Journal of Financial Econometrics, 2019, Vol. 17, No. 1, 91–117
doi: 10.1093/jjfinec/nby019
Advance Access Publication Date: 12 September 2018
Article

Hidden Markov and Semi-Markov Models with Multivariate Leptokurtic-Normal Components for Robust Modeling of Daily Returns Series

Antonello Maruotti[1,2], Antonio Punzo[3] and Luca Bagnato[4]

[1]Libera Università Ss Maria Assunta, [2]Southampton Statistical Sciences Research Institute, University of Southampton, [3]Università di Catania and [4]Università Cattolica del Sacro Cuore

Address correspondence to Luca Bagnato, Università Cattolica del Sacro Cuore, Via Emilia Parmense, 84 - 29122 Piacenza, Italy, or e-mail: luca.bagnato@unicatt.it.

Received February 21, 2017; revised July 3, 2018; editorial decision July 3, 2018; accepted July 10, 2018

Abstract

We introduce multivariate models for the analysis of stock market returns. Our models are developed under hidden Markov and semi-Markov settings to describe the temporal evolution of returns, whereas the marginal distribution of returns is described by a mixture of multivariate leptokurtic-normal (LN) distributions. Compared to the normal distribution, the LN has an additional parameter governing excess kurtosis and this allows us a better fit to both the distributional and dynamic properties of daily returns. We outline an expectation maximization algorithm for maximum likelihood estimation which exploits recursions developed within the hidden semi-Markov literature. As an illustration, we provide an example based on the analysis of a bivariate time series of stock market returns.

Key words: daily returns, elliptical distributions, EM algorithm, hidden Markov model, hidden semi-Markov model, kurtosis, multivariate time series

JEL classification: C32, C51

1 Introduction

In this article, we introduce new and tractable multivariate models to describe the evolution over time and the main stylized facts of stock market returns. The marginal distribution of returns is assumed to follow a mixture of multivariate leptokurtic-normal (LN) distributions and the temporal dependence structure is embodied by using hidden Markov and semi-Markov chains. Our goal is to analyze stock market returns using a multivariate setting by defining a robust hidden Markov model (HMM) and a hidden semi-Markov model (HSMM).

We provide evidence of the capabilities of our proposals to: reproduce the probability density function of returns, model extreme values, forecast future values [and risk measures such as the value-at-risk (VaR)], and replicate most of the stylized facts characterizing stock market returns, such as skewness and heavy tails of their distribution, heteroscedasticity, and time-varying correlations.

The starting point of our analysis is represented by the increasing interest in finance in determining whether the observed data are driven by hidden variables (e.g., the *state* of the market), and if so, how such dynamics can be estimated (for further details see, e.g., Mamon and Elliott, 2007, Mamon and Elliott, 2014, and Ang and Timmermann, 2012). In this context, HMMs have become very popular in the financial time series literature since the seminal works of Hamilton (1989, 1990). Interesting conclusions can be also found in Elliott, Aggoun, and Moore (2008) and Hardy (2001). Univariate applications of HMMs to asset allocation, stock returns, and financial data are discussed, for example, in Mergner and Bulla (2008), Bartolucci and Farcomeni (2010), Erlewein, Mamon, and Davison (2011), Langrock, MacDonald, and Zucchini (2012), De Angelis and Paas (2013), Cavaliere, Costa, and De Angelis (2014), and Nystrup, Madsen, and Lindstrm (2017). Multivariate extensions have been proposed in Bernardi, Maruotti, and Petrella (2017).

When analyzing financial returns, attention is commonly focused on HMMs with normal components (N-HMMs; Ang and Timmermann, 2012). Their success is mainly due to Rydén, Teräsvirta, and Ösbrink (1998), who show that N-HMMs can reproduce most of the stylized facts characterizing asset returns; see e.g. Granger and Ding (1995a, 1995b), and Granger, Spear, and Ding (2000). However, many empirical studies have shown strong evidence of heavy-taileness and leptokurtosis in returns' distribution (Harris and Küçüközmen, 2001; Gettinby et al., 2004). Under the HMM setting, Perez-Quiros and Timmermann (2001), Bulla (2011), Punzo and Maruotti (2016), and Abanto-Valle et al. (2017) introduced robust extensions of N-HMMs, allowing for univariate and multivariate heavy-tailed conditional distributions; see also Maruotti and Punzo (2017) when covariates are available.

A fundamental assumption of any HMM is that the sojourn time (also defined as state duration), that is, the number of consecutive time points that the Markov chain spends in a given state, is implicitly geometrically distributed. HSMMs are designed to relax this condition by allowing the sojourn time to be modeled by more flexible parametric/nonparametric distributions. A first attempt of using HSMMs is given in Bulla and Bulla (2006) who show that HSMMs with negative-binomially distributed sojourn times perform better than HMMs in reproducing the long-memory property of squared daily returns.

Motivated by these considerations, and following Bagnato, Punzo, and Zoia (2017), in this article we introduce both HMMs and HSMMs with the multivariate LN distribution as emission distribution (ED). We refer to the resulting models as LN–HMM and LN–HSMM, respectively. As for the multivariate t, the LN distribution is characterized by one additional parameter compared to the multivariate normal. However, advantageously with respect to the t distribution, the additional parameter in the LN distribution directly corresponds to the quantity of interest, namely, the excess kurtosis. Up to our knowledge, the LN–HSMM represents the first attempt to introduce a multivariate heavy-tailed distribution into a HSMM setting for the analysis of stock market returns. Compared to the existing and widely used multivariate N-HMM, our models benefit of the LN ED in reproducing various stylized facts of stock returns.

We estimate the parameters of the proposed models via maximum likelihood (ML) by using an expectation maximization (EM) algorithm. Compared to the EM algorithm

commonly used to fit HMMs and HSMMs, the M-step requires particular care. To avoid the use of a constrained maximization algorithm, a convenient reparameterization of the multivariate LN distribution is considered to simplify the maximization step. Using daily returns of the US stock market NASDAQ and Eurozone stock market STOXX50 indexes, we show that the proposed multivariate models are preferred according to classical model selection criteria and perform rather well in recovering most of the data features.

The remainder of this article is organized as follows. Section 2 introduces HMMs and HSMMs, and presents the multivariate LN distribution. In Section 3, we outline an *ad hoc* version of the EM algorithm to estimate the parameters of the more general LN–HSMM, extending the Baum–Welch iterative procedure (Baum et al., 1970) to deal with multivariate LN EDs. In the same section, for both the LN–HMM and LN–HSMM, we give also details about the initialization of the EM algorithm (Section 3.2) and the computation of standard errors by means of a parametric bootstrap procedure (Section 3.3). In Section A of the Supplementary Material, we investigate the parameter recovery of the EM algorithm via a simulation study. In Section 4, a maximum *a posteriori* (MAP) procedure to make forecast is presented. At last, in Section 5, we illustrate the proposal by describing the data in depth and investigating the behavior of the proposed models with respect to several data features. Conclusions and final remarks are discussed in Section 6.

2 Methodology

2.1 Hidden Markov models

A HMM is a particular kind of dependent mixture consisting of two parts: an underlying unobserved process and a state-dependent process (see Zucchini, MacDonald, and Langrock, 2016, Chapter 2). These processes are described below.

Let $\{S_t;\, t = 1, \ldots, T\}$ be a first-order Markov chain defined on the discrete state space $\{1, \ldots, k, \ldots, K\}$. The process $\{S_t\}$, which represents the underlying unobserved process of the HMM, fulfills the Markov property

$$P(S_t = s_t | S_{t-1} = s_{t-1}, \ldots, S_1 = s_1) = P(S_t = s_t | S_{t-1} = s_{t-1}),$$

meaning that the state at any given time t depends on the previous states only through the state at time t–1, $t = 2, \ldots, T$.

Let $\pi_k = P(S_1 = k)$ be the initial probability (IP) of state k, $k = 1, \ldots, K$. Moreover, let $\pi_{jk} = P(S_t = k | S_{t-1} = j)$, with $\sum_{k=1}^{K} \pi_{jk} = 1$ and $\pi_{jk} \geq 0$, denote the (time-homogeneous) transition probability (TP), that is, the probability of jumping from state j at time t–1 to state k at time t, $t = 2, \ldots, T$ and $j, k = 1, \ldots, K$. The K-dimensional vector π collects the IPs, whereas the $K \times K$ TP matrix Π collects the TPs, which govern the state switching behavior of the chain.

The number of time steps spent in a given state is called the sojourn time. Under an HMM with underlying first-order Markov chain, the sojourn distribution (SD), that is, the probability of spending u consecutive time steps in state k, $k = 1, \ldots, K$, is

$$d_k(u) = P(S_{t+u} \neq k, S_{t+u-1} = k, \ldots, S_{t+1} = k | S_t = k, S_{t-1} \neq k) = \pi_{kk}^{u-1}(1 - \pi_{kk}). \quad (1)$$

Hence, the sojourn time is geometrically distributed and the most likely sojourn time for every state is 1 (Zucchini, MacDonald, and Langrock, 2016, p. 165).

Let $\{\boldsymbol{X}_t;\, t = 1, \ldots, T\}$ denote a multivariate time series of length T, with each $\boldsymbol{X}_t$ taking values in $\mathbb{R}^P$. The process $\{\boldsymbol{X}_t\}$ represents the state-dependent process of the HMM and fulfills the conditional (on the hidden states) independence property

$$P(\boldsymbol{X}_t = \boldsymbol{x}_t | \boldsymbol{X}_1 = \boldsymbol{x}_1, \ldots, \boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1}, S_1 = s_1, \ldots, S_t = s_t) = P(\boldsymbol{X}_t = \boldsymbol{x}_t | S_t = s_t).$$

The conditional distribution of the observed variable $\boldsymbol{X}_t$ given the unobserved (or hidden) state $S_t = k$, $k = 1, \ldots, K$, is referred to as the ED and it is denoted by

$$f(\boldsymbol{x}_t | S_t = k) = P(\boldsymbol{X}_t = \boldsymbol{x}_t | S_t = k). \tag{2}$$

The HMM is specified by the triple (IP, TP, ED).

2.2 Hidden Semi-Markov Models

Due to the non-zero probability of self-transition of a nonabsorbing state, the sojourn time of an HMM implicitly follows a geometric distribution. To make the approach more flexible, a possible solution is to explicitly estimate the SD $d_k(u)$, producing what is referred as a HSMM. Thus, rather than having $d_k(u)$ defined by Π as in Equation (1), we model $d_k(u)$ explicitly. Note that the diagonal elements of Π for a HSMM are required to be zero, since the sojourn in the same state is now governed by the model chosen for the SD (see, e.g., Bulla, Bulla, and Nenadić, 2010, p. 612). Therefore, a HSMM is specified by the quadruple (IP, TP, SD, ED). In other words, the difference with respect to HMMs is that the probabilities of self-transitions π_{kk}, $k = 1, \ldots, K$, which determine the time spent in each state, are now modeled separately by the SD $d_k(u)$.

As in O'Connell and Højsgaard (2011), we use the shifted Poisson and gamma as models for the SD (for further examples of SDs commonly used, see Zucchini, MacDonald, and Langrock, 2016, Chapter 12.6). The shifted Poisson distribution has probability mass function

$$d_k(u; \lambda_k, \zeta_k) = \frac{e^{-\lambda_k} \lambda_k^{(u-\zeta_k)}}{(u - \zeta_k)!}, \quad u = \zeta_k, \zeta_k + 1, \ldots, \tag{3}$$

where $\lambda_k > 0$, while $\zeta_k \in \{1, 2, \ldots\}$ is an additional (with respect to the classical Poisson distribution) shift parameter setting the minimum sojourn time in state k, $k = 1, \ldots, K$. The gamma distribution has probability density function

$$d_k(u; \lambda_k, \zeta_k) = \frac{u^{\lambda_k - 1}}{\Gamma(\lambda_k) \zeta_k^{\lambda_k}} \exp\left(-\frac{u}{\zeta_k}\right), \quad u > 0, \tag{4}$$

where $\lambda_k > 0$ and $\zeta_k > 0$ are the shape and scale parameters, respectively. The first-order Markov chain and the SD together constitute a semi-Markov chain.

2.3 Multivariate Leptokurtic Normal Emission Distribution

As ED to be used in the definition of HMMs and HSMMs, we consider the P-variate LN (Bagnato, Punzo, and Zoia, 2017) with density

$$f_{LN}(\boldsymbol{x}_t; \mu_k, \Sigma_k, \beta_k) = q[(\boldsymbol{x}_t - \mu_k)' \Sigma_k^{-1} (\boldsymbol{x}_t - \mu_k); \beta_k] f_N(\boldsymbol{x}_t; \mu_k, \Sigma_k), \quad \boldsymbol{x}_t \in \mathbb{R}^P, \tag{5}$$

where $f_N(\cdot;\mu_k,\Sigma_k)$ is the density of a P-variate normal random vector with mean vector μ_k and covariance matrix Σ_k, and

$$q(y;\beta)=1+\frac{\beta}{8P(P+2)}[y^2-2(P+2)y+P(P+2)],\quad y\geq 0.$$

The use of this ED, in the definition of HMMs and HSMMs, gives rise to the proposals of this paper that we simply denote as LN–HMM and LN–HSMM. Advantageously, with respect to competing elliptical distributions, the parameters of the multivariate LN in Equation (5) are directly related to the theoretical moments of $\boldsymbol{X}_t|S_t=k$: μ_k represents the mean vector, Σ_k the covariance matrix, and $\beta_k\in[0,\beta_{\max}]$ the excess kurtosis, with $\beta_{\max}=\min[4P,4P(P+2)/5]$. In particular, the kurtosis of $\boldsymbol{X}_t|S_t=k$ equals $P(P+2)+\beta_k$, where $P(P+2)$ corresponds to the kurtosis of the P-variate normal distribution. As a special case, f_{LN} coincides with f_N for $\beta_k=0$. The constraint $\beta_k\in[0,\beta_{\max}]$ is the intersection of two constraints:

i. $\beta_k\in[0,4P]$, which assures that the density of $\boldsymbol{X}_t|S_t=k$ is positive elliptical;
ii. $\beta_k\in[0,4P(P+2)/5]$, which guarantees that the density of $\boldsymbol{X}_t|S_t=k$ is unimodal.

Figure B.1 of the Supplementary Material displays examples of bivariate LN densities having zero mean, identity covariance matrix, and different values of β. These plots show how, maintaining fixed mean and covariance matrix, peakedness and heavy tails move in line with β.

3 Maximum likelihood Estimation

As habit for mixture models, we adopt the EM algorithm (Baum et al., 1970; Dempster, Laird, and Rubin, 1977) to find ML estimates for the parameters of our models on the basis of the observed time series $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_T$. For the sake of brevity, we illustrate the EM algorithm (Section 3.1), along with its initialization strategy (Section 3.2), to fit the more general LN–HSMM; in principle, the same algorithm should work for the LN–HMM too by using a geometric SD (cf. Section 2.2). In Section 3.3, we also consider a parametric bootstrap procedure to estimate standard errors of the ML estimates.

3.1 EM Algorithm

Once the number of latent states K has been assigned/fixed, the algorithm basically works on the complete-data likelihood

$$\mathcal{L}_c(\vartheta)=\pi_{s_1^*}d_{s_1^*}(u_1;\lambda_{s_1^*},\zeta_{s_1^*})\left\{\prod_{r=2}^{R-1}\pi_{s_{r-1}^*s_r^*}d_{s_r^*}(u_r;\lambda_{s_r^*},\zeta_{s_r^*})\right\}\pi_{s_{R-1}^*s_R^*}D_{s_R^*}(u_R;\lambda_{s_R^*},\zeta_{s_R^*})\prod_{t=1}^{T}f_{LN}(\boldsymbol{x}_t;\mu_{s_t},\Sigma_{s_t},\beta_{s_t}),\tag{6}$$

where s_r^* is the rth visited state, u_r is the time spent in that state (i.e., the duration of the rth visit), R–1 is the number of state changes up to time T, and ϑ designates the vector of all parameters, that is, $\vartheta=\{\pi,\Pi,\lambda,\zeta,\mu,\Sigma,\beta\}$, where $\lambda=\{\lambda_k;k=1,\ldots,K\}$, $\zeta=\{\zeta_k;k=1,\ldots,K\}$, $\mu=\{\mu_k;k=1,\ldots,K\}$, $\Sigma=\{\Sigma_k;k=1,\ldots,K\}$, and $\beta=\{\beta_k;k=1,\ldots,K\}$. Guédon (2003) proposed using the survivor function

$$D_k(u;\lambda_k,\zeta_k)=\sum_{v\geq u}d_k(v;\lambda_k,\zeta_k)\tag{7}$$

for the sojourn time in the last state, so we do not have to assume the process is leaving a state immediately after time T. The use of this survivor function has two advantages: it improves parameter estimation and, perhaps more importantly, it provides a more accurate prediction of the last state visited which is important for applications where we wish/need to estimate the most recent visited state (O'Connell and Højsgaard, 2011, p. 7).

The EM algorithm iterates between two steps, one E-step and one M-step, until convergence. These steps, for the generic $(h+1)$th iteration of the algorithm, $h = 0, 1, \ldots$, are briefly outlined below.

3.1.1 E-step

In the E-step we calculate

$$Q(\vartheta|\vartheta^{(h)}) = E_{\vartheta^{(h)}}\{\ln[\mathcal{L}_c(\vartheta)]|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T\}, \tag{8}$$

the expected complete data log-likelihood given the observed data. In Equation (8), the expectation operator E has the subscript $\vartheta^{(h)}$ to explicitly convey that this expectation is being effected using $\vartheta^{(h)}$ for ϑ. We simply require the calculation of the following three quantities:

1. the probability

$$\gamma_t^{(h)}(k) = P_{\vartheta^{(h)}}(S_t = k|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T) \tag{9}$$

of being in state k at time t given the observed sequence;
2. the probability

$$\xi_t^{(h)}(j,k) = P_{\vartheta^{(h)}}(S_{t-1} = j, S_t = k|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T) \tag{10}$$

that the process left state j at time t–1 and entered state k at t given the observed sequence;
3. the expected number of times a process spends u time steps in state k,

$$\eta_{ku}^{(h)} = P_{\vartheta^{(h)}}(S_{1+u} \neq k, S_{1+u-v} = k, v = 1, \ldots, u|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$$

$$+\sum_{t=2}^{T} P_{\vartheta^{(h)}}(S_{t+u} \neq k, S_{t+u-v} = k, v = 1, \ldots, u, S_{t-1} \neq k|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T). \tag{11}$$

The first two quantities, $\gamma_t^{(h)}(j)$ and $\xi_t^{(h)}(j,k)$, can be calculated via a dynamic programming method known as the forward–backward algorithm (see, e.g., Rabiner, 1989). As concerns the updating of Equation (11), Guédon (2003) provides a version of the forward–backward algorithm which is implemented in the **mhsmm** package (O'Connell and Højsgaard, 2011) for R (R Core Team, 2017).

3.1.2 M-step

The M-step, on the same iteration, involves choosing $\vartheta^{(h+1)}$ as the value of ϑ that maximizes the conditional expectation $Q(\vartheta|\vartheta^{(h)})$. Fortunately, as well-documented in Guédon (2003), $Q(\vartheta|\vartheta^{(h)})$ for HSMMs can be typically rewritten as a sum of terms, each term depending on a given subset of parameters. This allows the updating formulas for each subset of parameters to be obtained by independently maximizing each of these terms.

In our case, the conditional expectation in Equation (8) can be rewritten as

$$Q(\vartheta|\vartheta^{(h)}) = Q_{\text{IP}}(\pi|\vartheta^{(h)}) + Q_{\text{TP}}(\Pi|\vartheta^{(h)}) + Q_{\text{SD}}(\lambda,\zeta|\vartheta^{(h)}) + Q_{\text{ED}}(\mu,\Sigma,\beta|\vartheta^{(h)}). \tag{12}$$

For details on the specific formulation of the first three terms on the right-hand side of Equation (12), see Guédon (2003).

The maximization of $Q_{\text{IP}}(\pi|\vartheta^{(h)})$ and $Q_{\text{TP}}(\Pi|\vartheta^{(h)})$ with respect to π and Π, respectively, subject to the constraints on these parameters, yields

$$\pi_k^{(h+1)} = \gamma_1^{(h)}(k) \quad \text{and} \quad \pi_{jk}^{(h+1)} = \frac{\sum_{t=2}^{T}\xi_t^{(h)}(j,k)}{\sum_{t=2}^{T}\sum_{j\neq k}\xi_t^{(h)}(j,k)}. \tag{13}$$

The maximization of $Q_{\text{SD}}(\lambda,\zeta|\vartheta^{(h)})$ with respect to λ and ζ depends from the chosen SD. If the shifted Poisson in considered, for each state k, $k = 1,\ldots,K$, λ_k is estimated for each possible shift parameter value $\zeta_k \in \{1,\ldots,\min(u : \eta_{ku}^{(h)} > 0)\}$ by classical point estimation procedures from the quantities $\eta_{ku}^{(h)}$ (Johnson, Kemp, and Kotz, 2005). The pair (λ_k,ζ_k) which gives the maximum value of $Q_{\text{SD}}(\lambda,\zeta|\vartheta^{(h)})$, say $(\lambda_k^{(h+1)},\zeta_k^{(h+1)})$, is retained. Guédon (2003) states that this *ad hoc* procedure works well in practice and this statement is further corroborated by O'Connell and Højsgaard (2011) via simulations. If the gamma is considered as SD, then the maximization of $Q_{\text{SD}}(\lambda,\zeta|\vartheta^{(h)})$ is performed following the methodology of Choi and Wette (1969) which is implemented by the `gammafit()` function of the **mhsmm** package (O'Connell and Højsgaard, 2011).

Finally, maximizing

$$Q_{\text{ED}}(\mu,\Sigma,\beta|\vartheta^{(h)}) = \sum_{k=1}^{K} Q_{\text{ED},k}(\mu_k,\Sigma_k,\beta_k)$$

with respect to μ, Σ, and β is equivalent to independently maximizing each of the K weighted log-likelihood functions

$$Q_{\text{ED},k}(\mu_k,\Sigma_k,\beta_k) = \sum_{t=1}^{T}\gamma_t^{(h)}(k)\ln[f_{LN}(\boldsymbol{x}_t;\mu_k,\Sigma_k,\beta_k)] \tag{14}$$

with respect to μ_k, Σ_k, and β_k, $k = 1,\ldots,K$. Maximizing $Q_{\text{ED},k}(\mu_k,\Sigma_k,\beta_k)$ in Equation (14) with respect to μ_k, Σ_k, and β_k, is a constrained problem due to constraints imposed by Σ_k and β_k. As suggested in Bagnato, Punzo, and Zoia (2017), to obtain an unconstrained optimization problem, we apply the following reparameterization for Σ_k and β_k. According to the Cholesky decomposition, we write

$$\Sigma_k = \Omega_k'\Omega_k,$$

where Ω_k is an upper triangular matrix. As concerns β_k we write

$$\beta_k = \beta_{\max}\frac{\exp(\tau_k)}{1+\exp(\tau_k)},$$

with $\tau_k \in \mathbb{R}$. Considering this parametrization, $Q_{\text{ED},k}(\mu_k,\Sigma_k,\beta_k)$ can be rewritten as

$$\begin{aligned}\tilde{Q}_{\text{ED},k}(\mu_k,\Omega_k,\tau_k) = &-\frac{1}{2}\sum_{t=1}^{T}\gamma_t^{(h)}(k)\ln|\Omega_k'\Omega_k| - \frac{PT_k^{(h)}}{2}\ln(2\pi)\\ &-\frac{1}{2}\sum_{t=1}^{T}\gamma_t^{(h)}(k)(\boldsymbol{x}_t-\mu_k)'\Omega_k^{-1}(\Omega_k')^{-1}(\boldsymbol{x}_t-\mu_k)\end{aligned}$$

$$+\sum_{t=1}^{T}\gamma_t^{(h)}(k)\ln\left\{1+\frac{\beta_{\max}}{8P(P+2)}\frac{\exp(\tau_k)}{1+\exp(\tau_k)}[((\boldsymbol{x}_t-\mu_k)'\Omega_k^{-1}(\Omega_k')^{-1}(\boldsymbol{x}_t-\mu))^2\right.$$

$$\left.-2(P+2)(\boldsymbol{x}_t-\mu_k)'\Omega_k^{-1}(\Omega_k')^{-1}(\boldsymbol{x}_t-\mu_k)+P(P+2)]\right\}, \tag{15}$$

where $T_k^{(h)}=\sum_{t=1}^{T}\gamma_t^{(h)}(k)$. By applying vector derivatives of Equation (15) with respect to μ_k, we obtain the equation

$$\frac{\partial\tilde{Q}_{\mathrm{ED},k}(\mu_k,\Omega_k,\tau_k)}{\partial\mu_k}=\sum_{t=1}^{T}\gamma_t^{(h)}(k)(1-\nu_t)\Omega_k^{-1}(\Omega_k')^{-1}(\boldsymbol{x}_t-\mu_k), \tag{16}$$

with

$$\nu_t=\frac{\beta_{\max}}{2P(P+2)}\frac{\exp(\tau_k)}{1+\exp(\tau_k)}\frac{(\boldsymbol{x}_i-\mu_k)'\Omega_k^{-1}(\Omega_k')^{-1}(\boldsymbol{x}_i-\mu_k)-(P+2)}{q\left((\boldsymbol{x}_t-\mu_k)'\Omega_k^{-1}(\Omega_k')^{-1}(\boldsymbol{x}_t-\mu_k);\beta_{\max}\frac{\exp(\tau_k)}{1+\exp(\tau_k)}\right)}.$$

By computing the derivative of Equation (15) with respect to Ω_k, we obtain

$$\frac{\partial\tilde{Q}_{\mathrm{ED},k}(\mu,\Omega_k,\tau_k)}{\partial\Omega_k}=-T_k^{(h)}(\Omega_k')^{-1}+\sum_{t=1}^{T}\gamma_t^{(h)}(k)(1-\nu_t)(\Omega_k')^{-1}(\boldsymbol{x}_t-\mu_k)(\boldsymbol{x}_t-\mu_k)'\Omega_k^{-1}(\Omega_k')^{-1}. \tag{17}$$

Finally, the derivative of Equation (15) with respect to τ_k is

$$\frac{\partial\tilde{Q}_{\mathrm{ED},k}(\mu_k,\Omega_k,\tau_k)}{\partial\tau_k}=\frac{1}{1+\exp(\tau_k)}\sum_{t=1}^{T}\gamma_t^{(h)}(k)\left[1-\frac{1}{q\left((\boldsymbol{x}_t-\mu_k)'\Omega_k^{-1}(\Omega_k')^{-1}(\boldsymbol{x}_i-\mu_k);\beta_{\max}\frac{\exp(\tau_k)}{1+\exp(\tau_k)}\right)}\right]. \tag{18}$$

Operationally, we perform unconstrained maximization of Equation (15) with respect to μ_k, Ω_k, and τ_k by the general-purpose optimizer `optim()` for R, included in the **stats** package. The BFGS method, passed to `optim()` via the argument method, is used for maximization. For the evaluation of the parameter recovery of the illustrated algorithm, see the results from the simulation study described in Section A of the Supplementary Material.

3.2 Initialization Strategy

The choice of the starting values for the EM algorithm applied to mixture models constitutes an important issue (see, e.g., Biernacki, Celeux, and Govaert, 2003, Karlis and Xekalaki, 2003, and Bagnato and Punzo, 2013). To fit the LN–HSMM with K states, we initialize the EM algorithm described in Section 3.1 by providing the initial parameters, say $\vartheta^{(0)}$, to the first E-step of the algorithm. The initial parameters are provided in the following way. An initial state partition, say $\{S_t^{(0)}\}$, is obtained by the K-means method as implemented by the `kmeans()` function of the **stats** package. Based on this partition, the off-diagonal elements of $\Pi^{(0)}$ are computed as proportions of transition. As concerns the initial parameters $\lambda_k^{(0)}$ and $\zeta_k^{(0)}$ of the considered SD (refer to Section 2.2), the sequence of sojourn times in state k is first computed from $\{S_t^{(0)}\}$ and, on such a sequence, $\lambda_k^{(0)}$ and $\zeta_k^{(0)}$ are estimated by the ML approach using the `fitdist()` function of the **fitdistrplus** package (Delignette-Muller and Dutang, 2015; Delignette-Muller, Dutang, and Siberchicot, 2017).

Finally, as concerns the initial parameters $\mu_k^{(0)}$, $\Sigma_k^{(0)}$, and $\beta_k^{(0)}$ of the emission LN distribution, the nested normal distribution if fitted by ML on the observations in state k. This means that we use the sample mean vector for $\mu_k^{(0)}$, the sample covariance matrix for $\Sigma_k^{(0)}$, and $\beta_k^{(0)} = 0$; the latter choice is due to the fact, as illustrated in Section 2.3, the $f_N(\mu_k^{(0)}, \Sigma_k^{(0)})$ coincides with $f_{LN}(\mu_k^{(0)}, \Sigma_k^{(0)}, \beta_k^{(0)} = 0)$.

3.3 Bootstrapped Standard Errors

The estimation procedure outlined above does not produce standard errors of the estimates, because approximations based on the observed information matrix do not provide numerically stable results for long time series. We estimate standard errors through a parametric bootstrap-based technique which can be roughly summarized as follows.

1. Fit the model, that is, estimate its parameters, and determine the number R–1 of state changes.
2. a Generate a P-variate time series with R–1 state changes from the fitted model, named bootstrap time series.
   b Estimate the parameters for the bootstrap time series and record them.
   c Repeat the previous two steps for a large number of times B.

The standard errors of the estimated parameters are then approximated by the sample standard deviation of the bootstrap estimates. Note that, at the step 2.a, the simulation of data from each P-variate LN ED is simplified by its elliptical stochastic representation detailed in Bagnato, Punzo, and Zoia (2017); for details about how to simulate data from elliptical distributions see Frahm (2004).

4 Forecasting

Forecasting daily values and, more in general, distributions of stock returns for the next day, given historical data, is an important task in financial applications. Some approaches have been already proposed to cope with this issue in the HMM setting (see, e.g., Hassan, 2009). For the LN–HMM, we use the MAP approach introduced by Gupta and Dhingra (2012). For the LN–SMMs, we extend this approach as follows. Given the historical data $\{\boldsymbol{X}_t; t = 1, \ldots, T\}$, and the ML estimate $\hat{\vartheta}$ of the HSMM-parameters ϑ on these data, we compute the MAP estimate $\hat{\boldsymbol{X}}_{T+1}$ of $\boldsymbol{X}_{T+1}$ at the next day $T+1$ as

$$\hat{\boldsymbol{X}}_{T+1} = \underset{\boldsymbol{X}_{T+1}}{\operatorname{argmax}} P_{\hat{\vartheta}}(\boldsymbol{X}_{T+1} | \boldsymbol{X}_1, \ldots, \boldsymbol{X}_T) = \underset{\boldsymbol{X}_{T+1}}{\operatorname{argmax}} \frac{P_{\hat{\vartheta}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T, \boldsymbol{X}_{T+1})}{P_{\hat{\vartheta}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T)}, \tag{19}$$

where $\boldsymbol{X}_{T+1}$ is allowed to vary over all possible values. Since the denominator in Equation (19) is constant with respect to $\boldsymbol{X}_{T+1}$, the MAP estimate simplifies as

$$\hat{\boldsymbol{X}}_{T+1} = \underset{\boldsymbol{X}_{T+1}}{\operatorname{argmax}} P_{\hat{\vartheta}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T, \boldsymbol{X}_{T+1}). \tag{20}$$

Operationally, we obtain $\hat{\boldsymbol{X}}_{T+1}$ in Equation (20) by the general-purpose optimizer `optim()`, included in the **stats** package. The BFGS method, passed to `optim()` via the argument method, is used for maximization. We compute the joint probability $P_{\hat{\vartheta}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T, \boldsymbol{X}_{T+1})$ by using the `predict.hsmm()` function of the **mhsmm** package (O'Connell and Højsgaard, 2011); this function also returns the jointly most likely configuration

$S_1, \ldots, S_T, S_{T+1}$ of the $T+1$ states, which is found using the Viterbi algorithm (Forney, 1973).

In addition to try to predict the next value $\boldsymbol{X}_{T+1}$ of the time series, the prediction of the most likely state S_{T+1} for $\hat{\boldsymbol{X}}_{T+1}$ allows us to define MAP prediction intervals, with coverage probability $1-\alpha$, for each of the P variables of $\hat{\boldsymbol{X}}_{T+1}$. With this aim, we first simulate M values from the ED for that state, and then we compute the quantiles of order $\alpha/2$ and $1-\alpha/2$ for each dimension. In other words, following Zucchini and Neumann (2001), we are focusing on the prediction of the VaR associated with financial assets on a daily basis.

5 Application

In this section, we illustrate the practical relevance of our methodology by jointly modelling the US stock market index NASDAQ and the Eurozone stock market index STOXX50; the latter collecting the largest and most liquid Eurozone stocks. Data description with basic statistics are provided in Section 5.1 while, in Section 5.2, different models are fitted with varying SDs (geometric, shifted Poisson, and gamma), EDs (multivariate normal and LN), and number of states K (from 2 to 6). At last, the ability of the fitted models in forecasting, modelling extreme values, and reproducing the stylized facts of daily returns series, is investigated (Section 5.3).

5.1 Data and Basic Statistics

We use $T=3392$ daily returns calculated for NASDAQ and STOXX50, spanning the period from January 2, 2003, to June 22, 2016. For each stock market index, daily returns are computed as $X_t = [\ln(P_t) - \ln(P_{t-1})] \times 100$, where P_t is the closing price on day t (the closing prices used to compute the daily returns are downloadable from http://finance.yahoo.com/). As we can note from Figure 1, data present different volatility periods, reflecting rapid (positive and negative) peak bursts at consecutive days and more stable periods.

Table 1 provides descriptive statistics for the NASDAQ and STOXX50 series. Both series are leptokurtic and the Jarque–Bera statistic—computed via the `jarque.bera.test()` function in the **tseries** package (Trapletti and Hornik, 2017)—confirms the departure from univariate normality at the 1‰ level. In Table 2, we investigate bivariate normality. Based on Mardia's bivariate skewness and kurtosis statistics—computed via the function `mardiaTest()` in the **MVN** package (Korkmaz, Goksuluk, and Zararsiz, 2015)—we reject the null hypothesis of bivariate normality at any reasonable significance level.

Similar conclusions can be drawn from the examination of the Q–Q plots in Figure 2: deviations from the benchmark straight line indicate possible departure from univariate [Figures 2(a)–2(b)] and bivariate [Figure 2(c)] normal distributions. The χ^2 Q–Q plot in Figure 2(c) is obtained via the function `hzTest()` in the **MVN** package.

5.2 Model Fitting and Comparison

As a first step of the empirical analysis, we perform model selection by comparing HSMMs based on the multivariate normal (N) and multivariate LN as EDs, and on the geometric, shifted Poisson and gamma as SDs. The combination of these choices gives rise to six different models, with those based on the geometric SD being HMMs (as discussed in Section 2). Each competing model is fitted for a number of states K varying from 2 to 6, resulting in a total of 30 fitted models.

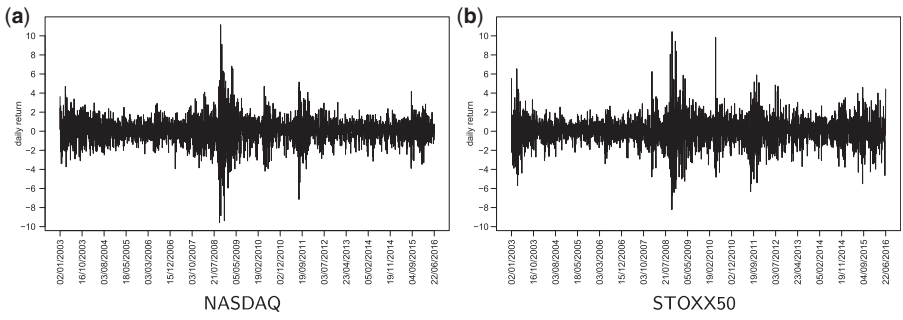

Figure 1. (a) NASDAQ and (b) STOXX50 daily returns from January 2, 2003, to June 22, 2016.

Table 1. Descriptive statistics, and Jarque–Bera normality tests (with *p*-values in brackets), for NASDAQ and STOXX50 series separately considered

	Mean	Std. dev	Skewness	Excess kurtosis	Jarque–Bera normality test
NASDAQ	0.038	1.328	−0.215	6.566	6119.9 (<0.001)
STOXX50	0.014	1.439	−0.055	5.039	3591.0 (<0.001)

Table 2. Mardia's bivariate normality test for the pair (*NASDAQ*, *STOXX*50)

	Statistic	χ^2-value	*p*-value
Skewness	0.206	116.797	<0.001
Kurtosis	23.263	111.119	<0.001

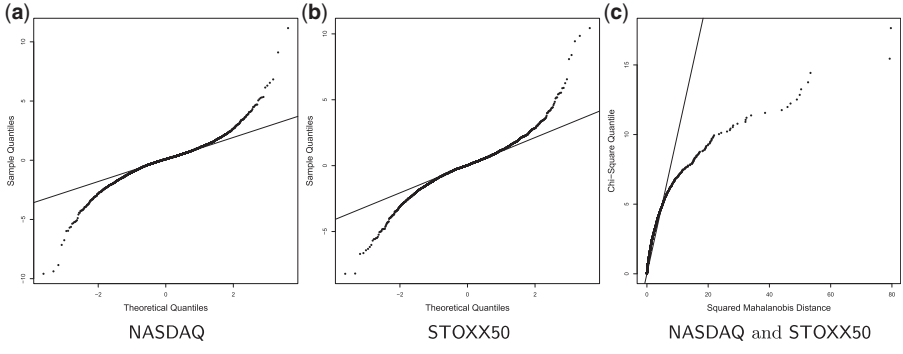

Figure 2. Univariate (normal) Q-Q plots for NASDAQ (a) and STOXX50 (b) daily returns, and bivariate (χ^2) Q–Q plot for NASDAQ and STOXX50 jointly considered (c).

We use three widely-used model selection criteria, namely the AIC, BIC, and ICL, to choose K; for the alternative use of likelihood-ratio tests, see Punzo, Browne, and McNicholas et al. (2016). The bold values in Table 3 highlight the best values of the considered criteria for each combination of sojourn and EDs. Note that AIC, BIC, and ICL as

Table 3. Model selection: log-likelihood, AIC, BIC, and ICL values for different HMMs and HSMMs with varying number of hidden states

	loglik	AIC	BIC	ICL	loglik	AIC	BIC	ICL
	N-HMM				LN–HMM			
$K=2$	−10242.97	−20511.94	−20591.62	−20743.55	−10148.03	−20326.05	−20417.99	−20523.15
$K=3$	−10047.82	−20141.65	−20282.62	**−20564.23**	−10014.82	−20081.64	**−20241.00**	**−20477.83**
$K=4$	−9997.81	−20065.63	**−20280.15**	−21288.60	−9989.78	−20057.55	−20296.59	−20869.06
$K=5$	−9957.91	−20013.83	−20314.16	−21072.51	−9938.40	−19984.81	−20315.78	−21137.15
$K=6$	−9900.81	**−19931.61**	−20330.01	−21318.37	−9891.57	**−19925.14**	−20360.31	−21261.74
	N-HSMM (shifted Poisson)				LN–HSMM (shifted Poisson)			
$K=2$	−10706.49	−21442.99	−21534.93	−22853.95	−10533.42	−21100.85	−21205.04	−22412.36
$K=3$	−10216.49	−20484.98	−20644.34	−22429.31	−10120.73	−20299.45	−20477.20	−21428.75
$K=4$	−10034.36	−20146.72	**−20385.76**	**−21407.48**	−10007.17	−20100.35	**−20363.90**	**−21236.97**
$K=5$	−9989.13	−20086.27	−20417.24	−21457.15	−9955.09	−20028.19	−20389.81	−21278.40
$K=6$	−9949.30	**−20040.60**	−20475.77	−21784.13	−9935.30	**−20024.59**	−20496.54	−21699.86
	N-HSMM (gamma)				LN–HSMM (gamma)			
$K=2$	−10246.02	−20522.04	−20613.98	−20721.21	−10148.10	−20330.19	−20434.39	−20533.06
$K=3$	−10049.66	−20151.32	−20310.68	−20576.63	−10013.96	−20085.93	**−20263.68**	**−20506.78**
$K=4$	−9993.94	−20065.89	**−20304.93**	**−20564.64**	−9984.26	−20054.53	−20318.08	−20556.54
$K=5$	−9939.53	−19987.05	−20318.03	−20565.60	−9922.46	**−19962.91**	−20324.53	−20544.42
$K=6$	−9920.25	**−19982.49**	−20417.66	−20706.26	−9909.18	−19972.36	−20444.31	−20728.87

used herein, are greater-is-better criteria. The AIC selects $K = 5$ states for the LN–HSMM with a gamma SD, and $K = 6$ states for all the other fitted models. However, it is well-known (see, e.g., Pohle et al., 2017) that AIC is unsatisfactory because it overestimates the number of mixture components and, for this reason, we will not consider it hereafter. BIC and ICL are more aligned in the choice of K, with the only difference being in the N-HMM, where the BIC selects $K = 4$ while the ICL $K = 3$. This is not surprising since the ICL further penalizes the BIC for the fuzziness of the classification (see, e.g., Biernacki, Celeux, and Govaert, 2000 and Baudry et al., 2010). Following this consideration and looking at Table 3, we can note that the LN–HMM with $K = 3$ states is the best fitted model according to both the considered criteria (BIC $= -20241.00$ and ICL $= -20477.83$). The second best is the LN–HSMM with $K = 3$ states and a gamma SD (BIC $= -20263.68$ and ICL $= -20506.78$).

For these two best models, Tables 4 and 5 show parameter estimates along with standard errors computed by using the parametric bootstrap technique illustrated in Section 3.3 where we choose $B = 500$ bootstrapped samples. For both the models we identify three different regimes with different levels of return: low, medium, and high. This aspect is highlighted by the estimated conditional means which are positive, almost zero, and negative, respectively. Moreover, it is well known that positive return shocks generate less volatility than negative return shocks (Engle and Ng, 1993). This stylized fact is corroborated by the estimated conditional means and covariance matrices: on the one hand, the estimated bivariate mean $\hat{\mu}_3$, having negative values (returns), is related to $\hat{\Sigma}_3$ which is the matrix with the highest values on the diagonal (i.e., with the highest trace); on the other hand, the bivariate mean $\hat{\mu}_1$, having positive values (returns), is related to $\hat{\Sigma}_1$ which is the matrix with the lowest values on the diagonal. The estimated parameters $\hat{\beta}_1 = 1.809$, $\hat{\beta}_2 = 0.450$, and $\hat{\beta}_3 = 2.466$ for the LN–HMM, and $\hat{\beta}_1 = 1.859$, $\hat{\beta}_2 = 0.419$, and $\hat{\beta}_3 = 2.416$ for the LN–HSMM with a gamma SD, indicate an excess of kurtosis in all EDs (compared to the normal case).

5.3 Further Aspects

Various aspects of the fitted models, such as the ability of reproducing the extreme values of the daily returns (Section 5.3.1), the capability to forecast the next daily value (Section 5.3.2), and the appropriateness in reproducing some stylized facts of stock returns (Section 5.3.3), are going to be analyzed. With this aim, some of the fitted models with $K = 3$ hidden states are compared: HMMs based on normal and LN EDs (respectively, denoted as N-HMM and LN–HMM), and HSMMs with a gamma SD and normal and LN EDs (respectively, denoted, for simplicity, as N-HSMM and LN–HSMM).

5.3.1 Extreme values modeling

For each of the P available time series, according to Breunig, Najarian, and Pagan (2003) and Bulla (2011), the capability of capturing extreme values can be evaluated by the test statistic

$$W = \frac{(\phi - \hat{\phi})^2}{\mathrm{Var}(\phi)}, \tag{21}$$

which conveniently compares ϕ and $\hat{\phi}$; these quantities denote, respectively, the empirical and estimated (according to the chosen model) proportions of observations lying outside the interval $(-r, r)$. It is possible to show that W is asymptotically distributed as a χ^2_1. In the fashion of Breunig, Najarian, and Pagan (2003) and Bulla (2011), we compute $\hat{\phi}$ and $\mathrm{Var}(\phi)$ by using 1000 simulations.

Table 4. Parameters estimates (with bootstrapped standard errors, over 500 replications, in brackets) of the LN–HMM with $K = 3$ states

Π	$\begin{bmatrix} 0.971 & 0.029 & 0.001 \\ (0.031) & (0.029) & (0.003) \\ 0.041 & 0.947 & 0.013 \\ (0.086) & (0.111) & (0.056) \\ 0.002 & 0.031 & 0.966 \\ (0.007) & (0.056) & (0.060) \end{bmatrix}$		
	$k = 1$	$k = 2$	$k = 3$
μ_k	$\begin{bmatrix} 0.146 \\ (0.066) \\ 0.101 \\ (0.076) \end{bmatrix}$	$\begin{bmatrix} -0.026 \\ (0.179) \\ -0.050 \\ (0.229) \end{bmatrix}$	$\begin{bmatrix} -0.170 \\ (0.172) \\ -0.145 \\ (0.215) \end{bmatrix}$
Σ_k	$\begin{bmatrix} 0.593 & 0.332 \\ (0.368) & (0.249) \\ 0.332 & 0.719 \\ (0.249) & (0.424) \end{bmatrix}$	$\begin{bmatrix} 1.583 & 0.978 \\ (0.403) & (0.197) \\ 0.978 & 1.965 \\ (0.197) & (0.366) \end{bmatrix}$	$\begin{bmatrix} 6.915 & 4.316 \\ (0.950) & (0.800) \\ 4.316 & 7.837 \\ (0.800) & (1.147) \end{bmatrix}$
β_k	$\begin{bmatrix} 1.809 \\ (0.547) \end{bmatrix}$	$\begin{bmatrix} 0.450 \\ (0.350) \end{bmatrix}$	$\begin{bmatrix} 2.466 \\ (0.672) \end{bmatrix}$

Table 5. Parameters estimates (with bootstrapped standard errors, over 500 replications, in brackets) of the LN–HSMM with $K = 3$ states and a gamma SD

Π	$\begin{bmatrix} 0.000 & 1.000 & 0.000 \\ (0.000) & (0.023) & (0.023) \\ 0.799 & 0.000 & 0.201 \\ (0.068) & (0.000) & (0.068) \\ 0.000 & 1.000 & 0.000 \\ (0.080) & (0.080) & (0.000) \end{bmatrix}$		
	$k = 1$	$k = 2$	$k = 3$
μ_k	$\begin{bmatrix} 0.152 \\ (0.020) \\ 0.107 \\ (0.025) \end{bmatrix}$	$\begin{bmatrix} -0.028 \\ (0.038) \\ -0.056 \\ (0.041) \end{bmatrix}$	$\begin{bmatrix} -0.186 \\ (0.136) \\ -0.156 \\ (0.147) \end{bmatrix}$
Σ_k	$\begin{bmatrix} 0.572 & 0.315 \\ (0.057) & (0.045) \\ 0.315 & 0.686 \\ (0.045) & (0.076) \end{bmatrix}$	$\begin{bmatrix} 1.562 & 0.997 \\ (0.084) & (0.072) \\ 0.997 & 1.986 \\ (0.072) & (0.111) \end{bmatrix}$	$\begin{bmatrix} 7.070 & 4.400 \\ (0.666) & (0.513) \\ 4.400 & 7.978 \\ (0.513) & (0.692) \end{bmatrix}$
β_k	$\begin{bmatrix} 1.859 \\ (0.305) \end{bmatrix}$	$\begin{bmatrix} 0.419 \\ (0.283) \end{bmatrix}$	$\begin{bmatrix} 2.416 \\ (0.640) \end{bmatrix}$
λ_k	$\begin{bmatrix} 0.841 \\ (6.795) \end{bmatrix}$	$\begin{bmatrix} 2.286 \\ (2.243) \end{bmatrix}$	$\begin{bmatrix} 1.026 \\ (1.671) \end{bmatrix}$
ζ_k	$\begin{bmatrix} 49.219 \\ (14.943) \end{bmatrix}$	$\begin{bmatrix} 11.996 \\ (5.998) \end{bmatrix}$	$\begin{bmatrix} 36.761 \\ (32.853) \end{bmatrix}$

Results for different values of r are shown in Table 6 where we report the estimates of ϕ, $\hat{\phi}$, and p-values associated to the test statistic W. The considered models are not rejected, regardless of the considered value of r, indicating their good performance in reproducing the extreme values of the daily return series. Using the p-values, we can also rank the competing models. We find that higher p-values (hence a better performance) are provided by the models with a LN ED, which are also the best two models according to BIC and ICL (cf. Table 3).

5.3.2 Out-of-sample analysis

We now investigate whether the competing models can be used to forecast future observations. In particular, we implement a rolling out-of-sample forecasting exercise where the value of the series at at time $T+1$ is predicted, for $T = 3000, \ldots, 3391$ and using 3000 of the most recent observations (rolling window). Hence, we derive 1-step ahead forecast for the period from December 2nd, 2014, to June 22nd, 2016. On each of the 392 training sets, we apply the procedure described in Section 4 (with $\alpha = 0.05$ and $M = 10{,}000$) and compare the MAP predictions with the empirical data (see Figure B.2 of the Supplementary Material). Regardless of the considered stock and/or model, the predicted values (middle line in the plots) are close to zero, while the predicted intervals approximately assume three widths, one for each state (for the LN–HMM and LN–HSMM, refer to the estimated covariance matrices in Tables 4 and 5).

Exceedances of the observed series with respect to the MAP predicted intervals are summarized in Table 7. As we can see, the fitted models lead to an excessive number of exceedances, especially on the STOXX50 series. In other words, the considered models underestimate the VaR values more often than was specified. This result is not surprising and in line with the findings of Zucchini and Neumann (2001). This downward behavior is particularly intensified for the models with a normal ED compared to the 2.5% of the lowest values of the STOXX50 series, where we have an excessive number of 36 (N-HMM) and 35 (N-HSMM) exceedances. This result further confirms the slight better performance of the models based on a LN ED.

To further compare the out-of-sample conditional quantile forecasts, we use the tick-loss function; see, for example, Giacomini and Komunjer (2005) and Brownlees and Gallo (2010). The estimated value of the tick-loss function for each series (NASDAQ and STOXX50), and each fitted model, is given in squared brackets in Table 7. The conditional quantile forecasts are quite similar, regardless of the considered stock market.

5.3.3 Stylized facts

Most of the empirical researches on modelling stock returns focuses on typical stylized facts which can have temporal or distributional nature and can be observed in both the univariate and multivariate contexts (Granger and Ding, 1995a, 1995b, Rydén, Teräsvirta, and Ösbrink, 1998; Chakraborti et al., 2011; McNeil, Frey, and Embrechts, 2015; Pfaff, 2016). Classical univariate stylized facts for a stock return series $\{X_t;\, t = 1, \ldots, T\}$ are:

U1. returns are not autocorrelated (except for, possibly, the first lag);

U2. the autocorrelation function (ACF) of the absolute and quadratic returns are slow decaying, and $\text{cor}(|X_t|, |X_{t-l}|) > \text{cor}(X_t^2, X_{t-l}^2)$, with l being a positive integer denoting the lag;

Table 6. Empirical (ϕ) and estimated ($\hat{\phi}$) proportions of observations lying outside the interval $(-r, r)$, and p-values from the test statistic W

	r	NASDAQ		STOXX50	
		ϕ		ϕ	
Empirical	1	0.336		0.367	
	1.5	0.182		0.217	
	2	0.104		0.128	
	2.5	0.058		0.075	
		$\hat{\phi}$	p-value	$\hat{\phi}$	p-value
N-HMM	1	0.346	0.583	0.385	0.380
	1.5	0.182	0.879	0.221	0.856
	2	0.103	0.922	0.129	0.941
	2.5	0.060	0.863	0.077	0.830
LN–HMM	1	0.337	0.914	0.376	0.705
	1.5	0.185	0.921	0.218	0.964
	2	0.103	0.950	0.129	0.965
	2.5	0.060	0.909	0.076	0.907
N-HSMM	1	0.348	0.524	0.387	0.359
	1.5	0.187	0.816	0.223	0.809
	2	0.104	0.997	0.131	0.872
	2.5	0.061	0.808	0.079	0.764
LN–HSMM	1	0.336	0.961	0.376	0.751
	1.5	0.184	0.964	0.217	0.994
	2	0.102	0.907	0.127	0.986
	2.5	0.059	0.966	0.075	0.966

Table 7. Percentage of exceedances of the observed time series with respect to the MAP prediction intervals from the fitted models ($1 - \alpha = 0.95$)

	NASDAQ		STOXX50	
	$< 2.5\%$	$> 97.5\%$	$< 2.5\%$	$> 97.5\%$
N-HMM	5.61% (22)	1.53% (6)	9.18% (36)	5.87% (23)
	[0.085]	[0.062]	[0.142]	[0.106]
LN–HMM	4.85% (19)	1.28% (5)	7.14% (28)	5.61% (22)
	[0.085]	[0.063]	[0.137]	[0.109]
N-HSMM	4.85% (19)	1.79% (7)	8.93% (35)	5.61% (22)
	[0.084]	[0.062]	[0.137]	[0.104]
LN–HSMM	4.59% (18)	1.28% (5)	7.40% (29)	5.61% (22)
	[0.084]	[0.063]	[0.136]	[0.108]

Note: The number of exceedances is given in round brackets while the estimated tick loss function is given in squared brackets

U3. autocorrelations of powers of absolute returns are highest at power one;
U4. autocorrelations of sign(X_t) are negligibly small;
U5. volatility appears to vary over time;
U6. $|X_t|$ and sign(X_t) are independent;
U7. the marginal distribution of $|X_t|$ is exponential;
U8. $|X_t|$ has the same mean and variance;
U9. return series are leptokurtic or heavy-tailed.

In the more general multivariate context, common stylized facts (see, e.g., McNeil, Frey, and Embrechts, 2015 and Pfaff, 2016) are:

M1. multivariate return series show little evidence of cross-correlation, except for contemporaneous returns;
M2. multivariate series of absolute returns show profound evidence of cross-correlation.

In what follows, we first verify if the above stylized facts hold for the NASDAQ and STOXX50 return series. Then, we investigate if they are also reproduced by the fitted models. In the following, when we will write about the results from the fitted models, we will refer to averages of the quantities of interest computed over 1000 bootstrapped series from each fitted model, according to a parametric bootstrap approach.

U1. We calculate the p-value of the test for the absence of autocorrelation of lag l, $l = 1, \ldots, 30$. If $\hat{\rho}_l$ is the sample autocorrelation at lag l, the p-value is calculated as $2\Phi(-\frac{|\hat{\rho}_l|}{\sqrt{T}})$, where $\Phi(\cdot)$ denotes the standard normal distribution function (see, e.g., Romano and Thombs, 1996 and Bagnato, Punzo, and Nicolis, 2012).
Figure 3 compares, separately for each series, "empirical" and "estimated" p-values. While a certain significance appears for the tests at lags 1 and 2 of NASDAQ, returns from the fitted models are not autocorrelated, regardless of the considered series and lag. This is in line with the requirements of **U1**.

U2. We first evaluate the slow decay of the sample and estimated ACF for absolute and squared series (see Figure 4). We find that the decay is slower for the models based on LN EDs. These models seem also to better fit the sample ACFs.
As a second aspect of **U2**, we evaluate if the difference $\text{cor}(|X_t|, |X_{t-l}|) - \text{cor}(X_t^2, X_{t-l}^2)$ is positive. For this purpose Figure 5 displays, separately for each series (NASDAQ and STOXX50), the empirical and estimated differences. The requirement on the difference does not seem to hold for the first lags of the observed series. Nevertheless, the estimated differences are always positive and show a decaying behavior. Interestingly, the positive difference is higher for the model based on LN EDs, regardless of the considered series and lag.

U3. This property is also known as the Taylor effect. For its evaluation, we compute $p_{1,\max}$, that is, the power of absolute returns which maximizes the autocorrelation of lag 1. This is done separately for each series (NASDAQ and STOXX50). The maximum is searched for over a grid of values for the power ranging from 0.1 to 2.0, with increments of 0.05. With this aim, we use the `teffectPlot()` function included in the **fBasics** package. Empirical and estimated values of $p_{1,\max}$ are all >1, with the empirical values of $p_{1,\max}$ being the greatest ones. Moreover, the results summarized in Table 8 highlight that $p_{1,\max}$ is closer to the empirical $p_{1,\max}$ for the models based on normal EDs, while $p_{1,\max}$ is closer to 1 for the models having LN EDs.

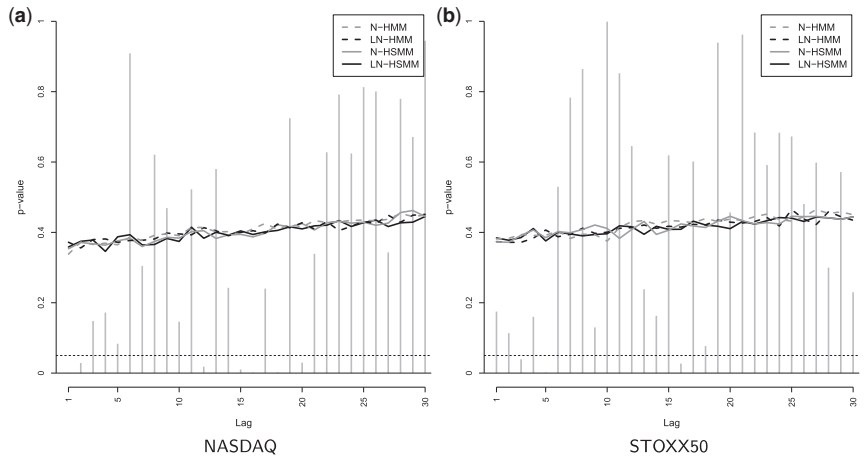

Figure 3. U1. *p*-values of the tests for the absence of autocorrelation of lag l, $l = 1, \ldots, 30$. Vertical bars denote *p*-values for the raw series while solid and dotted lines indicate *p*-values from the fitted models. An horizontal dashed line, at height $\alpha = 0.05$, marks the benchmark significance level.

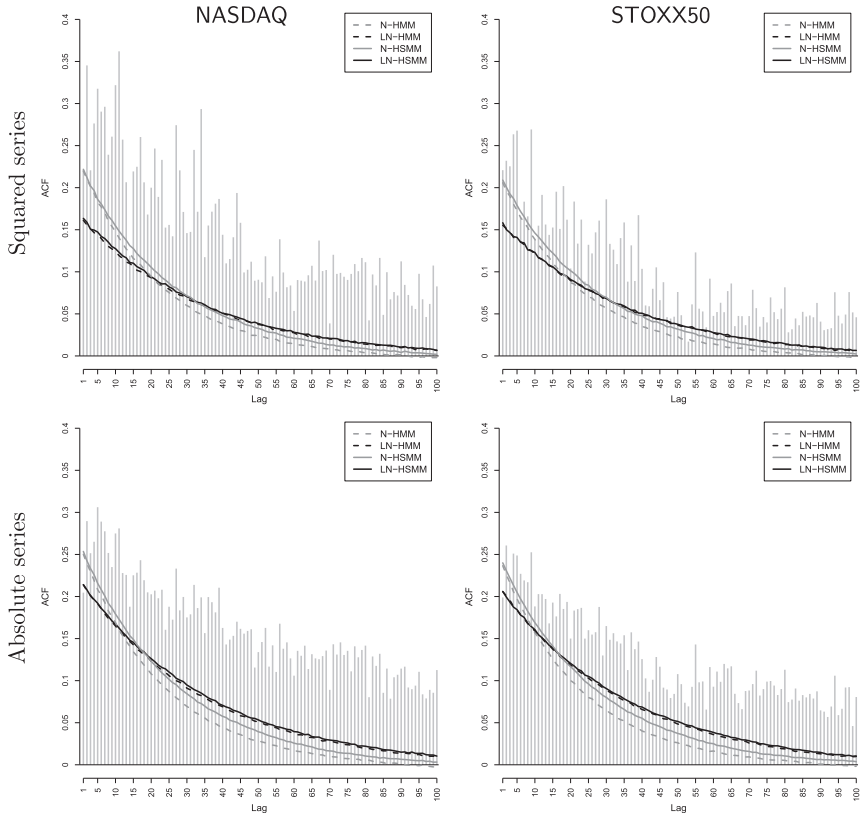

Figure 4. U2. Sample and estimated ACF of absolute (row 1) and squared returns (row 2).

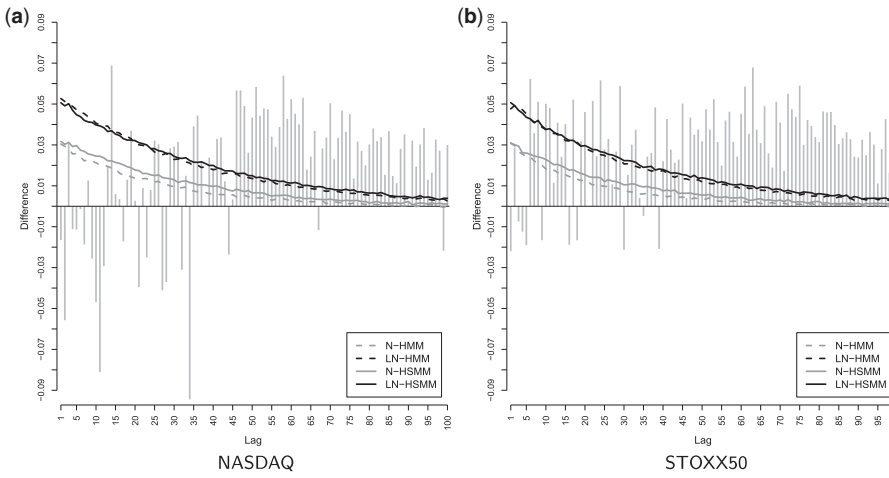

Figure 5. U2. Sample and estimated differences $\text{cor}(|X_t|, |X_{t-l}|) - \text{cor}(X_t^2, X_{t-l}^2)$, $l = 1, \ldots, 100$.

U4. We investigate **U4** in the same way we inspected **U1** but considering, instead of the raw series, the $\text{sign}(X_t)$ series. The results reported in Figure 6 confirm the validity of **U4**.

U5. This stylized fact is widely known as "volatility clustering," as "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes" (Mandelbrot, 1963). This phenomenon is also called conditional heteroscedasticity (see, e.g., Wang et al., 2005), and can be modeled by the classical ARCH model. Accordingly, if the time series exhibits autoregressive conditionally heteroscedasticity, we say it has the ARCH effect. To investigate for an ARCH effect, we apply the McLeod–Li test for conditional heteroscedascity (McLeod and Li, 1983) via the `McLeod.Li.test()` function of the **TSA** package (Chan and Ripley, 2012). The null hypothesis of absence of ARCH effects is always rejected with p-values which are practically null regardless of the considered lag and model; these p-values are not reported here for the sake of brevity. Based on these results, we can say that **U5** is fulfilled.

To further investigate **U5**, we focus on the clustering structure provided by the estimated models based on LN EDs. According to the parameters estimates in Tables 4 and 5, market states mainly differ in volatility, that is, different volatility clusters are identified. This becomes clearer by looking at Figure 7, where observations are clustered by using local decoding. In other words, as a by-product of the estimation step, we determine the state which is the most likely for each time. Persistence and different sojourn times can also be graphically identified, confirming the presence of volatility clustering (see also Tables 4 and 5).

U6. For simplicity, we only study whether or not the series $|X_t|$ and $\text{sign}(X_t)$ are correlated. We calculate the p-value of the test for absence of correlation between the series $|X_t|$ and $\text{sign}(X_t)$. The results are reported in Table 9. While for NASDAQ the null of absence of

Table 8. U3. Absolute returns: empirical and estimated power maximizing the first-order autocorrelation ($p_{1,\max}$)

	Empirical	N-HMM	LN–HMM	N-HSMM	LN–HSMM
NASDAQ	1.35	1.26	1.18	1.25	1.18
STOXX50	1.65	1.32	1.21	1.32	1.23

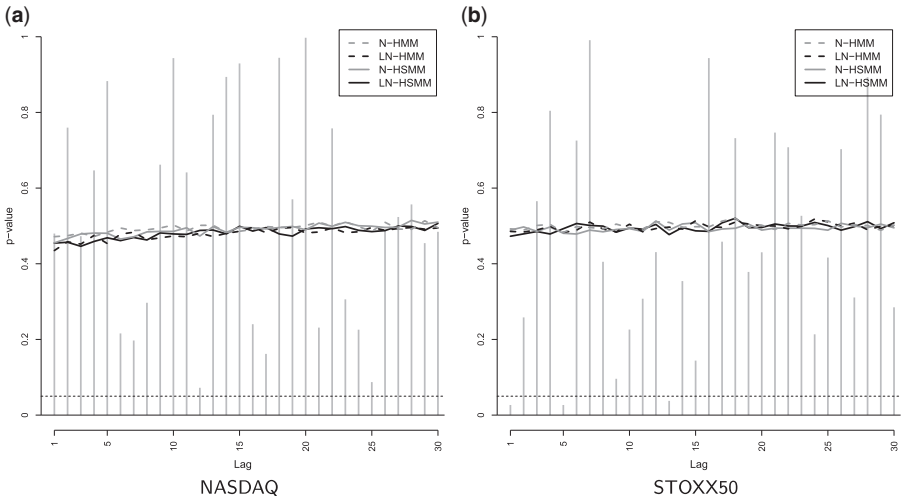

Figure 6. U4. p-values of the tests for the absence of autocorrelation of lag l, $l = 1, \ldots, 30$, for the sign(X_t) series of the NASDAQ (a) and STOXX50 (b) daily returns. Vertical bars denote p-values for the empirical series while solid and dotted lines indicate p-values from the fitted models. An horizontal dashed line, at height $\alpha = 0.05$, marks the benchmark significance level.

correlation is rejected, for STOXX50 the null is not. Estimated p-values in Table 9 highlight how **U6** holds for all the fitted models, regardless of the considered series.

U7. To compare the distribution of empirical and bootstrapped absolute series with the exponential distribution, we apply the Lilliefors-corrected Kolmogorov–Smirnov test (LKS) which results to be suitable when population parameters are unknown and must be estimated based on the sample (Lilliefors, 1967). For this aim, we use the `LcKS()` function of the **KScorrect** package. Table 10 reports the obtained p-values. The goodness-of-fit of the exponential model (i.e., the validity of **U7**) is rather poor for all the series, including also the empirical ones (with respect to $\alpha = 0.05$).

U8. Table 11 reports empirical and estimated mean-standard deviation ratios for the absolute series. The bootstrap confidence interval, with confidence level $1 - \alpha = 0.95$, is also reported in round brackets. To compute this interval, the empirical bootstrap distribution of the mean-standard deviation ratio is preliminarily constructed over 1000 bootstrap resamples with replacement from the absolute series. As we can note from the obtained results, all the empirical and estimated mean-standard deviation ratios are close to one; **U8** is also corroborated by the bootstrap confidence intervals which always contain the value one (the value under the null).

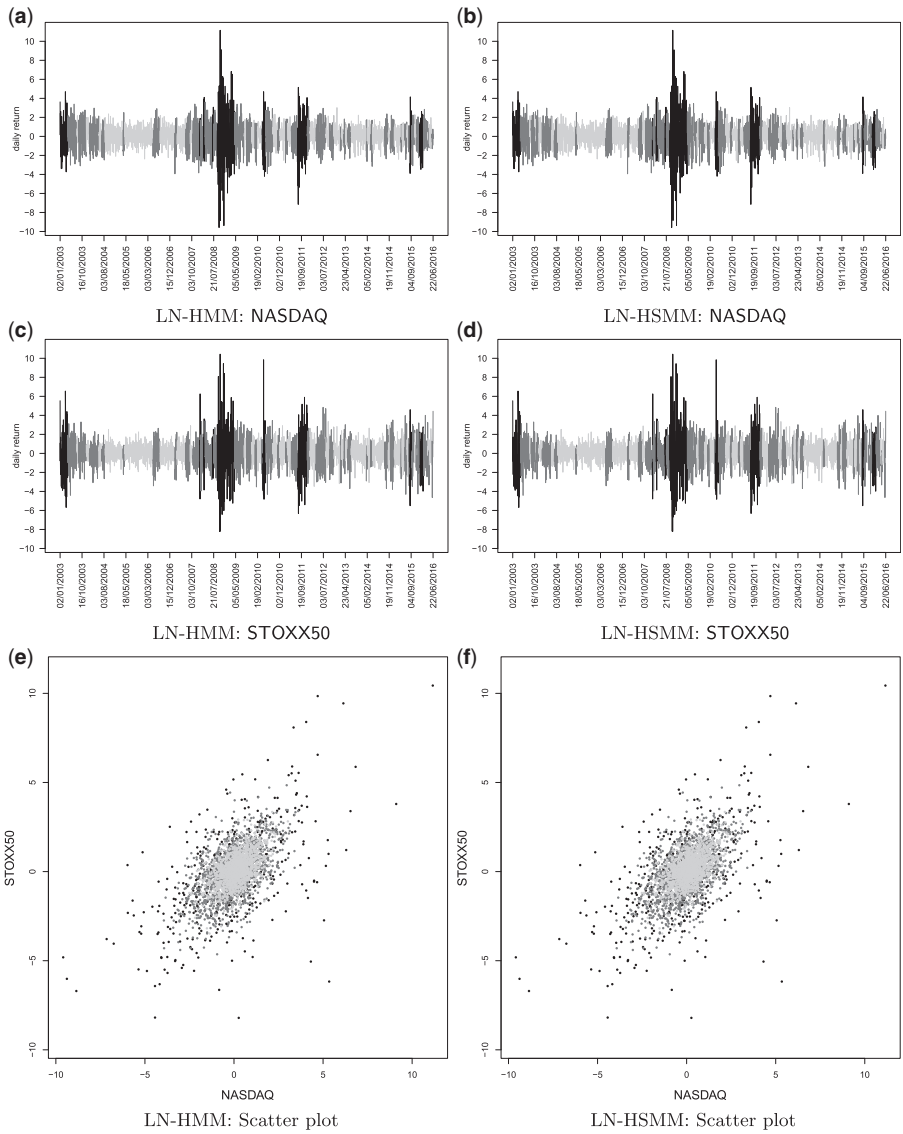

Figure 7. U5. Classification results: State 1 is in light gray; State 2 is in dark gray; State 3 is in black.

U9. Table 12 reports empirical and estimated kurtosis statistics along with p-values (in round brackets) of Mardia's test of mesokurtosis computed by the `mardia()` function of the **psych** package (Revelle, 2017). According to these p-values, the null hypothesis of mesokurtosis is rejected, at any significance level, in all of the considered cases, so confirming the validity of **U9**. However, the kurtosis of the models based on LN EDs matches better the empirical kurtosis.

Table 9. U6. p-values of the tests for the absence of correlation between the series $|X_t|$ and $\text{sign}(X_t)$

	Empirical	N-HMM	LN–HMM	N-HSMM	LN–HSMM
NASDAQ	0.010	0.335	0.260	0.330	0.252
STOXX50	0.634	0.293	0.266	0.299	0.264

Table 10. U7. Absolute returns: empirical and estimated p-values of the LKS test. The exponential distribution is used as null model

	Empirical	N-HMM	LN–HMM	N-HSMM	LN–HSMM
NASDAQ	0.014	0.000	0.001	0.000	0.001
STOXX50	0.031	0.000	0.002	0.000	0.001

Table 11. U8. Absolute returns: empirical and estimated mean-standard deviation ratios. Bootstrap confidence intervals (over 1000 bootstrap resamples and with confidence level 0.95) are reported in round brackets

	Empirical	N-HMM	LN–HMM	N-HSMM	LN–HSMM
NASDAQ	0.961	1.015	0.987	1.014	0.991
	(0.913, 1.012)	(0.977, 1.061)	(0.944, 1.034)	(0.975, 1.056)	(0.947, 1.037)
STOXX50	0.980	1.034	1.001	1.033	1.004
	(0.936, 1.026)	(0.997, 1.076)	(0.960, 1.048)	(0.996, 1.073)	(0.961, 1.048)

Table 12. U9. Returns: empirical and estimated kurtoses along with p-values (in round brackets) from the Mardia test of mesokurtosis

	Empirical	N-HMM	LN–HMM	N-HSMM	LN–HSMM
NASDAQ	9.566	7.403	8.568	7.404	8.496
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
STOXX50	8.039	6.662	7.949	6.668	7.916
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

M1. This stylized fact is a multivariate extension of **U1**. Therefore, in analogy with **U1**, we use the test for the absence of correlation between the NASDAQ return on day $t + h$ and the STOXX50 return on day t, with h assuming integer values from -15 to 15. Figure 8(a) reports the empirical (bar plot) and estimated (solid and dotted lines) p-values obtained for each choice of h. The estimated p-values are lower than 0.05 only for $h = 0$, confirming that the fitted models well-reproduce **M1**.

M2. This stylized fact, which is an extension of **U2**, is investigated in the same way of **M1** but considering, instead of raw series, the absolute series. Figure 8(b) shows the

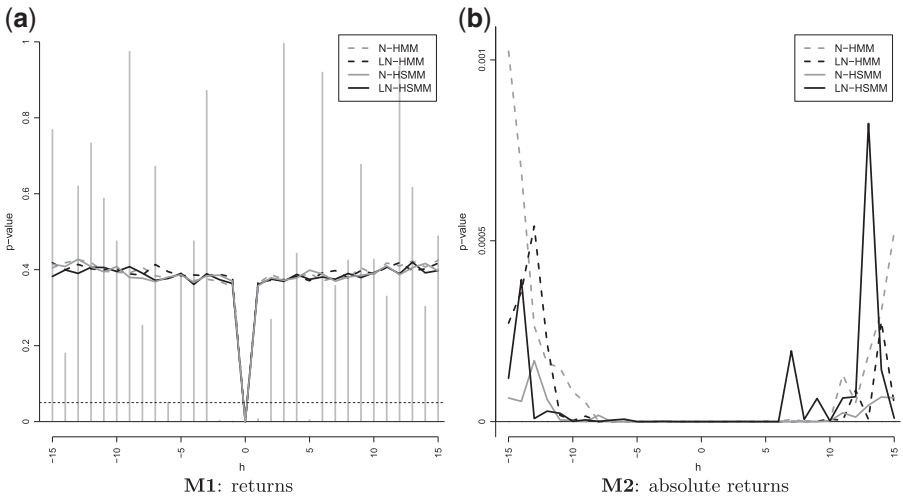

Figure 8. M1–M2. *p*-values of the tests for the absence of cross-correlation between returns (a) and absolute returns (b). Vertical bars denote *p*-values for the observed series while solid and dotted lines indicate *p*-values from the fitted models. A horizontal dashed line, at height $\alpha = 0.05$, marks the benchmark significance level.

obtained *p*-values. The estimated *p*-values are very close to zero, and this confirms that **M2** is valid for all the considered models. The empirical *p*-values are so close to zero that they are not visible on the plot.

6 Conclusions

The paper introduces new multivariate hidden Markov and hidden semi-Markov models tailored to model financial returns behavior. Although HSMMs have been successfully used in a variety of contexts, this is the first time, up to our knowledge, that they are employed in a multivariate daily returns setting. To handle the leptokurtosis of returns, we have considered the recently proposed multivariate LN as ED. We have outlined ML estimation of the parameters via the EM algorithm and a parametric bootstrap procedure has been implemented to obtain the standard errors. Finally, a MAP procedure has been introduced for forecasting purposes.

The parameter recovery of the fitting algorithm has been evaluated by artificial time series. The proposed models have been fitted to a publicly available financial bivariate time series and their performance has been measured and compared to HMM based on the multivariate normal ED. Further aspects, such as the ability of the fitted models in forecasting, modeling extreme values, and reproducing the stylized facts of daily returns series, have been evaluated too. Empirically, we find evidence that the introduced models are preferred by the commonly considered model selection criteria and they well reproduce most of the temporal and distributional properties of stock returns. Moreover, the proposed models reproduce extreme values and provide better 1-step ahead forecasts than the considered competing models. The reason for this improved performance may lie in the form of

the tails of the LN distribution which allows for greater robustness towards short periods of lower (or higher) volatility.

Supplementary Data

Supplementary data are available at *Journal of Financial Econometrics* online.

References

Abanto-Valle, C. A., R. Langrock, M.-H. Chen, and M. V. Cardoso. 2017. Maximum Likelihood Estimation for Stochastic Volatility in Mean Models with Heavy-Tailed Distributions. *Applied Stochastic Models in Business and Industry* 33: 394–408.

Ang, A., and A. Timmermann. 2012. Regime Changes and Financial Markets. *Annual Review of Financial Economics* 4: 313–337.

Bagnato, L., and A. Punzo. 2013. Finite Mixtures of Unimodal Beta and Gamma Densities and the *k*-Bumps Algorithm. *Computational Statistics* 28: 1571–1597.

Bagnato, L., A. Punzo, and O. Nicolis. 2012. The Autodependogram: A Graphical Device to Investigate Serial Dependences. *Journal of Time Series Analysis* 33: 233–254.

Bagnato, L., A. Punzo, and M. G. Zoia. 2017. The Multivariate Leptokurtic-Normal Distribution and Its Application in Model-Based Clustering. *Canadian Journal of Statistics* 45: 95–119.

Bartolucci, F., and A. Farcomeni. 2010. A Note on the Mixture Transition Distribution and Hidden Markov Models. *Journal of Time Series Analysis* 31: 132–138.

Baudry, J.-P., A. E. Raftery, G. Celeux, K. Lo, and R. Gottardo. 2010. Combining Mixture Components for Clustering. *Journal of Computational and Graphical Statistics* 19: 332–353.

Baum, L. E., T. Petrie, G. Soules, and N. Weiss. 1970. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The Annals of Mathematical Statistics* 41: 164–171.

Bernardi, M., A. Maruotti, and L. Petrella. 2017. Multiple Risk Measures for Multivariate Dynamic Heavy-Tailed Models. *Journal of Empirical Finance* 43: 1–32.

Biernacki, C., G. Celeux, and G. Govaert. 2000. Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22: 719–725.

Biernacki, C., G. Celeux, and G. Govaert. 2003. Choosing Starting Values for the EM Algorithm for Getting the Highest Likelihood in Multivariate Gaussian Mixture Models. *Computational Statistics & Data Analysis* 41: 561–575.

Breunig, R., S. Najarian, and A. Pagan. 2003. Specification Testing of Markov Switching Models. *Oxford Bulletin of Economics and Statistics* 65: 703–725.

Brownlees, C. T., and G. M. Gallo. 2010. Comparison of Volatility Measures: A Risk Management Perspective. *Journal of Financial Econometrics* 8: 29–56.

Bulla, J. 2011. Hidden Markov Models with *t* Components. Increased Persistence and Other Aspects. *Quantitative Finance* 11: 459–475.

Bulla, J., and I. Bulla. 2006. Stylized Facts of Financial Time Series and Hidden Semi-Markov Models. *Computational Statistics & Data Analysis* 51: 2192–2209.

Bulla, J., I. Bulla, and O. Nenadić. 2010. HSMM—an R Package for Analyzing Hidden Semi-Markov Models. *Computational Statistics & Data Analysis* 54: 611–619.

Cavaliere, G., M. Costa, and L. De Angelis. 2014. *Advances in Latent Variables, Chapter Investigating Stock Market Behavior Using Multivariate Markov-Switching Approach*, pp. 185–196. Berlin: Springer.

Chakraborti, A., I. M. Toke, M. Patriarca, and F. Abergel. 2011. Econophysics Review: I. Empirical Facts. *Quantitative Finance* 11: 991–1012.

Chan, K.-S., and B. Ripley. 2012. TSA: Time Series Analysis. *R Package Version 1.01 (2012-11-13)*.

Choi, S. C., and R. Wette. 1969. Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias. *Technometrics* 11: 683–690.

De Angelis, L., and L. Paas. 2013. A Dynamic Analysis of Stock Market Using a Hidden Markov Model. *Journal of Applied Statistics* 40: 1682–1700.

Delignette-Muller, M. L., and C. Dutang. 2015. fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software* 64: 1–34.

Delignette-Muller, M. L., C. Dutang, and A. Siberchicot. 2017. fitdistrplus: Help to Fit of a Parametric Distribution to Non-Censored or Censored Data. *R Package Version 1.0-9 (2017-03-24)*.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39: 1–38.

Elliott, R. J., L. Aggoun, and J. B. Moore. 2008. *Hidden Markov Models: Estimation and Control. Stochastic Modelling and Applied Probability*. New York: Springer.

Engle, R. F., and V. K. Ng. 1993. Measuring and Testing the Impact of News on Volatility. *The Journal of Finance* 48: 1749–1778.

Erlewein, C., R. Mamon, and M. Davison. 2011. An Examination of HMM-Based Investment Strategies for Asset Allocation. *Applied Stochastic Models in Business and Industry* 27: 204–221.

Forney, G. D. 1973. The Viterbi Algorithm. *Proceedings of the IEEE* 61: 268–278.

Frahm, G. 2004. Generalized Elliptical Distributions: Theory and Applications. PhD thesis, Universität zu Köln.

Gettinby, G. D., C. D. Sinclair, D. M. Power, and R. A. Brown. 2004. An Analysis of the Distribution of Extreme Share Returns in the UK from 1975 to 2000. *Journal of Business Finance & Accounting* 31: 607–646.

Giacomini, R., and I. Komunjer. 2005. Evaluation and Combination of Conditional Quantile Forecasts. *Journal of Business & Economic Statistics* 23: 416–431.

Granger, C. W. J., and Z. Ding. 1995a. Some Properties of Absolute Return: An Alternative Measure of Risk. *Annales D'Economie Et De Statistique* 40: 67–91.

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

Granger, C. W. J., S. Spear, and Z. Ding. 2000. "Stylized Facts on the Temporal and Distributional Properties of Absolute Returns: An Update." In W.-S. Chan, W. K. Li, and H. Tong (eds.), *Statistics and Finance: An Interface*, pp. 97–120. World Scientific. Proceedings of the Hong Kong International Workshop on Statistics in Finance. University of Hong Kong, 4–8 July 1999.

Guédon, Y. 2003. Estimating Hidden Semi-Markov Chains from Discrete Sequences. *Journal of Computational and Graphical Statistics* 12: 604–639.

Gupta, A., and B. Dhingra. 2012. "Stock Market Prediction Using Hidden Markov Models." In *Students Conference on Engineering and Systems (SCES)*, Allahabad, Uttar Pradesh, India, 16–18 March 2012, pp. 1–4. IEEE.

Hamilton, J. 1989. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* 57: 357–384.

Hamilton, J. 1990. Analysis of Time Series Subject to Changes in Regime. *Journal of Econometrics* 45: 39–70.

Hardy, M. 2001. A Regime Switching Model of Long Term Sotck Returns. *North American Acturial Journal* 2: 11–26.

Harris, R. D., and C. C. Küçüközmen. 2001. The Empirical Distribution of UK and US Stock Returns. *Journal of Business Finance & Accounting* 28: 715–740.

Hassan, M. R. 2009. A Combination of Hidden Markov Model and Fuzzy Model for Stock Market Forecasting. *Neurocomputing* 72: 3439–3446.

Johnson, N. L., A. W. Kemp, and S. Kotz. 2005. *Univariate Discrete Distributions*. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley.

Karlis, D., and E. Xekalaki. 2003. Choosing Initial Values for the EM Algorithm for Finite Mixtures. *Computational Statistics & Data Analysis* 41: 577–590.

Korkmaz, S., D. Goksuluk, and G. Zararsiz. 2015. *MVN: Multivariate Normality Tests. R Package Version 4.0*.

Langrock, R., I. L. MacDonald, and W. Zucchini. 2012. Some Nonstandard Stochastic Volatility Models and Their Estimation Using Structured Hidden Markov Models. *Journal of Empirical Finance* 19: 147–161.

Lilliefors, H. W. 1967. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association* 62: 399–402.

Mamon, R., and R. Elliott. 2007. *Hidden Markov Models in Finance*. New York: Springer.

Mamon, R., and R. Elliott. 2014. *Hidden Markov Models in Finance: Further Developments and Applications* (Volume II). New York: Springer.

Mandelbrot, B. 1963. The Variation of Certain Speculative Prices. *The Journal of Business* 36: 79–106.

Maruotti, A., and A. Punzo. 2017. Model-Based Time-Varying Clustering of Multivariate Longitudinal Data with Covariates and Outliers. *Computational Statistics & Data Analysis* 113: 475–496.

McLeod, A. I., and W. K. Li. 1983. Diagnostic Checking Arma Time Series Models Using Squared-Residual Autocorrelations. *Journal of Time Series Analysis* 4: 269–273.

McNeil, A. J., R. Frey, and P. Embrechts. 2015. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton, NJ: Princeton University Press.

Mergner, S., and J. Bulla. 2008. Time-Varying Beta Risk of Pan-European Industry Portfolios: A Comparison of Alternative Modeling Techniques. *The European Journal of Finance* 14: 771–802.

Nystrup, P., H. Madsen, and E. Lindstrm. 2017. Long Memory of Financial Time Series and Hidden Markov Models with Time-Varying Parameters. *Journal of Forecasting* 36: 989–1002.

O'Connell, J., and S. Højsgaard. 2011. Hidden Semi Markov Models for Multiple Observation Sequences: The mhsmm Package for R. *Journal of Statistical Software* 39: 1–22.

Perez-Quiros, G., and A. Timmermann. 2001. Business Cycle Asymmetries in Stock Returns: Evidence from Higher Order Moments and Conditional Densities. *Journal of Econometrics* 103: 259–306.

Pfaff, B. 2016. *Financial Risk Modelling and Portfolio Optimization with R*. Chichester, UK: John Wiley & Sons.

Pohle, J., R. Langrock, F. M. van Beest, and N. M. Schmidt. 2017. Selecting the Number of States in Hidden Markov Models: Pragmatic Solutions Illustrated Using Animal Movement. *Journal of Agricultural, Biological and Environmental Statistics* 22: 270–293.

Punzo, A., R. P. Browne, and P. D. McNicholas. 2016. Hypothesis Testing for Mixture Model Selection. *Journal of Statistical Computation and Simulation* 86: 2797–2818.

Punzo, A., and A. Maruotti. 2016. Clustering Multivariate Longitudinal Observations: The Contaminated Gaussian Hidden Markov Model. *Journal of Computational and Graphical Statistics* 25: 1097–1116.

Rabiner, L. R. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE* 77: 257–286.

R Core Team 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Revelle, W. 2017. psych: Procedures for Psychological, Psychometric, and Personality Research. *R Package Version 1.7.8 (2017-09-09)*.

Romano, J. P., and L. A. Thombs. 1996. Inference for Autocorrelations under Weak Assumptions. *Journal of the American Statistical Association* 91: 590–600.

Rydén, T., T. Teräsvirta, and S. Ösbrink. 1998. Stylized Facts of Daily Return Series and the Hidden Markov Model. *Journal of Applied Econometrics* 13: 217–244.

Trapletti, A., and K. Hornik. 2017. tseries: Time Series Analysis and Computational Finance. *R Package Version 0.10-43 (2018-02-09)*.

Wang, W., P. H. A. J. M. Van Gelder, J. K. Vrijling, and J. Ma. 2005. Testing and Modelling Autoregressive Conditional Heteroskedasticity of Streamflow Processes. *Nonlinear Processes in Geophysics* 12: 55–66.

Zucchini, W., I. L. MacDonald, and R. Langrock. 2016. *Hidden Markov Models for Time Series: An Introduction Using R (Second ed.), Volume 150 of Monographs on Statistics & Applied Probability*. Boca Raton, FL: CRC Press.

Zucchini, W., and K. Neumann. 2001. A Comparison of Several Time-Series Models for Assessing the Value at Risk of Shares. *Applied Stochastic Models in Business and Industry* 17: 135–148.