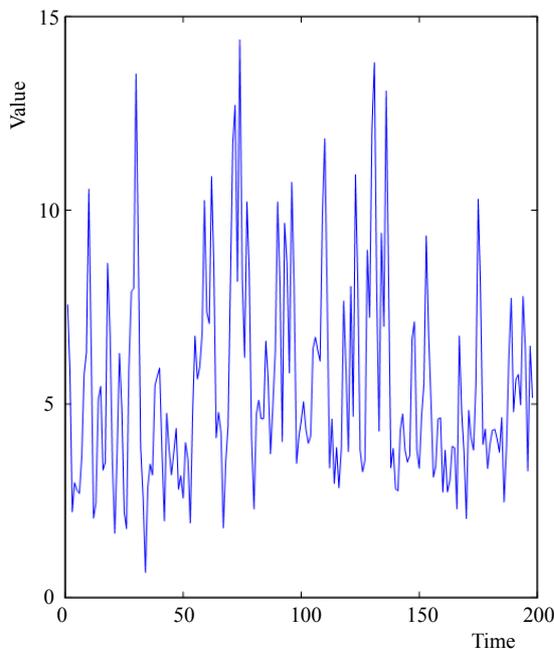


Figure 1. Illustration of a time series stimulated by the markov switching model.

To infer the value of X_t based on posterior samples, we use posterior probability > 0.5 as the cut-off point. Shown in **Figure 2**, the inferred regimes are compared with the true values, which shows that our model gives good recovery of the latent regimes (with the first 200 time points shown). Other model parameters are also correctly inferred as shown in **Table 1**, where posterior distributions cover the true values well.

4.2. Posterior Decision Making

The importance of inference of regimes when analyzing (non)stationary time series lies in the fact that commonly-used stationarity and cointegration tests (e.g. ADF unit root test and Engle-Granger cointegration tes [1]) may well give misleading results when regime switching

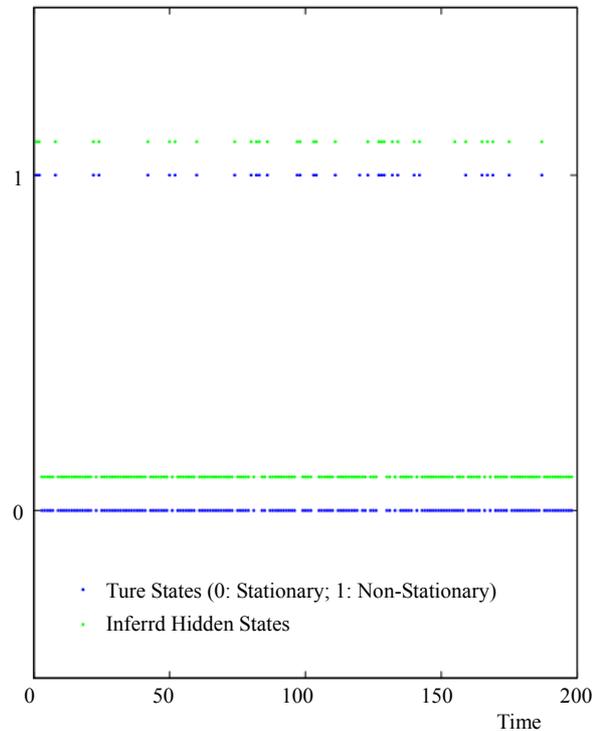


Figure 2. Inferred regimes X_t (in green) compared to the true values (in blue) show good inference.

Table 1. Posterior estimates of model parameters compared to the true values. The parameters are defined as in model 2 and specified in (3).

Parameter	Mean	STD	Truth
γ	-0.676	0.011	-0.68
$\beta_1^{(0)}$	0.287	0.013	-0.28
$\phi^{(0)}$	0.952	0.048	1
$\beta_1^{(1)}$	-0.323	0.041	-0.3
$\phi^{(1)}$	0.911	0.093	1

exists in the process. For illustration, a quick ADF test of the previously simulated data concludes that the null hypothesis with unit root is rejected at 99.9% confidence level, indicating the time series is stationary. If this time series were generated by the linear combination of two nonstationary time series, then the ADF test tells that these two are co-integrated, which is clearly wrong.

In the following part, we will use the context of pair trading to illustrate how the Markov regime-switching model can potentially help improve decision making in practice. Basically people do pair trading based on the traditional rule that you open a long-short position when the pair prices have diverged by more than two historical standard deviations. And you unwind the position when it returns to historical mean.

First of all, the model clearly allows more reasonable estimation of the historical mean and standard deviation, based solely on data in the stationary (cointegrated) regimes, rather than including data in the nonstationary (non-cointegrated) regimes. This difference can be observed in **Figure 3**, where the historical mean using data in the stationary regime is different from that using all data, and the standard deviation is also smaller.

Secondly, the identification of stationary (cointegrated) and nonstationary (non-cointegrated) regimes also help establish more rational decision making rules, which

should be: we open a position when it is both in the stationary state and has diverged from the historical stationary mean. It is apparently risky either to open a position when currently we are in a non-stationary state or the historical mean calculation involves non-stationary data.

Since people care much about the time points where values are at least 2 standard deviations away from the historical mean, the figure shows that we pick different time points using our model and decision making rules from those obtained using all historical data and traditional rules, which we believe are more reasonable choices. For example, many spikes in **Figure 3** are actually not good time points to open the position based on our Markov regime-switching model simply because those spikes are in the non-stationary (non-cointegrated) regime. However in comparison, the traditional approach considers them open positions whenever the values are 2 standard deviations away from the mean, which is a very risky decision not considering the regimes.

5. Cointegrated Price Series Analysis

An possible example of a pair of cointegrated time series is the gold ETF, GLD versus the gold miners ETF, GDX. GLD reflects the spot price of gold, and GDX is a basket of gold-mining stocks. It makes intuitive sense that their

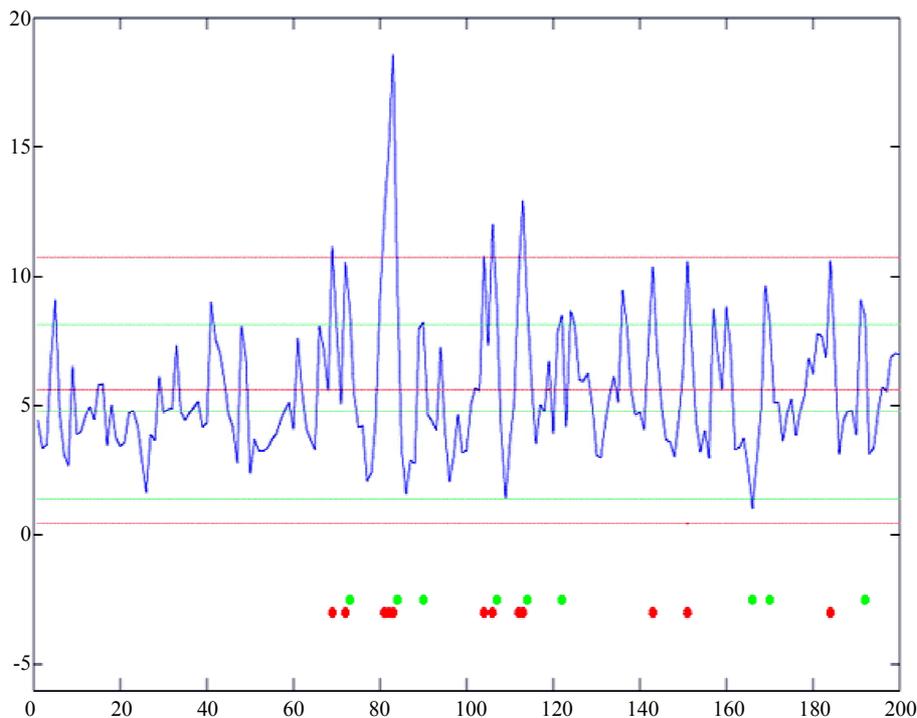


Figure 3. Results comparison between our Bayesian Markov regime-switching model and traditional cointegration test and analysis using all historical data. Red lines indicate the mean and mean $\pm 2SD$ using all historical data, which is a traditional way after you have done the ADF test to show the stationary property; Green lines indicate those using only historical data in stationary regimes. Red and green dots mark the time points where values at those points are at least 2SD away from the historical mean based on traditional and our Markov regime switching model respectively.

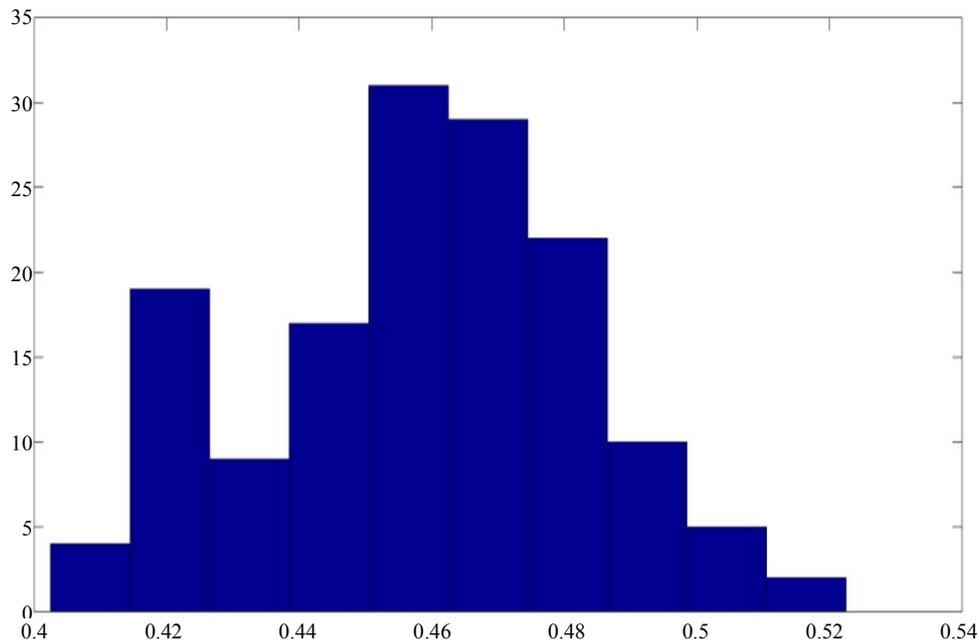


Figure 4. Distribution of the probability that X_t is in the cointegration regime ($t = 1, \dots, T$).

prices may move in tandem. Previous study via the two-step Engle-Granger method [1] identified that a portfolio with long 1 share of GLD and short 1.6766 share of GDX is likely a stationary time series, with lag 1 but the conclusion is later questioned by other studies [8]. To test the possible co-integration, the two-state Markov regime switching model is applied to the 05/23/06-11/30/07 GLD and GDX time series. A histogram shown the distribution of the probability of the time points being in the cointegration state is shown in **Figure 4**. According to the previous 0.5 cut-off point, the Markov regime switching model indicates that at most of the time, the two time series are not cointegrated with the 1.6766 cointegration ratio. This may serve as another counterexample (together with the simulation result) that the widely-used ADF test might provide misleading results when used to test co-integration regardless of possible regime switching.

6. Conclusions and Future Work

In this study, we proposed to use the Bayesian Markov regime-switching model as a flexible model for cointegration and stationarity analysis, where the latent regime-switching process is modeled via a Markov process. A strong message of this study is that, while identifying cointegration (or stationarity) is often hard globally, allowing local non-cointegration (or non-stationarity) and inferring the regime switching can provide much information for further analysis and decision making.

Several extensions of the study are still worth exploring, including relaxing the hidden Markov transition models and incorporating uncertainty about number of re-

gimes in the model. Hidden semi-Markov models are natural extensions of hidden Markov models. While the run-length distribution of the hidden Markov models implicitly follows a geometric distribution, hidden semi-Markov models allow for more general runlength distributions, and thus are more flexible to describe the time spend in a given regime. As for the cases with the number of regimes unknown, Bayesian inference through reversible jump MCMC methods [9] could be a viable alternative that both explores models with different number of regimes and estimation of regime-specific parameters.

REFERENCES

- [1] R. F. Engle and C. W. J. Granger, "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, Vol. 55, No. 2, 1987, pp. 251-276. [doi:10.2307/1913236](https://doi.org/10.2307/1913236)
- [2] H. Puspaningrum, "Pairs Trading Using Cointegration Approach," Ph.D. Thesis, University of Wollongong, Wollongong, 2012.
- [3] J. Campos and N. R. Ericsson and D. F. Hendry, "Cointegration Tests in the Presence of Structural Breaks," International Finance Discussion Papers 440, Board of Governors of the Federal Reserve System (US), 1996.
- [4] E. G. Gatev and W. Goetzmann and K. G. Rouwenhorst, "Pairs Trading: Performance of a Relative Value Arbitrage Rule," Boston College, Boston, 2006.
- [5] A. W. Gregory and B. E. Hansen, "Residual-Based Tests for Cointegration in Models with Regime Shifts," *Journal of Econometrics*, Vol. 70, No. 1, 1996, pp. 99-126. [doi:10.1016/0304-4076\(96\)01685-7](https://doi.org/10.1016/0304-4076(96)01685-7)
- [6] D. A. Dickey and W. A. Fuller, "Distribution of the Esti-

- mators for Autoregressive Time Series With a Unit Root,” *Journal of the American Statistical Association*, Vol. 74, No. 366, 1979, pp. 427-431. [doi:10.2307/2286348](https://doi.org/10.2307/2286348)
- [7] G. E. B. Archer and D. M. Titterton, “Parameter Estimation for Hidden Markov Chains,” *Journal of Statistical Planning and Inference*, Vol. 108, No. 1, 2002, pp. 365-390. [doi:10.1016/S0378-3758\(02\)00318-X](https://doi.org/10.1016/S0378-3758(02)00318-X)
- [8] E. P. Chan, “Quantitative Trading,” John Wiley and Sons, Hoboken, 2008.
- [9] C. P. Robert, T. Ryden and D. M. Titterton, “Bayesian Inference in Hidden Markov Models through the Reversible Jump Markov Chain Monte Carlo Method,” *Journal of The Royal Statistical Society Series B*, Vol. 62, No. 1, 2000, pp. 57-75. [doi:10.1111/1467-9868.00219](https://doi.org/10.1111/1467-9868.00219)