

Detection of Bearish and Bullish Markets in the DAX Using Hierarchical Hidden Markov Models

Faculty of Business Administration and Economics
Bielefeld University

Thesis

in Partial Fulfilment of the Requirements for the Academic Degree
Master of Science

Submitted by

Lennart Oelschläger

Born on November 16, 1994 in Herford

Matriculation Number: 2572042

Bielefeld, July 2019

Supervisor, 1. Reviewer: Prof. Dr. Roland Langrock
2. Reviewer: Dr. Tetyana Pasurek

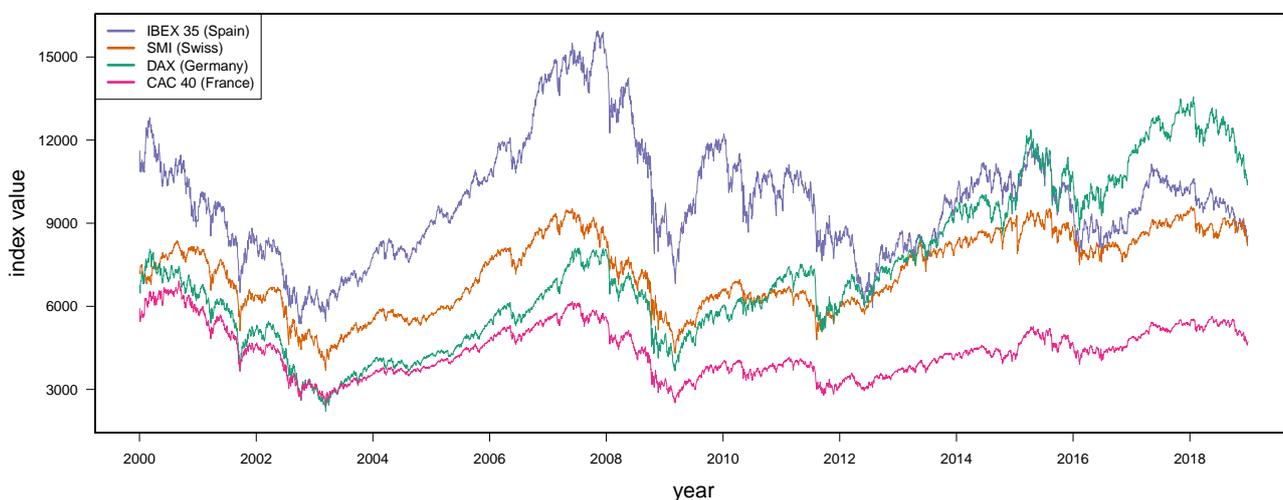
Confirmation

I, Lennart Oelschläger, hereby confirm that this master thesis is entirely my own work and that I have not used any additional assistance or resources other than indicated. All quotations, ideas and data sets that have been taken from other sources and other persons' work have been cited appropriately and provided with the corresponding references. Furthermore, I confirm that I have read the leaflet on dealing with plagiarism.

Date: _____ Signature: _____

Abstract

Almost any financial market exhibits alternating periods of rising and falling prices, as shown in the graphic below. Stock traders have to mind those trends to make profitable investment decisions. So far and among other models, the hidden Markov model has been used to quantify the trend changes in order to predict upcoming market behaviour. However, this model in its basic form is not capable of capturing short-term and long-term trends jointly. Extending the model by a hierarchical structure, this thesis fixes this deficit, aiming to draw a more comprehensive picture of the financial market. The improvement is exemplary investigated for the Deutscher Aktienindex. Modelling financial data with the hierarchical hidden Markov model can lead to more sophisticated trading strategies.



Contents

Confirmation	2
Abstract	3
1 Introduction	6
2 The DAX Data	8
3 Methodology	10
3.1 The Hidden Markov Model	10
3.2 Adding a Hierarchical Structure	11
3.3 The Likelihood Function	13
3.4 Challenges in the Maximization of the Likelihood	15
3.5 The Viterbi Algorithm for State Decoding	18
3.6 Pseudo-Residuals	19
4 Data Application	20
4.1 Model Selection	20
4.2 Model Checking	22
4.3 The Model Results	24
4.4 Impact of the Fine-Scale Time Horizon	27
4.5 Detection of Bearish and Bullish Markets	28
4.6 Accuracy of the State Decoding	30
4.7 Comparison to HMM Results	31
4.8 Thoughts on a HHMM-Based Trading Strategy	32
5 Discussion	34
A Proofs	37
B Parameter Estimates	41
Bibliography	43

List of Figures

1	Visualization of the DAX data	9
2	Dependence structure of the HHMM	13
3	Pseudo-residuals of the HHMM with normal distributions	22
4	Pseudo-residuals of the HHMM with t-distributions	24
5	Estimated state-dependent fine-scale distributions	25
6	Decoded states and a close-up of the financial crisis 2008	29
7	Bootstrapping of the Viterbi algorithm	30
8	Distributions, pseudo-residuals and decoded states of the HMM	31

List of Tables

1	Ingredients of the HMM	11
2	Ingredients of the HHMM	14
3	AIC and BIC for candidate HHMMs with normal distributions	21
4	AIC and BIC for candidate HHMMs with t-distributions	23
5	Estimates for $T^* = 25$ and $T^* = 35$	27
6	AIC and BIC for candidate HMMs	31

List of Algorithms

1	Compute the log-likelihood of a fine-scale HMM	16
2	Compute the log-likelihood of the HHMM	17
3	Decode the hidden states (Viterbi algorithm, logarithmic version)	19
4	Classify the upcoming log-returns	33

List of Abbreviations

DAX	Deutscher Aktienindex	log-	(natural) logarithm of
HMM	hidden Markov model	AIC	Akaike information criterion
HHMM	hierarchical hidden Markov model	BIC	Bayesian information criterion

1 Introduction

The formula for earning money at the stock market is simple: Buy and sell stocks at the proper moment. Generally speaking, stock traders want to invest at the beginning of an upwards trend (bullish market) and repel the shares just in time before the price drops (bearish market). As the stock price depends on a variety of environmental influences, luck is certainly involved in hitting those exact moments. However, an investigation of the market behaviour yields a better understanding how trends alternate and thereby increases the chance of making profitable investment decisions. Various models of market volatility have been considered, see for example [Pagan and Schwert, 1990]. This thesis aims to improve those investigations by applying the hierarchical extension of the hidden Markov model. The model will be used to detect past trends in the most important German stock index, the Deutscher Aktienindex (DAX), and to reveal their characteristics.

The hidden Markov model has been used several times to model financial data. In [Lihn, 2017], it is applied to the American S&P 500 index. The model states are identified as different levels of market volatility, aiming to verify the empirical law that return is negatively correlated to volatility. Another application to the same index can be found in [Nguyen, 2018], where four states are used to predict monthly closing prices. Based on the model results, a trading strategy is derived which outperforms the conventional buy-and-hold strategy. [Hassan and Nath, 2005] presents another field of application, applying the model to predict prices in the setting of airline stocks. All three papers conclude that the hidden Markov model is well suited to explain or forecast stock market data.

In its basic form however, the hidden Markov model operates on a single time scale and therefore it is not capable of capturing short-term and long-term trends jointly. This is a major deficit because the model can easily misinterpret short-term price fluctuations as a change in the long-term trend and hence might draw a distorted picture of the market behaviour. The main objective of this thesis lies in fixing this deficit by adding a hierarchical structure to the hidden Markov model and thereby improving the distinctness of short-term and long-term trends.

The hierarchical generalization of the hidden Markov model originates in the field of machine learning. In [Fine et al., 1998], the hierarchical hidden Markov model is used for handwriting recognition where hierarchical levels appear as letters, syllables and words. Recently, the model turned out to be beneficial for modelling animal movement data, cf. [Leos-Barajas et al., 2017] where the hierarchical structure allows for behavioural inference at two different time scales.

A slightly extended approach is done in [Adam et al., 2019], where a second observation layer is incorporated to model animal movement data on two different temporal resolutions. Both papers conclude that the hierarchical hidden Markov model improves the inference of animal movement behaviour in comparison to the hidden Markov model.

This work transfers the approach of [Adam et al., 2019] from animal movement inference to modelling financial data. Exemplary, the hierarchical hidden Markov model will be applied to closing prices of the DAX. We will investigate that the hierarchical model structure allows for a more comprehensive picture of stock market behaviour by minding the difference between short-term and long-term trends. The improvement can be used to develop more sophisticated trading strategies. A brief outlook is provided.

The thesis is structured as follows: Chapter 2 introduces the DAX data we investigate. Chapter 3 presents the methodology that will be applied to the data in Chapter 4. In Chapter 5, the findings will be discussed and possible future research will be addressed. Proofs and estimated parameters of candidate models are appended.

Concluding the introduction, I would like to express my gratitude towards Professor Roland Langrock for sharing his enthusiasm for hidden Markov models; towards Dr. Tetyana Pasurek for her offer to review my thesis and towards my parents Dagmar and Bernd Oelschläger for their support.

2 The DAX Data

For a start, this chapter briefly presents the background of the Deutscher Aktienindex and its data. The following information are based on [Janßen and Rudolph, 1992].

The DAX in its current form exists since July 1, 1988 and established itself as the most important German stock index. The index is computed as a weighted average over the share prices of the 30 largest publicly traded companies in Germany. Thereby, intra-company price changes are smoothed and the German market behaviour is well summarized. Formally, its value I_t at time point t is

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

with $i = 1, \dots, 30$ the included companies, t_0 the basis date (December 30, 1987), t^* the time point of the last adjustment, $p_{i,t}$ the share price and $q_{i,t}$ the capital of company i at time point t , respectively, and $c_{i,t}$ and K_{t^*} adjustment factors.

Nowadays, the index is updated every second during a trading day. Hence, a large amount of data on a very fine temporal resolution is available. This immediately poses the question of numerical processing effort, especially in our case: Beside short-term trends, we aim to capture long-term trends, which potentially persist over a time span of several years. To limit the processing effort, we only focus on the last update of each trading day (the daily closing prices¹). Furthermore, mind that the DAX is not updated at weekends or public holidays because shares are not traded on these days. Assuming that the market behaviour is not influenced at these days, we treat them as if they do not exist. Overall, this results in an easy-to-handle amount of 4822 daily closing prices from years 2000 to 2018, which is the time span investigated in this thesis.²

Instead of the daily closing prices $(I_t)_t$, we model the time series $(X_t)_t$ with

$$X_t = \log \left[\frac{I_t}{I_{t-1}} \right], \quad t \geq 2,$$

which we refer to as the logarithm of the returns (log-returns).³ This transformation reveals particular benefits from a modelling point of view, cf. Section 4.1. Obviously, results derived on $(X_t)_t$ can be back-transformed to the more natural representations $(\exp[X_t] - 1)_t$ or $(I_t)_t$.⁴

1. The DAX itself is not a share. To simplify matters, we use the expressions *price* and *return* anyway.

2. The data was received from www.finance.yahoo.com on March 21, 2019.

3. Throughout this thesis, \log refers to the natural logarithm.

4. The term $\exp[X_t] - 1 = [I_t - I_{t-1}]/I_{t-1}$ equals the relative return from time point $t - 1$ to t .

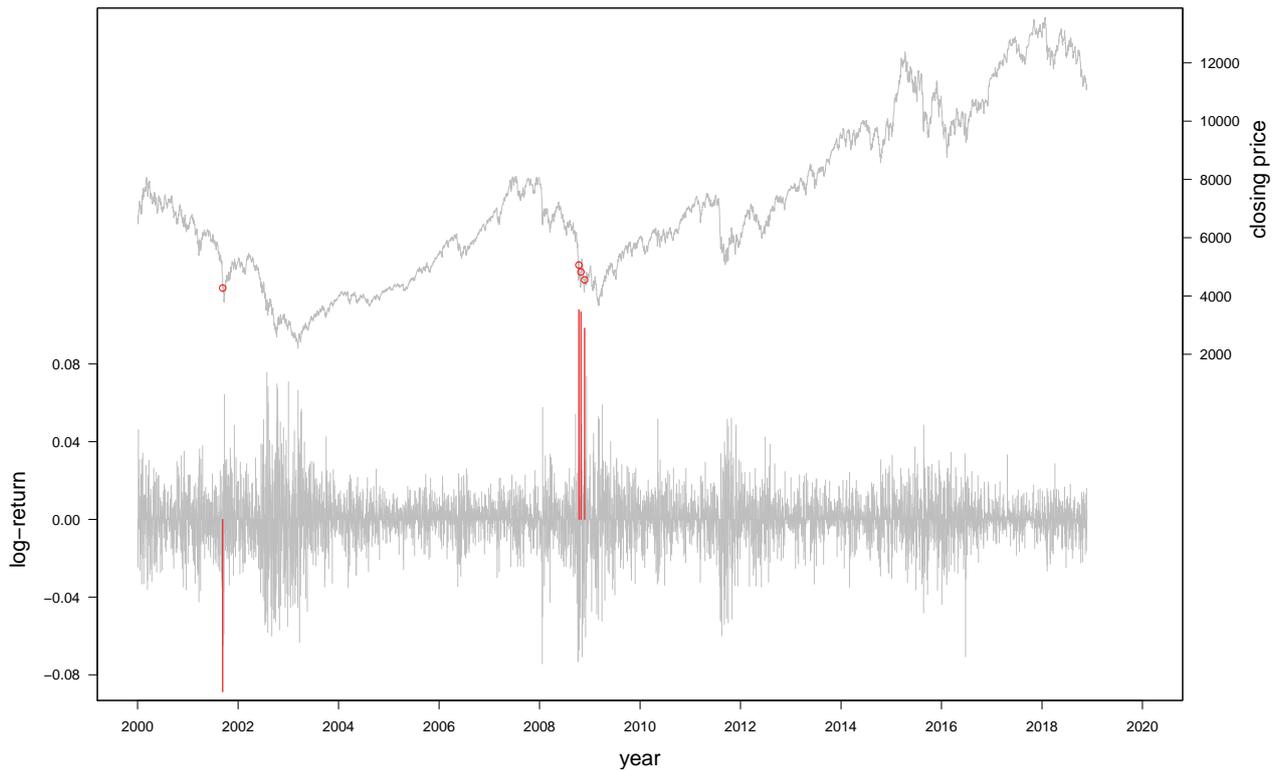


Figure 1: Visualization of the DAX data

Figure 1 is a visualization of $(I_t)_t$ on top and $(X_t)_t$ below. There, four outliers on September 11, 2001 and on three dates in October 2008 can be observed, marked in red. The former one corresponds to a reaction of the German stock market to the 9/11 terrorist attack in America, the latter ones can be interpreted as reactions to the German „Rettungsschirm“ (bailout) after the financial crisis. I decided to remove these data points from the investigation because they do not represent normal market behaviour. Summing up, we face the time series $(X_t)_t$ of log-returns with a total of 4817 observations.

3 Methodology

This chapter introduces the methodology of the thesis. Section 3.1 recaps the basic hidden Markov model, which is extended by a hierarchical structure in Section 3.2. Subsequently, Section 3.3 states a formula for the likelihood function of the hierarchical hidden Markov model. Section 3.4 discusses how to numerically maximize the likelihood function in order to get estimates for the model parameter. Section 3.5 presents the Viterbi algorithm, which enables us to decode the underlying states. Checking whether the model explains given data well can be done by computing pseudo residuals, which are introduced in Section 3.6. The concepts and notations of this chapter are based on [Zucchini et al., 2016] and [Adam et al., 2019].

3.1 The Hidden Markov Model

The hidden Markov model (HMM) is a statistical model for time series data. The model predicates the idea that the behaviour of the nature can be divided into a finite number of different states. It is not possible to directly determine which state is when active. 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 hidden state. This concept can be formulated more formally introducing two stochastic processes:

1. At each time point t of the discrete time space $\{1, \dots, T\}$, an underlying process $(S_t)_t$ selects one state from the state space $\{1, \dots, N\}$. We call $(S_t)_t$ the hidden state process.
2. Depending on which state is active at t , one of N distributions $f^{(1)}, \dots, f^{(N)}$ realises the observation X_t . The process $(X_t)_t$ is called the observed state-dependent process.

We make some assumptions on these processes:

1. We assume that $(S_t)_t$ is a Markov process of first order (that means S_{t+1} only depends on S_t) and that $(S_t)_t$ is time-homogeneous. Thus, the process is identified by its initial distribution δ and its transition probability matrix Γ .
2. The process $(X_t)_t$ is said to satisfy the conditional-independence assumption, that means conditionally on the current state S_t , the observation X_t is independent of all other states and observations.

Furthermore, from a practical point of view, it is reasonable to identify the initial distribution of $(S_t)_t$ with its stationary distribution¹ π for two reasons: On the one hand, the hidden state process has been evolving for some time before our attention focused on it and thus can be assumed to be stationary already. On the other hand, π is determined by Γ , and setting $\delta = \pi$ reduces the amount of parameters that have to be estimated.

In case of financial data, the hidden states can be interpreted as different moods² of the market. These moods cannot be observed directly. However, price changes are observed which clearly depend on the current mood. Thereby, using an underlying Markov process, we can detect which mood is active at which point in time and how they alternate. Depending on the current mood, a price change is realized by a unique distribution. These distributions characterise the moods in terms of expected return and volatility.

The notations and properties of the HMM ingredients are summarized in the following table:

Ingredients	Notations	Properties
time space	$\{1, \dots, T\}$	finite
state space	$\{1, \dots, N\}$	finite
hidden state process	$(S_t)_t$	first-order Markov process
initial probabilities	$\delta_i = \Pr(S_1 = i)$	$\delta_i \in [0, 1]$
initial distribution (row) vector	$\delta = (\delta_i)_i$	$\sum_i \delta_i = 1$
one-step transition probabilities	$\gamma_{ij} = \Pr(S_{t+1} = j \mid S_t = i)$	$\gamma_{ij} \in [0, 1]$, time-homogenous
transition probability matrix	$\Gamma = (\gamma_{ij})_{i,j}$	$\sum_j \gamma_{ij} = 1$
stationary distribution (row) vector	$\pi = (\pi_i)_i$	$\sum_i \pi_i = 1$, $\pi\Gamma = \pi$
state-dependent distributions	$f^{(1)}, \dots, f^{(N)}$	
observed state-dependent process	$(X_t)_t$	$X_t \sim f^{(i)}$ given $S_t = i$

Table 1: Ingredients of the HMM

3.2 Adding a Hierarchical Structure

The HMM can be extended by a hierarchical structure, which will be considered as hierarchical hidden Markov model (HHMM). In the following, a procedure is described how to incorporate a second hierarchy level. This procedure can recursively be applied several times in order to achieve any desired number of hierarchies. In this thesis, two hierarchies are considered.

-
1. If the Markov process is irreducible, it has a unique stationary distribution π , which solves $\pi\Gamma = \pi$. If additionally the Markov process is aperiodic, its state distribution converges to the stationary distribution, see [Norris, 1997]. Irreducibility and aperiodicity are usually satisfied assumptions in reality.
 2. The expression *mood* is used here as a synonym for the current market trend. If the trend is positive, the mood of the market is good, and vice versa.

It is assumed that we are dealing with two time series on two different time scales. For each observation of the time series on the coarser scale, we have several observations of the time series on the finer scale, for example monthly observations and corresponding daily observations. Following the concept of the HMM, 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 $\{1, \dots, T\}$, an underlying process $(S_t)_t$ selects one state from the coarse-scale state space $\{1, \dots, N\}$. We call $(S_t)_t$ the hidden coarse-scale state process.
2. Depending on which state is active at t , one of N distributions $f^{(1)}, \dots, f^{(N)}$ realises 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 in Section 3.1, 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 an amount of T^* fine-scale observations on every coarse-scale time point, we are facing T chunks of T^* fine-scale observations each. The hierarchical structure now evinces itself as we model each of the chunks again by one of N possible fine-scale HMMs. Each of the fine-scale HMMs has its own transition probability matrix $\Gamma^{*(i)}$, initial distribution $\delta^{*(i)}$, stationary distribution $\pi^{*(i)}$ and state-dependent distributions $f^{*(i,1)}, \dots, f^{*(i,N^*)}$. 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 $\{1, \dots, T^*\}$, the process $(S_{t,t^*}^*)_{t^*}$ selects one state from the fine-scale state space $\{1, \dots, N^*\}$. 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 $f^{*(i,1)}, \dots, f^{*(i,N^*)}$ realises the observation X_{t,t^*}^* . The process $(X_{t,t^*}^*)_{t^*}$ is called the observed fine-scale state-dependent process.

The state processes $(S_{1,t^*}^*)_{t^*}, \dots, (S_{T,t^*}^*)_{t^*}$ and the observation processes $(X_{1,t^*}^*)_{t^*}, \dots, (X_{T,t^*}^*)_{t^*}$ 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 .

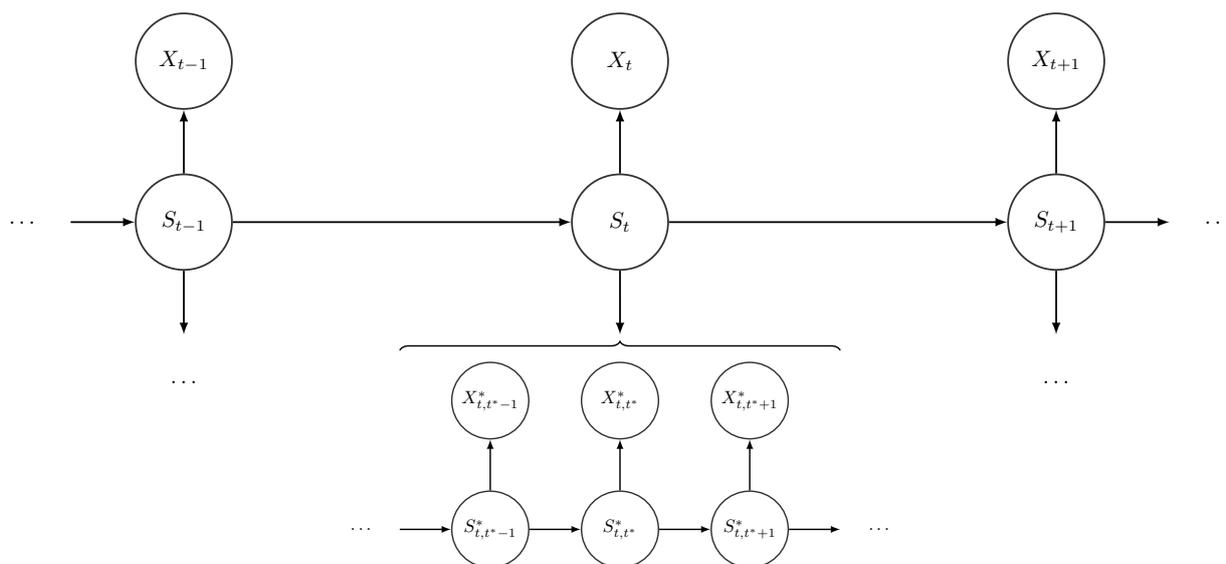


Figure 2: Dependence structure of the HHMM

The dependence structure of a HHMM with two hierarchies is visualised in Figure 2. The notations of the model ingredients are summarised in Table 2. As no additional properties emerged, the column with the properties is left out.

We now possess a model that allows for joint inference at two (or possibly more) temporal resolutions. In Chapter 4, we will identify the daily log-returns as the fine-scale resolution. For the coarse-scale resolution, we include the average over each chunk of fine-scale observations. The averaging smooths short-term fluctuations and thereby indicates long-term trends. Omitting these somewhat artificial coarse-scale observations is possible, leading to the framework of [Leos-Barajas et al., 2017]. In the case of financial data however, the truncated approach fails to conveniently capture short-term and long-term trends jointly without coarse-scale indicators. For that reason, the approach of [Adam et al., 2019] is the adequate choice. This discussion is resumed in Chapter 5.

3.3 The Likelihood Function

To fit a HHMM to the DAX, we require a procedure to estimate the model parameters for given observations. A natural approach is the maximum likelihood method, which suggests those parameters that yield the highest likelihood for observing the data. This section derives a formula for the likelihood function of a HHMM, which can be maximized numerically.

Conceptually, a HHMM can be treated as a HMM with two conditionally independent observations, the coarse-scale observation on the one hand and the corresponding chunk connected

Ingredients		Notations	
coarse-scale	time space	$\{1, \dots, T\}$	
	state space	$\{1, \dots, N\}$	
	hidden state process	$(S_t)_t$	
	initial probabilities	$\delta_i = \Pr(S_1 = i)$	
	initial distribution vector	$\delta = (\delta_i)_i$	
	one-step transition probabilities	$\gamma_{ij} = \Pr(S_{t+1} = j \mid S_t = i)$	
	transition probability matrix	$\Gamma = (\gamma_{ij})_{i,j}$	
	stationary distribution vector	$\pi = (\pi_i)_i$	
	state-dependent distributions	$f^{(1)}, \dots, f^{(N)}$	
	observed state-dependent process	$(X_t)_t$	
fine-scale	time space	$\{1, \dots, T^*\}$	
	state space	$\{1, \dots, N^*\}$	
	hidden state processes	$(S_{t,t^*}^*)_{t^*}$	$t = 1, \dots, T$
	initial probabilities	$\delta_k^{*(i)} = \Pr(S_1^{*(i)} = k)$	$i = 1, \dots, N$
	initial distribution vectors	$\delta^{*(i)} = (\delta_k^{*(i)})_k$	
	one-step transition probabilities	$\gamma_{jk}^{*(i)} = \Pr(S_{t,t^*+1}^{*(i)} = k \mid S_{t,t^*}^{*(i)} = j)$	
	transition probability matrices	$\Gamma^{*(i)} = (\gamma_{jk}^{*(i)})_{j,k}$	
	stationary distribution vectors	$\pi^{*(i)} = (\pi_k^{*(i)})_k$	
	state-dependent distributions	$f^{*(i,1)}, \dots, f^{*(i,N^*)}$	
	observed state-dependent processes	$(X_{t,t^*}^*)_{t^*}$	$t = 1, \dots, T$

Table 2: Ingredients of the HHMM

to a fine-scale HMM on the other hand. For a start, we derive the likelihood formula for the fine-scale HMMs. It is assumed that we want to fit the i -th fine-scale HMM with the model parameters $\theta^{*(i)} = (\delta^{*(i)}, \Gamma^{*(i)}, (f^{*(i,k)})_k)$ to the t -th chunk $(X_{t,t^*})_{t^*}$ of fine-scale observations. A formula for the likelihood function can be derived by applying the law of total probability:

$$\mathcal{L}^{HMM}(\theta^{*(i)} \mid (X_{t,t^*}^*)_{t^*}) = \sum_{S_{t,1}^*, \dots, S_{t,T^*}^*=1}^{N^*} \left(\prod_{t^*=1}^{T^*} f^{*(i, S_{t,t^*}^*)}(X_{t,t^*}^*) \right) \left(\delta_{S_{t,1}^*}^{*(i)} \prod_{t^*=2}^{T^*} \gamma_{S_{t,t^*-1}^* S_{t,t^*}^*}^{*(i)} \right). \quad (3.1)$$

As (3.1) is of exponential complexity, its computation is too expensive. However, a more efficient alternative exists. Consider the so-called fine-scale forward probabilities

$$\alpha_{k,t^*}^{*(i)} = f^{*(i)}(X_{t,1}^*, \dots, X_{t,t^*}^*, S_{t,t^*}^* = k),$$

where $t^* = 1, \dots, T^*$ and $k = 1, \dots, N^*$. Obviously,

$$\mathcal{L}^{HMM}(\theta^{*(i)} \mid (X_{t,t^*}^*)_{t^*}) = \sum_{k=1}^{N^*} \alpha_{k,T^*}^{*(i)}.$$

This representation improves formula (3.1) in terms of computation effort because the forward probabilities can be calculated in a recursive way of linear complexity:

$$\alpha_{k,1}^{*(i)} = \delta_k^{*(i)} f^{*(i,k)}(X_{t,1}^*), \quad (3.2)$$

$$\alpha_{k,t^*}^{*(i)} = f^{*(i,k)}(X_{t,t^*}^*) \sum_{j=1}^{N^*} \gamma_{jk}^{*(i)} \alpha_{j,t^*-1}^{*(i)}, \quad t^* = 2, \dots, T^*. \quad (3.3)$$

The transition from the likelihood function of a HMM to the likelihood function of a HHMM is straightforward: Consider the so-called coarse-scale forward probabilities

$$\alpha_{i,t} = f(X_1, \dots, X_t, (X_{1,t^*}^*)_{t^*}, \dots, (X_{t,t^*}^*)_{t^*}, S_t = i),$$

where $t = 1, \dots, T$ and $i = 1, \dots, N$. The likelihood function of the HHMM results from these variables as

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

The coarse-scale forward probabilities can again be calculated efficiently:

$$\alpha_{i,1} = \delta_i \mathcal{L}^{HMM}(\theta^{*(i)} | (X_{1,t^*}^*)_{t^*}) f^{(i)}(X_1), \quad (3.4)$$

$$\alpha_{i,t} = \mathcal{L}^{HMM}(\theta^{*(i)} | (X_{t,t^*}^*)_{t^*}) f^{(i)}(X_t) \sum_{j=1}^N \gamma_{ji} \alpha_{j,t-1}, \quad t = 2, \dots, T. \quad (3.5)$$

The proofs for the equations (3.1) to (3.5) are provided in Appendix A.

3.4 Challenges in the Maximization of the Likelihood

Maximization of the likelihood function is numerically feasible using the Newton-Raphson method. In practise, some technical issues arise which will be discussed in the following.

First, we address the problem of exceeding the computers' ability limit of processing very large numbers (numerical overflow) or numbers very close to zero (numerical underflow). We face this problem because the likelihood function contains products of probabilities³. Maximizing the logarithm of the likelihood and incorporating constants in a conducive way however solves the problem. Instead of computing the forward probabilities, we consider a logarithmic

3. In the case of modelling log-return data, numerical overflow is the primary problem. As we will investigate in Chapter 4, the distributions explaining the log-returns have a relatively small standard deviation. Thereby, the function values of the densities are large and products of them may exceed the computers' limit.

```

1: procedure  $\log \mathcal{L}^{HMM}(\theta^{*(i)} \mid (X_{t,t^*}^*)_{t^*})$ 
2:   for  $k = 1, \dots, N^*$  do
3:      $\phi_{k,1}^{*(i)} = \log[\delta_k^{*(i)}] + \log[f^{*(i,k)}(X_{t,1}^*)]$ 
4:   end for
5:   for  $t^* = 2, \dots, T^*$  do
6:      $c_{t^*-1} = \max\{\phi_{1,t^*-1}^{*(i)}, \dots, \phi_{N^*,t^*-1}^{*(i)}\}$ 
7:     for  $k = 1, \dots, N^*$  do
8:        $\phi_{k,t^*}^{*(i)} = \log[f^{*(i,k)}(X_{t,t^*}^*)] + \log\left[\sum_{j=1}^{N^*} \gamma_{jk}^{*(i)} \exp[\phi_{j,t^*-1}^{*(i)} - c_{t^*-1}]\right] + c_{t^*-1}$ 
9:     end for
10:  end for
11:   $c_{T^*} = \max\{\phi_{1,T^*}^{*(i)}, \dots, \phi_{N^*,T^*}^{*(i)}\}$ 
12:  return  $\log\left[\sum_{k=1}^{N^*} \exp[\phi_{k,T^*}^{*(i)} - c_{T^*}]\right] + c_{T^*}$ 
13: end procedure
    
```

Algorithm 1: Compute the log-likelihood of a fine-scale HMM

transformation thereof (log-forward probabilities):

$$\phi_{k,t^*}^{*(i)} = \log[\alpha_{k,t^*}^{*(i)}], \quad \phi_{i,t} = \log[\alpha_{i,t}].$$

The recursive form remains: For $t^* = 2, \dots, T^*$, the fine-scale log-forward probabilities satisfy

$$\phi_{k,1}^{*(i)} = \log[\delta_k^{*(i)}] + \log[f^{*(i,k)}(X_{t,1}^*)], \quad (3.6)$$

$$\phi_{k,t^*}^{*(i)} = \log[f^{*(i,k)}(X_{t,t^*}^*)] + \log\left[\sum_{j=1}^{N^*} \gamma_{jk}^{*(i)} \exp[\phi_{j,t^*-1}^{*(i)} - c_{t^*-1}]\right] + c_{t^*-1}, \quad (3.7)$$

where $c_{t^*-1} = \max\{\phi_{1,t^*-1}^{*(i)}, \dots, \phi_{N^*,t^*-1}^{*(i)}\}$. As it is easy to verify, the log-likelihood of a fine-scale HMM results from these variables as

$$\log \mathcal{L}^{HMM}(\theta^{*(i)} \mid (X_{t,t^*}^*)_{t^*}) = \log\left[\sum_{k=1}^{N^*} \exp[\phi_{k,T^*}^{*(i)} - c_{T^*}]\right] + c_{T^*},$$

where $c_{T^*} = \max\{\phi_{1,T^*}^{*(i)}, \dots, \phi_{N^*,T^*}^{*(i)}\}$. See Algorithm 1 for a pseudo-code of the computation. Similarly, the coarse-scale log-forward probabilities satisfy

$$\phi_{i,1} = \log[\delta_i] + \log \mathcal{L}^{HMM}(\theta^{*(i)} \mid (X_{1,t^*}^*)_{t^*}) + \log[f^{(i)}(X_1)], \quad (3.8)$$

$$\phi_{i,t} = \log \mathcal{L}^{HMM}(\theta^{*(i)} \mid (X_{t,t^*}^*)_{t^*}) + \log[f^{(i)}(X_t)] + \log\left[\sum_{j=1}^N \gamma_{ji} \exp[\phi_{j,t-1} - c_{t-1}]\right] + c_{t-1}, \quad (3.9)$$

1: $\theta = (\delta, \Gamma, (f^{(i)})_i)$	▷ initialize the coarse-scale parameters
2: for $i = 1, \dots, N$ do	
3: $\theta^{*(i)} = (\delta^{*(i)}, \Gamma^{*(i)}, (f^{*(i,k)})_k)$	▷ initialize the fine-scale parameters
4: end for	
5: procedure $\log \mathcal{L}^{HHMM}(\theta, (\theta^{*(i)})_i \mid (X_t)_t, ((X_{t,t^*}^*)_{t^*})_t)$	
6: for $i = 1, \dots, N$ do	
7: $\phi_{i,1} = \log[\delta_i] + \log \mathcal{L}^{HMM}(\theta^{*(i)} \mid (X_{1,t^*}^*)_{t^*}) + \log[f^{(i)}(X_1)]$	▷ $\log \mathcal{L}^{HMM}$ is defined in Algorithm 1
8: end for	
9: for $t = 2, \dots, T$ do	
10: $c_{t-1} = \max\{\phi_{1,t-1}, \dots, \phi_{N,t-1}\}$	
11: for $i = 1, \dots, N$ do	
12: $\phi_{i,t} = \log \mathcal{L}^{HMM}(\theta^{*(i)} \mid (X_{t,t^*}^*)_{t^*}) + \log[f^{(i)}(X_t)] + \log \left[\sum_{j=1}^N \gamma_{ji} \exp[\phi_{j,t-1} - c_{t-1}] \right] + c_{t-1}$	
13: end for	
14: end for	
15: $c_T = \max\{\phi_{1,T}, \dots, \phi_{N,T}\}$	
16: return $\log \left[\sum_{i=1}^N \exp[\phi_{i,T} - c_T] \right] + c_T$	
17: end procedure	

Algorithm 2: Compute the log-likelihood of the HHMM

where $c_{t-1} = \max\{\phi_{1,t-1}, \dots, \phi_{N,t-1}\}$ and $t = 2, \dots, T$. The log-likelihood of the HHMM results from these variables as

$$\log \mathcal{L}^{HHMM}(\theta, (\theta^{*(i)})_i \mid (X_t)_t, ((X_{t,t^*}^*)_{t^*})_t) = \log \left[\sum_{i=1}^N \exp[\phi_{i,T} - c_T] \right] + c_T,$$

where $c_T = \max\{\phi_{1,T}, \dots, \phi_{N,T}\}$. See Algorithm 2 for the pseudo-code. The proofs for the equations (3.6) to (3.9) are similar to those of (3.2) to (3.5). For the sake of completeness, they can be found in the appendix as well.

Secondly, we have to mind 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 transition probability matrices 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 $(\eta_{ij})_{i \neq j}$ for the non-diagonal entries of Γ 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 via 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 $N \times N$ -transition probability matrix. Furthermore, if for example a normal distribution is chosen as a state-dependent distribution,

the estimated standard deviation has to be strictly positive. This can be achieved by applying an exponential transformation to the unconstrained estimator.

A third source of conflicts lies in the circumstance that the likelihood is maximized with respect to a relatively large number of parameters. We must be aware that this generally yields to local maxima apart from the global maximum. Common Newton-Raphson-type optimisation routines are unable to distinguish local maxima from the global one. To avoid the trap of ending up at a local maximum, it is recommended to run the maximization routine several times at random starting points. The number of runs should increase with the number of parameters.

3.5 The Viterbi Algorithm for State Decoding

For a proper model interpretation, it is of interest to decode the hidden states. The term

$$\arg \max_{S_1, \dots, S_T} f(S_1, \dots, S_T \mid X_1, \dots, X_T) \quad (3.10)$$

represents the most-likely underlying state sequence $(S_t)_t$ of a HMM given the data $(X_t)_t$. Transforming the conditional probability reveals that (3.10) is equivalent to the term

$$\arg \max_{S_1, \dots, S_T} f(S_1, \dots, S_T, X_1, \dots, X_T),$$

which in turn can be computed using the so-called Viterbi algorithm, see [Zucchini et al., 2016, p. 88 ff.]. This algorithm is based on the variables

$$\xi_{i,t} = \max_{S_1, \dots, S_{t-1}} f(S_1, \dots, S_{t-1}, S_t = i, X_1, \dots, X_t),$$

$t = 1, \dots, T$ and $i = 1, \dots, N$, which can be calculated recursively:

$$\xi_{i,1} = \delta_i f^{(i)}(X_1), \quad (3.11)$$

$$\xi_{i,t} = \max_j (\xi_{j,t-1} \gamma_{ji}) f^{(i)}(X_t). \quad (3.12)$$

See Appendix A for proofs of (3.11) and (3.12). 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 back in time:

$$\begin{aligned} \hat{S}_T &= \arg \max_i \xi_{i,T}, \\ \hat{S}_t &= \arg \max_i \xi_{i,t} \gamma_{i\hat{S}_{t+1}}, \quad t = T - 1, \dots, 1. \end{aligned}$$

```

1:  $\theta = (\delta, \Gamma, (f^{(i)})_i)$ 
2: procedure VITERBI( $\theta, (X_t)_t$ )
3:   for  $i = 1, \dots, N$  do
4:      $\kappa_{i,1} = \log[\delta_i] + \log[f^{(i)}(X_1)]$ 
5:   end for
6:   for  $t = 2, \dots, T$  do
7:     for  $i = 1, \dots, N$  do
8:        $\kappa_{i,t} = \max_j (\kappa_{j,t-1} + \log[\gamma_{ji}]) + \log[f^{(i)}(X_t)]$ 
9:     end for
10:  end for
11:   $\hat{S}_T = \arg \max_i \kappa_{i,T}$ 
12:  for  $t = T - 1, \dots, 1$  do
13:     $\hat{S}_t = \arg \max_i \kappa_{i,t} \gamma_{i\hat{S}_{t+1}}$ 
14:  end for
15:  return  $(\hat{S}_t)_t$ 
16: end procedure

```

Algorithm 3: Decode the hidden states (Viterbi algorithm, logarithmic version)

Mind that \hat{S}_t incorporates the transition to the already decoded, subsequent state \hat{S}_{t+1} . As for the likelihood function, we need to prevent numerical conflicts (see Section 3.4). Therefore we apply the logarithm again, see Algorithm 3 where $\kappa_{i,t} = \log[\xi_{i,t}]$. Decoding a HHMM is straightforward by first decoding the coarse-scale process and using this information to decode the fine-scale process afterwards.

3.6 Pseudo-Residuals

Using so-called pseudo-residuals enables us to check whether a fitted HMM describes the data 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 in the following way:

$$\begin{aligned} &\text{If } X_t \text{ has the invertible distribution function } F_{X_t}, \text{ then } Z_t = \Phi^{-1}(F_{X_t}(X_t)) \\ &\text{is standard normally distributed,} \end{aligned} \tag{3.13}$$

where Φ denotes the cumulative distribution function of the standard normal distribution. The proof of statement (3.13) is appended. Now we can argue that the data $(X_t)_t$ is modelled well if the pseudo-residuals $(Z_t)_t$ are approximately standard normally distributed.

To check a HHMM, we first decode the coarse-scale process. Then, we assign each coarse-scale observation its estimated distribution function and perform the transformation (3.13). Using the decoded coarse-scale states, we treat the fine-scale observations analogously.

4 Data Application

In a next step, the HHMM is applied to the DAX. Section 4.1 discusses how we choose from the class of candidate models. In Section 4.2, we check the chosen model and adjust any noticeable lack of fit. In Section 4.3, the model results are analysed. Subsequently in Section 4.4, we discuss the impact of the fine-scale time horizon on the results. In Section 4.5, the Viterbi algorithm is applied to detect short- and long-term trends. Section 4.6 discusses the accuracy of the decoding by means of a bootstrapping procedure. Section 4.7 compares the model results to those we obtain from a basic HMM. In the end, Section 4.8 provides a brief outlook on how to derive a trading strategy based on the HHMM.

4.1 Model Selection

The HHMM is a relatively flexible model, allowing for any fine-scale time horizon, all kinds of state-dependent distributions and different numbers of states. Hence, selecting a good or in some sense *the best* model is a demanding task.

In the first place, we define two hierarchies. Remember from Chapter 2 that our data encompasses an amount of 4817 daily log-returns, which we identify as the fine-scale observations. We segment them into T distinct chunks, each with the same amount of T^* data points. Regarding a future model interpretation, it would be convenient to segment per month. However, since each month has a different amount of trading days, this does not lead to equally sized chunks. Therefore, we choose $T^* = 30$ for now¹, which yields an amount of $T = \lfloor 4817/30 \rfloor = 160$ chunks. Notice that we have to drop the 17 latest observations. For the coarse-scale hierarchy, we compute the average of each chunk and identify these $T = 160$ average log-returns as the coarse-scale observations. These data points smooth the fine-scale fluctuations and serve as an indicator for long-term trends.

Secondly, we discuss the choice of the state-dependent distributions. For now, we select normal distributions. This choice is motivated by mathematical finance: Stock prices are often described by a geometric Brownian motion, cf. [Black and Scholes, 1973]. This leads to log-normally distributed prices and thereby to normally distributed log-prices. Prices however

1. On average, the german stock market has 20 trading days per month, see www.boerse.de/boersenfeiertage (accessed on June 10, 2019). Hence, $T^* = 30$ represents about six weeks of trading, which seems to be a reasonable time span at which short-term trends can manifest themselves. Obviously, we later have to verify that the model results are independent from this arbitrary choice.

cannot be assumed to satisfy the conditional-independence assumption. Therefore, we model the log-returns rather than the log-prices.

Thirdly, we have to select the state space for the hidden state processes. A priori it is not clear in how many states the behaviour of a financial market can be divided. Therefore, we fit candidate HHMMs with $N = 2, 3, 4$ coarse-scale and $N^* = 2, 3, 4$ fine-scale states and compute the Akaike information criterion (AIC), cf. [Akaike, 1974], and the Bayesian information criterion (BIC), cf. [Schwarz, 1978]. These model selection criteria compare the candidate models, rewarding the goodness of fit (measured by the log-likelihood) while penalising the complexity. Theoretically, the relative best model corresponds to the lowest AIC or BIC value, respectively. The values for the candidate models are summarised in Table 3, where p denotes the number of parameters estimated by the model (which is the measure for the model complexity).² The minima are printed in bold. As we observe, AIC and BIC do not agree on the number of states:

N	N^*	p	$\log \mathcal{L}^{HHMM}$	AIC	BIC
2	2	18	15095,47	-30154,94	-30037,55
2	3	30	15146,60	-30233,20	-30037,55
2	4	46	15174,79	-30257,58	-29957,59
3	2	30	15149,44	-30238,88	-30043,23
3	3	48	15197,19	-30298,38	-29985,34
3	4	72	15228,01	-30312,02	-29842,47
4	2	44	15179,93	-30271,86	-29984,91
4	3	68	15231,61	-30327,22	-29883,75
4	4	100	15274,31	-30348,62	-29696,46

Table 3: AIC and BIC for candidate HHMMs with normal distributions

AIC prefers the maximum number while BIC prefers $N = 3$ and $N^* = 2$. As [Pohle et al., 2017] points out, we must be aware that especially AIC tends to select too many states in case of HMMs. In this respect, [Rydén et al., 1998] demonstrates that the HMM has a tendency of allocating exceptional returns in a separate state. [Pohle et al., 2017] therefore recommends to not simply rely on these criteria but to carefully inspect the estimated parameters. The model with $N = 3$ and $N^* = 3$ for example reveals an undesirable feature: For each coarse-scale state, a fine-scale state exists which seems to be redundant in the sense that its exit probability equals 1. Most likely, these states do only describe outliers. In contrast, all states of the model with $N = 3$ and $N^* = 2$ are persistent with reasonable transition probabilities. Therefore, we go with BIC and choose $N = 3$ coarse-scale and $N^* = 2$ fine-scale states. See Appendix B for the estimated parameters of both mentioned models.

2. Here, $p = p(N, N^*) = N(N - 1) + NN^*(N^* - 1) + N + NN^* + N + NN^*$.

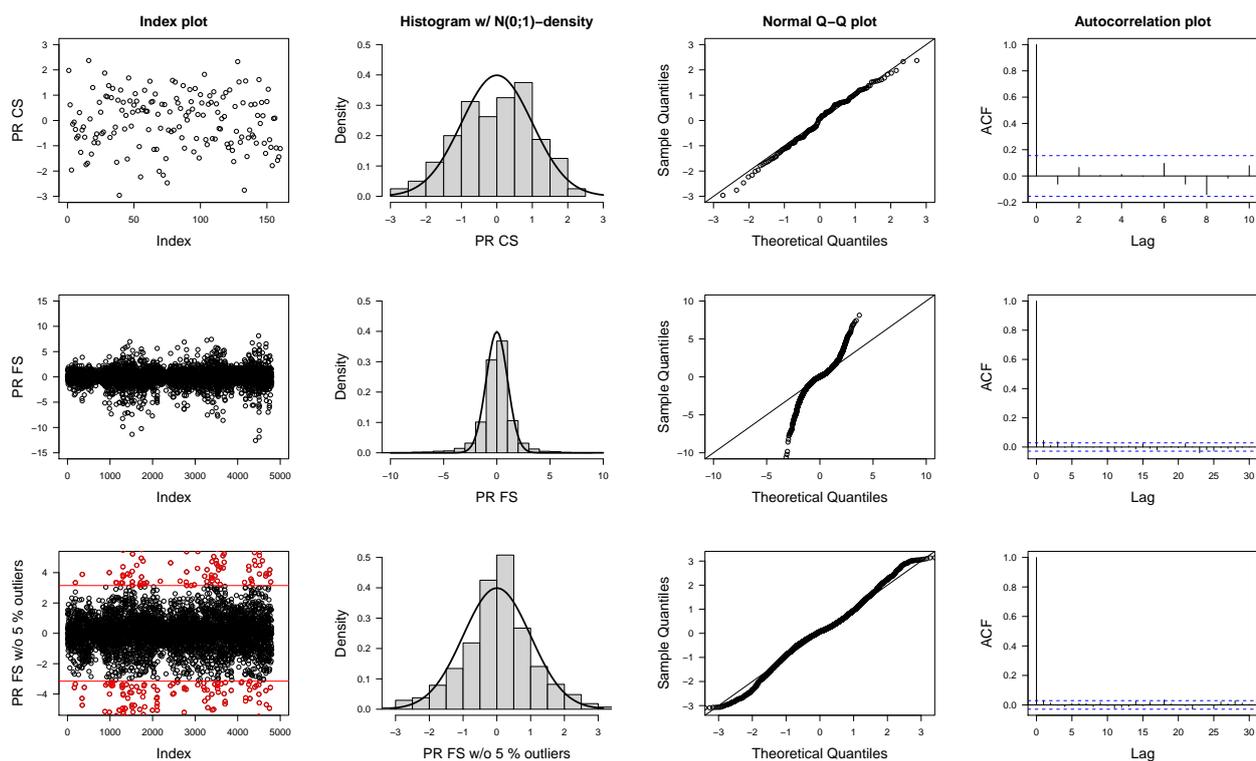


Figure 3: Pseudo-residuals of the HHMM with normal distributions (coarse-scale layer in first row, fine-scale layer in second and third row)

4.2 Model Checking

In the preceding section, normal state-dependent distributions and $N = 3$ coarse-scale and $N^* = 2$ fine-scale states were selected. Now, this particular model choice has to be checked. As mentioned in Section 3.6, we do so by computing and analysing the models' pseudo-residuals. See Figure 3 for different visualisations of the pseudo-residuals generated by the DAX data.

The first row visualises the pseudo-residuals of the coarse-scale observations. The index plot in the first, the histogram in the second and the normal quantile-quantile plot in the third column all confirm, that the pseudo-residuals are approximately standard normally distributed. Furthermore, the top right plot indicates the absence of significant autocorrelation. This is important to rely on the conditional independence-assumption.

The graphics in the second row correspond to the pseudo-residuals of the fine-scale observations. Here, the normal quantile-quantile plot clearly indicates a lack of fit: The quantiles do not form a straight line and hence the fine-scale observations are not well-modelled. However, only the tails seem to be problematic: The lower quantiles are smaller and the upper quantiles are greater than expected. In the third row, 5% of the most extreme values (marked in red) are removed. The remaining pseudo-residuals are roughly standard normally distributed.

We conclude that the state-dependent distributions have to be improved in terms of higher probability mass on the tails. This finding is consistent with the analysis in [Praetz, 1972] and [Harris and Küçüközmen, 2001]. Therefore, we generalise from the normal distribution \mathcal{N} to the t-distribution \mathcal{T} , which possesses an additional kurtosis parameter ν (commonly known as degrees of freedom). So far, the log-return X_t is modelled by the scaled normal distribution $\sigma_{S_t}\mathcal{N}(0; 1) + \mu_{S_t} =: \mathcal{N}(\mu_{S_t}; \sigma_{S_t})$. Now, we assume that X_t is stemming from the scaled t-distribution $\sigma_{S_t}\mathcal{T}_{\nu_{S_t}} + \mu_{S_t} =: \mathcal{T}(\mu_{S_t}; \sigma_{S_t}; \nu_{S_t})$. All parameters again depend on the current state S_t . The new distribution function F_{X_t} and density f_{X_t} of X_t are

$$F_{X_t}(x) = F_{\mathcal{T}_{\nu_{S_t}}}\left(\frac{x - \mu_{S_t}}{\sigma_{S_t}}\right) \quad \text{and} \quad f_{X_t}(x) = \sigma_{S_t}^{-1} f_{\mathcal{T}_{\nu_{S_t}}}\left(\frac{x - \mu_{S_t}}{\sigma_{S_t}}\right), \quad (4.1)$$

respectively. Here, $F_{\mathcal{T}_{\nu_{S_t}}}$ denotes the distribution function and $f_{\mathcal{T}_{\nu_{S_t}}}$ the density of the t-distribution with ν_{S_t} degrees of freedom. The proof of statement (4.1) is appended.

The class of candidate models is enlarged³ and the selection of the number of states has to be repeated. AIC and BIC prefer the same models as before, see Table 4. Again, HHMMs

N	N^*	p	$\log \mathcal{L}^{HHMM}$	AIC	BIC
2	2	24	15103,33	-30158,66	-30002,14
2	3	38	15150,56	-30225,12	-29977,30
2	4	56	15181,32	-30250,64	-29885,43
3	2	39	15171,78	-30265,56	-30011,22
3	3	60	15201,63	-30283,26	-29891,96
3	4	87	15230,88	-30287,76	-29720,28
4	2	56	15189,44	-30266,88	-29901,67
4	3	84	15246,82	-30325,64	-29777,83
4	4	120	15292,63	-30345,26	-29562,67

Table 4: AIC and BIC for candidate HHMMs with t-distributions

with too many states exhibit redundant states, see Appendix B for the estimated parameters of the model with $N = 3$ and $N^* = 3$. As a consequence, we stick to $N = 3$ coarse-scale and $N^* = 2$ fine-scale states. The estimated parameters of this model are presented in the following section. Computing the pseudo-residuals of the new model yields the results visualised in Figure 4. Especially the normal quantile-quantile plot in the second row highlights the improvement: Now, the empirical quantiles of the fine-scale pseudo-residuals are roughly equal to the ones of the standard normal distribution.

3. Letting $\nu_{S_t} \rightarrow \infty$ brings back the normal distribution.

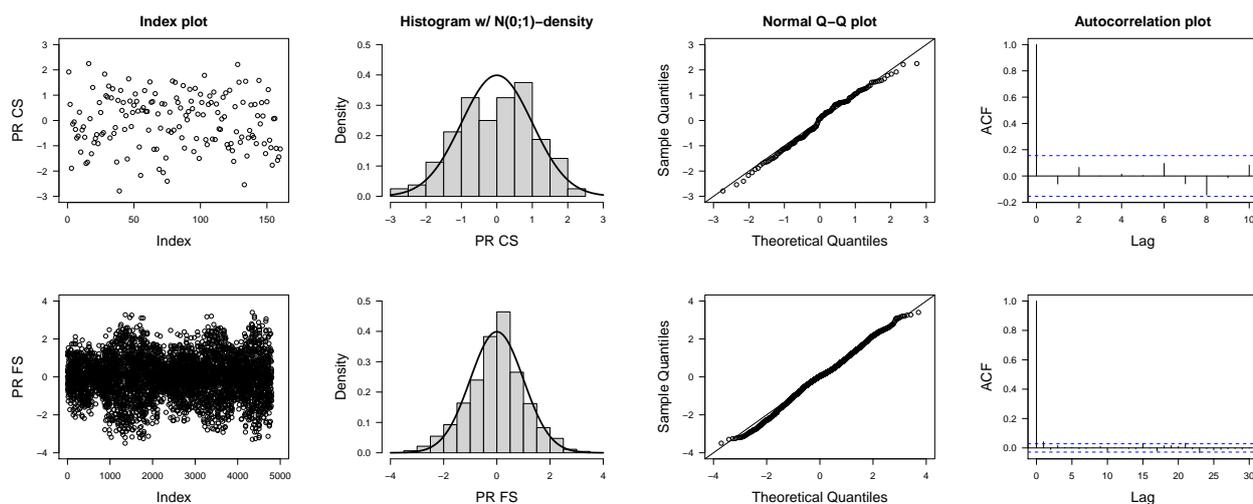


Figure 4: Pseudo-residuals of the HHMM with t -distributions (coarse-scale layer in first row, fine-scale layer in second row)

4.3 The Model Results

Inspecting the estimates of the HHMM, we account the coarse-scale layer first. The matrix Γ of transition probabilities and the stationary distribution vector π identify the hidden state process. The state-dependent distributions $f^{(1)}$, $f^{(2)}$ and $f^{(3)}$ reveal characteristics of the returns in the different states.

$$\Gamma = \begin{pmatrix} 0,9324 & 0,0428 & 0,0248 \\ 0,0772 & 0,8399 & 0,0829 \\ 0,0000 & 0,3201 & 0,6799 \end{pmatrix} \quad \begin{aligned} f^{(1)} &= \mathcal{T}(6,25 \cdot 10^{-4}; 1,28 \cdot 10^{-3}; 26) \\ f^{(2)} &= \mathcal{T}(1,68 \cdot 10^{-4}; 1,86 \cdot 10^{-3}; 17) \\ f^{(3)} &= \mathcal{T}(-20,33 \cdot 10^{-4}; 5,03 \cdot 10^{-3}; 28) \end{aligned}$$

$$\pi = (0,4587 \quad 0,4017 \quad 0,1396)$$

State 1 corresponds to the highest expected yield (return of $\exp[0.000625] - 1 \approx 0.06\%$), state 3 to the highest expected loss (return of $\exp[-0.002033] - 1 \approx -0.2\%$) while state 2 is in-between (return of $\exp[0.000168] - 1 \approx 0.02\%$). Furthermore, the higher the expected return, the lower the standard deviation turns out. The kurtosis parameters are relatively large, hence the t -distributions are very close to the shape of normal distributions. This comes as no surprise since Figure 3 already approved normal distributions for the coarse-scale observations. The matrix Γ indicates that all states are quite persistent. Being in state 1, it is roughly twice as likely to switch to state 2 than to state 3. Being in state 2, the probability to switch to state 1 is nearly the same as switching to state 3. Note that a direct transition from state 3 to state 1 has probability zero. The entries of π can be interpreted as the proportion of time the market

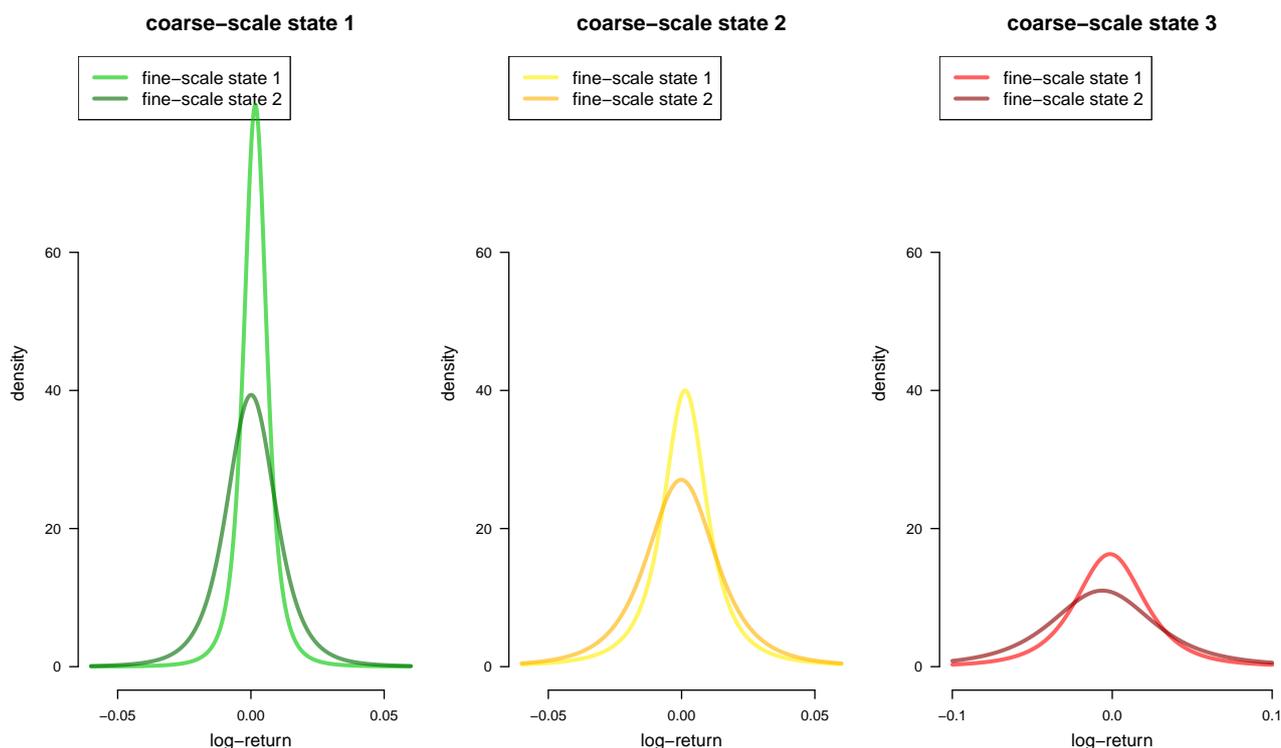


Figure 5: Estimated state-dependent fine-scale distributions

spends in the different states⁴. State 1 is leading, closely followed by state 2 while state 3 is rarer attained.

Within each coarse-scale state, a two-state HMM describes short-term trends. The estimated fine-scale distributions are visualised in Figure 5.

$$\begin{aligned} \Gamma^{*(1)} &= \begin{pmatrix} 0,9128 & 0,0872 \\ 0,0542 & 0,9458 \end{pmatrix} & f^{*(1,1)} &= \mathcal{T}(16,48 \cdot 10^{-4}; 4,52 \cdot 10^{-3}; 3) \\ \pi^{*(1)} &= (0,3833 \quad 0,6167) & f^{*(1,2)} &= \mathcal{T}(0,87 \cdot 10^{-4}; 9,53 \cdot 10^{-3}; 4) \end{aligned}$$

Within coarse-scale state 1, fine-scale state 1 represents short-term trends with a relatively high positive expected return ($\exp[0.001648] - 1 \approx 0.16\%$). In fine-scale state 2, the expected return is about 0. The latter state exhibits twice as much standard deviation as the former one. Both states are persistent. In 60% of the time, state 2 is active. The kurtosis parameters 3 and 4 indicate heavy tails.

4. This interpretation is motivated by the Ergodic theorem for Markov chains, see [Levin et al., 2017, p. 58]: The value of the stationary distribution almost-surely equals the proportion of time the process will spend in the corresponding state in the long run.

$$\begin{aligned}\Gamma^{*(2)} &= \begin{pmatrix} 0,9887 & 0,0113 \\ 0,0037 & 0,9963 \end{pmatrix} & f^{*(2,1)} &= \mathcal{T}(13,45 \cdot 10^{-4}; 8,84 \cdot 10^{-3}; 2) \\ \pi^{*(2)} &= (0,2467 \quad 0,7533) & f^{*(2,2)} &= \mathcal{T}(-1,18 \cdot 10^{-4}; 13,58 \cdot 10^{-3}; 3)\end{aligned}$$

In case of coarse-scale state 2, the model estimates fine-scale state 1 with a positive expected return ($\exp[0.001345] - 1 \approx 0.13\%$) and fine-scale state 2 with a slightly negative expected return ($\exp[-0.000118] - 1 \approx -0.01\%$). Again, the standard deviation corresponding to the state of lower expected return is larger. With values of 2 and 3, the kurtosis of both distributions is relatively high. Both states are persistent, while state 2 is active in 75% of the time.

$$\begin{aligned}\Gamma^{*(3)} &= \begin{pmatrix} 0,9537 & 0,0463 \\ 0,0757 & 0,9243 \end{pmatrix} & f^{*(3,1)} &= \mathcal{T}(-13,79 \cdot 10^{-4}; 22,55 \cdot 10^{-3}; 3) \\ \pi^{*(3)} &= (0,6205 \quad 0,3795) & f^{*(3,2)} &= \mathcal{T}(-62,26 \cdot 10^{-4}; 33,38 \cdot 10^{-3}; 3)\end{aligned}$$

Finally, for coarse-scale state 3, both fine-scale states exhibit high negative expected returns ($\exp[-0.001379] - 1 \approx -0.14\%$ for fine-scale state 1 and $\exp[-0.006226] - 1 \approx -0.62\%$ for fine-scale state 2) and high standard deviation. Again, kurtosis is high and both states are quite persistent. The time proportion for state 1 is slightly higher than for state 2.

Summing up, the model provides three indicators to distinguish between the states: Expected return, volatility and transition behaviour. On the coarser scale, state 1 represents a positive market mood while coarse-scale state 3 corresponds to periods of falling prices. For each coarse-scale state, fine-scale state 1 captures short-term upwards trends and fine-scale state 2 short-term downwards trends. Translating the findings to financial terms, we label coarse-scale state 1 *bullish market* and coarse-scale state 3 *bearish market*.⁵ Furthermore, we name coarse-scale state 2 *state of correction*, which is consistent with the expertise of the Hungarian investor André Kostolany: In [Kostolany, 2015], he argues that a financial market can be overrated or underrated, which causes periods of adaptation to the current proportion of supply and demand. The idea of capturing such periods in an own state seems convenient.

Although the labelling simplifies the model interpretation, we should not over-interpret the states. By no means we designed a universally valid definition of say a bullish market. All deductions have to be regarded in the context of the model formulation.

5. These terms are very popular in finance, their origin however is unclear. Most likely, they are simply metaphors: Prices can rise like bulls thrust upwards with their horns or drop like bears swipe downwards with their paws, see www.boerse.ard.de/boersenwissen (accessed on June 10, 2019).

4.4 Impact of the Fine-Scale Time Horizon

Up to now, we fixed a fine-scale time horizon of $T^* = 30$ trading days. Remember that this particular choice was not evidence-based and only served as a remedy because we could not segment by more convenient periods like months, see Section 4.1. We therefore have to discuss the impact of T^* on the model and check whether different choices lead to similar results.

In our model, T^* determines the segmentation of the log-returns into chunks: A small value leads to many chunks with only few observations each, while a larger value results in less chunks including more observations. Hence, a change in T^* affects the fine-scale HMMs fitted to each chunk and the averaging for the coarse-scale observations. Obviously, the extend of the effect depends on the type and amount of data. The effect on our data can be investigated by varying different values for T^* while fixing all other model parameters and repeating the parameter estimation. Table 5 gives an exemplary overview of the estimates for $T^* = 25$ and $T^* = 35$. Although we observe a minor discrepancy to the estimates for $T^* = 30$, the main

	$T^* = 25$	$T^* = 35$
Γ	$\begin{pmatrix} 0,8603 & 0,1330 & 0,0067 \\ 0,0596 & 0,8800 & 0,0604 \\ 0,0000 & 0,2574 & 0,7426 \end{pmatrix}$	$\begin{pmatrix} 0,8865 & 0,1071 & 0,0064 \\ 0,0972 & 0,8432 & 0,0596 \\ 0,0000 & 0,3560 & 0,6440 \end{pmatrix}$
$f^{(1)}$	$\mathcal{T}(7,05 \cdot 10^{-4}; 1,54 \cdot 10^{-3}; 27)$	$\mathcal{T}(7,06 \cdot 10^{-4}; 1,02 \cdot 10^{-3}; 24)$
$f^{(2)}$	$\mathcal{T}(2,19 \cdot 10^{-4}; 2,30 \cdot 10^{-3}; 22)$	$\mathcal{T}(1,39 \cdot 10^{-4}; 2,14 \cdot 10^{-3}; 17)$
$f^{(3)}$	$\mathcal{T}(-30,41 \cdot 10^{-4}; 4,93 \cdot 10^{-3}; 30)$	$\mathcal{T}(-22,33 \cdot 10^{-4}; 5,47 \cdot 10^{-3}; 25)$
$\Gamma^{*(1)}$	$\begin{pmatrix} 0,9537 & 0,0463 \\ 0,0747 & 0,9253 \end{pmatrix}$	$\begin{pmatrix} 0,9168 & 0,0832 \\ 0,0443 & 0,9557 \end{pmatrix}$
$f^{*(1,1)}$	$\mathcal{T}(16,76 \cdot 10^{-4}; 4,76 \cdot 10^{-3}; 7)$	$\mathcal{T}(15,34 \cdot 10^{-4}; 3,90 \cdot 10^{-3}; 3)$
$f^{*(1,2)}$	$\mathcal{T}(-1,03 \cdot 10^{-4}; 9,70 \cdot 10^{-3}; 8)$	$\mathcal{T}(0,17 \cdot 10^{-4}; 9,24 \cdot 10^{-3}; 7)$
$\Gamma^{*(2)}$	$\begin{pmatrix} 0,9714 & 0,0286 \\ 0,0125 & 0,9875 \end{pmatrix}$	$\begin{pmatrix} 0,9848 & 0,0152 \\ 0,0202 & 0,9798 \end{pmatrix}$
$f^{*(2,1)}$	$\mathcal{T}(12,03 \cdot 10^{-4}; 10,02 \cdot 10^{-3}; 6)$	$\mathcal{T}(8,81 \cdot 10^{-4}; 11,04 \cdot 10^{-3}; 4)$
$f^{*(2,2)}$	$\mathcal{T}(-0,41 \cdot 10^{-4}; 15,22 \cdot 10^{-3}; 3)$	$\mathcal{T}(-0,73 \cdot 10^{-4}; 16,54 \cdot 10^{-3}; 4)$
$\Gamma^{*(3)}$	$\begin{pmatrix} 0,9743 & 0,0257 \\ 0,1083 & 0,8917 \end{pmatrix}$	$\begin{pmatrix} 0,9581 & 0,0419 \\ 0,1515 & 0,8485 \end{pmatrix}$
$f^{*(3,1)}$	$\mathcal{T}(-26,77 \cdot 10^{-4}; 12,00 \cdot 10^{-3}; 9)$	$\mathcal{T}(-22,48 \cdot 10^{-4}; 14,18 \cdot 10^{-3}; 5)$
$f^{*(3,2)}$	$\mathcal{T}(-45,79 \cdot 10^{-4}; 27,76 \cdot 10^{-3}; 10)$	$\mathcal{T}(-37,99 \cdot 10^{-4}; 28,19 \cdot 10^{-3}; 7)$

Table 5: Estimates for $T^* = 25$ and $T^* = 35$

characteristics like transition behaviour or proportion of expected log-return to volatility do not change. Only if the fine-scale time horizon is chosen relatively small ($T^* \leq 20$) or relatively large ($T^* \geq 40$), the differences become more significant.

We conclude that there exists a range of values around $T^* = 30$ that lead to similar model results. Therefore, the effect of the segmentation is negligible as long as the choice for T^* is not too extreme. Our particular choice $T^* = 30$ equates to about six weeks of trading, which seems to be a reasonable time span in which short-term trends can manifest themselves. However, we have to avoid to interpret the estimates down to the last decimal place. Instead, they should be interpreted in relation to each other across the different states.

4.5 Detection of Bearish and Bullish Markets

The Viterbi algorithm (see Section 3.5) puts us in the position to detect past trends in the DAX. Figure 6 visualises the decoded states of years 2000 to 2018 in the graphic on top and a close-up of the financial crisis in 2008 below.

Evidently, the decoding fits to our expectations.⁶ From 2000 to 2002, the index drops by more than 3000 points. The model detects coarse-scale state 2. On two occasions, it switches from fine-scale state 2 to fine-scale state 1, indicating short-term upwards trends. The long-term downwards trend intensifies in 2003 with an increased volatility in the log-returns, leading to the detection of a bearish market. At the end of 2003, the market calms down. Until 2007, the index slowly but steadily rises, log-returns are positive on average with low volatility. Occasionally, the model detects a faster increase of the index, highlighted in light green. The bullish market comes to an end in 2007, where the model switches to the state of correction.

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 September 15, 2008. The model switches to coarse-scale state 3. Here, it would be an over-interpretation to emphasise the exact date of the coarse-scale state transition, since this is influenced by the choice of the fine-scale time horizon. However, it is noteworthy that the model immediately detects fine-scale state 2 within the bearish market (which represents most lossy periods) and switches to the calmer fine-scale state 1 as soon as more moderate log-returns occur.

Between 2008 and 2011, the index rises but shows increased volatility. Therefore, the model remains in the state of correction. At the end of 2011, another bearish market occurs, which lasts shortly. The market corrects itself quickly and turns into a bullish market, which persists until 2015. Then, the DAX enters another period of correction. From 2016 until the end of 2018, the index rises again. In 2017, we observe light green periods in which the DAX gains nearly 4000 points within just a few months. In 2018, the skyrocketing is stopped. As the volatility remains low, the model retains coarse-scale state 1. The last decoded state on November 29, 2018 is measured as fine-scale state 2 within coarse-scale state 1.

6. The trend changes are accompanied or possibly caused by historical events, see www.faz.net/-gv6-9brsq (accessed on June 20, 2019).

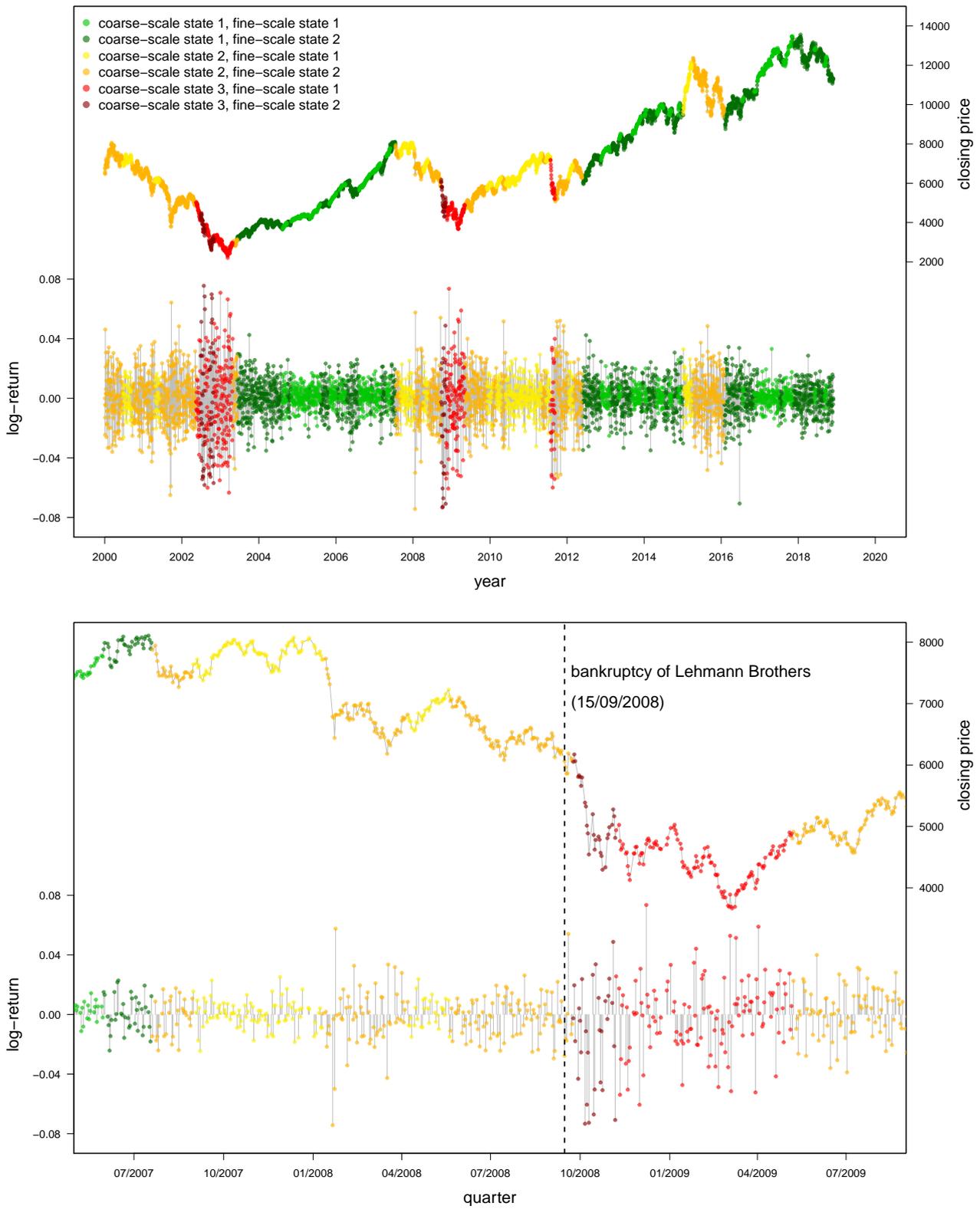


Figure 6: Decoded states and a close-up of the financial crisis 2008

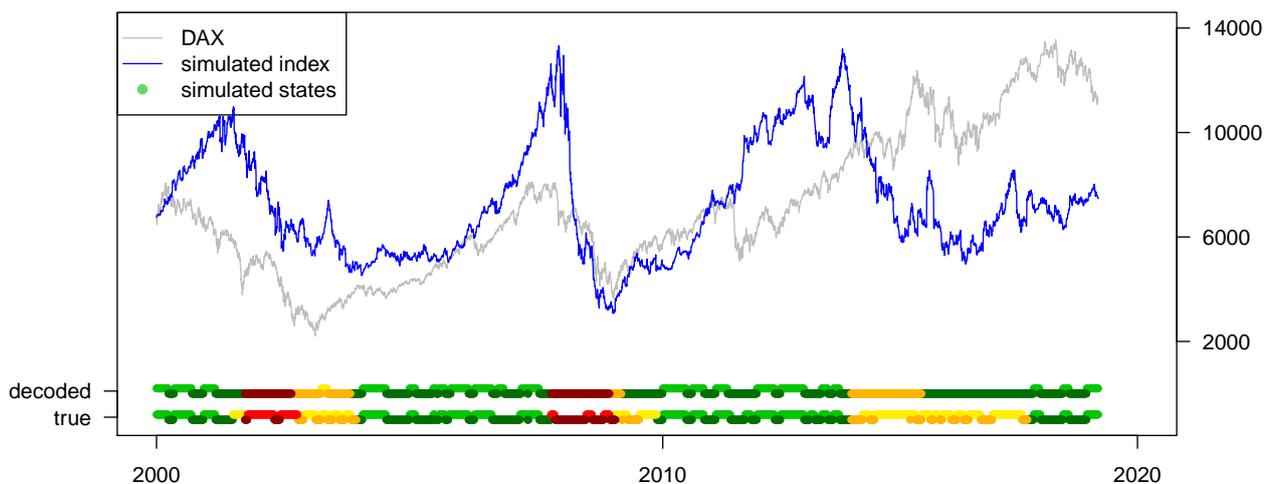


Figure 7: Bootstrapping of the Viterbi algorithm

4.6 Accuracy of the State Decoding

Analysing the accuracy of the state decoding is an essential part of the model interpretation, which can be carried out by means of a bootstrapping procedure. Based on the model parameters, we repeatedly simulate log-returns. For each simulation, we memorise the underlying state sequence (the *true* states). Then, the simulations are decoded by using the Viterbi algorithm. The decoded states can hence be compared to the true states. This comparison eventually reveals potential issues in the detection of short-term and long-term trends. Figure 7 shows a sample path⁷ in blue and for comparison purposes the DAX in grey.

On the coarse scale, state 1 and state 3 are distinguished very reliably. Most frequently, the Viterbi algorithm fails in separating state 1 from state 2: In about 30% of the cases, coarse-scale state 1 instead of the true state 2 is decoded and in about 20% of the cases, the reversed issue occurs. Taking this into consideration, the time span 2009 to 2012 has to be interpreted cautiously: Here, the state of correction is detected, see Figure 6, probably due to an increased standard deviation. However, we are now aware that a bullish market for this period is conceivable as well.

On the fine scale, state transitions are best detected within coarse-scale state 1. More frequently, the Viterbi algorithm fails to detect simulated short-term trends within coarse-scale state 2 and coarse-scale state 3. It often misleadingly decodes persistence in the fine-scale state processes. Most likely, this is due to the fact that the shapes of the estimated state-dependent distributions are relatively close to each other, see Figure 5.

Summing up, the Viterbi algorithm performs reasonably well decoding the model states. However, this analysis again sensitises to the danger of over-interpretation.

7. The sample path is generated by back-transforming the log-returns to index values, see Chapter 2.

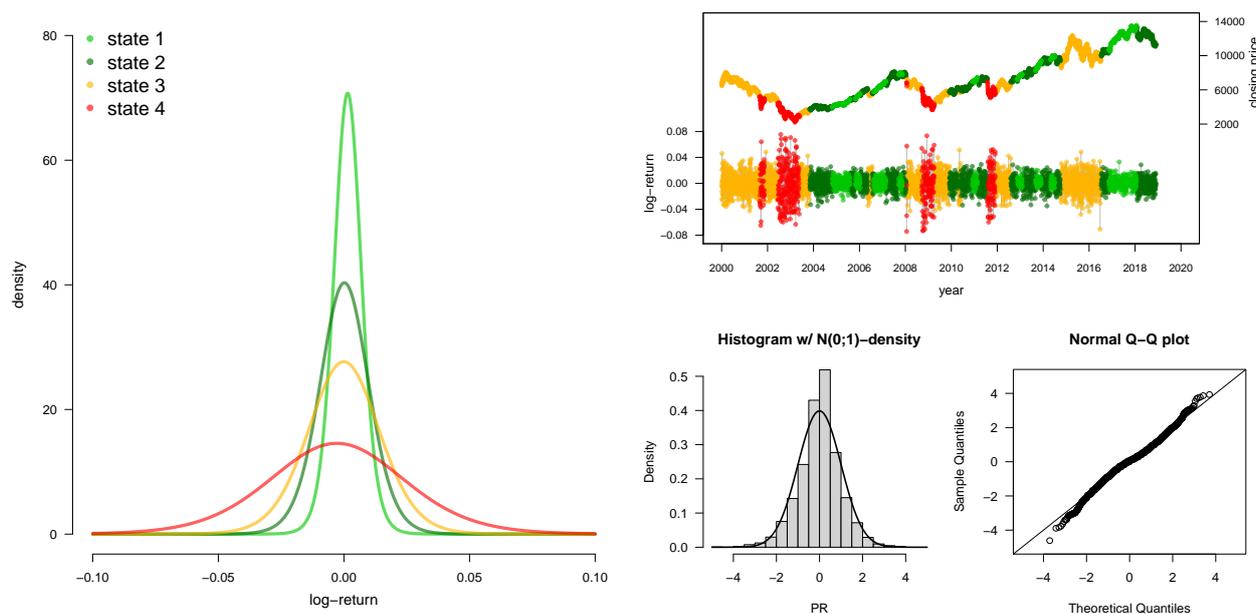


Figure 8: Estimated state-dependent distributions (left), decoded states (top-right) and pseudo-residuals (bottom-right, histogram and quantile-quantile plot) of the HMM

N	p	$\log \mathcal{L}^{HMM}$	AIC	BIC
2	8	14212.34	-28408.68	-28356.87
3	15	14334.49	-28638.98	-28541.83
4	24	14364.23	-28680.46	-28525.03
5	35	14392.64	-28715.28	-28488.61
6	48	14395.91	-28695.82	-28384.95
7	63	14410.11	-28694.22	-28286.21
8	80	14426.37	-28692.74	-28174.63

Table 6: AIC and BIC for candidate HMMs

4.7 Comparison to HMM Results

To spot the improvements in modelling financial data with the HHMM, we fit the basic HMM to the log-return data as well. As for the HHMM, state-dependent t -distributions are the adequate choice. Again, selecting the number of model states N is not straightforward. AIC prefers five and BIC three states respectively, see Table 6. As seen in [Nguyen, 2018] for the S&P 500, four states are selected here for the DAX, justified by the following arguments: First, three states seem to oversimplify the situation for a financial market. Secondly, in the case of five states, a redundant state with exit probability 1 occurs, see Appendix B. Thirdly, four states is a compromise between the model selection criteria elaborated on above.

The estimated parameters of the selected HMM with four states are appended. The model results are visualised in Figure 8. From the quantile-quantile plot we notice that the pseudo-residuals indicate a reasonable model fit. However, the state decoding by the Viterbi algorithm indicates that the difference between short- and long-term trends is not captured. This becomes clear as we back-transform the log-returns to index values, compare to Figure 6. Furthermore, the AIC and BIC values have significantly decreased in comparison to the values for the HHMMs.

4.8 Thoughts on a HHMM-Based Trading Strategy

The main motivation for modelling financial data lies in making predictions on the upcoming market behaviour and deriving investment guidance. This section gives a brief outlook on how to profitably exploit the advantage of the HHMM to distinguish between short-term and long-term trends regarding a trading strategy.

In general, trading strategies have to be adapted to subjective factors like risk-proclivity or pursued investment horizons. For the sake of convenience, the following case is considered: A risk-neutral stock trader wants to invest in the long term. Reasonably, the trader tries to earn money by buying shares at the beginning of a bullish market and selling them at the beginning of a bearish market. The trader aims to solely react to long-term trend changes and wants to ignore short-term price fluctuations as with every purchase and sale of shares fees are due. The HHMM is well suited to identify this difference.

Remember that our model distinguishes market states by three characteristics (expected return, volatility and transition probabilities). Assuming that these characteristics do not significantly change after the model calibration, it can be applied to classify upcoming index values into the model states. This classification in combination with financial knowledge on how to react to trend changes then leads to guidance when to buy, sell or hold shares.⁸

The classification can be carried out iteratively, see Algorithm 4 for a summary of the following procedure. Assume that the market is in coarse-scale state S_{t-1} and fine-scale state S_{t-1}^* at time point $t-1$.⁹ At time point t , we observe the log-return X_t^* and want to know whether this observation indicates a change in the long-term trend or is merely stemming from short-term price fluctuations. For that purpose, take the average \bar{X} over the T^* latest observations (in-

8. Such knowledge can be derived for example from [Kostolany, 2015]. The author states that financial markets exhibit six stages that repeat cyclically. For each stage, Kostolany gives recommendations for the stock exchange. His approach is called *Ei des Kostolany*.

9. The initial point in our case could be fine-scale state 2 within coarse-scale state 1 at the end of the model calibration, see Section 4.5.

```

1:  $\theta = (\delta, \Gamma, (f^{(i)})_i)$ 
2: for  $i = 1, \dots, N$  do
3:    $\theta^{*(i)} = (\delta^{*(i)}, \Gamma^{*(i)}, (f^{*(i,k)})_k)$ 
4: end for
5: procedure CLASSIFY( $X_t^* \mid S_{t-1}, S_{t-1}^*, ((X_{t,t^*})_{t^*})_t, \theta, (\theta^{*(i)})_i, T^*$ )
6:    $\bar{X} = 1/T^* \sum_{s=0}^{T^*-1} X_{t-s}^*$  ▷ compute the average over the fine-scale time horizon
7:    $\Lambda = \arg \max_i \gamma_{S_{t-1}i} f^{(i)}(\bar{X})$ 
8:   if  $\Lambda = S_{t-1}$  then ▷ assume no coarse-scale state transition
9:      $S_t = S_{t-1}$ 
10:     $S_t^* = \arg \max_j \gamma_{S_{t-1}^*j}^{*(S_t)} f^{*(S_t,j)}(X_t^*)$ 
11:   else if  $\Lambda \neq S_{t-1}$  then ▷ check upon a potential coarse-scale state transition
12:      $p^{\text{CS}} = \gamma_{S_{t-1}\Lambda} f^{(\Lambda)}(\bar{X})$ 
13:      $p^{\text{FS}} = \max_{j \neq S_{t-1}^*} \gamma_{S_{t-1}^*j}^{*(S_{t-1})} f^{*(S_{t-1},j)}(X_t^*)$ 
14:     if  $p^{\text{CS}} > p^{\text{FS}}$  then ▷ assume coarse-scale state transition
15:        $S_t = \Lambda$ 
16:        $S_t^* = \arg \max_j \delta_j^{*(S_t)} f^{*(S_t,j)}(X_t^*)$ 
17:     else if  $p^{\text{CS}} \leq p^{\text{FS}}$  then ▷ assume fine-scale state transition
18:        $S_t = S_{t-1}$ 
19:        $S_t^* = \arg \max_{j \neq S_{t-1}^*} \gamma_{S_{t-1}^*j}^{*(S_t)} f^{*(S_t,j)}(X_t^*)$ 
20:     end if
21:   end if
22:   return  $(S_t, S_t^*)$ 
23: end procedure
    
```

Algorithm 4: Classify the upcoming log-returns

cluding X_t^*) and identify which coarse-scale state most likely is accountable for \bar{X} by minding the transition from the subsequent state S_{t-1} . More formally, compute

$$\Lambda = \arg \max_i \gamma_{S_{t-1}i} f^{(i)}(\bar{X}), \text{ where } \bar{X} = 1/T^* \sum_{s=0}^{T^*-1} X_{t-s}^*.$$

If Λ equals S_{t-1} , we assume that no coarse-scale state transition occurred. A possible fine-scale state transition can be identified by means of

$$S_t^* = \arg \max_j \gamma_{S_{t-1}^*j}^{*(S_t)} f^{*(S_t,j)}(X_t^*).$$

However, if Λ differs from S_{t-1} , we might face a change in the long-term trend. In such a situation, we can rely on the maximum likelihood-principle: Compare the probabilities

$$p^{\text{CS}} = \gamma_{S_{t-1}\Lambda} f^{(\Lambda)}(\bar{X}) \text{ and } p^{\text{FS}} = \max_{j \neq S_{t-1}^*} \gamma_{S_{t-1}^*j}^{*(S_{t-1})} f^{*(S_{t-1},j)}(X_t^*)$$

for a coarse-scale and a fine-scale state transition and decide for the event that is more likely.

5 Discussion

The aim of the present thesis was to elaborate the hierarchical hidden Markov model as an improved framework for modelling financial data. The idea was motivated by [Adam et al., 2019], whose approach convinced for animal movement inference. Both fields share the characteristic of patterns occurring on different temporal resolutions. As pointed out in the introduction, the basic hidden Markov model has been used for modelling stock market indices before. However, the models' incapability of capturing trends on different time scales is a significant deficit: Short-term price fluctuations can easily be misinterpreted as a change in the long-term trend which leads to a distorted picture of the underlying mechanisms as a consequence. By adding a hierarchical structure, we paved the way to capture long-term and short-term trends jointly. A priori, it was expected that this extension leads to an improved inference of financial data which yields a better understanding of market behaviour.

We started by presenting the model formulation, a parameter estimation procedure based on the maximum likelihood method, the Viterbi algorithm for state decoding and the concept of pseudo-residuals as a tool for model checking. Subsequently, the methodology was exemplary applied to closing prices of the Deutscher Aktienindex. The thesis covered the chronological progression from the model selection to its interpretation. Obstacles in the choice of the state-dependent distributions and the number of model states were highlighted. The inherent impact of the fine-scale time horizon on the model results was discussed. Subsequently, applying the Viterbi algorithm, we were able to detect so-called bullish and bearish markets. We noticed that the decoding of the states fits well to historical events and a priori expectations. Because of the repeated use of the Viterbi algorithm throughout the thesis, we analysed its accuracy in the field of financial data by a bootstrapping procedure. We concluded that the algorithm performs reasonably well in most scenarios and highlighted problem cases. To spot the improvement of the hierarchical model extension, we also tested the non-hierarchical model. In terms of model interpretation and trend capturing on different resolutions, the HMM results were clearly inferior to the ones of the HHMM. The last section provided a brief outlook on a HHMM-based trading strategy, which profitably exploits the advantage that short-term and long-term trends can be distinguished.

Besides the success of capturing financial trends on two different resolutions, limitations of the approach emerged and some questions remained open. In the first place, it is highly debatable whether the independence assumptions of the model are fulfilled. Remember that we assumed conditional independence among the observations and that the fine-scale HMMs are indepen-

dent from the coarse-scale observations. These assumptions were made to achieve a likelihood function in simple product form. While the former assumption is justifiable in the case of returns, the latter one is critical because we used averages over the fine-scale data as coarse-scale observations. Choosing other indicators like trading volume would be no remedy as dependence across the layers is still conceivable. However, we can argue that the average log-returns indicate a change in the long-term trend which depends only on the current coarse-scale state and not on short-term fluctuations. Hence, independence may be practically plausible. As mentioned at the end of Section 3.2, it is possible to omit the coarse-scale observation layer which would prevent the dependence issue. Furthermore, data would enter the model only once which generally has to be preferred. However, it turned out that the truncated model fails to conveniently capture trends. This stands in contrast to [Leos-Barajas et al., 2017], where the approach was suitable for animal movement modelling. In the case of financial data however, coarse-scale indicators seem to be mandatory.

Furthermore, we should question the classification of financial market behaviour into a finite number of states. For animal movement data, discrete states that explain for example resting, foraging or travelling are fairly natural. The prospect of having similar proxies for financial data is desirable. However, classifying discretely here might in fact be unrealistic. Allowing gradual changes might bring us closer to reality. The HMM framework provides a solution to this scenario in terms of the state-space model, which possesses a continuous state space. Ideas like having discrete states on the coarse scale and the state-space model with infinite states on the fine scale or vice versa immediately come to mind. It would be interesting to find out whether such a model formulation would improve the results further, though we should expect the interpretation to become much harder. In contrast, discrete states are easier to interpret, despite being prone to over-interpretation.

In addition to state-space models, further extensions of the present work are conceivable. As stated in Section 3.2, any desired number of model hierarchies is theoretically feasible. Perhaps a third hierarchy capturing medium-term trends brings us closer to a complete description of financial market behaviour. Mind however that the model then becomes more instable and the numerical processing effort will increase rapidly. Remember that our final model with two hierarchies already possessed thirty-nine parameters. Choosing more layers or states would require an improvement to our approach of repeated numerical search for global likelihood maxima based on random starting points.

Another idea is to apply the model to stock market indices of other countries. Apparently, they share some patterns, see the graphic underneath the abstract. Trend changes in one index might indicate a trend change in others due to the given connectivity of today's financial markets.

Such information can in turn enter the field of trading strategies, which this thesis touched only briefly. Further research and testing is required on how the outlined ideas perform.

In summary, the HHMM has proven to be very suitable for investigation of financial market behaviour. It helps to understand how trends alternate and thereby increases the chance of making profitable investment decisions. Such guidance is valuable since following the mainstream will not pay off in the stock market. Instead, independent thinking is demanded, which this model can support. Last but not least, the underlying idea of the work is consistent with a famous principle of André Kostolany (loosely translated):

«Think long-term. Prices rise or fall over months and years. There is no need to let yourself be driven crazy by short-term fluctuations.»

Appendix A Proofs

Proof of equation (3.1):

$$\begin{aligned}
\mathcal{L}^{HMM}(\theta^{*(i)} | (X_{t,t^*}^*)_{t^*}) &= f^{*(i)}((X_{t,t^*}^*)_{t^*}) \\
&= \sum_{S_{t,1}^*, \dots, S_{t,T^*}^*=1}^{N^*} f^{*(i)}((X_{t,t^*}^*)_{t^*} | (S_{t,t^*}^*)_{t^*}) f^{*(i)}((S_{t,t^*}^*)_{t^*}) \quad (\text{law of total probability}) \\
&= \sum_{S_{t,1}^*, \dots, S_{t,T^*}^*=1}^{N^*} \left(\prod_{t^*=1}^{T^*} f^{*(i)}(X_{t,t^*}^* | S_{t,t^*}^*) \right) f^{*(i)}((S_{t,t^*}^*)_{t^*}) \quad (\text{conditional independence}) \\
&= \sum_{S_{t,1}^*, \dots, S_{t,T^*}^*=1}^{N^*} \left(\prod_{t^*=1}^{T^*} f^{*(i)}(X_{t,t^*}^* | S_{t,t^*}^*) \right) \left(f^{*(i)}(S_{t,1}^*) \prod_{t^*=2}^{T^*} f^{*(i)}(S_{t,t^*}^* | S_{t,t^*-1}^*) \right) \\
&\hspace{15em} (\text{law of total probability, Markov property}) \\
&= \sum_{S_{t,1}^*, \dots, S_{t,T^*}^*=1}^{N^*} \left(\prod_{t^*=1}^{T^*} f^{*(i, S_{t,t^*}^*)}(X_{t,t^*}^*) \right) \left(\delta_{S_{t,1}^*}^{*(i)} \prod_{t^*=2}^{T^*} \gamma_{S_{t,t^*-1}^* S_{t,t^*}^*}^{*(i)} \right) \quad (\text{notations})
\end{aligned}$$

Proof of equation (3.2):

$$\begin{aligned}
\alpha_{k,1}^{*(i)} &= f^{*(i)}(X_{t,1}^*, S_{t,1}^* = k) \\
&= f^{*(i)}(S_{t,1}^* = k) f^{*(i)}(X_{t,1}^* | S_{t,1}^* = k) \quad (\text{conditional probability}) \\
&= \delta_k^{*(i)} f^{*(i,k)}(X_{t,1}^*) \quad (\text{notations})
\end{aligned}$$

Proof of equation (3.3):

$$\begin{aligned}
\alpha_{k,t^*}^{*(i)} &= f^{*(i)}(X_{t,1}^*, \dots, X_{t,t^*}^*, S_{t,t^*}^* = k) \\
&= f^{*(i)}(X_{t,t^*}^* | X_{t,1}^*, \dots, X_{t,t^*-1}^*, S_{t,t^*}^* = k) f^{*(i)}(X_{t,1}^*, \dots, X_{t,t^*-1}^*, S_{t,t^*}^* = k) \quad (\text{conditional probability}) \\
&= f^{*(i)}(X_{t,t^*}^* | S_{t,t^*}^* = k) \sum_{j=1}^{N^*} f^{*(i)}(X_{t,1}^*, \dots, X_{t,t^*-1}^*, S_{t,t^*}^* = k | S_{t,t^*-1}^* = j) f^{*(i)}(S_{t,t^*-1}^* = j) \\
&\hspace{15em} (\text{conditional independence, law of total probability}) \\
&= f^{*(i)}(X_{t,t^*}^* | S_{t,t^*}^* = k) \sum_{j=1}^{N^*} f^{*(i)}(S_{t,t^*}^* = k | S_{t,t^*-1}^* = j) f^{*(i)}(X_{t,1}^*, \dots, X_{t,t^*-1}^*, S_{t,t^*-1}^* = j) \\
&\hspace{15em} (\text{conditional independence, conditional probability}) \\
&= f^{*(i,k)}(X_{t,t^*}^*) \sum_{j=1}^{N^*} \gamma_{jk}^{*(i)} \alpha_{j,t^*-1}^{*(i)} \quad (\text{notations})
\end{aligned}$$

Proof of equation (3.4):

$$\begin{aligned}
 \alpha_{i,1} &= f(X_1, (X_{1,t^*}^*)_{t^*}, S_t = i) \\
 &= f(S_t = i) f(X_1, (X_{1,t^*}^*)_{t^*} \mid S_t = i) && \text{(conditional probability)} \\
 &= f(S_t = i) f((X_{1,t^*}^*)_{t^*} \mid S_t = i) f(X_1 \mid S_t = i) && \text{(conditional independence)} \\
 &= \delta_i \mathcal{L}^{HMM}(\theta^{*(i)} \mid (X_{1,t^*}^*)_{t^*}) f^{(i)}(X_1) && \text{(notations)}
 \end{aligned}$$

Proof of equation (3.5):

$$\begin{aligned}
 \alpha_{i,t} &= f(X_1, \dots, X_t, (X_{1,t^*}^*)_{t^*}, \dots, (X_{t,t^*}^*)_{t^*}, S_t = i) \\
 &= f(X_t, (X_{t,t^*}^*)_{t^*} \mid X_1, \dots, X_{t-1}, (X_{1,t^*}^*)_{t^*}, \dots, (X_{t-1,t^*}^*)_{t^*}, S_t = i) \\
 &\quad \cdot f(X_1, \dots, X_{t-1}, (X_{1,t^*}^*)_{t^*}, \dots, (X_{t-1,t^*}^*)_{t^*}, S_t = i) && \text{(conditional probability)} \\
 &= f(X_t, (X_{t,t^*}^*)_{t^*} \mid S_t = i) \\
 &\quad \cdot \sum_{j=1}^N f(X_1, \dots, X_{t-1}, (X_{1,t^*}^*)_{t^*}, \dots, (X_{t-1,t^*}^*)_{t^*}, S_t = i \mid S_{t-1} = j) f(S_{t-1} = j) \\
 &\hspace{15em} \text{(conditional independence, law of total probability)} \\
 &= f((X_{t,t^*}^*)_{t^*} \mid S_t = i) f(X_t \mid S_t = i) \\
 &\quad \cdot \sum_{j=1}^N f(S_t = i \mid S_{t-1} = j) f(X_1, \dots, X_{t-1}, (X_{1,t^*}^*)_{t^*}, \dots, (X_{t-1,t^*}^*)_{t^*}, S_{t-1} = j) \\
 &\hspace{15em} \text{(conditional independence, conditional probability)} \\
 &= \mathcal{L}^{HMM}(\theta^{*(i)} \mid (X_{t,t^*}^*)_{t^*}) f^{(i)}(X_t) \sum_{j=1}^N \gamma_{ji} \alpha_{j,t-1} && \text{(notations)}
 \end{aligned}$$

Proof of equation (3.6):

$$\begin{aligned}
 \phi_{k,1}^{*(i)} &= \log[f^{*(i)}(X_{t,1}^*, S_{t,1}^* = k)] \\
 &= \log[f^{*(i)}(S_{t,1}^* = k) f^{*(i)}(X_{t,1}^* \mid S_{t,1}^* = k)] && \text{(conditional probability)} \\
 &= \log[\delta_k^{*(i)}] + \log[f^{*(i,k)}(X_{t,1}^*)] && \text{(logarithmic identity, notations)}
 \end{aligned}$$

Proof of equation (3.7):

$$\begin{aligned}
 \phi_{k,t^*}^{*(i)} &= \log[f^{*(i)}(X_{t,1}^*, \dots, X_{t,t^*}^*, S_{t,t^*}^* = k)] \\
 &= \log[f^{*(i)}(X_{t,t^*}^* \mid X_{t,1}^*, \dots, X_{t,t^*-1}^*, S_{t,t^*}^* = k) f^{*(i)}(X_{t,1}^*, \dots, X_{t,t^*-1}^*, S_{t,t^*}^* = k)] && \text{(conditional probability)} \\
 &= \log[f^{*(i)}(X_{t,t^*}^* \mid S_{t,t^*}^* = k) \sum_{j=1}^{N^*} f^{*(i)}(X_{t,1}^*, \dots, X_{t,t^*-1}^*, S_{t,t^*}^* = k \mid S_{t,t^*-1}^* = j) f^{*(i)}(S_{t,t^*-1}^* = j)] \\
 &\hspace{15em} \text{(conditional independence, law of total probability)}
 \end{aligned}$$

$$\begin{aligned}
 &= \log[f^{*(i)}(X_{t,t^*}^* | S_{t,t^*}^* = k) \sum_{j=1}^{N^*} f^{*(i)}(S_{t,t^*}^* = k | S_{t,t^*-1}^* = j) f^{*(i)}(X_{t,1}^*, \dots, X_{t,t^*-1}^*, S_{t,t^*-1}^* = j)] \\
 &\hspace{20em} \text{(conditional independence, conditional probability)} \\
 &= \log[f^{*(i,k)}(X_{t,t^*}^*)] + \log \left[\sum_{j=1}^{N^*} \gamma_{jk}^{*(i)} \exp[\phi_{j,t^*-1}^{*(i)} - c_{t^*-1}] \right] + c_{t^*-1} \\
 &\hspace{20em} \text{(logarithmic and exponential identities, notations)}
 \end{aligned}$$

Proof of equation (3.8):

$$\begin{aligned}
 \phi_{i,1} &= \log[f(X_1, (X_{1,t^*}^*)_{t^*}, S_t = i)] \\
 &= \log[f(S_t = i) f(X_1, (X_{1,t^*}^*)_{t^*} | S_t = i)] && \text{(conditional probability)} \\
 &= \log[f(S_t = i) f((X_{1,t^*}^*)_{t^*} | S_t = i) f(X_1 | S_t = i)] && \text{(conditional independence)} \\
 &= \log[\delta_i] + \log[\mathcal{L}^{HMM}(\theta^{*(i)} | (X_{1,t^*}^*)_{t^*})] + \log[f^{(i)}(X_1)] && \text{(logarithmic identity, notations)}
 \end{aligned}$$

Proof of equation (3.9):

$$\begin{aligned}
 \phi_{i,t} &= \log[f(X_1, \dots, X_t, (X_{1,t^*}^*)_{t^*}, \dots, (X_{t,t^*}^*)_{t^*}, S_t = i)] \\
 &= \log[f(X_t, (X_{t,t^*}^*)_{t^*} | X_1, \dots, X_{t-1}, (X_{1,t^*}^*)_{t^*}, \dots, (X_{t-1,t^*}^*)_{t^*}, S_t = i) \\
 &\quad \cdot f(X_1, \dots, X_{t-1}, (X_{1,t^*}^*)_{t^*}, \dots, (X_{t-1,t^*}^*)_{t^*}, S_t = i)] && \text{(conditional probability)} \\
 &= \log[f(X_t, (X_{t,t^*}^*)_{t^*} | S_t = i) \\
 &\quad \cdot \sum_{j=1}^N f(X_1, \dots, X_{t-1}, (X_{1,t^*}^*)_{t^*}, \dots, (X_{t-1,t^*}^*)_{t^*}, S_t = i | S_{t-1} = j) f(S_{t-1} = j)] \\
 &\hspace{20em} \text{(conditional independence, law of total probability)} \\
 &= \log[f((X_{t,t^*}^*)_{t^*} | S_t = i) f(X_t | S_t = i) \\
 &\quad \cdot \sum_{j=1}^N f(S_t = i | S_{t-1} = j) f(X_1, \dots, X_{t-1}, (X_{1,t^*}^*)_{t^*}, \dots, (X_{t-1,t^*}^*)_{t^*}, S_{t-1} = j)] \\
 &\hspace{20em} \text{(conditional independence, conditional probability)} \\
 &= \log \mathcal{L}^{HMM}(\theta^{*(i)} | (X_{t,t^*}^*)_{t^*}) + \log[f^{(i)}(X_t)] + \log \left[\sum_{j=1}^N \gamma_{ji} \exp[\phi_{j,t-1} - c_{t-1}] \right] + c_{t-1} \\
 &\hspace{20em} \text{(logarithmic and exponential identities, notations)}
 \end{aligned}$$

Proof of equation (3.11):

$$\begin{aligned}
 \xi_{i,1} &= f(S_1 = 1, X_1) \\
 &= f(S_1 = i) f(X_1 | S_1 = i) && \text{(conditional probability)} \\
 &= \delta_i f^{(i)}(X_1) && \text{(notations)}
 \end{aligned}$$

Proof of equation (3.12):

$$\begin{aligned}
 \xi_{i,t} &= \max_{S_1, \dots, S_{t-1}} f(S_1, \dots, S_{t-1}, S_t = i, X_1, \dots, X_t) \\
 &= \max_{S_1, \dots, S_{t-1}} f(X_t | S_1, \dots, S_{t-1}, S_t = i, X_1, \dots, X_{t-1}) f(S_1, \dots, S_{t-1}, S_t = i, X_1, \dots, X_{t-1}) \\
 &\hspace{25em} \text{(conditional probability)} \\
 &= \max_{S_1, \dots, S_{t-1}} f(X_t | S_t = i) f(S_t = i | S_1, \dots, S_{t-1}, X_1, \dots, X_{t-1}) f(S_1, \dots, S_{t-1}, X_1, \dots, X_{t-1}) \\
 &\hspace{15em} \text{(conditional independence, conditional probability)} \\
 &= \max_{S_1, \dots, S_{t-1}} f(X_t | S_t = i) f(S_t = i | S_{t-1}) f(S_1, \dots, S_{t-1}, X_1, \dots, X_{t-1}) \hspace{2em} \text{(Markov property)} \\
 &= \max_{S_1, \dots, S_{t-2}, j} f(X_t | S_t = i) f(S_t = i | S_{t-1} = j) f(S_1, \dots, S_{t-2}, S_{t-1} = j, X_1, \dots, X_{t-1}) \\
 &= \max_j (\xi_{j,t-1} \gamma_{ji}) f^{(i)}(X_t) \hspace{15em} \text{(notations)}
 \end{aligned}$$

Proof of statement (3.13): Let $U_t = F_{X_t}(X_t)$ and assume that F_{X_t} is invertible. Then:

$$F_{U_t}(u) = \Pr(U_t \leq u) = \Pr(F_{X_t}(X_t) \leq u) = \Pr(X_t \leq F_{X_t}^{-1}(u)) = F_{X_t}(F_{X_t}^{-1}(u)) = u,$$

which implies that $U_t = F_{X_t}(X_t) \sim \mathcal{U}[0, 1]$. Let now $Z_t = \Phi^{-1}(U_t)$. Then:

$$F_{Z_t}(z) = \Pr(Z_t \leq z) = \Pr(\Phi^{-1}(U_t) \leq z) = \Pr(U_t \leq \Phi(z)) = F_{U_t}(\Phi(z)) = \Phi(z),$$

which implies that $Z_t = \Phi^{-1}(U_t) = \Phi^{-1}(F_{X_t}(X_t)) \sim \mathcal{N}(0, 1)$.

Proof of statement (4.1): If $X_t \sim \sigma_{S_t} \mathcal{T}_{\nu_{S_t}} + \mu_{S_t}$, then:

$$\begin{aligned}
 F_{X_t}(x) &= \Pr(X_t \leq x) = \Pr\left(\frac{X_t - \mu_{S_t}}{\sigma_{S_t}} \leq \frac{x - \mu_{S_t}}{\sigma_{S_t}}\right) = F_{\mathcal{T}_{\nu_{S_t}}}\left(\frac{x - \mu_{S_t}}{\sigma_{S_t}}\right), \\
 f_{X_t}(x) &= \frac{d}{dx} F_{X_t}(x) = \frac{d}{dx} F_{\mathcal{T}_{\nu_{S_t}}}\left(\frac{x - \mu_{S_t}}{\sigma_{S_t}}\right) = \sigma_{S_t}^{-1} f_{\mathcal{T}_{\nu_{S_t}}}\left(\frac{x - \mu_{S_t}}{\sigma_{S_t}}\right).
 \end{aligned}$$

Appendix B Parameter Estimates

Estimated parameters of the HHMM with $T^ = 30$, $N = 3$, $N^* = 2$ and normal distributions¹:*

$$\begin{aligned}
 \Gamma &= \begin{pmatrix} 0,9326 & 0,0376 & 0,0298 \\ 0,0764 & 0,8418 & 0,0818 \\ 0,0000 & 0,3310 & 0,6690 \end{pmatrix} & f^{(1)} &= \mathcal{N}(6,08 \cdot 10^{-4}; 1,31 \cdot 10^{-3}) \\
 \Gamma^{*(1)} &= \begin{pmatrix} 0,9832 & 0,0168 \\ 0,0070 & 0,9930 \end{pmatrix} & f^{(2)} &= \mathcal{N}(1,85 \cdot 10^{-4}; 1,93 \cdot 10^{-3}) \\
 \Gamma^{*(2)} &= \begin{pmatrix} 0,8424 & 0,1576 \\ 0,0771 & 0,9229 \end{pmatrix} & f^{(3)} &= \mathcal{N}(-21,28 \cdot 10^{-4}; 5,17 \cdot 10^{-3}) \\
 \Gamma^{*(3)} &= \begin{pmatrix} 0,9817 & 0,0183 \\ 0,0143 & 0,9857 \end{pmatrix} & f^{*(1,1)} &= \mathcal{N}(17,46 \cdot 10^{-4}; 4,45 \cdot 10^{-3}) \\
 \pi &= (0,4566 \quad 0,4028 \quad 0,1406) & f^{*(1,2)} &= \mathcal{N}(0,43 \cdot 10^{-4}; 10,16 \cdot 10^{-3}) \\
 \pi^{*(1)} &= (0,2941 \quad 0,7059) & f^{*(2,1)} &= \mathcal{N}(12,02 \cdot 10^{-4}; 9,59 \cdot 10^{-3}) \\
 \pi^{*(2)} &= (0,3285 \quad 0,6715) & f^{*(2,2)} &= \mathcal{N}(-2,30 \cdot 10^{-4}; 15,08 \cdot 10^{-3}) \\
 \pi^{*(3)} &= (0,4387 \quad 0,5613) & f^{*(3,1)} &= \mathcal{N}(-14,50 \cdot 10^{-4}; 24,62 \cdot 10^{-3}) \\
 & & f^{*(3,2)} &= \mathcal{N}(-57,16 \cdot 10^{-4}; 35,78 \cdot 10^{-3})
 \end{aligned}$$

Estimated parameters of the HHMM with $T^ = 30$, $N = 3$, $N^* = 3$ and normal distributions:*

$$\begin{aligned}
 \Gamma &= \begin{pmatrix} 0,8991 & 0,0609 & 0,0400 \\ 0,1345 & 0,7331 & 0,1324 \\ 0,0766 & 0,2851 & 0,6383 \end{pmatrix} & f^{(1)} &= \mathcal{N}(5,70 \cdot 10^{-4}; 1,36 \cdot 10^{-3}) \\
 \Gamma^{*(1)} &= \begin{pmatrix} 0,0000 & 1,0000 & 0,0000 \\ 0,1017 & 0,7320 & 0,1663 \\ 0,1073 & 0,0000 & 0,8927 \end{pmatrix} & f^{(2)} &= \mathcal{N}(1,85 \cdot 10^{-4}; 1,95 \cdot 10^{-3}) \\
 \Gamma^{*(2)} &= \begin{pmatrix} 0,0000 & 1,0000 & 0,0000 \\ 0,1895 & 0,6367 & 0,1738 \\ 0,2102 & 0,7898 & 0,0000 \end{pmatrix} & f^{(3)} &= \mathcal{N}(-17,97 \cdot 10^{-4}; 4,83 \cdot 10^{-3}) \\
 \Gamma^{*(3)} &= \begin{pmatrix} 0,0000 & 0,8375 & 0,1625 \\ 0,0000 & 0,9823 & 0,0177 \\ 0,0372 & 0,0000 & 0,9628 \end{pmatrix} & f^{*(1,1)} &= \mathcal{N}(108,60 \cdot 10^{-4}; 5,60 \cdot 10^{-3}) \\
 \pi &= (0,5298 \quad 0,3013 \quad 0,1689) & f^{*(1,2)} &= \mathcal{N}(8,37 \cdot 10^{-4}; 4,11 \cdot 10^{-3}) \\
 \pi^{*(1)} &= (0,0951 \quad 0,3549 \quad 0,5500) & f^{*(1,3)} &= \mathcal{N}(-14,10 \cdot 10^{-4}; 10,34 \cdot 10^{-3}) \\
 \pi^{*(2)} &= (0,1615 \quad 0,7144 \quad 0,1241) & f^{*(2,1)} &= \mathcal{N}(169,94 \cdot 10^{-4}; 9,81 \cdot 10^{-3}) \\
 \pi^{*(3)} &= (0,0133 \quad 0,6292 \quad 0,3575) & f^{*(2,2)} &= \mathcal{N}(0,06 \cdot 10^{-4}; 9,99 \cdot 10^{-3}) \\
 & & f^{*(2,3)} &= \mathcal{N}(-206,21 \cdot 10^{-4}; 8,87 \cdot 10^{-3}) \\
 & & f^{*(3,1)} &= \mathcal{N}(587,74 \cdot 10^{-4}; 9,07 \cdot 10^{-3}) \\
 & & f^{*(3,2)} &= \mathcal{N}(1,91 \cdot 10^{-4}; 19,85 \cdot 10^{-3}) \\
 & & f^{*(3,3)} &= \mathcal{N}(-74,76 \cdot 10^{-4}; 29,93 \cdot 10^{-3})
 \end{aligned}$$

1. Remember that $\mathcal{N}(\mu; \sigma)$ is the notation of the normal distribution with mean μ and standard deviation σ .

Estimated parameters of the HHMM with $T^* = 30$, $N = 3$, $N^* = 3$ and t -distributions²:

$$\begin{aligned}
 \Gamma &= \begin{pmatrix} 0,8773 & 0,1132 & 0,0095 \\ 0,1872 & 0,7592 & 0,0536 \\ 0,0025 & 0,2568 & 0,7407 \end{pmatrix} & f^{(1)} &= \mathcal{T}(6,13 \cdot 10^{-4}; 1,33 \cdot 10^{-3}; 29) \\
 \Gamma^{*(1)} &= \begin{pmatrix} 0,0157 & 0,5848 & 0,3995 \\ 0,0000 & 0,8529 & 0,1471 \\ 0,1939 & 0,0008 & 0,8053 \end{pmatrix} & f^{(2)} &= \mathcal{T}(-0,66 \cdot 10^{-4}; 2,10 \cdot 10^{-3}; 24) \\
 \Gamma^{*(2)} &= \begin{pmatrix} 0,8499 & 0,1335 & 0,0166 \\ 0,0086 & 0,9780 & 0,0134 \\ 0,0133 & 0,8894 & 0,0973 \end{pmatrix} & f^{(3)} &= \mathcal{T}(-20,14 \cdot 10^{-4}; 5,36 \cdot 10^{-3}; 26) \\
 \Gamma^{*(2)} &= \begin{pmatrix} 0,0973 & 0,8894 & 0,0133 \\ 0,0134 & 0,9780 & 0,0086 \\ 0,0166 & 0,1335 & 0,8499 \end{pmatrix} & f^{*(1,1)} &= \mathcal{T}(129,48 \cdot 10^{-4}; 5,00 \cdot 10^{-3}; 10) \\
 \Gamma^{*(3)} &= \begin{pmatrix} 0,0435 & 0,9515 & 0,0050 \\ 0,0006 & 0,9865 & 0,0129 \\ 0,6923 & 0,0091 & 0,2986 \end{pmatrix} & f^{*(1,2)} &= \mathcal{T}(16,35 \cdot 10^{-4}; 4,52 \cdot 10^{-3}; 11) \\
 \pi &= \begin{pmatrix} 0,5480 & 0,3579 & 0,0941 \end{pmatrix} & f^{*(1,3)} &= \mathcal{T}(-26,48 \cdot 10^{-4}; 8,65 \cdot 10^{-3}; 11) \\
 \pi^{*(1)} &= \begin{pmatrix} 0,0992 & 0,3972 & 0,5036 \end{pmatrix} & f^{*(2,1)} &= \mathcal{T}(17,50 \cdot 10^{-4}; 0,23 \cdot 10^{-3}; 2) \\
 \pi^{*(2)} &= \begin{pmatrix} 0,0148 & 0,9306 & 0,0546 \end{pmatrix} & f^{*(2,2)} &= \mathcal{T}(1,93 \cdot 10^{-4}; 13,62 \cdot 10^{-3}; 7) \\
 \pi^{*(3)} &= \begin{pmatrix} 0,0136 & 0,9685 & 0,0179 \end{pmatrix} & f^{*(2,3)} &= \mathcal{T}(-40,65 \cdot 10^{-4}; 21,09 \cdot 10^{-3}; 8) \\
 & & f^{*(3,1)} &= \mathcal{T}(334,34 \cdot 10^{-4}; 0,85 \cdot 10^{-3}; 2) \\
 & & f^{*(3,2)} &= \mathcal{T}(-17,53 \cdot 10^{-4}; 24,76 \cdot 10^{-3}; 3) \\
 & & f^{*(3,3)} &= \mathcal{T}(-523,16 \cdot 10^{-4}; 5,53 \cdot 10^{-3}; 7)
 \end{aligned}$$

Estimated parameters of the HMM with $N = 4$ and t -distributions:

$$\begin{aligned}
 \Gamma &= \begin{pmatrix} 0,9555 & 0,0445 & 0,0000 & 0,0000 \\ 0,0291 & 0,9603 & 0,0106 & 0,0000 \\ 0,0000 & 0,0112 & 0,9833 & 0,0055 \\ 0,0000 & 0,0000 & 0,0164 & 0,9836 \end{pmatrix} & f^{(1)} &= \mathcal{T}(14,92 \cdot 10^{-4}; 5,50 \cdot 10^{-3}; 10) \\
 \pi &= \begin{pmatrix} 0,2241 & 0,3427 & 0,3244 & 0,1088 \end{pmatrix} & f^{(2)} &= \mathcal{T}(1,80 \cdot 10^{-4}; 9,72 \cdot 10^{-3}; 14) \\
 & & f^{(3)} &= \mathcal{T}(-0,49 \cdot 10^{-4}; 14,19 \cdot 10^{-3}; 15) \\
 & & f^{(4)} &= \mathcal{T}(-27,85 \cdot 10^{-4}; 26,87 \cdot 10^{-3}; 13)
 \end{aligned}$$

Estimated parameters of the HMM with $N = 5$ and t -distributions:

$$\begin{aligned}
 \Gamma &= \begin{pmatrix} 0,0000 & 0,5169 & 0,0000 & 0,0000 & 0,4831 \\ 0,0000 & 0,8923 & 0,0000 & 0,0000 & 0,1077 \\ 0,0101 & 0,0000 & 0,9847 & 0,0052 & 0,0000 \\ 0,0000 & 0,0000 & 0,0162 & 0,9838 & 0,0000 \\ 0,1982 & 0,0000 & 0,0145 & 0,0000 & 0,7873 \end{pmatrix} & f^{(1)} &= \mathcal{T}(129,68 \cdot 10^{-4}; 5,25 \cdot 10^{-3}; 14) \\
 \pi &= \begin{pmatrix} 0,0515 & 0,2472 & 0,3476 & 0,1116 & 0,2421 \end{pmatrix} & f^{(2)} &= \mathcal{T}(15,99 \cdot 10^{-4}; 4,76 \cdot 10^{-3}; 7) \\
 & & f^{(3)} &= \mathcal{T}(-0,38 \cdot 10^{-4}; 13,97 \cdot 10^{-3}; 16) \\
 & & f^{(4)} &= \mathcal{T}(-26,90 \cdot 10^{-4}; 25,77 \cdot 10^{-3}; 8) \\
 & & f^{(5)} &= \mathcal{T}(-27,81 \cdot 10^{-4}; 8,62 \cdot 10^{-3}; 11)
 \end{aligned}$$

2. Remember that $\mathcal{T}(\mu; \sigma; \nu)$ is the notation of the scaled t -distribution $\sigma\mathcal{T}_\nu + \mu$ with ν degrees of freedom.

Bibliography

- [Adam et al., 2019] Adam, T., Griffiths, C. A., Leos-Barajas, V., Meese, E. N., Lowe, C. G., Blackwell, P. G., Righton, D., and Langrock, R. (2019). *Joint modelling of multi-scale animal movement data using hierarchical hidden Markov models*. *Methods in Ecology and Evolution*, forthcoming.
- [Akaike, 1974] Akaike, H. (1974). *A New Look at the Statistical Model Identification*. *IEEE Transactions on Automatic Control*, volume 19.
- [Black and Scholes, 1973] Black, F. and Scholes, M. (1973). *The Pricing of Options and Corporate Liabilities*. *Journal of Political Economy*, volume 81.
- [Bulla and Bulla, 2006] Bulla, J. and Bulla, I. (2006). *Stylized facts of financial time series and hidden semi-Markov models*. *Computational Statistics & Data Analysis*, volume 51.
- [Fine et al., 1998] Fine, S., Singer, Y., and Tishby, N. (1998). *The Hierarchical Hidden Markov Model: Analysis and Applications*. *Machine Learning*, volume 32.
- [Harris and Küçüközmen, 2001] Harris, R. D. F. and Küçüközmen, C. C. (2001). *The Empirical Distribution of UK and US Stock Returns*. *Journal of Business Finance & Acc.*, volume 28.
- [Hassan and Nath, 2005] Hassan, M. R. and Nath, B. (2005). *Stock Market Forecasting Using Hidden Markov Model: A New Approach*.
- [Janßen and Rudolph, 1992] Janßen, B. and Rudolph, B. (1992). *Der Deutsche Aktienindex DAX*. Knapp Verlag.
- [Kostolany, 2015] Kostolany, A. (2015). *Die Kunst über Geld nachzudenken*. Ullstein Verlag, fifth edition.
- [Leos-Barajas et al., 2017] Leos-Barajas, V., Gangloff, E. J., Adam, T., Langrock, R., van Beest, F. M., Nabe-Nielsen, J., and Morales, J. M. (2017). *Multi-scale modeling of animal movement and general behavior data using hidden Markov models with hierarchical structures*. *Journal of Agricultural, Biological and Environmental Statistics*, volume 22.
- [Levin et al., 2017] Levin, D. A., Peres, Y., and Wilmer, E. L. (2017). *Markov Chains and Mixing Times*. American Mathematical Society, second edition.

- [Lihn, 2017] Lihn, S. H.-T. (2017). *Hidden Markov Model for Financial Time Series and Its Application to S&P 500 Index*. Quantitative Finance, forthcoming.
- [Nguyen, 2018] Nguyen, N. (2018). *Hidden Markov Model for Stock Trading*. International Journal of Financial Studies, volume 6.
- [Norris, 1997] Norris, J. R. (1997). *Markov Chains*. Cambridge University Press.
- [Pagan and Schwert, 1990] Pagan, A. and Schwert, G. (1990). *Alternative models for conditional stock volatility*. Journal of Econometrics, volume 45.
- [Pohle et al., 2017] Pohle, J., Langrock, R., van Beest, F. M., and Schmidt, N. M. (2017). *Selecting the Number of States in Hidden Markov Models: Pragmatic Solutions Illustrated Using Animal Movement*. Journal of Agricultural, Biological and Environmental Statistics, volume 22.
- [Praetz, 1972] Praetz, P. D. (1972). *The Distribution of Share Price Changes*. The Journal of Business, volume 45.
- [Rydén et al., 1998] Rydén, T., Teräsvirta, T., and Åsbrink, S. (1998). *Stylized facts of daily return series and the hidden Markov model*. Journal of Applied Econometrics, volume 13.
- [Schwarz, 1978] Schwarz, G. (1978). *Estimating the Dimension of a Model*. The Annals of Statistics, volume 6.
- [Zucchini et al., 2016] Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R*. CRC Press, second edition.