DTU Library

Stylised facts of financial time series and hidden Markov models in continuous time

Nystrup, Peter; Madsen, Henrik; Lindström, Erik

Link back to DTU Orbit

Stylised facts of financial time series and hidden Markov models in continuous time

Peter Nystrup*†, Henrik Madsen†, and Erik Lindström†‡

†Department of Applied Mathematics and Computer Science, Technical University of
Denmark, Richard Petersens Plads, Building 324, 2800 Kgs. Lyngby, Denmark
‡Centre for Mathematical Sciences, Lund University, Box 118, SE-221 00 Lund, Sweden

Hidden Markov models are often applied in quantitative finance to capture the stylised facts of financial returns. They are usually discrete-time models and the number of states rarely exceeds two because of the quadratic increase in the number of parameters with the number of states. This paper presents an extension to continuous time where it is possible to increase the number of states with a linear rather than quadratic growth in the number of parameters. The possibility of increasing the number of states leads to a better fit to both the distributional and temporal properties of daily returns.

Keywords: Hidden Markov models; Continuous time; Daily returns; Leptokurtosis; Volatility clustering; Long memory

JEL Classification: C01—Econometrics, C16—Specific Distributions, C22—Time-series models, C52—Model evaluation and testing

1. Introduction

The normal distribution is well-known as being a poor fit to most financial returns. Mixtures of normal distributions provide a much better fit as they are able to reproduce both the skewness and leptokurtosis often observed (Cont 2001). Markov switching mixture models, also referred to as hidden Markov models (HMMs), are a natural extension in order to also capture the temporal properties of financial returns. In an HMM, the distribution that generates an observation depends on the state of an underlying and unobserved Markov chain.

The ability of an HMM to reproduce most of the stylised facts of daily return series introduced by Granger and Ding (1995a,b) was illustrated by Rydén *et al.* (1998). They found that the one stylised fact that cannot be reproduced by an HMM is the slow decay of the autocorrelation function (ACF) of squared daily returns, which is of great importance, for instance, in financial risk management.

According to Bulla and Bulla (2006), the lack of flexibility of an HMM to model this temporal higher order dependence can be explained by the implicit assumption of geometrically distributed sojourn times in the hidden states. Silvestrov and Stenberg (2004), among others, argued that the memoryless property of the geometric distribution is inadequate from an empirical perspective, although it is consistent with the no-arbitrage principle.

Bulla and Bulla (2006) considered hidden semi-Markov models (HSMMs) in which the sojourn time distribution is modelled explicitly for each hidden state so that the Markov property is

---

*Corresponding author. Email: pnys@dtu.dk

transferred to the embedded first-order Markov chain. They showed that HSMMs with negative binomial sojourn time distributions are able to reproduce most of the stylised facts comparably well, and often better, than the HMM. Specifically, they found HSMMs to reproduce the long-memory property of squared daily returns much better than HMMs. However, they did not consider the complicated problem of selecting the most appropriate sojourn time distributions and following the approach by Rydén *et al.* (1998) they only considered models with two hidden states.

Bulla (2011) later showed that HMMs with *t*-distributed components reproduce most of the stylised facts as well or better than the Gaussian HMM, at the same time as increasing the persistence of the visited states and the robustness to outliers. Bulla (2011) also found that models with three states provided a better fit than models with two states.

Many different stylised facts have been established for financial returns. See, for example, Granger and Ding (1995a,b), Granger *et al.* (2000), Cont (2001), Malmsten and Teräsvirta (2004). This paper focuses on the stylised facts relating to the long memory of the ACF and examines the importance of the number of hidden states on the ability to fit the slowly decaying ACF of squared daily returns. An extension of HMMs to continuous time is presented as a flexible alternative to the discrete-time models.

Two hidden states are found to be too few to reproduce the slowly decaying ACF as well as the observed skewness and leptokurtosis. A major limitation of discrete-time HMMs and HSMMs is the quadratic increase in the number of parameters with the number of states. This limitation does not apply to HMMs in continuous time as it can reasonably be assumed that the only possible transitions in an infinitesimally short time interval are to the neighboring states.

This assumption leads to a linear rather than quadratic growth in the number of parameters with the number of states and consequently a significant reduction in the number of parameters for higher order models. With the added flexibility, the number of states can be considered a parameter that needs to be estimated.[1] In addition, it is possible to incorporate temporal inhomogeneity without a dramatic increase in the number of parameters using a continuous-time formulation.

Section 2 gives an introduction to the main theory relating to HMMs and HSMMs. Section 3 introduces HMMs where the underlying Markov chain is a continuous-time Markov chain. Section 4 contains a description of the data used. The empirical results are reported in section 5 and section 6 concludes. All parameter estimates can be found in appendix A. The results of the analysis of the FTSE 100 index can be found in appendix B.

2.   Hidden Markov models in discrete time

In hidden Markov models, the probability distribution that generates an observation depends on the state of an underlying and unobserved Markov process. HMMs are a particular kind of dependent mixture and are therefore also referred to as Markov switching mixture models.

A sequence of discrete random variables $\{S_t : t \in \mathrm{N}\}$ is said to be a Markov chain if, for all $t \in \mathrm{N}$, it satisfies the Markov property:

$$\mathrm{Pr}\left(S_{t+1}|\, S_t, \ldots, S_1\right) = \mathrm{Pr}\left(S_{t+1}|\, S_t\right). \tag{1}$$

The conditional probabilities $\mathrm{Pr}\left(S_{u+t} = j |\, S_u = i\right) = \gamma_{ij}\left(t\right)$ are called transition probabilities. The Markov chain is said to be homogeneous if the transition probabilities are independent of u, otherwise inhomogeneous.

A Markov chain with transition probability matrix $\mathbf{\Gamma}\left(t\right) = \{\gamma_{ij}\left(t\right)\}$ has stationary distribution $\boldsymbol{\pi}$ if $\boldsymbol{\pi}\mathbf{\Gamma} = \boldsymbol{\pi}$ and $\boldsymbol{\pi}\mathbf{1} = 1$. The Markov chain is termed stationary if $\boldsymbol{\pi} = \boldsymbol{\delta}$, where $\boldsymbol{\delta}$ is the initial

---

[1]See Cappé *et al.* (2005) for a perspective on order estimation.

distribution, that is $\delta_i = \Pr\left(S_1 = i\right)$.

If the Markov chain $\{S_t\}$ has m states, then $\{X_t : t \in \mathrm{N}\}$ is called an m-state HMM. With $\boldsymbol{X}^{(t)}$ and $\boldsymbol{S}^{(t)}$ representing the sequence of values from time 1 to time t, the simplest model of this kind can be summarised by

$$\Pr\left(\left.S_t\right| \boldsymbol{S}^{(t-1)}\right) = \Pr\left(\left.S_t\right| S_{t-1}\right), \;\; t = 2, 3, \ldots, \tag{2a}$$

$$\Pr\left(\left.X_t\right| \boldsymbol{X}^{(t-1)}, \boldsymbol{S}^{(t)}\right) = \Pr\left(\left.X_t\right| S_t\right), \quad t \in \mathrm{N}. \tag{2b}$$

When the current state S_t is known, the distribution of X_t depends only on S_t. This causes the autocorrelation of X_t to be strongly dependent on the persistence of S_t.

A specific observation can usually arise from more than one state as the support of the conditional distributions overlaps. The unobserved state process $\{S_t\}$ is, therefore, not directly observable through the observation process $\{X_t\}$, but can only be estimated.

As an example, consider the two-state model with Gaussian conditional distributions:

$$X_t = \mu_{S_t} + \varepsilon_{S_t}, \;\; \varepsilon_{S_t} \sim N\left(0, \sigma_{S_t}^2\right),$$

where

$$\mu_{S_t} = \begin{cases} \mu_1, & \text{if } S_t = 1, \\ \mu_2, & \text{if } S_t = 2, \end{cases} \quad \sigma_{S_t}^2 = \begin{cases} \sigma_1^2, & \text{if } S_t = 1, \\ \sigma_2^2, & \text{if } S_t = 2, \end{cases} \quad \text{and } \boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} & 1 - \gamma_{11} \\ 1 - \gamma_{22} & \gamma_{22} \end{bmatrix}.$$

For this model, the value of the autocorrelation function at lag k is

$$\rho_{X_t}\left(\left.k\right| \theta\right) = \frac{\pi_1\left(1 - \pi_1\right)\left(\mu_1 - \mu_2\right)^2}{\sigma^2}\lambda^k$$

and the autocorrelation function for the squared process is

$$\rho_{X_t^2}\left(\left.k\right| \theta\right) = \frac{\pi_1\left(1 - \pi_1\right)\left(\mu_1^2 - \mu_2^2 + \sigma_1^2 - \sigma_2^2\right)^2}{\mathrm{E}\left[\left.X_t^4\right| \theta\right] - \mathrm{E}\left[\left.X_t^2\right| \theta\right]^2}\lambda^k,$$

when θ denotes the model parameters, $\sigma^2 = \mathrm{Var}\left[\left.X_t\right| \theta\right]$ is the unconditional variance, and $\lambda = \gamma_{11} + \gamma_{22} - 1$ is the second largest eigenvalue of $\boldsymbol{\Gamma}$ (Frühwirth-Schnatter 2006). It is evident from these expressions, as noted by Rydén *et al.* (1998), that HMMs can only reproduce an exponentially decaying autocorrelation structure.

The ACF of the first-order process becomes zero if the means are equal, whereas persistence in the squared process can be induced either by a difference in the mean values, as for a mixed effects model, or by a difference in the variances across the states. In both cases, the persistence increases with the combined persistence of the states as measured by λ (Ang and Timmermann 2011).

The parameters of an HMM are typically estimated using the Maximum Likelihood method. Under the assumption that successive observations are independent, the likelihood is given by

$$L_T\left(\theta\right) = \Pr\left(\left.\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}\right| \theta\right) = \boldsymbol{\delta}\mathbf{P}\left(x_1\right)\boldsymbol{\Gamma}\mathbf{P}\left(x_2\right)\cdots\boldsymbol{\Gamma}\mathbf{P}\left(x_T\right)\mathbf{1}, \tag{3}$$

where $\mathbf{P}\left(x\right)$ is a diagonal matrix with the state-dependent conditional densities $p_i\left(x\right) = \Pr\left(\left.X_t = x\right| S_t = i\right)$, $i \in \{1, 2, \ldots, m\}$, as entries. The conditional distribution of X_t may be

either discrete or continuous, univariate or multivariate. In mixtures of continuous distributions, the likelihood can be unbounded in the vicinity of certain parameter combinations.[1]

The likelihood function of an HMM is in general a complicated function of the parameters with several local maxima. The two most popular approaches to maximising the likelihood are direct numerical maximisation and the Baum–Welch algorithm, a special case of the Expectation Maximisation (EM) algorithm (Cappé *et al.* 2005, Zucchini and MacDonald 2009). All discrete-time models are estimated using the R-package `hsmm` due to Bulla *et al.* (2010) that implements the EM algorithm in the version presented by Guédon (2003).

In HMMs, the sojourn times are implicitly assumed to be geometrically distributed:

$$\Pr\left(\text{'staying } t \text{ time steps in state } i\text{'}\right) = \gamma_{ii}^{t-1}\left(1 - \gamma_{ii}\right). \tag{4}$$

The geometric distribution is memoryless, implying that the time until the next transition out of the current state is independent of the time spent in the state.

2.1 *Hidden semi-Markov models*

If the assumption of geometrically distributed sojourn times is unsuitable, then hidden semi-Markov models can be applied. HMMs and HSMMs differ only in the way that the state process is defined. In HSMMs, the sojourn time distribution is modelled explicitly for each state i:

$$d_i\left(u\right) = \Pr\left(S_{t+u+1} \neq i,\, S_{t+u-v} = i,\, v = 0, \ldots, u-2 \middle|\, S_{t+1} = i,\, S_t \neq i\right) \tag{5}$$

and the transition probabilities are defined as

$$\gamma_{ij} = \Pr\left(S_{t+1} = j \middle|\, S_{t+1} \neq i,\, S_t = i\right) \tag{6}$$

for each $j \neq i$ with $\gamma_{ii} = 0$ and $\sum_j \gamma_{ij} = 1$.

The conditional independence assumption for the observation process is similar to a simple HMM, but the semi-Markov chain associated with HSMMs does not have the Markov property at each time t. This property is transferred to the embedded first-order Markov chain, that is the sequence of visited states (Bulla and Bulla 2006). In other words, the future states are only conditionally independent of the past states when the process changes state.

3.   Hidden Markov models in continuous time

In continuous-time Markov chains, transitions may occur at all times rather than at discrete and equidistant time points. There is no smallest time step and the quantities of interest become the transition probabilities

$$p_{ij}\left(\Delta t\right) = \Pr\left(S\left(t + \Delta t\right) = j \middle|\, S\left(t\right) = i\right) \tag{7}$$

as $\Delta t \to 0$. Clearly, $p_{ij}\left(0\right) = 0$ for different states i and j, and it can be shown that under certain regularity conditions

$$\lim_{t \to 0} \mathbf{P}\left(t\right) = \mathbf{I}. \tag{8}$$

---

[1]If the conditional distribution is normal, then the likelihood can be made arbitrarily large by setting the mean equal to one of the observations and letting the conditional variance tend to zero (Frühwirth-Schnatter 2006).

Assuming that $p_{ij}(\Delta t)$ is differentiable at 0, the transition rates are defined as

$$p'_{ij}(0) = \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t) - p_{ij}(0)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Pr\left(S\left(t + \Delta t\right) = j \mid S\left(t\right) = i\right)}{\Delta t} = q_{ij} \tag{9}$$

with the additional definition $q_{ii} = q_i = -\sum_{j \neq i} q_{ij}$. The transition intensity matrix $\mathbf{Q} = \{q_{ij}\}$ has non-negative off-diagonal elements q_{ij}, non-positive diagonal entries q_i, and all rows sum to zero.

The stationary distribution $\boldsymbol{\pi}$, if it exists, is found by solving the system of equations

$$\begin{cases} \boldsymbol{\pi}\mathbf{Q} = \mathbf{0} \\ \boldsymbol{\pi}\mathbf{1} = 1. \end{cases} \tag{10}$$

If it has a strictly positive solution (all elements in $\boldsymbol{\pi}$ are strictly positive), then the stationary distribution exists and is independent of the initial distribution.

The matrix of transition probabilities $\mathbf{P}(t) = \{p_{ij}(t)\}$ can be found as the solution to Kolmogorov's differential equation

$$\frac{\mathrm{d}\mathbf{P}(t)}{\mathrm{d}t} = \mathbf{P}(t)\mathbf{Q} \tag{11}$$

with the initial condition $\mathbf{P}(0) = \mathbf{I}$. The solution being

$$\mathbf{P}(t) = e^{\mathbf{Q}t}\mathbf{P}(0) = e^{\mathbf{Q}t}. \tag{12}$$

When the process enters state i, it remains there according to an exponential distribution with parameter $-q_i > 0$ before it instantly jumps to another state $j \neq i$ with probability $-q_{ij}/q_i$. A continuous-time Markov chain is fully characterised by its initial distribution $\boldsymbol{\delta}$ and the transition intensity matrix $\mathbf{Q}$.

It follows that the transition intensity matrix $\mathbf{Q}$ can in principle be found by taking the logarithm of the one-step transition probability matrix

$$\mathbf{P}(t) = e^{\mathbf{Q}t} \Rightarrow \mathbf{Q} = \log\mathbf{P}(1). \tag{13}$$

Taking the logarithm of a matrix that has a lot of elements that are close to zero or zero is not a trivial operation. Instead, an intuitive estimate of the transition rate q_i can be based on the discrete transition probability γ_{ii} as

$$\tilde{q}_i = -\log\hat{\gamma}_{ii}. \tag{14}$$

This estimate does not take into account that the process might change from a given state and back within the sampling interval. Thus, the simple estimate will underestimate q_i, but the error will be small for $q_i \ll 1$ (Madsen *et al.* 1985).

The exponential distribution is memoryless just like its discrete analogue, the geometric distribution. By introducing dummy states that are indistinguishable from one or more of the original states, it is possible to allow for non-exponentially distributed sojourn times (see e.g. Madsen *et al.* 1985, Iversen *et al.* 2013). The distribution of sojourn times will then be a mixture of exponential distributions, which is a phase-type distribution, and the Markov property will be transferred to the embedded Markov chain as for the HSMM. Phase-type distributions can be used to approximate any positive-valued distribution with arbitrary precision (for details, see Nielsen 2013). Similarly, Langrock and Zucchini (2011) showed how discrete-time HMMs can be structured to fit any sojourn time distribution with arbitrary precision.

It is often convenient to assume that in a short time interval Δt, the only possible transitions are to the neighboring states:

$$\left.\begin{array}{r}p_{ij} = o\left(\Delta t\right), \;\; |i - j| \geq 2, \\ p_{ii}\left(\Delta t\right) = 1 - q_i \Delta t + o\left(\Delta t\right), \\ p_{i,i-1}\left(\Delta t\right) = w_i q_i \Delta t + o\left(\Delta t\right), \\ p_{i,i+1}\left(\Delta t\right) = \left(1 - w_i\right) q_i \Delta t + o\left(\Delta t\right), \\ i \in \{1, 2, \ldots, m\},\end{array}\right\} \tag{15}$$

where $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$. The notation includes transitions from state 1 to m and reverse with the definition that state 0 = state m and state $(m+1)$ = state 1.

It should be noted that, even though, the process cannot go straight from state i to state $i+2$ without going through state $i+1$, there is no limit to how fast a transition from state i to state $i+2$ can occur.

Under this assumption, the matrix of transition intensities has the structure

$$\mathbf{Q} = \begin{bmatrix} -q_1 & (1-w_1)\,q_1 & 0 & \cdots & 0 & w_1 q_1 \\ w_2 q_2 & -q_2 & (1-w_2)\,q_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ (1-w_m)\,q_m & 0 & 0 & \cdots & w_m q_m & q_m \end{bmatrix}. \tag{16}$$

The number of parameters increases linearly with the number of states. Thus, a continuous-time Markov chain yields a parameter reduction over its discrete-time analogue if the number of states exceeds three. The higher the number of states, the larger the reduction. In addition, it is possible to incorporate inhomogeneity without a dramatic increase in the number of parameters using splines, harmonic functions, or similar.[1]

Another advantage of a continuous-time formulation is the flexibility to use data with any sampling interval as the data is not assumed to be equidistantly sampled. In a discrete-time model, weekends and bank holidays are ignored so that the trading days are aggregated, meaning that Friday is followed by Monday in a normal week. Using a continuous-time model, it is possible to model the sampling times and thereby recognise that there is a longer time span between Friday and Monday. The so-called weekend effect in returns have been studied empirically for decades (see e.g. French 1980, Rogalski 1984, Asai and McAleer 2007). There are two main effects, first that returns are higher on Fridays and lower on Mondays than on other days and secondly that the variance is larger on Fridays and lower on Mondays. There are several plausible explanations for this, but all of them are more complicated to model than just treating Saturdays and Sundays as missing observations. Below, the observations are assumed to be equidistantly sampled in order to facilitate a comparison to the discrete-time models using model selection criteria.

The continuous-time hidden Markov models (CTHMMs) are estimated using the R-package `msm` due to Jackson (2011) that is based on direct numerical maximisation of the likelihood function.

4.   The data

The data analysed is daily log-returns of the S&P 500 and the FTSE 100 total return index covering the period from 23 July 1993 to 22 July 2013. The log-returns are calculated using $r_t = \log\left(P_t\right) - \log\left(P_{t-1}\right)$, where P_t is the closing price of the index on day t and log is the natural logarithm. The focus will be on the log-returns of the S&P 500 index as the analysis of the FTSE

---

[1]See Iversen *et al.* (2013) for an example of the use of splines to reduce the number of parameters in an inhomogeneous Markov model.
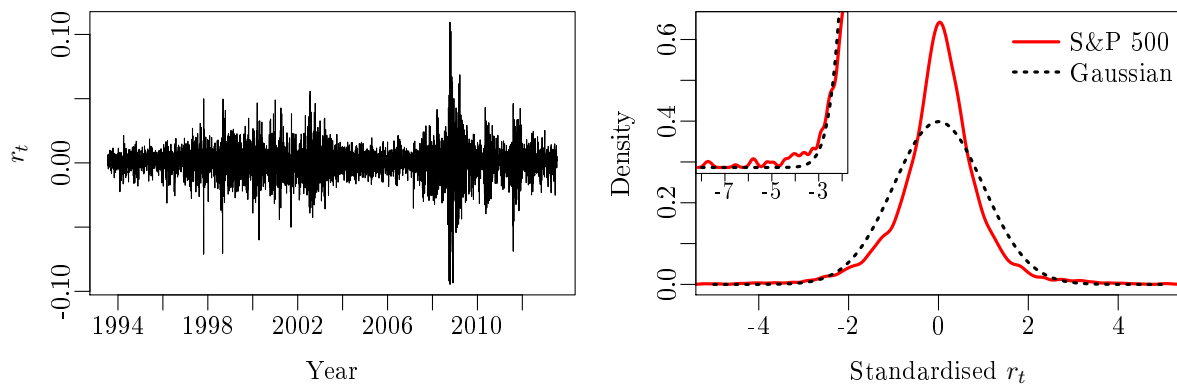
Figure 1. The daily log-returns of the S&P 500 total return index and a kernel estimate of the density of the standardised daily log-returns together with the density function for the standard normal distribution.

Table 1. The first four moments of the S&P 500 log-returns and the Jarque–Bera test statistic together with bootstrapped 95%-confidence intervals.

Mean	Std. deviation	Skewness	Kurtosis	JB
0.00034	0.0121	-0.24	11.3	14372
$[0.00001; 0.00068]$	$[0.0116; 0.0127]$	$[-0.75; 0.30]$	$[8.5; 14.1]$	$[6356; 25803]$

100 index showed similar results. The results of the analysis of the FTSE returns can be found in appendix B.

The 5040 log-returns of the S&P 500 index are shown in figure 1. The volatility is seen to form clusters as large price movements tend to be followed by large price movements and vice versa. Volatility clustering is a consequence of the persistence of the ACF of the squared returns (Cont 2001).

The first four moments of the daily log-returns are shown in table 1 together with approximate 95%-confidence intervals based on bootstrapping 500,000 series of length 5040 from the log-returns with replacement. The confidence intervals for the mean and skewness are very wide, whereas the estimates of the standard deviation and kurtosis are more certain. The distribution is left skew and leptokurtic with an excess kurtosis of 8.3 compared to the normal distribution. The Jarque–Bera test statistic[1] rejects the normal distribution at a 0.1% level of significance.

The excess kurtosis is evident from the plot of the density function in figure 1. There is too much mass centered right around the mean and in the tails compared to the normal distribution. There are 81 observations that deviate more than three standard deviations from the mean compared to an expectation of 14 if the returns were normally distributed.

Figure 2 shows the ACF of the absolute returns raised to different positive powers. It is a stylised fact that autocorrelations of positive powers of absolute returns are highest at power one. This is called the Taylor effect. The results generally agree with the Taylor effect although the effect is not clear-cut at the lowest lags.

5. Empirical results

The empirical autocorrelation function of the squared log-returns is shown in figure 3 together with ACFs of simulated squared returns from the fitted models. Of the two-state models, the HMM with normal conditional distributions is seen to be the best fit at the lowest lags, whereas the HSMM with normal conditional distributions is the best fit from lag 40 and upwards. The

---

[1]The Jarque–Bera test statistic is defined as $JB = T\left(\frac{\text{Skewness}^2}{6} + \frac{(\text{Kurtosis}-3)^2}{24}\right)$, where T is the number of observations.
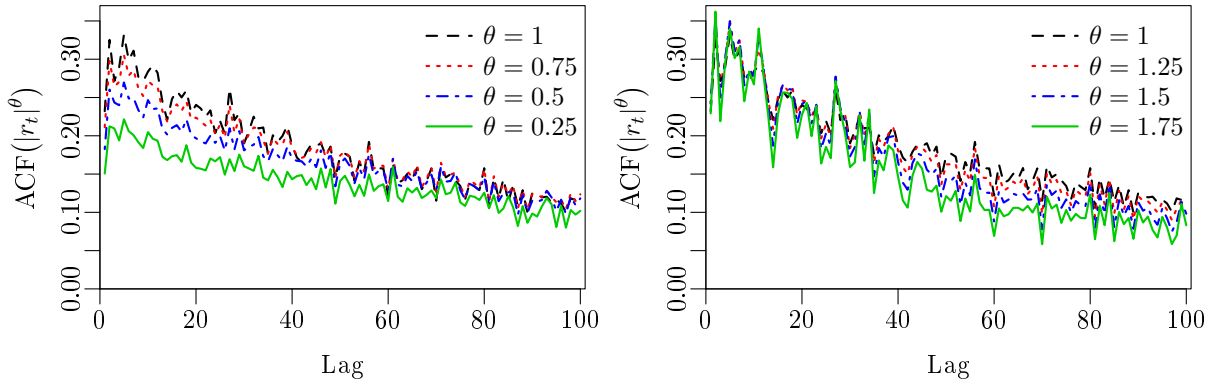
Figure 2. The empirical autocorrelation function of the absolute log-returns of the S&P 500 total return index raised to different positive powers.
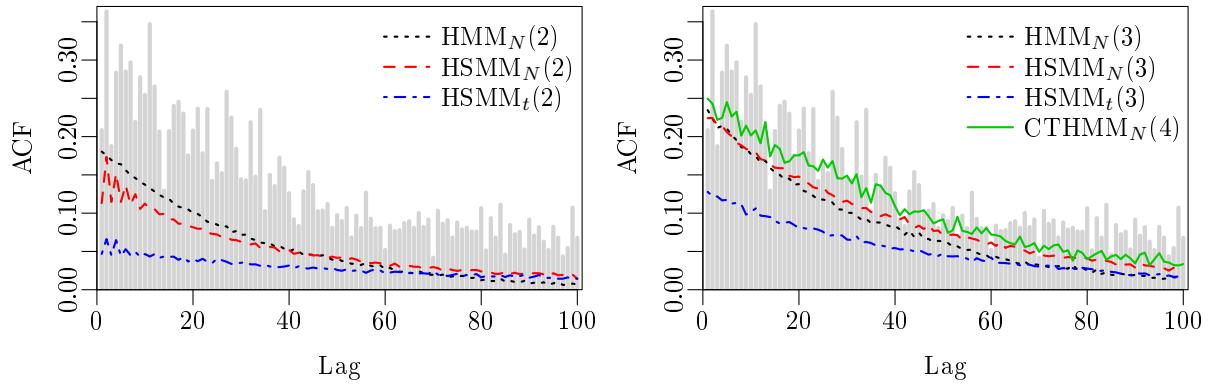

Figure 3. The empirical autocorrelation function of the squared log-returns at lag 1–100 together with simulated autocorrelation functions for the fitted models.

HSMM with t components is seen to be very persistent, but at too low of a level and it provides a poor fit overall.

As the HSMM with normal conditional distributions provides the best fit at the highest lags, this is also the model that best reproduces the stylised fact relating to the persistence of the ACF. This is true when looking at two-state models, as concluded by Bulla and Bulla (2006), but a much better fit can be obtained by increasing the number of states to three.

The HSMM with normal conditional distributions is also a better fit than the HMM when looking at the three-state models in figure 3, as the ACF for the HMM decays too fast. The ACF for the HSMM with conditional t distributions is again at too low of a level.

The fit of a CTHMM with three states with normal conditional distributions is similar to that of the three-state HMM. This appears from the mean squared error and the weighted mean squared error of the ACF of the squared returns for the fitted models in table 2. The weighted mean squared error reweights the error at lag k by $0.95^{(100-k)}$ to increase the influence of higher order lags following the approach by Bulla and Bulla (2006).

A CTHMM with four states with normal conditional distributions is seen to provide a better fit to the ACF of the squared returns than the three-state HSMM with normal conditional distributions. This observation is supported by the computed mean squared errors and weighted mean squared errors.

The first four moments of the log-returns are shown in table 3 together with the estimated moments for the fitted models based on 500,000 Monte Carlo simulations. Two states with normal conditional distributions are not enough to adequately capture the excess kurtosis of the log-returns. The two-state model with conditional t distributions is able to reproduce the excess kurtosis, but this model was not a good fit to the ACF.

Table 2. The mean squared error and the weighted mean squared error of the autocorrelation function of the squared returns and the outlier-corrected squared returns for the fitted models.

	Original data		Outlier-corrected data	
	MSE$\times 10^3$	WMSE$\times 10^3$	MSE$\times 10^3$	WMSE$\times 10^3$
HMM$_N$ (2)	7.9	4.6	13.1	9.7
HSMM$_N$ (2)	9.1	3.6	14.0	8.0
HSMM$_t$ (2)	15.7	4.6	15.1	6.5
HMM$_N$ (3)	4.4	3.2	6.1	6.7
HSMM$_N$ (3)	3.3	2.0	4.1	4.2
HSMM$_t$ (3)	9.2	3.4	5.1	4.0
CTHMM$_N$ (3)	4.2	3.2	6.1	6.5
HMM$_N$ (4)	4.0	2.2	3.4	3.3
HSMM$_N$ (4)	2.4	1.0	1.8	2.0
HSMM$_t$ (4)	3.4	1.1	1.4	1.2
CTHMM$_N$ (4)	1.9	1.3	1.3	1.6

Table 3. The first four moments of the log-returns together with bootstrapped 95%-confidence intervals and simulated moments for the fitted models.

Model	Mean	Std. dev.	Skewness	Kurtosis
r_t	0.00034	0.0121	-0.24	11.3
	$[0.00001; 0.00068]$	$[0.0116; 0.0127]$	$[-0.75; 0.30]$	$[8.5; 14.1]$
HMM$_N$ (2)	0.00034	0.0122	-0.17	5.6
HSMM$_N$ (2)	0.00035	0.0121	-0.24	6.5
HSMM$_t$ (2)	0.00042	0.0122	-0.14	12.0
HMM$_N$ (3)	0.00036	0.0120	-0.18	8.2
HSMM$_N$ (3)	0.00032	0.0121	-0.26	8.4
HSMM$_t$ (3)	0.00037	0.0123	-0.19	14.0
CTHMM$_N$ (3)	0.00025	0.0120	-0.20	8.7
HMM$_N$ (4)	0.00033	0.0122	-0.34	10.3
HSMM$_N$ (4)	0.00033	0.0122	-0.30	10.5
HSMM$_t$ (4)	0.00035	0.0125	-0.33	11.5
CTHMM$_N$ (4)	0.00037	0.0124	-0.30	10.7

The three-state models all provide a reasonable fit to the empirical moments. The kurtosis is still a little too low, with the exception of the HSMM with t components. The four-state models all provide a good fit to the empirical moments.

5.1 *Correcting for outliers*

Figure 4 shows the empirical ACF of the squared outlier-corrected log-returns together with the ACFs of the squared outlier-corrected simulated log-returns for the fitted models. Following the approach by Granger and Ding (1995a), values outside the interval $\bar{r}_t \pm 4\hat{\sigma}$ are set equal to the nearest boundary.

Restraining the impact of outliers reduces the amount of noise in the empirical ACF significantly. The noise reduction reveals a weekly variation that could suggest the need for an inhomogeneous, yet continuous, Markov model. The flexibility of a continuous-time model would be necessary to incorporate inhomogeneity without a dramatic increase in the number of parameters.

The conclusions regarding the fit of the different models to the empirical ACF are still valid when looking at the outlier-corrected returns. The outperformance of the HSMMs relative to the HMMs at the high lags is even more apparent when looking at the outlier-corrected data. What is also more apparent is the outperformance of the four-state CTHMM with normal conditional distributions relative to the three-state HSMMs. The ACF for the four-state CTHMM still decays
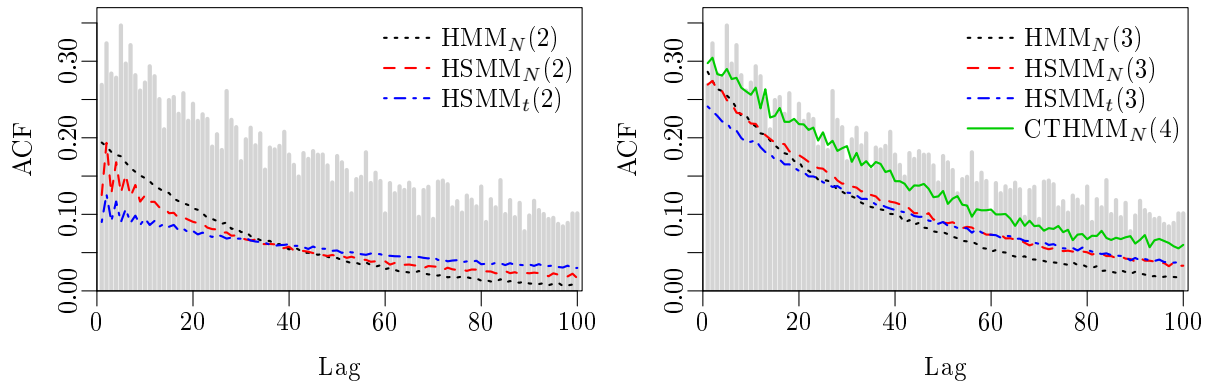
Figure 4.   The empirical autocorrelation function of the squared outlier-corrected log-returns at lag 1–100 together with autocorrelation functions of the squared outlier-corrected simulated log-returns for the fitted models.

Table 4.   Model selection based on the Akaike information criterion and the Bayesian information criterion.

Model	No. of parameters	Log-lik	AIC	BIC
HMM_N (2)	7	15984	-31954	-31908
HSMM_N (2)	9	16062	-32107	-32048
HSMM_t (2)	11	16137	-32251	-32180
HMM_N (3)	14	16214	-32400	-32309
HSMM_N (3)	17	16227	-32419	-32308
HSMM_t (3)	20	16245	-32449	-32319
CTHMM_N (3)	13	16209	-32391	-32306
HMM_N (4)	23	16262	-32478	-32328
HSMM_N (4)	27	16273	-32492	-32316
HSMM_t (4)	31	16284	-32505	-32303
CTHMM_N (4)	19	16256	-32474	-32350

too fast from lag 40 and onwards, but it clearly provides a better fit than the HSMMs with a similar number of parameters.

5.2   Model selection

Model selection involves both the choice of an appropriate number of states and the choice between competing state-dependent distributions. Likelihood ratio tests cannot be applied to models with different numbers of states as these are not hierarchically nested. Instead, penalised likelihood criteria can be used to select the model that is estimated to be closest to the 'true' model, as suggested by Zucchini and MacDonald (2009). The disadvantage is that model selection criteria provide no information about the confidence in the selected model relative to others.

A four-state HSMM fits the data as well as the four-state CTHMM with normal conditional distributions but it has 8 or 12 more parameters (see table 4). The four-state CTHMM is, therefore, preferred to both the three and the four-state HSMMs according to the Bayesian information criterion[1]. Akaike's information criterion[2] selects the four-state HSMM with t components as it puts less emphasis on the number of parameters. However, various simulation studies have shown that AIC tends to select models with too many states (Bacci $et\ al.$ 2014).

The parameter estimates are more uncertain the higher the number of states because of the quadratic increase in the number of parameters for the discrete-time models. Rydén $et\ al.$ (1998)

---

[1]The Bayesian information criterion is defined as $\mathrm{BIC} = -2\log L + p\log T$, where T is the number of observations and p is the number of parameters.

[2]The Akaike information criterion is defined as $\mathrm{AIC} = -2\log L + 2p$.
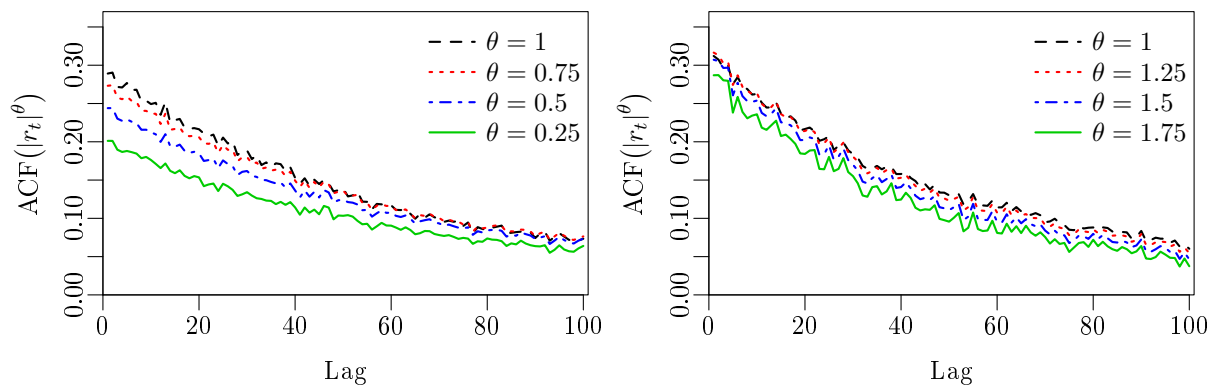
Figure 5. The autocorrelation function of the absolute value of 500,000 returns simulated from the estimated four-state CTHMM raised to different positive powers.

investigated HMMs with two and three states and found that the three-state models were "less similar to each other" and that "the estimation results seemed heavily dependent on outlying observations". That was also the reason why Bulla and Bulla (2006) only considered HSMMs with two-states. There is a strong preference for models with fewer parameters as a four-state HSMM with over 30 parameters is likely to be overfitting the data.

It is problematic to fit a five-state CTHMM to the log-returns. The likelihood function appears to be highly multimodal and it is easy to find several local maxima by using different starting values. This indicates that the model is overfitting the data. It was not possible to find a five-state CTHMM with a lower BIC-value than the four-state CTHMM.

The ability of the four-state CTHMM to capture the Taylor effect is illustrated in figure 5. The ACF of the absolute value of 500,000 simulated returns raised to different positive powers is seen to be highest at power one.

6.   Conclusion

HSMMs were found to be better at reproducing the slowly decaying ACF of squared daily returns of the S&P 500 and the FTSE 100 total return index than HMMs when looking at two and three-state models in agreement with the finding by Bulla and Bulla (2006). A much better fit to the slowly decaying ACF and the empirical moments was obtained by increasing the number of hidden states from two to three.

An extension to continuous time was presented and it was shown that a CTHMM with four states provides a better fit than the discrete-time models with three states with a similar number of parameters and, even more so, after restraining the impact of outliers. There was no indication that the memoryless property of the sojourn time distribution is inconsistent with the long-memory property of the squared returns.

Different models were preferred by the different selection criteria, but the four-state CTHMM with normal conditional distributions was selected by the Bayesian information criterion that was believed to be the most reliable. Finally, it was argued that the four-state CTHMM is preferred to the four-state HSMMs due to the significantly lower number of parameters resulting from the continuous-time formulation that makes the model less likely to be overfitting the data.

Acknowledgement

References

Ang, A. and Timmermann, A., Regime Changes and Financial Markets. Working Paper 17182, National Bureau of Economic Research, 2011.

Asai, M. and McAleer, M., Non-trading day effects in asymmetric conditional and stochastic volatility models. *Econometrics Journal*, 2007, **10**, 113–123.

Bacci, S., Pandolfi, S. and Pennoni, F., A comparison of some criteria for states selection in the latent Markov model for longitudinal data. *Advances in Data Analysis and Classification*, 2014, **8**, 125–145.

Bulla, J., Hidden Markov models with t components. Increased persistence and other aspects. *Quantitative Finance*, 2011, **11**, 459–475.

Bulla, J. and Bulla, I., Stylized facts of financial time series and hidden semi-Markov models. *Computational Statistics and Data Analysis*, 2006, **51**, 2192–2209.

Bulla, J., Bulla, I. and Nenadic, O., hsmm − An R package for analyzing hidden semi-Markov models. *Computational Statistics and Data Analysis*, 2010, **54**, 611–619.

Cappé, O., Moulines, E. and Rydén, T., *Inference in hidden Markov models*, 2005, Springer.

Cont, R., Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 2001, **1**, 223–236.

French, K.R., Stock Returns and the Weekend Effect. *J. Financial Economics*, 1980, **8**, 55–69.

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

Granger, C.W.J. and Ding, Z., Some Properties of Absolute Return: An Alternative Measure of Risk. *Annales D'Economie Et Statistique*, 1995a, **40**, 67–92.

Granger, C.W.J. and Ding, Z., Stylized Facts on the Temporal and Distributional Properties of Daily Data from Speculative Markets. Unpublished paper, Department of Economics, University of California, San Diego, 1995b.

Granger, C.W.J., Spear, S. and Ding, Z., Stylized Facts on the Temporal and Distributional Properties of absolute returns: An update. *Proceedings of the Hong Kong International Workshop on Statistics in Finance*, 2000, pp. 97–120 Imperial College Press.

Guédon, Y., Estimating hidden semi-Markov chains from discrete sequences. *J. Computational and Graphical Statistics*, 2003, **12**, 604–639.

Iversen, E.B., Møller, J.K., Morales, J.M. and Madsen, H., Inhomogeneous Markov Models for Describing Driving Patterns. IMM-Technical Report-2013 02, Technical University of Denmark, 2013.

Jackson, C.H., Multi-State Models for Panel Data: The msm Package for R. *J. Statistical Software*, 2011, **38**, 1–29.

Langrock, R. and Zucchini, W., Hidden Markov models with arbitrary state dwell-time distributions. *Computational Statistics and Data Analysis*, 2011, **55**, 715–724.

Madsen, H., Spliid, H. and Thyregod, P., Markov Models in Discrete and Continuous Time for Hourly Observations of Cloud Cover. *J. Applied Meteorology and Climatology*, 1985, **24**, 629–639.

Malmsten, H. and Teräsvirta, T., Stylized Facts of Financial Time Series and Three Popular Models of Volatility. Working Paper 563, Department of Economic Statistics, Stockholm School of Economics, 2004.

Nielsen, B.F., Matrix Analytic Methods in Applied Probability with a View towards Engineering Applications. Technical University of Denmark, Doctoral thesis, 2013.

Rogalski, R.J., New findings regarding day-of-the-week returns over trading and non-trading periods. *J. Finance*, 1984, **39**, 1603–1614.

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

Silvestrov, D. and Stenberg, F., A pricing process with Stochastic volatility controlled by a semi-Markov process. *Communications in Statistics - Theory and Methods*, 2004, **33**, 591–608.

Zucchini, W. and MacDonald, I.L., *Hidden Markov Models for Time Series: An Introduction Using R*, 2 , 2009, Chapman & Hall.

Appendix A: Parameter estimates

Table A1. Parameter estimates for the fitted m-state HMMs with normal conditional distributions together with bootstrapped standard errors based on 250 simulations of the model.

m	$\mathbf{\Gamma}$				$\mu \times 10^4$	$\sigma^2 \times 10^4$	δ
2	0.990	0.010			8.3	0.52	1.0
		(0.002)			(1.1)	(0.01)	(0.2)
	0.021	0.979			−6.9	3.47	0.0
	(0.004)				(4.9)	(0.14)	
3	0.982	0.018	0.000		10.1	0.32	1.0
		(0.004)	(0.001)		(1.3)	(0.01)	(0.1)
	0.015	0.978	0.006		0.5	1.28	0.0
	(0.003)		(0.002)		(2.5)	(0.04)	(0.1)
	0.000	0.030	0.970		−12.4	7.08	0.0
	(0.003)	(0.010)			(12.8)	(0.50)	
4	0.979	0.021	0.000	0.000	10.8	0.29	1.0
		(0.005)	(0.006)	(0.00)	(1.5)	(0.01)	(0.2)
	0.020	0.970	0.009	0.001	3.3	0.96	0.0
	(0.005)		(0.003)	(0.001)	(2.7)	(0.05)	(0.2)
	0.000	0.017	0.976	0.007	−2.9	2.39	0.0
	(0.000)	(0.006)		(0.004)	(5.7)	(0.14)	(0.1)
	0.000	0.000	0.049	0.951	−29.9	12.16	0.0
	(0.000)	(0.002)	(0.026)		(31.2)	(1.73)	

Table A2. Parameter estimates for the fitted m-state HSMMs with normal conditional distributions together with bootstrapped standard errors based on 250 simulations of the model. p and r are the parameters of the negative binomial sojourn time distribution.

m	$\mathbf{\Gamma}$				$1-p$	$r \times 10$	$\mu \times 10^4$	$\sigma^2 \times 10^4$	δ
2	0	1			0.994	0.4	9.0	0.48	1.0
					(0.002)	(0.1)	(1.3)	(0.01)	(0.4)
	1	0			0.975	0.5	−10.8	4.00	0.0
					(0.010)	(0.2)	(5.7)	(0.18)	
3	0	1.000	0.000		0.993	0.3	10.5	0.31	1.0
		(0.000)			(0.003)	(0.1)	(1.3)	(0.01)	(0.3)
	0.972	0	0.028		0.945	1.7	−1.0	1.51	0.0
	(0.013)				(0.037)	(1.7)	(3.0)	(0.07)	(0.3)
	0.000	1.000	0		0.983	5.5	−14.3	7.26	0.0
	(0.000)				(0.052)	(33.0)	(12.7)	(0.54)	
4	0	1.000	0.000	0.000	0.985	0.5	11.6	0.25	1.0
		(0.000)	(0.000)		(0.006)	(0.5)	(1.4)	(0.01)	(0.3)
	0.970	0	0.022	0.008	0.965	1.0	2.1	1.05	0.0
	(0.022)		(0.020)		(0.024)	(0.7)	(2.8)	(0.06)	(0.2)
	0.000	0.675	0	0.325	0.937	31.2	−2.4	2.38	0.0
	(0.000)	(0.161)			(0.124)	(124.8)	(5.2)	(0.17)	(0.2)
	0.000	0.000	1.000	0	0.977	4.2	−30.6	11.81	0.0
	(0.000)	(0.000)			(0.098)	(68.9)	(31.8)	(2.26)	

Table A3. Parameter estimates for the fitted m-state HSMMs with Student t conditional distributions together with bootstrapped standard errors based on 250 simulations of the model. p and r are the parameters of the negative binomial sojourn time distribution and t is the degrees of freedom for the conditional t distributions.

m	$\mathbf{\Gamma}$	$1-p$	$r \times 10$	$\mu \times 10^4$	$\sigma^2 \times 10^4$	t	δ
2	0 1	0.997	0.2	9.5	0.34	7.2	1.0
		(0.002)	(0.1)	(1.2)	(0.02)	(1.2)	(0.4)
	1 0	0.984	0.5	−6.0	2.10	5.6	0.0
		(0.008)	(0.2)	(4.8)	(0.17)	(1.0)	
3	0 1.000 0.000	0.990	7.4	10.5	0.25	6.7	1.0
	(0.042)	(0.009)	(7.8)	(1.1)	(0.01)	(1.4)	(0.3)
	0.630 0 0.370	0.979	10.5	1.3	1.16	22.4	0.0
	(0.144)	(0.047)	(22.2)	(2.7)	(0.07)	(467.9)	(0.3)
	0.000 1.000 0	0.983	5.2	−12.2	4.96	7.2	0.0
	(0.045)	(0.055)	(28.5)	(13.1)	(0.75)	(7.4)	
4	0 1.000 0.000 0.000	0.987	8.9	10.6	0.23	6.8	1.0
	(0.000) (0.000)	(0.013)	(10.2)	(1.4)	(0.01)	(1.5)	(0.2)
	0.610 0 0.296 0.094	0.957	18.6	4.1	0.86	24.8	0.0
	(0.103) (0.106)	(0.052)	(34.2)	(2.2)	(0.05)	(16.4)	(0.2)
	0.000 0.724 0 0.276	0.931	34.3	−1.9	2.22	49.0	0.0
	(0.000) (0.148)	(0.115)	(115.9)	(5.2)	(0.16)	(98.2)	(0.0)
	0.000 0.000 1.000 0	0.981	3.9	−29.3	9.56	13.8	0.0
	(0.000) (0.000)	(0.110)	(61.9)	(31.6)	(2.44)	(11968)	

Table A4. Parameter estimates for the fitted m-state CTHMMs with normal conditional distributions together with approximate standard errors based on the Hessian. The three-state model has a dummy state as the second and the third state are indistinguishable. No standard errors are given for the initial distributions as the Hessian is unreliable for this purpose.

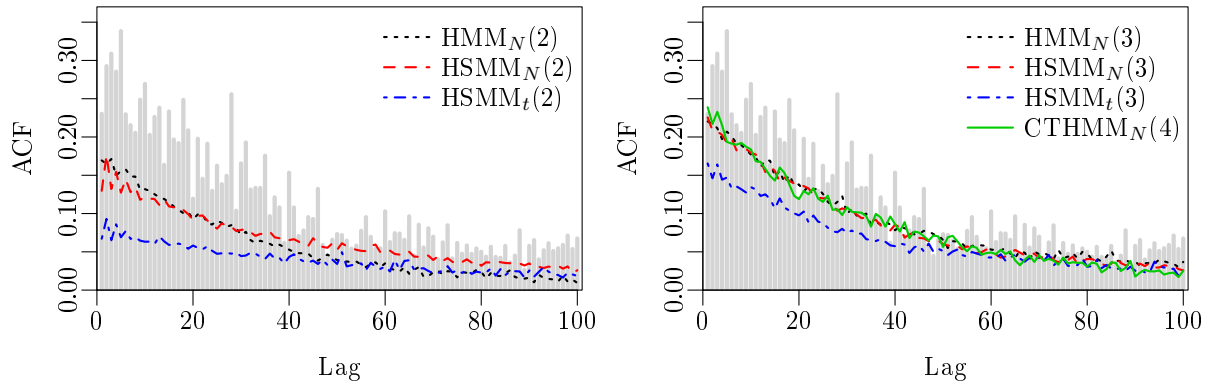| m	$\mathbf{Q}$	$\mu \times 10^4$	$\sigma^2 \times 10^4$	δ
3	−0.014 0.014 0 0	10.6	0.32	1.0
	(0.003)	(1.4)	(0.01)	
	0 −0.020 0.020 0	0.8	1.29	0.0
	(0.003)	(2.5)	(0.03)	
	0.048 0 −0.068 0.020	0.8	1.29	0.0
	(0.019)			
	0.005 0.019 0 −0.024	−14.6	7.12	0.0
	(0.003) (0.003)	(12.1)	(0.26)	
4	−0.018 0.017 0 0.001	11.1	0.29	1.0
	(0.005) (0.001)	(1.6)	(0.01)	
	0.015 −0.020 0.005 0	3.6	0.95	0.0
	(0.004) (0.002)	(2.6)	(0.03)	
	0 0.010 −0.015 0.005	−3.2	2.39	0.0
	(0.003) (0.002)	(5.2)	(0.10)	
	0.005 0 0.029 −0.034	−29.2	12.29	0.0
	(0.005) (0.013)	(25.9)	(0.82)	

Appendix B: FTSE results

Figure B1. The empirical autocorrelation function of the squared FTSE 100 log-returns at lag 1–100 together with simulated autocorrelation functions for the fitted models.

Table B1. The first four moments of the FTSE 100 log-returns together with bootstrapped 95%-confidence intervals and simulated moments for the fitted models.

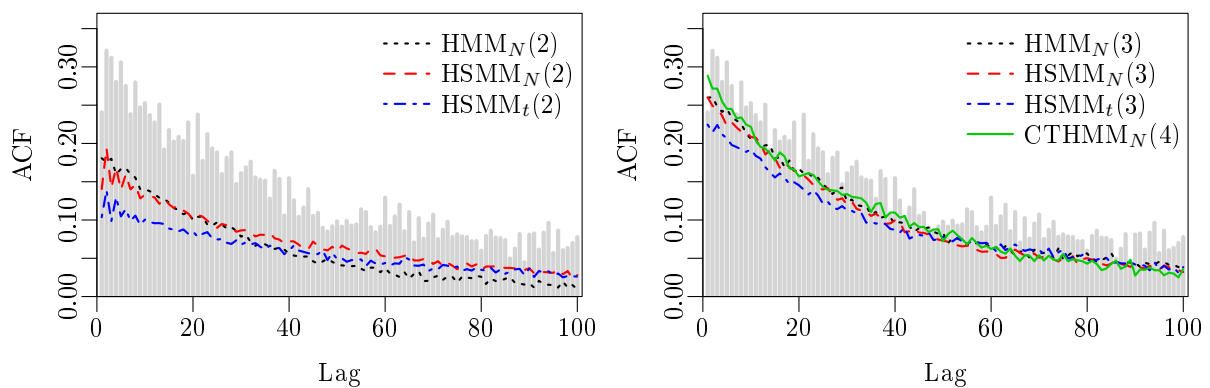| Model	Mean	Std. dev.	Skewness	Kurtosis
r_t	0.00031	0.0118	-0.16	8.9
	$[-0.00002; 0.00063]$	$[0.0113; 0.0123]$	$[-0.56; 0.24]$	$[7.0; 10.9]$
$\text{HMM}_N(2)$	0.00026	0.0119	-0.16	5.4
$\text{HSMM}_N(2)$	0.00032	0.0115	-0.18	6.0
$\text{HSMM}_t(2)$	0.00038	0.0116	-0.19	8.5
$\text{HMM}_N(3)$	0.00028	0.0116	-0.21	7.2
$\text{HSMM}_N(3)$	0.00026	0.0118	-0.23	7.1
$\text{HSMM}_t(3)$	0.00037	0.0118	-0.15	9.1
$\text{CTHMM}_N(4)$	0.00026	0.0123	-0.33	8.0

Figure B2. The empirical autocorrelation function of the squared outlier-corrected FTSE 100 log-returns at lag 1–100 together with autocorrelation functions of the squared outlier-corrected simulated log-returns for the fitted models.

15

Table B2.   Model selection based on the Akaike information criterion and the Bayesian information criterion.

Model	No. of parameters	Log-lik	AIC	BIC
$\mathrm{HMM}_N(2)$	7	16054	-32093	-32047
$\mathrm{HSMM}_N(2)$	9	16093	-32167	-32108
$\mathrm{HSMM}_t(2)$	11	16125	-32227	-32156
$\mathrm{HMM}_N(3)$	14	16220	-32413	-32321
$\mathrm{HSMM}_N(3)$	17	16224	-32414	-32303
$\mathrm{HSMM}_t(3)$	20	16235	-32429	-32299
$\mathrm{CTHMM}_N(4)$	19	16252	-32466	-32342