www.oeaw.ac.at

ÖAW AUSTRIAN ACADEMY OF SCIENCES

Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

RICAM
JOHANN-RADON-INSTITUTE
FOR COMPUTATIONAL AND APPLIED MATHEMATICS

Markov Chain Monte Carlo Methods for Parameter Estimation in Multidimensional Continuous Time Markov Switching Models

M. Hahn, S. Frühwirth-Schnatter, J. Sass

RICAM-Report 2007-09

www.ricam.oeaw.ac.at

Markov Chain Monte Carlo Methods for Parameter Estimation in Multidimensional Continuous Time Markov Switching Models*

Markus Hahn†, Sylvia Frühwirth-Schnatter‡, Jörn Sass†

April 5, 2007

Abstract

We present Markov chain Monte Carlo methods for estimating parameters of multidimensional, continuous time Markov switching models. The observation process can be seen as a diffusion, where drift and volatility coefficients are modeled as continuous time, finite state Markov chains with a common state process. The states for drift and volatility and the rate matrix of the underlying Markov chain have to be estimated. Applications to simulated data indicate that the proposed algorithm can outperform the expectation maximization algorithm for difficult cases, e.g. for high rates. Application to financial market data shows that the Markov chain Monte Carlo method indeed provides sufficiently stable estimates.

Keywords: Bayesian inference, data augmentation, hidden Markov model, switching diffusion

JEL classification codes: C11, C13, C15, C32

1 Introduction

Typical dynamics of a price process $S = (S_t)_{t\in[0,T]}$ of n stocks are

$$dS_t = \mathrm{Diag}(S_t)(\mu_t\,dt + \sigma_t\,dW_t), \quad S_0 = s_0, \tag{1}$$

where $\mathrm{Diag}(S_t)$ denotes the diagonal matrix with diagonal $S_t = (S_t^1, \dots, S_t^n)$. Here $W = (W_t)_{t\in[0,T]}$ is an n-dimensional Brownian motion, s_0 is the initial price vector, $\mu = (\mu_t)_{t\in[0,T]}$ the drift process, $\sigma = (\sigma_t)_{t\in[0,T]}$ the volatility

*An updated version of this paper has appeared in the IFAS research paper series as no. 2009-41 and is available for download at `http://www.ifas.jku.at`

†Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Linz, Austria

‡Department for Applied Statistics and Econometrics, University Linz, Austria

process and $T > 0$ the terminal trading time. Suppose that μ and σ can take d possible values and that switching between these values is governed by a state process Y which is a continuous time Markov chain with state space $\{1, \ldots, d\}$ and rate matrix Q. Then the corresponding return process $R = (R_t)_{t \in [0,T]}$, defined by $dR_t = (\mathrm{Diag}(S_t))^{-1}\, dS_t$, satisfies

$$dR_t = \mu_t\, dt + \sigma_t\, dW_t \tag{2}$$

and represents a continuous time Markov switching model (MSM). In case of constant σ we call it a hidden Markov model (HMM). For a MSM in continuous time, in principle the state process Y can be observed via the quadratic variation of R. This is not possible for a HMM, Y is hidden.

Continuous time MSMs (e.g. Buffington and Elliott (2002), Guo (2001)), discrete time MSMs (e.g. Engel and Hamilton (1990)), and discrete as well as continuous time HMMs (e.g. Liechty and Roberts (2001), Sass and Haussmann (2004)) are widely used in finance, but are also applied in many other areas such as in biophysics (e.g. Rosales et al. (2001), Ball et al. (1999)).

In the present paper, we consider joint modeling of several stock prices using MSMs. Although stock prices are observed only in discrete time, like in many other inference problems encountered in finance, it is more convenient to use continuous time models since these allow the derivation of closed form solutions. For example, Sass and Haussmann (2004) derive explicit optimal trading strategies for a portfolio optimization problem based on a continuous time HMM. The practical application of these strategies, however, requires known values for the parameters of the continuous time stock model, and one faces the problem of estimating parameters of a continuous time model from discrete observations. Since in continuous time S and R generate the same filtration, we can assume that we observe R, hence we have to estimate parameters of a MSM like (2), or a HMM, if σ is constant.

One might use a moment based method, see e.g. Elliott et al. (2006), which yields good estimates, provided that the number of observations is very large (depending on the noise level, they present results for 10 000 to 20 000 observations for simulated data and 30 000 to 150 000 for market data).

In discrete time the expectation maximization (EM) algorithm (Dempster et al. (1977)) can be used to estimate the drift and covariance coefficients as well as the rate matrix governing the switching of the states, see e.g. Elliott et al. (1997) or Engel and Hamilton (1990). For low rates and low volatility the EM algorithm provides very accurate results, but estimation results in Sass and Haussmann (2004) indicate that for high rates, where a jump occurs between two observations with high probability, the EM algorithm performs poorly. For a continuous time HMM one can use a continuous time version of the EM algorithm as presented in James et al. (1996). For a continuous time MSM, however, no finite-dimensional filters are known, making it impossible to employ the EM algorithm to estimate the volatility jointly with the other parameters since the change of measure involved in deriving the filters used in the EM algorithm requires known σ (cf. Elliott et al. (1995)).

In this paper, we consider a Bayesian approach to parameter estimation, using Markov chain Monte Carlo (MCMC) methods, which is capable of dealing both with continuous time HMMs as well as continuous time MSMs. A Bayesian approach has been considered, for instance, by Rosales et al. (2001) and Boys et al. (2000) for discrete time HMMs, Ball et al. (1999) and Liechty and Roberts (2001) for continuous time HMMs, and Kim and Nelson (1999) for discrete time MSMs, dealing with problems in biophysics, bioinformatics and economics.

It is the aim of this paper to extend these methods to multidimensional continuous time HMMs and MSMs. We are particularly interested to estimate parameters in a multidimensional MSM with high rates, considerable noise, based on not too many observations (less than, say, 5000), as this is the typical situation one faces for many financial time series. We assume, however, that the number of states is given, and refer to Otranto and Gallo (2002) and Frühwirth-Schnatter (2006, Chapter 4 and 5) for the matter of model selection. As comparisons between the discrete and continuous models seem to be hardly treated in the relevant literature, we treat both the case of a discrete time approximation as well as methods for continuous time. We develop a Metropolis-Hastings sampler employing data augmentation by adding the state process to the unknowns. Hence, for the update of the rate matrix a Gibbs step can be used. In the discrete case, a Gibbs step is used also for the state process update, while in continuous time proposing the state process conditioning only on the rate matrix is much faster.

In Section 2, the stock return model which is a continuous time multidimensional MSM is introduced in more detail. In Section 3, MCMC estimation for continuous time state processes (referred to as CMCMC) is described. In Section 4, we present the approximating discrete time model and the corresponding MCMC algorithm (referred to as DMCMC). For both algorithms, we give the a priori and conditional distributions and describe the proposals and sampling methods used. We deal with the approximation error and the problem of computing the rate matrix corresponding to a transition matrix. In Section 5, we deal with issues concerning the application of the methods presented in Sections 3 and 4. Finally, we show numerical results for simulated as well as market data from daily stock index quotes.

2 Continuous Time Markov Switching Model

In this section we present the market model which is a multidimensional continuous time MSM. On a filtered probability space $(\Omega, \mathcal{F} = (\mathcal{F}_t)_{t\in[0,T]}, \mathrm{P})$

we observe in $[0,T]$ the n-dimensional process $R=(R_t)_{t\in[0,T]}$,

$$R_t = \int_0^t \mu_s\,ds + \int_0^t \sigma_s\,dW_s, \tag{3}$$

where $W=(W_t)_{t\in[0,T]}$ is an n-dimensional Brownian motion with respect to $\mathcal{F}$. The drift process $\mu=(\mu_t)_{t\in[0,T]}$ and the volatility process $\sigma=(\sigma_t)_{t\in[0,T]}$, taking values in $\mathbb{R}^n$ and $\mathbb{R}^{n\times n}$, respectively, are continuous time, time homogeneous, irreducible Markov processes with d states, adapted to $\mathcal{F}$ and independent of W, driven by the same state process $Y=(Y_t)_{t\in[0,T]}$. We denote the possible values of μ and σ by $B=(\mu^{(1)},\dots,\mu^{(d)})$ and $\Sigma=(\sigma^{(1)},\dots,\sigma^{(d)})$, respectively. We assume the volatility matrices $\sigma^{(k)}$ to be nonsingular and use the notation $C^{(k)}=\sigma^{(k)}\,(\sigma^{(k)})^\top$. Sometimes we will interpret B as a matrix, i.e. $B_{ik}=\mu_i^{(k)}$. The state process Y, which is a continuous time, time homogeneous, irreducible Markov chain adapted to $\mathcal{F}$, independent of W, with state space $\{1,\dots,d\}$, allows for the representations

$$\mu_t = \mu^{(Y_t)} = \sum_{k=1}^d \mu^{(k)}\,\mathbb{I}_{\{Y_t=k\}}, \qquad \sigma_t = \sigma^{(Y_t)} = \sum_{k=1}^d \sigma^{(k)}\,\mathbb{I}_{\{Y_t=k\}}. \tag{4}$$

The state process Y is characterized by the rate matrix $Q\in\mathbb{R}^{d\times d}$ as follows: For $\lambda_k=-Q_{kk}=\sum_{l=1,l\neq k}^d Q_{kl}$, $k=1,\dots,d$, in state k the waiting time for the next jump is λ_k-exponentially distributed, and the probability of jumping to state $l\neq k$ when leaving k is given by Q_{kl}/λ_k.

Starting from a prior distribution of the unknown parameters $\mathrm{P}(Q,B,\Sigma)$ we wish to determine $\mathrm{P}(Q,B,\Sigma\,|\,(V_m)_{m=1,\dots,N})$, the posterior distribution of these parameters given the observed data

$$V_m = \Delta R_{m\,\Delta t} = \int_{(m-1)\,\Delta t}^{m\,\Delta t} \mu_s\,ds + \int_{(m-1)\,\Delta t}^{m\,\Delta t} \sigma_s\,dW_s, \quad m=1,\dots,N. \tag{5}$$

Remark 1. One cannot distinguish between the pairs (σ,W) and $(\bar\sigma,\bar W)$, where $\bar\sigma$ is a square-root of $\sigma\sigma^\top$ and $\bar W=\bar\sigma^{-1}\sigma W$. However, without loss of generality we can assume $\sigma^{(k)}$ to be a lower triangular matrix, i.e. $\sigma^{(k)}(\sigma^{(k)})^\top$ equals the Cholesky factorization of the covariance matrix.

3 MCMC for Continuous Time State Process

In this section, we describe an MCMC algorithm for the continuous time model (referred to as CMCMC) to estimate the parameters Q, B and Σ given stock returns, $V=(V_m)_{m=1,\dots,N}$, observed at fixed observation times Δt, $2\Delta t,\dots,N\,\Delta t=T$. This method is easily extended to deal with non-equidistant observations, see Remark 3 below. We allow for jumps of the hidden state process at any time and especially for any number of jumps within each observation interval.

3.1 Data Augmentation

The state process Y, which is allowed to jump any time, is described by the process of jump times, $J = (J_h)_{h=0,\dots,H}$, and the sequence of states visited, $Z = (Z_h)_{h=0,\dots,H}$, where H is the number of jumps of Y in $[0, T]$, i.e. $J_0 = 0$ and $Z_0 = Y_0$, J_h is the time of the h-th jump, and Z_h is the state Y jumps to at the h-th jump. Hence the inter-arrival time $\Delta J_h = J_h - J_{h-1}$ is exponentially distributed with parameter $\lambda_{Z_{h-1}}$. Notice that J_{h+1} and Z_{h+1} are independent given J_h and Z_h.

For parameter estimation, we augment the parameter space by adding the state process Y, and determine the joint posterior distribution of Q, B, Σ, and Y given the observed data V.

3.2 Prior Distributions

Prior distributions have to be chosen for Q, B, Σ, and Y_0. We consider two prior specifications, differing in the prior assumptions concerning the initial state Y_0. The first prior is based on assuming prior independence among all parameters.

Assumption 1. Q, B, Σ, and Y_0 are a priori independent, i.e.

$$\mathrm{P}(Q, B, \Sigma, Y_0) = \mathrm{P}(Q)\ \mathrm{P}(B)\ \mathrm{P}(\Sigma)\ \mathrm{P}(Y_0). \tag{6}$$

Further, for $i = 1, \dots, n$, $k, l = 1, \dots, d$, $l \neq k$,

$$\begin{aligned} Q_{kl} &\sim \Gamma(f_{kl}, g_{kl}), \\ \mu_i^{(k)} &\sim \mathrm{N}(m_{ik}, s_{ik}^2), \\ C^{(k)} &\sim \mathrm{IW}(\Xi^{(k)}, \nu_k), \\ Y_0 &\sim \mathrm{U}(\{1, \dots, d\}). \end{aligned} \tag{7}$$

Also, the off-diagonal elements of Q are independent. With Γ, N, IW, and U we refer to the Gamma, normal, inverted Wishart, and uniform distribution, respectively.

Remark 2. For the inverted Wishart distribution, we use the parametrization where the density is given through

$$f_{\mathrm{IW}}(C; \Xi, \nu) \propto (\det C)^{-\nu-(n+1)/2} \exp\big(-\operatorname{tr}(\Xi C^{-1})\big) \tag{8}$$

and the expected value is given as $\mathrm{E}[C] = \Xi\,(\nu - (n+1)/2)^{-1}$.

If we think of time 0 as the beginning of our observations after the process has already run for some time, it may be reasonable to alter the prior distribution as follows:

Assumption 1a. Q and Y_0 are a priori dependent:

$$\mathrm{P}(Q, B, \Sigma, Y_0) = \mathrm{P}(Q)\ \mathrm{P}(Y_0 \,|\, Q)\ \mathrm{P}(B)\ \mathrm{P}(\Sigma), \tag{9}$$

and the state process starts from its ergodic probability ω, i.e. $\mathrm{P}(Y_0 \,|\, Q) = \omega$. All other prior assumptions are the same as in Assumption 1.

Under the second assumption, Y given Q is a stationary Markov chain, i.e. $\mathrm{P}(Y_t \,|\, Q) = \omega$ for all $t \in [0, T]$.

3.3 Full Conditional Posterior Distributions

To sample from the joint posterior distribution of Q, B, Σ, and Y given the observed data V, we partition the unknowns into three blocks, namely Q, (B, Σ), Y, and draw each of them from the appropriate conditional density.

3.3.1 Complete-Data Likelihood Function

As given B, Σ, and Y, the price process S is Markov and the returns $(V_m)_{m=1,\dots,N}$ are independent, the likelihood function is given by

$$\mathrm{P}(V \,|\, Q, B, \Sigma, Y) = \prod_{m=1}^{N} \varphi(V_m, \bar{\mu}_m, \bar{C}_m), \tag{10}$$

where φ denotes the density of a multivariate normal distribution with mean vector $\bar{\mu}_m$ and covariance matrix $\bar{C}_m$ given by:

$$\bar{\mu}_m = \int_{(m-1)\,\Delta t}^{m\,\Delta t} \mu^{(Y_s)}\, ds, \qquad \bar{C}_m = \int_{(m-1)\,\Delta t}^{m\,\Delta t} C^{(Y_s)} ds. \tag{11}$$

Remark 3. The algorithm presented in the following can be easily extended to non-equidistant observation times $0 = t_0 < t_1 < \cdots < t_N = T$ with distances $\Delta t_m = t_m - t_{m-1}$ by a slight adaptation in the data likelihood: In Equation (10), $\bar{\mu}_m$ and $\bar{C}_m$ have to be replaced by $\tilde{\mu}_m = \int_{t_{m-1}}^{t_m} \mu^{(Y_s)}\, ds$ and $\tilde{C}_m = \int_{t_{m-1}}^{t_m} C^{(Y_s)}\, ds$, respectively.

3.3.2 Drift and Volatility

The conditional joint distribution of drift and volatility is simply given by:

$$\mathrm{P}(B, \Sigma \,|\, V, Q, Y) \propto \mathrm{P}(V \,|\, B, \Sigma, Y)\ \mathrm{P}(B)\ \mathrm{P}(\Sigma). \tag{12}$$

3.3.3 State Process

The prior distribution of the state process Y_t for $t \in]0, T]$ is determined by the distribution of Y_0 and the rate matrix Q, and is independent of B and Σ. Therefore we obtain the following full conditional posterior:

$$\mathrm{P}(Y \,|\, V, Q, B, \Sigma) \propto \mathrm{P}(V \,|\, B, \Sigma, Y)\; \mathrm{P}(Y \,|\, Q). \tag{13}$$

The probability of Y given Q equals (cf. Ball et al. (1999))

$$\begin{aligned}
\mathrm{P}(Y \,|\, Q) &= \mathrm{P}(Y_0 \,|\, Q) \prod_{h=1}^{H} \left(\lambda_{Z_{h-1}} e^{-\lambda_{Z_{h-1}} \Delta J_h} \frac{Q_{Z_{h-1}, Z_h}}{\lambda_{Z_{h-1}}} \right) e^{-\lambda_{Z_H}(T - J_H)} &(14)\\
&= \mathrm{P}(Y_0 \,|\, Q) \prod_{k=1}^{d} \prod_{\substack{l=1\\ l \neq k}}^{d} \left(e^{-Q_{kl}\, O_T^k}\, Q_{kl}^{N_{kl}} \right), &(15)
\end{aligned}$$

where O_T^k denotes the occupation time of state k, and N_{kl} denotes the number of jumps from state k to l,

$$O_T^k = \int_0^T \mathbb{I}_{\{Y_t = k\}}\, dt, \qquad N_{kl} = \sum_{h=1}^{H} \mathbb{I}_{\{Z_{h-1} = k, Z_h = l\}}. \tag{16}$$

3.3.4 Rate Matrix

Using (15) and the independence of the priors of Q_{kl}, we have

$$\mathrm{P}(Q \,|\, V, B, \Sigma, Y) \propto \mathrm{P}(Y_0 \,|\, Q) \prod_{k=1}^{d} \prod_{\substack{l=1\\ l \neq k}}^{d} \psi_{kl}(Q_{kl}) \tag{17}$$

(cf. Ball et al. (1999)), where ψ_{kl} is a Gamma distribution with parameters $f_{kl} + N_{kl}$ and $g_{kl} + O_T^k$. If Y_0 is independent of Q, we can drop the term $\mathrm{P}(Y_0 \,|\, Q)$ and the off-diagonal elements Q_{kl}, $k \neq l$, are Gamma distributed.

3.4 Proposal Distributions

3.4.1 Drift and Volatility

For the update of B and Σ, a joint normal random walk proposal, reflected at 0 for the diagonal entries of $\sigma^{(k)}$, is used:

$$B' = B + r^B \phi, \quad \sigma_{ij}^{(k)\prime} = \sigma_{ij}^{(k)} + r^\sigma \psi_{ij}^{(k)}, \quad \sigma_{ii}^{(k)\prime} = \left| \sigma_{ii}^{(k)} + r^\sigma \psi_{ii}^{(k)} \right|, \tag{18}$$

where ϕ and ψ are $n \times d$ and $n \times n$ matrices of independent standard normal variates, $i = 1, \ldots, n$, $j = 1, \ldots, i-1$, $k = 1, \ldots, d$, and r^B and r^σ are parameters scaling the step widths.

As the transition kernel is symmetric, we have a Metropolis step with acceptance probability $\alpha_{B,\Sigma} = \min\{1, \bar{\alpha}_{B,\Sigma}\}$, where:

$$\bar{\alpha}_{B,\Sigma} = \frac{\mathrm{P}(V \mid B', \Sigma', Y)\ \mathrm{P}(B')\ \mathrm{P}(\Sigma')}{\mathrm{P}(V \mid B, \Sigma, Y)\ \mathrm{P}(B)\ \mathrm{P}(\Sigma)}. \tag{19}$$

Remark 4. The update of σ rather than $C = \sigma\sigma^\top$ guarantees that the resulting covariance matrix is positive semi-definite, as a symmetric matrix is positive definite if and only if there is a Cholesky decomposition: For every non-null vector v we have $v^\top C v = v^\top \sigma\sigma^\top v = (\sigma^\top v)^\top(\sigma^\top v) \geq 0$; if σ is non-singular, then C is positive definite.

Remark 5. A common problem that occurs in Bayesian methods for the analysis of Markov switching models is label switching (for a detailed discussion see e.g. Frühwirth-Schnatter (2006, Section 3.5)). Although due to the prior distribution (7) the posterior is not invariant to relabeling in our case, it is more effective to constrain the parameter space appropriately.

This can be done by proposing only values for B' that keep up a certain order, e.g. for $n = 1$ we can simply demand that $B'_{11} > ... > B'_{1d}$. Constraints for the multidimensional case can be introduced as follows. If we assume that the drift for each asset return R_i can take d_i different values $b_i^{l_i}$, $l_i = 1, \ldots, d_i$, we set the total number of states to be $d = \prod_{i=1}^n d_i$. Then $B = (\mu^{(1)}, \ldots, \mu^{(d)})$ consists of all possible vectors $(b_1^{l_1}, \ldots, b_n^{l_n})^\top$, where $l_i \in \{1, \ldots, d_i\}$ for $i \in \{1, \ldots, n\}$. The drift vectors $\mu^{(k)}$, $k = 1, \ldots, d$, can be sorted e.g. in lexicographical order, so $\mu^{(1)} = (b_1^1, \ldots, b_n^1)^\top$ contains the highest and $\mu^{(d)} = (b_1^{d_1}, \ldots, b_n^{d_n})^\top$ contains the lowest drift for each asset. This ordering can be easily maintained in each update, guaranteeing a unique labeling.

Even with constraints on the parameter space, label switching can still constitute a problem (cf. Stephens (2000)). In our examples, we never encountered such problems. However, if the posterior densities show evidence of label switching like multi-modality, methods like relabeling (Stephens (2000), Celeux (1998)), random permutation (Frühwirth-Schnatter (2001b) and Frühwirth-Schnatter (2001a)), or the use of invariant loss functions (Celeux et al. (2000), Hurn et al. (2003)) are necessary; a review of these methods is found in Jasra et al. (2005).

3.4.2 State Process

For updating the block $(Y_t)_{t\in[t_0,t_1[}$, $0 < t_0 < t_1 < T$, we generate a proposal $(Y'_t)_{t\in[0,t'[}$, $t' = t_1 - t_0$, as follows: First, we set $Z'_0 = Y_{t_0}$. Then we simulate the waiting time until the next jump time and the state the chain jumps to given the rate matrix Q. This is repeated until the jump time is greater than t', which is assumed to happen after $H'+1$ steps, i.e. there are H' jumps in $[0, t'[$. In order to fit the proposal Y' to Y, we have to consider three cases.

If $Z'_{H'} = Y_{t_1}$ we are done. If $Z'_{H'} \neq Y_{t_1}$ and $H' > 0$, we enforce $Z'_{H'} = Y_{t_1}$, possibly removing the last jump, if the chain was in state Y_{t_1} before the jump. Finally, if $Z'_{H'} \neq Y_{t_1}$ and $H' = 0$, we just start over.

So what is the probability of proposing some given Y'? Denote the originally proposed parameters by $\tilde{Y}$, $\tilde{J}$, $\tilde{Z}$, $\tilde{H}$, the adapted proposals by Y', J', Z', H', and $\overline{Y} = (Y_t)_{t \in [t_0, t_1[}$.

First assume $t_0 > 0$, $t_1 < T$, and $H' > 0$. A possible adaptation of $\tilde{Y}$ affects only the time interval $[J'_{H'}, t'[$. We distinguish between the cases $H' = \tilde{H}$ and $H' = \tilde{H} - 1$ to obtain

$$
\begin{aligned}
q(\overline{Y}, Y') = \prod_{h=1}^{H'-1} & \left(e^{-\lambda_{Z'_{h-1}} \Delta J'_h} Q_{Z'_{h-1}, Z'_h} \right) e^{-\lambda_{Z'_{H'-1}} \Delta J'_{H'}} \\
\times & \Bigg(\sum_{j \neq Z'_{H'-1}} Q_{Z'_{H'-1}, j}\, e^{-\lambda_j (t' - J'_{H'})} \\
& + Q_{Z'_{H'-1}, Z'_{H'}} \sum_{j \neq Z'_{H'}} Q_{Z'_{H'}, j} f(\lambda_{Z'_{H'}}, \lambda_j, t' - J'_{H'}) \Bigg),
\end{aligned} \tag{20}
$$

where

$$
f(\lambda_1, \lambda_2, t) = \begin{cases} \frac{e^{-\lambda_1 t} - e^{-\lambda_2 t}}{\lambda_2 - \lambda_1} & \text{if } \lambda_1 \neq \lambda_2, \\ t\, e^{-\lambda_1 t} & \text{if } \lambda_1 = \lambda_2. \end{cases}
$$

For $H' = 0$, where final and initial states coincide and $\tilde{H} \in \{0, 1\}$, q takes the simpler form

$$
q(\overline{Y}, Y') = e^{-\lambda_{Z'_0} t'} + \sum_{j \neq Z'_0} Q_{Z'_0, j}\, f(\lambda_{Z'_0}, \lambda_j, t'). \tag{21}
$$

For updating $(Y_t)_{t \in [0, t_1[}$, Y'_0 is sampled from the initial distribution of the state process and in (20) and (21) the factor $\mathrm{P}(Y'_0 \,|\, Q)$ enters.

For updating $(Y_t)_{t \in [t_0, T]}$, no adaptations are needed, i.e. $\tilde{Y} = Y'$; in (20) the second line is replaced with $Q_{Z'_{H'-1}, Z'_{H'}}\, e^{-\lambda_{Z'_{H'}} (t' - J'_{H'})}$ and in (21) the sum is dropped.

Set $\underline{Y} = (Y_t)_{t \in [0,T] \setminus [t_0, t_1[}$ and let $\overline{V}$ denote the set of observed data for time $[t_0, t_1[$. Then the conditional probability of the proposal restricted to this interval is given by

$$
\mathrm{P}(Y' \,|\, \overline{V}, B, \Sigma, Q, \underline{Y}) = \\
\mathrm{P}(\overline{V} \,|\, B, \Sigma, Y') \prod_{h=1}^{H'} \left(e^{-\lambda_{Z'_{h-1}} \Delta J'_h} Q_{Z'_{h-1}, Z'_h} \right) e^{-\lambda_{Z'_{H'}} (t' - J'_{H'})}. \tag{22}
$$

Combining (10), (22), and (20) (or (21)), we compute the acceptance probability $\alpha_Y = \min\{1, \bar{\alpha}_Y\}$, where:

$$\bar{\alpha}_Y = \frac{\mathrm{P}(\overline{V} \,|\, B, \Sigma, Y')}{\mathrm{P}(\overline{V} \,|\, B, \Sigma, \overline{Y})} \frac{\mathrm{P}(Y' \,|\, Q)}{q(\overline{Y}, Y')} \frac{q(Y', \overline{Y})}{\mathrm{P}(\overline{Y} \,|\, Q)}.$$

Comparing (14) and (20), note that most terms in $\mathrm{P}(Y' \,|\, Q)/q(\overline{Y}, Y')$ and $q(Y', \overline{Y})/\,\mathrm{P}(\overline{Y} \,|\, Q)$ cancel. For $t_0 = 0$, $\bar{\alpha}_Y$ is replaced by $\bar{\alpha}_Y^{(0)}$, while for $t_1 = T$, $\bar{\alpha}_Y$ simplifies to $\bar{\alpha}_Y^{(T)}$, where

$$\bar{\alpha}_Y^{(0)} = \frac{\mathrm{P}(Y_0' \,|\, Q)}{\mathrm{P}(\overline{Y}_0 \,|\, Q)} \bar{\alpha}_Y, \qquad \bar{\alpha}_Y^{(T)} = \frac{\mathrm{P}(\overline{V} \,|\, B, \Sigma, Y')}{\mathrm{P}(\overline{V} \,|\, B, \Sigma, \overline{Y})}.$$

3.4.3 Rate Matrix

To update the rate matrix, we sample from a Gamma distribution. For $k, l = 1, \ldots, d$, $l \neq k$, the proposal

$$Q'_{kl} \sim \Gamma(f_{kl} + N_{kl}, g_{kl} + O_T^k), \qquad Q'_{kk} = -\sum_{\substack{l=1 \\ l \neq k}}^{d} Q'_{kl} \tag{23}$$

is used. If the initial distribution of the state process $\mathrm{P}(Y_0)$ is independent of Q, then (23) is already a draw from the appropriate full conditional distribution, and we obtain a Metropolis-Hastings step with acceptance rate 1, i.e. a Gibbs step. Otherwise, using Assumption 1a, we accept the draw with probability $\alpha_Q = \min\{1, \bar{\alpha}_Q\}$, where $\bar{\alpha}_Q$ equals the ratio of the ergodic probabilities of Y_0 given the new and old rate matrix, i.e.

$$\bar{\alpha}_Q = \frac{\mathrm{P}(Y_0 \,|\, Q')}{\mathrm{P}(Y_0 \,|\, Q)} = \frac{\omega'}{\omega}. \tag{24}$$

4 MCMC for Discrete Time Approximation

In this section, we describe an algorithm (referred to as DMCMC) to estimate the parameters Q, B, and Σ given the stock returns at fixed observation times $\Delta t, 2\Delta t, \ldots, N\,\Delta t = T$ assuming that the state process jumps only at the end of these observation times. That means that over each observation time interval the drift of the return is constant. While this gives a good approximation of the continuous time model if the rates are not too high (compared to the time step Δt), it allows a better adaption of the algorithm to the model and can lead to more stable results. Finally, we give some considerations for the approximation error and discuss the problem of how to compute the rate matrix corresponding to some transition matrix.

4.1 The Model

To allow jumps of the state process at the observation times only, we replace the rate matrix Q by the transition matrix $X \in \mathbb{R}^{d \times d}$, where $X_{kl} = \mathrm{P}(Y_{t+\Delta t} = l \,|\, Y_t = k)$, i.e. $X = \exp(Q\,\Delta t)$. Then the probability of leaving state k is simply $1 - X_{kk}$ and the time spent in state k is geometrically distributed with parameter X_{kk}. The step process Y is fully described by its values at times $m\,\Delta t$, $m = 0, \ldots, N-1$, and the unknown state process reduces to $Y = (Y_m)_{m=0,\ldots,N-1}$; the same notation is used for μ and σ.

Starting from a prior distribution of the unknown parameters $\mathrm{P}(X, B, \Sigma)$, we determine the augmented posterior distribution $\mathrm{P}(X, B, \Sigma, Y \,|\, V)$, given the observed data $V = (V_m)_{m=1,\ldots,N}$, where in comparison to the previous section, V_m simplifies to

$$V_m = \mu_{m-1}\,\Delta t + \sigma_{m-1}\,(W_{m\,\Delta t} - W_{(m-1)\,\Delta t}). \tag{25}$$

4.2 Prior Distributions

Prior distributions have to be chosen for X, B, Σ, and Y_0, and as in Subsection 3.2, we consider two prior specifications, differing in the prior assumptions concerning the initial state Y_0. The independent priors are essentially the same as in (6), with $\mathrm{P}(Q)$ being substituted by $\mathrm{P}(X)$.

Assumption 2. X, B, Σ, and Y_0 are independent, i.e.

$$\mathrm{P}(X, B, \Sigma, Y_0) = \mathrm{P}(X)\ \mathrm{P}(B)\ \mathrm{P}(\Sigma)\ \mathrm{P}(Y_0), \tag{26}$$

where for $k = 1, \ldots, d$, the rows $X_{k\cdot}$ of X are assumed to be independent and to follow a Dirichlet distribution:

$$X_{k\cdot} \sim \mathrm{D}(g_{k1}, \ldots, g_{kd}). \tag{27}$$

The stationary prior is essentially the same as in (9), with $\mathrm{P}(Q)\ \mathrm{P}(Y_0 \,|\, Q)$ being substituted by $\mathrm{P}(X)\ \mathrm{P}(Y_0 \,|\, X)$.

Assumption 2a. X and Y_0 are a priori dependent:

$$\mathrm{P}(X, B, \Sigma, Y_0) = \mathrm{P}(X)\ \mathrm{P}(Y_0 \,|\, X)\ \mathrm{P}(B)\ \mathrm{P}(\Sigma), \tag{28}$$

and the state process starts from its ergodic probability ω, i.e. $\mathrm{P}(Y_0 \,|\, X) = \omega$. $\mathrm{P}(B)$ and $\mathrm{P}(\Sigma)$ are the same as in Assumption 1, $\mathrm{P}(X)$ is the same as in Assumption 2.

4.3 Sampling from the Full Conditional Posterior Distributions

To sample from the posterior, we partition the unknowns into X, (B, Σ), Y.

4.3.1 Complete-Data Likelihood Function

The complete-data likelihood function (10) reduces to:

$$\mathrm{P}(V \mid X, B, \Sigma, Y) = \prod_{m=1}^{N} \varphi(V_m, \mu^{(Y_{m-1})}\,\Delta t, C^{(Y_{m-1})}\,\Delta t), \tag{29}$$

where φ denotes the density of a multivariate normal distribution with mean vector $\mu^{(k)}$ and covariance matrix $C^{(k)}$, whenever $Y_{m-1} = k$.

4.3.2 Drift and Volatility

The conditional distribution of drift and volatility is the same as in (12). For the update of B and Σ, we proceed exactly as in the continuous case, using a joint random walk proposal. Note, however, that a more efficient Gibbs update would also be available, as a joint update of all elements of B conditional on Σ and of all elements of Σ conditional on B is feasible.

4.3.3 State Process

The prior distributions of the state process Y_m, $m = 1, \ldots, N-1$ is determined by the distribution of Y_0, and the transition probabilities X, and is independent of B and Σ. Therefore the full conditional posterior reads:

$$\mathrm{P}(Y \mid V, B, \Sigma, X) \propto \mathrm{P}(V \mid B, \Sigma, Y)\ \mathrm{P}(Y \mid X). \tag{30}$$

Defining $N_{kl} = \sum_{m=1}^{N-1} \mathbb{I}_{\{Y_{m-1}=k,\, Y_m=l\}}$, the number of transitions from state k to l, the probability of Y given X is

$$\mathrm{P}(Y \mid X) = \mathrm{P}(Y_0 \mid X) \prod_{m=1}^{N-1} \mathrm{P}(Y_m \mid Y_{m-1}, X) = \mathrm{P}(Y_0 \mid X) \prod_{k,l=1}^{d} X_{kl}^{N_{kl}}. \tag{31}$$

By forward-filtering-backward-sampling, see Frühwirth-Schnatter (2006), we may update Y drawing from the full conditional posterior $\mathrm{P}(Y \mid V, B, \Sigma, X)$. One such update, however, requires roughly $N \times d$ evaluations of an n-dimensional normal distribution and the generation of N discrete random numbers. Thus, this step (in particular, computing the filtered probabilities) is by far the most costly part of one update step. So we do not update the whole process Y in each step, but we randomly choose a block of approximately exponentially distributed lengths. Fixing the average block length has to be done weighing higher computation time in each step against faster mixing (compare Section 5.1).

4.3.4 Transition Matrix

For the update of the transition matrix, we use the method of sampling from a Dirichlet distribution as described e.g. in Frühwirth-Schnatter (2006). For each row $k = 1, \dots, d$ of the transition matrix, the proposal

$$X'_k \sim \mathrm{D}(g_{k1} + N_{k1}, \dots, g_{kd} + N_{kd}), \tag{32}$$

is used. If the initial distribution of the state process $\mathrm{P}(Y_0)$ is independent of X, then X'_k is a sample from the appropriate full conditional distribution, and we obtain a Gibbs step with acceptance rate 1. Otherwise (using Assumption 2a) we accept the draw with probability $\alpha_X = \min\{1, \bar{\alpha}_X\}$, where:

$$\bar{\alpha}_X = \frac{\mathrm{P}(Y_0 \,|\, X')}{\mathrm{P}(Y_0 \,|\, X)} = \frac{\omega'}{\omega}. \tag{33}$$

4.4 Discretization Error

The algorithm presented in this section is tailored to a discrete time model. Hence, we give some considerations about the error that arises from ignoring jumps within observation times.

The estimation of Q from a given (true) transition matrix X should introduce no error: the computation of Q from X via the matrix logarithm takes into account possible jumps occurring between the observation times.

The error in the estimated drift parameters occurs as follows: In the continuous model, B represents the instantaneous rates of return. If we assume Y, Q, and Σ to be known and denote the occupation time of state k in $[0, t]$ with O^k_t, we estimate in the discrete algorithm

$$\bar{B}_{ik} = \mathrm{E}[V^i_m \,\Delta t^{-1} \,|\, Y_{m-1} = k, Q, \Sigma] = \sum_{l=1}^{d} B_{il}\, \mathrm{E}\Big[\frac{O^l_{\Delta t}}{\Delta t} \,\Big|\, Y_0 = k, Q\Big]. \tag{34}$$

If the rates are high compared to the observation time interval, i.e. if the number of expected jumps within one period gets high, $\bar{B}_{ik}$ approaches $\sum_{l=1}^{d} B_{il}\, \omega_l$ regardless of k, while for low rates $\bar{B}_{ik}$ is close to B_{ik}.

Next we give a rough analysis of the covariance estimate. As Y and W are independent, the observed covariance of the returns is the sum of the covariances resulting from state jumps within observation intervals and the Brownian motion. As $\int_0^t \mu^i_s ds = \sum_{k=1}^{d} B_{ik}\, O^k_t$, we have given Y_0, Q, and B

$$\bar{C}^{(k)} = B \;\mathrm{Cov}[O_{\Delta t} \,|\, Y_0 = k, Q]\, B^\top \,\Delta t^{-1} + \sum_{l=1}^{d} C^{(l)}\, \mathrm{E}\Big[\frac{O^l_{\Delta t}}{\Delta t} \,\Big|\, Y_0 = k, Q\Big]. \tag{35}$$

However, estimating all parameters together, there is much interaction, e.g. if the drifts are underestimated, the volatility is overestimated.

4.5 Finding Generators for Transition Matrices

Employing DMCMC, we face the problem of computing the generator matrix Q corresponding to some transition matrix X for a fixed time step Δt. The problem of the existence of an adequate rate matrix (embedding problem) was already addressed by Elfving (1937) and in more detail by Kingman (1962), however, only recently the problem regained interest in the context of credit risk modeling, see e.g. Israel et al. (2001) for a collection of theoretical results and Kreinin and Sidelnikova (2001) for regularization algorithms for the computation of an (approximating) generator. Bladt and Sørensen (2005) describe how to find maximum likelihood estimators for Q using the EM algorithm or MCMC methods.

The problem turns out to be non-trivial for matrices of dimension greater than two. In general, there may exist no, one or more than one matrices Q such that $X = \exp(Q\,\Delta t)$ and Q is a valid generator. If the transition matrix is strictly diagonally dominant, than there is at most one generator (Israel et al., 2001, Remark 2.2), that is, the matrix logarithm of X exists uniquely, but need not yield a valid generator. A result in Culver (1966) implies the same if all eigenvalues of X are positive and distinct. If the (unique) matrix logarithm of X contains negative off-diagonal elements, we have to look for an approximating generator. Kreinin and Sidelnikova (2001) propose to regularize such invalid generators by projecting onto the space of valid generators. Another possibility is setting all negative off-diagonal entries, which are very small usually, to zero and adjusting all other elements proportional to their magnitude. Another problem occurs if there are multiple valid generators: Israel et al. (2001) state that choosing different generators results in different values $X_t(Q) = \exp(Q\,t)$ for most t. Then one can choose the generator that represents the transition matrix best in some sense, see Israel et al. (2001, Remark 5.1), for some considerations.

However, in our application there is another possibility to avoid these problems by restricting the parameter space for X to all uniquely embeddable transition matrices. This can be accomplished by rejecting and redrawing updates for X that do not have a unique corresponding generator.

5 Applications

In this section, we describe some details on the implementation. Then we present numerical results of the proposed algorithms both for simulated data and historical prices. Regarding the small number of observations and the relatively high volatility, we cannot expect to get very accurate results (especially for the rates, which are most difficult to estimate). However, we get reasonable results that are not visible to the naked eye. In particular, our methods turn out to have some advantages over the widely used EM algorithm.

5.1 Notes on the Implementation

We discuss how parameters of the prior and proposal distributions as well as initial values can be chosen.

5.1.1 Choosing the Prior

Although, asymptotically, the hyperparameters of the prior distributions have vanishing influence on the results, they should be chosen with care, as we are dealing with a limited number of observations, in order not to introduce some bias or predetermine the results too strongly. Slightly data dependent priors can be used to define the prior for the drift and volatility parameters, see, for instance, the market data example discussed in Subsection 5.2.2. For the transition matrix in DMCMC, the vector $g_{k\cdot}$ equals the a priori expectation of $X_{k\cdot}$ times a constant that determines the variance. If X^0 denotes our prior expectation of X, we may set $g_{k\cdot} = X^0_{k\cdot}\, c_k$. Then c_k can be interpreted as the number of observations for jumps out of state k in the prior distribution (added to the information contained in the data). The prior of the rate matrix for CMCMC is chosen in such a way that f_{kl} and g_{kl} in (7) are equal to the prior expectation of the number of jumps from k to l and the occupation time in state k, respectively, both times the same factor. Hence, denoting the expected rate matrix by Q^0, $f_{k\cdot}$ is set to $Q^0_{k\cdot}\, c_k$ and g_{kl} to c_k, where the constant c_k determines the variance of the prior distribution. Similar as in the discrete case, c_k can be interpreted as the time we observe state k a priori.

For simulated data hyperparameters can be chosen such that the prior expectation equals the true values in order to avoid a bias.

5.1.2 Running MCMC

To implement the Metropolis-Hastings algorithm, the scaling factors r^B and r^σ have to be selected in (18). We found that selecting r^B around 1 to 5 % of the minimal difference between two adjacent initial drift values worked well whereas r^σ is set to 1 % of the maximal element of the initial volatility matrix.

When updating the state process in CMCMC, the acceptance rate tends to be very low for proposal blocks that are too long. Hence we choose the average block length such that about 25 % of the proposals are accepted. Additionally, we found it useful to fix an upper bound for the block length. For the parameter determining the average and the maximal block length, suitable values were found by monitoring the acceptance in dependence on the block length in various test runs. In our numerical experiments an average of about 3 % and a maximal length of 7 % of the data available turned out to be a good choice resulting in an average acceptance about 25 %.

The same block length works well in DMCMC to reduce the computational costs of updating the whole state process.

Finally, starting from initial values for the trend, volatility, and rates (e.g. prior means), we generate an initial value for the state process by simulating from the smoothed discrete time state estimates.

5.2 Numerical Results

For 2 assets, 4 states, and 1500 observations, about 10 000 steps per minute can be performed with CMCMC. The discrete version is a little bit slower, as the smoothed sampling of the states is costly. Depending on the data, 50 000 up to 2 000 000 steps are needed for reliable results. Generally, the quality of the estimates is proportional to the difference between the drifts $\mu_i^{(k)} - \mu_i^{(k+1)}$ and indirect proportional to the magnitude of the variances $(C^{(k)})_{ii}$ and the rates λ_k. Clearly, the more data available the better the results are; this also implies that estimates for parameters for states that are visited less frequently are less reliable and have higher variance.

5.2.1 Comparison for Simulated Data

To compare DMCMC, CMCMC, and the EM algorithm we consider 500 samples of simulated prices ($N = 1500$, $T = 6$, $\Delta t = 1/250$) with continuous state processes and constant volatility.

For the EM algorithm we use initial values $(1, -1)$ for the drifts of both assets, whereas the off-diagonal elements of the initial rate matrix are set to 32. The covariances are pre-estimated using regressing for approximations of the quadratic variation process of the return for different step widths. This method was developed in Sass and Haussmann (2004), where we also refer to for a detailed description of the continuous time EM algorithm using robust filters for HMMs based on James et al. (1996) used here. 250 steps are sufficient for the EM algorithm to converge.

In the MCMC samplers 100 000 iterations are performed where the first 25 000 values are discarded. For the MCMC algorithms we use the following a priori distributions: For the drifts of both assets the mean values are set to $(2, -2)$ and the standard deviations to 2. For the volatility, the flat prior $p(C) \propto c$ is used for $C = \sigma\sigma^\top$, and we start from initial volatility $\mathrm{Diag}(0.2, 0.2)$. The mean transition matrix is set to its true value, the mean rate matrix has off-diagonal entries 17.2; the standard deviations for X and Q are

$$\begin{pmatrix} 0.23 & 0.18 & 0.15 & 0.11 \\ 0.15 & 0.22 & 0.15 & 0.11 \\ 0.11 & 0.15 & 0.22 & 0.15 \\ 0.11 & 0.15 & 0.18 & 0.23 \end{pmatrix}, \quad \begin{pmatrix} 0 & 12.9 & 10.0 & 6.9 \\ 10.6 & 0 & 10.0 & 6.9 \\ 6.9 & 10.0 & 0 & 10.6 \\ 6.9 & 10.0 & 12.9 & 0 \end{pmatrix}, \tag{36}$$

respectively. The state process is initialized by sampling from the smoothed state estimates for a discrete process given the initial values for the remaining parameters.

Table 1 provides a comparison of DMCMC, CMCMC, and the EM algorithm with respect to their statistical efficiency in parameter estimation. For all algorithms we evaluate the estimators of the drift, the covariance matrix and the rate matrix, by computing the average of all estimators over the 500 replications as well as their root mean square errors. It turns out that DMCMC yields appealing results for all parameters. The estimates for the drifts reflect the approximation error. Both CMCMC and the EM algorithm fail to provide exact estimates for the rates, but they manage to identify the underlying "structure". However, they are clearly outperformed by DMCMC.

The results in Table 1 suggest that CMCMC tends to over-estimate the rates. Indeed, further tests showed that especially for processes with high rates, the rate estimates sometimes do not enter into a stationary distribution but continue growing which corresponds to inserting more and more jumps to the state process estimate. As this does not necessarily interfere with the average state estimation at each time (which works well nevertheless), the likelihood function appears to carry insufficient information. Notice that to get aware of this phenomenon it is necessary to have the MCMC algorithm run for quite some time.

5.2.2 Daily Stock Index Data

As an example for financial market data we consider daily data for the Dow Jones Industrial Average Index (DJIA) from 1957 to 2006 (see Figure 1). The data was organized in 45 overlapping blocks of six consecutive years' quotes, each comprising about 1500 data points.

For each set of six years' quotes we employed DMCMC to fit a MSM. Assuming that there are $d = 3$ states seems to be a reasonable choice, as it allows for an up-, down-, and zero-state of the drift, while the number of parameters is still moderate.

For each run, the parameters for the prior information of the drift are set to

$$m_1 = \hat{q}_{0.85}\,\Delta t^{-1}, \qquad m_2 = \hat{m}\,\Delta t^{-1}, \qquad m_3 = \hat{q}_{0.15}\,\Delta t^{-1}, \qquad (37)$$

and $s_k = s = 0.25(m_1 - m_3)$, where $\hat{m}$, $\hat{q}_{0.15}$, and $\hat{q}_{0.85}$ denote mean and 15% as well as 85% quantiles of the daily returns under investigation, respectively. For the volatilities we set

$$\Xi^{(1)} = \Xi^{(3)} = 1.44\,\hat{\sigma}^2\Delta t^{-1}(\nu - 1), \qquad \Xi^{(2)} = 0.64\,\hat{\sigma}^2\Delta t^{-1}(\nu - 1), \qquad (38)$$

and $\nu_k = \nu = 3$, where $\hat{\sigma}^2$ denotes the variance of the daily returns. For the transition matrix, the matrix of parameters for the Dirichlet distribution

$\sigma\sigma^\top$		$B^\top$		Q			
		1.50	1.50	−94.1	49.5	30.3	14.3
0.0225	0.0075	1.50	−1.50	33.6	−78.2	30.3	14.3
0.0075	0.0250	−1.50	1.50	14.3	30.3	−78.2	33.6
		−1.50	−1.50	14.3	30.3	49.5	−94.1
DMCMC							
		1.43	1.38	−99.8	49.7	34.6	15.5
0.0225	0.0076	1.43	−1.40	35.8	−77.6	24.6	17.2
0.0076	0.0251	−1.43	1.38	17.1	24.5	−75.4	33.9
		−1.43	−1.40	15.1	36.8	50.3	−102.1
		0.17	0.24	32.8	27.6	24.0	18.5
0.0012	0.0008	0.17	0.23	20.3	21.6	14.3	14.5
0.0008	0.0015	0.18	0.24	14.3	15.1	19.3	18.2
		0.18	0.23	16.5	28.4	29.0	38.0
CMCMC							
		1.47	1.38	−124.9	67.7	38.4	18.7
0.0233	0.0073	1.47	−1.37	51.0	−100.3	29.8	19.4
0.0073	0.0266	−1.46	1.38	19.1	29.7	−100.3	51.5
		−1.46	−1.37	18.5	38.8	68.0	−125.3
		0.16	0.25	32.3	19.6	10.0	6.0
0.0013	0.0008	0.16	0.25	19.0	24.0	5.9	7.1
0.0008	0.0022	0.16	0.25	6.9	5.8	24.1	19.5
		0.16	0.25	5.5	10.3	19.9	32.6
EM							
		1.40	1.33	−71.7	31.1	26.3	14.3
0.0229	0.0074	1.48	−1.50	29.8	−61.0	14.8	16.4
0.0074	0.0255	−1.50	1.44	18.6	15.4	−60.2	26.1
		−1.41	−1.37	14.5	24.6	31.2	−70.2
		0.54	0.64	32.2	27.4	18.3	15.3
0.0026	0.0021	0.39	0.46	21.9	27.2	19.7	14.6
0.0021	0.0031	0.38	0.41	16.4	19.4	25.0	18.1
		0.54	0.63	15.6	19.3	27.3	34.1

Table 1: Comparison of DMCMC, CMCMC and EM for simulated continuous data (top: true values, below: average and mean root square errors of estimators over 500 samples for each method)

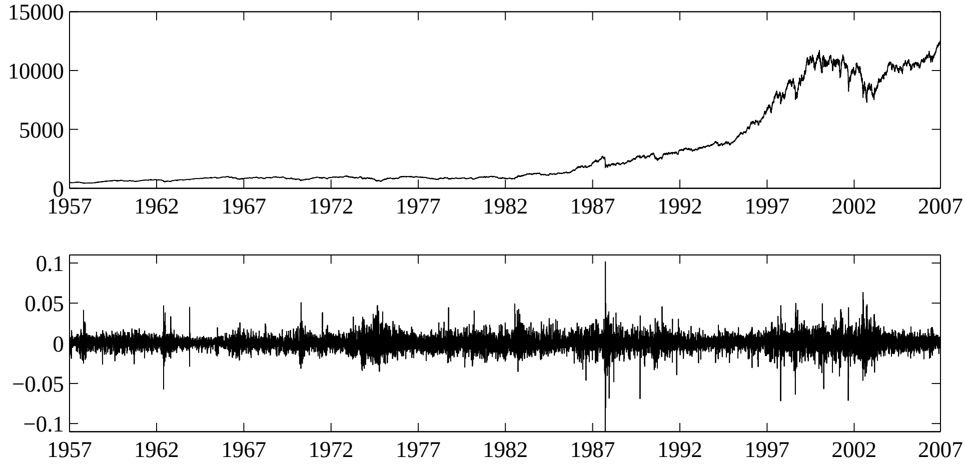

Figure 1: DJIA quotes (top) and daily returns (bottom) for 01/1957–12/2006

(compare (27)) is set to

$$g = \begin{pmatrix} 0.64 & 0.23 & 0.13 \\ 0.06 & 0.88 & 0.06 \\ 0.13 & 0.23 & 0.64 \end{pmatrix} c, \tag{39}$$

where $c = 75$. This means that on average we expect jumps after 3 observations in state 1 and 3 and after 8 observations in state 2.

Running DMCMC, 1 500 000 MCMC steps are performed of which the first 300 000 are discarded. The resulting mean estimates (using data from the year annotated and the five years before) are given in Table 2 and Figures 2 to 5, where the dashed lines give 10% and 90% quantiles. Visual inspection of the sample paths indicates quick and stable convergence to the equilibrium distribution. For a more rigorous justification, we performed multiple runs from different initial values and computed the potential scale reduction factor $\sqrt{\hat{R}}$ (Gelman and Rubin, 1992) for each single component as well as the multivariate potential scale reduction factor (Brooks and Gelman, 1998) for all components of B, Σ, and Q together; for all tests, we get $\sqrt{\hat{R}} < 1.05$, and in fact we mostly get $\sqrt{\hat{R}} < 1.01$.

Typical values of ± 1.5 for the drift and 0.15 for the volatility correspond to average daily returns about $\pm 0.6\%$ with a volatility about 1%, showing a rather low signal-to-noise ratio. For the rates, values between 60 and 120 as in states 1 and 3 correspond to average sojourn times of 2 to 4 days, values between 15 and 20 as in state 2 correspond to average sojourn times of 13 to 17 days.

Time	$\mu^{(1)}$	$\mu^{(2)}$	$\mu^{(3)}$	$\sigma^{(1)}$	$\sigma^{(2)}$	$\sigma^{(3)}$	Q_{11}	Q_{12}	Q_{13}	Q_{21}	Q_{22}	Q_{23}	Q_{31}	Q_{32}	Q_{33}
1962	1.09	0.29	-1.53	0.233	0.084	0.108	-116.0	62.1	53.9	1.5	-15.8	14.3	60.5	46.5	-107.0
1963	1.06	0.31	-1.61	0.249	0.082	0.088	-113.0	62.0	51.1	1.2	-14.3	13.1	62.5	58.2	-120.8
1964	1.04	0.27	-1.44	0.233	0.075	0.094	-112.5	62.4	50.2	1.5	-14.6	13.2	62.1	48.8	-110.9
1965	1.03	0.25	-1.27	0.228	0.070	0.098	-106.0	56.6	49.4	1.9	-13.6	11.7	53.5	42.6	-96.1
1966	1.14	0.22	-1.28	0.232	0.071	0.110	-101.3	54.1	47.2	1.6	-12.2	10.6	45.9	40.8	-86.7
1967	1.04	0.23	-1.28	0.224	0.069	0.101	-94.3	52.1	42.1	1.9	-14.0	12.1	46.6	42.3	-88.9
1968	1.13	0.25	-1.06	0.160	0.065	0.090	-111.8	67.0	44.8	3.6	-17.6	14.0	41.4	38.7	-80.1
1969	1.36	0.21	-1.20	0.110	0.062	0.091	-119.2	69.6	49.6	7.8	-27.2	19.4	47.2	34.0	-81.2
1970	1.46	0.19	-1.49	0.162	0.074	0.099	-108.1	66.5	41.7	6.3	-26.5	20.3	48.2	46.8	-95.0
1971	1.36	0.25	-1.45	0.180	0.080	0.094	-100.2	61.9	38.3	5.4	-31.4	26.0	41.3	56.7	-98.1
1972	1.17	0.39	-1.19	0.199	0.080	0.083	-92.9	60.1	32.8	5.0	-36.8	31.9	28.2	67.4	-95.6
1973	1.04	0.24	-1.83	0.217	0.087	0.087	-77.2	43.9	33.2	3.9	-25.9	21.9	47.6	70.8	-118.4
1974	2.05	0.05	-2.55	0.214	0.101	0.140	-107.5	38.0	69.5	5.3	-12.6	7.3	73.2	32.4	-105.7
1975	2.30	0.09	-2.54	0.199	0.105	0.142	-108.1	38.0	70.1	6.6	-16.2	9.6	80.8	30.3	-111.1
1976	2.20	0.13	-2.46	0.198	0.109	0.142	-107.4	39.4	68.1	5.8	-15.2	9.4	78.1	34.1	-112.2
1977	2.24	0.05	-2.62	0.199	0.114	0.145	-107.3	39.4	67.9	5.2	-12.1	6.9	78.9	33.3	-112.2
1978	2.38	0.04	-2.72	0.218	0.124	0.143	-115.7	51.0	64.7	5.9	-14.1	8.2	72.5	39.8	-112.3
1979	2.30	0.07	-2.48	0.216	0.117	0.145	-111.6	53.3	58.3	5.5	-13.6	8.1	63.9	44.0	-107.9
1980	1.83	0.17	-1.69	0.175	0.112	0.133	-119.1	68.9	50.3	8.3	-26.1	17.8	58.8	61.6	-120.4
1981	1.63	0.12	-1.62	0.176	0.113	0.134	-131.8	78.0	53.8	8.5	-23.9	15.3	50.0	68.8	-118.8
1982	1.78	0.05	-1.75	0.242	0.118	0.141	-110.2	55.2	55.0	5.7	-15.0	9.3	53.2	64.2	-117.5
1983	1.97	0.12	-1.72	0.244	0.125	0.145	-113.9	57.7	56.2	5.2	-14.2	9.0	55.5	61.1	-116.6
1984	2.06	0.03	-1.44	0.229	0.124	0.152	-122.9	65.1	57.8	8.1	-16.9	8.8	57.2	63.5	-120.7
1985	2.08	0.06	-1.35	0.227	0.122	0.150	-123.6	65.4	58.2	8.3	-16.9	8.5	58.6	59.9	-118.5
1986	1.89	0.06	-1.20	0.233	0.121	0.160	-120.1	64.9	55.2	8.6	-18.0	9.5	57.1	70.1	-127.2
1987	1.15	0.13	-2.05	0.271	0.129	0.763	-61.2	39.6	21.6	5.3	-7.2	1.8	53.1	63.4	-116.5
1988	1.43	0.16	-2.06	0.269	0.131	0.750	-92.1	58.6	33.5	6.1	-8.1	2.0	58.1	63.0	-121.1
1989	1.33	0.18	-2.01	0.265	0.126	0.750	-92.4	60.3	32.0	7.1	-9.5	2.4	58.1	65.0	-123.1
1990	0.73	0.26	-2.01	0.260	0.125	0.732	-74.2	49.3	25.0	8.1	-10.7	2.6	57.0	66.8	-123.8
1991	0.91	0.22	-2.11	0.269	0.131	0.729	-84.1	56.3	27.8	8.4	-11.4	3.0	58.0	69.8	-127.8
1992	0.80	0.16	-2.05	0.260	0.121	0.746	-73.5	49.8	23.7	8.2	-10.8	2.5	56.7	67.1	-123.8
1993	1.19	0.10	-1.35	0.173	0.099	0.271	-121.8	80.9	40.9	14.7	-27.6	12.9	66.2	75.7	-141.9
1994	1.38	0.13	-1.29	0.182	0.095	0.215	-129.9	83.6	46.3	9.4	-21.6	12.2	57.9	68.0	-125.9
1995	1.44	0.17	-1.35	0.180	0.088	0.174	-124.5	74.7	49.8	7.0	-19.4	12.4	59.4	59.0	-118.4
1996	1.31	0.18	-1.03	0.148	0.084	0.162	-122.8	74.7	48.1	8.9	-23.0	14.1	60.8	64.1	-124.9
1997	1.17	0.19	-1.04	0.145	0.086	0.207	-111.8	63.5	48.3	7.8	-20.2	12.4	64.9	54.4	-119.3
1998	0.62	0.25	-1.27	0.153	0.089	0.292	-68.9	35.8	33.1	11.1	-18.0	6.9	63.8	52.4	-116.2
1999	1.12	0.32	-1.03	0.243	0.102	0.231	-101.1	56.4	44.7	8.4	-21.1	12.7	51.0	50.2	-101.2
2000	1.42	0.38	-0.80	0.308	0.111	0.216	-113.9	59.7	54.2	5.4	-20.0	14.7	41.3	37.3	-78.6
2001	1.97	0.20	-1.94	0.234	0.145	0.335	-126.3	76.0	50.3	5.3	-13.5	8.2	59.6	51.1	-110.7
2002	2.48	0.14	-2.45	0.352	0.164	0.332	-117.0	63.0	54.1	3.0	-8.9	5.9	60.7	45.7	-106.4
2003	2.40	0.13	-2.39	0.335	0.160	0.309	-111.9	57.5	54.4	3.5	-9.7	6.1	61.7	45.7	-107.4
2004	2.13	0.08	-1.97	0.309	0.146	0.276	-110.4	54.5	55.9	4.0	-10.6	6.6	60.6	40.7	-101.4
2005	1.16	0.06	-0.62	0.334	0.114	0.229	-84.2	32.7	51.5	3.8	-10.2	6.4	33.3	19.9	-53.3
2006	1.24	0.10	-0.46	0.353	0.103	0.189	-82.5	33.2	49.3	2.6	-8.7	6.2	26.1	18.5	-44.6

Table 2: Estimation results for DJIA data: mean estimates for drift, volatility, and rate matrix

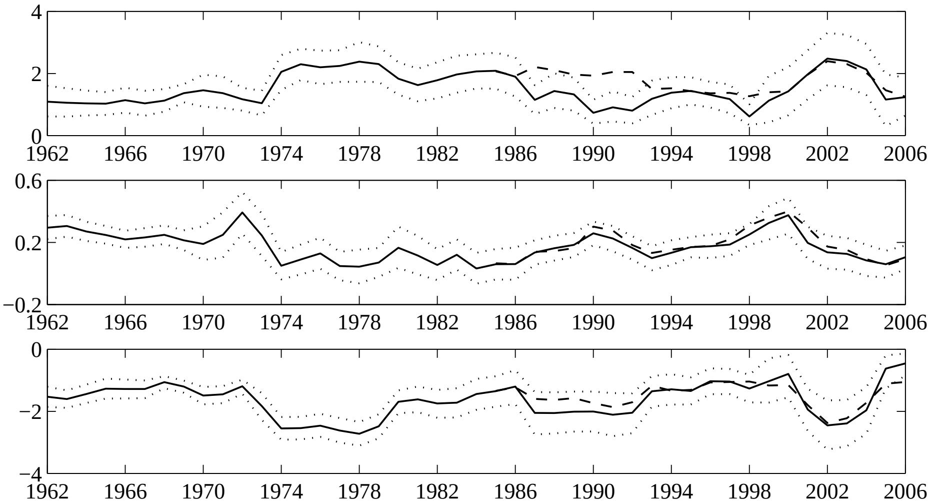

Figure 2: Mean estimates (solid), 10% and 90% quantiles (dotted), and mean estimates ignoring jumps (dashed) for drift (top: $\mu^{(1)}$, middle: $\mu^{(2)}$, bottom: $\mu^{(3)}$)

How can the results be interpreted? States 1 and 3 correspond to extreme up and down movements, respectively, both with short sojourn times. They come together with a conspicuously higher volatility than state 2. Considering the history of the DJIA, the results are quite plausible. The years before 1974 where a period of rather low volatility, followed by a bullish period (note the higher estimates for drift and volatility in the up-state). The crash in October 1987 with a drop of 23% only on October 19 dramatically influences the estimates for 1987 and the following years (also including the Kuweit crisis in 1990) – the volatility for the down state explodes. Afterwards, markets recovered until in the late nineties, with the drop in August 1998 a very volatile period started, where from 1998–2002 movements greater than 400 points happened 8 times and greater than 250 points 56 times.

In fact, a simple 3 state MSM cannot account for such pronounced jumps. So we re-estimated the period from 1985 to 2006 discarding daily returns greater than 5%. This results in more moderate estimates evolving more smoothly over time (see the dashed lines in the plots). However, for a more thorough analysis of the periods containing large movements, one could introduce additional states reserved for extreme values of drift and volatility and very short sojourn times.

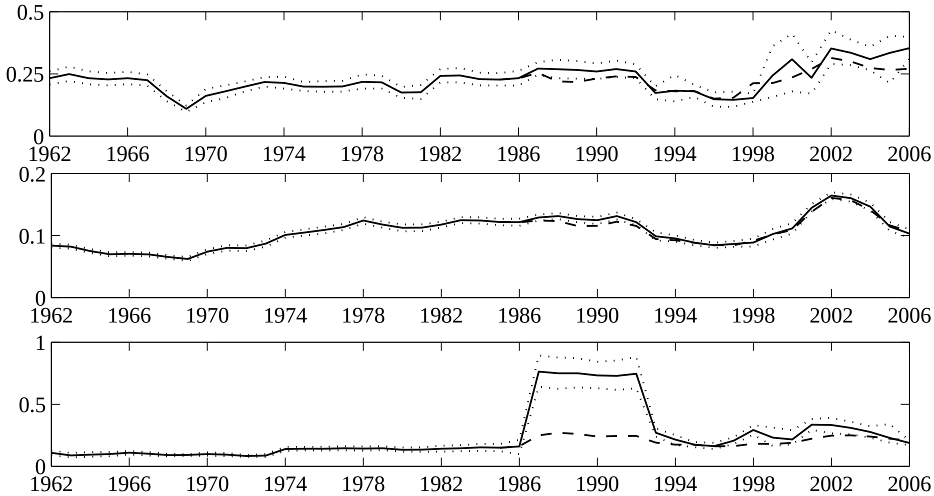

Figure 3: Mean estimates (solid), 10% and 90% quantiles (dotted), and mean estimates ignoring jumps (dashed) for volatility (top: $\sigma^{(1)}$, middle: $\sigma^{(2)}$, bottom: $\sigma^{(3)}$)

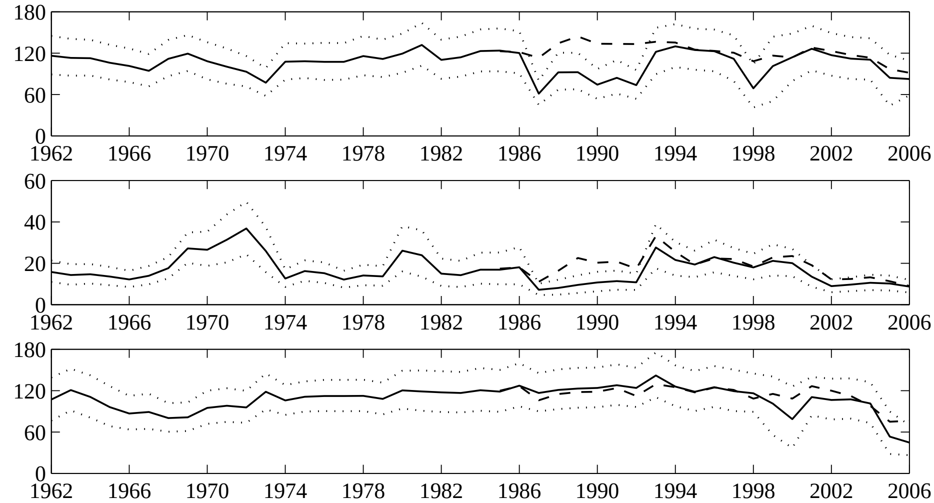

Figure 4: Mean estimates (solid), 10% and 90% quantiles (dotted), and mean estimates ignoring jumps (dashed) for rates (top: λ_1, middle: λ_2, bottom: λ_3)

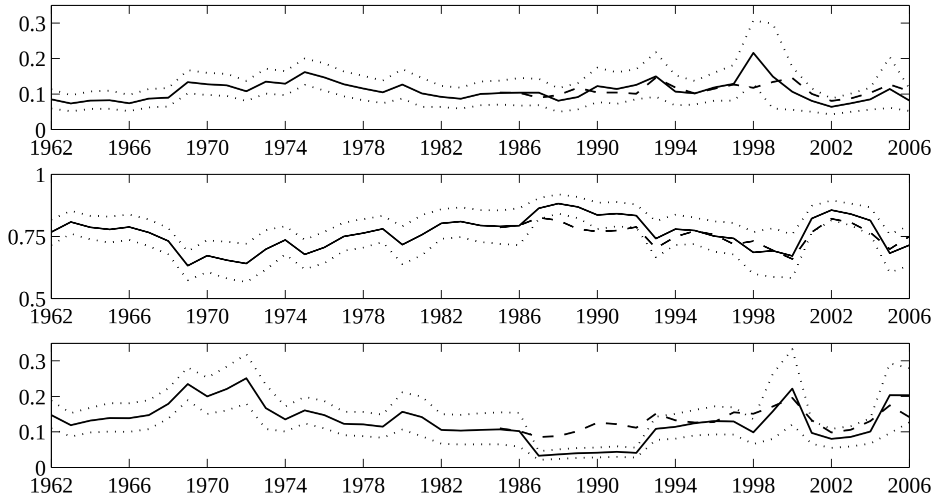

Figure 5: Mean estimates (solid), 10% and 90% quantiles (dotted), and mean estimates ignoring jumps (dashed) for stationary distribution (top: ω_1, middle: ω_2, bottom: ω_3)

6 Conclusion

We present a discrete and a continuous time MCMC algorithm for parameter estimation in multidimensional continuous time MSMs. We apply the methods to simulated as well as historical stock data and compare the results with those of the EM algorithm. In the cases where the EM algorithm works fine (low rates, low volatility, much data), it needs few steps and is much faster. But as soon as we deal with more challenging parameters, there often seems to be no convergence to plausible values, while the MCMC sampler still works well.

The continuous time MCMC algorithm is attractive for its easy extensibility to non-equidistant observation times. Moreover, it allows estimation of switching volatility without introducing a discretization error. However, while working well for moderate parameters, it turns out not to be sufficiently robust for extreme parameters. On the other hand, the discrete method outperforms the EM algorithm giving better results for simulated data with high rates and few observations. Further, working more stable, the presented discrete algorithm manages to provide plausible results for financial data with moderate volatility, e.g. stock indices.

Acknowledgment

The first author acknowledges financial support by the Austrian Science Fund (FWF-projects S8305 and P17947-N12).

References

F. G. Ball, Y. Cai, J. B. Kadane, and A. O'Hagan. Bayesian inference for ion-channel gating mechanisms directly from single-channel recordings, using Markov chain Monte Carlo. *Proceedings of the Royal Society Series A*, 455:2879–2932, 1999.

M. Bladt and M. Sørensen. Statistical inference for discretely observed Markov jump processes. *Journal of the Royal Statistical Society Series B*, 67(3):295–410, 2005.

R. J. Boys, D. A. Henderson, and D. J. Wilkinson. Detecting homogeneous segments in DNA sequences by using hidden Markov models. *Applied Statistics*, 49:269–285, 2000.

S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998.

J. Buffington and R. J. Elliott. American options with regime switching. *International Journal of Theoretical and Applied Finance*, 5:497–514, 2002.

G. Celeux. Bayesian inference for mixtures: The label-switching problem. In R. Payne and P. J. Green, editors, *COMPSTAT 98 – Proceedings in Computational Statistics*, pages 227–232, Heidelberg, 1998. Physica.

G. Celeux, M. Hurn, and C. P. Robert. Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95:957–970, 2000.

W. J. Culver. On the existence and uniqueness of the real logarithm of a matrix. *Proceedings of the American Mathematical Society*, 17:1146–1151, 1966.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, 39(1):1–38, 1977.

G. Elfving. Zur Theorie der Markoffschen Ketten. *Acta Social Science Fennicae n. Series A*, 2(8), 1937.

R. J. Elliott, L. Aggoun, and J. B. Moore. *Hidden Markov Models – Estimation and Control*, volume 29. Springer-Verlag, New York, 1995.

R. J. Elliott, W. C. Hunter, and B. M. Jamieson. Drift and volatility estimation in discrete time. *Journal of Economic Dynamics and Control*, 22: 209–218, 1997.

R. J. Elliott, V. Krishnamurthy, and J. Sass. Moment based regression algorithm for drift and volatility estimation in continuous time Markov switched models, 2006.

C. Engel and J. D. Hamilton. Long swings in the dollar: Are they in the data and do markets know it? *American Economic Review*, 80(4):689–713, 1990.

S. Frühwirth-Schnatter. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96(453):194–209, 2001a.

S. Frühwirth-Schnatter. Fully Bayesian analysis of switching Gaussian state space models. *Annals of the Institute of Statistical Mathematics*, 53(1): 31–49, 2001b.

S. Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models.* Springer, New York, 2006.

A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–72, 1992.

X. Guo. An explicit solution to an optimal stopping problem with regime switching. *Stochastic Processes and their Applications*, 38:464–481, 2001.

M. Hurn, A. Justel, and C. P. Robert. Estimating mixtures of regressions. *Journal of Computational and Graphical Statistics*, 12:55–79, 2003.

R. B. Israel, J. S. Rosenthal, and J. Z. Wei. Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings. *Mathematical Finance*, 11:245–265, 2001.

M. James, V. Krishnamurthy, and F. Le Gland. Time discretization of continuous-time filters and smoothers for HMM parameter estimation. *IEEE Transactions on Information Theory*, 42:593–605, 1996.

A. Jasra, C. C. Holmes, and D. A. Stephens. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, 20(1):50–67, 2005.

C.-J. Kim and C. R. Nelson. Has the U.S. economy become more stable? A Bayesian approach on a Markov-switching model of the business cycle. *The Review of Econometrics and Statistics*, 81:608–616, 1999.

J. F. C. Kingman. The imbedding problem for finite Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 1:14–24, 1962.

A. Kreinin and M. Sidelnikova. Regularization algorithms for transition matrices. *Algo Research Quarterly*, 4:23–40, 2001.

J. C. Liechty and G. O. Roberts. Markov chain Monte Carlo methods for switching diffusion models. *Biometrika*, 88(2):299–315, 2001.

E. Otranto and G. M. Gallo. A nonparametric Bayesian approach to detect the number of regimes in Markov switching models. *Econometric Reviews*, 21(4):477–496, 2002.

R. Rosales, J. A. Stark, W. J. Fitzgerald, and S. B. Hladky. Bayesian restoration of ion channel records using hidden Markov models. *Biophysical Journal*, 80(3):1088–1103, 2001.

J. Sass and U. G. Haussmann. Optimizing the terminal wealth under partial information: The drift process as a continuous time Markov chain. *Finance and Stochastics*, 8(4):553–577, 2004.

M. Stephens. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society Series B*, 62(4):795–809, 2000.