Bayesian estimation for Markov Modulated Asset Prices.

Richard Emilion*, Hafedh Faires †
MAPMO, UMR 6628 Université d’Orléans-CNRS,
BP. 6759 45067 Orléans Cedex 2, France,
and
Srikanth K. Iyer‡
Department of Mathematics Indian Institute of Science
Bangalore - 560012, India.

June 8, 2009

Abstract

A Stochastic Differential Equation (SDE) appearing in mathematical finance is extended in random environment by assuming that its two parameters are switched by an unobserved continuous-time Markov chain whose states represent the states of the market environment. A Dirichlet process is placed as a prior

*corresponding author: richard.emilion@labomath.univ-orleans.fr
†hafedh.faires@univ-orleans.fr
‡Research supported in part by UGC SAP IV, skiyer@math.iisc.ernet.in.
on the space of the sample paths of this chain, leading to a hierarchical Dirichlet model whose estimation is done both on simulated data and on a real data set from the Indian market.

**Keywords:** Dirichlet process, Markov regime switching, random environment, Stochastic differential equation, Geometric Brownian motion, Asset price.

## 1 Introduction

Models in which parameters move between a fixed number of regimes with switching controlled by an unobserved stochastic process, are very popular in a great variety of domains (Finance, Biology, Meteorology, Networks, etc.), notably because they take into account random regime change of the environment. In this paper we consider the estimation problem for a model described by a stochastic differential equation (SDE) with Markov regime-switching (MRS), i.e., with parameters controlled by a finite state continuous-time Markov chain (CTMC) as done, for example, in (Deshpande and Ghosh, 2008). In such a setting, the parameter estimation problem poses a real challenge, mainly due to the fact that the paths of the CTMC are unobserved. A standard approach consists in using the celebrated EM algorithm (Dempster, Laird, Rubin, 1977) as proposed for example in (Hamilton, 1990).

In the present paper, our estimation approach is Bayesian, the aim being to find a pair (parameters, CTMC path) with likelihood as large as possible. Standard priors are placed on the parameters space but, as the CTMC paths are unobserved, a large number of paths are drawn from a Dirichlet process placed as a prior on the path space of the CTMC. The complete model then appears as a Hierarchical Dirichlet Model (HDM), as in Ishwaran, James and Sun (2000) and Ishwaran and James (2002),
whose estimation procedure requires some rather nontrivial computations of posterior distributions due to the temporal level induced by the specific SDE and the CTMC. Using the well-known stick-breaking approximation, each set of iterations selects the pair with largest likelihood and then the Dirichlet process is updated in order to look for other paths which can further improve the likelihood.

The considered SDE is that of a geometric Brownian motion, a popular model for asset prices in mathematical finance which depends on two parameters, the trend and volatility. It is extended to a MRS setting so that the CTMC transitions correspond to regime changes in the market.

The rest of the paper is structured as follows. In Section 2 we present the stock price SDE with MRS and the complete HDM. Section 3 is devoted to the posterior computations. The estimation algorithm is described in Section 4. Numerical results are presented in Section 5 for one simulated data and on a data from the Indian market. We conclude with a summary in the last Section.

## 2 Markov regime switching with Dirichlet prior

Our model is specified in a mathematical finance setting but of course it can be extended similarly in many other cases. We model the stock price using a geometric Brownian motion with drift and variance depending on the state of the market, the latter being modeled by a continuous time Markov chain. In what follows, the notation $\sigma$ will be used to denote the variance rather than the standard deviations.

The following notations will be adopted:

1. $n$ will denote the number of observed data and also the length of an observed path.
2. $M$ will denote the number of states of the Markov chain.

3. The state space of the chain will be denoted by $S = \{ i : 1 \leq i \leq M \}$.

4. $N$ will denote the number of simulated paths.

5. Given a path, $m$ will denote the number of distinct states in that path.

We now describe the model:

- Let $(X_t)$ be a CTMC taking values in the set $S = \{ i : 1 \leq i \leq M \}$ which represents $M$ possible states of the market. The transition probabilities of this chain are denoted by $p_{ij}, i, j \in S$ and the transition rate matrix is $Q = (q_{ij})_{i,j \in S}$ with
  \[
  \lambda_i > 0, \quad q_{ij} = \lambda_i p_{ij} \quad \text{if} \quad i \neq j, \quad \text{and} \quad q_{ii} = -\sum_{j \neq i} q_{ij}, \quad i, j \in S.
  \]

- The stock price follows the following SDE:
  \[
  \frac{dS_t}{S_t} = \beta(X_t)dt + \sqrt{\sigma(X_t)}dB_t, \quad t \geq 0,
  \]
  where $B_t$ is a standard Brownian motion. By Ito’s formula, the process $Z_t = \log(S_t)$ satisfies the SDE,
  \[
  dZ_t = \mu(X_t)dt + \sqrt{\sigma(X_t)}dB_t, \quad t \geq 0,
  \]
  where $\mu(X_t) = \beta(X_t) - \frac{1}{2} \sigma(X_t)$. Let the observed data be denoted by $Z_0, Z_1, \ldots, Z_n$, and define the log-returns as, $Y_t = Z_t - Z_{t-1} = \log(S_t/S_{t-1}), t = 1, 2, \ldots, n$. Given the path $X = \{ X_s, 0 \leq s \leq n \}$, let $T_j(t)$ be the time spent by the path $X$ in state $j$ in the time interval $[t - 1, t]$. Define
  \[
  \mu(t) := \sum_{j=1}^M \mu_j T_j(t); \quad \sigma(t) := \sum_{j=1}^M \sigma_j T_j(t). \quad (1)
  \]
Then, conditionally to path $X$, the random variables $Y_t$, $t = 1, 2, \ldots, n$ are i.i.d. $\mathcal{N}(\mu(t), \sigma(t))$, where $\mathcal{N}(\mu, \sigma)$ denotes a Normal distribution with mean $\mu$ and variance $\sigma$.

• For each $i = 1, 2, \ldots, M$, let the priors on $\mu_i$ and $\sigma_i$ are given by

$$
\mu_i \overset{\text{ind}}{\sim} \mathcal{N}(\theta, \tau^\mu), \quad \text{with} \quad \theta \sim \mathcal{N}(0, A), \quad A > 0, \quad (2)
$$

$$
\sigma_i \overset{\text{ind}}{\sim} \Gamma(\nu_1, \nu_2), \quad (3)
$$

where $\Gamma(\nu_1, \nu_2)$ denotes a Gamma distribution with shape parameter $\nu_1$ and scale parameter $\nu_2$.

• We now place a Dirichlet prior $\mathcal{D}(\alpha \ H)$, on the path space of the CTMC $(X_t)$, with precision parameter $\alpha > 0$ and mean $H$ which is a probability measure governing a CTMC on the path space $D([0, \infty), S)$, the set of cadlag functions. The initial distribution according to $H$ is the uniform distribution $\pi_0 = (1/M, \ldots, 1/M)$, and the transition rate matrix is $Q_0$ with $p_{ij} = 1/(M-1)$ and $\lambda_i = \lambda > 0$. Thus the Markov chain under $Q_0$ will spend an exponentially distributed time with mean $1/\lambda$ in any state $i$ and then jump to state $j \neq i$ with probability $1/(M-1)$.

A realization of the Markov chain from the above prior is generated as follows: Generate a large number of paths $X_i = \{x^i_s : 0 \leq s \leq n\}$, $i = 1, 2, \ldots, N$, from $H$. Generate the vector of probabilities $(p_i, i = 1, \ldots, N)$ from a stick-breaking scheme with parameter $\alpha$. Then draw a realization of the Markov chain from the distribution

$$
p = \sum_{i=1}^N p_i \delta_{X_i}, \quad (4)
$$

The parameter $\lambda$ is chosen to be small so that the variance is large and hence we obtain a large variety of paths to sample from at a later stage. The prior for $\alpha$
is given by,

\[ \alpha \sim \Gamma(\eta_1, \eta_2). \] (5)

3 Estimation

Roughly speaking, estimation will be done by simulating a large number of paths of the Markov chain, selecting one path according to a probability vector generated by stick-breaking, and then using the blocked Gibbs sampling technique. This technique requires the posterior distribution of the each parameter conditioned on the current values of the other parameters.

We denote by \( \mu \) and \( \sigma \), the current values of the vectors \((\mu_1, \mu_2, \ldots, \mu_n)\) and \((\sigma_1, \sigma_2, \ldots, \sigma_n)\), respectively. Let \( Y \) be the vector of observed data \((Y_1, \ldots, Y_n)\). Given the current path \( X = (x_s, 0 \leq s \leq n) \) the Markov chain, let \( X^* = (x_1^*, \ldots, x_m^*) \) be the distinct values in \( X \).

3.1 Modifying the observed data set

In order to obtain the conditional distribution of the parameters, we first need to extract the change in the log-returns between the jump times of the Markov chain. Let \( 0 = t_0 < t_1 < t_2 < \ldots t_J \) be the times at which the path \( X \) changes state. Define the log-returns between the jump times, \( W_k = \log(S_{t_k}/S_{t_{k-1}}), \ k = 1, 2, \ldots, J. \) To obtain realizations of the \( W_k \) from the observed \( Y \) process, we need to simulate Gaussian random variables conditioned on their sums.

Consider any \( t \in \{0, 1, \ldots, n\} \) for which the chain changes state atleast once in the time interval \([t - 1, t]\). So for some \( p, k \) we have \( t_{k-1} < t - 1 \leq t_k < \ldots < t_{k+p} < t < t_{k+p+1}. \)
3.2 Gibbs sampling procedure

Let \( V_1^t = \log(S_t S_{t-1}) \) and \( V_2^t = \log(S_t S_{t+p}) \). Then,

\[
Y_t = V_1^t + \sum_{i=1}^{p} W_{k+i} + V_2^t. \tag{6}
\]

Suppose that the chain \( X \) is in state \( j_i \) in the time interval \([t_{k+i-1}, t_{k+i}]\), \( i = 0, 1, \ldots, p+1 \). Set \( s_0 = t_k - t - 1, s_i = t_{k+i} - t_{k+i-1}, i = 1, 2, \ldots, p \), and \( s_{p+1} = t - t_{k+p} \). Let \( m_j = \mu_j s_i \) and \( v_j = \sigma_j s_i, i = 0, 1, \ldots, p + 1 \). Recall that \( Y_t \sim \mathcal{N}(\mu(t), \sigma(t)) \), where \( \mu(t), \sigma(t) \) are as defined in (1). It is easy to see that the joint conditional density of \((V_1^t, W_{k+1}, \ldots, W_{k+p})\) given \( Y_t = y \) will be

\[
f(u_0, u_1, \ldots, u_p) = C \prod_{i=0}^{p} \exp \left( -\frac{1}{2} \frac{u_i + v_{p+1}}{v_i v_{p+1}} \left( u_i - \frac{v_{p+1} m_i + v_i (y - m_{p+1})}{v_i + v_{p+1}} \right)^2 \right), \tag{7}
\]

where \( C \) is a constant that depends on \( y \) and the parameters. Thus, one can simulate the variables \( V_1^t, W_k, W_{k+1}, \ldots, W_{k+p} \) from independent Gaussians and then obtain \( V_2^t \) using (6).

Using the above procedure, we can obtain a realization for all \( W_k \) for which \([t_{k-1}, t_k] \subseteq [t-1, t] \), for some \( t \in \{0, 1, \ldots, n\} \). Now for any \( k \) for which there is a \( q \geq 0 \), such that \( t - 1 \leq t_{k-1} < t < t + 1 < \ldots < t + q \leq t_k < t + q + 1 \), we can obtain \( W_k \) using the relation

\[
W_k = V_2^t + \sum_{i=1}^{q} Y_{t+i} + V_1^{t+q+1}. \tag{8}
\]

Note that the \( W \) values depend on the path \( X \) and the parameter values \( \mu, \sigma \) and hence are to be computed in each iteration of the Gibbs sampling procedure which we describe next.

3.2 Gibbs sampling procedure

We are now ready to estimate the posterior distributions of the parameters using Gibbs sampling. The procedure consists of two nested iterations. The outer loop is initiated
using the parameters chosen for the priors. Each iteration of the outer loop consists of a large number of iterations of the inner loop in which a sample is drawn recursively for each parameter conditioned on the current values of the other parameters and the data. At the end of an iteration of the inner loop we obtain one realization of the parameters from their approximate posterior distribution.

We now derive the conditional distribution for each parameter conditioned on the other parameters and the data. Recall the $X^*$ is the set of distinct values observed in the path $X$ of the CTMC.

- **Conditional for $\mu$.** For each $j \in X^*$ draw

  \[ (\mu_j | \theta, \tau^\mu, \sigma, X, W)^{ind} \sim N(\mu^*_j, \sigma^*_j), \]  

  where

  \[ \mu^*_j = \sigma^*_j \left( \sum_{k: X_{k-1} = j} \frac{W_k}{\sigma_j (t_k - t_{k-1}) + \theta / \tau^\mu} \right), \]

  \[ \sigma^*_j = \left( \frac{n_j}{\sigma_j} + \frac{1}{\tau^\mu} \right)^{-1}, \]

  and $n_j$ being the number of times $j$ occurs in $X$. For each $j \in X \setminus X^*$, independently simulate $\mu_j \sim N(\theta, \tau^\mu)$.

- **Conditional for $\sigma$.** For each $j \in X^*$ draw

  \[ (\sigma_j | \mu, \nu, X, W)^{ind} \sim \Gamma(\nu_1 + \frac{n_j}{2}, \nu^*_2, j), \]  

  where

  \[ \nu^*_2, j = \nu_{2, j} + \sum_{k: X_{k-1} = j} \frac{(W_k - \mu_j (t_k - t_{k-1}))^2}{2(t_k - t_{k-1})}. \]

  Also for each $j \in X \setminus X^*$, independently simulate $\sigma_j \sim \Gamma(\nu_1, \nu_2)$. 

3.2 Gibbs sampling procedure

- **Conditional for** $X$.

$$(X|p) \sim \sum_{i=1}^{N} p_i^* \delta_{X_i},$$  \hspace{1cm} (11)

where

$$p_i^* \propto \frac{1}{\prod_{j=1}^{m} \left( \prod_{\{k: x_{i,k-1}^* = j\}} 2\pi \sigma_j (t_k - t_{k-1}))^{1/2} e^{-\frac{1}{2\sigma_j} (W_i^j(t_k - t_{k-1}))^2} \right) p_i},$$  \hspace{1cm} (12)

where $(x_{i,1}^*, \ldots, x_{i,m_i}^*)$ denote the current $m = m(i)$ unique values of the states and $t_k^i, W_k^i$ are as defined in subsection 3.1 for the path $X_i, i = 1, \ldots, N$.

- **Conditional for** $p$.

$$p_1 = V_1^*, \text{ and } p_k = (1 - V_1^*) \cdots (1 - V_{k-1}^*)V_k^*, \ k = 2, 3, \ldots, N-1,$$  \hspace{1cm} (13)

where

$$V_k^* \sim \beta \left( 1 + r_k, \alpha \right),$$

$r_k$ equal 1 if $i = k$ and 0 otherwise.

- **Conditional for** $\alpha$.

$$(\alpha|p) \sim \Gamma \left( N + \eta_1 - 1, \eta_2 - \sum_{i=1}^{N-1} \log(1 - V_i^*) \right),$$

where the $V^*$ values are those obtained in the simulation of $p$ in the above step.

- **Conditional for** $\theta$.

$$(\theta|\mu) \sim \mathcal{N}(\theta^*, \tau^*),$$  \hspace{1cm} (14)

where

$$\theta^* = \frac{\tau^*}{\tau^\mu} \sum_{j=1}^{M} \mu_j,$$

and

$$\tau^* = \left( \frac{M}{\tau^\mu + \frac{1}{A}} \right)^{-1}.$$
Proof.

(a) The computations of the posterior distributions for $\mu$, $\sigma$ and $\theta$ follow in the same manner as in Ishwaran and James (2002) and Ishwaran and Zarepour (2000). Here, $X_t = s$ means that the class variable is equal to $s$.

(b) Conditional for $X$:

$$P\{X = X_i \mid p, \mu, \sigma, W\} = P\{W \mid p, \sigma, X = X_i, \mu\}P\{X = X_i \mid \sigma, \mu, p\}P\{\mu, \sigma\} \propto \prod_{j=1}^{m} \left( \prod_{\{k : x_{i}^{*}_{k-1} = j\}} \frac{1}{(2\pi\sigma_j(t_k - t_{k-1}))^{1/2}} e^{-\frac{1}{2\sigma_j}(W_{i} - \mu_j(t_k - t_{k-1}))^2} \right) p_i$$

where $X_i = (x_1^i, \ldots, x_n^i)$ and $(x_{1}^{i*}, \ldots, x_{m}^{i*})$ denote the current $m$ unique values in the path $X_i$.

(c) Conditional for $p$: The Sethuraman stick-breaking scheme can be extended to the two-parameter Beta distributions, see Walker and Muliere (1997, 1998):

Let $V_k \overset{ind}{\sim} \beta(a_k, b_k)$, for each $k = 1, \ldots, N$. Let

$$p_1 = V_1, \text{ and } p_k = (1 - V_1) \cdot \ldots \cdot (1 - V_{k-1})V_k, \quad k = 2, 3, \ldots, N - 1.$$ 

We will write the above random vector, in short as

$$p \sim SB(a_1, b_2, \ldots, a_{N-1}, b_{N-1}).$$

By Connor and Mosimann (1969), the density of $p$ is

$$\left( \prod_{k=1}^{N-1} \frac{\Gamma(a_k - b_k)}{\Gamma(a_k)\Gamma(b_k)} \right) p_{a_1-1}^{a_1} \ldots p_{a_{N-1}-1}^{a_{N-1}-1} p_{b_{N-1}-1}^{b_{N-1}-1} \times$$

$$\times (1 - P_1)^{b_1 - (a_2 - b_2)} \ldots (1 - P_{N-2})^{b_{N-2} - (a_{N-1} - b_{N-1})},$$
where \( P_k = p_1 + \ldots + p_k \).

From this, it easily follows that the distribution is conjugate for multinomial sampling, and consequently the posterior distribution of \( p \) given \( X \), when \( a_k = 1 \) and \( b_k = \alpha \) for each \( k \), is

\[
SB(a_1^*, b_2^*, \ldots, a_{N-1}^*, b_{N-1}^*),
\]

where

\[
b_k^* = \alpha \\
a_k^* = 1 + r_k,
\]

and \( r_k \) equal 1 if \( i = k \) and 0 otherwise, \( k=1, \ldots, N-1 \).

\[\square\]

4 Implementation

The algorithm presented in the previous section was implemented in C language. The implementation includes:

- functions that simulate standard probability distributions: Uniform, Normal, Gamma, Beta, Exponential.
- a function that returns an index \( \in \{1, \ldots, n\} \) according to a vector of probability \( p_1, \ldots, p_n \).
- a function that simulates a probability vector according to stick-breaking scheme.
- a function that simulates \( n \) paths of a Markov chain.
- a function that records the number of times a state appears in a path.
- a function that chooses one of the paths according to a vector of probability.
- a function that modifies the parameters of prior distributions according to the formulas of the above posteriori distributions.
We first simulated a large number of paths given $\pi_0$ and $Q_0$. Then for each iteration a path is selected according to the weights drawn by a stickk breaking scheme and the parameters are updated using posteriori formulas. The pair (parameters, path) is kept memorized if it improves the likelihood. After a large number of iteration we use the pair with maximum likelihood to re-estimate $\pi$ and $Q_0$ from the corresponding path. The above procedure is repeated a large number of times expecting an improvement of the likelihood.

### 4.1 Simulated data

We fit the model, using the algorithm developed above, to a simulated series of length $n = 480$, $M = 4$ regimes, the mean and variance in each state being chosen as follows:

\[
\begin{align*}
(\mu_1, \sigma_1) &= (-1.15, 0.450) \\
(\mu_2, \sigma_2) &= (-0.93, 0.450) \\
(\mu_3, \sigma_3) &= (-0.60, 0.440) \\
(\mu_4, \sigma_4) &= (1.40, 0.500).
\end{align*}
\]

We carry out the estimation procedure for the series simulated using the above parameters with number of states $M = 10$, number of paths $N = 100$ and run it for 25,000 iterations. It was observed that the algorithm was able to put most of the mass (in terms of the stationary distribution of the MC) on 4 regimes, which are close to the ones chosen above.

At the end of the iterations we compute a confidence interval for the mean and for the variance in each regime. The confidence intervals for the mean and the variance are given below.

**Regime 1:**

\[
I_m = [-1.208, -1.12423] \quad \text{and} \quad I_v = [0.431, 0.4738].
\]
4.2 Indian National Stock Exchange

Regime 2:

\[ I_m = [-0.9351, -0.9296] \text{ and } I_v = [0.442, 0.4538]. \]

Regime 3:

\[ I_m = [-0.63446, -0.5140] \text{ and } I_v = [0.4319, 0.4491]. \]

Regime 4:

\[ I_m = [1.30114, 1.43446] \text{ and } I_v = [0.4949, 0.5081]. \]

Based on the confidence intervals given above, we can conclude that the algorithm is able to identify the parameters of the simulated data set.

4.2 Indian National Stock Exchange

We applied our algorithm to the daily closing prices of the mid-cap index of the Bombay stock exchange from 21/12/2006 to 15/11/2007 (www.bseindia.com). For this data set we have, \( n = 250 \), and we run our algorithm with \( N = 100 \) paths and choose a Gamma(2, 4) distribution as the prior for \( \alpha \).

With the above choice, we obtain 6 regimes for which the estimates for the mean, variance and stationary probabilities are as follows:

<table>
<thead>
<tr>
<th></th>
<th>R 1</th>
<th>R 2</th>
<th>R 3</th>
<th>R 4</th>
<th>R 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>-0.00124</td>
<td>0.00092</td>
<td>-0.0032</td>
<td>0.0014091143</td>
<td>-0.00339</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.1982 e-4</td>
<td>1.0917e-4</td>
<td>2.9624e-4</td>
<td>0.6382e-4</td>
<td>1.0502e-4</td>
</tr>
<tr>
<td>( \pi )</td>
<td>0.21</td>
<td>0.12</td>
<td>0.098</td>
<td>0.43</td>
<td>0.12</td>
</tr>
</tbody>
</table>

The most frequent Markov chain path, its parameters \( \lambda_s \)s and the matrix of transition probability \( (p_{i,j})_{i,j=1,...,6,i\neq j} \) are respectively equal to:
It is interesting to note that in the high volatility states, the index has a negative drift as is usually observed in analysis of empirical data. A by-product of our algorithm is the distribution of the current state of the volatility, which is required to compute the price of an option (see [2] and references therein).

**Acknowledgment:** We would like to thank Chiranjit Mukhopadhyay for sharing his data with us.

### References


