Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2006) 2192 – 2209
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

# Stylized facts of financial time series and hidden semi-Markov models

Jan Bulla[a,∗], Ingo Bulla[b]

[a] *Institut für Statistik und Ökonometrie, Georg-August-Universität, 37073 Göttingen, Germany*
[b] *Département de Mathématiques, Université de Bretagne Occidentale, Brest 29238, France*

Available online 14 August 2006

**Abstract**

Hidden Markov models reproduce most of the stylized facts about daily series of returns. A notable exception is the inability of the models to reproduce one ubiquitous feature of such time series, namely the slow decay in the autocorrelation function of the squared returns. It is shown that this stylized fact can be described much better by means of hidden semi-Markov models. This is illustrated by examining the fit of two such models to 18 series of daily sector returns.


*Keywords:* Daily return series; Hidden Markov model; Hidden semi-Markov model; Sojourn time distribution; EM algorithm; Right-censoring

## 1. Introduction

Applications related to Financial Econometrics like risk measurement, pricing of derivatives, margin setting and many other financial indicators rely on a suitable modeling of the distributional and temporal properties of the daily return series of stocks, indices or other assets.

The normal distribution with stationary parameters has often been chosen to model daily return series in financial theory. After the classical paper of Fama (1965), which observed more kurtosis and higher peaks contradicting the assumption of normality, many authors proposed solutions to overcome this drawback. For example, Praetz (1972) and Blattberg and Gonedes (1974) preferred the *t* distribution while Mittnik and Rachev (1993) examined various stable distributions. Gettinby et al. (2004) and Harris and Küçüközmen (2001) provide a good overview over the several approaches available in the literature.

As with the distributional properties, the temporal properties of daily return series have also been examined by many authors. The most popular models of the last decade fall into the class of ARCH-type models; for a review, see Bollerslev et al. (1992) or Franses and van Dijk (2000). Other approaches include the stochastic volatility models introduced by Taylor (1986) and have been applied in many contexts, e.g., by Koopman et al. (2005) and Yu (2002). State-space models based on the Kalman filter were investigated inter alia by Faff et al. (2000) and Yao and Gao (2004). The large class of Markov-switching models was introduced by Hamilton (1989, 1990). Turner et al. (1989) first considered a

∗ Corresponding author. Tel.: +49 551 397282; fax: +49 551 397279.
*E-mail addresses:* janbulla@yahoo.com (J. Bulla), ingobulla@gmx.de (I. Bulla).

0167-9473/$ - see front matter 
doi:10.1016/j.csda.2006.07.021

Markov mixture of normal distributions, and many other studies followed, e.g., Cecchetti et al. (1990), Linne (2002), and Bialkowski (2003).

In this paper, we focus on modeling the distributional and temporal properties of daily return series by hidden semi-Markov models (HSMM). The HSMM is a generalization of the well-known hidden Markov Model (HMM) insofar as it allows for more flexible distributions of the sojourn time in the visited states. Rydén et al. (1998) show that a HMM—mixing normal variables according to the states of an unobserved Markov chain—reproduces most of the stylized facts for daily return series listed by Granger and Ding (1995a,b). However, the analysis of Rydén et al. (1998), henceforth abbreviated RY, also illustrates that the stylized fact of the very slowly decaying autocorrelation for absolute (or squared) returns cannot be described by a HMM. The lack of flexibility of a HMM to model the temporal higher order dependence can be explained by the implicit geometric distributed sojourn time in the hidden states.

The two HSMMs explored in this article are generalizations of the model presented by RY. After fitting them to daily return series from 18 Pan-European sector indices, we show a significantly improved fit of the autocorrelation function.

The remainder of this article is organized as follows: Section 2 provides a brief introduction to HMMs. In Section 3 we give the theoretical basis for HSMMs and in Section 4 we present the estimation procedures for our particular models. Section 5 contains a short description of the data. In Section 6 the results are analyzed while Section 7 concludes. Finally, Appendix A presents some tables with detailed results and Appendix B contains certain mathematical details relating to the contents presented in the body of the article.

## 2. Hidden Markov models

HMMs are a class of models for time series $\{X_0, X_1, \ldots, X_{\tau-1}\}$, where the probability distribution of $X_t$ is determined by the unobserved states of a homogeneous and irreducible finite-state Markov chain $S_t$. They are also known as regime switching models. The implicit assumption of models switching between different regimes is that the data result from a process that undergoes abrupt changes, which are induced, e.g., by political or environmental events. The switching behavior is governed by a *transition probability matrix* (TPM). Under the assumption of a model with two states, the TPM is of the form

$$\mathbf{\Pi} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix},$$

where $p_{11}$ denotes the probability of staying in the first state from period $t$ to period $t+1$ and $p_{12}$ is the probability of switching from the first to the second state. The second row can be interpreted analogously.

The distribution of the observation at time $t$ is specified by the *conditional* or *component distributions* $P(X_t = x_t | S_t = s_t)$ associated with the “state” or “regime” $s_t$. Assuming, for instance, a two-state model with normal component distributions yields

$$X_t = \mu_{s_t} + \eta_{s_t}, \quad \eta_{s_t} \sim N\left(0, \sigma^2_{s_t}\right),$$

where

$$\mu_{s_t} = \begin{cases} \mu_1 & \text{if } s_t = 1, \\ \mu_2 & \text{if } s_t = 2 \end{cases} \quad \text{and} \quad \sigma^2_{s_t} = \begin{cases} \sigma^2_1 & \text{if } s_t = 1, \\ \sigma^2_2 & \text{if } s_t = 2. \end{cases}$$

The parameters of a HMM are generally estimated using the method of maximum-likelihood. The likelihood function is available in a convenient form:

$$L(\theta) = \boldsymbol{\delta} \boldsymbol{P}(x_0)\, \mathbf{\Pi} \boldsymbol{P}(x_1) \mathbf{\Pi} \cdots \boldsymbol{P}(x_{\tau-2})\, \mathbf{\Pi} \boldsymbol{P}(x_{\tau-1})\, \mathbf{1}, \tag{1}$$

where $\boldsymbol{P}(x_t)$ represents a diagonal matrix with the state-dependent conditional distributions as entries. The initial distribution of the Markov chain is denoted by $\boldsymbol{\delta}$, the model parameters by $\theta$, and $\mathbf{1} = (1, \ldots, 1)^{\mathrm{T}}$. The log-likelihood can be easily evaluated even for very long sequences of observations and so the parameters can be estimated by direct numerical maximization of the log-likelihood function using, e.g., Newton-type methods (see MacDonald and Zucchini, 1997). A popular and routinely used alternative is the Baum–Welch algorithm, a special case of what subsequently became known as the EM algorithm (Baum et al., 1970; Dempster et al., 1977; Rabiner, 1989).

## 3. Hidden semi-Markov models

The HMM presented in the previous section provides flexible, general-purpose models for univariate and multivariate time series. However, a major limitation is the implicit geometric distribution of the sojourn times as a consequence of the Markov property, i.e.,

$$P\,(\text{‘sojourn in state } j \text{ of length } u\text{’}) = p_{jj}^{u-1}\left(1 - p_{jj}\right).$$

The HSMM is a generalization of the HMM that allows to use more general sojourn time (or state occupancy) distributions. Non-parametric distributions were first proposed in the field of speech recognition by Ferguson (1980). After this pioneering work, the model was further investigated by different authors, e.g., Levinson (1986), Guédon and Cocozza-Thivent (1990), Sansom and Thomson (2001), Yu and Kobayashi (2003).

In this article, we extend the theory for right-censored models (see Guédon, 2003) to deal with different parametric distributions for the sojourn times and the observations. In contrast to the original estimation procedures of Ferguson (1980), the right-censored model does not assume that the last observation coincides with an exit from the last visited state. That assumption, though mathematically convenient, is clearly unrealistic; it is also not possible to check in practice.

### 3.1. Notation and definitions

We will denote the observed sequence of length $\tau$, $X_0 = x_0, \ldots, X_{\tau-1} = x_{\tau-1}$ by $X_0^{\tau-1} = x_0^{\tau-1}$. The same convention is used for the state sequence $S_t$. The complete set of parameters in the model is denoted by $\theta$.

A HSMM consists of a pair of discrete-time stochastic processes $\{S_t\}$ and $\{X_t\}$. Similar to HMMs, the observed process $\{X_t\}$ is related to the unobserved state process $\{S_t\}$ by the conditional distributions. The state process itself is a finite-state semi-Markov chain constructed as follows: the transitions between states are modeled by a homogeneous Markov chain with $J$ states, which we label $0, 1, \ldots, J-1$. The chain is specified by the *initial probabilities*

$$\pi_j := P\,(S_0 = j) \quad \text{with} \sum_j \pi_j = 1,$$

and the *transition probabilities* for the state $i$: for each $j \neq i$

$$p_{ij} := P\,(S_{t+1} = j \mid S_{t+1} \neq i, S_t = i) \quad \text{with} \sum_j p_{ij} = 1 \quad \text{and} \quad p_{ii} = 0.$$

Associated with each state is a *sojourn time distribution*, which models the number of time periods the chain remains in the state:

$$d_j(u) := P\,(S_{t+u+1} \neq j,\; S_{t+u-v} = j,\; v = 0, \ldots, u-2 \mid S_{t+1} = j, S_t \neq j)\,.$$

In contrast to HMMs, the diagonal elements of the TPM of a HSMM all have to be zero. (We assume that the states are non-absorbing.)

The sojourn time in the last visited state is subject to a right-censoring and is modeled by the *survivor function*

$$D_j(u) := \sum_{v \geqslant u} d_j(v).$$

The survivor function is the key to extend the original algorithms of Ferguson (1980), i.e. to drop the assumption that there is a change of state immediately after the last observation.

Thus HMMs and HSMMs differ only in the way that their state processes $\{S_t\}$ are defined. The semi-Markov chain associated with HSMMs does not have the Markov property at each time $t$; that property is transferred to the level of the embedded first-order Markov chain.

The observed process $\{X_t\}$ at time $t$ is associated to the state process $\{S_t\}$ by the *component distributions*

$$b_j(x_t) = P(X_t = x_t \mid S_t = j).$$

For the observation component, the conditional independence property known from the HMM is fulfilled, i.e.,

$$P(X_t = x_t \mid S_t = j) = P\left(X_t = x_t \mid X_0^{\tau-1} = x_0^{\tau-1}, S_0^{t-1} = s_0^{t-1}, S_t = j, S_{t+1}^{\tau-1} = s_{t+1}^{\tau-1}\right),$$

which implies, the output process at time $t$ depends only on the value of $S_t$.

### 3.2. The likelihood of a HSMM

The likelihood of a HMM is available in a tractable form given by Eq. (1). Unfortunately, such a convenient representation does not exist for the likelihood of a HSMM. We first have to introduce the *complete-data likelihood* which also plays an important role for the parameter estimation in Section 3.3. The likelihood of the complete data, i.e., the observations $x_0^{\tau-1}$ as well as the unobserved sequence $s_0^{\tau-1+u}$, is given by

$$\begin{aligned} L_C\left(s_0^{\tau-1+u}, x_0^{\tau-1} \mid \theta\right) = P\Big(&S_0^{\tau-1} = s_0^{\tau-1}, S_{\tau-1+v} = s_{\tau-1}, \ v = 1, \ldots, u-1, \\ &S_{\tau-1+u} \neq s_{\tau-1}, X_0^{\tau-1} = x_0^{\tau-1} \mid \theta\Big). \end{aligned} \tag{2}$$

The last visited state is left at time $\tau-1+u$ and therefore the completed state sequence stops at this time, instead of $\tau-1$ for algorithms without right-censoring (see Guédon, 2003). The contribution of the state sequence to the complete-data likelihood is given by

$$\pi_{\tilde{s}_0} d_{\tilde{s}_0}(u_0) \prod_{r=1}^{R} p_{\tilde{s}_{r-1}\tilde{s}_r} d_{\tilde{s}_r}(u_r)\, I\left(\sum_{r=0}^{R-1} u_r < \tau \leqslant \sum_{r=0}^{R} u_r\right), \tag{3}$$

where $\tilde{s}_0, \ldots, \tilde{s}_R$ denote the $R+1$ states visited by $s_0^{\tau-1+u}$. Combining Eqs. (2) and (3), the likelihood of the observed sequence can be calculated by summing the complete-data likelihood over all admissible paths. Compared to the classical likelihood of Ferguson (1980), the likelihood function here involves an additional sum over all possible prolongations of the state sequence $s_0^{\tau-1}$ up to the exit from the last visited state:

$$L(\theta) = \sum_{s_0,\ldots,s_{\tau-1}} \sum_{u_{\tau+}} L_C\left(s_0^{\tau-1+u}, x_0^{\tau-1} \mid \theta\right), \tag{4}$$

where $\sum_{s_0,\ldots,s_{\tau-1}}$ denotes the summation over every possible state sequence of length $\tau$ and $\sum_{u_{\tau+}}$ denotes the sum over every supplementary duration from time $\tau$ spent in the state occupied at time $\tau-1$.

### 3.3. The EM algorithm for HSMM

The likelihood of a HSMM given in Eq. (4) could be easily evaluated if the state sequence was known. However, this is not the case and we are hence confronted with a missing data problem. A popular method of dealing with this type of problem is the expectation maximization (EM) algorithm, an iterative procedure which increases the likelihood monotonically until it reaches a stationary point (see, for instance, Baum et al., 1970; Dempster et al., 1977). After assigning initial values to the parameters, the EM algorithm is implemented by successively iterating the following two steps:

*E-step*: Compute the $Q$-function

$$Q\left(\theta, \theta^{(k)}\right) = \mathbf{E}\left[L_C\left(s_0^{\tau-1+u}, x_0^{\tau-1} \mid \theta\right) \mid X_0^{\tau-1} = x_0^{\tau-1}, \theta^{(k)}\right], \tag{5}$$

the conditional expectation of the complete-data log-likelihood, where $\theta^{(k)}$ denotes the current estimate of the parameter vector $\theta$.

*M-step*: Compute $\theta^{(k+1)}$, the parameter values that maximize the function $Q$ w.r.t. $\theta$, i.e.,

$$\theta^{(k+1)} = \underset{\theta}{\operatorname{argmax}}\ Q\left(\theta, \theta^{(k)}\right).$$

The two steps are repeated until a stationary point is reached. In our case, the EM algorithm maximizes $L(\theta)$ from Eq. (4). The implementation of Guédon (2003) serves as basis for our algorithms because of its appealing properties such as immunity to numerical underflow and low complexity. We focus on the extension to different parametric sojourn time and component distributions. Re-estimation formulae for the models considered are provided in Appendix B. For a detailed survey on the EM algorithm for right-censored HSMMs, we refer to Guédon (2003).

The main difficulty of the EM algorithm is the *E-step*. To obtain a mathematically tractable formulation of the $Q$-function, the conditional expectation has to be rewritten pathwise. The conditional distribution of the missing observations is given by $P\left(S_0^{\tau-1+u} = s_0^{\tau-1+u} \mid X_0^{\tau-1} = x_0^{\tau-1}, \theta\right)$ and the distribution of the complete-data by $P\left(S_0^{\tau-1+u} = s_0^{\tau-1+u}, X_0^{\tau-1} = x_0^{\tau-1} \mid \theta\right)$. Hence, as a first step Eq. (5) is transformed to

$$Q\left(\theta, \theta^{(k)}\right) = \sum_{s_1,\ldots,s_{\tau-1}} \sum_{u_{\tau+}} \left[\log L_C\left(s_0^{\tau-1+u}, x_0^{\tau-1} \mid \theta\right) P\left(S_0^{\tau-1+u} = s_0^{\tau-1+u} \mid X_0^{\tau-1} = x_0^{\tau-1}, \theta^{(k)}\right)\right]. \tag{6}$$

Secondly, consider the contribution of a specific path to the likelihood of a HSMM given by Eq. (3). Adding the contribution of the observed sequence, the complete-data likelihood becomes

$$L_C\left(s_0^{\tau-1+u}, x_0^{\tau-1} \mid \theta\right) = \pi_{\tilde{s}_0} d_{\tilde{s}_0}(u_0) \prod_{r=1}^{R} p_{\tilde{s}_{r-1}\tilde{s}_r} d_{\tilde{s}_r}(u_r) \prod_{t=0}^{\tau-1} b_{s_t}(x_t). \tag{7}$$

Taking the logarithm splits Eq. (7) into four independent terms and Eq. (6) becomes

$$\begin{aligned} Q\left(\theta, \theta^{(k)}\right) = \sum_{s_1,\ldots,s_{\tau-1}} \sum_{u_{\tau+}} \Bigg[&\log \pi_{\tilde{s}_0} + \left(\sum_{r=1}^{R} \log p_{\tilde{s}_{r-1}\tilde{s}_r}\right) + \left(\sum_{r=0}^{R} \log d_{\tilde{s}_r}(u_r)\right) \\ &+ \left(\sum_{t=0}^{\tau-1} \log b_{s_t}(x_t)\right)\Bigg] P\left(S_0^{\tau-1+u} = s_0^{\tau-1+u} \mid X_0^{\tau-1} = x_0^{\tau-1}, \theta^{(k)}\right). \end{aligned} \tag{8}$$

The four terms in Eq. (8) correspond to the initial, the transition, the sojourn time and the observation probabilities. In a last step, the summation over all paths is marginalized, and the four terms can be written as

$$\sum_{j=0}^{J-1} P\left(S_0 = j \mid X_0^{\tau-1} = x_0^{\tau-1}, \theta^{(k)}\right) \log \pi_j, \tag{9}$$

$$\sum_{i=0}^{J-1} \sum_{j \neq i} \sum_{t=0}^{\tau-2} P\left(S_{t+1} = j, S_t = i \mid X_0^{\tau-1} = x_0^{\tau-1}, \theta^{(k)}\right) \log p_{ij}, \tag{10}$$

$$\begin{aligned} \sum_{u} \Bigg\{ &\sum_{t=0}^{\tau-2} \left(S_{t+u+1} \neq j, S_{t+u-v} = j,\ v = 0, \ldots, u-1, S_t \neq j \mid X_0^{\tau-1} = x_0^{\tau-1}, \theta^{(k)}\right) \\ &+ P\left(S_u \neq j, S_{u-v} = j,\ v = 1, \ldots, u \mid .X_0^{\tau-1} = x_0^{\tau-1}, \theta^{(k)}\right) \Bigg\} \log d_j(u), \end{aligned} \tag{11}$$

and

$$\sum_{j=0}^{J-1} \sum_{t=0}^{\tau-1} P\left(S_t = j \mid X_0^{\tau-1} = x_0^{\tau-1}, \theta^{(k)}\right) \log b_j(x_t). \tag{12}$$

The re-estimation formulae for these four quantities can be obtained by maximizing each of the terms separately. The results are given in Appendix B.

## 4. The models for daily return series

RY fitted HMMs with normal component distributions to subseries of the well-known S&P 500 index. In particular, they noticed that the autocorrelation function of the estimated model does not capture the behavior of the empirical autocorrelation function satisfactorily, mainly due to the much slower decay of the latter. The temporal dependence properties of a HMM depend on the values of the transition probability matrix (MacDonald and Zucchini, 1997, Chapter 2.4) but the sojourn times are always geometrically distributed. In contrast to many other applications, e.g., speech recognition, there exists no test data for financial time series where the ‘real’ sojourn time distribution is known. Sansom and Thomson (2001) fitted HSMMs with non-parametric state occupation in a first step to deduce the shape of a parametric distribution in the context of rainfall data. This approach does not yield satisfactory results for daily return series. We generalize the model of RY by fitting a HSMM with negative binomial sojourn time distributions of the form

$$d_j(u) = \binom{u-2+r_j}{u-1} p_j^{r_j} \left(1-p_j\right)^{u-1}, \quad u = 1, 2, \ldots .$$

The number of parameters only increases by one per state, and for $r_j = 1$, $j \in 0, \ldots, J-1$ our model reduces to a HMM.

While Granger and Ding (1995a,b) suggested a double exponential distribution to characterize daily returns, RY proposed mixtures of normal variables. We fit HSMMs with normal and $t$ distributed variables, respectively. For the remainder of this article, the HMM of RY will be denoted by $M_{\mathrm{RY}}$, the HSMM with normal conditional distributions by $SM_N$ and the HSMM with conditional $t$ distributions by $SM_t$. RY investigated models with two and three states. They noticed that the three-state models “are less similar to each other” and that “the estimation results seem heavily dependent on outlying observations” (Rydén et al., 1998). These findings were confirmed in our own preliminary analysis. Therefore, all models considered in this article have two states.

Finally, it may be noted that the EM algorithm showed a high stability. Exploring the effect of different initial values by grid searches, we discovered a stable convergence to the global maximum for all cases investigated except when very poor initial guesses were used.

## 5. The data series

The data used here are the daily returns calculated for 18 Pan-European industry portfolios, covering the period from 1st January 1987 to 5th September 2005. All sector indices are from STOXX Ltd. and the common currency used is Euro. The daily returns of the period from $t$ to $t-1$ are computed continuously by

$$R_t = \ln\left(P_t\right) - \ln\left(P_{t-1}\right),$$

where $P_t$ represents the index closing price on day $t$ and ln is the natural logarithm. All data are obtained from Thomson financial datastream.

Descriptive statistics for the data are provided in Table 1. It was found that all sector indices are leptokurtic and negatively skewed. The Jarque–Bera statistic confirms the departure from normality for all return series at the 1% level of significance.

The kurtosis of all series is high (Table 1). Thus, the data might contain extreme outliers, which jeopardize the specification power of the ACF (Chan, 1995). To reduce outlier effects we followed the approach of Granger and Ding (1995a) by setting values outside the interval $\left[\bar{r}_t - 4\hat{\sigma}, \bar{r}_t + 4\hat{\sigma}\right]$ equal to the value of the closest interval boundary; $\hat{\sigma}$ denotes the estimated standard deviation. For the empirical analysis in Section 6, we study the original as well as the outlier-corrected data.

Table 1
Descriptive statistics of daily sector returns

| Sector | $N$[a] | Mean × $10^4$ | S.D. × $10^2$ | Skew. | Kurt. | JB[b] |
|---|---|---|---|---|---|---|
| Automobiles | 4824 | 0.534 | 1.48 | −0.363 | 8.99 | 7315 |
| Banks | 4824 | 2.28 | 1.2 | −0.29 | 9.98 | 9863 |
| Basics | 4824 | 2.09 | 1.25 | −0.396 | 9.82 | 9474 |
| Chemicals | 4824 | 2.54 | 1.29 | −0.125 | 8.6 | 6314 |
| Construction | 4824 | 2.88 | 1.09 | −0.56 | 9.39 | 8478 |
| Financials | 4824 | 1.99 | 1.16 | −0.508 | 11 | 13 217 |
| Food | 4824 | 2.84 | 1.1 | −0.247 | 17.2 | 40 877 |
| Healthcare | 4824 | 3.54 | 1.31 | −0.388 | 9.3 | 8096 |
| Industrials | 4824 | 2.74 | 1.16 | −0.416 | 9.42 | 8423 |
| Insurance | 4824 | 1.25 | 1.45 | −0.265 | 10.4 | 11 041 |
| Media | 4824 | 2.42 | 1.49 | −0.443 | 10.3 | 10 990 |
| Oil and gas | 4824 | 3.89 | 1.24 | −0.251 | 7.48 | 4098 |
| Personal | 4824 | 3.32 | 1.05 | −0.162 | 7.78 | 4627 |
| Retail | 4824 | 1.68 | 1.87 | −0.258 | 8.26 | 5614 |
| Technology | 4824 | 3.33 | 1.63 | −0.134 | 7.27 | 3683 |
| Telecom | 3521 | 2.25 | 1.48 | −0.363 | 8.38 | 4339 |
| Travel | 3521 | 2.38 | 1.48 | −0.186 | 5.47 | 916 |
| Utilities | 3521 | 2.07 | 1.19 | −0.428 | 10.2 | 7740 |

This table summarizes the daily returns data of the 18 DJ STOXX sector indices, covering the period from 1 January 1987 to 5 September 2005.

[a]In September 2004, STOXX Ltd. replaced the sectors cyclical goods and services, non-cyclical goods and services, and retail (old) by the new sectors travel and leisure, personal and household goods, and retail (new), respectively. The history of the newly formed sectors (with 3521 observations) dates back to 31st December 1991.

[b]JB is the Jarque–Bera statistic for testing normality.

## 6. Empirical results

### 6.1. Basic results

As a first step of the empirical analysis we present some basic statistics for the return series and the fitted models. All estimation results for the three models are reported in Tables A.1–A.6 of Appendix A. All three models considered, fit the marginal distributions of the returns reasonably well. As expected for daily return series, the empirical mean and the mean of the fitted models lie very close to zero for all 18 sectors. The standard deviation almost coincides with the empirical value for every model, the skewness is slightly underestimated. The three models exhibit a clearer tendency towards the kurtosis. Even though all models are subject to excess kurtosis, $SM_t$ and $SM_N$ provide the best results for the original and for the outlier-corrected series, respectively. The average empirical kurtosis of the 18 sectors is 9.41 (outlier-corrected data: 5.91) and the average kurtosis of $SM_t$ is 9.98 (6.22). $SM_N$ and $M_{\mathrm{RY}}$ achieve an average kurtosis of 6.46 (5.78) and 5.95 (5.4), respectively.

The average log-likelihood of $M_{\mathrm{RY}}$ is 14 200 (14 267). It increases to 14 236 (14 299) and 14 271 (14 311) for $SM_N$ and $SM_t$, respectively. A standard procedure for comparing two nested models is the likelihood ratio test (LRT). As the three models are hierarchically nested, the LRT may be applied with the null hypothesis of $r_1, r_2 = 1$ for the comparison $M_{\mathrm{RY}}/SM_N$, and $v_1, v_2 = \infty$ for $SM_N/SM_t$. Using the original data, $SM_N$ is better than $M_{\mathrm{RY}}$ at 0.1% level of significance for each of the 18 sectors. The same statement holds true for $SM_N/SM_t$, indicating that $SM_t$ provides the best fit to the data. The results for the outlier-corrected data are similar, with the limitation that level of significance is 1% for the comparison $SM_N/SM_t$. The only exception is the utilities sector, where the test is not significant. The preference for the HSMMs is supported by the Akaike information criterion which, on an average, decreases from −28 388 (−28 523) for the $M_{\mathrm{RY}}$ to −28 456 (28 582) and −28 522 (−28 601) for $SM_N$ and $SM_t$, respectively.

### 6.2. The stylized facts

The main focus of the second step of our empirical analysis lies on absolute and squared daily returns. Granger and Ding (1995a,b) described several temporal and distributional properties for daily return series for the S&P 500 index and extended their analysis to other return series in a more recent paper (Granger et al., 2000). RY showed that the HMM reproduces most of these properties or stylized facts. The four temporal properties are as follows:

TP1: Returns $r_t$ are not autocorrelated (except for, possibly, at lag one).
TP2: $|r_t|$ and $r_t^2$ are 'long-memory', i.e., their autocorrelation functions decay slowly starting from the first autocorrelation, and $\text{corr}\,(|r_t|\,, |r_{t-k}|) > \text{corr}\left(r_t^2, r_{t-k}^2\right)$. The autocorrelations remain positive for many lags and the decay is much slower than the exponential rate of a typical stationary ARMA model.
TP3: The Taylor effect $\text{corr}\,(|r_t|\,, |r_{t-k}|) > \text{corr}\left(|r_t|^\theta, |r_{t-k}|^\theta\right)$, $\theta \neq 1$ (Taylor, 1986). Autocorrelations of powers of absolute returns are highest at power one.
TP4: The autocorrelations of sign $(r_t)$are negligibly small.

The three distributional properties are:

DP1: $|r_t|$ and sign $(r_t)$ are independent.
DP2: Mean $|r_t|$ = standard deviation $|r_t|$.
DP3: The marginal distribution of $|r_t|$ is exponential (after outlier correction).

Note that an exponentially distributed variable (DP3) $x_t$ has the following properties.

PED1: $E\,(x_t) = Var\,(x_t)$ (same as DP2).
PED2: $E(x_t - E\,(x_t))^3 = 2$.
PED3: $E(x_t - E\,(x_t))^4 = 9$.

RY showed that a HMM with normal component distributions and means equal to zero satisfies TP1, and that TP4 is not violated in practice. DP1 holds by construction of their model. The models fitted in our analysis allow for means unequal to zero. However, the estimation results show that all means take values very close to zero (Tables A.1–A.6). Preliminary analysis, which are not displayed here, showed that none of the estimated models violates TP1, TP4 or DP1. We defer the discussion of TP2 and TP3 until later and focus on the remaining distributional properties PED1–PED3.

The mean-standard deviation ratio, skewness, and kurtosis of the absolute returns estimated from the 18 data series and from the three fitted models are presented in Table 2 and 3 for both the original and the outlier-corrected data, respectively. The ratio of mean and standard deviation (PED1/DP2) lies close to one for the majority of the original and outlier-corrected series. This stylized fact is reproduced very well by $SM_t$, whereas the two models with normal components tend to slightly overestimate the ratio. This confirms the analysis of RY, who noted that PED1 "has to be relaxed somewhat (the mean has to be allowed to be slightly larger than the standard deviation) if we at the same time want PED2 and PED3 to be satisfied". For the original data, the skewness and the kurtosis are reproduced very well by $SM_t$. The average skewness of the data (model) is 2.81 (2.81), and the average kurtosis is 17.2 (18.7). The other two models tend to underestimate both skewness and kurtosis. Considering the outlier-corrected data, the differences between the models reduce. Skewness and kurtosis are reproduced well by all models, though $SM_N$ performs slightly better than the other two. $M_{\text{RY}}$ and $SM_t$ slightly underestimate and overestimate these moments, respectively. Summarizing the above results it can be stated that the HSMMs reproduce PED1–PED3 comparably well, or better, than the HMM. For the original data, $SM_t$ is the best model, and $SM_N$ is preferable for outlier-corrected data.

One of the remaining stylized facts to check is TP3, the Taylor effect. We estimated the coefficient $\theta$ for each data series by maximizing the first-order autocorrelation of $|r_t|^\theta$ by means of a quasi-Newton algorithm. Following the approach of RY, the value of $\theta$ maximizing the first-order autocorrelation for the models was estimated over the range $\{0.1, 0.2, \ldots, 2.0\}$ by Monte-Carlo approximation. The results are summarized in Table 4. We can confirm the observation of RY that the outlier-correction seems to weaken the Taylor effect. The average value of $\theta$ maximizing the first-order autocorrelation increases significantly from 1.13 to 1.42 following outlier-correction. For the original data, the maximizing values of $\theta$ calculated for the three models coincide reasonably well with those calculated from observations, though $SM_t$ slightly underestimates $\theta$. For the outlier-corrected data all models perform comparably and underestimate $\theta$ a little. However, the results are still generally satisfactory.

Table 2
Statistics of the absolute returns and the estimated models, original data mean-standard deviation ratio, skewness and kurtosis of the absolute returns estimated from the 18 data series and from the three fitted models $M_{RY}$, $SM_N$, and $SM_t$ (by Monte Carlo approximation)

| Sector | Mean/S.D. | | | | Skewness | | | | Kurtosis | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Data | $M_{RY}$ | $SM_N$ | $SM_t$ | Data | $M_{RY}$ | $SM_N$ | $SM_t$ | Data | $M_{RY}$ | $SM_N$ | $SM_t$ |
| Automobiles | 0.965 | 1.04 | 1.01 | 0.965 | 2.81 | 2.19 | 2.33 | 2.83 | 16.2 | 9.83 | 10.7 | 17.3 |
| Banks | 0.896 | 0.962 | 0.931 | 0.892 | 2.97 | 2.39 | 2.51 | 3.14 | 16.2 | 10.7 | 11.3 | 20.4 |
| Basics | 0.955 | 1.03 | 1 | 0.953 | 2.96 | 2.12 | 2.22 | 3.1 | 18.2 | 9.2 | 9.69 | 24.1 |
| Chemicals | 0.97 | 1.05 | 1.02 | 0.973 | 2.67 | 1.97 | 2.1 | 2.66 | 15.4 | 8.21 | 8.86 | 16.1 |
| Construction | 1 | 1.08 | 1.05 | 1.01 | 2.9 | 2.03 | 2.2 | 2.93 | 18.8 | 9.08 | 10.2 | 21.1 |
| Financials | 0.886 | 0.956 | 0.926 | 0.883 | 3.14 | 2.4 | 2.6 | 3.36 | 18.6 | 10.7 | 12.1 | 24.6 |
| Food | 0.918 | 1.02 | 0.986 | 0.938 | 4.08 | 2.27 | 2.47 | 3.85 | 37.7 | 10.2 | 11.6 | 44.4 |
| Healthcare | 0.992 | 1.06 | 1.04 | 0.997 | 2.8 | 1.89 | 1.98 | 2.56 | 18.4 | 7.66 | 8.15 | 15.4 |
| Industrials | 0.966 | 1.03 | 1.01 | 0.962 | 2.88 | 2.02 | 2.08 | 2.85 | 17.4 | 8.39 | 8.69 | 19 |
| Insurance | 0.881 | 0.953 | 0.918 | 0.88 | 3.03 | 2.44 | 2.67 | 3.23 | 16.8 | 11 | 12.6 | 20.9 |
| Media | 0.908 | 0.973 | 0.956 | 0.903 | 3.06 | 2.32 | 2.39 | 3.24 | 17.6 | 10.2 | 10.6 | 22.9 |
| Oil and gas | 1.01 | 1.08 | 1.06 | 1.01 | 2.44 | 1.87 | 1.97 | 2.42 | 13.3 | 7.71 | 8.29 | 13.2 |
| Personal | 1.02 | 1.09 | 1.06 | 1.02 | 2.55 | 2.11 | 2.29 | 2.59 | 14.5 | 9.89 | 11.1 | 15 |
| Retail | 0.893 | 0.954 | 0.931 | 0.889 | 2.52 | 2.05 | 2.13 | 2.56 | 12.2 | 8.02 | 8.42 | 13 |
| Technology | 0.996 | 1.08 | 1.04 | 0.996 | 2.31 | 1.79 | 1.89 | 2.22 | 12.7 | 7.14 | 7.48 | 10.6 |
| Telecom | 0.974 | 1.05 | 1.02 | 0.977 | 2.62 | 2.03 | 2.17 | 2.62 | 14.9 | 8.65 | 9.41 | 14.9 |
| Travel | 1.08 | 1.14 | 1.11 | 1.08 | 1.93 | 1.78 | 1.84 | 1.98 | 8.88 | 7.74 | 7.82 | 8.83 |
| Utilities | 0.994 | 1.06 | 1.03 | 1 | 2.93 | 2.01 | 2.19 | 2.58 | 21.7 | 8.64 | 9.79 | 14.7 |

Table 3
Statistics of the absolute returns and the estimated models, outlier-corrected data

| Sector | Mean/S.D. | | | | Skewness | | | | Kurtosis | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Data | $M_{RY}$ | $SM_N$ | $SM_t$ | Data | $M_{RY}$ | $SM_N$ | $SM_t$ | Data | $M_{RY}$ | $SM_N$ | $SM_t$ |
| Automobiles | 1.02 | 1.07 | 1.04 | 1.02 | 2.09 | 2 | 2.13 | 2.14 | 8.74 | 8.66 | 9.44 | 9.37 |
| Banks | 0.949 | 0.998 | 0.967 | 0.95 | 2.28 | 2.2 | 2.32 | 2.33 | 9.55 | 9.5 | 10.1 | 10.3 |
| Basics | 1.01 | 1.07 | 1.04 | 1.01 | 2.1 | 1.91 | 1.99 | 2.21 | 8.78 | 7.91 | 8.27 | 10.4 |
| Chemicals | 1.02 | 1.08 | 1.05 | 1.02 | 2.01 | 1.81 | 1.94 | 2.09 | 8.3 | 7.26 | 7.94 | 9.32 |
| Construction | 1.06 | 1.12 | 1.09 | 1.07 | 2 | 1.83 | 1.94 | 2.06 | 8.53 | 7.82 | 8.38 | 9.47 |
| Financials | 0.947 | 0.997 | 0.969 | 0.952 | 2.3 | 2.21 | 2.37 | 2.36 | 9.63 | 9.52 | 10.6 | 10.5 |
| Food | 1.02 | 1.09 | 1.07 | 1.02 | 2.14 | 1.85 | 1.96 | 2.36 | 9.35 | 7.66 | 8.27 | 12.9 |
| Healthcare | 1.05 | 1.09 | 1.08 | 1.05 | 1.92 | 1.75 | 1.78 | 2 | 7.84 | 6.96 | 7.03 | 8.94 |
| Industrials | 1.02 | 1.06 | 1.05 | 1.02 | 2.06 | 1.88 | 1.93 | 2.14 | 8.56 | 7.65 | 7.85 | 9.76 |
| Insurance | 0.933 | 0.989 | 0.954 | 0.937 | 2.36 | 2.25 | 2.46 | 2.45 | 9.91 | 9.77 | 11.2 | 11.1 |
| Media | 0.972 | 1.01 | 0.993 | 0.978 | 2.23 | 2.17 | 2.24 | 2.24 | 9.31 | 9.35 | 9.7 | 9.76 |
| Oil and gas | 1.04 | 1.1 | 1.08 | 1.05 | 1.91 | 1.77 | 1.85 | 1.96 | 7.76 | 7.23 | 7.56 | 8.43 |
| Personal | 1.06 | 1.11 | 1.09 | 1.07 | 1.94 | 1.91 | 2.04 | 2.07 | 8.12 | 8.49 | 9.23 | 9.38 |
| Retail | 0.919 | 0.966 | 0.944 | 0.922 | 2.14 | 2.01 | 2.09 | 2.18 | 8.32 | 7.85 | 8.2 | 8.99 |
| Technology | 1.02 | 1.09 | 1.05 | 1.03 | 1.84 | 1.73 | 1.8 | 1.89 | 7.08 | 6.8 | 7.01 | 7.64 |
| Telecom | 1.02 | 1.07 | 1.04 | 1.02 | 2.02 | 1.92 | 2.04 | 2.05 | 8.23 | 8.01 | 8.59 | 8.7 |
| Travel | 1.1 | 1.15 | 1.12 | 1.09 | 1.7 | 1.73 | 1.79 | 1.87 | 6.69 | 7.4 | 7.5 | 7.98 |
| Utilities | 1.04 | 1.09 | 1.06 | 1.05 | 2 | 1.86 | 2.01 | 2.02 | 8.28 | 7.74 | 8.6 | 8.7 |

Mean-standard deviation ratio, skewness and kurtosis of the absolute returns estimated from the 18 data series and from the three fitted models $M_{RY}$, $SM_N$, and $SM_t$ (by Monte Carlo approximation).

The last remaining stylized fact to check is TP2. The slow decay of the ACF for series of absolute/squared daily return is difficult to model. RY stated that this stylized fact cannot be reproduced by the HMM because the decay of the autocorrelations is (much) faster than that observed in reality. He considered this stylized fact to be "the most difficult [...] to reproduce with a HMM". Fig. 1 shows the empirical ACF of squared returns as well as the ACF of the

Table 4
Taylor coefficients of the absolute returns and the estimated models, outlier-corrected data

| Sector | Original data | | | | Outlier-corrected data | | | |
|---|---|---|---|---|---|---|---|---|
| | Data | $M_{RY}$ | $SM_N$ | $SM_t$ | Data | $M_{RY}$ | $SM_N$ | $SM_t$ |
| Automobiles | 1.27 | 1.2 | 1.1 | 1.0 | 1.58 | 1.2 | 1.2 | 1.1 |
| Banks | 1.13 | 1.0 | 1.0 | 0.8 | 1.43 | 1.0 | 1.0 | 1.0 |
| Basics | 1.06 | 1.1 | 1.1 | 0.8 | 1.33 | 1.1 | 1.0 | 1.0 |
| Chemicals | 1.08 | 1.1 | 1.0 | 0.8 | 1.31 | 1.0 | 1.0 | 0.9 |
| Construction | 1.08 | 1.3 | 1.2 | 1.0 | 1.5 | 1.3 | 1.2 | 1.2 |
| Financials | 1.05 | 1.0 | 1.0 | 0.8 | 1.29 | 1.0 | 1.0 | 1.0 |
| Food | 1.12 | 1.1 | 1.1 | 0.8 | 1.55 | 1.1 | 1.1 | 0.9 |
| Healthcare | 0.946 | 1.0 | 1.0 | 0.8 | 1.25 | 1.0 | 1.0 | 0.9 |
| Industrials | 0.89 | 1.0 | 1.0 | 0.8 | 1.07 | 1.0 | 1.0 | 0.9 |
| Insurance | 1 | 1.0 | 1.0 | 0.8 | 1.18 | 1.0 | 1.0 | 1.0 |
| Media | 1.15 | 1.0 | 1.0 | 0.8 | 1.32 | 1.0 | 1.0 | 1.0 |
| Oil and gas | 1.58 | 1.1 | 1.1 | 0.9 | 1.92 | 1.1 | 1.1 | 1.0 |
| Personal | 1.26 | 1.4 | 1.3 | 1.2 | 1.68 | 1.4 | 1.3 | 1.3 |
| Retail | 0.744 | 0.8 | 0.8 | 0.7 | 0.82 | 0.8 | 0.8 | 0.8 |
| Technology | 0.95 | 1.0 | 0.9 | 0.8 | 1.05 | 1.0 | 0.9 | 0.9 |
| Telecom | 1.22 | 1.1 | 1.1 | 1.0 | 1.35 | 1.1 | 1.1 | 1.1 |
| Travel | 1.58 | 1.4 | 1.2 | 1.2 | 2.25 | 1.4 | 1.2 | 1.2 |
| Utilities | 1.22 | 1.1 | 1.1 | 0.9 | 1.6 | 1.1 | 1.1 | 1.1 |

Values of $\theta$ maximizing the first-order autocorrelation of $|r_t|^\theta$ estimated from the 18 original and outlier-corrected data series and the six fitted models.

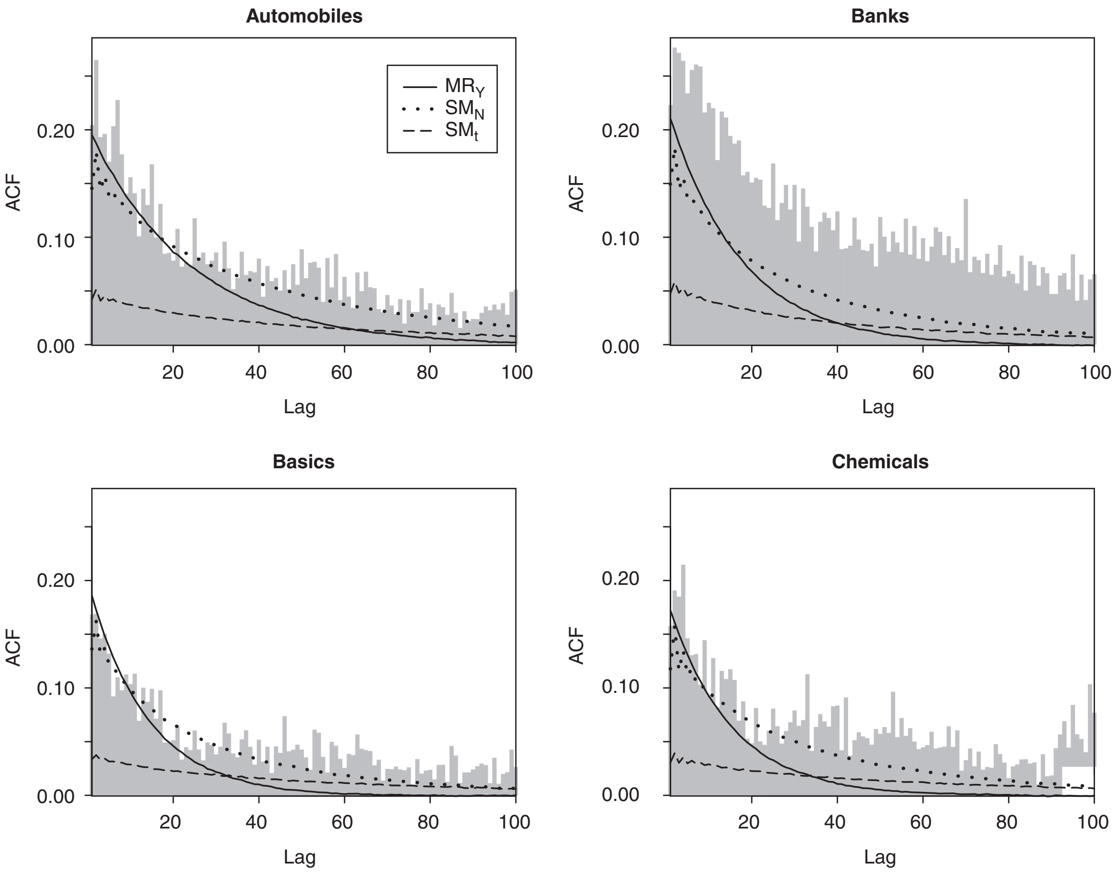


Fig. 1. Empirical (gray bars) and model ACF at lag 1-100 of $r_t^2$ for the first four sectors.

Table 5
Average mean squared error and weighted mean squared error for the ACF of the 18 sectors

| Criterion | Original data | | | Outlier-corrected data | | |
|---|---|---|---|---|---|---|
| | $M_{\mathrm{RY}}$ | $SM_N$ | $SM_t$ | $M_{\mathrm{RY}}$ | $SM_N$ | $SM_t$ |
| $\overline{MSE} \times 10^3$ | 12 | 8.87 | 17.5 | 13.6 | 10.2 | 19.5 |
| $\overline{wMSE} \times 10^3$ | 2.34 | 1.6 | 2.19 | 2.63 | 1.81 | 2.55 |

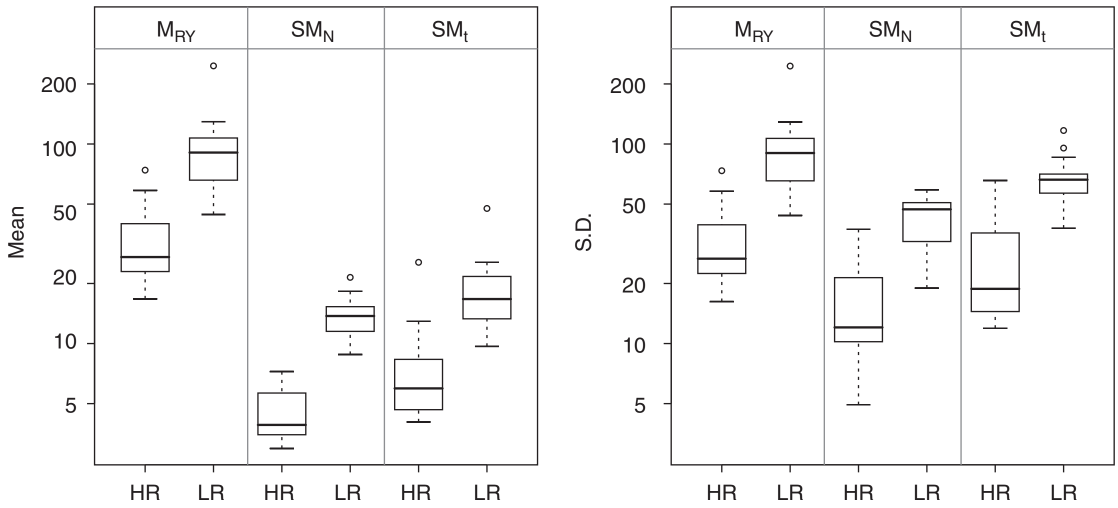


Fig. 2. Mean and standard deviation of the sojourn time distributions of all the 18 sectors, grouped by model and high-risk (HR)/low-risk (LR) states. The *y*-axis is logarithmic.

three models (estimated by Monte Carlo approximation) for the first four sectors. The other fourteen sectors and the outlier-corrected series show similar results and are therefore omitted. The solid line represents the ACF of $M_{\mathrm{RY}}$, while the dotted and dashed lines represent $SM_N$ and $SM_t$, respectively.

The HMM shows the typical strong decay of the autocorrelations and is far from the gray empirical ACF, which confirms the results of RY. Both $SM_N$ and $SM_t$ reproduce this stylized fact much better than the HMM. However, the $SM_t$ looses some of its credibility due to the bad fit for the lags of lower order. Here, $SM_N$ performs clearly better. To measure the fit of the ACF, we calculate the mean squared error (*MSE*) of the models and a weighted mean squared error (*wMSE*). The *wMSE* reweights the error at lag $i$ by $0.95^{(100-i)}$ to increase the influence of higher order lags. The results reported in Table 5 confirm the visual impression that $SM_N$ provides the best fit with respect to both criteria. Compared to $M_{\mathrm{RY}}$, the *MSE* of the ACF for $SM_N$ is reduced by approximately 25%.

Summarizing, the two HSMMs perform comparably well or better than $M_{\mathrm{RY}}$ w.r.t. most temporal and distributional properties. While $SM_t$ is the first choice regarding the model selection criteria AIC and LRT, $SM_N$ reproduces the shape of the empirical ACF much better than $M_{\mathrm{RY}}$ and $SM_N$.

### *6.3. Evaluation: the sojourn time distribution*

The difference between HMMs and HSMMs is the sojourn time distribution, which we study in this subsection. The results of the original and the outlier-corrected data do not differ substantially, and we therefore restrict our remarks to the analysis of the original data. In the HMM, mean and variance of the sojourn times (in each state) are controlled by the parameters of the geometric distributions. In the HSMM, the additional parameter allows more flexibility of mean and variance. The results are displayed in Fig. 2.

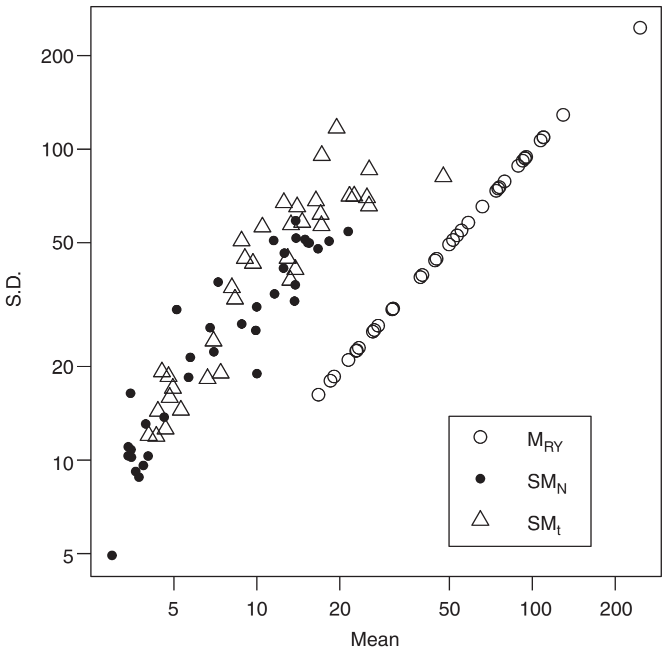


Fig. 3. Mean and standard deviation of the sojourn times. (Both axis are logarithmic.)

For every model, the expected sojourn time is higher in the ‘low-risk state’, where risk is measured in terms of variance of the respective conditional distribution. This seems reasonable because periods of high volatility reflect a nervous market and are historically less persistent than periods of low volatility.

It is remarkable that the average sojourn times for the HSMMs are significantly lower than for $M_{\mathrm{RY}}$, i.e., the persistence of both the high- and the low-risk state is much lower. The higher sojourn times of $SM_t$ w.r.t. $SM_N$ are a consequence of the heavier tails of the component distributions—in most cases the degrees of freedom take values between 5 and 10. On the other hand, the standard deviations of the sojourn time distributions show a smaller difference between the models.

Taking both mean and standard deviation into account, the coefficient of variation equals $1 - \pi_1$ and $1 - \pi_2$ for the two states of $M_{\mathrm{RY}}$. Due to the high persistence of the states, this yields values close to one. However, the average coefficient of variation of the $SM_N$ $(SM_t)$ is 3.23 (3.72). Plotting mean and standard deviation of the sojourn time against each other shows a clear separation of the HMM and the two HSMMs, as shown in Fig. 3.

## 7. Conclusion

In this paper we present a generalization of the approach of RY to model daily return series, and improve the distributional and temporal dependence properties described by Granger and Ding (1995a). In particular, we show that the one stylized fact, which could not be reproduced by a HMM, namely the slowly decaying autocorrelation function, can be described by a HSMM with negative binomial sojourn time and normal conditional distributions.

The negative binomial sojourn time distribution is only one of the many possibilities to extend a HMM. It certainly seems to be a more appropriate choice than the geometric sojourn time distribution of a HMM. Future research may show that other, parametric or non-parametric alternatives are even better.

## Acknowledgments

We would like to thank Y. Guédon, P. Thomson, and W. Zucchini for their helpful comments and Karthinathan Thangavelu for his editorial assistance. Ingo Bulla was supported by a fellowship within the Postdoc-Programme of the German Academic Exchange Service (DAAD).

## Appendix A. Estimation results

All estimation results for the three models are reported in Tables A.1–A.6.

Table A.1
Parameter estimates for the HMM, original data

| Sector | $1-p_1$ | $1-p_2$ | $\mu_1 \times 10^3$ | $\mu_2 \times 10^3$ | $\sigma_1^2 \times 10^4$ | $\sigma_2^2 \times 10^4$ | S.D. $\times 10^2$ | Skew. | Kurt. | AIC |
|---|---|---|---|---|---|---|---|---|---|---|
| Automobiles | 0.968 | 0.991 | −1.78 | 0.585 | 6.29 | 0.978 | 1.47 | −0.203 | 6.16 | −28 473 |
| Banks | 0.957 | 0.987 | −1.2 | 0.667 | 4.43 | 0.502 | 1.19 | −0.234 | 7.08 | −31 114 |
| Basics | 0.948 | 0.982 | −1.49 | 0.803 | 4.16 | 0.635 | 1.25 | −0.237 | 5.99 | −30 061 |
| Chemicals | 0.953 | 0.98 | −0.925 | 0.761 | 3.95 | 0.657 | 1.29 | −0.164 | 5.51 | −29 645 |
| Construction | 0.962 | 0.989 | −1.42 | 0.766 | 3.26 | 0.601 | 1.09 | −0.229 | 5.6 | −31 058 |
| Financials | 0.957 | 0.987 | −1.23 | 0.64 | 4.17 | 0.46 | 1.16 | −0.243 | 7.17 | −31 483 |
| Food | 0.946 | 0.985 | −0.87 | 0.608 | 3.65 | 0.514 | 1.1 | −0.18 | 6.48 | −31 398 |
| Healthcare | 0.963 | 0.981 | −0.622 | 0.842 | 3.83 | 0.663 | 1.31 | −0.141 | 5.25 | −29 407 |
| Industrials | 0.975 | 0.989 | −0.845 | 0.757 | 3.29 | 0.506 | 1.16 | −0.183 | 5.7 | −30 931 |
| Insurance | 0.964 | 0.989 | −1.14 | 0.496 | 6.79 | 0.739 | 1.45 | −0.167 | 7.3 | −29 360 |
| Media | 0.968 | 0.989 | −1.27 | 0.74 | 6.52 | 0.778 | 1.49 | −0.199 | 6.81 | −28 989 |
| Oil and gas | 0.983 | 0.992 | −0.195 | 0.651 | 3.48 | 0.666 | 1.24 | −0.0801 | 5.14 | −29 955 |
| Personal | 0.956 | 0.991 | −1.3 | 0.681 | 3.37 | 0.599 | 1.05 | −0.208 | 5.83 | −31 418 |
| Retail | 0.978 | 0.987 | −1.17 | 0.923 | 8.17 | 0.822 | 1.87 | −0.164 | 6.07 | −27 222 |
| Technology | 0.981 | 0.989 | −0.546 | 0.843 | 5.51 | 1.01 | 1.63 | −0.101 | 4.98 | −27 351 |
| Telecom | 0.94 | 0.977 | −0.493 | 0.495 | 5.56 | 0.922 | 1.48 | −0.0826 | 5.67 | −20 559 |
| Travel | 0.986 | 0.996 | −0.962 | 0.598 | 5.27 | 1.25 | 1.48 | −0.102 | 4.81 | −20 337 |
| Utilities | 0.975 | 0.991 | −0.743 | 0.547 | 3.59 | 0.627 | 1.19 | −0.133 | 5.58 | −22 230 |

Estimated parameters of the HMM with normal component distributions for the 18 DJ STOXX sector indices. For state $i$, $i=1,2$, $p_i$ is the parameter of the sojourn time distribution, $\mu_i$ and $\sigma_i^2$ model the component distribution. Skewness and kurtosis in this and the subsequent tables were estimated by Monte Carlo approximation.

Table A.2
Parameter estimates for the HMM, outlier-corrected data

| Sector | $1-p_1$ | $1-p_2$ | $\mu_1 \times 10^3$ | $\mu_2 \times 10^3$ | $\sigma_1^2 \times 10^4$ | $\sigma_2^2 \times 10^4$ | S.D. $\times 10^2$ | Skew | Kurt. | AIC |
|---|---|---|---|---|---|---|---|---|---|---|
| Automobiles | 0.971 | 0.99 | −1.36 | 0.586 | 5.14 | 0.932 | 1.41 | −0.164 | 5.52 | −28 607 |
| Banks | 0.962 | 0.986 | −0.91 | 0.658 | 3.6 | 0.479 | 1.14 | −0.189 | 6.34 | −31 261 |
| Basics | 0.956 | 0.98 | −1.08 | 0.835 | 3.28 | 0.588 | 1.19 | −0.194 | 5.3 | −30 222 |
| Chemicals | 0.964 | 0.98 | −0.685 | 0.775 | 3.23 | 0.607 | 1.24 | −0.137 | 4.99 | −29 786 |
| Construction | 0.967 | 0.988 | −1 | 0.789 | 2.6 | 0.571 | 1.05 | −0.183 | 4.97 | −31 205 |
| Financials | 0.964 | 0.987 | −0.907 | 0.653 | 3.35 | 0.442 | 1.1 | −0.201 | 6.37 | −31 650 |
| Food | 0.971 | 0.988 | −0.446 | 0.629 | 2.34 | 0.467 | 1.02 | −0.12 | 5.06 | −31 670 |
| Healthcare | 0.971 | 0.982 | −0.414 | 0.855 | 3.22 | 0.629 | 1.27 | −0.113 | 4.84 | −29 569 |
| Industrials | 0.978 | 0.989 | −0.621 | 0.788 | 2.72 | 0.476 | 1.11 | −0.159 | 5.23 | −31 099 |
| Insurance | 0.969 | 0.989 | −0.94 | 0.512 | 5.5 | 0.703 | 1.39 | −0.152 | 6.52 | −29 503 |
| Media | 0.97 | 0.989 | −0.975 | 0.752 | 5.43 | 0.76 | 1.41 | −0.165 | 6.21 | −29 161 |
| Oil and gas | 0.986 | 0.993 | −0.0357 | 0.632 | 3.12 | 0.652 | 1.21 | −0.0602 | 4.86 | −30 064 |
| Personal | 0.961 | 0.989 | −1.12 | 0.726 | 2.73 | 0.571 | 1.02 | −0.192 | 5.22 | −31 514 |
| Retail | 0.978 | 0.987 | −1.06 | 0.907 | 7.62 | 0.807 | 1.83 | −0.155 | 5.92 | −27 317 |
| Technology | 0.984 | 0.989 | −0.461 | 0.863 | 5.01 | 0.977 | 1.6 | −0.0935 | 4.77 | −27 444 |
| Telecom | 0.944 | 0.976 | −0.264 | 0.486 | 4.84 | 0.891 | 1.44 | −0.0624 | 5.28 | −20 644 |
| Travel | 0.988 | 0.996 | −0.816 | 0.604 | 4.92 | 1.24 | 1.46 | −0.0904 | 4.63 | −20 368 |
| Utilities | 0.978 | 0.991 | −0.524 | 0.537 | 3.01 | 0.595 | 1.15 | −0.107 | 5.1 | −22 325 |

Estimated parameters of the HMM with normal component distributions for the 18 DJ STOXX sector indices. For state $i$, $i=1,2$, $p_i$ is the parameter of the sojourn time distribution, $\mu_i$ and $\sigma_i^2$ model the component distribution.

Table A.3
Parameter estimates for the HSMM with normal conditional distributions for original data

| Sector | $1-p_1$ | $1-p_2$ | $r_1 \times 10$ | $r_2 \times 10$ | $\mu_1 \times 10^3$ | $\mu_2 \times 10^3$ | $\sigma_1^2 \times 10^4$ | $\sigma_2^2 \times 10^4$ | S.D. $\times 10^2$ | Skew. | Kurt. | AIC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Automobiles | 0.983 | 0.995 | 0.518 | 0.746 | −1.87 | 0.576 | 6.76 | 0.924 | 1.47 | −0.226 | 6.67 | −28 537 |
| Banks | 0.969 | 0.993 | 0.847 | 0.775 | −1.1 | 0.625 | 4.66 | 0.459 | 1.19 | −0.226 | 7.67 | −31 203 |
| Basics | 0.969 | 0.991 | 0.922 | 0.966 | −1.57 | 0.812 | 4.37 | 0.592 | 1.24 | −0.266 | 6.39 | −30 132 |
| Chemicals | 0.979 | 0.99 | 0.544 | 0.821 | −1.11 | 0.804 | 4.21 | 0.611 | 1.28 | −0.196 | 5.96 | −29 736 |
| Construction | 0.965 | 0.994 | 0.995 | 0.824 | −1.55 | 0.748 | 3.61 | 0.575 | 1.09 | −0.257 | 6.2 | −31 104 |
| Financials | 0.971 | 0.994 | 0.89 | 0.848 | −1.41 | 0.631 | 4.65 | 0.448 | 1.15 | −0.282 | 8.03 | −31 569 |
| Food | 0.919 | 0.995 | 1.75 | 0.629 | −0.95 | 0.581 | 4.11 | 0.5 | 1.09 | −0.192 | 7.28 | −31 464 |
| Healthcare | 0.981 | 0.991 | 0.703 | 0.844 | −0.801 | 0.897 | 4.04 | 0.631 | 1.31 | −0.169 | 5.58 | −29 468 |
| Industrials | 0.988 | 0.993 | 0.728 | 1.08 | −0.94 | 0.793 | 3.38 | 0.475 | 1.16 | −0.198 | 5.96 | −30 979 |
| Insurance | 0.977 | 0.995 | 0.558 | 0.622 | −1.23 | 0.469 | 7.63 | 0.707 | 1.44 | −0.185 | 8.28 | −29 464 |
| Media | 0.992 | 0.993 | 0.471 | 1.43 | −1.33 | 0.749 | 6.71 | 0.742 | 1.48 | −0.212 | 7.12 | −29 055 |
| Oil and gas | 0.986 | 0.996 | 0.644 | 0.476 | −0.367 | 0.705 | 3.7 | 0.633 | 1.23 | −0.11 | 5.49 | −30 006 |
| Personal | 0.976 | 0.993 | 0.615 | 1.18 | −1.62 | 0.711 | 3.71 | 0.575 | 1.04 | −0.262 | 6.45 | −31 454 |
| Retail | 0.996 | 0.99 | 0.28 | 1.23 | −1.32 | 0.96 | 8.53 | 0.777 | 1.86 | −0.187 | 6.42 | −27 331 |
| Technology | 0.99 | 0.987 | 0.495 | 1.18 | −0.503 | 0.825 | 5.7 | 0.873 | 1.63 | −0.103 | 5.32 | −27 423 |
| Telecom | 0.991 | 0.975 | 0.231 | 2.31 | −0.568 | 0.502 | 5.99 | 0.859 | 1.48 | −0.0957 | 6.17 | −20 640 |
| Travel | 0.996 | 0.988 | 0.183 | 1.55 | −0.996 | 0.71 | 5.07 | 1.07 | 1.47 | −0.126 | 5.03 | −20 361 |
| Utilities | 0.98 | 0.996 | 0.488 | 0.429 | −0.793 | 0.522 | 3.97 | 0.6 | 1.17 | −0.148 | 6.19 | −22 279 |

This table reports the estimated parameters of the HSMM with normal component distributions for the 18 DJ STOXX sector indices. For state $i$, $i=1,2$, $\pi_i$ and $r_i$ are the parameters of the sojourn time distribution, $\mu_i$ and $\sigma_i^2$ model the component distribution.

Table A.4
Parameter estimates for the HSMM with normal conditional distributions for outlier-corrected data

| Sector | $1-p_1$ | $1-p_2$ | $r_1 \times 10$ | $r_2 \times 10$ | $\mu_1 \times 10^3$ | $\mu_2 \times 10^3$ | $\sigma_1^2 \times 10^4$ | $\sigma_2^2 \times 10^4$ | S.D. $\times 10^2$ | Skew. | Kurt. | AIC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Automobiles | 0.984 | 0.994 | 0.535 | 0.757 | −1.54 | 0.593 | 5.59 | 0.892 | 1.41 | −0.194 | 5.98 | −28 661 |
| Banks | 0.973 | 0.993 | 0.795 | 0.71 | −0.833 | 0.617 | 3.8 | 0.439 | 1.14 | −0.188 | 6.85 | −31 342 |
| Basics | 0.979 | 0.989 | 0.776 | 1.07 | −1.16 | 0.856 | 3.42 | 0.546 | 1.19 | −0.215 | 5.61 | −30 287 |
| Chemicals | 0.982 | 0.99 | 0.52 | 0.742 | −0.895 | 0.803 | 3.55 | 0.58 | 1.24 | −0.172 | 5.45 | −29 866 |
| Construction | 0.979 | 0.992 | 0.755 | 1.03 | −1.02 | 0.785 | 2.75 | 0.536 | 1.05 | −0.195 | 5.32 | −31 240 |
| Financials | 0.975 | 0.994 | 0.907 | 0.853 | −1.07 | 0.652 | 3.69 | 0.43 | 1.09 | −0.229 | 7.03 | −31 719 |
| Food | 0.981 | 0.993 | 0.722 | 0.706 | −0.529 | 0.635 | 2.51 | 0.446 | 1.02 | −0.137 | 5.43 | −31 712 |
| Healthcare | 0.994 | 0.987 | 0.334 | 1.24 | −0.499 | 0.925 | 3.23 | 0.584 | 1.26 | −0.136 | 4.95 | −29 627 |
| Industrials | 0.992 | 0.991 | 0.554 | 1.33 | −0.695 | 0.814 | 2.8 | 0.451 | 1.11 | −0.174 | 5.42 | −31 143 |
| Insurance | 0.977 | 0.995 | 0.608 | 0.56 | −1.06 | 0.491 | 6.21 | 0.68 | 1.38 | −0.169 | 7.38 | −29 590 |
| Media | 0.993 | 0.992 | 0.42 | 1.51 | −1.01 | 0.757 | 5.58 | 0.722 | 1.41 | −0.177 | 6.48 | −29 223 |
| Oil and gas | 0.993 | 0.995 | 0.397 | 0.609 | −0.171 | 0.694 | 3.23 | 0.61 | 1.21 | −0.0814 | 5.11 | −30 110 |
| Personal | 0.98 | 0.991 | 0.632 | 1.41 | −1.35 | 0.769 | 2.91 | 0.542 | 1.01 | −0.233 | 5.61 | −31 543 |
| Retail | 0.996 | 0.99 | 0.27 | 1.3 | −1.23 | 0.961 | 7.93 | 0.763 | 1.82 | −0.175 | 6.23 | −27 422 |
| Technology | 0.992 | 0.984 | 0.425 | 1.37 | −0.392 | 0.84 | 5.14 | 0.823 | 1.6 | −0.0947 | 5.06 | −27 513 |
| Telecom | 0.99 | 0.969 | 0.254 | 2.57 | −0.336 | 0.504 | 5.17 | 0.816 | 1.44 | −0.0748 | 5.72 | −20 719 |
| Travel | 0.996 | 0.985 | 0.16 | 1.82 | −0.849 | 0.723 | 4.74 | 1.05 | 1.46 | −0.116 | 4.87 | −20 393 |
| Utilities | 0.98 | 0.996 | 0.559 | 0.364 | −0.527 | 0.508 | 3.26 | 0.566 | 1.13 | −0.109 | 5.57 | −22 365 |

Estimated parameters of the HSMM with normal component distributions for the 18 DJ STOXX sector indices. For state $i$, $i=1,2$, $\pi_i$ and $r_i$ are the parameters of the sojourn time distribution, $\mu_i$ and $\sigma_i^2$ model the component distribution.

Table A.5
Parameter estimates for the HSMM with Student $t$ conditional distributions for original data

| Sector | $1-p_1$ | $1-p_2$ | $r_1 \times 10$ | $r_2 \times 10$ | $\mu_1 \times 10^3$ | $\mu_2 \times 10^3$ | $\sigma_1^2 \times 10^4$ | $\sigma_2^2 \times 10^4$ | $v_1$ | $v_2$ | S.D. $\times 10^2$ | Skew. | Kurt. | AIC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Automobiles | 0.985 | 0.997 | 0.588 | 0.511 | −1.56 | 0.622 | 4.52 | 0.769 | 7.55 | 10.9 | 1.46 | −0.191 | 9.54 | −28 594 |
| Banks | 0.977 | 0.996 | 0.791 | 0.464 | −0.685 | 0.628 | 3.04 | 0.382 | 7.05 | 10.3 | 1.18 | −0.162 | 11.5 | −31 258 |
| Basics | 0.983 | 0.996 | 0.962 | 0.55 | −1.04 | 0.816 | 2.28 | 0.482 | 5.53 | 9.94 | 1.24 | −0.201 | 12 | −30 224 |
| Chemicals | 0.987 | 0.995 | 0.516 | 0.401 | −0.585 | 0.739 | 2.5 | 0.485 | 6.71 | 8.86 | 1.27 | −0.122 | 8.76 | −29 814 |
| Construction | 0.982 | 0.996 | 1.15 | 0.935 | −1.03 | 0.805 | 1.97 | 0.49 | 5.9 | 13.5 | 1.08 | −0.188 | 10.1 | −31 170 |
| Financials | 0.979 | 0.996 | 0.902 | 0.678 | −1.01 | 0.671 | 2.81 | 0.366 | 6.33 | 11 | 1.15 | −0.208 | 12.9 | −31 625 |
| Food | 0.994 | 0.993 | 1.4 | 3.28 | −0.263 | 0.658 | 1.35 | 0.37 | 4.37 | 10.4 | 1.08 | −0.0471 | 20.2 | −31 639 |
| Healthcare | 0.997 | 0.991 | 0.238 | 1.05 | −0.32 | 0.932 | 2.3 | 0.516 | 6.61 | 11.9 | 1.3 | −0.112 | 8.26 | −29 564 |
| Industrials | 0.994 | 0.997 | 0.722 | 0.823 | −0.441 | 0.787 | 2 | 0.401 | 6.21 | 12.1 | 1.14 | −0.136 | 10.1 | −31 066 |
| Insurance | 0.979 | 0.997 | 0.659 | 0.399 | −0.892 | 0.516 | 4.9 | 0.575 | 7.2 | 11.1 | 1.44 | −0.158 | 11.9 | −29 512 |
| Media | 0.993 | 0.995 | 0.499 | 1.2 | −1.08 | 0.767 | 4.38 | 0.642 | 6.32 | 12.3 | 1.48 | −0.162 | 12.1 | −29 121 |
| Oil and gas | 0.994 | 0.998 | 0.397 | 0.289 | −0.0489 | 0.64 | 2.52 | 0.526 | 8.15 | 10.6 | 1.23 | −0.0714 | 7.43 | −30 066 |
| Personal | 0.977 | 0.996 | 0.868 | 0.863 | −1.46 | 0.71 | 2.62 | 0.495 | 9.26 | 12.8 | 1.04 | −0.228 | 7.94 | −31 483 |
| Retail | 0.996 | 0.995 | 0.328 | 0.815 | −1.08 | 0.983 | 6.55 | 0.631 | 9.35 | 7.99 | 1.85 | −0.164 | 8.53 | −27 386 |
| Technology | 0.99 | 0.997 | 0.616 | 0.294 | −0.538 | 0.834 | 4.36 | 0.709 | 10.5 | 6.83 | 1.6 | −0.101 | 6.65 | −27 497 |
| Telecom | 0.99 | 0.992 | 0.34 | 0.981 | −0.0699 | 0.465 | 4.35 | 0.742 | 8.29 | 7.9 | 1.47 | −0.0464 | 8.27 | −20 679 |
| Travel | 0.989 | 0.999 | 0.419 | 0.25 | −1.04 | 0.619 | 5.31 | 0.987 | 47 | 10.2 | 1.44 | −0.111 | 5.52 | −20 382 |
| Utilities | 0.984 | 0.997 | 0.562 | 0.285 | −0.288 | 0.5 | 2.43 | 0.503 | 7.98 | 20.7 | 1.16 | −0.0792 | 8.04 | −22 313 |

Estimated parameters of the two-state HSMM with $t$ component distributions for the 18 DJ STOXX sector indices. For state $i$, $i=1,2$, $\pi_i$ and $r_i$ are the parameters of the sojourn time distribution, $\mu_i$, $\sigma_i^2$ and $v_i$ model the component distribution.

Table A.6
Parameter estimates for the HSMM with Student $t$ conditional distributions for outlier-corrected data

| Sector | $1-p_1$ | $1-p_2$ | $r_1 \times 10$ | $r_2 \times 10$ | $\mu_1 \times 10^3$ | $\mu_2 \times 10^3$ | $\sigma_1^2 \times 10^4$ | $\sigma_2^2 \times 10^4$ | $v_1$ | $v_2$ | S.D. $\times 10^2$ | Skew. | Kurt. | AIC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Automobiles | 0.983 | 0.997 | 0.581 | 0.521 | −1.66 | 0.609 | 5.64 | 0.772 | 344 | 10.9 | 1.41 | −0.197 | 6.11 | −28 686 |
| Banks | 0.973 | 0.996 | 0.839 | 0.492 | −0.863 | 0.629 | 3.92 | 0.387 | 347 | 10.2 | 1.13 | −0.191 | 7.04 | −31 363 |
| Basics | 0.978 | 0.996 | 1.01 | 0.521 | −1.13 | 0.806 | 2.84 | 0.487 | 13.3 | 10.1 | 1.19 | −0.196 | 6.53 | −30 308 |
| Chemicals | 0.984 | 0.995 | 0.55 | 0.399 | −0.741 | 0.736 | 3.02 | 0.492 | 16.9 | 8.63 | 1.24 | −0.144 | 6.04 | −29 890 |
| Construction | 0.978 | 0.995 | 1.04 | 0.795 | −1.11 | 0.791 | 2.49 | 0.489 | 22.5 | 14.1 | 1.04 | −0.195 | 5.78 | −31 251 |
| Financials | 0.976 | 0.996 | 0.91 | 0.676 | −1.07 | 0.658 | 3.72 | 0.371 | 345 | 11.2 | 1.09 | −0.227 | 7.12 | −31 736 |
| Food | 0.992 | 0.995 | 0.83 | 1.09 | −0.32 | 0.652 | 1.67 | 0.372 | 8.12 | 10.8 | 1.02 | −0.105 | 7.15 | −31 745 |
| Healthcare | 0.996 | 0.991 | 0.28 | 1.01 | −0.384 | 0.926 | 2.66 | 0.518 | 13.6 | 12.2 | 1.26 | −0.116 | 5.73 | −29 643 |
| Industrials | 0.993 | 0.996 | 0.702 | 0.742 | −0.53 | 0.793 | 2.37 | 0.404 | 14.9 | 12.6 | 1.11 | −0.147 | 6.24 | −31 156 |
| Insurance | 0.978 | 0.997 | 0.635 | 0.438 | −1.01 | 0.505 | 6.23 | 0.585 | 349 | 11.2 | 1.37 | −0.165 | 7.48 | −29 615 |
| Media | 0.993 | 0.995 | 0.51 | 1.19 | −1.02 | 0.755 | 5.62 | 0.645 | 216 | 12.6 | 1.4 | −0.176 | 6.62 | −29 235 |
| Oil and gas | 0.992 | 0.998 | 0.48 | 0.273 | −0.0723 | 0.639 | 2.96 | 0.528 | 26 | 10.5 | 1.2 | −0.0684 | 5.54 | −30 131 |
| Personal | 0.975 | 0.996 | 0.838 | 0.78 | −1.33 | 0.702 | 3.04 | 0.492 | 343 | 12.4 | 1.01 | −0.211 | 5.79 | −31 556 |
| Retail | 0.995 | 0.995 | 0.353 | 0.726 | −1.16 | 0.983 | 7.57 | 0.634 | 34.1 | 7.55 | 1.81 | −0.172 | 6.68 | −27 447 |
| Technology | 0.986 | 0.997 | 0.694 | 0.274 | −0.498 | 0.818 | 5.19 | 0.722 | 64.3 | 6.92 | 1.58 | −0.0991 | 5.43 | −27 553 |
| Telecom | 0.99 | 0.992 | 0.336 | 1.04 | −0.247 | 0.474 | 5.39 | 0.744 | 343 | 7.87 | 1.43 | −0.0561 | 5.83 | −20 734 |
| Travel | 0.991 | 0.998 | 0.466 | 0.418 | −0.947 | 0.631 | 5.16 | 0.981 | 346 | 10.4 | 1.44 | −0.104 | 5.14 | −20 409 |
| Utilities | 0.981 | 0.997 | 0.561 | 0.329 | −0.505 | 0.503 | 3.18 | 0.522 | 107 | 21.5 | 1.13 | −0.106 | 5.68 | −22 366 |

Estimated parameters of the two-state HSMM with $t$ component distributions for the 18 DJ STOXX sector indices. For state $i$, $i=1,2$, $\pi_i$ and $r_i$ are the parameters of the sojourn time distribution, $\mu_i$, $\sigma_i^2$ and $v_i$ model the component distribution.

## Appendix B. Re-estimation formulae

The *M-step* of the EM algorithm requires the maximization of the $Q$-function to determine the next set of parameters. As shown in Section 3.3, the $Q$-function from Eq. (8) can be decomposed into the four different terms (9), (10), (11), and (12) each of which can be maximized separately.

For a more convenient notation, we introduce some auxiliary variables. Let

$$\eta_{ju}^{(k)} := \sum_{t=0}^{\tau-2} P\left(S_{t+u+1} \neq j, S_{t+u-v} = j,\ v = 0, \ldots, u-1, S_t \neq j | X_0^{\tau-1} = x_0^{\tau-1}, \theta^{(k)}\right)$$
$$+ P\left(S_u \neq j, S_{u-v} = j,\ v = 1, \ldots, u \mid X_0^{\tau-1} = x_0^{\tau-1}, \theta^{(k)}\right),$$

$$L_j(t) := P\left(S_t = j \mid X_0^{\tau-1} = x_0^{\tau-1}\right),$$

$$L1_j(t) := P\left(S_{t+1} \neq j, S_t = j \mid X_0^{\tau-1} = x_0^{\tau-1}\right),$$

$$F_j(t) := P\left(S_{t+1} \neq j, S_t = j \mid X_0^t = x_0^t\right),$$

$$G_j(t+1) := \frac{P\left(X_{t+1}^{\tau-1} = x_{t+1}^{\tau-1} \mid S_{t+1} = j, S_t \neq j\right)}{P\left(X_{t+1}^{\tau-1} = x_{t+1}^{\tau-1} \mid X_0^t = x_0^t\right)}.$$

The computation of the auxiliary variables can be combined with the forward-backward algorithm. Details are given in Guédon (2003).

### B.1. The semi-Markov chain

The re-estimation formula for the initial parameters of the semi-Markov chain is

$$\pi_j^{(k+1)} = P\left(S_0 = j \mid X_0^{\tau-1} = x_0^{\tau-1}, \theta^{(k)}\right) = L_j(0),$$

while the transition probabilities of the embedded Markov chain can be written as

$$p_{ij}^{(k+1)} = \frac{\sum_{t=0}^{\tau-2} P\left(S_{t+1} = j, S_t = i \mid X_0^{\tau-1} = x_0^{\tau-1}, \theta^{(k)}\right)}{\sum_{t=0}^{\tau-2} P\left(S_{t+1} \neq i, S_t = i \mid X_0^{\tau-1} = x_0^{\tau-1}, \theta^{(k)}\right)}$$
$$= \frac{\sum_{t=0}^{\tau-2} G_j(t+1) p_{ij} F_i(t)}{\sum_{t=0}^{\tau-2} L1_i(t)}.$$

The estimators for the parameters of the negative binomial distribution sojourn times cannot be computed directly. One obtains the system of equations

$$p_j = \frac{r_j \sum_{u=1}^{\tau-1} \eta_{ju}^{(k)}}{\sum_{u=1}^{\tau-1} \eta_{ju}^{(k)} \left(r_j + u - 1\right)},$$

$$\sum_{u=1}^{\tau-1} \eta_{ju}^{(k)} \left(\psi\left(u - 1 + r_j\right) - \psi\left(r_j\right) + \log\left[\frac{r_j \sum_{v=1}^{\tau-1} \eta_{jv}^{(k)}}{\sum_{v=1}^{\tau-1} \eta_{jv}^{(k)} \left(r_j + v - 1\right)}\right]\right) = 0,$$

where $\Gamma(\cdot)$ denotes the Gamma-function, $\psi(s) := \partial \log \Gamma(s)/\partial s$ the Digamma function and $r_j > 0$, $p_j \in (0, 1)$ the parameters of the distribution function. This equation has to be solved numerically, e.g., by a bisection algorithm.

### B.2. The observation component

Re-estimation formulae for normal component distributions, i.e., $b_j(\cdot) \sim N\left(\mu_j, \sigma_j^2\right)$ are given by Sansom and Thomson (2001)

$$\mu_j^{(k+1)} = \frac{\sum_{t=0}^{\tau-1} L_j(t) x_t}{\sum_{t=0}^{\tau-1} L_j(t)},$$

$$\sigma_j^{2\,(k+1)} = \frac{\sum_{t=0}^{\tau-1} L_j(t) \left(x_t - \mu_j\right)^2}{\sum_{t=0}^{\tau-1} L_j(t)}.$$

The derivation and maximization of the $Q$-function for $t$ distributed variables is somewhat demanding. However, the techniques presented by Peel and McLachlan (2000) for the estimation of mixtures of $t$-distributions can be adopted to the case of a HSMM.

The density of the $t$ distribution with location parameter $\boldsymbol{\mu}$, $v$ degrees of freedom and positive definite inner product matrix $\boldsymbol{\Sigma}$ is given by

$$b_j\left(\boldsymbol{x_t}\right) = \frac{\Gamma\left((v_j + p)/2\right) |\boldsymbol{\Sigma}_j|^{-1/2}}{(\pi v)^{1/2p} \Gamma\left(v_j/2\right) \left\{1 + \delta(\boldsymbol{x}_t, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)/v_j\right\}^{1/2(v_j+p)}},$$

where $\delta(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the Mahalanobis distance, defined by

$$\delta(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\boldsymbol{x} - \boldsymbol{\mu}) \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}),$$

and $p$ the dimension of the observations.

The re-estimation formulae for $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ can be derived explicitly as

$$\boldsymbol{\mu}_j^{(k+1)} = \frac{\sum_{t=0}^{\tau-1} L_j(t) u_{jt}^{(k)} \boldsymbol{x}_t}{\sum_{t=0}^{\tau-1} L_j(t) u_{jt}^{(k)}}$$

and

$$\boldsymbol{\Sigma}_j^{(k+1)} = \frac{\sum_{t=0}^{\tau-1} L_j(t) u_{jt}^{(k)} \left(\boldsymbol{x}_t - \boldsymbol{\mu}_j^{(k+1)}\right) \left(\boldsymbol{x}_t - \boldsymbol{\mu}_j^{(k+1)}\right)^{\mathrm{T}}}{\sum_{t=0}^{\tau-1} L_j(t)}, \tag{B.1}$$

where

$$u_{jt}^{(k)} := \frac{v_j^{(k)} + p}{v_j^{(k)} + \delta\left(\boldsymbol{x}_t^{(k)}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}\right)}.$$

Note that, in the case of a single component $t$ distribution, the denominator of (B.1) can also be replaced by $\sum_{t=0}^{\tau-1} L_j(t) u_{jt}^{(k)}$ to increase the speed of convergence (Kent et al., 1994).

The estimator $v_j^{(k+1)}$ is the (unique) solution of the equation

$$-\psi\left(\frac{1}{2} v_j^{(k)}\right) + \log\left(\frac{1}{2} v_j^{(k)}\right) + 1 + \frac{1}{\sum_{t=0}^{\tau-1} L_j(t)} \left[\sum_{t=0}^{\tau-1} L_j(t) \left(\log u_{jt}^{(k)} - u_{jt}^{(k)}\right)\right]$$
$$+\psi\left(\frac{v_j^{(k)} + p}{2}\right) - \log\left(\frac{v_j^{(k)} + p}{2}\right) = 0.$$

This can be solved, e.g., by a bisection algorithm or quasi-Newton methods as the function on the left hand side is monotonically increasing in $v_j^{(k)}$.

## References


Baum, L.E., Petrie, T., Soules, G., Weiss, N., 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann. Math. Statist. 41, 164–171.

Bialkowski, J., 2003. Modelling returns on stock indices for western and central European stock exchanges—Markov switching approach. Southeast European J. Econom. 2 (2), 81–100.

Blattberg, R.C., Gonedes, N.J., 1974. A comparison of the stable and student distributions as statistical models for stock prices. J. Business 52 (1–2), 244–280.

Bollerslev, T., Chou, R.Y., Kroner, K.F., 1992. Arch modeling in finance: a review of the theory and empirical evidence. J. Econometrics 47 (2), 5–59.

Cecchetti, S.G., Lam, P.-S., Mark, N.C., 1990. Mean reversion in equilibrium asset prices. Amer. Econom. Rev. 80 (3), 398–418.

Chan, W.-S., 1995. Time series outliers and spurious autocorrelations. J. Appl. Statist. Sci. 2 (2), 153–162.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Statist. Soc. Ser. B 39 (1), 1–38 (with discussion).

Faff, R.W., Hillier, D., Hillier, J., 2000. Time varying beta risk: an analysis of alternative modelling techniques. J. Business Finance Account 27 (5), 523–554.

Fama, E.F., 1965. The behavior of stock-market prices. J. Business 38 (1), 34–105.

Ferguson, J.D., 1980. Variable duration models for speech. Proceedings of the Symposium on the Applications of Hidden Markov Models to Text and Speech, Princeton, New Jersey, pp. 143–179.

Franses, P.H., van Dijk, D., 2000. Non-linear Time Series Models in Empirical Finance. Cambridge University Press, Cambridge, UK.

Gettinby, G., Sinclair, C., Power, D., Brown, R., 2004. An analysis of the distribution of extreme share returns in the UK from 1975 to 2000. J. Business Finance Account 31 (5), 607–646.

Granger, C.W., Ding, Z., 1995a. Some properties of absolute return: an alternative measure of risk. Ann. Econom. Statist. 40, 67–91.

Granger, C.W., Ding, Z., 1995b. Stylized facts on the temporal and distributional properties of daily data from speculative markets, Department of Economics, University of California, San Diego, unpublished paper.

Granger, C.W., Ding, Z., Spear, S., 2000. Stylized facts on the temporal and distributional properties of absolute returns: an update. Proceedings of the Hong Kong International Workshop on Statistics and Finance: An Interface, Imperial College Press, pp. 97–120.

Guédon, Y., 2003. Estimating hidden semi-Markov chains from discrete sequences. J. Comput. Graph. Statist. 12 (3), 604–639.

Guédon, Y., Cocozza-Thivent, C., 1990. Explicit state occupancy modelling by hidden semi-Markov models: application of Derin's scheme. Comput. Speech Language 4, 167–192.

Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57 (2), 357–384.

Hamilton, J.D., 1990. Analysis of time series subject to changes in regime. J. Econometrics 45 (1–2), 39–70.

Harris, R.D., Küçüközmen, C.C., 2001. The empirical distribution of UK and US stock returns. J. Business Finance Account 28 (5–6), 715–740.

Kent, J.T., Tyler, D.E., Vardi, Y., 1994. A curious likelihood identity for the multivariate $t$-distribution. Comm. Statist. Simulation Comput. 23 (2), 441–453.

Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the s&p 100 stock index using historical, realised and implied volatility measurements. J. Empirical Finance 12 (3), 445–475.

Levinson, S.E., 1986. Continuously variable duration hidden Markov models for automatic speech recognition. Comput. Speech Language 1, 29–45.

Linne, T., 2002. A Markov switching model of stock returns: an application to the emerging markets in central and eastern Europe. In: East European Transition and EU Enlargement. Physica-Verlag, pp. 371–384.

MacDonald, I.L., Zucchini, W., 1997. Hidden Markov and Other Models for Discrete-valued Time Series. Monographs on Statistics and Applied Probability, vol. 70. Chapman & Hall, London.

Mittnik, S., Rachev, S.T., 1993. Modeling asset returns with alternative stable distributions. Econometric Rev. 12 (3), 261–389.

Peel, D., McLachlan, G.J., 2000. Robust mixture modelling using the $t$-distribution. Statist. Comput. 10, 339–348.

Praetz, P.D., 1972. The distribution of share price changes. J. Business 45 (1), 49–55.

Rabiner, L., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. IEEE Trans. Inform. Theory 77 (2), 257–284.

Rydén, T., Terasvirta, T., Asbrink, S., 1998. Stylized facts of daily return series and the hidden Markov model. J. Appl. Econom. 13 (3), 217–244.

Sansom, J., Thomson, P., 2001. Fitting hidden semi-Markov models to breakpoint rainfall data. J. Appl. Probab. 38A, 142–157.

Taylor, S.J., 1986. Modelling Financial Time Series. Wiley, Chichester, UK.

Turner, C.M., Startz, R., Nelson, C.R., 1989. A Markov model of heteroskedasticity, risk, and learning in the stock market. J. Financial Econom. 25 (1), 3–22.

Yao, J., Gao, J., 2004. Computer-intensive time-varying model approach to the systematic risk of Australian industrial stock returns. Austral. J. Manage. 29 (1), 121–146.

Yu, J., 2002. Forecasting volatility in the New Zealand stock market. Appl. Financial Econom. 12, 193–202.

Yu, S.-Z., Kobayashi, H., 2003. An efficient forward–backward algorithm for an explicit-duration hidden Markov model. Institute of Electrical and Electronics Engineers. Signal Process. Lett. 10 (1), 11–14.