

Contents lists available at [ScienceDirect](#)

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Filterbased stochastic volatility in continuous-time hidden Markov models

Vikram Krishnamurthy^a, Elisabeth Leoff^b, Jörn Sass^{c,*}

^a Department of Electrical and Computer Engineering, Cornell Tech, Cornell University, USA

^b Department of Financial Mathematics, Fraunhofer Institute for Industrial Mathematics ITWM, Kaiserslautern, Germany

^c Department of Mathematics, University of Kaiserslautern, Kaiserslautern, Germany

ARTICLE INFO

Article history:

Received 5 February 2016
Revised 14 October 2016
Accepted 14 October 2016
Available online xxx

Keywords:

Markov switching model
Non-constant volatility
Stylized facts
Portfolio optimization
Social learning

ABSTRACT

Regime-switching models, in particular Hidden Markov Models (HMMs) where the switching is driven by an unobservable Markov chain, are widely-used in financial applications, due to their tractability and good econometric properties. In continuous time, properties of HMMs with constant and of HMMs with switching volatility can be quite different. To have a realistic model with unobservable Markov chain in continuous time and good econometric properties, a regime-switching model where the volatility depends on the filter for the underlying chain is introduced and the filtering equations are stated. Such models are motivated by agent based social learning models in economics. An approximation result for a fixed information filtration is proved and further motivation is provided by considering social learning arguments. The relation to the switching volatility model is analyzed in detail and a convergence result for the discretized model is given. Econometric properties are illustrated by numerical simulations.

© 2016 ECOSTA ECONOMETRICS AND STATISTICS. Published by Elsevier B.V. All rights reserved.

1. Introduction

Regime-switching models are very popular in the field of mathematical finance to describe return processes with time-changing drift or volatility parameters. They are a possible way to generalize the classical Black–Scholes log normal stock price model by making the parameters dependent on a Markov chain with finitely many states. This Markov chain can be interpreted as encoding all the information in the market into one single process that describes the underlying state of the economy. As such this chain is assumed to be unobservable, the normal investor does not have access to all the possible information. The only observation that can be made is the return process itself.

In this work, we consider both the Hidden Markov Model (HMM) and what we call the Markov Switching Model (MSM) in continuous time. Both models have been studied extensively and are popular for applications. The difference between them is only that the HMM has constant volatility, while in the MSM the volatility jumps with the Markov chain. This difference leads to a difference in the behavior of information between both models, since the volatility is observable via the quadratic variation in continuous time. Such a distinction is not present in discrete time, cf. [Elliott et al. \(2008\)](#), [Elliott and Siu \(2012\)](#), [Guo \(2001\)](#), [Siu \(2011\)](#). For comparison of estimation procedures for the continuous-time MSM we refer to [Hahn et al. \(2010\)](#).

* Corresponding author.

E-mail addresses: vikramk@cornell.edu (V. Krishnamurthy), elizabeth.leoff@itwm.fraunhofer.de (E. Leoff), sass@mathematik.uni-kl.de (J. Sass).

<http://dx.doi.org/10.1016/j.ecosta.2016.10.007>

2452-3062/© 2016 ECOSTA ECONOMETRICS AND STATISTICS. Published by Elsevier B.V. All rights reserved.

In finance, one often needs models that are defined in continuous time since these allow for explicit results. To be applicable these have to be close to the results for the real-world discrete-time setting. For example, in portfolio optimization the aim is to find the strategy that maximizes the expected utility at terminal time. These optimal strategies can be derived explicitly only in continuous time, but in the HMM they are close to the discrete-time strategies, cf. the detailed discussion in Section 2.2. Consequently, the continuous-time HMM is such a model which is very well tractable and yields results close to the discrete-time model. Thus we may call it discretization-consistent.

The volatility in the continuous-time HMM is assumed to be constant. Empirical observations of financial data across a wide range of instruments, markets and time periods indicate that for many real data sets this is not true, as was already noted by e.g., Fama (1965). This leads to the concept of stochastic volatility in model building. There have been many different modeling approaches leading to various models where the volatility itself is a stochastic process. Popular models are e.g., Engle's famous Autoregressive Conditional Heteroskedasticity model (ARCH) from 1982, see Engle (1982), and its generalization GARCH, see Bollerslev (1986) and also cf. Taylor (1994), or in continuous time the well-known Heston model (Heston, 1993).

In the context of regime-switching models, the natural way to introduce a non-constant volatility is to use the Markov chain to control the volatility, as usually done in discrete time when considering regime switching models, see Hamilton (1989). Switching can also be introduced in the volatility models mentioned above. Switching between constant values for the volatility leads in continuous time to the MSM, which has better econometric properties than the HMM for many applications. Some stylized facts such as volatility clustering cannot be reproduced in the HMM. However, it turns out that in the continuous-time MSM the Markov chain can be observed using the quadratic variation of the return process, see Proposition 2.1 below. Therefore, results obtained in the MSM can be far away from results obtained in the corresponding discrete-time model. Thus we do not have the discretization-consistency for the continuous-time MSM as we have it for the HMM.

To get a model which is both consistent and has good econometric properties, the idea now is to introduce stochastic volatility in a continuous-time HMM. Haussmann and Sass (2004) propose a regime-switching model where the volatility depends on an observable diffusion. They prove that despite the non-constant volatility in this model the chain is still hidden and derive its filtering equation which is again finite-dimensional. The question of how to actually model the volatility process is still left open. Due to the better econometric properties of the MSM our goal is to find a volatility process that leads to an HMM that in a sense approximates the continuous-time MSM.

Thus, in this work, we introduce an HMM in continuous time where the volatility is a linear function of the (normalized) filter. For a given filtration we construct the volatility process adapted to this filtration that results in the best approximation of the MSM-returns in the mean squared sense. This result motivates the introduction of the Filterbased-Volatility Hidden Markov Model (FB-HMM), where the volatility is a linear function of the normalized filter. In this model, the returns and the filter constitute a system of equations coupled by the volatility process. In a sense, this model lies between the HMM and MSM: the chain is still hidden as in the HMM, but the volatility is non-constant similar to the MSM. The function that connects the filter to the volatility process in the FB-HMM is even the same function that connects the chain to the volatility in the MSM. Another motivation for the model comes from an instantaneous iteration of observing the returns and trading (resulting in a new volatility) in a financial market, which is similar to a social learning model used in economics, see Remark 3.3.

But as a function of the unnormalized filter, the volatility in the FB-HMM does not satisfy the assumptions in Haussmann and Sass (2004). In this work, we state the filtering equations for the FB-HMM, following the methodology from Elliott (1993), and see that the filter is as tractable as in the HMM from Haussmann and Sass (2004). Further, the issue of how the model behaves under time discretization is very important for practical applications, as we have already mentioned. For the question of approximation and consistency in the FB-HMM we consider the discretized returns. Here, the issue of consistency is a different one than what is typically investigated in statistics. The discretized FB-HMM depends on the continuous-time filter at the discretization points through its volatility, so it is not truly a discrete-time model. For consistency we consider the convergence of the discretized returns instead and prove that already for the simple Euler discretization the global discretization error converges to 0 in L^2 .

The paper is organized as follows: in Section 2, we introduce the HMM and MSM in continuous time, prove the observability of the Markov chain in the continuous-time MSM and consider portfolio optimization in regime-switching models. Section 3 introduces the FB-HMM and prove an approximation result for a fixed information filtration. We further motivate the model by considering social learning arguments and shortly investigate the stylized facts present in the returns. Section 4 proves that the Euler-discretization of the FB-HMM returns converges to the continuous-time returns. We conclude the paper by presenting some numerical results and discuss an extension to a jump diffusion model.

2. The HMM and MSM in continuous time

Consider a return process $(R_t)_{t \in [0, T]}$ with dynamics

$$dR_t = \mu_t dt + \sigma_0 dW_t,$$

where the unobservable drift process $\mu_t = \mu(Y_t)$ jumps between d states following a process Y . The volatility $\sigma_0 \in \mathbb{R}_{>0}$ is constant, where we denote with $\mathbb{R}_{d>0}$ the strictly positive real numbers. $(W_t)_{t \in [0, T]}$ is a standard Brownian motion defined

on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Since μ_t has only finitely many states, we can represent it as $\mu_t = \mu^T Y_t$ with $\mu \in \mathbb{R}^d$ the so-called state vector. Y is assumed to be an irreducible, time-homogenous Markov chain on $\{e_1, \dots, e_d\}$ with $d \times d$ generator matrix Q (so $Q_{ii} < 0$ and $\sum_j Q_{ij} = 0$). We assume that $Y_0 \sim \nu$ with ν the invariant distribution of Y that exists and is unique due to the irreducibility and finite state space of Y .

The above model is reasonable if the drift is influenced by discrete events such as the arrival of news or shocks that change the overall state of the economy. The Markov chain then models this underlying state of the economy that influences the behavior of stock returns.

We assume that the Markov chain Y cannot be observed, only the returns can be observed by the investor. Thus we are in the setting of incomplete information and thus the model is an instance of a continuous time Hidden Markov Model (HMM). Any investment decision can be made only with respect to the information available from observing R , i.e., must be adapted to the filtration $\mathcal{F}_t^R = \sigma(R_s | s \leq t)$. In the terminology of filtering theory, R is the observation while Y is the signal and we are interested in the conditional expectation $\hat{Y}_t := \mathbb{E}[Y_t | \mathcal{F}_t^R]$ called filter and providing the posterior probabilities.

The filtering equation for estimating the state process Y in innovations form is given by the Wonham filter, see [Wonham \(1965\)](#). It is a finite dimensional equation describing the dynamics of \hat{Y}_t :

$$d\hat{Y}_t = Q^T \hat{Y}_t dt + (\text{diag}(g) \hat{Y}_t - (g^T \hat{Y}_t) \hat{Y}_t) dV_t,$$

with $V_t = \int_0^t \sigma_0^{-1} \mu^T (Y_s - \hat{Y}_s) ds + W_t$ the innovation process, an \mathcal{F}^R -adapted \mathbb{P} -Brownian motion, and $g^T = \sigma_0^{-1} \mu^T$. The monograph by [Lipster and Shiryaev \(2001\)](#) provides some discussion on innovation processes.

Since this equation is not linear in the filter \hat{Y} one typically uses the Zakai equation for the so-called unnormalized filter which is the conditional expectation under a new measure.

For these filters we need the stochastic exponential process

$$Z_t = \exp \left(- \int_0^t \sigma_0^{-1} \mu^T Y_s dW_s - \frac{1}{2} \int_0^t \|\sigma_0^{-1} \mu^T Y_s\|^2 ds \right)$$

and with this the new measure $\tilde{\mathbb{P}}$ via $\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = Z_T$. Due to the boundedness of $\sigma_0^{-1} \mu^T Y_t$ Girsanov's theorem guarantees that \tilde{W} defined as

$$d\tilde{W} = \sigma_0^{-1} \mu^T Y_t dt + dW_t$$

is a $\tilde{\mathbb{P}}$ -Brownian motion. Using the new reference measure we can define the unnormalized filter ρ as

$$\rho_t = \tilde{\mathbb{E}}[Z_t^{-1} Y_t | \mathcal{F}_t^R]$$

from which the normalized filter can be derived using the Kallianpur–Striebel formula, which is an abstract version of Bayes' formula:

$$\hat{Y}_t = \frac{\rho_t}{\tilde{\mathbb{E}}[Z_t^{-1} | \mathcal{F}_t^R]}.$$

Observing that

$$1 = \mathbb{E}[1 | \mathcal{F}_t^R] = \mathbb{E}[1^T Y_t | \mathcal{F}_t^R] = 1^T \mathbb{E}[Y_t | \mathcal{F}_t^R] = 1^T \hat{Y}_t = \frac{1^T \rho_t}{\tilde{\mathbb{E}}[Z_t^{-1} | \mathcal{F}_t^R]}$$

we see that

$$\hat{Y}_t = \frac{\rho_t}{1^T \rho_t}.$$

Thus for filtering purposes it suffices to compute the unnormalized filter ρ which has the advantage that it follows a linear equation driven by the observation process (see [Elliott \(1993\)](#); [Elliott et al. \(1995\)](#) who use a reference probability method to obtain this Zakai equation):

$$d\rho_t = Q^T \rho_t dt + \frac{1}{\sigma_0^2} \text{diag}(\mu) \rho_t dR_t.$$

One drawback of the HMM is that it assumes the returns evolve with a constant volatility. Since we model the drift as depending on the Markov chain describing the current state of the economy it is natural to model the volatility as depending on the same Markov chain. This leads us to what we call the Markov Switching Model (MSM), as one does usually in the discrete-time model.

2.1. The Markov Switching Model

In the MSM, we assume that the volatility evolves with time depending on the Markov chain Y in the same way the drift does, i.e.,

$$dR_t = \mu_t dt + \sigma_t dW_t,$$

with

$$\mu_t = \mu^T Y_t, \quad \mu \in \mathbb{R}^d, \quad \text{and} \quad \sigma_t = \sigma^T Y_t, \quad \sigma \in \mathbb{R}_{>0}^d,$$

where $\mathbb{R}_{>0}^d$ denotes the d -dimensional vectors with strictly positive entries.

The main feature of the MSM is that it allows for a changing volatility process over time. For the case of $\sigma_i = \sigma_j$ for all i, j it can be seen as a generalization of the HMM. One property of the model is that in continuous time the dependence of the volatility on the same Markov chain as the drift makes the Markov chain observable, as we prove in the following proposition.

Proposition 2.1. *Suppose $dR_t = \mu^T Y_t dt + \sigma^T Y_t dW_t$ and that $\sigma_i \neq \sigma_j$ for $i \neq j$. Then the process Y is adapted to \mathcal{F}^R . Therefore $\mathbb{E}[Y_t | \mathcal{F}_t^R] = Y_t$ implying that estimating Y_t is a trivial filtering problem.*

Proof. Fix $t \in [0, T]$ and let $n \in \mathbb{N}$. Due to the exponentially distributed waiting times of Y for almost all $\omega \in \Omega$ we can find some $\varepsilon_\omega > 0$ such that

$$Y_t(\omega) = Y_{t+\varepsilon}(\omega)$$

for $0 \leq \varepsilon \leq \varepsilon_\omega$. Thus the following limit equality holds almost surely

$$\begin{aligned} (\sigma^T Y_t)^2 &= \lim_{m \rightarrow \infty} \frac{1}{\varepsilon_m} \int_t^{t+\varepsilon_m} (\sigma^T Y_s)^2 ds = \lim_{m \rightarrow \infty} \frac{1}{\varepsilon_m} \int_t^{t+\varepsilon_m} (\sigma^T Y_s)^2 ds \\ &= \lim_{m \rightarrow \infty} \frac{1}{\varepsilon_m} ([R]_{t+\varepsilon_m} - [R]_t), \end{aligned}$$

where ε_m is some sequence converging to 0 from above and $[R]_t = \int_0^t (\sigma^T Y_s)^2 ds$ is the quadratic variation of R . There exist M_n such that $\varepsilon_m < \frac{1}{n}$ for all $m > M_n$, so $\frac{1}{\varepsilon_m} ([R]_{t+\varepsilon_m} - [R]_t)$ is measurable with respect to $\mathcal{F}_{t+\frac{1}{n}}^R$ for $m > M_n$. In particular, it follows that the limit $(\sigma^T Y_t)^2$ is $\mathcal{F}_{t+\frac{1}{n}}^R$ -measurable. Since $\sigma_i \neq \sigma_j$ for $i \neq j$ and $\sigma_i > 0$ this implies that Y_t is measurable with respect to $\mathcal{F}_{t+\frac{1}{n}}^R$ for all $n \in \mathbb{N}$, so also with respect to the intersection $\bigcap_{n \in \mathbb{N}} \mathcal{F}_{t+\frac{1}{n}}^R$. But \mathcal{F}^R is a right-continuous filtration, so

$$\bigcap_{n \in \mathbb{N}} \mathcal{F}_{t+\frac{1}{n}}^R = \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}^R = \mathcal{F}_t^R$$

and Y is indeed adapted to \mathcal{F}^R . \square

This result has useful implications for fitting the continuous-time MSM to real data: we can observe (discretized) returns on the stock market, but there is no process available that directly corresponds to the Markov chain.

The chain describes the general “state of the economy”, a quantity that is influenced by several different factors with different characteristics, like high-level political decisions, central banks changing their interest rates, the development of real estate prices and similar indices. It is not feasible to quantify all these factors and distill their information into one process Y , so for our model we just assumed the chain to be unobservable. But Proposition 2.1 tells us that this is not possible for a regime-switching model in continuous time with switching volatilities, the theory always implies that the chain is adapted to the observation filtration. This difference between the HMM and MSM is not present in discrete time. In the discrete-time models, Y is not observable for both constant and switching volatility. Thus, even with careful discretization the information from observing the discrete-time MSM might be far away from the continuous-time model. Consequently, the results obtained in the continuous-time model might not be close the results in the discrete-time MSM. In this sense, the MSM is not consistent.

2.2. Portfolio Optimization in HMM and MSM

To illustrate the difference of HMM and MSM as motivated in the introduction we consider as application in finance a portfolio optimization problem. We view the process R above as stock returns and introduce further a money market account with constant interest rate $r = 0$. We fix a time horizon $T > 0$ and describe trading by the fraction π_t of the portfolio value (wealth) X_t^π invested at $t \in [0, T]$ in stocks. For initial capital x_0 , the wealth controlled by $\pi = (\pi_t)_{t \in [0, T]}$ evolves as

$$dX_t^\pi = X_t^\pi (1 - \pi_t) r dt + X_t^\pi \pi_t dR_t, \quad X_0^\pi = x_0.$$

An investor whose preferences are given by a utility function U , wishes to maximize expected utility of terminal wealth,

$$\mathbb{E}[U(X_T^\pi)],$$

over all *admissible* strategies π . Typical utility functions are

$$U_\alpha(x) = \frac{x^\alpha}{\alpha}, \quad \alpha < 1, \alpha \neq 1 \quad \text{or} \quad U_0(x) = \log(x).$$

Let $\hat{Z}_t = \mathbb{E}[Z_t | \mathcal{F}_t^R]$ denote the conditional density. Then [Sass and Haussmann \(2004\)](#) show that in the HMM for utility function U_α

$$\pi_t^{\text{HMM}} = \frac{1}{(1 - \alpha) \mathbb{E}[\hat{Z}_{t,T}^{\frac{\alpha-1}{\alpha}} | \rho_t]} \left\{ \frac{1}{\sigma_0^2} \mu^T \hat{Y}_t \mathbb{E}[\hat{Z}_{t,T}^{\frac{2\alpha-1}{\alpha}} | \rho_t] + \frac{1}{\sigma_0} \mathbb{E}[\hat{Z}_{t,T}^{\frac{2\alpha-1}{\alpha}} \int_t^T (D_t \rho_{t,s}) \mu \frac{1}{\sigma_0^2} dR_s | \rho_t] \right\}$$

is optimal, where $\hat{Z}_{t,T} = \hat{Z}_T / \hat{Z}_t$ and $\rho_{t,s} = \rho_s / \hat{Z}_t^{-1}$. Further $D_t \rho_{t,s}$ is the Malliavin derivative which can be shown to satisfy a linear SDE due to the linearity of the Zakai equation for ρ_t . Note that for $U_0 = \log$ we get

$$\pi_t^{\text{HMM}} = \sigma_0^{-2} \mu^T \hat{Y}_t.$$

In the MSM, [Bauerle and Rieder \(2004\)](#) show that for observable Y (which is the case for an MSM due to [Proposition 2.1](#))

$$\pi_t^{\text{MSM}} = \frac{1}{1 - \alpha} \frac{\mu^T Y_t}{(\sigma^T Y_t)^2}$$

is optimal for U_α with $\alpha = 0$ corresponding to $U_0 = \log$.

Without providing the filtering equations and the model formulation of the more classical discrete-time models here, let us denote the filters in the discretized HMM with constant volatility and in the discretized MSM with switching volatility by \hat{Y}_k^{HMM} and \hat{Y}_k^{MSM} , respectively. We obtain for mild parameters, where *mild* means that the optimal fractions stay in $[0,1]$, that of order Δt the optimal strategies for maximizing logarithmic utility are given by risky fractions

$$\pi_k^{\text{HMM}} = \frac{\mu^T \hat{Y}_k^{\text{HMM}}}{\sigma_0^2} + o(\Delta_t) \quad \text{and} \quad \pi_k^{\text{MSM}} = \frac{\mu^T \hat{Y}_k^{\text{MSM}}}{(\sigma^T \hat{Y}_k^{\text{MSM}})^2} + o(\Delta_t),$$

where Δ_t denotes the discretization step, see [Taksar and Zeng \(2007\)](#). In fact, these optimal fractions correspond to those obtained by a one-period mean variance optimization with corresponding risk aversion parameter. When only looking at portfolio optimization as performance measure for the model, one may use other one-period optimization problems here. But the strategies above are also consistent for the continuous-time and the multi-period setting, respectively.

Note that in π_k^{MSM} the filter \hat{Y}_k^{MSM} appears while the corresponding continuous-time strategy π_t^{MSM} is based on the Markov chain Y_t itself. This leads to a large deviation of the corresponding optimization results which is not desirable. For the HMM both are based on the filter, the discrete-time filter being the discretization of the continuous-time filter, see [James et al. \(1996\)](#). The optimization results for the continuous-time HMM are thus close to the discrete time model, if we look at strategies constrained to $[0, 1]$ meaning that neither short selling nor borrowing are allowed. In this sense, the HMM is discretization-consistent.

This predicament of the MSM, the lack of discretization-consistency, is the main motivation for introducing a filterbased volatility model in the next section.

3. A Filterbased Volatility Model

This section considers a continuous time HMM for stock returns that was introduced by [Haussmann and Sass \(2004\)](#). As in the classical HMM the return process R depends on the Markov chain Y with state space $\{e_1, \dots, e_d\}$ via

$$R_t = \int_0^t \mu^T Y_s ds + \int_0^t \sigma_s dW_s,$$

with W a Brownian motion. The drift process of the returns is again given by $\mu^T Y_t$, with the state vector $\mu \in \mathbb{R}^d$ containing its d possible states.

The difference to the classical HMM is that the volatility process σ_t is no longer a constant over time. We assume that σ_t is uniformly bounded with uniformly bounded inverse σ_t^{-1} .

Since σ_t^{-1} is uniformly bounded the process $\theta_t = \sigma_t^{-1} \mu^T Y_t$, which is the market price of risk, is uniformly bounded as well and the density process $dZ_t = -Z_t \theta_t dW_t$ is a martingale. Thus we can define the reference measure $\tilde{\mathbb{P}}$ for filtering in this model as $d\tilde{\mathbb{P}} = Z_t d\mathbb{P}$. Similar to the HMM with constant volatility Girsanov's Theorem guarantees that $d\tilde{W} = \theta_t dt + dW_t$ is a $\tilde{\mathbb{P}}$ -Brownian motion.

For the volatility σ , [Haussmann and Sass \(2004\)](#) specify a Markovian model with an observable diffusion ξ driven by \tilde{W} :

$$d\xi_t = \nu(\xi_t) dt + \tau(\xi_t) d\tilde{W}_t,$$

with ν, τ continuously differentiable functions of the appropriate dimensions with bounded partial derivatives. The volatility process is then given by

$$\sigma_t = \bar{\sigma}(\xi_t),$$

where the function $\bar{\sigma}$ has bounded partial derivatives.

This choice of volatility model preserves the non-observability of the Markov chain from the HMM while having non-constant volatility as in the MSM. In [Haussmann and Sass \(2004\)](#), the filtering equation for this model was formulated:

Theorem 3.1. *In the HMM with non-constant volatility, the unnormalized filter ρ satisfies*

$$d\rho_t = Q^T \rho_t dt + \frac{1}{\sigma_t^2} \text{diag}(\mu) \rho_t dR_t.$$

Using that $dR_t = \sigma_t d\tilde{W}_t$ we see that

$$d\rho_t = Q^T \rho_t dt + \text{diag}(\mu) \rho_t \sigma_t^{-1} d\tilde{W}_t,$$

so the unnormalized filter (posterior) itself is a diffusion with respect to \tilde{W} and σ_t can be chosen to be a function of the filter.

The intuition behind this choice for ξ is that one interpretation of volatility is that it arises due to the accumulated actions of traders, and traders can only act on the information available to them. The information that is available to market participants about the state of the economy is encoded in the filter, and thus it makes sense to model the volatility as depending on this filter.

The volatility process $\sigma_t = \bar{\sigma}(\xi_t)$ is modeled as a function of an observable diffusion. This model allows for a wide range of possible volatility processes, incorporating e.g., level dependency via $\sigma_t = \bar{\sigma}(S_t)$ and dependencies on other \mathcal{F}^R -adapted processes that capture information about the state of the economy. From the form of the filtering equation in this setting we can see that we can also choose the unnormalized filter as the diffusion, since $\bar{\sigma}$ is assumed to be continuously differentiable with bounded derivatives. This leads to the question of how to choose ξ and $\bar{\sigma}$ in a good way.

3.1. The optimal approximating model given the filtration

We want a model for the stock prices that is as close as possible to the MSM while still allowing for a filtering problem. As we are modeling a return process, closeness here means that the output from both models, the return processes, are close to each other in a quantifiable way. A typical measure is the expected squared distance, also known in statistics as the *mean-squared error* (MSE). Since the MSE is given by an expectation it is not always easy to investigate or clear how to decrease it.

Consider a continuous-time MSM with parameters μ , σ^{MSM} and Q . The return process R^{MSM} is

$$R_t^{\text{MSM}} = \int_0^t \mu^T Y_s ds + \int_0^t (\sigma^{\text{MSM}})^T Y_s dW_s.$$

Assume that we have some filtration \mathcal{G} that describes the available information. We then want to choose a regime-switching model $R^{\mathcal{G}}$ driven by the same Y , W and with observable volatility process $\sigma_t^{\mathcal{G}}$ that approximates the MSM-returns R^{MSM} . The best approximation would be the model \tilde{R} with volatility $\tilde{\sigma}_t$ which minimizes the MSE for all times t over \mathcal{G}_t -measurable $\sigma_t^{\mathcal{G}}$, i.e.,

$$\tilde{\sigma}_t = \arg \min_{\sigma_t^{\mathcal{G}} \mathcal{G}_t\text{-mbl.}} \mathbb{E} \left[|R_t^{\text{MSM}} - R_t^{\mathcal{G}}|^2 \right].$$

Then the minimizers $\tilde{\sigma}_t$ form a \mathcal{G} -adapted volatility process $\tilde{\sigma}$. Such a minimization problem is typically not easy to solve. Fortunately, it turns out that in our situation we can reduce the problem to the minimization problem that has a solution in terms of conditional expectations:

Theorem 3.2. *For a fixed filtration \mathcal{G} the minimizer of*

$$\min_{\sigma_t^{\mathcal{G}} \mathcal{G}_t\text{-mbl.}} \mathbb{E} \left[|R_t^{\text{MSM}} - R_t^{\mathcal{G}}|^2 \right]$$

is given by

$$\tilde{\sigma}_t = \sigma^T \mathbb{E}[Y_t | \mathcal{G}_t].$$

Proof. The difference between the return process R^{MSM} and any $R^{\mathcal{G}}$ is

$$\begin{aligned} R_t^{\text{MSM}} - R_t^{\mathcal{G}} &= \int_0^t \mu^T Y_s ds + \int_0^t \sigma^T Y_s dW_s - \int_0^t \mu^T Y_s ds - \int_0^t \sigma_s^{\mathcal{G}} dW_s \\ &= \int_0^t \sigma^T Y_s - \sigma_s^{\mathcal{G}} dW_s, \end{aligned}$$

so with Itô's isometry it follows that

$$\mathbb{E}[(R_t^{\text{MSM}} - R_t^{\mathcal{G}})^2] = \mathbb{E} \left[\left(\int_0^t \sigma^T Y_s - \sigma_s^{\mathcal{G}} dW_s \right)^2 \right] = \mathbb{E} \left[\int_0^t (\sigma^T Y_s - \sigma_s^{\mathcal{G}})^2 ds \right].$$

With Fubini we can reformulate this as

$$\mathbb{E}[(R_t^{\text{MSM}} - R_t^{\mathcal{G}})^2] = \int_0^t \mathbb{E}[(\sigma^T Y_s - \sigma_s^{\mathcal{G}})^2] ds.$$

The integral is minimized if we can find a \mathcal{G} -measurable $\tilde{\sigma}$ such that at any time s $\mathbb{E}[(\sigma^T Y_s - \tilde{\sigma}_s)^2]$ is smaller than $\mathbb{E}[(\sigma^T Y_s - \sigma_s^\mathcal{G})^2]$ for any possible $\sigma_s^\mathcal{G}$.

But we know that minimizing the L^2 -distance over the set of random variables measurable w.r.t. a certain σ -algebra is solved by the conditional expectation w.r.t. this σ -algebra, so

$$\arg \min_{\sigma_s^\mathcal{G}\text{-mbl.}} \mathbb{E}[(\sigma^T Y_s - \sigma_s^\mathcal{G})^2] = \mathbb{E}[\sigma^T Y_s | \mathcal{G}_s]$$

and

$$\mathbb{E}[\sigma^T Y_s | \mathcal{G}_s] = \sigma^T \mathbb{E}[Y_s | \mathcal{G}_s].$$

So the choice of $\tilde{\sigma}_t = \sigma^T \mathbb{E}[Y_t | \mathcal{G}_t]$ minimizes $\mathbb{E}[(R_t^{\text{MSM}} - R_t^\mathcal{G})^2]$ over all processes that are \mathcal{G} -adapted. \square

3.2. Iteration and social learning

We see that for a given information filtration \mathcal{G} the best approximation of the MSM would be given by

$$\tilde{R}_t = \int_0^t \mu^T Y_s ds + \int_0^t \sigma^T \mathbb{E}[Y_s | \mathcal{G}_s] dW_s.$$

But in practice one chooses $\mathcal{G}_t = \mathcal{F}_t^R$ which depends on the “real” volatility process. With this choice of \mathcal{G} changing the model for the volatility changes the available information to $\mathcal{G}^1 = \mathcal{F}^{\tilde{R}}$. We end up with a new minimization problem

$$\tilde{\sigma}_t^{(2)} = \arg \min_{\sigma_t^{\mathcal{G}_t^1}\text{-mbl.}} \mathbb{E} \left[|R_t^{\text{MSM}} - R_t^{\mathcal{G}_t^1}|^2 \right]$$

that is again solved by

$$\tilde{\sigma}_t^{(2)} = \sigma^T \mathbb{E}[Y_t | \mathcal{G}_t^1]. \tag{3.1}$$

This gives rise to another filtration generated by $d\tilde{R}_t^{(2)} = \mu^T Y_t dt + \tilde{\sigma}_t^{(2)} dW_t$, $\mathcal{G}^2 = \mathcal{F}^{\tilde{R}^{(2)}}$. For this filtration we can solve for the best approximation of the MSM using the information \mathcal{G}^2 and end up with another approximating model $\tilde{R}^{(3)}$ and a new information filtration.

In such a manner, we can iteratively construct a sequence of models that approximate the MSM and depend on different information filtrations. These models all have the serious drawback that their volatilities $\tilde{\sigma}^{(i)}$ are not adapted to the filtrations \mathcal{G}^i generated by the corresponding returns. By construction the volatilities are instead adapted to the filtration \mathcal{G}^{i-1} generated by the returns from the previous step. Thus, we end up with models that are not in the class of HMMs from [Haussmann and Sass \(2004\)](#). In particular, the results from [Haussmann and Sass \(2004\)](#) cannot be applied to calculate their filters.

Remark 3.3 (Bayesian social learning). The above iterative model involving the filtered estimates can also be motivated from a social learning perspective, such models are used widely in economics to model multi-agent systems see [Chamley \(2004\)](#); [Krishnamurthy \(2016\)](#); [Krishnamurthy et al. \(2014\)](#).

Let us first describe the classical social learning model in discrete time. Consider a Markov chain Y with discrete time observations

$$R_{t_k} = \mu^T Y_{t_k} + \sigma_{t_k} w_{t_k}, \quad k = 1, 2, \dots$$

where w denotes an independent and identically distributed process. The dynamics of the social learning model proceeds iteratively as follows: at time t_k , the public belief $\hat{Y}_{k-1} = \mathbb{E}[Y_{t_{k-1}} | a_1, \dots, a_{k-1}]$ is available. At time t_k , given observation R_{t_k} , agent k chooses action

$$a_k = \arg \min_{a \in \mathcal{A}} \mathbb{E}[c(Y_{t_k}, a) | a_1, \dots, a_{k-1}, R_k]$$

for some specified cost function $c(y, a)$ and finite action set \mathcal{A} . Given this new action a_k , the public belief is updated as

$$\hat{Y}_k = \mathbb{E}[Y_{t_k} | a_1, \dots, a_k],$$

and the model proceeds to time t_{k+1} , and so on. An interesting property of the above social learning model is that agents can form herds and information cascades [Chamley \(2004\)](#).

We now construct a modified version of the model considered in this paper which is similar in spirit to the above social learning model. Suppose the first trader observes the returns, filters and trades accordingly yielding a new volatility since volatility is generated by trading. The next trader sees the returns coming from this new volatility, computes the filter (which may now be different) and trades accordingly, yielding again a new volatility. This iteration continues and might end at some model as defined below (if we reach a fixed point of this iteration). In practice, such a transition of information would occur with small time-steps. Here the volatility is rather motivated as arising from an instantaneous updating which can be seen as instantaneous or degenerate social learning. For trading it might further be motivated from

high-frequency trading which allows for a lot of iterations within short time intervals. In [Krishnamurthy and Bhatt \(2016\)](#), we have justified such models with change detection and a CVaR risk measure.

Considering the iterative construction of the models in (3.1) we potentially end up with an infinite sequence. It is not clear whether the iteration converges.

Remark 3.4. In the discrete time social learning model outlined above, assuming a finite action set and finite observation set with the Markov chain having identity transition matrix, an elementary application of the martingale convergence theorem shows that the public belief will converge to a constant with probability one in finite time – this is called an information cascade.

For convergence in our model we would need a volatility process such that \mathcal{G}^{i-1} and \mathcal{G}^i coincide. The sequence obviously depends strongly on the initial information \mathcal{G} , thus the question is also how to choose it in a reasonable way. Considering our original goal of finding a regime-switching model with better properties compared to the HMM it makes sense to first start with information coming from HMM returns, $\mathcal{G} = \mathcal{F}^{\text{HMM}}$. Then in the next step we have

$$\tilde{\sigma}_t = \sigma^T \hat{Y}_t^{\text{HMM}},$$

where $\hat{Y}_t^{\text{HMM}} = \mathbb{E}[Y_t | \mathcal{F}_t^{\text{HMM}}]$ is the HMM-filter. But we still have the problem that in this model we cannot compute the filter. We can view this model as the first step on the way from an HMM to an HMM with non-constant volatility, but we do not truly reach the class of models from [Haussmann and Sass \(2004\)](#).

3.3. The Filterbased Volatility Model

To make the above iterative approach useful, we now construct a model that is invariant under the iteration (3.1). Consider the following model

$$dR_t^{\text{FB}} = \mu^T Y_t dt + \sigma^T \mathbb{E}[Y_t | \mathcal{F}_t^{\text{FB}}] dW_t,$$

where $\mathcal{F}_t^{\text{FB}} = \mathcal{F}_t^{R^{\text{FB}}}$, i.e., the volatility is adapted to the observation filtration generated by the returns. This is now a coupling between the volatility and the observations as in [Haussmann and Sass \(2004\)](#) and it guarantees that applying the iteration (3.1) does not change the model, since then

$$\tilde{\sigma}_t = \sigma^T \mathbb{E}[Y_t | \mathcal{F}_t^{\text{FB}}].$$

With this choice we now have an approximation of the MSM that is in this sense consistent and minimizes the MSE over all other volatility processes adapted to \mathcal{F}^{FB} . This gives us a model that truly lies between the HMM and MSM. The distance to the MSM is given by

$$\mathbb{E} \left[\left| R_t^{\text{MSM}} - R_t^{\text{FB}} \right|^2 \right] = \int_0^t \mathbb{E} \left[(\sigma^T Y_s - \sigma^T \mathbb{E}[Y_s | \mathcal{F}_s^{\text{FB}}])^2 \right] ds.$$

This result has deep implications for the volatility choice. In the formulation of the model we started with an arbitrary observable diffusion and only assumed the usual conditions on the functions in the diffusion to ensure the existence of a strong solution. Only by looking at the filtering equation we see that we can e.g., choose the unnormalized filter ρ as this diffusion. At this point the motivation for this is only economic intuition since one typically interprets the volatility as arising because of traders whose actions have to be adapted to the available information. As we have pointed out before, the filter collects this information into the best estimate of the true state of the economy.

The iterative method for approximating the MSM now implies that using the filter as diffusion ξ is not only an economically reasonable choice but the choice for approximating the MSM where the iteration stops. And furthermore, it also answers the next question after choosing the diffusion, that is how to choose the function σ that connects the diffusion to the volatility process. It suggests that the observable volatility process that leads to the best approximation of the MSM is the function with which the volatility in the MSM depends on the chain now applied to the best estimate of the chain, that is, the filter \hat{Y} .

The function $\tilde{\sigma}$ is the scalar product of the vector σ with the normalized filter. As such it has the nice property that it is linear in \hat{Y} . That does not seem very surprising looking at the simple volatility structure of the MSM we want to approximate, but considering the structure of the HMM in [Haussmann and Sass \(2004\)](#) this is a very useful result. This HMM in general allows for much more complicated volatility functions, but [Theorem 3.2](#) tells us that already the simple scalar product of the filter with the volatility vector from the MSM is the best approximation.

Now, after pointing out these advantageous practical implications, we consider the theoretical intricacies arising in this model for the volatility: the volatility in the HMM from [Haussmann and Sass \(2004\)](#) depends on a \tilde{W} -diffusion, and the filtering equation implies that the unnormalized filter is such a diffusion. [Theorem 3.2](#) suggests the volatility as a function of the *normalized* filter, so the model actually reads

$$dR_t = \mu^T Y_t dt + \sigma^T \frac{\rho_t}{1^T \rho_t} dW_t.$$

But the function $\sigma^T \frac{x}{1^T x}$ does not have bounded derivatives w.r.t the x_i , so the Zakai equation proven in [Hausmann and Sass \(2004\)](#) cannot be applied. The model is still well-defined, since for x with strictly positive entries the function $\sigma^T \frac{x}{1^T x}$ is not only bounded, but also bounded away from 0:

$$0 < \min_i \sigma_i \leq \sigma^T \frac{x}{1^T x} \leq \max_i \sigma_i$$

as σ has only positive entries. This implies that Novikov's condition is fulfilled and we can do the usual measure change leading to the unnormalized filter ρ_t . With the existence of the unnormalized filter ensured we see that the model with $\sigma_t = \sigma^T \frac{\rho_t}{1^T \rho_t}$ is indeed well-defined. This leads to the introduction of a new model:

Definition 3.5. Let the return process R have the dynamics

$$dR_t = \mu^T Y_t dt + \sigma^T \hat{Y}_t dW_t,$$

where Y is an irreducible, time-homogenous Markov chain with state space $\{e_1, \dots, e_d\}$, $\mu \in \mathbb{R}^d$, $\sigma \in \mathbb{R}_{>0}^d$ and $\hat{Y}_t = \mathbb{E}[Y_t | \mathcal{F}_t^R]$. This model is then called the Filterbased-Volatility Hidden Markov Model (FB-HMM).

As we have already pointed out, at first sight, this looks like a special case of a HMM since

$$dR_t = \mu^T Y_t dt + \sigma^T \frac{\rho_t}{1^T \rho_t} dW_t.$$

However, as we cannot apply [Theorem 3.1](#) and it is not clear whether ρ_t is truly a \tilde{W} -diffusion. One of the major advantages of regime-switching models is that their filters are actually given by a finite system of equations. So to make our model tractable we have to derive the dynamics of the filter, which we can be done following the methods from [Elliott \(1993\)](#).

Theorem 3.6. In the Filterbased Volatility Model, the unnormalized filter ρ satisfies

$$d\rho_t = Q^T \rho_t dt + \sigma_t^{-2} \text{diag}(\mu) \rho_t dR_t.$$

Proof. The function $\langle \sigma, \frac{\rho_t}{1^T \rho_t} \rangle$ is bounded from above in ρ_t and also bounded away from 0. Using these properties the claim can be proven using similar methods as in [Elliott \(1993\)](#). \square

From [Theorem 3.6](#) it also follows that for the case of $\mu \equiv 0$ the FB-HMM is an HMM with volatility $\sigma^T \nu$. This is due to the fact that for vanishing μ the diffusion part in the filtering equation also vanishes, and due to $\rho_0 = \nu$ the first part does not contribute either. Thus, the volatility stays constant. If we instead use the MSM-filter for the volatility, which corresponds to one step in the iteration procedure outlined in [Section 3.2](#), then we obtain a model that for $\mu \equiv 0$ still lies between the HMM and MSM. But this is not a model from the class of FB-HMM.

3.4. Stylized facts

Real data often exhibit certain characteristics, known as *stylized facts*, see e.g., [Cont \(2001\)](#); [Ryden et al. \(1998\)](#). This behavior is common across a wide range of instruments, markets and time periods. Thus an approach for assessing the quality of a model is to see which stylized facts can be reproduced by a return process coming from the model. The main motivation for the introduction of the FB-HMM were the better econometric properties of the MSM due to its switching volatility. The MSM can e.g., capture volatility clustering, which is the observation that high volatility events tend to cluster in time, followed by phases of calmer movement of the prices.

Volatility clustering. In the FB-HMM the volatility process is controlled by the filter. So any clustering of high volatility events must come from time periods where the filter puts a comparably large probability weight on states where the corresponding entry in σ is higher than the others. The filter is an estimate of the true state of Y , so if Y is in the state with large entry in σ and the filter performs well, then there will be a phase of high volatility. Volatility clustering in the filterbased model is thus connected to not only the values in σ directly, as it is in the MSM via the Markov chain, but also to the performance of the filter.

Leverage effect. The so-called leverage effect concerns an asymmetric relation between the volatility and the returns of a financial asset. In particular, a statistical leverage effect is the fact that the correlation between past returns and future volatility is mostly observed to be negative, which is in accord with one's economic intuition. In the MSM, one can derive a heuristic for choosing the parameters that leads to a negative value of the approximated covariance. Due to the tractability of the transition probabilities in the 2-state model, one can then prove that in this case the heuristic leads to a negative covariance, see [Leoff \(2016\)](#). Since the FB-HMM approximates the MSM, it is intuitive to apply the same heuristic also here. But the coupled structure makes it impossible to analytically derive the covariance even for $d = 2$. Numerical examples still indicate a negative covariance, but just as for volatility clustering, the extent to what this stylized fact can be introduced depends directly on the performance of the filter.

Asymmetry. The next stylized fact which we consider is the frequently observed asymmetry of the return distribution. Large drawdowns in prices are more often observed than equally large upward movements. This asymmetry arises naturally e.g., from the existence of thresholds below which positions must be cut unconditionally for regulatory reasons. Declining stock prices are more likely to give rise to massive portfolio rebalancing (and thus volatility) than increasing stock prices, which connects the asymmetry of return distributions to the leverage effect.

Cont noted in (Cont, 2001, Section 4.1) that one must use more than only the standard deviation to capture the variability of the returns due to the non-Gaussian nature of most of the empirical data. One popular approach is to consider higher moments as a measure of dispersion. These moments are very involved in the continuous-time MSM due to the unbounded number of possible jumps in every finite interval. Instead we can again employ a heuristic choice that uses the different entries in σ to push more of the probability mass to the left tail of the distribution. This choice can be made in a way that is consistent with the heuristic for the leverage effect. We can apply the same choice for the FB-HMM and underline with numerical examples in Section 5 that these parameters then indeed lead to an asymmetry in the histograms.

Autocorrelation function. The autocorrelation functions (acf) are used to describe and quantify the long-term dependence between increments of the return processes. Their properties are similar over a wide range of empirical data and thus for model-building it is of importance to know their characteristics. In particular, one is interested in the linear autocorrelations and the autocorrelations of the absolute value of the returns.

In Frühwirth-Schnatter (2006) the autocorrelations of the discrete-time HMM and MSM are presented following the work of Timmermann (2000). In Leoff (2016), we applied similar methods for the continuous-time case. For the FB-HMM we have the following lemma:

Lemma 3.7. *If R follows the FB-HMM, then for $0 \leq t < s \leq T$ with $t + \Delta_t < s$*

$$\begin{aligned} \text{Cov}(\Delta R_t, \Delta R_s) &= \int_s^{s+\Delta_t} \int_t^{t+\Delta_t} \sum_{i,j=1}^d \mu_i \mu_j (e^{Q(r-u)})_{ij} v_i v_j du dr \\ &\quad + \mathbb{E} \left[\int_t^{t+\Delta_t} \sigma^T \hat{Y}_r dW_r \int_s^{s+\Delta_t} \mu^T Y_r dr \right] - \Delta_t^2 (\mu^T v)^2. \end{aligned}$$

Proof. Evaluating the covariance and using $\mathbb{E}[\hat{Y}_t] = \mathbb{E}[Y_t] = v$ implies

$$\begin{aligned} \text{Cov}(\Delta R_t, \Delta R_s) &= \mathbb{E} \left[\left(\int_t^{t+\Delta_t} \mu^T Y_r dr + \int_t^{t+\Delta_t} \sigma^T \hat{Y}_r dW_r - \Delta_t \mu^T v \right) \right. \\ &\quad \times \left. \left(\int_s^{s+\Delta_t} \mu^T Y_r dr + \int_s^{s+\Delta_t} \sigma^T \hat{Y}_r dW_r - \Delta_t \mu^T v \right) \right] \\ &= \mathbb{E} \left[\int_t^{t+\Delta_t} \mu^T Y_r dr \int_s^{s+\Delta_t} \mu^T Y_r dr \right] \\ &\quad + \mathbb{E} \left[\int_t^{t+\Delta_t} \mu^T Y_r dr \int_s^{s+\Delta_t} \sigma^T \hat{Y}_r dW_r + \int_t^{t+\Delta_t} \sigma^T \hat{Y}_r dW_r \int_s^{s+\Delta_t} \mu^T Y_r dr \right] \\ &\quad + \mathbb{E} \left[\int_t^{t+\Delta_t} \sigma^T \hat{Y}_r dW_r \int_s^{s+\Delta_t} \sigma^T \hat{Y}_r dW_r \right] - (\Delta_t \mu^T v)^2. \end{aligned}$$

The first summand depends on Y which has only finitely many states, thus we can reformulate it as

$$\mathbb{E} \left[\int_t^{t+\Delta_t} \mu^T Y_r dr \int_s^{s+\Delta_t} \mu^T Y_r dr \right] = \int_s^{s+\Delta_t} \int_t^{t+\Delta_t} \sum_{i,j=1}^d \mu_i \mu_j (e^{Q(r-u)})_{ij} v_i v_j du dr.$$

For the last expectation we apply Itô's isometry to obtain

$$\begin{aligned} \mathbb{E} \left[\int_t^{t+\Delta_t} \sigma^T \hat{Y}_r dW_r \int_s^{s+\Delta_t} \sigma^T \hat{Y}_r dW_r \right] &= \mathbb{E} \left[\int_0^{s+\Delta_t} \mathbb{1}_{[t,t+\Delta_t]}(r) \sigma^T \hat{Y}_r \mathbb{1}_{[s,s+\Delta_t]} \sigma^T \hat{Y}_r dr \right] \\ &= 0. \end{aligned}$$

Now it is left to show that $\mathbb{E} \left[\int_t^{t+\Delta_t} \mu^T Y_r dr \int_s^{s+\Delta_t} \sigma^T \hat{Y}_r dW_r \right] = 0$. Observe that since \hat{Y} is bounded $\int_0^u \sigma^T \hat{Y}_r dW_r$ is a true martingale w.r.t. $\mathcal{F}_u^{Y,W}$, so

$$\begin{aligned} \mathbb{E} \left[\int_s^{s+\Delta_t} \sigma^T \hat{Y}_r dW_r | \mathcal{F}_s^{Y,W} \right] &= \mathbb{E} \left[\int_0^{s+\Delta_t} \sigma^T \hat{Y}_r dW_r | \mathcal{F}_s^{Y,W} \right] - \mathbb{E} \left[\int_0^s \sigma^T \hat{Y}_r dW_r | \mathcal{F}_s^{Y,W} \right] \\ &= \int_0^s \sigma^T \hat{Y}_r dW_r - \int_0^s \sigma^T \hat{Y}_r dW_r \\ &= 0 \end{aligned}$$

and since $t + \Delta_t < s$

$$\begin{aligned} \mathbb{E}\left[\int_t^{t+\Delta_t} \mu^T Y_r dr \int_s^{s+\Delta_t} \sigma^T \hat{Y}_r dW_r\right] &= \mathbb{E}\left[\mathbb{E}\left[\int_t^{t+\Delta_t} \mu^T Y_r dr \int_s^{s+\Delta_t} \sigma^T \hat{Y}_r dW_r \mid \mathcal{F}_s^{Y,W}\right]\right] \\ &= \mathbb{E}\left[\int_t^{t+\Delta_t} \mu^T Y_r dr \mathbb{E}\left[\int_s^{s+\Delta_t} \sigma^T \hat{Y}_r dW_r \mid \mathcal{F}_s^{Y,W}\right]\right] \\ &= 0. \end{aligned}$$

□

This representation of the linear autocovariances is useful for numerical estimation. In Leoff (2016), we also estimated the autocorrelations of the absolute value, the sign and the squares of the return increments. A plot of the results is presented in Section 5. We observed that the autocorrelations seem to converge to 0, but slower than in both the MSM and HMM. For the linear acf this is probably due to the additional summand $\mathbb{E}[\int_t^{t+\Delta_t} \sigma^T \hat{Y}_r dW_r \int_s^{s+\Delta_t} \mu^T Y_r dr]$, and we suspect a similar effect for the other acf. The autocorrelations of the squares were significantly positive in the simulations, which underlines the fact that there is volatility clustering in the FB-HMM.

3.5. Comparison with stochastic volatility models

The volatility in the FB-HMM depends on the filter which is driven by the stock returns, so there is no additional source of randomness apart from the Markov chain and the underlying one-dimensional Brownian motion W . Under the reference measure only the Brownian motion \tilde{W} is driving the unnormalized filter.

In a strict sense, we speak of *stochastic volatility* if the volatility depends on some additional independent stochastic process. A generic stochastic volatility model can be introduced as follows. Instead of an observable factor process $d\xi_t = \nu_t(\xi_t)dt + \tau(\xi_t)d\tilde{W}_t$ as we used it in Section 3 to motivate the FB-HMM, we can define a factor process ζ for a second Brownian motion W^ζ , which is independent of W , as follows:

$$dR_t = \mu^T Y_t dt + \sigma^R(\zeta_t)dW_t$$

and

$$d\zeta_t = (\mu^\zeta)^T Y_t dt + \sigma^\zeta(\xi_t)dW_t^\zeta + \tau^\zeta(\zeta_t)dW_t.$$

Here R_t is observable but ζ_t may not. The observability of ζ_t can be guaranteed by assuming that the map

$$z \mapsto (\sigma^R(z))^2 \text{ is one-to-one,}$$

cf. Assumption 5.1 in Sass (2007). Then ζ can be observed via the quadratic variation of R and we can write the observation process as

$$d\bar{R}_t = B Y_t dt + \Sigma_t d\bar{W}_t,$$

where

$$\bar{R}_t = \begin{pmatrix} R_t \\ \zeta_t \end{pmatrix}, \quad \bar{W}_t = \begin{pmatrix} W_t \\ W_t^\zeta \end{pmatrix}, \quad \Sigma_t = \begin{pmatrix} \sigma^R(\zeta_t) & 0 \\ \sigma^\zeta(\zeta_t) & \tau^\zeta(\zeta_t) \end{pmatrix}$$

and B is the matrix with rows μ and μ^ζ . Under suitable conditions on σ^R , σ^ζ , τ^ζ and the observability assumption above we then are in a standard multivariate HMM setting and obtain the unnormalized filter

$$d\rho_t = Q^T \rho_t dt + \text{diag}(\rho_t) B^T (\Sigma_t \Sigma_t^T)^{-1} d\bar{R}_t.$$

While filtering under the observability assumption thus works straightforwardly, the second term in the representation of ζ , the stochastic integral with respect to the new Brownian motion W^ζ , leads to an incomplete financial market. This means that not all future payoffs can be replicated.

E.g., for portfolio optimization a solution is still known to exist, if the value of the problem is finite, but an explicit form can only be obtained via a dual problem and there is no guarantee that the minimizer in the dual problem is attained. For a case in which a solution can be obtained we refer to (Sass, 2007, Section 5). Also pricing becomes more difficult since there may exist no replication strategy and there is more than one arbitrage-free price.

These difficulties serve as motivation for models such as the FB-HMM which still lead to a complete financial market model in which all observable payoffs can be replicated. This allows for more explicit results for several problems in finance, e.g., for portfolio optimization, see Remark 4.3 below. On the other hand it satisfies much more of the stylized facts than the simple HMM with constant volatility, see Sections 3.4 and 5.

Note that the FB-HMM can be extended by adding a term depending on a new Brownian motion to the dynamics of the FB-volatility. While filtering would still work, this would yield an incomplete market as outlined above. Which model is to be preferred depends on the specific setting and the problems under consideration.

4. Consistency of the discretized model

To explore the relationship between discrete- and continuous-time models, we will now examine a notion of consistency between the discretized and the continuous-time returns in our new model.

The FB-HMM in discrete time can be defined in a similar way as a discrete-time HMM as

$$R_k = \mu^T Y_{k-1} + \sigma^T \hat{Y}_{k-1} w_k, \quad k = 0, 1, \dots,$$

with the w_k i.i.d. $\mathcal{N}(0, 1)$ -distributed.

If we want to obtain a model that is purely discrete time, \hat{Y}_{k-1} has to be the estimate of Y_{k-1} at time $k-1$ using only the discrete-time returns, i.e., $\hat{Y}_{k-1} = \mathbb{E}[Y_{k-1} | \mathcal{F}_{k-1}^R]$.

Instead of this discrete-time model we can also discretize the returns R in the continuous-time model directly as

$$\bar{R}_k = \mu^T Y_{k-1} \Delta t + \sigma^T \hat{Y}_{k-1} (W_{t_k} - W_{t_{k-1}}), \quad (4.2)$$

where we now use the discretized continuous-time filter $\hat{Y}_{k-1} = \mathbb{E}[Y_{t_{k-1}} | \mathcal{F}_{t_{k-1}}^R]$. In the HMM and MSM these discretizations would be the same: in the HMM we would use $\sigma^T v$ for both R_k and \bar{R}_k , and in the MSM we would directly use the chain Y at time t_{k-1} . This kind of consistency is not given for the FB-HMM.

Discretizing the continuous-time returns directly does not lead to a time series that follows a discrete-time FB-HMM. Instead we consider the consistency in the returns themselves by investigating the behavior of the discretized returns for a discretization step converging to 0. In the following, we will prove that if we aggregate the \bar{R}_k over time, the resulting sum does indeed converge to the continuous-time process.

Let R follow a continuous-time FB-HMM with parameters μ , σ and Q

$$dR_t = \mu^T Y_t dt + \sigma^T \hat{Y}_t dW_t.$$

For $n \in \mathbb{N}$ consider the discretization $0 = t_0^n, t_1^n, \dots, t_n^n = T$ of the time interval $[0, T]$ given by

$$\Delta t^n = \frac{T}{n}, \quad t_k^n = k \Delta t^n,$$

and define the discretization of R as

$$R_k^n = R_{k-1}^n + \mu^T Y_{t_{k-1}^n} \Delta t^n + \sigma^T \hat{Y}_{k-1}^n \Delta W_k^n, \quad R_0^n = 0,$$

where \hat{Y}_k^n, Y_k^n and ΔW_k^n are now the discretizations of \hat{Y}, Y and $W_{t_k} - W_{t_{k-1}}$ respectively, i.e.,

$$Y_k^n = Y_{t_k^n}, \quad \hat{Y}_k^n = \hat{Y}_{t_k^n} \quad \text{and} \quad \Delta W_k^n = W_{t_k^n} - W_{t_{k-1}^n}.$$

This discretization R^n is just the cumulative sum of the usual discretizations. We use the sum here since we want to approximate a continuous-time process that is given by an integral. This kind of method is usually known as the (explicit) *Euler scheme*. It has the advantage that it is recursive and thus easy to implement.

To quantify the global discretization error we introduce the piecewise constant approximations of Y, \hat{Y} on this grid as

$$Y_t^n = Y_{t_k^n} \quad \text{and} \quad \hat{Y}_t^n = \hat{Y}_{t_k^n}$$

for $t \in [t_k^n, t_{k+1}^n)$. Then the difference between the discretization R^n and the continuous-time process at the final time is given by

$$\varepsilon^n := R_T - R_n^n = \int_0^T \mu^T Y_t dt + \int_0^T \sigma^T \hat{Y}_t - \sum_{k=1}^n \mu^T Y_{k-1}^n \Delta t^n + \sigma^T \hat{Y}_{k-1}^n \Delta W_k^n.$$

Using Y^n and \hat{Y}^n this can be reformulated as

$$\varepsilon^n = \int_0^T \mu^T (Y_t - Y_t^n) dt + \int_0^T \sigma^T (\hat{Y}_t - \hat{Y}_t^n) dW_t.$$

We now show that this global error actually converges to zero if the grid becomes finer. This means that already the simple Euler discretization of the continuous-time returns is convergent and there is no need for higher-order methods to achieve consistency.

Theorem 4.1. *For arbitrarily fine grids, the global discretization error at T tends to 0 in L^2 , that is*

$$\mathbb{E}[\|\varepsilon^n\|^2] \rightarrow 0, \quad (n \rightarrow \infty).$$

Proof. First we gather some observations that we will use in the proof:

(i) For $a, b \in \mathbb{R}$

$$(a + b)^2 \leq (|a| + |b|)^2 \quad \text{and} \quad (|a| + |b|)^2 \leq 4(a^2 + b^2).$$

(ii) For any real valued function f

$$\left(\int_0^t f(s)ds \right)^2 \leq \left(\int_0^t |f(s)|ds \right)^2.$$

(iii) For $N \in \mathbb{N}$ and arbitrary $a_k \in \mathbb{R}$ Jensen's inequality implies

$$\left(\frac{1}{N} \sum_{k=1}^N a_k \right)^2 \leq \frac{1}{N} \sum_{k=1}^N a_k^2.$$

Now we proceed with proving the theorem: the squared L^2 -norm of the error ε^n can be estimated using the reformulation $\varepsilon^n = \int_0^T \mu^T(Y_t - Y_t^n)dt + \int_0^T \sigma^T(\hat{Y}_t - \hat{Y}_t^n)dW_t$ as

$$\begin{aligned} \|\varepsilon^n\|_{L^2}^2 &= \mathbb{E} \left[\left(\int_0^T \mu^T(Y_t - Y_t^n)dt + \int_0^T \sigma^T(\hat{Y}_t - \hat{Y}_t^n)dW_t \right)^2 \right] \\ &\stackrel{(i)}{\leq} \mathbb{E} \left[\left(\left| \int_0^T \mu^T(Y_t - Y_t^n)dt \right| + \left| \int_0^T \sigma^T(\hat{Y}_t - \hat{Y}_t^n)dW_t \right| \right)^2 \right] \\ &\stackrel{(ii)}{\leq} 4\mathbb{E} \left[\left(\int_0^T \mu^T(Y_t - Y_t^n)dt \right)^2 \right] + 4\mathbb{E} \left[\left(\int_0^T \sigma^T(\hat{Y}_t - \hat{Y}_t^n)dW_t \right)^2 \right]. \end{aligned}$$

Thus it suffices to show that both of the summands converge to 0.

First, consider that

$$\mathbb{E} \left[\left(\int_0^T \mu^T(Y_t - Y_t^n)dt \right)^2 \right] \stackrel{(ii)}{\leq} \mathbb{E} \left[\left(\int_0^T |\mu^T(Y_t - Y_t^n)dt| \right)^2 \right]$$

and by Cauchy-Schwartz

$$\begin{aligned} \int_0^T |\mu^T(Y_t - Y_t^n)dt| &= \sum_{k=1}^n \mathbb{1}_{\{Y \text{ jumps on } [t_{k-1}^n, t_k^n]\}} \int_{t_{k-1}^n}^{t_k^n} |\mu^T(Y_t - Y_t^n)|dt \\ &\leq 2\|\mu\| \Delta_t^n \sum_{k=1}^n \mathbb{1}_{\{Y \text{ jumps on } [t_{k-1}^n, t_k^n]\}}. \end{aligned}$$

So

$$\begin{aligned} \mathbb{E} \left[\left(\int_0^T \mu^T(Y_t - Y_t^n)dt \right)^2 \right] &\leq \mathbb{E} \left[\left(2\|\mu\| \Delta_t^n \sum_{k=1}^n \mathbb{1}_{\{Y \text{ jumps on } [t_{k-1}^n, t_k^n]\}} \right)^2 \right] \\ &= 4\|\mu\|^2 (\Delta_t^n)^2 \mathbb{E} \left[\left(\sum_{k=1}^n \mathbb{1}_{\{Y \text{ jumps on } [t_{k-1}^n, t_k^n]\}} \right)^2 \right] \\ &\stackrel{(iii)}{\leq} 4\|\mu\|^2 (\Delta_t^n)^2 n^2 \mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{Y \text{ jumps on } [t_{k-1}^n, t_k^n]\}} \right] \\ &= 4\|\mu\|^2 \frac{T^2}{n} \sum_{k=1}^n \mathbb{P}(Y \text{ jumps on } [t_{k-1}^n, t_k^n]) \\ &= 4\|\mu\|^2 T^2 \mathbb{P}(Y \text{ jumps on } [0, t_1^n]), \end{aligned}$$

and since $t_1^n \rightarrow 0$ for $n \rightarrow \infty$, $\mathbb{P}(Y \text{ jumps on } [0, t_1^n]) \rightarrow 0$.

For the second summand we have with Itô and Cauchy-Schwartz

$$\begin{aligned} \mathbb{E} \left[\left(\int_0^T \sigma^T(\hat{Y}_t - \hat{Y}_t^n)dW_t \right)^2 \right] &= \mathbb{E} \left[\int_0^T (\sigma^T(\hat{Y}_t - \hat{Y}_t^n))^2 dt \right] \\ &\leq \mathbb{E} \left[\int_0^T \|\sigma\|^2 \|\hat{Y}_t - \hat{Y}_t^n\|^2 dt \right] \\ &\leq \|\sigma\|^2 \mathbb{E} \left[\sum_{k=1}^n \Delta_t^n \sup_{s \in [t_{k-1}^n, t_k^n]} \|\hat{Y}_t - \hat{Y}_t^n\|^2 \right] \end{aligned}$$

$$\begin{aligned} &\leq \|\sigma\|^2 \Delta_t^n n \mathbb{E} \left[\sup_{s \in [0, T]} \|\hat{Y}_t - \hat{Y}_t^n\|^2 \right] \\ &= \|\sigma\|^2 T \mathbb{E} \left[\sup_{s \in [0, T]} \|\hat{Y}_t - \hat{Y}_t^n\|^2 \right]. \end{aligned}$$

This expectation converges to 0, because $\sup_{s \in [0, T]} \|\hat{Y}_s - \hat{Y}_s^n\|^2$ converges to 0 almost surely. \square

Theorem 4.1 shows that the discretized returns from an FB-HMM are consistent with the continuous-time returns in the sense that in the limit where $\Delta_t \rightarrow 0$ the Euler discretizations converge in L^2 to the continuous-time returns. This is particularly beneficial since the explicit Euler scheme is very easy to implement.

Remark 4.2. It is important to note that **Theorem 4.1** does not deal with consistency of the filters. As we have mentioned earlier, the discretization R^n depends on the discretized filter \hat{Y}_t^n , not on the discrete-time filter, so R^n does not follow a truly discrete-time model.

The Euler discretization of the HMM is $\bar{R}_k^{\text{HMM}} = \mu^T Y_{t_{k-1}} \Delta_t + \sigma_0 (W_{t_k} - W_{t_{k-1}})$. Comparing this to (4.2) we see the discretization of the HMM indeed follows a discrete-time HMM.

Remark 4.3 (Conclusion). Following [Hausmann and Sass \(2004\)](#), optimal trading strategies π^* can be computed in the FB-HMM and would for logarithmic utility $U_0(x) = \log(x)$ be of the form that in [Section 2.2](#) we just replace σ_0 by $\sigma^T \hat{Y}_t$ in π_t^{HMM} or equivalently Y_t by \hat{Y}_t in π_t^{MSM} . This yields in both cases

$$\pi_t^* = (\sigma^T \hat{Y}_t)^{-2} \mu^T \hat{Y}_t$$

which again illustrates that the FB-HMM lies between HMM and MSM. An extension to power utility is possible. The consistency shown in [Theorem 4.1](#) can then be used to show that optimization results in the continuous-time FB-HMM are close to the results obtained in the discretized model (for fractions within $[0, 1]$) providing the problem-specific discretization consistency for the FB-HMM. Thus, while having on one hand better econometric properties than the HMM by approximating the MSM, it has on the other hand better computational and consistency properties than the MSM.

5. Numerical examples

HMM and MSM sample paths

In [Fig. 1](#), we give an example of the returns of MSM and HMM, depending on the same Markov chain and Brownian motion. One can clearly see the effect of the jumping volatility in the MSM and the resulting volatility clustering.

Filtering in HMM and MSM

In [Proposition 2.1](#), we have proven that the HMM and MSM behave differently during the transition to continuous time. This difference can already be seen by considering discretizations with very fine grids. To illustrate the different behavior for increasing grid size we consider an HMM and MSM depending on the same Y and W , discretized with $n = 10,000$ grid points over $T = 1$ year. The example paths are presented in [Fig. 2](#) for the MSM and [Fig. 3](#) for the HMM. The difference even at this pathwise level is staggering. The performance of the MSM-filter improves significantly, while the HMM-filter does not. Looking at the returns we see intuitively why this is the case: with so many grid points we can strongly see the effect of the changing volatility in the MSM. This allows us to guess the state of the Markov chain quite well for a lot of grid points just by considering the amplitudes of the returns. Recalling the proof of [Proposition 2.1](#), where we proved the observability of the signal in the continuous-time MSM via the quadratic variation, we can see the implications of this theoretical connection now even at the pathwise level. We can also observe the problems of trying to exploit this theoretical feature in practice: the chain can be estimated well only for an unrealistically large number of grid points. So this effect is not as relevant for e.g., daily data, which for $T = 1$ leads to only 250 data points. But it is important and has a large impact for investigating the transition to continuous time.

Problems for the filter and the observer arise when, for example, quite extreme values for the normal distribution are simulated. These values imply a larger volatility than is really present and thus adversely affect the filter. But with a high probability these outliers do not persist over a longer time, since the Brownian increments are independent of each other. Thus over more time points the filter returns to a better estimate of the drift and does not really believe in a jump in the underlying chain.

In the HMM we do not have the help of a changing volatility. Looking at the returns and the performance of the filter we can see this clearly.

Filterbased HMM

Due to the structure of the volatility in the FB-HMM it is necessary to compute a discretization of the filter \hat{Y}_k for simulating discretized returns R_k . This is again different to the HMM or MSM, where the path of the filter is not necessary for simulating the returns.

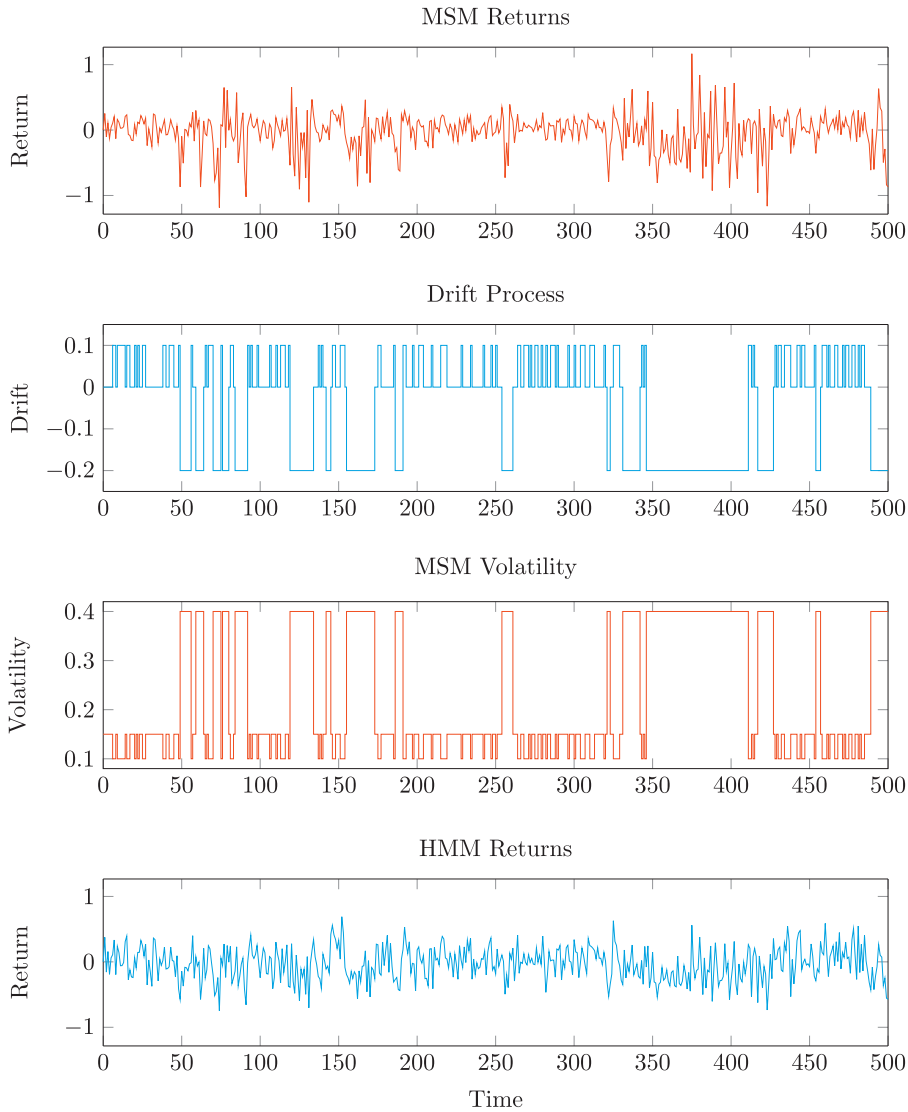


Fig. 1. Example of an HMM and MSM return series depending on the same Markov chain.

We start with

$$R_0 = 0, \quad \rho_0 = \nu, \quad \tilde{\sigma}_0 = \sigma^T \nu$$

and then simulate the next return increment as

$$R_1 = \mu^T Y_{t_0} \Delta_t + \tilde{\sigma}_0 \sqrt{\Delta_t} w_1$$

with $w_1 \sim \mathcal{N}(0, 1)$. We see that for simulating R_2 we need an estimate of $\hat{Y}_1 = \frac{\rho_1}{1^T \rho_1}$ to update the volatility process to its current value $\tilde{\sigma}_1 = \sigma^T \hat{Y}_1$.

The numerical method is thus as follows: initialize $R_0, \rho_0, \tilde{\sigma}_0$ as above.

If we assume that we have simulated the returns and the filter up to time $k - 1$ we now need a way to simulate R_k and calculate the filter. We discretize the return at time t_k as

$$R_k = \mu^T Y_{t_{k-1}} \Delta_t + \tilde{\sigma}_{k-1} \sqrt{\Delta_t} w_k,$$

where $\tilde{\sigma}_{k-1} = \sigma^T \hat{Y}_{k-1} = \frac{\rho_{k-1}}{1^T \rho_{k-1}}$.

Making use of the recursive nature of the filtering equation we can update the filter in the usual recursive way: to use the robust discretization of the filtering equation introduced in James et al. (1996) we first set

$$\tilde{\phi}_k^i := \exp \left(\tilde{\sigma}_{k-1}^{-2} \mu_i R_k - \frac{1}{2} \tilde{\sigma}_{k-1}^{-2} \mu_i^2 \Delta_t \right)$$

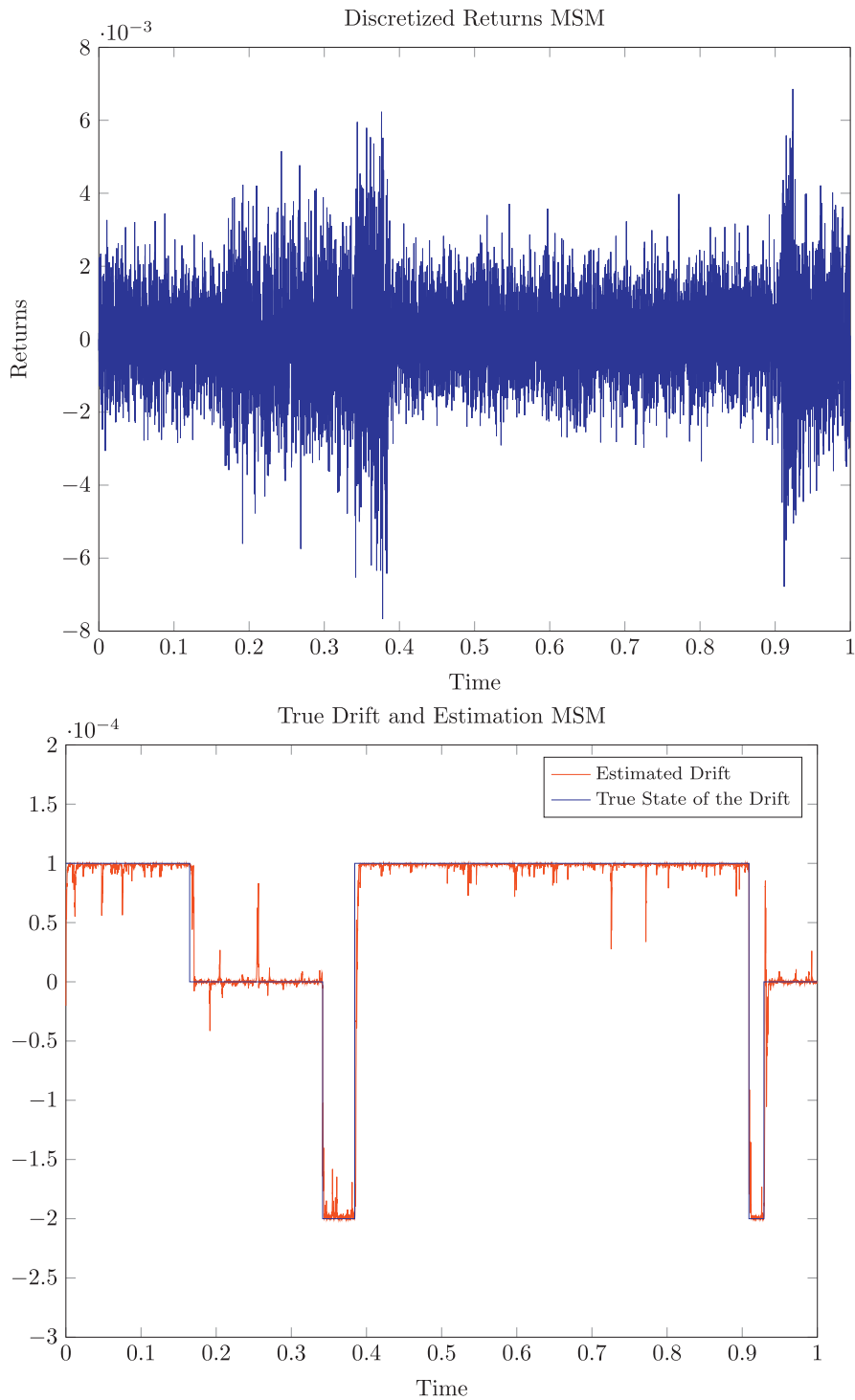


Fig. 2. Example path MSM with 10,000 grid points.

and

$$\tilde{\psi}_k = \text{diag}(\tilde{\phi}_k^1, \dots, \tilde{\phi}_k^d).$$

Then the next value of the discretized filter is given by

$$\rho_k = \tilde{\psi}_k (I + \Delta_t Q^T) \rho_{k-1}$$

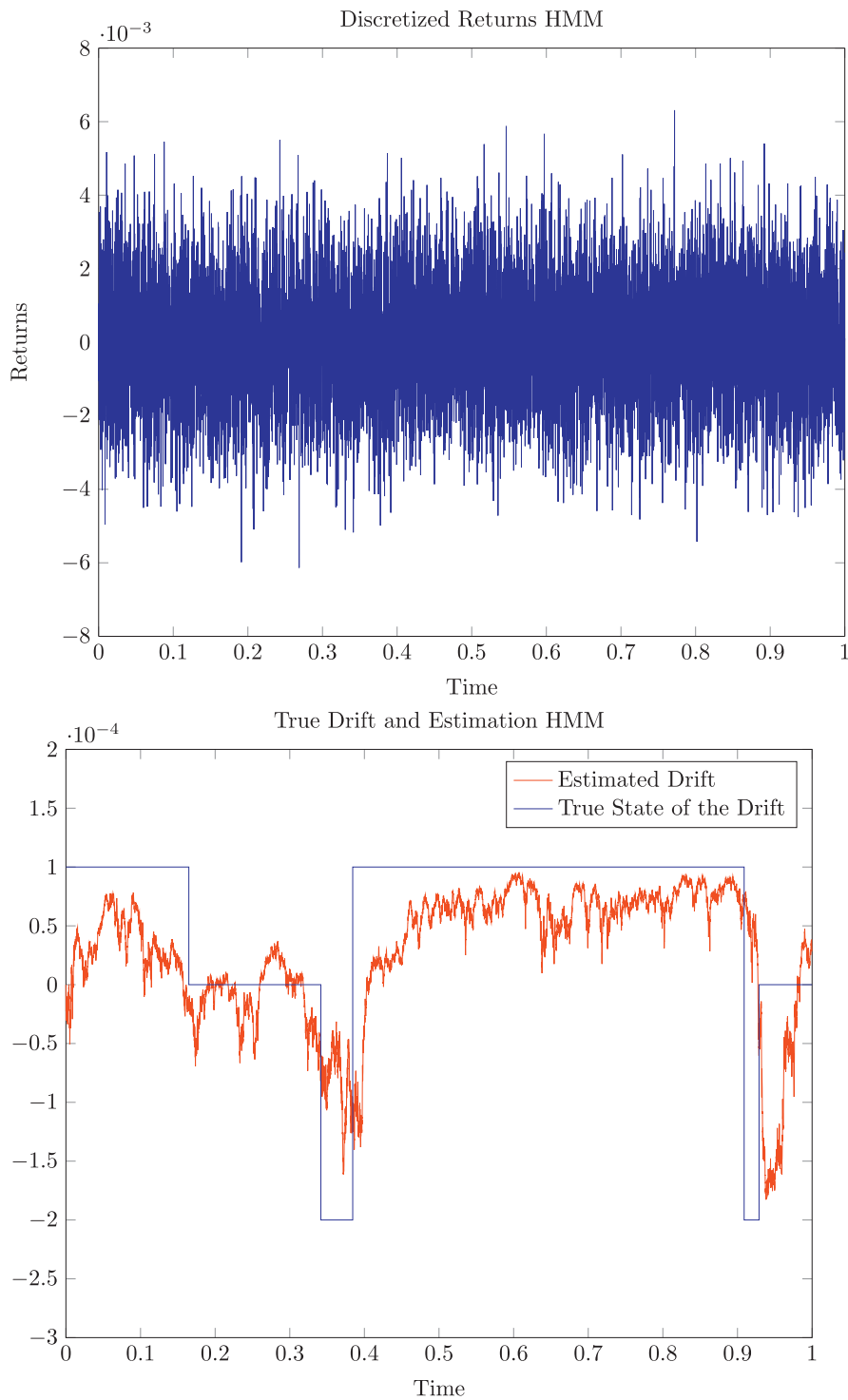


Fig. 3. Example path HMM with 10,000 grid points.

and the volatility is updated as

$$\tilde{\sigma}_k = \sigma^T \frac{\rho_k}{\mathbf{1}^T \rho_k}.$$

With this numerical scheme we can simulate the discretized returns.

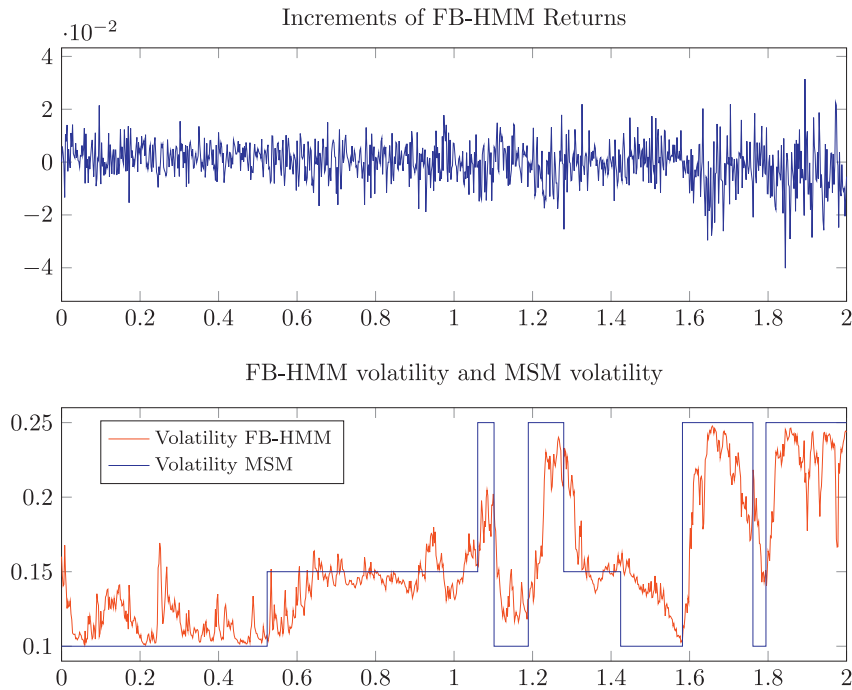


Fig. 4. Return increments and volatility in the FB-HMM.

Some stylized facts in the FB-HMM

Fig. 4 displays a simulated path of the FB-HMM and the volatility processes both of the FB-HMM and the corresponding MSM. The parameters chosen were

$$Q = \begin{pmatrix} -7 & 4 & 3 \\ 2 & -4 & 2 \\ 3 & 5 & -8 \end{pmatrix}, \quad \mu = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} \quad \text{and} \quad \sigma = \begin{pmatrix} 0.10 \\ 0.15 \\ 0.25 \end{pmatrix}.$$

Both in the volatility process and the return increments it can be seen that the returns exhibit volatility clustering.

In Fig. 5, we simulated increments of the FB-HMM over $T = 4$ years with 250 discretization points each. We can clearly see that it is asymmetric. The left tail has more probability mass than the right tail. Q was chosen as before and the parameters for drift and volatility were

$$\mu = \begin{pmatrix} 0.5 \\ 0 \\ -1 \end{pmatrix} \quad \text{and} \quad \sigma = \begin{pmatrix} 0.05 \\ 0.08 \\ 0.12 \end{pmatrix}.$$

Choosing the largest volatility for the state with negative drift leads to this more pronounced left tail.

In Fig. 6 we see the estimated acf in the FB-HMM. They converge to 0, but slower than in simulations for the HMM and MSM with the same parameters. The acf of the squared return increments is positive, which is a measure for the volatility clustering in the FB-HMM. Comparing the estimates for the HMM and MSM to the FB-HMM, we have another interesting observation: in the MSM the linear acf has the same order of magnitude as the absolute acf, while in the HMM the linear acf is much larger than the absolute acf. In the FB-HMM we observed that as in the HMM the linear acf is larger than the absolute acf, but the effect is much less pronounced than in the original HMM. This again supports the intuition with which we introduced the FB-HMM: in a sense the FB-HMM lies between the MSM and the HMM.

6. Extending the model: jump diffusions

A possible extension to make the model more realistic is to include jumps in the observation process. To illustrate the ideas, we consider a stylized model where these jumps come from a single Poisson process. This can be generalized in a straightforward manner to sums of Poisson processes, see Remark 3.4 below, and to compound Poisson processes where the distribution of the jumps has no support around 0. For estimation in a regime switching Lévy model cf. Chevallier and Goutte (2016).

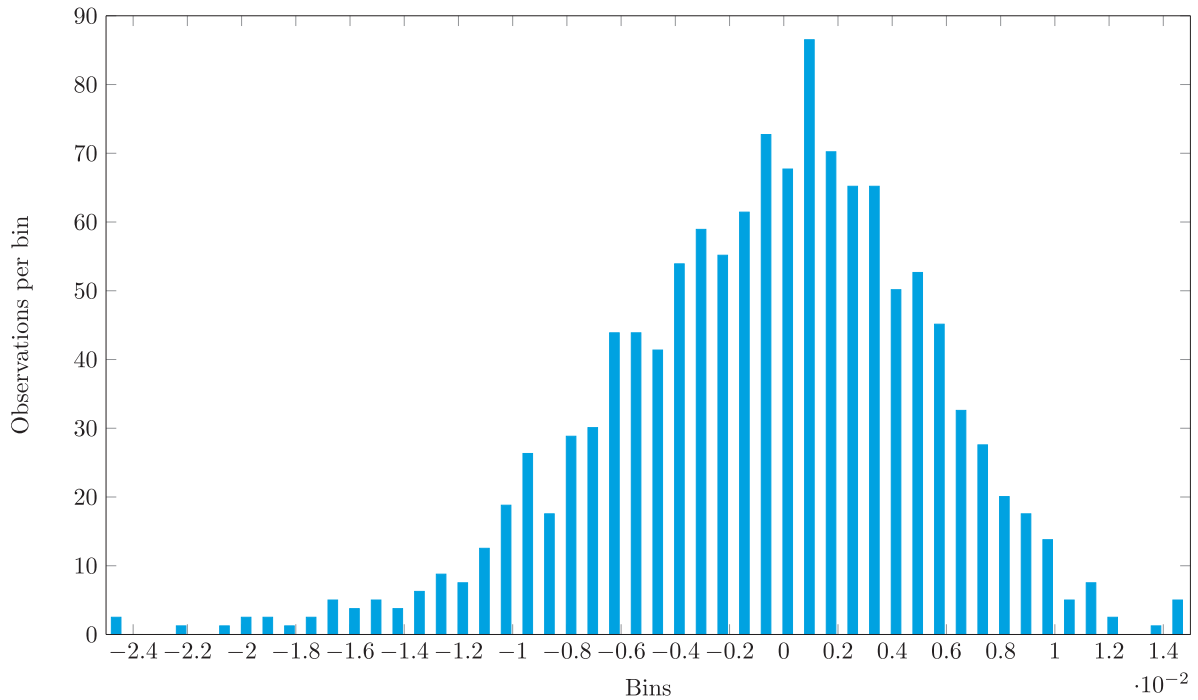


Fig. 5. Histogram of increments in the FB-HMM.

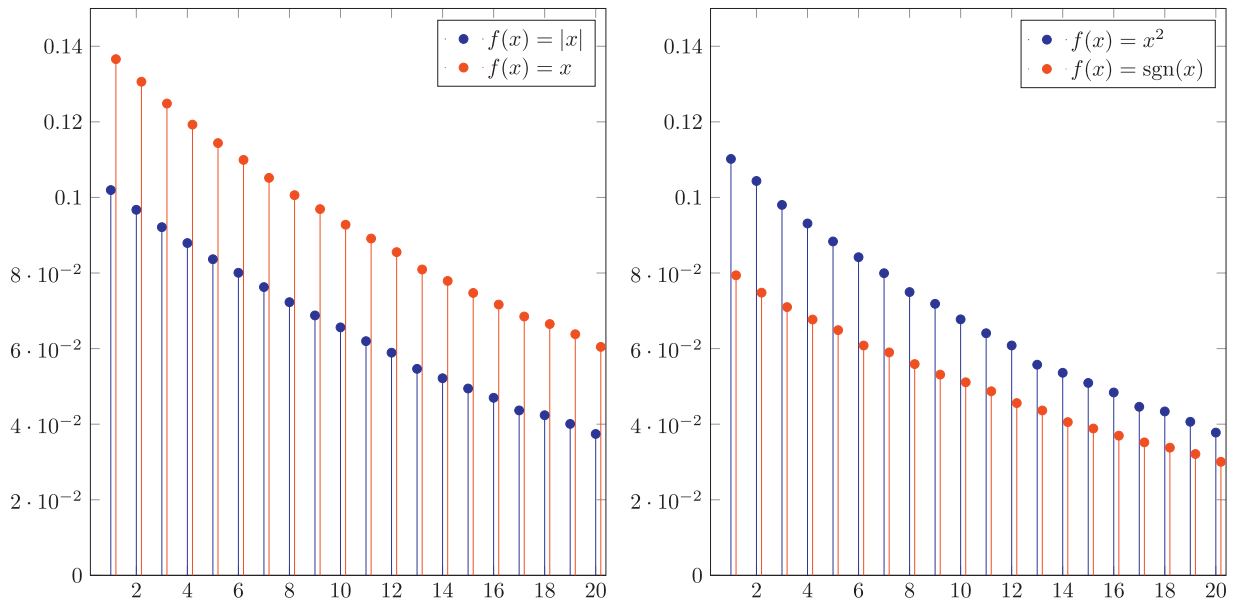


Fig. 6. Autocorrelations of increments in the FB-HMM.

Model with Poisson jumps

We introduce in the dynamics of the FB-HMM a compound Poisson process, whose rate depends on the Markov chain. The return process then reads

$$dR_t^{FB} = \mu^T Y_t dt + \sigma^T \mathbb{E}[Y_t | \mathcal{F}_t^{FB}] dW_t + \gamma dM_t,$$

where M is a compensated Poisson process independent of W . Its intensity process $\lambda = (\lambda_t)_{t \in [0, T]}$ depends on the Markov chain via the d -dimensional vector $\lambda \in \mathbb{R}_{>0}^d$ which has strictly positive entries, i.e.,

$$\lambda_t = \lambda^T Y_t > 0.$$

We denote with N the corresponding Poisson process with intensities λ_t , so

$$M_t = N_t - \int_0^t \lambda^T Y_s ds.$$

The jump size $\gamma > -1$, $\gamma \neq 0$ is assumed to be constant.

Filtering

Note that the filter $\mathbb{E}[Y_t | \mathcal{F}_t^{\text{FB}}]$ and thus also the volatility process $\sigma_t = \sigma^T \mathbb{E}[Y_t | \mathcal{F}_t^{\text{FB}}]$ is still well-defined, as it is still simply a conditional expectation of an integrable random variable, now with respect to the observation filtration generated by an R^{FB} that is a jump diffusion.

With this return process we can do a similar measure change as in the original model, where under the new measure $\tilde{\mathbb{P}}$ we now also have that $\tilde{M}_t = N_t - t$ is a standard compensated Poisson process. Due to the jump component in R^{FB} the new Brownian motion is

$$d\tilde{W}_t = dW_t + \sigma_t^{-1} (\mu^T Y_t - \gamma (\lambda^T Y_t - 1)) dt$$

and the density $Z_T = \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}$ fullfils

$$dZ_t = -Z_t \sigma_t^{-1} (\mu^T Y_t - \gamma (\lambda^T Y_t - 1)) dW_t - Z_t \frac{\lambda^T Y_t - 1}{\lambda^T Y_t} dM_t.$$

For the unnormalized filter $\rho_t = \tilde{\mathbb{E}}[Z_t^{-1} Y_t | \mathcal{F}_t^R]$ one can prove the following theorem.

Theorem 6.1. *In the Filterbased Volatility Model with jumps the unnormalized filter ρ satisfies*

$$d\rho_t = Q^T \rho_t dt + \sigma_t^{-2} \text{diag}(\mu - \gamma(\lambda - 1)) \rho_t d\tilde{W}_t + \text{diag}(\lambda - 1) \rho_t d\tilde{M}_t.$$

Proof. In Wu and Elliott (2005), the filtering equations for an HMM with a jump diffusion were proven with a small typo in the formulation. The proof works analogously to Theorem 3.1 by using the product rule for jump processes. \square

Remark 6.2. Due to the Poisson process as an additional source of randomness, the financial market with jump diffusion dynamics is incomplete, cf. Section 3.5. This implies that there is more than one reference measure. These may lead to different unnormalized filters, but after normalizing by the Kallianpur–Striebel formula they all lead to the same filter. For portfolio optimization we face similar difficulties as in Section 3.5 due to the incompleteness: we have to minimize in a dual approach over all these reference measures. For maximizing expected logarithmic utility of terminal wealth, this leads for Poisson jumps to solving two quadratic equations specifying the optimal strategy and the minimizing reference measure in the dual problem. If these equations, which depend on the filter, have an admissible solution, an optimal strategy can be given in terms of these.

Convergence

The convergence of Theorem 4.1 also holds in the model with jumps due to the constant jump size. For defining the Euler discretization in this model one has to make a distinction between the formulation with M and with N . Using the notation from Section 4, we define the first discretization as

$$R_k^n = R_{k-1}^n + \mu^T Y_{t_{k-1}}^n \Delta_t^n + \sigma^T \hat{Y}_{k-1}^n \Delta W_k^n + \gamma \Delta M_k^n,$$

with $\Delta M_k^n = M_{t_k}^n - M_{t_{k-1}}^n$. The global discretization error can be calculated as

$$\begin{aligned} R_T - R_n^n &= \int_0^T \mu^T Y_t dt + \int_0^T \sigma^T \hat{Y}_t dW_t + \gamma M_T - \sum_{k=1}^n \left(\mu^T Y_{t_{k-1}}^n \Delta_t^n + \sigma^T \hat{Y}_{k-1}^n \Delta W_k^n + \gamma \Delta M_k^n \right) \\ &= \varepsilon^n + \gamma M_T - \sum_{k=1}^n \gamma \Delta M_k^n \\ &= \varepsilon^n + \gamma \left(M_T - \sum_{k=1}^n M_{t_k}^n - M_{t_{k-1}}^n \right) \\ &= \varepsilon^n + \gamma (M_T - M_{t_n}^n + M_0) \\ &= \varepsilon^n, \end{aligned}$$

so due to the constant jump size this is the same expression as in Theorem 4.1. Thus we have the following corollary:

Corollary 6.3. *For grid size converging to 0 the global discretization error at T tends to 0 in L^2 , that is*

$$\mathbb{E}[\|R_T - R_n^n\|^2] \rightarrow 0, \quad (n \rightarrow \infty).$$

If we discretize R in the form $dR_t = \mu^T Y_t dt + \sigma^T \mathbb{E}[Y_t | \mathcal{F}_t^{\text{FB}}] dW_t + \gamma (dN_t - \lambda^T Y_t dt)$ the discretization error reads

$$\begin{aligned} R_T - \tilde{R}_n^n &= \int_0^T \mu^T (Y_t - Y_t^n) dt + \int_0^T \sigma^T (\hat{Y}_t - \hat{Y}_t^n) dW_t + \gamma \int_0^T \lambda^T (Y_t - Y_t^n) dt + \gamma \left(N_T - \sum_{k=1}^n N_{t_k^n} - N_{t_{k-1}^n} \right) \\ &= \int_0^T (\mu + \gamma \lambda)^T (Y_t - Y_t^n) dt + \int_0^T \sigma^T (\hat{Y}_t - \hat{Y}_t^n) dW_t + \gamma \left(N_T - \sum_{k=1}^n N_{t_k^n} - N_{t_{k-1}^n} \right) \\ &= \int_0^T (\mu + \gamma \lambda)^T (Y_t - Y_t^n) dt + \int_0^T \sigma^T (\hat{Y}_t - \hat{Y}_t^n) dW_t. \end{aligned}$$

So with the same arguments as in [Theorem 4.1](#), now for $\mu + \gamma \lambda$ instead of only μ we obtain:

Corollary 6.4. For grid size converging to 0 the global discretization error of the discretization \tilde{R} at T vanishes in L^2 , i.e.,

$$\mathbb{E}[\|R_T - \tilde{R}_n^n\|^2] \rightarrow 0, \quad (n \rightarrow \infty).$$

Remark 6.5. It is also possible to consider finitely many Poisson processes M^i that each have a constant jump size γ_i

$$dR_t^{\text{FB}} = \mu^T Y_t dt + \sigma^T \mathbb{E}[Y_t | \mathcal{F}_t^{\text{FB}}] dW_t + \sum_{i=1}^m \gamma_i dM_t^i.$$

The filtering techniques and the discretization results then still hold. Thus the assumption of constant jump size is not as restrictive as it seems.

References

- Bäuerle, N., Rieder, U., 2004. Portfolio optimization with Markov-modulated stock prices and interest rates. *IEEE Trans. Automatic Control* 29 (3), 442–447.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *J. Econom.* 31 (3), 307–327.
- Chamley, C., 2004. *Rational Herds: Economic Models of Social Learning*. Cambridge University Press, Cambridge.
- Chevallier, J., Goutte, S., 2016. On the estimation of regime-switching Lévy models. *Stud. Nonlin. Dyn. Econom. doi:10.1515/snde-2016-0048*.
- Cont, R., 2001. Empirical properties of asset returns: stylized facts and statistical issues. *Quant. Finance* 1, 223–236.
- Elliott, R., 1993. New finite-dimensional filters and smoothers for noisily observed Markov chains. *IEEE Trans. Inf. Theory* 39 (1), 265–271.
- Elliott, R., Aggoun, L., Moore, J., 1995. *Hidden Markov Models. Estimation and Control*. Springer-Verlag, New York.
- Elliott, R., Krishnamurthy, V., Sass, J., 2008. Moment based regression algorithms for drift and volatility estimation in continuous-time Markov switching models. *Econom. J.* 11, 244–270.
- Elliott, R., Siu, T., 2012. An HMM approach for optimal investment of an insurer. *Int. J. Robust Nonlin. Control* 22 (7), 778–807.
- Engle, R., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50 (4), 987–1007.
- Fama, E., 1965. The behavior of stock market prices. *J. Bus.* 38 (1), 34–105.
- Frühwirth-Schnatter, S., 2006. *Finite mixture and Markov switching models*. Springer Series in Statistics. Springer, New York.
- Guo, X., 2001. Information and option pricings. *Quant. Finance* 1 (1), 38–44.
- Hahn, M., Frühwirth-Schnatter, S., Sass, J., 2010. Markov chain Monte Carlo methods for parameter estimation in multidimensional continuous time Markov switching models. *J. Financial Econom.* 8, 88–121.
- Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57 (2), 357–384.
- Haussmann, U., Sass, J., 2004. Optimal terminal wealth under partial information for HMM stock returns. In: Yin, G., Zhang, Q. (Eds.), *Mathematics of Finance: Proceedings of an AMS-IMS-SIAM Summer Conference, June 22–26, 2003*, vol. 351. Utah, AMS Contemporary Mathematics, pp. 171–185.
- Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financial Stud.* 6 (2), 327–343.
- James, M.R., Krishnamurthy, V., Gland, F.L., 1996. Time discretization of continuous-time filters and smoothers for HMM parameter estimation. *IEEE Trans. Inform. Theory* 42, 593–605.
- Krishnamurthy, V., 2016. *Partially Observed Markov Decision Processes. From Filtering to Controlled Sensing*. Cambridge University Press, Cambridge.
- Krishnamurthy, V., Bhatt, S., 2016. Sequential detection of market shocks with risk-averse CVaR social sensors. *IEEE J. Sel. Top. Signal Process* 10 (6), 1061–1072.
- Krishnamurthy, V., Gharehshiran, O., Hamdi, M., 2014. *Interactive sensing and decision making in social networks*. Foundations and Trends in Signal Processing. Now Publishers, Boston.
- Leoff, E., 2016. *Stochastic Filtering in Regime-Switching Models: Econometric Properties, Discretization and Convergence*. TU Kaiserslautern (Ph.D. thesis).
- Lipster, R., Shiryaev, A., 2001. *Statistics of Random Processes: II. Applications*. Springer-Verlag, Berlin.
- Ryden, T., Teräsvirta, T., Asbrink, S., 1998. Stylized facts of daily return series and the hidden Markov model. *J. Appl. Econom.* 13, 217–244.
- Sass, J., 2007. Utility maximization with convex constraints and partial information. *Acta Applicandae Mathematicae* 97, 221–238.
- Sass, J., Haussmann, U., 2004. Optimizing the terminal wealth under partial information: the drift process as a continuous time Markov chain. *Finance Stoch.* 8 (4), 553–577.
- Siu, T., 2011. Long-term strategic asset allocation with inflation risk and regime switching. *Quant. Finance* 11 (10), 1565–1580.
- Taksar, M., Zeng, X., 2007. Optimal terminal wealth under partial information: both the drift and the volatility driven by a discrete-time Markov chain. *SIAM J. Optim. Control* 46 (4), 1461–1482.
- Taylor, S., 1994. *Modelling Financial Time Series*. Wiley, Chichester.
- Timmermann, A., 2000. Moments of Markov switching models. *J. Econom.* 96 (1), 75–111.
- Wonham, W.M., 1965. Some applications of stochastic differential equations to optimal nonlinear filtering. *J. Soc. Ind. Appl. Math. Ser. A Control* 2 (3), 347–369.
- Wu, P., Elliott, R., 2005. Hidden Markov chain filtering for a jump diffusion model. *Stoch. Anal. Appl.* 23 (1), 153–163.