Detecting bearish and bullish markets in financial time series using hierarchical hidden Markov models

Lennart Oelschläger¹ and Timo Adam¹,²
¹Department of Business Administration and Economics, Bielefeld University, Bielefeld, Germany
²School of Mathematics and Statistics, University of St Andrews, St Andrews, United Kingdom

Abstract: Financial markets exhibit alternating periods of rising and falling prices. Stock traders seeking to make profitable investment decisions have to account for those trends, where the goal is to accurately predict switches from bullish to bearish markets and vice versa. Popular tools for modelling financial time series are hidden Markov models, where a latent state process is used to explicitly model switches among different market regimes. In their basic form, however, hidden Markov models are not capable of capturing both short- and long-term trends, which can lead to a misinterpretation of short-term price fluctuations as changes in the long-term trend. In this article, we demonstrate how hierarchical hidden Markov models can be used to draw a comprehensive picture of market behaviour, which can contribute to the development of more sophisticated trading strategies. The feasibility of the suggested approach is illustrated in two real-data applications, where we model data from the Deutscher Aktienindex and the Deutsche Bank stock. The proposed methodology is implemented in the R package fHMM, which is available on CRAN.

Key words: decoding market behaviour, hidden Markov models, state-space models, temporal resolution, time series modelling

Received 5 October 2019; Revised 5 April 2021; Accepted 5 June 2021

1 Introduction

Earning money with stock trading is simple: one only needs to buy and sell stocks at the right moment. In general, stock traders seek to invest at the beginning of upward trends (hereon termed as bullish markets) and repel their stocks just in time before the prices fall again (hereon termed as bearish markets). As stock prices depend on a variety of environmental factors (Humpe and Macmillan, 2009; Cohen et al., 2013), chance certainly plays a fundamental role in hitting those exact moments. However, investigating market behaviour can lead to a better
understanding of how trends alternate and thereby increases the chance of making profitable investment decisions. In this article, we aim at contributing to those investigations by applying the hierarchical extension of hidden Markov models (HMMs) (Fine et al., 1998) to detect bearish and bullish markets in financial time series.

Over the last decades, various HMM-type models have emerged as popular tools for modelling financial time series that are subject to state-switching over time (Schaller and van Norden, 1997; Dias et al., 2009; Ang and Timmermann, 2012; De Angelis and Paas, 2013). Rydåén et al. (1998), Bulla and Bulla (2006), and Nystrup et al. (2015), for example, used HMMs to derive stylized facts of stock returns, while Hassan and Nath (2005) and Nystrup et al. (2017) demonstrated that HMMs can prove useful for economic forecasting. More recently, Lihn (2017) applied HMMs to the Standard and Poor’s 500 (S&P 500), where HMMs were used to identify different levels of market volatility, aiming at providing evidence for the conjecture that returns exhibit negative correlation with volatility. Another application to the S&P 500 can be found in Nguyen (2018), where HMMs were used to predict monthly closing prices to derive an optimal trading strategy, which was shown to outperform the conventional buy-and-hold strategy. Further applications, which involve HMM-type models for asset allocation and portfolio optimization, can be found in Ang and Bekaert (2002), Bulla et al. (2011), and Nystrup et al. (2015, 2018), to name but a few examples. All these applications demonstrate that HMMs constitute a versatile class of time series models that naturally accounts for the dynamics typically exhibited by financial time series.

However, in their basic form HMMs operate on a single time scale, with observations made, for example, on a quarterly, monthly, or daily basis. As a consequence, conventional HMMs are not capable of capturing both short- and long-term trends. Furthermore, the temporal resolution of the data strongly determines the kind of inference that can be made. This is to be regarded as a major deficit, as short-term price fluctuations can easily be misinterpreted as changes in the long-term trend and hence may draw a distorted picture of market behaviour. While models for mixed frequency modelling exist (see Guerin and Marcellino, 2013), or approaches that penalize state-switches to force the model to focus on changes in the long-term trend rather than short-term price fluctuations (see Nystrup et al., 2020), we propose that this issue can to some extent be overcome by adding a hierarchical structure to the basic HMM, which improves the model’s capability for distinguishing between short- and long-term trends and allows to interpret market dynamics at multiple time scales.

The hierarchical extension of HMMs originates from supervised machine learning. In Fine et al. (1998), hierarchical HMMs (HHMMs) were first applied to solve handwriting recognition tasks, where hierarchical levels appear as single letters, syllables, and words. More recently, Leos-Barajas et al. (2017) and Adam et al. (2019) used HHMMs for animal movement modelling, where the hierarchical structure was exploited to jointly infer animals’ behavioural modes at fine and coarse scales, respectively. Here, we demonstrate that the same model structure has also strong potential for modelling financial time series, where it is often of major interest to
distinguish short-term (e.g., daily) price fluctuations from long-term trends (e.g., bullish and bearish markets lasting several months).

The article is structured as follows: In Section 2, we introduce the methodology, discuss how HHMMs can be fitted via numerical likelihood maximization, and briefly outline related topics such as state decoding and model checking. In Section 3, we illustrate the feasibility of the suggested approach by modelling time series of the DAX and the Deutsche Bank stock, followed by a simulation study and a comparison to the basic HMM. In Section 4, we conclude with a discussion and outline possible avenues for future research. Pseudo-code for the proposed methodology is appended, while an implementation is provided in the R package fHMM (Oelschläger and Adam, 2021).

2 Methodology

In this section, we introduce the model formulation and dependence structure, state a formula for the likelihood, and discuss its numerical maximization to obtain estimates for the model parameters. In addition, we outline the Viterbi algorithm, which allows to decode the hidden states underlying the observations. Model checking based on pseudo-residuals is discussed at the end of the section.

2.1 Model formulation and dependence structure

HMMs constitute a versatile class of statistical models for time series, see, for example, Zucchini et al. (2016) for a comprehensive introduction. They predicate that the behaviour of the nature can be divided into a finite number of states, where the state that is active cannot be directly observed. However, at each point in time, a data point is observed, which depends on the current state of the nature and thus yields information on the latter. More formally, this concept can be formulated by introducing two stochastic processes:

1. At each time point $t$ of the discrete time space $\{1, \ldots, T\}$, an underlying process $(S_t)_t$ selects one state from the state space $\{1, \ldots, N\}$. We call $(S_t)_t$ the hidden state process.

2. Depending on which state is active at $t$, one of $N$ distributions $f^{(1)}, \ldots, f^{(N)}$ generates the observation $X_t$. The process $(X_t)_t$ is called the observed state-dependent process.

We make the following assumptions on these processes:

1. We assume that $(S_t)_t$ is a time-homogeneous Markov process of first order. The process is therefore identified by its initial distribution $\delta$ and its transition probability matrix (t.p.m.) $\Gamma$. 
2. The process \((X_t)_t\) is said to satisfy the conditional independence assumption, that is, conditionally on the current state \(S_t\), the observation \(X_t\) is independent of all other states and observations.

From a practical point of view, it is reasonable to identify the initial distribution of \((S_t)_t\) with its stationary distribution \(\pi\) (which we assume to exist): On the one hand, the hidden state process has been evolving for some time before we start to observe it and hence can be assumed to be stationary. On the other hand, \(\pi\) is determined by \(\Gamma\) through the equation \(\pi \Gamma = \pi\), where setting \(\delta = \pi\) reduces the number of parameters that need to be estimated, which is convenient from a computational perspective (Zucchini et al., 2016).

In case of financial data, the hidden states can be interpreted as different moods of the market. Even though these moods cannot be observed directly, price changes—which clearly depend on the current mood of the market—can be observed. Thereby, using an underlying Markov process, we can detect which mood is active at any point in time and how the different moods alternate. Depending on the current mood, a price change is generated by a different distribution. These distributions characterize the moods in terms of expected return and volatility.

The HMM can be extended by a hierarchical structure, resulting in the HHMM. Throughout this article, two hierarchies are considered. Assume that we are dealing with two time series observed on two different time scales. For each observation of the time series on the coarser scale, we have several observations of the times series on the finer scale, for example, monthly observations and corresponding daily observations. Following the concept of HMMs, we can model both state-dependent time series jointly. First, we treat the time series on the coarser scale as stemming from an ordinary HMM, which we refer to as the coarse-scale HMM:

1. At each time point \(t\) of the coarse-scale time space \(\{1, \ldots, T\}\), an underlying process \((S_t)_t\) selects one state from the coarse-scale state space \(\{1, \ldots, N\}\). We call \((S_t)_t\) the hidden coarse-scale state process.

2. Depending on which state is active at \(t\), one of \(N\) distributions \(f^{(1)}, \ldots, f^{(N)}\) realizes the observation \(X_t\). The process \((X_t)_t\) is called the observed coarse-scale state-dependent process.

The processes \((S_t)_t\) and \((X_t)_t\) have the same properties as before, namely \((S_t)_t\) is a first-order Markov process and \((X_t)_t\) satisfies the conditional independence assumption.

Subsequently, we segment the observations of the fine-scale time series into \(T\) distinct chunks, each of which contains all data points that correspond to the \(t\)th coarse-scale time point. Assuming that we have \(T^*\) fine-scale observations on every coarse-scale time point, we face \(T\) chunks comprising of \(T^*\) fine-scale observations each. The hierarchical structure now evinces itself as we model each of the chunks by one of \(N\) possible fine-scale HMMs. Each of the fine-scale HMMs has its own
Figure 1  Dependence structure of the HHMM

t.p.m. $\Gamma^{*(i)}$, initial distribution $\delta^{*(i)}$, stationary distribution $\pi^{*(i)}$, and state-dependent distributions $f^{*(i,1)}, \ldots, f^{*(i,N^*)}$. Which fine-scale HMM is selected to explain the $t$th chunk of fine-scale observations depends on the hidden coarse-scale state $S_t$. The $i$th fine-scale HMM explaining the $t$th chunk of fine-scale observations consists of the following two stochastic processes:

1. At each time point $t^*$ of the fine-scale time space $\{1, \ldots, T^*\}$, the process $(S_{t^*,t^*})_{t^*}$ selects one state from the fine-scale state space $\{1, \ldots, N^*\}$. We call $(S_{t^*,t^*})_{t^*}$ the hidden fine-scale state process.

2. Depending on which state is active at $t^*$, one of $N^*$ distributions $f^{*(i,1)}, \ldots, f^{*(i,N^*)}$ realizes the observation $X_{t^*,t^*}$. The process $(X_{t^*,t^*})_{t^*}$ is called the observed fine-scale state-dependent process.

The fine-scale processes $(S_{1,t^*})_{t^*}, \ldots, (S_{T^*,t^*})_{t^*}$ and $(X_{1,t^*})_{t^*}, \ldots, (X_{T^*,t^*})_{t^*}$ satisfy the Markov property and the conditional independence assumption, respectively, as well. Furthermore, it is assumed that the fine-scale HMM explaining $(X_{t^*,t^*})_{t^*}$ only depends on $S_t$. The dependence structure of the HHMM with two hierarchies is visualized in Figure 1.
2.2 Likelihood evaluation and numerical maximization

Conceptually, an HHMM can be treated as an HMM with two conditionally independent observations, the coarse-scale observation on the one hand and the corresponding chunk of fine-scale observations connected to a fine-scale HMM on the other hand. To derive the likelihood of an HHMM, we start by stating the likelihood formula for the fine-scale HMMs.

Assume that we want to fit the $i$th fine-scale HMM, with model parameters $\theta^{(i)} = (\delta^{(i)}, \Gamma^{(i)}, (f^{*(i,k)})_k)$, to the $t$th chunk of fine-scale observations, $(X_{t,t^*})_{t^*}$. Consider the so-called fine-scale forward probabilities $\alpha^{*(i)}_{k,t^*} = f^{*(i)}(X_{t,1}^*, \ldots, X_{t,t^*}^*, S_{t,t^*}^* = k)$, where $t^* = 1, \ldots, T^*$ and $k = 1, \ldots, N^*$. Obviously,

$$
\mathcal{L}^{\text{HHMM}}(\theta^{(i)} | (X^*_{t,t^*})_{t^*}) = \sum_{k=1}^{N^*} \alpha^{*(i)}_{k,T^*}.
$$

The forward probabilities can be calculated in a recursive way of linear complexity:

$$
\alpha^{*(i)}_{k,1} = \delta^{*(i)}_k f^{*(i,k)}(X^*_{t,1}) \quad \text{and} \quad \alpha^{*(i)}_{k,t^*} = f^{*(i,k)}(X^*_{t,t^*}) \sum_{j=1}^{N^*} \gamma^{*(i)}_{j,k} \alpha^{*(i)}_{j,t^*-1}, \quad t^* = 2, \ldots, T^*.
$$

The transition from the likelihood function of an HMM to the likelihood function of an HHMM is straightforward: Consider the so-called coarse-scale forward probabilities $\alpha_{i,t} = f(X_1, \ldots, X_t, (X^*_{t,t^*})_{t^*}, \ldots, (X^*_{T,t^*})_{t^*}, S_t = i)$, where $t = 1, \ldots, T$ and $i = 1, \ldots, N$. The likelihood function of the HHMM results as

$$
\mathcal{L}^{\text{HHMM}}(\theta, (\theta^{(i)})_i | (X_t)_t, ((X^*_t, t^*_t)_t)_t) = \sum_{i=1}^{N} \alpha_{i,T}.
$$

The coarse-scale forward probabilities can be calculated similarly by applying the recursive scheme

$$
\alpha_{i,1} = \delta_i \mathcal{L}^{\text{HHMM}}(\theta^{(i)} | (X^*_{1,t^*})_{t^*}) f^{(i)}(X_1),
$$

$$
\alpha_{i,t} = \mathcal{L}^{\text{HHMM}}(\theta^{(i)} | (X^*_{t,t^*})_{t^*}) f^{(i)}(X_t) \sum_{j=1}^{N} \gamma_{ij} \alpha_{j,t-1}, \quad t = 2, \ldots, T.
$$

Maximization of the likelihood function is numerically feasible using the Newton-Raphson method. In practice, we often face the technical issues such as numerical under- or overflow, which can be addressed by maximizing the logarithm of the likelihood and incorporating constants in a conducive way. Instead of computing the forward probabilities directly, we consider the logarithmic transformation $\phi^{*(i)}_{k,t^*} = \log[\alpha^{*(i)}_{k,t^*}]$ and $\phi_{i,t} = \log[\alpha_{i,t}]$ thereof (log-forward probabilities). The recursive
form described above remains: The fine-scale log-forward probabilities satisfy
\[
\phi_{k,1}^{*}(i) = \log[\delta_{k}^{*}(i)] + \log[f^{*}(i,k)(X_{1,*})],
\]
\[
\phi_{k,t}^{*}(i) = \log[f^{*}(i,k)(X_{t,*})] + \log \left[ \sum_{j=1}^{N_{*}} \gamma_{j,k}^{*}(i) \exp[\phi_{j,t-1}^{*} - c_{t-1}] \right] + c_{t-1},
\]
where \( c_{t-1} = \max\{\phi_{1,t-1}^{*}, \ldots, \phi_{N_{*},t-1}^{*}\} \) and \( t^{*} = 2, \ldots, T^{*} \). The log-likelihood of a fine-scale HMM results from these variables as
\[
\log \mathcal{L}^{\text{HMM}}(\theta^{*}(i) \mid (X_{t,*})_{t^{*}}) = \log \left[ \sum_{k=1}^{N_{*}} \exp[\phi_{k,T^{*}}^{*} - c_{T^{*}}] \right] + c_{T^{*}},
\]
where \( c_{T^{*}} = \max\{\phi_{1,T^{*}}, \ldots, \phi_{N_{*},T^{*}}\} \). See Algorithm 4 in the appendix for pseudo-code of the computation. The coarse-scale log-forward probabilities satisfy
\[
\phi_{i,1} = \log[\delta_{i}] + \log \mathcal{L}^{\text{HMM}}(\theta^{*}(i) \mid (X_{1,*}, t_{*}^{*})_{t^{*}}) + \log[f^{*}(i)(X_{1})],
\]
\[
\phi_{i,t} = \log \mathcal{L}^{\text{HMM}}(\theta^{*}(i) \mid (X_{t,*})_{t^{*}}) + \log[f^{*}(i)(X_{t})] + \log \left[ \sum_{j=1}^{N} \gamma_{j,i} \exp[\phi_{j,t-1} - c_{t-1}] \right] + c_{t-1},
\]
where \( c_{t-1} = \max\{\phi_{1,t-1}, \ldots, \phi_{N,t-1}\} \) and \( t = 2, \ldots, T \). The log-likelihood of the HHMM results from these variables as
\[
\log \mathcal{L}^{\text{HHMM}}(\theta, (\theta^{*}(i))_{i}, (X_{t})_{t}, ((X_{t,*})_{t^{*}})_{i}) = \log \left[ \sum_{i=1}^{N} \exp[\phi_{i,T} - c_{T}] \right] + c_{T},
\]
where \( c_{T} = \max\{\phi_{1,T}, \ldots, \phi_{N,T}\} \). See Algorithm ?? in the appendix for a pseudo-code.

Additionally, we have to consider that certain model parameters must satisfy constraints, namely the transition probabilities and potentially parameters of the state-dependent distributions. Using parameter transformations serves the purpose. To ensure that the entries of the t.p.m.s fulfill non-negativity and the unity condition, we use a bijective transformation from the real numbers to the unity interval. Rather than estimating the probabilities \( (\gamma_{ij})_{i,j} \) directly, we estimate unconstrained values \( (\eta_{ij})_{i,j} \) for the non-diagonal entries of \( \Gamma \) and derive the probabilities using the multinomial logit link
\[
\gamma_{ij} = \frac{\exp[\eta_{ij}]}{1 + \sum_{k \neq i} \exp[\eta_{ik}]} , \quad i \neq j.
\]
The diagonal entries result from the unity condition \( \gamma_{ii} = 1 - \sum_{j \neq i} \gamma_{ij} \). Noteworthy, not \( N^{2} \) but \( N(N - 1) \) parameters have to be estimated for an \( N \times N \)-t.p.m. Furthermore,
variances are strictly positive, which can be achieved by applying an exponential transformation to the unconstrained estimator. For basic HMMs, Alexandrovich et al. (2016) show that identifiability holds when the state-dependent distributions are distinct and the t.p.m. is ergodic and has full rank, conditions that are usually fulfilled in practice. Given that the fine-scale HMMs are distinct, this also holds for HHMMs (for a discussion of identifiability in HHMMs, see Leos-Barajas, 2019).

A third source of conflicts arises from the fact that the likelihood is maximized with respect to a relatively large number of parameters, which can lead to local maxima apart from the global maximum. Common Newton-Raphson-type optimization routines are unable to distinguish local maxima from the global one. To avoid the trap of ending up at a local maximum, we recommended to run the maximization routine multiple times from different, possibly randomly selected starting points, and to choose the model that corresponds to the highest likelihood. A reasonable set of starting points can be chosen based on the observed data, for example, using the method of moments estimator. Due to the increasingly complex likelihood surface, the number of optimization runs should increase with the number of parameters.

2.3 State decoding

In practice, the primary interest often lies in decoding the hidden states. The term \( \arg \max_{S_1, \ldots, S_T} f(S_1, \ldots, S_T | X_1, \ldots, X_T) \) represents the most-likely underlying state sequence \((S_t)_t\) of an HMM given the data \( (X_t)_t \), which is equivalent to the expression \( \arg \max_{S_1, \ldots, S_T} f(S_1, \ldots, S_T, X_1, \ldots, X_T) \). This can be computed using the Viterbi algorithm (see Zucchini et al., 2016), which is based on the variables

\[
\xi_{i,t} = \max_{S_1, \ldots, S_{t-1}} f(S_1, \ldots, S_{t-1}, S_t = i, X_1, \ldots, X_t),
\]

\( t = 1, \ldots, T \) and \( i = 1, \ldots, N \), which can be calculated recursively via

\[
\xi_{i,1} = \delta_i f^{(i)}(X_1) \quad \text{and} \quad \xi_{i,t} = \max_j (\xi_{j,t-1} \gamma_{ji}) f^{(i)}(X_t).
\]

Obtaining the most-likely state sequence \((\hat{S}_t)_t\) is feasible using these variables, starting at the end of the time horizon and going backwards in time:

\[
\hat{S}_T = \arg \max_i \xi_{i,T} \quad \text{and} \quad \hat{S}_t = \arg \max_i \xi_{i,t} \gamma_{\hat{S}_{t+1}}, \quad t = T - 1, \ldots, 1.
\]

As for the likelihood function, we need to prevent numerical conflicts. Therefore, we again apply a logarithmic transformation, see Algorithm 4 in the appendix, where \( \kappa_{i,t} = \log[\xi_{i,t}] \).

State decoding in HHMMs is straightforward by first decoding the coarse-scale state process and then, using this information, to decode the fine-scale state process afterwards (see Adam et al., 2019).
2.4 Model checking

Analysing so-called pseudo-residuals enables us to check whether a fitted HMM describes the data sufficiently well. This cannot be done by standard residual analysis since the observations are explained by different distributions, depending on the active state. Therefore, all observations have to be transformed on a common scale, which can be achieved in the following way: If $X_t$ has the invertible distribution function $F_{X_t}$, then $Z_t = \Phi^{-1}(F_{X_t}(X_t))$ is standard normally distributed, where $\Phi$ denotes the cumulative distribution function of the standard normal distribution. The observations, $(X_t)_t$, are modelled well if the pseudo-residuals, $(Z_t)_t$, are approximately standard normally distributed (Zucchini et al., 2016).

For HHMMs, we first decode the coarse-scale state process using the Viterbi algorithm (see Section 2.3). Subsequently, we assign each coarse-scale observation its associated distribution function under the fitted model and perform the transformation described above. Using the Viterbi-decoded coarse-scale states, we then treat the fine-scale observations analogously.

3 Application to stock market data

Stock market indices are computed as weighted averages over the stock prices of several companies in the market, thereby constituting a convenient measure of the overall market behaviour. In this section, HHMMs are applied to the DAX and the Deutsche Bank stock, pursuing the goal of detecting long-term trends along with short-term dynamics. The data were downloaded from www.finance.yahoo.com on 26 March 2021.

3.1 DAX

The DAX averages the stock prices of the 30 largest publicly traded companies in Germany. Formally, its value $I_t$ at time point $t$ equals

$$I_t = \frac{\sum_i p_{i,t} \cdot q_{i,t^*} \cdot c_{i,t}}{\sum_i p_{i,t_0} \cdot q_{i,t_0} \cdot K_{t^*}} \cdot 1000,$$

where $i = 1, \ldots, 30$ denotes the included companies, $t_0$ is the basis date (30 December 1987), $t^*$ denotes the time point of the last adjustment, $p_{i,t}$ is the stock price and $q_{i,t}$ denotes the capital of company $i$ at time point $t$, respectively, and $c_{i,t}$ and $K_{t^*}$ are adjustment factors, see Janßen and Rudolph (1992).

Instead of modelling the process $(I_t)_t$ directly, we consider the time series

$$X_t = \log \left[ \frac{I_t}{I_{t-1}} \right], \quad t \geq 2,$$
which we refer to as the logarithm of the daily returns (log-returns). This transformation yields two important benefits from a modelling point of view: First, conditional independence becomes a reasonable assumption, which is required to preserve the first-order Markov property of the hidden state process. Second, financial theory suggests a distribution for log-returns, namely the \( t \)-distribution, see Platen et al. (2008).

For our first case study, the fine-scale observation process is identified by the daily log-returns corresponding to the years 2000 to 2019. For the coarse-scale observations, we considered the average log-returns over 30 trade days, which appears to be a reasonable time span at which short-term trends can manifest themselves. While this choice is somewhat arbitrary, we found that choosing different fine-scale time horizons between 20 and 40 days had no major impact on the estimation results. For model selection, we considered the BIC, which favoured 3 coarse-scale states and 2 fine-scale states.

The t.p.m. associated with the coarse-scale state process was estimated as

\[
\hat{\Gamma} = \begin{pmatrix}
0.936 & 0.053 & 0.011 \\
0.073 & 0.828 & 0.099 \\
0.000 & 0.319 & 0.681
\end{pmatrix},
\]

which implies the stationary distribution \((0.458, 0.402, 0.140)\). The stationary state probabilities can be regarded as the long-term proportion of time that the coarse-scale state process spends in the different states. The t.p.m.s associated with the fine-scale state process were estimated as

\[
\hat{\Gamma}^{*1} = \begin{pmatrix}
0.924 & 0.076 \\
0.070 & 0.930
\end{pmatrix}, \quad \hat{\Gamma}^{*2} = \begin{pmatrix}
0.923 & 0.077 \\
0.077 & 0.923
\end{pmatrix}, \quad \hat{\Gamma}^{*3} = \begin{pmatrix}
0.932 & 0.068 \\
0.052 & 0.948
\end{pmatrix},
\]

implying the stationary distributions \((0.479, 0.521)\), \((0.500, 0.500)\) and \((0.433, 0.567)\), respectively.

The estimated state-dependent scaled \( t \) distributions of fine-scale log-returns are visualized in Figure 2(a). Coarse-scale state 1 (bullish market) corresponds to the lowest marginal volatility \((7.5 \cdot 10^{-3})\) and highest marginal expected return \((5.9 \cdot 10^{-4})\), while coarse-scale state 3 (bearish market) corresponds to the highest marginal volatility \((28.1 \cdot 10^{-3})\) and lowest marginal expected return \((-20.6 \cdot 10^{-4})\). Coarse-scale state 2 can be interpreted as a stable state, which is characterized by a moderate volatility. According to the stationary distribution under the fitted model, the market was in a bearish state in about 14.0\% of the time, whereas the bullish market was active in about 45.8\% of the time.

Figure 2(c) displays the Viterbi-decoded time series, with the decoding performed using the Viterbi algorithm as described in Section 2.3. In the autumn of 2008, the DAX is marked by the global financial crisis, which led to the bankruptcy of the US investment bank Lehman Brothers on 15 September 2008. Noteworthy, the model detects fine-scale state 2 within the bearish market (which represents most lossy periods) and switches to calmer fine-scale states as soon as the log-returns become...
more moderate. In 2017, we observe a high proportion of light green periods, in which the DAX gained nearly 4000 points within just a few months. In 2018, this skyrocketing was stopped. However, as the volatility remained low, the model retained coarse-scale state 1.

Figure 2(b) shows that the pseudo-residuals can be considered as independently normal distributed, indicating a reasonable model fit. In case of severe deviations from a normal distribution, more flexible state-dependent distributions such as the normal inverse Gaussian distribution (De Angelis and Viroli, 2017) can be
considered. When experiencing severe residual correlation, autoregressive terms can be incorporated into the observation process. In that regard, testing for differences in the corresponding coefficients among the coarse-scale states can provide a starting point for investigating the LeBaron effect (LeBaron, 1992).

### 3.2 Deutsche Bank

In our second case study, we demonstrate how HHMMs can be used to detect bearish and bullish markets by jointly modelling indices and individual stocks, where the fine-scale observation process is identified by the daily log-returns of the Deutsche Bank stock between 2000 and March 2021. Instead of considering the monthly averages of the daily log-returns for the coarse-scale observation process (as it was done in Section 3.1), we considered the monthly returns of the DAX, and present the model with 2 coarse-scale states and 2 fine-scale states.

The t.p.m. associated with the coarse-scale state process was estimated as

\[
\hat{\pi} = \begin{pmatrix}
0.909 & 0.091 \\
0.101 & 0.899
\end{pmatrix},
\]

which implies the stationary distribution (0.527, 0.473). The t.p.m.s associated with the fine-scale state process were estimated as

\[
\hat{\pi}^{(1)} = \begin{pmatrix}
0.978 & 0.022 \\
0.011 & 0.989
\end{pmatrix}, \quad \hat{\pi}^{(2)} = \begin{pmatrix}
>0.999 & <0.001 \\
<0.001 & >0.999
\end{pmatrix},
\]

implying the stationary distributions (0.323, 0.677) and (0.852, 0.148), respectively.

Figure 3 displays the Viterbi-decoded time series. In September 2001, after the 9/11 terrorist attacks, the Deutsche Bank stock faces a short period of substantial losses. In the autumn of 2008, the stock is marked by the global financial crisis in general and the bankruptcy of Lehman Brothers on 15 September in particular. Between 2010 and 2020, the volatility remains moderate, except for two short transitions to coarse-scale state 2. On 27 January the first confirmed COVID-19 case in Germany led to another switch to coarse-scale state 2, which is characterized by a high volatility especially in the coarse-scale observations and lasts until the end of the time series.

The estimated parameters, as well as 95% confidence intervals (CIs), which were computed based on 999 bootstrap samples, are reported in Table 1. From the estimated means of the coarse-scale state-dependent distributions, \(\hat{\mu}^{(1)} = 0.0073\) and \(\hat{\mu}^{(2)} = -0.0016\), we can conclude that coarse-scale state 1 can be linked to bullish markets, whereas coarse-scale state 2 can be interpreted as a bearish market. Noteworthy, the state-dependent distribution for fine-scale state 2 within coarse-scale state 1 has a negative estimated mean \(\hat{\mu}^{(1,2)} = -0.0008\), which we interpret as short-term corrections within bullish periods (this is also supported by the Viterbi-decoded time series visualized in Figure 3). Furthermore, since the CIs for \(\sigma^{(1)}\) do not overlap the CIs for \(\sigma^{(2)}\), and the CIs for \(\sigma^{(1,1)}\) and \(\sigma^{(1,2)}\) do not overlap
Figure 3 Visualization of the HHMM results for DAX and Deutsche Bank

Table 1 Estimated parameters of the state-dependent distributions and corresponding 95% CIs

|                  | Parameter       | Estimate  | 95% CI            |                  |
|------------------|-----------------|-----------|-------------------|------------------|
| Coarse scale     | $\mu^{(1)}$     | 0.0073    | [-0.0002, 0.0141] |                  |
|                  | $\mu^{(2)}$     | -0.0016   | [-0.0126, 0.0111] |                  |
|                  | $\sigma^{(1)}$  | 0.0371    | [0.0314, 0.0414]  |                  |
|                  | $\sigma^{(2)}$  | 0.0697    | [0.0562, 0.0809]  |                  |
|                  | $df^{(1)}$      | 7.023.00  | [7.96, $10^3$]    |                  |
|                  | $df^{(2)}$      | 12.59     | [4.55, $10^3$]    |                  |
| Fine scale in coarse-scale state 1 | $\mu^{(1,1)}$ | 0.0010    | [0.0001, 0.0021]  |                  |
|                  | $\mu^{(1,2)}$   | -0.0008   | [-0.0019, 0.0001] |                  |
|                  | $\sigma^{(1,1)}$| 0.0093    | [0.0081, 0.0106]  |                  |
|                  | $\sigma^{(1,2)}$| 0.0158    | [0.0145, 0.0174]  |                  |
|                  | $df^{(1,1)}$    | 6.80      | [4.37, 76.87]     |                  |
|                  | $df^{(1,2)}$    | 10.82     | [6.96, 35.94]     |                  |
| Fine scale in coarse-scale state 2 | $\mu^{(2,1)}$ | -0.0001   | [-0.0001, 0.0007] |                  |
|                  | $\mu^{(2,2)}$   | -0.0035   | [-0.0092, 0.0032] |                  |
|                  | $\sigma^{(2,1)}$| 0.0231    | [0.0218, 0.0244]  |                  |
|                  | $\sigma^{(2,2)}$| 0.0537    | [0.0471, 0.0612]  |                  |
|                  | $df^{(2,1)}$    | 7.50      | [5.76, 11.27]     |                  |
|                  | $df^{(2,2)}$    | 9.03      | [4.85, $10^3$]    |                  |

the CIs for $\sigma^{(2,1)}$ and $\sigma^{(2,2)}$, we can conclude that both the DAX in general and
the Deutsche Bank stock in particular exhibit a significantly higher volatility in coarse-scale state 2 than in coarse-scale state 1.

### 3.3 Simulation study

To assess the properties of HHMMs, we conducted the following simulation study: In each of 100 simulation runs, we first simulated 256 realizations from a 2-state Markov chain (coarse-scale state process). For each of these realizations, we then simulated, on average, 22.2 realizations from another 2-state Markov chain (fine-scale state process). Conditional on the simulated coarse- and fine-scale states, the observations were then drawn from scaled $t$-distributions. Thus, we simulated 256 coarse-scale and, on average, 5,683.2 fine-scale observations in total (the sample sizes were chosen based on the sample sizes of the data modelled in Section 3.2). As true parameters, we considered the estimated parameters from the model presented in Section 3.2 (see Table 1).

To assess the differences between the true and the estimated parameters, we considered the sample mean absolute bias, which was computed as

$$\text{Mean bias} = \frac{1}{100} \sum_{i=1}^{100} |\hat{\theta}_i - \theta|,$$

where $\theta$ denotes the true parameter and $\hat{\theta}_i$ is the corresponding estimate obtained in the $i$th simulation run.

--

**Figure 4** Estimated state-dependent distributions for the coarse-scale observations (left panel) and the fine-scale observations within coarse-scale state 1 (middle panel) and 2 (right panel) obtained in 100 simulation runs. True state-dependent distributions are indicated by dashed lines.
The estimated state-dependent distributions obtained in 100 simulations runs are visualized in Figure 4. As the estimated distributions are very close to the true ones, there is no indication for substantial biases in the estimated parameters, indicating that the suggested approach is capable of adequately capturing the true state-dependent distributions.

The sample mean estimates and the sample mean absolute biases are reported in Table 2. The mean estimates are very close to the true values (see Table 1). Due to the much smaller sample size, the biases are, in general, larger for the estimated parameters of the coarse-scale state-dependent distributions than for the fine-scale state-dependent distributions. When the true degrees of freedom are relatively large, then the corresponding estimates are substantially biased, which is due to the fact that the degrees of freedom do not affect the shape of the state-dependent distributions anymore when \(>10^3\), thus leading to a high estimator variance.

### 3.4 Comparison to basic HMMs

As demonstrated in Sections 3.1, 3.2 and 3.3, HHMMs are, by definition, capable of jointly modelling short-term price fluctuations and long-term trends. In contrast, basic HMMs are limited to modelling state-switching dynamics at only one temporal resolution. This can lead to characteristics that are difficult to interpret, which is to be regarded as a major deficit of basic HMMs (see also Lihn, 2017 and Nguyen, 2018). To demonstrate this issue, we compared the HHMM presented in Section 3.2 to a basic HMM. As the model presented in Section 3.2 has effectively 4 states to model the fine-scale observation process, we considered a 4-state HMM for comparison.
Figure 5 visualizes the Viterbi-decoded time series of daily closing prices of the Deutsche Bank stock under the fitted 4-state HMM.

Figure 5 visualizes the Viterbi-decoded time series of daily closing prices. Compared to the Viterbi-decoded time series under the fitted HHMM displayed in Figure 3, the fitted HMM is characterized by more state switches. Noteworthy, the expected state dwell times under the fitted HMM for states 2, 3 and 4 lie between 1.6 and 4.7 days, which also indicates a high state transition frequency. Many of these state transitions clearly relate to short-term price fluctuations instead of actual regime switches and, as a consequence, overfit market behaviour, which, as demonstrated in Section 3.2, can be overcome using HHMMs.

4 Discussion

In this article, we proposed HHMMs as a versatile extension of basic HMMs for detecting bearish and bullish markets in financial time series. By adding a hierarchical structure to the basic HMM, we paved the way for jointly inferring long-term trends and short-term dynamics from multi-scale time series data. In two real-data applications, we demonstrated that HHMMs can help to improve our understanding of how trends alternate, which in turn can help to make profitable investment decisions. In this last section, we discuss some limitations of the suggested approach and outline possible directions for future research.

First, the dependence assumptions of the proposed model are to be questioned. Recall that we assumed conditional independence across observations and that the fine-scale HMMs are independent of the coarse-scale observations. While the former assumption is usually justified in the specific case of log-returns, the latter one can be regarded as more critical, as we used averages over the fine-scale observations (Section 3.1) and returns of a stock index that contains the stock modelled at the fine scale (Section 3.2) as coarse-scale observations. However, although inducing some dependence, we argue that the coarse-scale observation indicate changes in the long-term trend, which only depend on the current coarse-scale state and not on
short-term price fluctuations. Furthermore, the dependence structure can be extended by incorporating semi-Markov chains into the state process, see Bulla and Bulla (2006) and Maruotti (2019) for details.

Second, one can put into question whether market behaviour can be classified into a finite number of states. For animal movement data, for example discrete states explaining resting, foraging, or travelling behaviour seem natural (Leos-Barajas et al., 2017). While the prospect of having similar proxies for financial data certainly is desirable, allowing for gradual changes may bring us closer to reality. In that regard, having discrete states on the coarse scale and a state-space model with an infinite number of states on the fine scale would be interesting to explore. Since we should expect the interpretation of the states to become more complicated, such an extension can help us to overcome the deficit that discrete states are often prone to over-interpretation.

While we focused on understanding past market behaviour, it would be interesting to investigate the performance of trading strategies that are based on our model. A naive strategy could be to buy stocks whenever a switch to a bullish market is predicted, and repel them whenever a switch to a bearish market is predicted. Similar strategies for basic HMMs can be found in Lihn (2017) and Nguyen (2018), and it would be interesting to compare the performance of HHMMs to these approaches.

On a final note, we would like to highlight that, analogous to handwriting recognition, any desired number of hierarchies would be theoretically feasible. However, HHMMs with more than two hierarchies would certainly not be as straightforward to implement and to handle, and the problems outlined in Section 2.2 would likely be exacerbated due to the much larger number of parameters. Nonetheless, investigating whether additional hierarchies, capturing, for example, medium-term or intra-day patterns, can help to draw a more complete picture of market dynamics seems intriguing and is therefore to be regarded as a promising avenue for future research.

Supplementary materials

We implemented the proposed methodology in the R package fHMM, which is available from https://CRAN.R-project.org/package=fHMM.

Acknowledgements

The authors wish to thank Roland Langrock for valuable discussions and helpful advice that considerably improved the article.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.
Funding

The authors received no financial support for the research, authorship and/or publication of this article.

**APPENDIX**

Algorithm 1 Computing the log-likelihood of a fine-scale HMM

```
1: procedure log $\mathcal{L}_{\text{HMM}}(\theta^{(i)} | (X_{t}^{*})_{t=1}^{T})$
2:     for $k = 1, \ldots, N^*$ do
3:         $\phi_{k,1}^{(i)} = \log[\delta_k] + \log[f^{(i)}(X_{1}^{*})]$
4:     end for
5:     for $t^* = 2, \ldots, T^*$ do
6:         $c_{t^*-1} = \max[\phi_{1,t^*-1}^{(i)}, \ldots, \phi_{N^*,t^*-1}^{(i)}]$
7:     end for
8:     return $\log[\sum_{i=1}^{N^*} \exp[\phi_{i,t^*}^{(i)} - c_{t^*-1}]] + c_{t^*}$
9: end procedure
```

Algorithm 2 Computing the log-likelihood of a HHMM

```
1: $\theta = (\delta, \Gamma, (f^{(i)}))$
2:     for $i = 1, \ldots, N$ do
3:         $\theta^{(i)} = (\delta^{(i)}, \Gamma^{(i)}, (f^{(i,k)}))$
4:     end for
5: procedure log $\mathcal{L}_{\text{HHMM}}(\theta, (\theta^{(i)}), (X_{t})_{t=1}^{T})$
6:     for $i = 1, \ldots, N$ do
7:         $\phi_{i,1}^{(i)} = \log[\delta_i] + \log \mathcal{L}_{\text{HMM}}(\theta^{(i)} | (X_{1})_{t=1}^{T}) + \log[f^{(i)}(X_{1})]$
8:     end for
9:     for $t = 2, \ldots, T$ do
10:        $c_{t-1} = \max[\phi_{1,t-1}, \ldots, \phi_{N,t-1}]$
11:    end for
12:    $c_T = \max[\phi_{1,T}, \ldots, \phi_{N,T}]$
13:    return $\log[\sum_{i=1}^{N} \exp[\phi_{i,T} - c_T]] + c_T$
14: end procedure
```
1: $\theta = (\delta, \Gamma, (f^0_i),)$
2: procedure Viterbi($\theta, (X_t)_t$)
3: \hspace{1em} for $i = 1, \ldots, N$ do
4: \hspace{2em} $\kappa_{i,1} = \log[\delta_i] + \log[f^0_1(X_1)]$
5: \hspace{2em} end for
6: \hspace{1em} for $t = 2, \ldots, T$ do
7: \hspace{2em} for $i = 1, \ldots, N$ do
8: \hspace{3em} $\kappa_{i,t} = \max_j (\kappa_{j,t-1} + \log[\gamma_{ji}]) + \log[f^0_t(X_t)]$
9: \hspace{2em} end for
10: \hspace{1em} end for
11: $\hat{S}_T = \arg\max_i \kappa_{i,T}$
12: \hspace{1em} for $t = T - 1, \ldots, 1$ do
13: \hspace{2em} $\hat{S}_t = \arg\max_i \kappa_{i,t} \gamma_{\hat{S}_{t+1}}$
14: \hspace{1em} end for
15: return $(\hat{S}_i)_t$
16: end procedure

**Algorithm 3** Decoding the hidden states

**References**

Adam T, Grith CA, Leos-Barajas V, Meese EN, Lowe CG, Blackwell PG, Righton D and Langrock R (2019) Joint modelling of multi-scale animal movement data using hierarchical hidden Markov models. *Methods in Ecology and Evolution*, 10, 1536–1550.

Alexandrovich G, Holzmann H and Leister L (2016): Nonparametric identification and maximum likelihood estimation for hidden Markov models. *Biometrika*, 103, 423–434.

Ang A and Bekaert G (2002) International asset allocation with regime shifts. *The Review of Financial Studies*, 15, 1137–1187.

Ang A and Timmermann A (2012) Regime changes and financial markets. *Annual Review of Financial Economics*, 4, 313–337.

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

Bulla J, Mergner S, Bulla I, Sesboüé A and Chesneau C (2011) Markov switching asset allocation: Do profitable strategies exist? *Journal of Asset Management*, 12, 310–321.

Cohen L, Dietker K and Malloy C (2013) Legislating stock prices. *Journal of Financial Economics*, 110, 574–595.

De Angelis L and Paas LJ (2013) A dynamic analysis of stock markets using a hidden Markov model. *Journal of Applied Statistics*, 40, 1682–1700.

De Angelis L and Viroli C (2017) A Markov-switching regression model with non-Gaussian innovations: Estimation and testing. *Studies in Nonlinear Dynamics and Econometrics*, 21, 1–22.

Dias JG, Vermunt JK and Ramos S (2009) Mixture hidden Markov models in finance research. In *Advances in Data Analysis, Data Handling and Business Intelligence*, pages 451–459. Berlin, Heidelberg: Springer.

Fine S, Singer Y and Tishby N (1998) The hierarchical hidden Markov model: Analysis and applications. *Machine Learning*, 32, 41–62.
Guerin P and Marcellino M (2013) Markov-switching MIDAS models. *Journal of Business and Economic Statistics*, 31, 45–56.

Hassan MR and Nath B (2005) Stock market forecasting using hidden Markov model: A new approach. *5th International Conference on Intelligent Systems Design and Applications*, 192–196.

Humpe A and Macmillan P (2009) Can macroeconomic variables explain longterm stock market movements? A comparison of the US and Japan. *Applied Financial Economics*, 19, 111–119.

Janßen B and Rudolph B (1992) *Der Deutsche Aktienindex DAX [The Deutscher Aktienindex DAX]*. Frankfurt: Knapp Verlag.

LeBaron B (1992) Some relations between volatility and serial correlations in stock market returns. *Journal of Business*, 65, 199–219.

Leos-Barajas V (2019) Incorporating multi-scale structures and physiological processes into the modelling of animal movement, Dissertation. URL https://lib.dr.iastate.edu/etd/17497/

Leos-Barajas V, Ganglo EJ, Adam T, Langrock R, van Beest FM, NabeNielsen J and Morales JM (2017) Multi-scale modelling of animal movement and general behavior data using hidden Markov models with hierarchical structures. *Journal of Agricultural, Biological and Environmental Statistics*, 22, 232–248.

Lihn SH (2017) Hidden Markov model for financial time series and its application to S&P 500 index. *Quantitative Finance*. URL https://mran.microsoft.com/snapshot/2018-03-03/web/packages/ldhmm/vignettes/ldhmm-spx.pdf

Maruotti A, Punzo A and Bagnato L (2019) Hidden Markov and semi-Markov models with multivariate leptokurtic-normal components for robust modelling of daily returns series. *Journal of Financial Econometrics*, 17, 91–117.

Nguyen N (2018) Hidden Markov model for stock trading. *International Journal of Financial Studies*, 6, 36.

Nystrup P, Hansen BW, Madsen H and Lindström E (2015) Regime-based versus static asset allocation: Letting the data speak. *The Journal of Portfolio Management*, 42, 103–109.

Nystrup P, Lindström E and Madsen H (2020) Learning hidden Markov models with persistent states by penalizing jumps. *Expert Systems with Applications*, 150, 113307.

Nystrup P, Madsen H and Lindström E (2017) Long memory of financial time series and hidden Markov models with time-varying parameters. *Journal of Forecasting*, 36, 989–1002.

Nystrup P, Madsen H and Lindström E (2018) Dynamic portfolio optimization across hidden market regimes. *Quantitative Finance*, 18, 83–95.

Oelschlager L and Adam T (2021) *fHMM: Fitting hidden Markov models to financial data*. R package, 0.3.0. URL https://CRAN.R-project.org/package=fHMM

Platen E and Rendek R (2008) Empirical evidence on Student-t log-returns of diversified world stock indices. *Journal of Statistical Theory and Practice*, 2, 233–251.

Ryden T, Teräsvirta T and Asbrink S (1998) Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics*, 13, 217–244.

Schaller H and van Norden SV (1997) Regime switching in stock market returns. *Applied Financial Economics*, 7, 177–191.

Zucchini W, MacDonald IL and Langrock R (2016) *Hidden Markov models for time series: An introduction using R, 2nd edition*. Boca Raton, FL: Chapman and Hall/CRC.