Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2009-41

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

Markus Hahn[a], Sylvia Frühwirth-Schnatter[b], Jörn Sass[c]

March 2009

A previous version of this research report appeared as RICAM-Report No. 2007-09 in May 2007.

[a]Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Linz
[b]Department of Applied Statistics and Econometrics, Johannes Kepler Universität Linz
[c]Department of Mathematics, University of Kaiserslautern

Abstract

We consider a multidimensional, continuous time model where the observation process is a diffusion with drift and volatility coefficients being modeled as continuous time, finite state Markov chains with a common state process. For the econometric estimation of the states for drift and volatility and the rate matrix of the underlying Markov chain, we develop both an exact continuous time as well as an approximate discrete time MCMC sampler and compare these approaches with ML estimation. For simulated data, MCMC outperforms ML estimation for difficult cases like high rates. Finally, for daily stock index quotes from Argentina, Brazil, Mexico, and the US we identify four states differing not only in the volatility of the various assets, but also in their correlation.

Keywords: Bayesian inference, data augmentation, hidden Markov model, switching diffusion
JEL classification codes: C11, C13, C15, C32

1 Introduction

Discrete time Markov switching models have been vastly used in many areas including econometrics, biosciences, image processing, and speech processing, see Frühwirth-Schnatter (2006) for a recent overview. Applications in financial econometrics, where the latent state process Y is usually interpreted as some unobservable underlying economic variable, include modeling of exchange rate dynamics (Engel and Hamilton, 1990), interest rates and term structures (Garcia and Perron, 1996; Bansal and Zhou, 2002), stock market returns (Schaller and Van Norden, 1997), and the business cycle (Kim and Nelson, 1999). Typically, these applications deal with univariate time series, exceptions are the recent papers by Guidolin and Timmermann (2006, 2007), where multivariate modeling and asset allocation are considered. Rydén et al. (1998) demonstrate that even very simple discrete time Markov switching models are able to capture a number of stylized facts (both distributional and temporal) observed in asset returns. Investigating potential patterns of volatility, skewness, kurtosis, and autocovariance, Timmermann (2000) extends the analysis of Rydén et al. (1998) and approves that this model is in many respects very well suited to the application to financial time series.

In recent years, also interest in continuous time Markov switching models has been increasing. Assume, for instance, that the dynamics of a price process $S = (S_t)_{t\in[0,T]}$ of n stocks are describes as

$$\mathrm{d}S_t = \mathrm{Diag}(S_t)(\mu_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t), \quad S_0 = s_0,$$

where $\mathrm{Diag}(S_t)$ denotes the diagonal matrix with diagonal $S_t = (S_t^1, \ldots, S_t^n)$. 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 process, and $T > 0$ the time horizon. 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 $\mathrm{d}R_t = (\mathrm{Diag}(S_t))^{-1}\,\mathrm{d}S_t$, satisfies

$$\mathrm{d}R_t = \mu_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t \tag{1}$$

and represents a multivariate continuous time Markov switching model (MSM). Although in most applications data is observed only in discrete time, in a variety of problems it is convenient to use such a continuous time model since it allows the derivation of closed form solutions.

At first sight, such a continuous time MSM might seem too simplistic for applications in finance like modeling stock returns, as it obviously does not allow e.g. for jumps in the price process. However, inheriting the properties of the discrete time model analyzed in Rydén et al. (1998) and Timmermann (2000), it turns out to be surprisingly flexible and powerful; in particular, it represents a substantial improvement over the still heavily used Black-Scholes model with constant drift and volatility. Recent applications of this model include short rate models (Elliott et al., 2001), option pricing (Guo, 2001a,b; Buffington and Elliott, 2002a,b; Chan et al., 2005; Liu et al., 2006; Yao et al., 2006), portfolio optimization (Honda, 2003; Zhou and Yin, 2003; Sass and Haussmann, 2004; Bäuerle and Rieder, 2005), investment problems (Zhang, 2001; Guo et al., 2005), and risk measures for derivatives (Elliott et al., 2008).

Concerning estimation of MSMs, there exists by now a huge literature on this issue for discrete time models, see Cappé et al. (2005) for a recent overview. Rydén et al. (1998), for instance, propose to maximize the likelihood using optimization methods. Another maximum likelihood (ML) approach is the expectation maximization (EM, Dempster et al., 1977) algorithm (used e.g. in Engel and Hamilton, 1990; Elliott et al., 1997). Robert et al. (1993) implement a Bayesian framework relying on Markov chain Monte Carlo (MCMC) methods. Recent work in that direction includes Rosales et al. (2001); Scott (2002).

In contrast to that, relatively few papers are dealing with the estimation for continuous time MSM, despite the increasing interest in this model. EM algorithms are described in James et al. (1996) and Sass and Haussmann (2004). However, this approach requires constant and known volatility σ: even for constant but unknown σ it is 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). Furthermore, for a general continuous time MSM given discrete observations, no finite dimensional filters are known and hence the conditional expectations used in the EM algorithm cannot be computed. Roberts and Ephraim (2008) develop an EM algorithm for a slightly different model, where the state process Y is modeled in continuous time, however, using observation intervals of length Δt, drift and volatility are allowed to depend only on the Δt-skeleton of Y, i.e. just as in the discrete time model, μ and σ can jump only at times $t = m\,\Delta t$ for $m = 0, 1, \ldots, N$. Elliott et al. (2008) propose a moment based regression method, 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). MCMC approaches for the same model as considered in Roberts and Ephraim (2008) for univariate observations are pursued in Liechty and Roberts (2001) for constant σ and more general in Ball et al. (1999).

Hence, we see a gap in the literature: neither there are efficient methods available that are exactly tailored to the multivariate continuous time MSM (1) nor there are investigations to which extent discrete time approximations are reliable and how problems arising with them can be overcome. In the present paper, we try to fill exactly this gap. We consider joint modeling of several stock prices using multivariate continuous time MSMs and take a Bayesian approach to estimate parameters using MCMC methods. Extending the work of Ball et al. (1999) and Liechty and Roberts (2001), we construct a sampler tailored to a multivariate continuous time MSM. Second, we adapt a discrete time sampler as described e.g. in Frühwirth-Schnatter (2006) to serve as an approximation for the continuous time model. We look at the approximation error and discuss extensively the problem of obtaining the (continuous time) rate matrix corresponding to a (discrete time) transition matrix. Such a rate matrix might not exist. We show that the MCMC approach allows for a nice solution of this embedding problem. We compare the proposed discrete and continuous time methods on simulated data. We are particularly interested in estimating parameters in multidimensional MSMs with high rates and considerable noise, based on not too many observations (less than, say, 5 000), as this is the typical situation one faces for many financial time series.

The rest of the paper is organized as follows. 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 define prior 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 practical application of the methods presented in Sections 3 and 4. We discuss the selection of the number of states and the problem of label switching. Furthermore, we show numerical results for simulated data using different estimation methods. Finally, in Section 6, we consider market data from daily stock index quotes from Argentina, Brazil, Mexico, and the US. The estimated complexity (number of states) of this multivariate setting is in line with the findings in similar settings known from the literature where only data from one country is analyzed. The detailed estimates also have a straightforward economic interpretation.

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 $(\Omega, \mathcal{F} = (\mathcal{F}_t)_{t\in[0,T]}, \mathrm{P})$, over time $[0, T]$ we assume the n-dimensional return process $R = (R_t)_{t\in[0,T]}$ to evolve like

$$R_t = \int_0^t \mu_s \,\mathrm{d}s + \int_0^t \sigma_s \,\mathrm{d}W_s,$$

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 notations $C^{(k)} = \sigma^{(k)}\,(\sigma^{(k)})^\top$ for the covariance matrices, $\tau_i^{(k)} = \big(C_{ii}^{(k)}\big)^{1/2}$ for the volatilities of a single asset, $\tau^{(k)} = (\tau_1^{(k)}, \dots, \tau_n^{(k)})^\top$ for the vector of volatilities, and $\rho_{ij}^{(k)} = C_{ij}^{(k)}/\big(\tau_i^{(k)}\tau_j^{(k)}\big)$ for the correlation between two assets. Sometimes we will interpret B as a matrix with elements $B_{ik} = \mu_i^{(k)}$.

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

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

The state process Y is characterized by the rate matrix $Q \in \mathbb{R}^{d\times d}$ as follows: Setting $\lambda_k = -Q_{kk} = \sum_{l=1, l\neq k}^{d} Q_{kl}$, in state k the waiting time for the next jump is λ_k-exponentially distributed and $\mathrm{P}(Y_t = l \,|\, Y_{t^-} = k, Y_t \neq Y_{t^-})$, 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 (e.g. daily returns)

$$V_m = \Delta R_{m\,\Delta t} = \int_{(m-1)\Delta t}^{m\,\Delta t} \mu_s\,\mathrm{d}s + \int_{(m-1)\,\Delta t}^{m\,\Delta t} \sigma_s\,\mathrm{d}W_s, \quad m = 1, \dots, N, \tag{2}$$

which is equivalent to observing the (log) price or return process at discrete times.

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 with positive diagonal, i.e. $\sigma^{(k)}(\sigma^{(k)})^\top$ equals the Cholesky factorization of the covariance matrix.

Remark 2. An important special case occurs when σ is constant and only μ switches. Note that with switching σ, in principle the state process Y can be observed via the quadratic variation of R, as $\mathrm{d}[R]_t = \sigma_t\sigma_t^\top\,\mathrm{d}t$. This is not possible with constant σ, where Y is hidden even if continuous observations are available.

The algorithms presented in the following sections assume that both μ and σ are switching. However, algorithms tailored to switching exclusively either in μ or in σ are obtained with straightforward adaptations by updating only $\sigma^{(1)}$ (or $\mu^{(1)}$) and replacing $\sigma^{(k)}$ ($\mu^{(k)}$), $k = 2, \dots, d$, with copies of $\sigma^{(1)}$ ($\mu^{(1)}$).

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 return data, $V = (V_m)_{m=1,\dots,N}$, 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 4 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. $J_0 = 0$, $Z_0 = Y_0$, 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. One prior is based on assuming prior independence among all parameters, i.e.

$$\mathrm{P}(Q, B, \Sigma, Y_0) = \mathrm{P}(Q)\ \mathrm{P}(B)\ \mathrm{P}(\Sigma)\ \mathrm{P}(Y_0), \tag{3}$$

where $\mathrm{P}(Y_0) = 1/d$. However, 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 assume that the state process starts from its ergodic probability ω, making Q and Y_0 dependent apriori:

$$\mathrm{P}(Q, B, \Sigma, Y_0) = \mathrm{P}(Q)\ \mathrm{P}(B)\ \mathrm{P}(\Sigma)\ \mathrm{P}(Y_0 \,|\, Q), \tag{4}$$

where $\mathrm{P}(Y_0 \,|\, Q) = \omega$. Under the second prior, Y given Q is a stationary Markov chain, i.e. $\mathrm{P}(Y_t \,|\, Q) = \omega$ for all $t \in [0, T]$.

Concerning the remaining parameters, we assume that the off-diagonal elements of Q as well as the elements of B are apriori mutually independent as are the volatility matrices. Furthermore, for $i = 1, \dots, n$, $k, l = 1, \dots, d$, $l \neq k$, we assume:

$$Q_{kl} \sim \Gamma(f_{kl}, g_{kl}), \qquad \mu_i^{(k)} \sim \mathrm{N}(m_{ik}, s_{ik}^2), \qquad C^{(k)} \sim \mathrm{IW}(\Xi^{(k)}, \nu_k). \tag{5}$$

With Γ, N, and IW we refer to the Gamma, normal, and inverted Wishart distribution, respectively. We use the notation $m_{\cdot k} = (m_{1k}, \dots, m_{nk})^\top$ and $s_{\cdot k}^2 = (s_{1k}^2, \dots, s_{nk}^2)^\top$ to denote the vectors of prior means and prior variances of $\mu^{(k)}$.

Remark 3. For the inverted Wishart distribution $C \sim \mathrm{IW}(\Xi, \nu)$, we use the parameterization where the density is given through $f_{\mathrm{IW}}(C; \Xi, \nu) \propto (\det C)^{-\nu-(n+1)/2} \exp\big(-\mathrm{tr}(\Xi C^{-1})\big)$ and the expected value is given as $\mathrm{E}[C] = \Xi\,(\nu - (n+1)/2)^{-1}$.

3.3 Complete-Data Likelihood Function

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

$$\mathrm{P}(V \mid Q, B, \Sigma, Y) = \mathrm{P}(V \mid B, \Sigma, Y) = \prod_{m=1}^{N} \varphi(V_m, \bar{\mu}_m, \bar{C}_m), \tag{6}$$

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

$$\bar{\mu}_m = \int_{(m-1)\Delta t}^{m\,\Delta t} \mu^{(Y_s)}\,\mathrm{d}s, \qquad \bar{C}_m = \int_{(m-1)\Delta t}^{m\,\Delta t} C^{(Y_s)}\,\mathrm{d}s.$$

Remark 4. 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 complete-data likelihood. In Equation (6), $\bar{\mu}_m$ and $\bar{C}_m$ have to be replaced by $\tilde{\mu}_m = \int_{t_{m-1}}^{t_m} \mu^{(Y_s)}\,\mathrm{d}s$ and $\tilde{C}_m = \int_{t_{m-1}}^{t_m} C^{(Y_s)}\,\mathrm{d}s$, respectively.

3.4 Proposal Distributions

To sample from the joint posterior distribution of Q, B, Σ, and Y given the observed data V, we partition the unknowns into the blocks Q, $\mu^{(k)}$, $C^{(k)}$, Y, and draw each block from the appropriate conditional distribution.

3.4.1 Drifts

For the update of $\mu^{(k)}$ for each state k, a Gibbs step can be performed as follows. First, we introduce the notation $B^{-k} = \left(\mu^{(1)}, \ldots, \mu^{(k-1)}, \mu^{(k+1)}, \ldots, \mu^{(d)}\right)$ and

$$o_m^k = \int_{(m-1)\Delta t}^{m\,\Delta t} \mathbb{I}_{\{Y_s = k\}}\,\mathrm{d}s, \qquad \bar{\mu}_m^{-k} = \sum_{l=1, l\neq k}^{d} \mu^{(l)}\, o_m^l.$$

Then we have

$$\mathrm{P}\left(\mu^{(k)} \mid V, B^{-k}, \Sigma, Y\right) \propto \varphi\big(\mu^{(k)};\, m_{\cdot k}, \mathrm{Diag}(s_{\cdot k}^2)\big) \prod_{m=1}^{N} \varphi\big(V_m - \bar{\mu}_m^{-k};\, \mu^{(k)} o_m^k, \bar{C}_m\big)\,,$$

and hence $\mu^{(k)} \mid V, B^{-k}, \Sigma, Y \sim \mathrm{N}\left(a^{(k)}, S^{(k)}\right)$, where

$$S^{(k)} = \left(\mathrm{Diag}(s_{\cdot k}^2)^{-1} + \sum_{m=1}^{N} \bar{C}_m^{-1} (o_m^k)^2\right)^{-1},$$

$$a^{(k)} = S^{(k)} \left(\mathrm{Diag}(s_{\cdot k}^2)^{-1}\, m_{\cdot k} + \sum_{m=1}^{N} \bar{C}_m^{-1} (V_m - \bar{\mu}_m^{-k}) o_m^k\right).$$

3.4.2 Volatilities

The Metropolis-Hastings update for the single covariance matrices $C^{(k)}$ is inspired by the discrete time Gibbs update in Section 4.1.1. The proposals are based on the observations containing no jumps: We draw $C^{(k)'} \sim \mathrm{IW}\left(\Xi^{(k)} + \Xi^{(k)'}, \nu_k + \nu'_k\right)$, where

$$\begin{aligned}
\Xi^{(k)'} &= \frac{1}{2\,\Delta t} \sum_{m=1}^{N} \mathbb{I}_{\{o^k_m = \Delta t\}} (V_m - \mu^{(k)}\,\Delta t)(V_m - \mu^{(k)}\,\Delta t)^\top, \\
\nu'_k &= \frac{1}{2} \sum_{m=1}^{N} \mathbb{I}_{\{o^k_m = \Delta t\}}.
\end{aligned}$$

Then the acceptance depends only on the observations with occupation time in state k greater than zero but less than Δt: Defining $\bar{V}^k = (V_m)_{\{1 \le m \le N \,|\, 0 < o^k_m < \Delta t\}}$, we have

$$\alpha_{C^{(k)}} = \min\left\{1, \frac{\mathrm{P}(\bar{V}^k \,|\, B, \Sigma', Y)}{\mathrm{P}(\bar{V}^k \,|\, B, \Sigma, Y)}\right\}.$$

Remark 5. For the special case where the volatility is constant, a Gibbs step is available by drawing $C' \sim \mathrm{IW}\left(\Xi + \Xi', \nu + \nu'\right)$, where $\Xi' = \frac{1}{2\,\Delta t} \sum_{m=1}^{N} (V_m - \bar{\mu}_m)(V_m - \bar{\mu}_m)^\top$ and $\nu' = N/2$.

3.4.3 State Process

We first consider the full conditional probability distribution $\mathrm{P}(Y \,|\, V, Q, B, \Sigma)$. The prior distribution of the state process Y_t for $t > 0$ is determined by the distribution of Y_0 and the rate matrix Q, and is independent of B and Σ. Therefore we obtain:

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

The probability of Y given Q equals

$$\begin{aligned}
\mathrm{P}(Y \,|\, Q) &= \mathrm{P}(Y_0 \,|\, Q) \prod_{h=1}^{H} \left(\lambda_{Z_{h-1}} e^{-\lambda_{Z_{h-1}}\,\Delta J_h} \frac{Q_{Z_{h-1},Z_h}}{\lambda_{Z_{h-1}}}\right) e^{-\lambda_{Z_H}(T - J_H)} && (7) \\
&= \mathrm{P}(Y_0 \,|\, Q) \prod_{k=1}^{d} \prod_{\substack{l=1 \\ l \neq k}}^{d} \left(e^{-Q_{kl}\, O^k_T}\, Q_{kl}^{N_{kl}}\right), && (8)
\end{aligned}$$

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

$$O^k_T = \int_0^T \mathbb{I}_{\{Y_t = k\}}\,\mathrm{d}t, \qquad N_{kl} = \sum_{h=1}^{H} \mathbb{I}_{\{Z_{h-1} = k, Z_h = l\}}.$$

For the update of Y, we draw from the conditional distribution given Q, which simplifies the acceptance probability. To obtain good rates of acceptance, we do not update the whole process at one time but break it into a number of blocks of approximately exponentially distributed length, which are updated independent of each other. For a discussion of a suitable choice of the corresponding tuning parameters see Section 5.3.

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

$$
\begin{aligned}
q(\overline{Y}, Y') = \prod_{h=1}^{H'-1} \left(e^{-\lambda_{Z'_{h-1}} \Delta J'_h} Q_{Z'_{h-1}, Z'_h} \right) e^{-\lambda_{Z'_{H'-1}} \Delta J'_{H'}} \\
\left(\sum_{j \neq Z'_{H'-1}} Q_{Z'_{H'-1}, j}\, e^{-\lambda_j (t' - J'_{H'})} + Q_{Z'_{H'-1}, Z'_{H'}} \sum_{j \neq Z'_{H'}} Q_{Z'_{H'}, j} f(\lambda_{Z'_{H'}}, \lambda_j, t' - J'_{H'}) \right),
\end{aligned}
\tag{9}
$$

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

$$
q(\overline{Y}, Y') = e^{-\lambda_{Z'_0} t'} + \sum_{j \neq Z'_0} Q_{Z'_0, j}\, f(\lambda_{Z'_0}, \lambda_j, t'). \tag{10}
$$

For updating $(Y_t)_{t\in[0,t_1[}$, Y'_0 is sampled from the initial distribution of the state process and in (9) and (10) 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 (9) the second line is replaced with $Q_{Z'_{H'-1}, Z'_{H'}}\, e^{-\lambda_{Z'_{H'}} (t' - J'_{H'})}$ and in (10) 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

$$
\mathrm{P}(Y' \,|\, \overline{V}, B, \Sigma, Q, \underline{Y}) = \mathrm{P}(\overline{V} \,|\, B, \Sigma, Y') \prod_{h=1}^{H'} \left(e^{-\lambda_{Z'_{h-1}} \Delta J'_h} Q_{Z'_{h-1}, Z'_h} \right) e^{-\lambda_{Z'_{H'}} (t' - J'_{H'})}. \tag{11}
$$

Combining (6), (11), and (9) (or (10)), 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 (7) and (9), 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

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

3.4.4 Rate Matrix

Using (8) and the fact that all elements Q_{kl}, where $l \neq k$, are apriori independent, we obtain:

$$\mathrm{P}(Q \,|\, V, B, \Sigma, Y) \propto \mathrm{P}(Y_0 \,|\, Q) \prod_{k=1}^{d} \prod_{\substack{l=1 \\ l \neq k}}^{d} \psi_{kl}(Q_{kl}) \tag{12}$$

(cf. Ball et al., 1999), where ψ_{kl} is a Gamma distribution with parameters $f_{kl} + N_{kl}$ and $g_{kl} + O_T^k$. To update the rate matrix, we propose a rate matrix Q' with elements

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

where $k \neq l$. If the initial distribution of the state process $\mathrm{P}(Y_0)$ is independent of Q, then we can drop the term $\mathrm{P}(Y_0 \,|\, Q)$ in (12) and Q' is already a draw from the appropriate full conditional distribution, which is accepted with probability 1. However, if Y_0 starts from the ergodic distribution, we accept Q' 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.

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

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 returns observed at fixed observation times Δt, $2\Delta t, \ldots, N\,\Delta t = T$ assuming that the state process jumps only at the end of these observation times. 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 to the approximation error and discuss the problem of how to compute the rate matrix corresponding to some transition matrix.

4.1 The Discrete Time Approximation

4.1.1 Discrete Time Markov Mixture of Multivariate Normal Distributions

We assume that the state process jumps only at the end of the observation times $m\,\Delta t$, $m = 0, \ldots, N-1$. That means that drift and volatility of the return process R_t are constant over each observation time interval $[(m-1)\,\Delta t, m\,\Delta t[$, taking value μ_{m-1} and σ_{m-1}, respectively. In comparison to representation (2), V_m simplifies to

$$V_m = \mu_{m-1}\,\Delta t + \sigma_{m-1}\,(W_{m\,\Delta t} - W_{(m-1)\,\Delta t}). \tag{13}$$

In the same way, the step process Y is fully described by its values at the times $m\,\Delta t$ and the unknown state process reduces to $Y = (Y_m)_{m=0,\ldots,N-1}$.

Since we 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}, i.e. the average duration of state k is equal to $1/(1 - X_{kk})$.

The resulting model is a discrete time Markov mixture of multivariate normal distributions, because $V_m \,|\, Y_{m-1} = k \sim \mathrm{N}(\mu^{(k)}\,\Delta t, C^{(k)}\,\Delta t)$, and Y_m is a hidden Markov chain with transition matrix X. For such a model, MCMC estimation is implemented easily, see Subsection 4.3, however, we have to take care of the so called embedding problem discussed in the next subsection.

4.1.2 Finding Rate Matrices for Transition Matrices

When we use a discrete time MSM as an approximation to the continuous time MSM, we face the so-called embedding problem. This means we have to compute the (continuous time) rate matrix Q corresponding to some (discrete-time) transition matrix X for a fixed time step Δt. The problem of the existence of an adequate rate matrix was already addressed by Elfving (1937) and in more detail by Kingman (1962). 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) rate matrix. Bladt and Sørensen (2005) describe how to find ML estimators for Q for observable Markov chains 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 matrix Q such that $X = \exp(Q\,\Delta t)$ and Q is a valid rate matrix.

If the transition matrix is strictly diagonally dominant (Israel et al., 2001, Remark 2.2), or all eigenvalues of X are positive and distinct (cf. Culver, 1966, Theorem 2), then $\log X$, the matrix logarithm of X, exists uniquely; however, it need not yield a valid rate matrix, but complex or negative off-diagonal elements may occur. Israel et al. (2001, Theorem 3.1) state various sufficient conditions, when a valid rate matrix does not exist. In such cases, Kreinin and Sidelnikova (2001) propose to regularize $\log X$ by projecting onto the space of valid rate matrices. It is also possible to set all negative off-diagonal entries, which are very small usually, to zero and to adjust all other elements proportional to their magnitude.

Another problem is uniqueness of Q, see Israel et al. (2001, Theorem 5.1) for conditions guaranteeing that X allows for a unique rate matrix. If there are multiple valid rate matrices, Israel et al. (2001) observe that choosing different rate matrices results in different values $X_t(Q) = \exp(Q\,t)$ for most t and suggest to choose the rate matrix that represents the transition matrix best in some sense.

Within the Bayesian approach we pursue in this paper, alternative methods are available to handle this problem. Since we are using sampling based methods, it is easy to restrict the parameter space for X to all uniquely embeddable transition matrices, simply by rejecting and redrawing updates for X that do not have a unique corresponding rate matrix, see Subsection 4.3. Furthermore, as outlined in Subsection 4.2, it is possible to impose regularity through choosing an informative Dirichlet prior on X.

4.2 Prior Distributions

Prior distributions have to be chosen for X, B, Σ, and Y_0. We choose the same priors for B and Σ as in Subsection 3.2. Again, we consider two prior assumptions concerning the initial state Y_0, one where X and Y_0 are a priori independent and one where the state process starts from the ergodic probability ω corresponding to X, i.e. $\mathrm{P}(Y_0 \,|\, X) = \omega$.

In contrast to Subsection 3.2, we choose a prior for the transition matrix X rather than for the rate matrix Q. As usual for Markov mixture models, the d rows $X_{k\cdot}$ of X, where $k = 1, \dots, d$, are assumed to be independent, each following a Dirichlet distribution:

$$X_{k\cdot} \sim \mathrm{D}(g_{k1}, \dots, g_{kd}). \tag{14}$$

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 $g_{k\cdot} = X^0_{k\cdot}\, c_k$ and 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).

We observed an interesting relationship between choosing the hyperparameter c_k and the embedding problem discussed in Subsection 4.1.2. We found that the larger the values $c_1, \dots, c_d$ the larger is the a priori fraction of matrices X that have a unique and valid corresponding rate matrix. For illustration, consider a model with $d = 4$ states. Assume that $c_1 = \ldots = c_d = c$ and that the diagonal elements X^0_{kk} are equal to 0.6, while the off-diagonal elements X^0_{kl}, $l \neq k$ are equal to 0.4/3. Table 1 shows for various values of c how the a priori fraction of regular matrices increases with c. Thus one way to handle the embedding problem discussed in Subsection 4.1.2 is to choose an informative prior on X.

4.3 The Discrete Time Sampler

Starting from the prior distribution discussed in Subsection 4.2, we partition the unknowns into B, Σ, Y, and X and use a four step MCMC sampler to draw from the augmented posterior distribution $\mathrm{P}(X, B, \Sigma, Y \,|\, V)$. The complete-data likelihood function (6) of the observed data $V = (V_m)_{m=1,\dots,N}$ given X, B, Σ, and Y reduces to:

$$\mathrm{P}(V \mid X, B, \Sigma, Y) = \prod_{m=1}^{N} \varphi(V_m, \mu^{(Y_{m-1})}\,\Delta t, C^{(Y_{m-1})}\,\Delta t), \tag{15}$$

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

4.3.1 Sampling Drift, Volatility, and State Process

Sampling the drift B, the volatility Σ, and the state process X using efficient Gibbs steps is entirely standard, because we are dealing with a Markov mixture model and data augmentation as discussed e.g. in (Frühwirth-Schnatter, 2006, Chapter 11) is easily implemented. We update all elements of Σ jointly conditional on B:

$$C^{(k)} \mid B, Y, V \sim \mathrm{IW}\left(\Xi^{(k)} + \frac{1}{2\,\Delta t} \sum_{m:Y_{m-1}=k} (V_m - \mu^{(k)}\,\Delta t)(V_m - \mu^{(k)}\,\Delta t)^{\top},\ \nu_k + \frac{N_k}{2}\right),$$

where $N_k = \sum_{m=0}^{N-1} \mathbb{I}_{\{Y_m = k\}}$, and update all elements of B jointly conditional on Σ:

$$\mu^{(k)} \mid \Sigma, Y, V \sim \mathrm{N}(a_k, S_k),$$

where

$$\begin{aligned} S_k &= \left(\mathrm{Diag}(s_{\cdot k}^2)^{-1} + \Delta t\, N_k (C^{(k)})^{-1}\right)^{-1}, \\ a_k &= S_k\left(\mathrm{Diag}(s_{\cdot k}^2)^{-1} m_{\cdot k} + (C^{(k)})^{-1} \sum_{m:Y_{m-1}=k} V_m\right). \end{aligned}$$

To update Y we draw from the full conditional posterior $\mathrm{P}(Y \mid V, B, \Sigma, X)$ by forward-filtering-backward-sampling, see Frühwirth-Schnatter (2006).

4.3.2 Sampling the Transition Matrix

Sampling the transition matrix X in the present context is less standard because of the embedding problem discussed in Subsection 4.1.2. To ensure that any posterior draw X has a unique and valid rate matrix Q, we use rejection sampling. We propose transition matrices X' by sampling each row $k = 1, \ldots, d$ independently from the proposal

$$X_k' \sim \mathrm{D}(g_{k1} + N_{k1}, \ldots, g_{kd} + N_{kd}), \tag{16}$$

where $N_{kl} = \sum_{m=1}^{N-1} \mathbb{I}_{\{Y_{m-1}=k,\, Y_m = l\}}$ is equal to the number of transitions from state k to l. This step is repeated until we obtain a proposal X' that has a unique and valid rate matrix Q.

If the initial distribution of the state process $\mathrm{P}(Y_0)$ is independent of the transition matrix, then X' is a draw from the appropriate full conditional posterior density restricted to the space of all uniquely embeddable transition matrices and is accepted with probability 1.

If the state process starts from the ergodic distribution ω of X, then we add a Metropolis-Hastings step and accept X' with probability $\alpha_X = \min\{1, \bar{\alpha}_X\}$, where:

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

Usually only a few steps of rejection sampling are necessary to force regularity of X'. For illustration, we determine the fraction of regular matrices X' when the rows are sampled from (16) for a hidden Markov chain Y with following transition counts:

$$\begin{pmatrix} N_{11} & \cdots & N_{14} \\ \vdots & & \vdots \\ N_{41} & \cdots & N_{44} \end{pmatrix} = \begin{pmatrix} 816 & 65 & 4 & 12 \\ 52 & 213 & 12 & 3 \\ 13 & 1 & 536 & 22 \\ 16 & 1 & 20 & 797 \end{pmatrix}. \tag{17}$$

Additionally, the hyperparameters $g_{kj} = cX^0_{kj}$ vary as in Subsection 4.2. Table 1 shows for various values of c that the aposteriori fraction of regular matrices is larger than the prior fraction. Even in cases where the prior is rather non-informative, e.g. for $c = 4$ about 50 % of the matrices X drawn from the conditional posterior (16) are regular, meaning that on average only two rejection steps are necessary for this particular path of the hidden Markov chain Y.

4.3.3 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 bias: 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

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

If the rates are high compared to the observation time interval, then the number of expected jumps within one period gets high and $\bar{B}_{ik}$ 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 $\int_0^t \mu^i_s \,\mathrm{d}s = \sum_{k=1}^{d} B_{ik}\, O^k_t$, we have given Y_0, Q, and B

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

Whenever estimating all parameters together, there is much interaction, for instance, if the drifts are underestimated, then the volatility is overestimated.

5 Practical Considerations and Application to Simulated Data

In this section, we describe some details on the implementation. Then we present numerical results of the proposed algorithms for simulated data. 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 Label Switching and Post-Processing MCMC

Like for any (Markov) mixture model, label switching may occur during MCMC sampling either for the continuous time or the discrete time model, see e.g. Jasra et al. (2005) and (Frühwirth-Schnatter, 2006, Section 3.5). In some applications it is sufficient to constrain the parameter space appropriately; e.g. for $n = 1$ we can simply demand that $\mu^{(1)} > ... > \mu^{(d)}$ or $\sigma^{(1)} > ... > \sigma^{(d)}$. However, in general suitable constraints are not always available, and even with constraints on the parameter space, label switching can still constitute a problem (cf. Stephens, 2000).

To make sure that all labeling subspaces are explored, one can add a random permutation step as in Frühwirth-Schnatter (2001a) to the MCMC scheme introduced in Subsections 3.4 and 4.3 and perform post-processing of the MCMC output to handle label switching. The subsequent discussion applies both to the MCMC output of CMCMC as well as DMCMC.

Frühwirth-Schnatter (2001a) suggested to use a point process representation of the MCMC draws, by producing scatter plots of pairs of state specific parameters. A visual inspection of these plots allows to study the difference in the state specific parameters and to formulate an identifiability constraint. Although this works quite well in lower dimensions, it is difficult or even impossible to extend this method to higher dimensional problems.

Following Celeux (1998) we use standard k-means clustering in the point process representation of the MCMC draws to identify the MSM, however, as opposed to Celeux (1998), clustering is performed in a post-processing manner. This method is described in (Frühwirth-Schnatter, 2006, p. 96f) for finite mixture models, but applies to MSMs as well. For MSMs this method not only allows to identify the state specific parameters, but also to estimate the hidden Markov chain. In the present context, we apply k-means clustering to all posterior draws of the vector $\left(\mu^{(1)}, \ldots, \mu^{(d)}, \tau^{(1)}, \ldots, \tau^{(d)}\right)^\top$.

The whole method is based on the idea that MCMC draws belonging to the same state will cluster around the same point in the point process representation. Even if label switching occurred between two draws, the classification sequence resulting from k-means clustering indicates how to rearrange the state specific parameters. In cases where the simulation clusters are well-separated all classification sequences are a permutation of the labels $\{1, \ldots, d\}$ and show how to relabel the MCMC draws in order to obtain draws from an identified model.

5.2 Selecting the Number of States

Selecting the number of components of a MSM is quite a challenge, see Otranto and Gallo (2002) for a non-parametric approach and (Frühwirth-Schnatter, 2006, Chapters 4 and 5) for a recent review. The choice may be based on the posterior probability $\mathrm{P}(\mathcal{M}_d \,|\, V) \propto \mathrm{P}(V \,|\, \mathcal{M}_d)\,\mathrm{P}(\mathcal{M}_d)$ of a MSM with d states. However, the computation of these posterior probabilities, either using reversible jump MCMC as in Robert et al. (2000) or computing the marginal likelihoods $\mathrm{P}(V \,|\, \mathcal{M}_d)$ using bridge sampling as in Frühwirth-Schnatter (2004), turns out to be rather challenging.

Instead, we use BIC_d which is an asymptotic approximation to $-2 \log \mathrm{P}(V \,|\, \mathcal{M}_d)$:

$$\mathrm{BIC}_d = -2 \log \mathrm{P}(V \,|\, \hat{\theta}_d, \mathcal{M}_d) + p_d \log N, \tag{18}$$

where $p_d = d(d-1) + dn + dn(n+1)/2$ is the number of unknown parameters and $\hat{\theta}_d$ is the ML estimator of $\theta = (B, \Sigma, X)$ in a model with d states obtained by maximizing the log-likelihood function $\log \mathrm{P}(V \,|\, \theta, \mathcal{M}_d)$.

5.3 Notes on the Implementation

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

5.3.1 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.

The prior of the rate matrix for CMCMC is chosen in such a way that f_{kl} and g_{kl} in (5) 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 described in Subsection 4.2 for the discrete case, c_k can be interpreted as the time we observe state k a priori. Possible choices for priors on drifts and covariance matrices can be found in Subsection 5.4.

5.3.2 Running MCMC

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. In our numerical experiments, proposals containing about 3 to 5 jumps on average and 10 to 15 at maximum turned out to be a good choice resulting in an average acceptance about 25 %. The choice of these parameters can be refined by monitoring the acceptances in various test runs.

Finally, starting from initial values for drift, 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.

5.4 Numerical Results for Simulated Data

Generally, the quality of the estimates is proportional to the difference between the drifts $\mu_i^{(k)} - \mu_i^{(l)}$ and indirectly proportional to the magnitude of the variances $C_{ii}^{(k)}$ 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.

In a first example, we consider 200 samples of bivariate simulated prices ($N = 1500$, $\Delta t = 1$) with constant volatility and four different states in the drift. Aside from CMCMC and DMCMC, we run a discrete time ML method using optimization methods (referred to as DMLO) proposed in Rydén et al. (1998), and a continuous time EM algorithm (referred to as CEM), which is not available for the general MSM with switching both in drift and volatility. For CEM, the covariance matrix is pre-estimated using a linear regression for approximations of the quadratic variation process of the return for different step widths as presented in Sass and Haussmann (2004).

In the MCMC samplers 5 000 iterations are performed of which the first 500 are discarded. The following prior distributions are used: The prior for the state-specific drift is $\mu_i^{(k)} \sim \mathrm{N}(m_i, s_i^2)$ for $k = 1, \dots, d$, where $m_i = (\min_m V_m^i + \max_m V_m^i)/2$ is the midpoint and $s_i = \max_m V_m^i - \min_m V_m^i$ is the range of the i-th time series. The prior for the state-specific covariance matrix is $C^{(k)} \sim \mathrm{IW}(0.5\,\Xi, 3)$, where Ξ equals the empirical covariance matrix. The mean transition matrix has diagonal entries 0.7 and off-diagonal entries 0.1, and $c_k = c = 2.5$, cf. Subsection 5.3.1, i.e. standard deviations are 0.24 (diagonal) and 0.16 (off-diagonal); the corresponding mean rate matrix has off-diagonal entries 0.128 and the standard deviations are 0.224. For all methods, as initial values we used the prior means for C and Q (or X) and combinations of the 0.7 and 0.3 quantiles for the drifts.

Table 2 provides a comparison of CMCMC, DMCMC, CEM, and DMLO with respect to their statistical efficiency in parameter estimation. For all algorithms we show estimators of volatilities τ_i, correlation ρ_{12}, drifts $\mu_i^{(k)}$, and rates Q_{kl}, by computing the average of all estimators over the 200 replications as well as the corresponding root mean squared errors (RMSEs). On average, CMCMC gives very precise results, clearly outperforming the discrete time methods as well as CEM. For all approaches, RMSEs are slightly higher for series 2 (where volatility is higher than for series 1) and for most parameters for states 1 and 4, where state switching is faster than in states 2 and 3.

In a second example, we consider 200 samples of bivariate simulated prices ($N = 2500$, $\Delta t = 1$) with switching in both drift and volatility with three states. We run CMCMC, DMCMC, a discrete time EM algorithm similar as presented in Engel and Hamilton (1990) referred to as DEM, and DMLO; a continuous time EM algorithm is not available for this setting.

In the MCMC samplers, again 5 000 iterations (with a burn-in of 500) are performed. Priors for B and C are as in the preceeding example. The mean transition matrix has diagonal entries 0.9 and off-diagonal entries 0.05, $c_k = c = 3.33$, i.e. standard deviations are 0.14 (diagonal) and 0.1 (off-diagonal); the corresponding mean rate matrix has off-diagonal entries 0.05 and the standard deviations are 0.12. Initial values for B are 0.7, 0.5, and 0.3 quantiles, initial volatilities are set to the square-root of the empirical covariance matrix times 1.0, 0.7, and 1.1 for states 1, 2, and 3. Initial values for Q (or X) are the prior means.

Table 3 provides a comparison of DMCMC, CMCMC, DEM, and DMLO for estimators of volatilities $\tau_i^{(k)}$, correlation ρ_{12}, drifts $\mu_i^{(k)}$, and rates Q_{kl}. For this setting where state switching is less frequent than in the previous example, again CMCMC results are most precise on average, however, the discretization error is low and RMSEs are of similar order of magnitude for all methods. The most accurate results are obtained for state 2, where volatility and rates are lowest.

For initial values rather close to true values as used here, DEM and DMLO encountered no problems with convergence to local maxima or degenerate parameter values. Hence, only one run of DEM and DMLO was performed for each data set. Also the computation of a regular rate matrix from the transition matrix worked without complications. Note that the situation is quite different for the application to market data in the following Section. There, not only several runs from different randomly chosen initial values are necessary, but also the embedding-problem requires a suitable treatment. Hence, for a comparison of run times we refer to Section 6.4.

6 Application to Modeling Stock Indices

In this section, we examine daily data ranging from January 2, 1998 to December 31, 2007 from the following four indices: S&P 500 (US), IPC (Mexico), MerVal (Argentina), and Bovespa (Brazil). For estimation, we use daily returns multiplied with 100, i.e. daily movements in percent, as shown in Figure 1.

6.1 Estimating a Multivariate MSM

MCMC estimation of a joint MSM for all four indices is carried out with an increasing number d of states. For a fixed number of states, 10 000 MCMC draws were generated after a burn-in of 5 000 draws using DMCMC as described in Subsection 4.3. Alternative estimation methods are discussed in Subsection 6.4.

Estimation is based on the following prior distributions which are invariant to relabeling the states. The prior for the state-specific drift is chosen as in Subsection 5.4. The prior for the state-specific variance-covariance matrix is $C^{(k)} \sim \mathrm{IW}(\Xi, \nu)$ with $\nu = 2.5 + \frac{n-1}{2}$. Since Ξ is likely to be influential we consider a hierarchical prior where Ξ is a random parameter with a prior of its own: $\Xi \sim \mathrm{W}(G_0, g_0)$ where $g_0 = 0.5 + \frac{n-1}{2}$ and $G_0 = \frac{100 g_0}{\nu} \operatorname{Diag}(1/s_1^2, \ldots, 1/s_n^2)$. Under this prior, an additional step has to be added to the DMCMC scheme to sample $\Xi \sim \mathrm{W}(G_0 + \sum_{k=1}^{d} (C^{(k)})^{-1}, g_0 + d\nu)$. Finally, the rows $X_{k\cdot}$ of X are assumed to be independent and follow a Dirichlet distribution as in (14) with $g_{kk} = 4$ and $g_{kj} = 1/(d-1)$ for $j \neq k$. This choice is based on (Frühwirth-Schnatter, 2006, p. 335) and leads to a prior that is invariant to relabeling the states.

To select the number of states in the multivariate MSM, BIC is computed as in (18), where $\hat{\theta}_d$ is the approximate ML estimator obtained by maximizing the log likelihood function $\log \mathrm{P}(V \mid \theta_d, \mathcal{M}_d)$ over all MCMC draws. The resulting values are reported in Table 4 and suggest a model with $d = 4$ states.

The four-state MSM is identified as described in Subsection 5.1. The estimated state-specific drift vector $(\mu^{(1)}, \ldots, \mu^{(4)})$, the estimated state-specific volatility vector $(\tau^{(1)}, \ldots, \tau^{(4)})$, as well as the estimated state-specific correlations of the various indices are reported in Table 5. Furthermore, the estimated discrete time transition matrix X and the estimated continuous time rate matrix Q are reported together with the expected duration of each state in Table 6. Mean and standard deviation of the elements of Q are derived from the transformed MCMC draws $\log X$, where log is the matrix logarithm of X. The duration of state k is estimated from X using the transformed MCMC draws $1/(1 - X_{kk})$ and from Q using the transformed MCMC draws $-1/Q_{kk}$.

Figure 2 shows the smoothed state probabilities. Figure 3 shows for each index the time-varying mean and the time-varying volatility. Finally, Figure 4 shows all time-varying correlations. A detailed economic interpretation of these estimation results is given in Section 6.3.

6.2 Individual MSMs

For comparison, we also consider individual modeling of each stock index using univariate MSMs with an increasing number d of states. This means, that each index i is driven by an independent hidden Markov chain Y^i with rate matrix Q^i (or transition matrix X^i).

For each index i, DMCMC was carried out under the same priors for $\mu_i^{(k)}$ and X^i as in Subsection 6.1, however, the state-specific variance $C_{ii}^{(k)}$ follows a hierarchical prior where $C_{ii}^{(k)} \sim \mathrm{IG}(\xi_i, \nu)$ with $\nu = 2.5$ and $\xi_i \sim \Gamma(G_0, g_0)$ with $g_0 = 0.5$ and $G_0 = \frac{100 g_0}{s_i^2 \nu}$. Again, a step has to be added to the DMCMC scheme to sample $\xi_i \sim \Gamma(G_0 + \sum_{k=1}^{d} 1/C_{ii}^{(k)}, g_0 + d\nu)$. For a fixed number of states 10 000 MCMC draws were generated after a burn-in of 5 000 draws for each index.

We considered BIC to select the number of states for each univariate MSM. The BIC values computed through (18) are reported in Table 4 and suggest a model with $d = 3$ states for each of the indices. Each three-state MSM is identified as described in Subsection 5.1. The estimated state-specific drift $\mu_i^{(k)}$ and the estimated state-specific volatility $\tau_i^{(k)}$ are reported for each index i in Table 7. The aggregated BIC reported in the last line of Table 4 corresponds to the BIC of a multivariate model where each index i is modeled independently by an individual MSM with d states. Evidently, the BIC of any such model is considerably larger than the optimal BIC of the multivariate MSM, providing further evidence for our multivariate specification.

6.3 Economic Interpretation

The number of regimes (4 for the multivariate time series and 3 for each of the individual time series) seems reasonable in view of preceding studies in the literature. Guidolin and Timmermann (2006, 2007) investigate monthly data from 1954–1999 for a portfolio of large caps, a portfolio of small caps, and a portfolio of long-term bonds (all US). They report that univariate dynamics for stock and bond requires 2 or 3 states each. Joint modeling requires 4 states termed crash (negative drift, high volatility, low persistence), recovery (high drift, high volatility, low persistence), slow growth (slightly positive drift, low volatility, high persistence), bull (moderate drift, low volatility, high persistence). Correlation between large and small caps is highest in the crash state (0.82) and lowest in the recovery state (0.5). And Rydén et al. (1998) investigate (univariate) daily data from the S&P 500 index from 1928 to 1991 (outlier reduced $\pm 4\hat{\sigma}$). Splitting the data into 10 subseries and fitting models with switching volatility but constant mean, they found 7 of the subseries to follow 3 regimes and 3 subseries to follow 2 regimes.

As in Guidolin and Timmermann (2006, 2007), also for our data from four countries it turns out that 4 states are sufficient to describe the multivariate model. The estimates in Section 6.1 (Table 5) show that state 2 is the only one with a positive drift for all countries, all other drift components are not as significant but they have a certain tendency (state 1: down, state 3: S&P 500 and IPC performing better, state 4: S&P 500 performing worst). Generally, volatility is lower for S&P 500 than for the other indices, while volatility is higher for MerVal and Bovespa than for IPC. Evidently, state 2 is the low volatility state, while state 3 is the high volatility state. Both states 1 and 4 are states of medium volatility with the exception of MerVal where state 4 is also a high volatility state in combination with relatively lower correlation with all other indices. Thus our interpretation of the states is a little different. We might look at state 2 as a slow growth or even a bull state while state 1 is an indifferent or slightly bearish state. These correspond to the characterization above. States 3 and 4 may be seen as mixed states, the first corresponding to the financial crisis in Argentina starting 1997/98 (cf. Figure 2) which also affected Brazil, and where MerVal and Bovespa had a poorer performance than S&P 500 and IPC. In fact, Table 5 shows that in this state the correlation between MerVal and Bovespa is extremely high (0.72). State 4 corresponds to S&P 500 performing on average worst of all indices. In this state, the correlation of S&P 500 with the other indices has its minimum and MerVal is very lowly correlated with all other indices. Looking at Figure 2, we see that this state was dominant in the years after 9/11 where S&P 500 had a poor performance, while MerVal was recovering from the crisis.

While other interesting effects might be observed, essentially the interpretation shows that these four states are reasonable for the years 1998–2007. Looking at longer time periods, one might expect that more states are needed if there are extreme events which let the indices move in different directions. E.g. we have no state which would explain a year long down movement of IPC while S&P 500 increases. In fact we see that the correlation between S&P 500 and IPC is more or less the same over all states, while it changes considerably for all other indices, see Table 5 and Figure 4. Using more infrequent data for a longer horizon, in particular when using monthly data, we might get to the four regimes proposed in Guidolin and Timmermann (2006, 2007) as we see them in states 1 and 2 already. The analysis of the single indices in Section 6.2 supports these observations.

6.4 Comparison of DMCMC to Other Estimation Methods

Implementation of all estimation methods was carried out using Matlab (Version 7.3.0) on a notebook with a 2.0 GHz processor. Using DMCMC, both for the univariate as well as for the multivariate models we used 10 000 MCMC draws after a burn-in of 5 000. For comparison, we implemented CMCMC to produce 20 000 MCMC draws after a burn-in of 5 000 draws using priors similar as for DMCMC. Finally, we also employed two discrete time ML approaches as described in Section 5.4. We ran DEM, for which a resulting non-embeddable transition matrix can be handled only by regularizing the corresponding (invalid) rate matrix, as well as DMLO, which we adapted by setting the log-likelihood to $-\infty$ for non-embeddable transition matrices. We repeatedly started DEM and DMLO from randomly chosen initial values (see Table 8), as these methods often get stuck at local maxima of the likelihood function or at degenerate parameter values. The stopping criterion for the single runs was a maximal change in the parameter values of 1 %. Then, we selected the parameters giving the maximum log-likelihood.

Computing times are summarizes for the various methods in Table 8, where we averaged over the different indices for univariate models. When comparing the MCMC approaches, we find that DMCMC is not only much faster than CMCMC, but also more efficient, see Table 9 which compares inefficiency factors computed as in Geweke (1992) for selected parameters. This is not surprising, because DMCMC is a full Gibbs sampler while CMCMC is, at least partly, a Metropolis-Hastings sampler. Additionally, the construction of the updates is more involved in the continuous time setting. For instance, to update the drifts, both algorithms use a Gibbs step, however, the number of matrix inversions required for DMCMC equals d, while it is much larger for CMCMC, depending on the speed of state switching.

When we turn to the discrete time ML approaches, we find that DEM is faster than DMLO and CMCMC. DMLO works well for moderate numbers of parameters, but slows down for more states and more assets, while runtimes for CMCMC are growing only very slowly for increasing numbers of states. DEM and DMCMC show similar computation times, DEM being a bit faster for lower numbers of parameters and DMCMC being faster for more complex models. Concerning the runtimes per single runs for DEM, one should note that these are not necessarily increasing with the numbers of parameters, since for more complex models, DEM tends to abort early at irregular parameter values more often. This comparison should be seen with caution, though. The number of iterations from different starting points for DEM and DMLO is somewhat arbitrary; we tried to be rather sparing (Rydén et al., 1998, used 200 iterations for 2 states and 500 for 3 states for univariate data). Moreover, more sophisticated methods for the optimization could improve the speed of DMLO.

As the speed of state switching is low (except for state 3 with short duration and high volatility), we expect the discretization error for the discrete time methods to be negligible. Indeed, when comparing the estimates of DEM, DMLO, and DMCMC with CMCMC, we found that both for joint and individual data, results were very close to each other except for the parameters corresponding to the highly volatile states, but also for these there was good accordance, see Table 10 which reports results of CMCMC, DEM, and DMLO for selected parameters.

Thus if one is interested in point estimates only, none of the methods can be viewed as superior regarding the quality of the estimates. However, the Bayesian approach provides a lot of additional econometric inference without much additional effort which we feel is a definite advantage. Examples are easy computation of confidence regions and standard errors for all parameters of interest like state-specific drifts, volatilities, or correlation matrices, see Table 5 to Table 6; regarding the relatively high parameter uncertainty, this information is of great importance. Also time-varying moments like means, volatilities or correlations are immediately available, see Figure 3 to Figure 4.

7 Conclusion

We present a continuous time MCMC algorithm for parameter estimation in multidimensional continuous time MSMs as well as a discrete time sampler refined to serve as an approximation method. We apply the methods to simulated as well as historical stock data and compare the results with those of more classical discrete time approaches.

The EM algorithm and the ML approach using optimization methods generally require a lot of runs from different starting values which makes them time consuming especially for higher numbers of parameters. This dependence on initial values does not hold for the MCMC samplers, which moreover give a lot of important additional information like confidence intervals without further effort. The continuous time MCMC algorithm is attractive as it introduces no discretization error and is easily extended to non-equidistant observation times; however, the computational effort is rather big. On the other hand, the discrete time MCMC method adapted to the continuous time setting gives good results within reasonable time. The problem of finding a continuous time rate matrix corresponding to some discrete time transition matrix, arising with all discrete time methods, can be handled most flexibly and elegantly within the MCMC approach.

Applying the algorithms to financial time series, stable results are obtained only for data with moderate volatility like stock indices. For some data sets it may be sufficient to introduce a "jump" state with huge volatility and short duration as observed in Kaufmann und Frühwirth (2002) for a switching ARCH model. However, in general achieving good fits for highly volatile data definitely requires a more refined model for the volatility. For a single asset, Hahn et al. (2007) look at a model combining switching drift with the non-constant volatility model of Hobson and Rogers (1998). A possibility to extend this model to multiple dimension might be to employ Markov switching in the correlation matrix as proposed by Pelletier (2006), and using Hobson and Rogers (1998) for the marginal volatilities. Compared to the MSM presented here, such a model increases the number of parameters only slightly (essentially dimension times two) and should be tractable for estimation and applications, while offering a substantially better fit. Detailed investigations into that direction must however remain as future work.

Acknowledgment

The first author acknowledges financial support by the Austrian Science Fund (FWF-projects S8305 and P17947-N12). The third author acknowledges support by the Heisenberg Programme of the German Research Foundation DFG. We thank the referees, the Associate Editor, and the Editor for their comments and suggestions which led to a considerable improvement of the paper.

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.

R. Bansal and H. Zhou. Term structure of interest rates with regime shifts. *The Journal of Finance*, 57:1997–2043, 2002.

N. Bäuerle and U. Rieder. Portfolio-optimization with unobservable Markov-modulated drift process. *Journal of Applied Probability*, 42:362–378, 2005.

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.

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

J. Buffington and R.J. Elliott. Regime switching and european options. In B. Pasik-Duncan, editor, *Stochastic Theory and Control*, pages 73–82. Springer, Berlin Heidelberg, 2002b.

O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models.* Springer, New York, 2005.

G. Celeux. Bayesian inference for mixtures: The label-switching problem. In R. Payne and P. J. Green, editors, *COMPSTAT 98*, 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.

L. L. Chan, R. J. Elliott, and T. K. Siu. Option pricing and Esscher transform under regime switching. *Annals of Finance*, 1:423–432, 2005.

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.* 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, W. C. Hunter, and B. M. Jamieson. Financial signal processing – a self calibrating model. *International Journal of Theoretical and Applied Finance*, 4:567–584, 2001.

R. J. Elliott, V. Krishnamurthy, and J. Sass. Moment based regression algorithm for drift and volatility estimation in continuous time Markov switching models. *Econometrics Journal*, 11:244–270, 2008.

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. Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *The Econometrics Journal*, 7:143–167, 2004.

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

R. Garcia and P. Perron. An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics*, 78:111–125, 1996.

J. Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting*, pages 169–188. Oxford University Press, Oxford, 1992.

M. Guidolin and A. Timmermann. An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. *Journal of Applied Econometrics*, 21(1):1–22, 2006.

M. Guidolin and A. Timmermann. Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control*, 31:3503–3544, 2007.

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

X. Guo. Information and option pricings. *Quantitative Finance*, 1:38–44, 2001b.

X. Guo, J. Miao, and E. Morellec. Irreversible investment with regime shifts. *Journal of Economic Theory*, 122:37–59, 2005.

M. Hahn, W. Putschögl, and J. Sass. Portfolio optimization with non-constant volatility and partial information. *Brazilian Journal of Probability and Statistics*, 21:27–61, 2007.

D. G. Hobson and L. C. G. Rogers. Complete models with stochastic volatility. *Mathematical Finance*, 8(1):27–48, 1998.

T. Honda. Optimal portfolio choice for unobservable and regime-switching mean returns. *Journal of Economic Dynamics and Control*, 28:45–78, 2003.

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.

S. Kaufmann and S. Frühwirth. Bayesian analysis of switching ARCH models. *Journal of Time Series Analysis*, 23:426–458, 2002.

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.

R. H. Liu, Q. Zhang, and G. Yin. Option pricing in a regime-switching model using the fast Fourier transform. *Journal of Applied Mathematics and Stochastic Analysis*, 2006:1–22, 2006.

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.

D. Pelletier. Regime switching for dynamic correlations. *Journal of Econometrics*, 131:445–473, 2006.

C. P. Robert, G. Celeux, and J. Diebolt. Bayesian estimation of hidden Markov chains: A stochastic implementation. *Statistics & Probability Letters*, 16(1): 77–83, 1993.

C. P. Robert, T. Rydén, and D. M. Titterington. Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *Journal of the Royal Statistical Society Series B*, 62:57–75, 2000.

W. J. J. Roberts and Y. Ephraim. An EM algorithm for ion-channel current estimation. *IEEE Transactions on Signal Processing*, 56:26–33, 2008.

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.

T. Rydén, T. Teräsvirta, and S. Åsbrink. Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics*, 13:217–244, 1998.

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.

H. Schaller and S. Van Norden. Regime switching in stock market returns. *Applied Financial Economics*, 7(2):177–191, 1997.

S. L. Scott. Bayesian methods for hidden Markov models: Recursive computing in the 21st century. *Journal of the American Statistical Association*, 97:337–351, 2002.

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

A. Timmermann. Moments of Markov switching models. *Journal of Econometrics*, 96:75–111, 2000.

D. Yao, Q. Zhang, and X. Y. Zhou. *Stochastic Processes, Optimization, and Control Theory: Applications in Financial Engineering, Queueing Networks, and Manufacturing Systems*, chapter A Regime-Switching Model for European Options. Springer, New York, 2006.

Q. Zhang. Stock trading: An optimal selling rule. *SIAM Journal on Control and Optimization*, 40:64–87, 2001.

X. Y. Zhou and G. Yin. Markowitz's mean-variance portfolio selection with regime switching: A continuous-time model. *SIAM Journal on Control and Optimization*, 42:1466–1482, 2003.

Tables and Figures

c	fraction prior (in %)	fraction posterior (in %)
4	2.1	50.1
5	3.7	55.5
10	17.3	76.9
20	52.8	94.5
50	95.1	99.9
75	99.3	100.0

Table 1: Fraction of regular transition matrices X with a unique and valid rate matrix Q when sampling the rows of X from the Dirichlet prior $\mathrm{D}(X^0_{k1}\,c, \ldots, X^0_{k4}\,c)$, with $X^0_{kk} = 0.6$ and $X^0_{kl} = 0.4/3$ for $l \neq k$ for various values of c. The last column displays this fraction for the corresponding Dirichlet posterior $\mathrm{D}(X^0_{k1}\,c + N_{k1}, \ldots, X^0_{k4}\,c + N_{k4})$, where the counts N_{kl} are given by (17). All fractions were determined by Monte Carlo simulations.

	τ_1 $\times 10^3$	τ_2 $\times 10^3$	ρ_{12}	μ_1 $\times 10^3$	μ_2 $\times 10^3$	Q			
true values	3.00	4.00	-0.100	6.00	6.00	-0.377	0.198	0.121	0.057
				6.00	-6.00	0.134	-0.313	0.121	0.057
				-6.00	6.00	0.057	0.121	-0.313	0.134
				-6.00	-6.00	0.057	0.121	0.198	-0.377
CMCMC									
average	3.02	4.03	-0.102	5.98	5.97	-0.370	0.190	0.117	0.063
				5.98	-5.96	0.130	-0.306	0.120	0.057
				-5.97	5.94	0.060	0.117	-0.308	0.131
				-5.97	-5.94	0.062	0.118	0.193	-0.373
RMSE	0.07	0.10	0.033	0.22	0.34	0.056	0.048	0.036	0.028
				0.18	0.25	0.029	0.039	0.026	0.019
				0.18	0.27	0.019	0.025	0.033	0.027
				0.24	0.35	0.031	0.036	0.042	0.052
DMCMC									
average	3.16	4.12	-0.126	5.63	5.46	-0.334	0.173	0.124	0.038
				5.55	-5.41	0.117	-0.270	0.094	0.059
				-5.56	5.43	0.057	0.098	-0.271	0.117
				-5.63	-5.45	0.039	0.123	0.173	-0.335
RMSE	0.17	0.15	0.041	0.42	0.62	0.057	0.043	0.032	0.030
				0.48	0.63	0.029	0.051	0.035	0.018
				0.47	0.62	0.019	0.032	0.051	0.029
				0.44	0.64	0.029	0.030	0.045	0.058
CEM									
average	3.24	4.34	-0.112	5.67	5.48	-0.268	0.132	0.099	0.037
				5.66	-5.55	0.090	-0.220	0.082	0.048
				-5.72	5.50	0.050	0.084	-0.224	0.090
				-5.74	-5.53	0.038	0.096	0.136	-0.270
RMSE	0.41	0.51	0.129	0.42	0.63	0.113	0.072	0.031	0.027
				0.39	0.54	0.048	0.095	0.042	0.015
				0.34	0.57	0.015	0.040	0.092	0.048
				0.38	0.63	0.025	0.034	0.068	0.111
DMLO									
average	3.15	4.11	-0.125	5.65	5.45	-0.331	0.176	0.120	0.035
				5.55	-5.41	0.117	-0.274	0.096	0.060
				-5.54	5.43	0.059	0.095	-0.270	0.116
				-5.65	-5.47	0.037	0.121	0.174	-0.332
RMSE	0.16	0.15	0.043	0.42	0.64	0.060	0.042	0.032	0.031
				0.48	0.63	0.031	0.051	0.033	0.018
				0.49	0.61	0.018	0.034	0.052	0.030
				0.42	0.62	0.031	0.034	0.041	0.060

Table 2: Comparison of CMCMC, DMCMC, CEM, and DMLO for bivariate simulated data with constant volatility and four different states in the drift (top: true values, below: average and root mean squared errors (RMSE) of estimators over 200 samples for each method)

	τ_1 $\times 10^3$	τ_2 $\times 10^3$	ρ_{12}	μ_1 $\times 10^3$	μ_2 $\times 10^3$		Q	
true values	3.00	2.50	0.300	4.00	2.00	-0.300	0.180	0.120
	2.20	2.00	0.400	0.00	0.00	0.090	-0.180	0.090
	3.50	3.00	0.500	-2.00	-4.00	0.120	0.180	-0.300
CMCMC								
average	3.00	2.50	0.299	4.02	1.99	-0.310	0.194	0.116
	2.19	1.99	0.395	0.00	0.02	0.096	-0.190	0.094
	3.49	2.97	0.506	-1.98	-3.99	0.120	0.185	-0.305
RMSE	0.13	0.10	0.054	0.28	0.19	0.045	0.050	0.037
	0.08	0.06	0.033	0.10	0.10	0.024	0.029	0.023
	0.13	0.12	0.038	0.21	0.23	0.034	0.047	0.042
DMCMC								
average	2.94	2.49	0.290	3.93	1.91	-0.279	0.186	0.093
	2.23	2.01	0.389	0.04	-0.02	0.088	-0.175	0.088
	3.43	2.87	0.484	-1.90	-3.89	0.093	0.186	-0.279
RMSE	0.14	0.09	0.049	0.22	0.18	0.045	0.041	0.038
	0.07	0.06	0.035	0.10	0.08	0.019	0.023	0.017
	0.14	0.17	0.043	0.23	0.24	0.039	0.038	0.042
DEM								
average	2.95	2.50	0.292	3.90	1.90	-0.278	0.184	0.094
	2.22	2.01	0.393	-0.02	-0.02	0.088	-0.173	0.085
	3.45	2.91	0.492	-1.88	-3.85	0.099	0.172	-0.271
RMSE	0.15	0.09	0.055	0.27	0.21	0.047	0.041	0.038
	0.07	0.06	0.033	0.09	0.08	0.021	0.026	0.017
	0.14	0.15	0.040	0.25	0.27	0.035	0.037	0.046
DMLO								
average	2.95	2.49	0.293	3.92	1.91	-0.274	0.184	0.090
	2.23	2.01	0.395	0.03	-0.03	0.085	-0.172	0.087
	3.45	2.88	0.489	-1.90	-3.87	0.095	0.179	-0.274
RMSE	0.13	0.09	0.047	0.24	0.18	0.049	0.044	0.041
	0.08	0.06	0.031	0.10	0.09	0.020	0.024	0.017
	0.13	0.17	0.040	0.20	0.24	0.037	0.038	0.044

Table 3: Comparison of CMCMC, DMCMC, DEM, and DMLO for bivariate simulated data where drift and volatility are switching between three states (top: true values, below: average and root mean squared errors (RMSE) of estimators over 200 samples for each method)

	Number of states d				
	1	2	3	4	5
Multivariate MSM	37 358.3	35 483.9	35 120.9	**34 829.3**	34 855.2
S&P 500 – individual MSM	7 929.7	7 399.1	**7 342.0**	7 377.0	7 445.3
IPC – individual MSM	9 597.9	9 005.8	**8 996.7**	9 024.6	9 072.2
MerVal – individual MSM	11 429.8	10 720.1	**10 688.1**	10 719.6	NaN
Bovespa – individual MSM	11 348.7	10 694.9	**10 619.9**	10 645.8	10 687.7
Independent MSM	40 306.1	37 819.9	**37 646.7**	37 767.7	NaN

Table 4: Selecting the number of states for various MSMs using BIC

		State k			
		1	2	3	4
S&P 500	$\mu_1^{(k)}$	-0.07 (0.04)	0.10 (0.02)	0.12 (0.13)	-0.01 (0.05)
	$\tau_1^{(k)}$	1.19 (0.04)	0.60 (0.02)	1.90 (0.10)	1.10 (0.04)
IPC	$\mu_2^{(k)}$	-0.09 (0.07)	0.24 (0.03)	0.10 (0.20)	0.09 (0.05)
	$\tau_2^{(k)}$	1.67 (0.06)	0.90 (0.03)	2.88 (0.15)	1.05 (0.04)
MerVal	$\mu_3^{(k)}$	-0.10 (0.06)	0.23 (0.05)	-0.09 (0.24)	0.19 (0.15)
	$\tau_3^{(k)}$	1.49 (0.05)	1.30 (0.04)	3.59 (0.20)	3.23 (0.12)
Bovespa	$\mu_4^{(k)}$	-0.05 (0.07)	0.29 (0.05)	-0.01 (0.29)	0.07 (0.09)
	$\tau_4^{(k)}$	1.88 (0.06)	1.33 (0.04)	4.53 (0.25)	1.80 (0.06)
S&P 500/IPC	$\rho_{12}^{(k)}$	0.59 (0.03)	0.56 (0.03)	0.59 (0.04)	0.55 (0.03)
S&P 500/MerVal	$\rho_{13}^{(k)}$	0.53 (0.03)	0.35 (0.04)	0.49 (0.05)	0.04 (0.05)
S&P 500/Bovespa	$\rho_{14}^{(k)}$	0.54 (0.03)	0.55 (0.03)	0.48 (0.05)	0.46 (0.04)
IPC/MerVal	$\rho_{23}^{(k)}$	0.56 (0.03)	0.31 (0.04)	0.62 (0.04)	0.13 (0.04)
IPC/Bovespa	$\rho_{24}^{(k)}$	0.58 (0.03)	0.55 (0.03)	0.63 (0.04)	0.39 (0.04)
MerVal/Bovespa	$\rho_{34}^{(k)}$	0.60 (0.03)	0.43 (0.03)	0.72 (0.03)	0.11 (0.04)

Table 5: Estimated drift $\mu_i^{(k)}$, estimated volatility $\tau_i^{(k)}$, and estimated correlation $\rho_{ij}^{(k)}$ between the various indices in the four different states of a multivariate MSM (standard errors are given in parenthesis)

	State k			
	1	2	3	4
X_{1k}	0.92 (0.01)	0.02 (0.01)	0.06 (0.01)	0.01 (0.01)
X_{2k}	0.02 (0.01)	0.97 (0.01)	0.00 (0.00)	0.01 (0.01)
X_{3k}	0.16 (0.05)	0.01 (0.01)	0.80 (0.05)	0.03 (0.02)
X_{4k}	0.02 (0.01)	0.02 (0.01)	0.01 (0.01)	0.95 (0.02)
Duration	12.8 (2.4)	33.8 (9.9)	5.4 (1.4)	23.9 (9.1)
Q_{1k}	-0.091 (0.02)	0.019 (0.01)	0.065 (0.02)	0.007 (0.01)
Q_{2k}	0.019 (0.01)	-0.033 (0.01)	0.002 (0.01)	0.011 (0.01)
Q_{3k}	0.187 (0.06)	0.009 (0.01)	-0.227 (0.06)	0.031 (0.02)
Q_{4k}	0.024 (0.01)	0.019 (0.01)	0.006 (0.01)	-0.049 (0.02)
Duration	11.5 (2.4)	33.0 (9.9)	4.8 (1.4)	23.3 (9.1)

Table 6: Estimated discrete time transition matrix X, estimated continuous time rate matrix Q, and corresponding estimated durations of each state of a multivariate MSM (standard errors are given in parenthesis)

		State k		
		1	2	3
S&P 500	$\mu_1^{(k)}$	0.08 (0.02)	-0.04 (0.13)	0.00 (0.03)
	$\tau_1^{(k)}$	0.61 (0.02)	2.07 (0.14)	1.12 (0.04)
	Duration	76.4 (39.1)	22.7 (9.9)	50.2 (15.0)
IPC	$\mu_2^{(k)}$	0.53 (0.51)	-0.13 (0.09)	0.17 (0.03)
	$\tau_2^{(k)}$	3.58 (0.51)	1.82 (0.12)	0.94 (0.02)
	Duration	6.5 (7.4)	22.9 (11.6)	68.0 (20.9)
MerVal	$\mu_3^{(k)}$	0.18 (0.34)	0.14 (0.04)	-0.08 (0.11)
	$\tau_3^{(k)}$	4.60 (0.34)	1.26 (0.04)	2.09 (0.18)
	Duration	5.4 (2.6)	44.4 (13.0)	10.9 (4.1)
Bovespa	$\mu_4^{(k)}$	0.18 (0.04)	-0.13 (0.11)	0.27 (1.18)
	$\tau_4^{(k)}$	1.47 (0.04)	2.61 (0.17)	7.30 (1.41)
	Duration	71.8 (22.0)	27.2 (9.0)	38.6 (63.0)

Table 7: Estimated drift $\mu_i^{(k)}$, estimated volatility $\tau_i^{(k)}$, and estimated duration for the three different states of each univariate MSM (standard errors are given in parenthesis)

	CMCMC		DMCMC		DEM		DMLO	
states	$n=1$	$n=4$	$n=1$	$n=4$	$n=1$	$n=4$	$n=1$	$n=4$
2	104	302	10	11	3 (100)	7 (100)	4 (100)	62 (100)
3	104	298	16	17	11 (100)	25 (200)	17 (100)	415 (200)
4	108	310	19	20	30 (100)	47 (300)	84 (200)	827 (300)
5	110	312	41	39	36 (200)	76 (800)	186 (200)	8250 (800)

Table 8: Total runtimes in CPU minutes for different numbers of states for univariate ($n = 1$) and multivariate ($n = 4$) MSMs; the total number of draws is 25 000 for CMCMC and 15 000 for DMCMC; numbers of runs used for DEM and DMLO in parentheses)

	DMCMC				CMCMC			
	$k=1$	$k=2$	$k=3$	$k=4$	$k=1$	$k=2$	$k=3$	$k=4$
$\mu_1^{(k)}$	1.51	1.80	1.45	2.20	8.32	6.15	9.54	4.55
$\mu_2^{(k)}$	1.44	2.95	1.41	3.72	12.36	6.01	16.38	6.91
$\mu_3^{(k)}$	1.76	2.28	1.26	2.68	11.72	5.81	7.32	3.14
$\mu_4^{(k)}$	1.74	2.30	1.31	1.72	10.92	6.87	6.05	7.25
$\tau_1^{(k)}$	1.80	5.10	2.80	2.87	29.13	16.61	31.36	12.17
$\tau_2^{(k)}$	1.35	3.97	2.08	3.42	34.78	11.86	38.76	14.36
$\tau_3^{(k)}$	1.50	3.94	2.17	3.26	29.23	19.59	37.58	18.01
$\tau_4^{(k)}$	2.14	4.97	1.79	4.15	28.25	6.90	38.78	10.99
Q_{1k}	8.96	10.82	9.77	4.64	36.96	33.66	40.18	39.97
Q_{2k}	8.72	9.00	4.61	11.22	27.91	37.01	20.71	37.34
Q_{3k}	7.85	4.12	6.34	2.33	42.27	41.29	42.42	38.33
Q_{4k}	7.14	12.29	4.31	11.74	40.01	39.93	32.72	36.53

Table 9: Inefficiency factors for the posterior draws of the drift $\mu_i^{(k)}$, the volatility $\tau_i^{(k)}$ and the elements Q_{jk} of the rate matrix in the four different states of a multivariate MSM based on DMCMC (left hand side) and CMCMC (right hand side)

		State k							
		1		2		3		4	
CMCMC	$\mu_1^{(k)}$	-0.06	(0.05)	0.10	(0.02)	0.15	(0.16)	-0.01	(0.05)
	$\tau_1^{(k)}$	1.21	(0.05)	0.61	(0.02)	2.04	(0.11)	1.11	(0.04)
	$\mu_2^{(k)}$	-0.10	(0.07)	0.23	(0.03)	0.21	(0.27)	0.08	(0.05)
	$\tau_2^{(k)}$	1.70	(0.08)	0.94	(0.03)	3.10	(0.21)	1.07	(0.04)
	$\mu_3^{(k)}$	-0.12	(0.06)	0.22	(0.05)	-0.00	(0.30)	0.16	(0.15)
	$\tau_3^{(k)}$	1.52	(0.06)	1.29	(0.04)	3.97	(0.25)	3.26	(0.13)
	$\mu_4^{(k)}$	-0.06	(0.08)	0.28	(0.05)	0.03	(0.36)	0.06	(0.09)
	$\tau_4^{(k)}$	1.92	(0.07)	1.36	(0.04)	4.85	(0.31)	1.82	(0.07)
DEM	$\mu_1^{(k)}$	-0.06		0.10		0.13		-0.01	
	$\tau_1^{(k)}$	1.17		0.58		1.90		1.10	
	$\mu_2^{(k)}$	-0.09		0.24		0.11		0.10	
	$\tau_2^{(k)}$	1.66		0.88		2.88		1.04	
	$\mu_3^{(k)}$	-0.10		0.23		-0.08		0.18	
	$\tau_3^{(k)}$	1.47		1.29		3.59		3.20	
	$\mu_4^{(k)}$	-0.05		0.29		0.00		0.06	
	$\tau_4^{(k)}$	1.87		1.30		4.50		1.79	
DMLO	$\mu_1^{(k)}$	-0.06		0.10		0.12		-0.01	
	$\tau_1^{(k)}$	1.17		0.58		1.90		1.10	
	$\mu_2^{(k)}$	-0.09		0.24		0.11		0.10	
	$\tau_2^{(k)}$	1.66		0.88		2.88		1.04	
	$\mu_3^{(k)}$	-0.10		0.23		-0.08		0.18	
	$\tau_3^{(k)}$	1.47		1.29		3.59		3.20	
	$\mu_4^{(k)}$	-0.04		0.29		0.00		0.06	
	$\tau_4^{(k)}$	1.86		1.31		4.50		1.79	

Table 10: Estimated drift $\mu_i^{(k)}$ and estimated volatility $\tau_i^{(k)}$ in the four different states of a multivariate MSM using CMCMC (standard errors in parenthesis), DEM, and DMLO

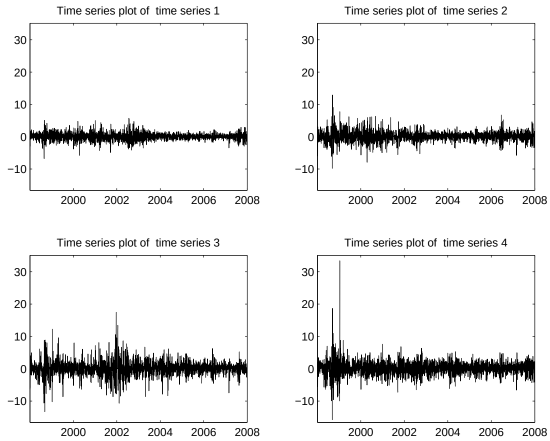

Figure 1: Daily returns of four stock indices (1 . . . S&P 500, 2 . . . IPC, 3 . . . MerVal, 4 . . . Bovespa) from Jan 2, 1998 to Dec 31, 2007

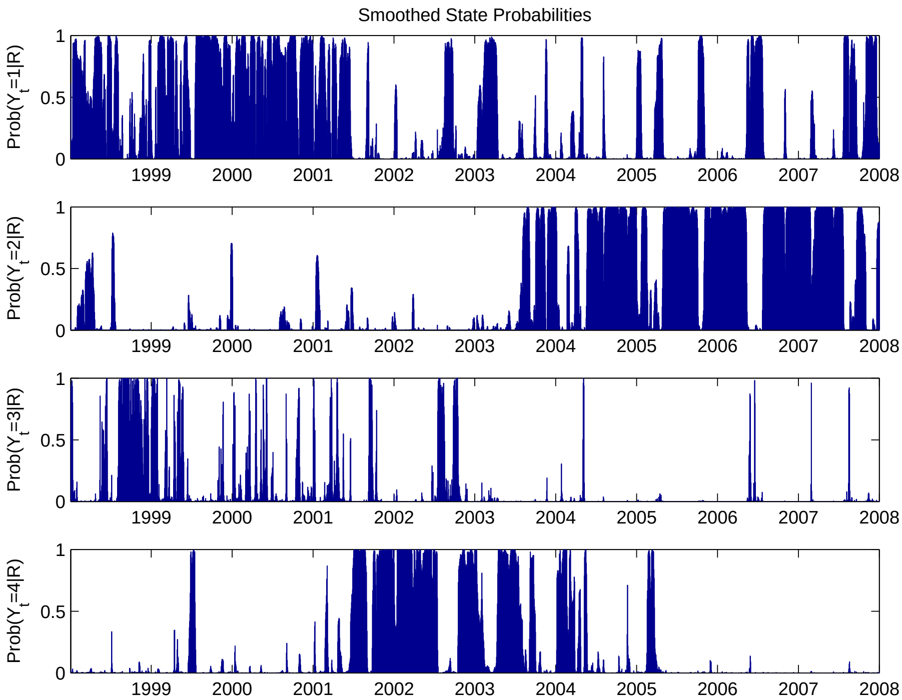

Figure 2: Smoothed state probabilities from Jan 2, 1998 to Dec 31, 2007

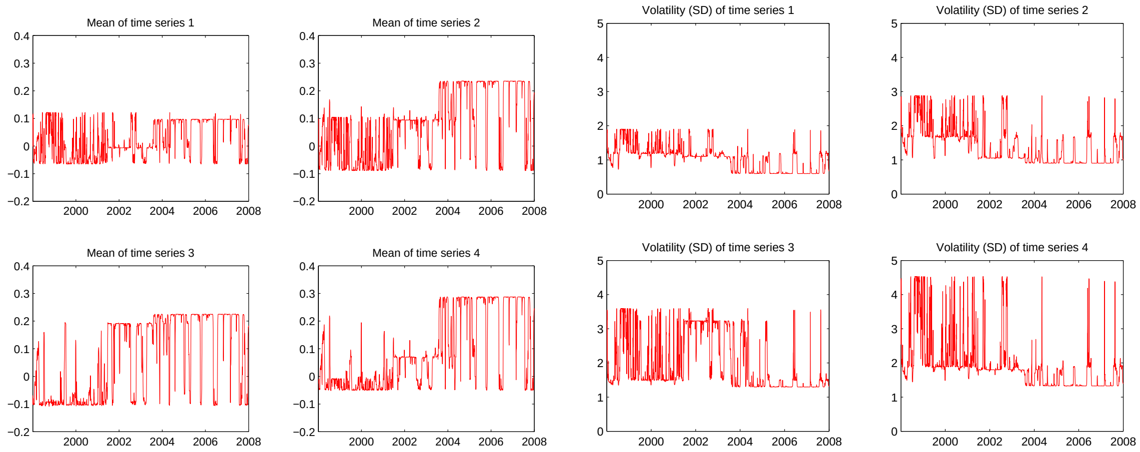

Figure 3: Time-varying mean (left hand side) and volatility (right hand side)

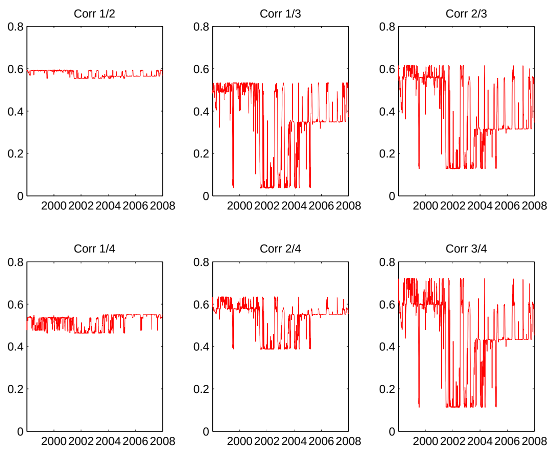

Figure 4: Time-varying correlation