

# Efficient Risk Factor Allocation with Regime Based Models

Oscar Axelsson

June 2017

## Abstract

It is widely accepted that financial market behaviour is characterized by periodicity. However, in academia and practice financial markets are often modeled as time consistent, resulting in static investment strategies that are assumed to be efficient. This has great implications on investors' allocation decisions and in valuation of risks and performances.

In this thesis we consider different risk factors and their behaviour over time. By using a hidden market model, which allow the risk factors to jump between 2 different regimes, we are able to identify distinctly different behaviour for all the risk factors in the different regimes. We are then able to design different investment strategies and use an online implementation of the model that yield significant better returns than equivalent static investment strategies.

Our results give further proof of the benefits of using regime based models when working with portfolio decisions. We also suggest a shrinkage approach to the covariance matrix estimations that gives increased stability to the model and makes it highly applicable in practice.

**Keywords:** Risk Factors, Hidden Markov Models, Regime Models, Asset Allocation, Efficient Frontier, Shrinkage Factor, Baum-Welch Algorithm, EM algorithm.

## Acknowledgements

The research included in this thesis could not have been performed if not for the assistance, and generosity of Bodenholm Capital. Bodenholm has shown great patient and support throughout the process. I extend my sincere thanks to Per Johansson, Erik Karlsson, and John Lindberg for my stimulating and developing time at Bodenholm Capital.

I would also like to extend my gratitude to my supervisor, Professor Erik Lindström at the department of Mathematical Statistics, Lund University. Erik has been extremely helpful and has provided me with many stimulating discussions which has been crucial for the development of this thesis. Also Erik has shown great patient throughout the process.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Objective of the Thesis . . . . .	3
1.3	Thesis Outline . . . . .	3
<b>2</b>	<b>Theory and Concepts</b>	<b>5</b>
2.1	Financial Theory . . . . .	5
2.1.1	Multi-factor models . . . . .	5
2.1.2	Regime Based Asset allocation . . . . .	7
2.2	Model Concepts and Theory . . . . .	8
2.2.1	The hidden Markov model . . . . .	8
2.2.2	Viterbi's path . . . . .	9
2.2.3	Baum-Welch algorithm . . . . .	10
2.2.4	Shrinkage of covariance matrices . . . . .	12
2.2.5	Frobenius norm . . . . .	13
2.2.6	Skewness and Kurtosis . . . . .	13
<b>3</b>	<b>Data</b>	<b>15</b>
3.1	Factors construction . . . . .	15
3.2	Data Analysis . . . . .	17
<b>4</b>	<b>Modeling</b>	<b>20</b>
4.1	3 factor shrinkage model . . . . .	21
4.2	Efficient Frontier Static Model . . . . .	24

<b>5</b>	<b>Time varying parameters and regime decision</b>	<b>26</b>
<b>6</b>	<b>Asset Allocation</b>	<b>29</b>
6.1	Time Varying Efficient Frontier . . . . .	29
6.2	Results . . . . .	30
<b>7</b>	<b>Discussion and Conclusions</b>	<b>35</b>
7.1	Model approach . . . . .	35
7.2	Online application . . . . .	35
7.3	Student's t-distribution . . . . .	36
7.4	t-distribution in filter probabilities . . . . .	37
7.5	In Summary . . . . .	37
<b>8</b>	<b>Appendices</b>	<b>42</b>
8.1	QQ-plots . . . . .	42
8.2	Shrinkage weights . . . . .	44
8.3	Regime Performance . . . . .	45

# 1 Introduction

## 1.1 Background

One of the key issues for any investor is how to allocate resources between different assets. In finance using mean-variance analysis is one of the most common mathematical frameworks. The framework, often mentioned as modern portfolio theory, had its first contributor in Markowitz (1952) and Markowitz (1959). The core of the theory is the relationship between risk in the form of variance, and expected return for assets in a finite universe of tradable assets. Considering the covariance matrix and the expected returns Markowitz proved the theory of *the efficient frontier*, with an optimal portfolio construction (in the sense of variance), for every desired expected return. Figure 1 shows a typical example of the efficient frontier. Along the x-axis is the standard deviation of the assets, and along the y-axis is the expected return of the assets. Each red dot denote an asset from the universe under consideration, and the blue line is the efficient frontier. As we can note all assets is located below the blue line, meaning that the efficient frontier is the optimal portfolio construction with respect to variance with a fixed expected return.

The mean-variance efficiency concept is still one of the most important cornerstones in modern portfolio theory, and put emphasis on the importance of diversification in portfolios. A common extension of the model is to include an additional assumption of the existence of a risk free rate that investors can loan and invest in. The consequence of this assumption is that the investors' preference for risk is no longer relevant for the construction of the optimal portfolio. This is called the separation theory and gives rise to the optimal portfolio being one single tangency portfolio. The tangency portfolio is the portfolio on the efficient frontier were the expected return minus the risk free rate and divided by the standard deviation is maximal. If the investor desires to have an increased or decreased level of risk she can simply choose to invest or loan at the risk free rate, i.e. use the concept of leverage or deleverage. This theory reduce the investment decision to a linear problem (see Elton and Gruber (1997) et al.).

Based on the concept of a tangency portfolio William F. Sharpe (1964), Lintner (1965) et al. developed a single linear model explaining the relation between an asset's expected return and the expected return of the tangency portfolio. It has the form of,

$$r_i(t) - r_f(t) = \alpha_i + \beta_i \cdot (r_m(t) - r_f(t)) + \epsilon_i \quad \epsilon_i \sim \mathcal{N}(0, \sigma_i). \quad (1)$$

The model is called the *Capital Asset Pricing Model*, from here on referred to as *CAPM*, and explains the cross-return of any asset as a linear relation between the return of the tangency portfolio and the asset's return.  $r_i(t)$  is the return of the asset, and  $r_m(t)$  is the return of the

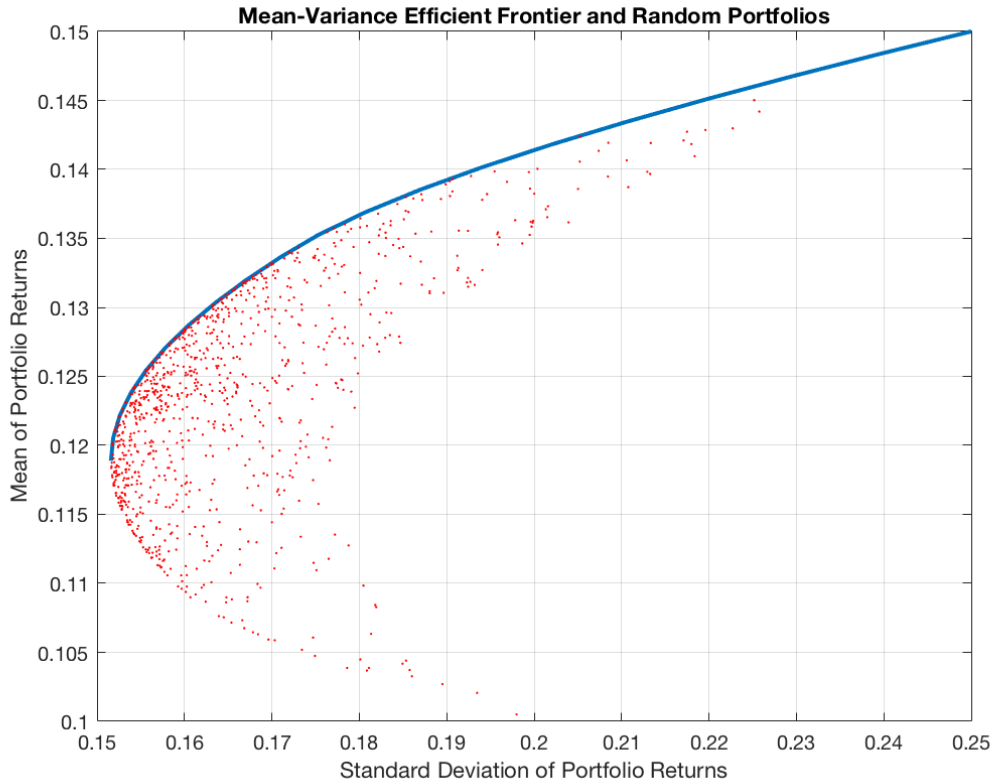


Figure 1: A simulated example of the efficient frontier. The blue line is the frontier and each red dot is an asset. The figure shows a typical relation between expected risk and expected return that can be observed in financial markets

tangency portfolio, which is often referred to as the market portfolio.  $r_f(t)$  is the risk free rate which investors can borrow and invest in,  $\beta_i$  is the constant that gives the sensitivity relation between the asset and the market.  $\alpha_i$  is the unique expected return that can not be explained by market exposure, and  $\epsilon_i$  is a residual term, often assumed to be normally distributed with expected value 0.

*CAPM* has had a huge impact on the financial markets and the valuation of risk, and is a common tool in performance evaluation of investment strategies or in evaluation of active investors with the alpha term being an indicator of an investor's skill. The Sharpe ratio from William F Sharpe (1966) is arguably the most common used measure for risk adjusted return. It is directly deduced from *CAPM*, and makes it possible to compare the expected return of assets that carries different amount of risk. It is defined as,

$$S_i(t) = \frac{\mathbb{E}(r_i(t) - r_f(t))}{\sqrt{\mathbb{V}(r_i(t))}} \quad (2)$$

with  $\mathbb{V}(r_i(t))$  being the variance of the returns of asset  $i$ . If we consider the efficient frontier in Figure 1 it is not hard to convince oneself that the tangency portfolio, which is the tangency point on the *efficient frontier* which crosses the y-axis at the risk free rate, is the same point as the portfolio with the highest Sharpe ratio,  $S_i(t)$ .

## 1.2 Objective of the Thesis

The main objective of this thesis can be divided into two parts. The first is to find a suitable regime based model for modeling market risk factors which burst from Fama and French (1992) and Carhart (1997). Our target will be a two state model, allowing the market to either be in a low volatility regime or a high volatility regime. For a model to be suitable it has to demonstrate regime stability, meaning that it does not jump back and forth between regimes too frequently. It also has to identify regimes with distinguished characteristics. Hence, the behaviour of the factor returns have to differ significantly between the regimes for the model approach to make sense.

The second part of the thesis is to identify and evaluate an online application of the model which can be used in practice by investors. As basis for the online regime decisions the filter probabilities for the different states are used. Furthermore, for the approach to be applicable in any time period the parameters are updated daily based on a set of fixed length of previous observations. Thus, the main target for this part of the thesis is to find a suitable online regime decision approach that uses the time varying parameters in order to take optimal asset allocation decisions, i.e. finding *the efficient frontier*.

The thesis will use daily returns of the risk factor from Fama and French (1992) and Carhart (1997), and only consider the US based factors. The regime based model will be in the shape of a Hidden Markov Model.

## 1.3 Thesis Outline

The thesis outline is as follows,

**Section 2** gives first a theoretical background on the financial theory and concept that is central for the approach used in the thesis. Then the section gives a mathematical description of the Hidden Markov Model, and an extensive review of the algorithms used in parameter estimation and probability calculations.

**Section 3** gives a thorough review of the data used in the thesis, and how the different risk factors are constructed. The section also provides a short analysis of the data with emphasis



on correlation and distribution of the factor returns.

**Section 4** presents the procedure and results of the process of finding a suitable model when the whole data set is considered. The aim is to find a model which satisfactorily fulfill the requirements for a regime based model stated in the objective of the thesis.

**Section 5** presents the online application of the model, and time varying parameter estimation.

**Section 6** presents the results when the online model is applied in practice, and we evaluate the performance against static investment strategies.

**Section 7** discuss our results and highlights strengths and weaknesses of our model.

## 2 Theory and Concepts

### 2.1 Financial Theory

The central concepts of financial theory in this thesis can be divided into two parts. The first is the theory around multi-factor models which is direct extension of the *CAPM* framework. The second part is the research around regime based modeling of financial markets. This section will provide a comprehensive summary of the findings from published research around regime based models of the financial market.

#### 2.1.1 Multi-factor models

Fama and French (2004) gives an extensive empirical evaluation of *CAPM*'s performance, and concludes that their findings give little support for the application of *CAPM* that is often seen in academia and in practice. Fama and French argue that even though there exists a clear relation between the market exposure of an asset and its expected return, *CAPM* consistently overestimate the returns of companies with high beta, and underestimates the returns of companies with low beta.

The shortcomings of *CAPM*, and *CAPM*'s wide use in practice, have encouraged researchers to search for simple investment criteria that overtime yield alpha in the *CAPM* framework. Suddhasatwa (1977) investigates the returns of a portfolio constructed of stocks and their *P/E* ratios (price divided by earnings). He concluded that the risk adjusted return of stocks with a low *P/E* ratio outperformed stocks with a higher *P/E* ratio. In a similar way Banz (1981) proves that smaller companies' stocks outperform bigger companies'. While Suddhasatwa (1977) interprets the results as an indication of the market being inefficient. Banz (1981), however, argues that this is an excess conclusion, and suggests instead that *CAPM* fails to capture all risk dimensions that are not diversifiable and that investors are concerned with.

In line with Banz (1981) arguments, Fama and French (1992) presents a three factor model, similar to *CAPM*, but with two additional risk factors. The interpretation of their model is that the stock market consists of several risk factors, with the market risk factor used in *CAPM*, only being one risk dimension. They suggest a size risk dimension in line with Banz (1981), and a value risk dimension, first identified by Bhandari (1988). Their model has the form,

$$\begin{aligned} r_i(t) - r_f(t) = & \alpha_i + \beta_{i,m} \cdot (r_m(t) - r_f(t)) + \dots \\ & \dots + \beta_{i,s} \cdot r_{\text{size}}(t) + \beta_{i,v} \cdot r_{\text{value}}(t) + \epsilon_{i,t} \quad \epsilon_{i,t} \sim \mathcal{N}(0, \sigma_{i,t}). \end{aligned} \tag{3}$$

In Fama and French (1993), they extend their work from Fama and French (1992) to a more general application. They investigate other financial assets than equity, such as fixed income, and also other regions than the US. They conclude that the model seems to perform well regardless of asset class and geographical region.

Another multi-factor linear model that has received big interest is Carhart (1997)'s 4 factor model. Carhart (1997) focuses on explaining the returns of mutual funds, and he finds that the three factor model from Fama and French (1992) misses what Carhart (1997) characterize as the momentum effect. Therefore he suggests an additional factor in the form of price momentum which Carhart (1997) proves gives a significant increase in explaining fund managers' returns.

Among practitioners multi-factor models have been of big importance, both in performance valuation and in portfolio construction. While *CAPM* has been the most common way to evaluate the skill of an active investor, the multi-factor models have given a framework to shed additional light on active investors ability to deliver alpha. From the portfolio managers' perspective the risk exposure towards different factors have consequently been an increased concern. Hedge funds that traditionally aimed to be beta neutral in the *CAPM* framework, now have an increased pressure to stay beta neutral in a multi-dimension framework.

The multi-factor frameworks have also had an impact on the passive investments. Traditional passive investments have been associated with cap-weighted indices which traditionally gives exposure to the market risk premium. With an increased interest from investors to capture other risk premiums in the market, many banks and other financial institutions offer so called *smart beta* products which offer indices with a tilt towards one or more factors, with value and momentum arguably being the two most common. Amenc and Goltz (2013) also shows that many value-weighted market indices, that is often used as proxies for the market risk factor, in fact often displays significant exposure towards other risk factors, such as exposure towards growth over value and large size rather than small size.

There has also been empirical evidence that over time exposure towards some risk factors can yield significant higher risk adjusted returns compared to exposure towards other risk factors. Fama and French (1998) shows that value stocks outperformed growth stocks in most financial markets in the time period between 1975 and 1995. Similarly Asness, Moskowitz, and Pedersen (2013) investigates the excess returns of momentum and value, and provides comprehensive evidence on the excess return of momentum and value based strategies globally across assets, and also highlights the negative correlation between the value and the momentum factor.

### 2.1.2 Regime Based Asset allocation

Within financial returns there are several aggravating features that differ from the common approach with normally distributed returns (or log-normal returns). Perhaps the most obvious are volatility clustering, skewness and fat tails (kurtosis). Another feature of financial market returns is periodicity in market correlation. Erb, Harvey, and Viskanta (1994) argues that correlation within, and between equity markets are related to the business cycle, and that stocks tend to correlate more during bear markets than under normal market conditions. In order to address this issue Ang and Bekaert (2002) uses a regime based model approach to capture the asymmetric correlation in equities, and with promising results. Their approach is greatly inspired by Hamilton (1989), which is arguably the most considerable paper in regime based modelling of financial markets. Hamilton (1989) uses a regime shifting model to capture business cycle behaviour in modeling of real GNP, and found strong support for the models ability to capture and forecast expansions and recessions in the real economy. In line with Hamilton (1989), Ang and Bekaert (2002) identifies two regimes, one normal market regime with relatively low correlation and low variance, and one bear market, or high variance regime with high correlations, high variances, and negative or at least lower expected market returns. They broaden their research in Ang and Bekaert (2004) and concludes that the existence of regimes greatly affects how investors should think of efficient asset allocation. Their findings of market regimes have great implications on the optimal mean-variance portfolio, and it implies a requirement of a dynamic portfolio to be the efficient frontier by Markowitz (1959).

With the help of increased computer power and progress in machine learning algorithm efficiency, more recently notable progress has been done in regime based market modeling. Bulla et al. (2011) investigates the existence of profitable regime based asset allocation strategies based on daily returns. Older research has more commonly been based on monthly returns. Bulla et al. (2011) concludes that a simple stock/cash strategy (hold a stock index in a low volatility regime and hold cash in a high volatility regime) outperforms a static buy and hold strategy both in absolute returns and risk adjusted returns. Further evidence has been given to the support of profitability in regime based investment strategies, for instance by Nystrup, Hansen, et al. (2015) which investigates different strategies involving long/short equity and long/short equity/bonds.

Ammann and Verhofen (2006) presents a regime model based on different risk factors (to the author's chagrin). They use the 4 factor model introduced by Carhart (1997) and the factors' monthly returns. Ammann and Verhofen (2006) find that in a low volatility regime the momentum and the market factors offers best returns, but in a high volatility regime only the value factor has good performance.

In conclusion the research published on regime based modeling of financial markets indicates

promising results. There exists no clear superior argument for how many regimes that should be used, but for simplicity two regimes are often sufficient (Nystrup, Madsen, and Lindström 2016). A two regime approach often tend to result in one low volatility regime and one high volatility regime. Where the high volatility regime is associated with a bear market, and unsatisfactory expected returns. Consequently most regime based investment strategies focus on being long stocks in low volatility regimes, and have a cash or fixed income exposure in high volatility regimes. Ang and Timmermann (2012) gives an comprehensive summary of the benefits with a regime based approach in modelling financial markets and in asset allocation.

## 2.2 Model Concepts and Theory

This section provides review of the most central concepts, and mathematical and statistical theory used in the thesis. The description of the hidden Markov model is the most necessary theory to be familiar with.

### 2.2.1 The hidden Markov model

The hidden Markov model, *HMM*, has found important application with numerous fields, and has been used widely in financial areas. The dynamic of the model is to assume that a system consists of an observable sequence  $(\mathbf{x}_t)_{1:T}$ , which follows a distribution that is determined by an unobserved state sequence  $(Q_t)_{1:T}$  of a finite-state Markov chain. The Markov chain is driven by a transition probability matrix,  $\mathbf{A}$ , which expresses the probability to stay in one state or move to another state conditional on the value of the current state. For a two state process the transition matrix can be expressed as,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad (4)$$

$$\mathbb{P}(Q_{t+1} = j | Q_t = i) = a_{ij} \quad i, j \in [1, 2]$$

Another important property for the Markov chain is that it is "memoryless", meaning that the probability for being in a state in the future is only dependent on the current state value and have no additional dependence on any previous states. This can be expressed as,

$$\begin{aligned} \mathbb{P}(Q_{t+1} = j | Q_t = i, \dots, Q_1 = i) &= \mathbb{P}(Q_{t+1} = j | Q_t = i) = a_{ij}, \quad \forall t > 1 \\ \mathbb{P}(Q_1 = i) &= \pi_i \end{aligned} \quad (5)$$

A Markov chain which length goes to infinity will converge to its steady state distribution,  $\mathbf{T}$ ,

which will give the percentage of the total time the process will be in the different states. Given a two state process it is given by,

$$\mathbf{T} = \lim_{N \rightarrow \infty} \mathbf{A}^N = \lim_{N \rightarrow \infty} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^N = \begin{pmatrix} t_1 & t_2 \\ t_1 & t_2 \end{pmatrix} \quad (6)$$

In the *HMM* the distribution of the observable sequence,  $(\mathbf{x}_t)_{1:T}$ , is defined with respect to its corresponding Markov state. Assuming that the observations follow a normal distribution we can express it as,

$$\mathbf{x}_t | (Q_t = i) = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma}_i), \quad (7)$$

with  $i$  indicating the corresponding Markov state,  $\boldsymbol{\mu}_i$  the expected value, and  $\boldsymbol{\Sigma}_i$  the covariance matrix.

The corresponding density function to equation (7) is  $p_i(\mathbf{x}_t | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, Q_t = i)$ , from here on  $p_i(\mathbf{x}_t)$ . For a two state *HMM* the conditional density vector can be expressed as,

$$\mathbf{p}(\mathbf{x}_t) = \begin{pmatrix} p_1(\mathbf{x}_t) & p_2(\mathbf{x}_t) \end{pmatrix}, \quad (8)$$

with  $p_i(\mathbf{x}_t)$  being given by the standard multivariate normal probability density function,

$$p_i(\mathbf{x}_t) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}_t - \boldsymbol{\mu}_i) \boldsymbol{\Sigma}_i^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_i)\right) \quad (9)$$

with  $d$  being the dimension of the joint observations  $\mathbf{x}_t$ .

Based on the above equations, the likelihood function for a *HMM* with an observed series  $(\mathbf{x}_t)_{1:T}$  can be expressed as,

$$l(\mathbf{x}) = \boldsymbol{\pi}^* \circ \mathbf{p}(\mathbf{x}_1) \cdot \mathbf{A} \circ \mathbf{p}(\mathbf{x}_2) \cdot \dots \cdot \mathbf{A} \circ \mathbf{p}(\mathbf{x}_T) \cdot \mathbf{1}^* \quad (10)$$

with  $\circ$  denoting the Hadamard product.

### 2.2.2 Viterbi's path

Given an observed sequence  $(\mathbf{x}_t)_{1:T}$  that is assumed to follow a *HMM* with  $N$  hidden states, we want to establish the most probable corresponding hidden Markov chain. We assume that

all parameters are given and denote them  $\Theta$ . The most probable Markov sequence is given by solving,

$$Q_{1:T}^{opt} = \arg \max_{Q_{1:T}} \mathbb{P}(Q_{1:T} | \mathbf{x}_1, \dots, \mathbf{x}_T, \Theta). \quad (11)$$

Solving this problem is the same as maximize  $\mathbb{P}(Q_{1:T}, \mathbf{x}_{1:T} | \Theta)$ , which the famous Viterbi's algorithm does by induction (Viterbi (1967)).

- First step,  $\delta_1(i) = \pi_i \cdot p_i(\mathbf{x}_1)$
- Induction step,  $\delta_t(j) = \max_i \delta_{t-1}(i) \cdot a_{ij} p_j(\mathbf{x}_t)$ .

$\delta_T(i)$  is the optimal path for ending up in state  $i$  at time  $T$ . Hence, the solution for equation (11), also called Viterbi's path is,

$$Q_{1:T}^{opt} = \arg \max_i \delta_i(T) \quad (12)$$

### 2.2.3 Baum-Welch algorithm

In order to estimate the parameters,  $\Theta$ , in the *HMM* the straight forward approach is to maximize the likelihood function given by equation (10). A popular algorithm for doing this is a variant of the famous EM-algorithm, called the Baum-Welch algorithm. A more extensive review of the Baum-welch algorithm can be found in Bilmes (1998), details regarding the algorithm's efficiency and convergence can be found in Dempster, Laird, and Rubin (1977).

As in the previous sections the parameters are denoted  $\Theta = (\mathbf{A}, \mu, \Sigma, \pi)$ .  $p_i(x_t | Q_t = i)$  is the normal density function given by equation (9) and corresponding to the normal distribution  $\mathcal{N}(\mu_i, \sigma_i)$ .

**Recursive calculations**, the Baum Welch algorithm uses the E and M step from the EM-algorithm, with recursive calculation, and with one forward procedure and one backward procedure.

#### E-step

**Forward procedure** We define  $\alpha_i(t) = p(x_1, \dots, x_t | Q_t = i, \theta)$ , with  $i \in [1 : N]$ . It is calculated recursively for all  $t \in [1 : T]$ ,

1.  $\alpha_i(1) = \pi_i p(x_1 | Q_t = i)$

$$2. \alpha_j(t+1) = p(x_{t+1}|Q_t = j) \sum_{i=1}^n \alpha_i(t) a_{ij}.$$

Recall that  $a_{ij}$  is an element in the transition matrix  $\mathbf{A}$  defined as  $a_{ij} = \mathbb{P}(Q_{t+1} = j|Q_t = i)$ .

**Backward procedure** Then we define  $\beta_i(t) = p(x_{t+1}, \dots, x_T|Q_t = i, \theta)$ , which is also recursively calculated for  $t \in [1 : T]$ ,

$$1. \beta_i(T) = 1$$

$$2. \beta_i(t) = \sum_{j=1}^N \beta_j(t+1) a_{ij} p(x_{t+1}|Q_t = j)$$

We now define jet another variable  $\gamma_i(t) = \mathbb{P}(Q_t = i|x_{1:T}, \theta)$ . By Bayes' theorem we can use  $\alpha_i(t)$  and  $\beta_i(t)$  in order to calculate,

$$\begin{aligned} \alpha_i(t)\beta_i(t) &= p(x_{1:T}, Q_t = i|\theta) \quad \forall \quad i \in [1 : N] \\ p(x_{1:T}|\Theta) &= \sum_{j=1}^N p(x_{1:T}, Q_t = j) = \sum_{j=1}^N \alpha_j(t)\beta_j(t) \end{aligned}$$

and we calculate  $\gamma_i(t)$  for all  $i \in [1 : N]$  and for all  $t \in [1 : T]$ ,

$$\gamma_i(t) = \mathbb{P}(Q_t = i|x_{1:T}, \Theta) = \frac{p(x_{1:T}, Q_t = i|\Theta)}{p(\mathbf{x}|\Theta)} = \frac{\alpha_i(t)\beta_i(t)}{\sum_{j=1}^N \alpha_j(t)\beta_j(t)}. \quad (13)$$

In a similar way we also define  $\xi_{ij}(t) = \mathbb{P}(Q_t = i, Q_{t+1} = j|x_{1:T}, \theta)$ , which can be calculated as,

$$\begin{aligned} \xi_{ij}(t) &= \mathbb{P}(Q_t = i, Q_{t+1} = j|x_{1:T}, \Theta) = \frac{\mathbb{P}(Q_t = i|\mathbf{x}, \Theta) p(x_{t+1}, \dots, x_T|Q_t = i, \Theta)}{p(x_t, \dots, x_T|Q_t = i, \Theta)} \\ &= \frac{\gamma_i(t) a_{ij} p(o_{t+1}|Q_t = j, \Theta) \beta_j(t+1)}{\beta_i(t)} \end{aligned} \quad (14)$$

## M-step

By the definition of  $\gamma_i(t)$  and  $\xi_i(t)$  from equation (13) and (14), we can obtain the maximum likelihood estimation of the model parameters.

We start by estimating the paramters for the hidden Markov chain. We define  $I_t(i)$  as a indicator variable with the property,  $I_t(i) = 1$  if the model is in regime  $i$  at time  $t$  and zero for all other  $t$ . We also define the indicator variable  $I_t(i, j)$ , with the property of  $I_t(i, j) = 1$  if the model moves from regime  $i$  to  $j$  at time  $t$ , otherwise it is zero. It is intuitively true that  $\mathbb{E}(\sum_{t=1}^T I_t(i))$  and  $\mathbb{E}(\sum_{t=1}^T I_t(i, j))$  is given by,



$$\begin{aligned}
\sum_{t=1}^T \gamma_i(t) &= \mathbb{E}\left(\sum_{t=1}^T I_t(i)\right), \\
\sum_{t=1}^T \xi_i(t) &= \mathbb{E}\left(\sum_{t=1}^T I_t(i, j)\right).
\end{aligned} \tag{15}$$

by our estimates of  $\mathbb{E}(\sum_{t=1}^T I_t(i))$  and  $\mathbb{E}(\sum_{t=1}^T I_t(i, j))$  we can estimate the transition elements as,

$$\begin{aligned}
\pi_i &= \gamma_i(1) \\
\hat{a}_{ij} &= \frac{\sum_{t=1}^T \xi_{ij}(t)}{\sum_{t=1}^T \gamma_i(t)} \quad i, j \in [1 : N]
\end{aligned} \tag{16}$$

Now we estimate the the parameters in the normal distribution in the following way,

$$\begin{aligned}
\hat{\mu}_i &= \frac{\sum_{t=1}^T \gamma_i(t) \cdot x_t}{\sum_{t=1}^T \gamma_i(t)} \quad i \in [1 : N] \\
\hat{\sigma}_i^2 &= \frac{\sum_{t=1}^T \gamma_i(t) \cdot (x_t - \hat{\mu}_i)^2}{\sum_{t=1}^T \gamma_i(t)} \quad i \in [1 : N].
\end{aligned} \tag{17}$$

In the iterative procedure of the Baum-Welch algorithm the likelihood from equation (10) is evaluated, and we return to the **M-step** with the updated estimates of  $\Theta$ . The procedure is updated until the increase in the likelihood function is sufficiently flat.

#### 2.2.4 Shrinkage of covariance matrices

In many financial applications a covariance matrix needs to be estimated, and not seldomly the size of the covariance matrix is relatively large to the sample size. This results in an ill-conditioned matrix, sometimes even not invertible. Ledoit and Wolf (2004) suggests a linear combination with the sample covariance matrix,  $\Sigma^S$  and the identity matrix,

$$\Sigma^* = \rho_1 \cdot I + \rho_2 \cdot \Sigma^S. \tag{18}$$

This approach allow for a guarantee that the eigenvalues are positive and therefore that the matrix  $\Sigma^*$  is invertible.

An extension of the idea is to shrink the covariance matrix towards the identity matrix. The shrinkage intensity is decided by a quadratic loss function,  $\mathbb{E}(\|\Sigma^* - \Sigma\|^2)$ . With  $\Sigma$  being the true

covariance matrix. Based on the well-conditioned estimator of Ledoit and Wolf (2004), Fiecas et al. (2017) suggests an estimation of  $\Sigma^*$  for the EM algorithm, which they express as,

$$\Sigma_k^* = (1 - w) \cdot \Sigma_k^S + w \cdot \alpha \cdot I_p. \quad (19)$$

In this thesis we will use  $\alpha = 1/p \cdot \text{tr}(\Sigma_k^S)$  which is deduced by Fiecas et al. (2017) for minimizing the quadratic loss function. The  $w$  is left as a free parameter to tune the model to convenient behaviour.

### 2.2.5 Frobenius norm

If we consider a matrix of dimension  $m \times n$  with each element in the matrix denoted as  $a_{ij}$ , we define the Frobenius norm as,

$$\|A\|_F = \sqrt{\text{tr}(AA^T)} \quad (20)$$

### 2.2.6 Skewness and Kurtosis

The first and second moment is often considered when investigating the distribution of a stochastic process, but often higher moments also contribute to important information about the distribution. In this thesis also the third and fourth moment will be of importance, and are therefore described below.

Skewness can be interpreted as a measure of asymmetry in a distribution and is given by,

$$s = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}\right)^3} \quad (21)$$

Under the assumption of the underlying process being normally distribution the skewness convergence in distribution when the sample size  $n$  goes to infinity as,

$$\sqrt{n} \cdot s \rightarrow \mathcal{N}(0, 6) \quad (22)$$

Kurtosis is a measure of a process' spread, as the variance will give a sense of the spread by most of the a process' values, kurtosis is a measure of how wide out the most extreme values are in a distribution. It is given by,

$$k = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}\right)^4} \quad (23)$$

As for skewness the kurtosis converge in distribution when the sample size  $n$  goes to infinity based on the underlying process being normally distributed,

$$\sqrt{n} \cdot k \rightarrow \mathcal{N}(3, 24) \quad (24)$$

## 3 Data

The focus of this thesis is US stocks, and the logarithm transformation of daily return time series defined by  $r_t = \ln(p_t) - \ln(p_{t-1})$  with  $p_t$  being the closing price at day  $t$ . The different portfolios considered are called factors and are based on the approach from Fama and French (1992), and Carhart (1997). The detail of how the factors are constructed is given in section 3.1. The data is collected from *Kenneth R. French* n.d., which keeps data related to factors in different asset universes available on his web page. The data in this thesis is from the time period 1 January 1998 to 31 January 2017 which implies a total of 4801 trading days.

### 3.1 Factors construction

Market risk factors is an academic approach to market price behaviour, and are not directly observed. In academic papers indices or different kind of portfolios are normally used as proxies for different market risk factors. In this paper we take use of the portfolios and factors constructed by Fama and French (1992) and Carhart (1997). The stock universe is defined as all NYSE, AMEX and NASDAQ stocks available at CRSP and Compustat.

The most intuitively factor is the market factor, and the market factor is simply a value-weighted portfolio of all stocks in the defined universe minus the risk free rate, as it was already defined in William F. Sharpe (1964). The remaining factors are less intuitive. For the multidimensional factor model to be efficient, the factors should be linearly independent, i.e. orthogonal, or the factorial space could possibly be reduced. In order to address this Fama and French (1992) starts with constructing portfolios based on size and value criteria and Carhart (1997) follows a similar approach with his additional momentum factor.

The size criteria, presented in Fama and French (1992) is simply based on current market value (stock price  $\times$  number of stocks). The value criteria also from Fama and French (1992) uses the book-to-market ratio, B/M, which gives a ratio between a company's booked value, i.e. its assets on the balance sheet divided by the valuation based on its current stock price. A high B/M indicates a so called value stock, while a low B/M indicates a growth stock, terms often used in practice. For the additional factor by Carhart (1997), the momentum criteria, Carhart (1997) uses the price returns on the 2-12 last months.

As one may expect, there is high correlation between portfolios based on the different criteria. In order to address this problem two steps are taken, the first is to sort stocks with a 2 dimensional criteria base. The first dimension being size, with the stocks being sorted into small and big, and the other dimension being value or momentum. Table 1 and 2 show how the value-, and

momentum portfolios are constructed. The portfolios are updated and re-weighted every day, and the portfolios are all value-weighted.

	Low B/M	Neutral	High B/M
Small	$SG$	$SN$	$SV$
Big	$BG$	$BN$	$BV$

Table 1: Portfolios  $2 \times 3$  based on Size-Value

	Low	Neutral	High
Small	$SL$	$SN$	$SH$
Big	$BL$	$BN$	$BH$

Table 2: Portfolios  $2 \times 3$  based on Size-Momentum

The breaking values for the value and momentum criteria are the 30:th and 70:th percentiles for the NYSE, which allows the portfolios to be well diversified from idiosyncratic behaviour. As limit for the size criterion the median size of the NYSE stocks is used.

The second step is to use the portfolios in order to construct the factors with low correlation. Equation (25) displays how the size factor is constructed, and can simply explained as small companies minus big companies,  $SMB$ . Equation (26) shows how the value factor is constructed, and can be described as high B/M minus low B/M companies,  $HML$ . Last, equation (27) displays how the momentum factor is constructed with high momentum minus low momentum,  $MOM$ .

$$\begin{aligned}
SMB_{B/M} &= \frac{1}{3} \cdot (SV + SN + SG) - \frac{1}{3} \cdot (BV + BN + BG), \\
SMB_{OP} &= \frac{1}{3} \cdot (SR + SN + SW) - \frac{1}{3} \cdot (BR + BN + BW), \\
SMB_{INV} &= \frac{1}{3} \cdot (SC + SN + SA) - \frac{1}{3} \cdot (BC + BN + BA), \\
SMB &= \frac{1}{3} \cdot (SMB_{B/M} + SMB_{OP} + SMB_{INV}).
\end{aligned} \tag{25}$$

$$HML = \frac{1}{2} \cdot (SV + BV) - \frac{1}{2} \cdot (SG + BG). \tag{26}$$

$$MOM = \frac{1}{2} \cdot (SH + BH) - \frac{1}{2} \cdot (SL + BL). \tag{27}$$

The approach described in equations (25), (26) and (27) describe how the portfolios from table 1 and 2 are used in attempt to construct orthogonal factors, with low correlation. The subtraction between two different sets of portfolio can be interpreted as one long, and one short position with the implication of a net neutral market exposure.

### 3.2 Data Analysis

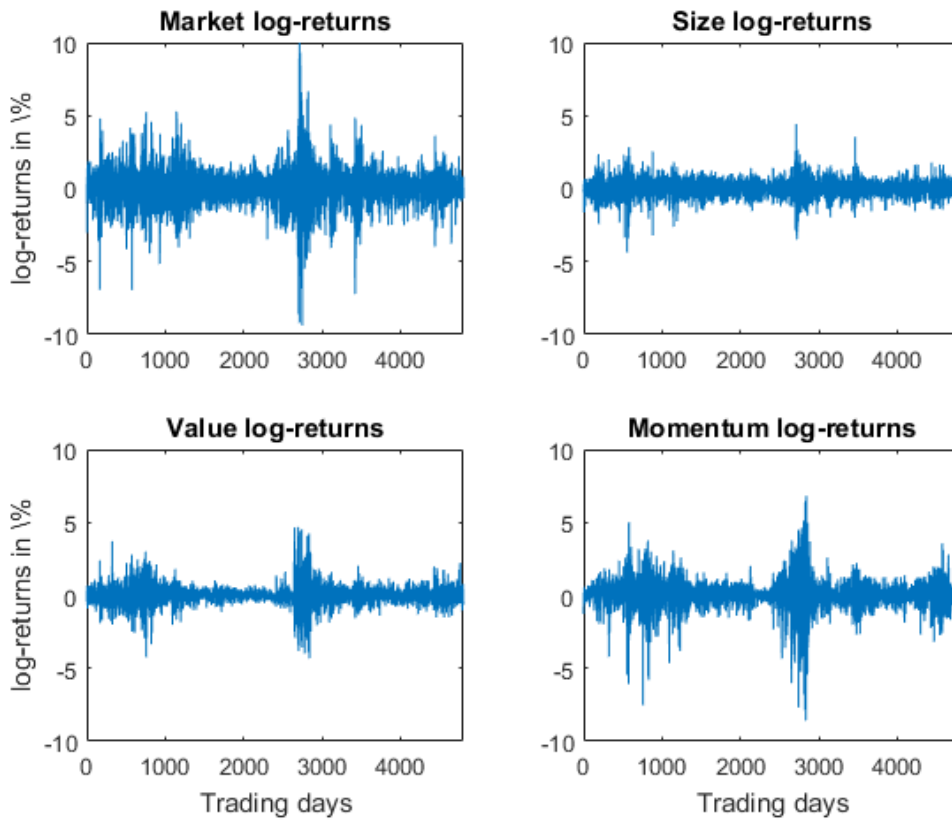


Figure 2: The log-returns over the entire time set of data (1 January 1998 to 31 January 2017) for the 4 factors that initially are considered in this thesis.

In this section a short analysis is done on the log-returns in order to compare with the results received after modelling.

Table 3 shows the correlation between the log-transformed data. As we can see the factors from Fama and French (1992) have negligible correlation, while the momentum factor has significant negative correlation with both value and market.

	Mkt - RF	SMB	HML	MOM
Mkt - RF	1	0,0658	0,00023	-0,2826
SMB	0,0658	1	0,0569	0,0344
HML	0,00023	0,0569	1	-0,3407
MOM	-0,2826	0,0344	-0,3407	1

Table 3: Correlation Matrix between log-transformed factor returns

As financial returns often are modeled as normally or log-normally distributed it is of interest to

investigate if this seems feasible. Table 4 displays the mean, standard deviation, skewness and kurtosis of the log-returns. For a normal distributed time series we would expect skewness to be 0 and the kurtosis to be 3. Hence, we can conclude that the assumption of normal distribution before modelling seems like a poor fit, with overall negative skewness and very "fat tails", i.e. high kurtosis.

	Mean	SD	Skewness	Kurtosis
Mkt - RF	0,0267	1,2462	-0,0824	9,8778
SMB	0,0119	0,6166	-0,1597	6,3806
HML	0,0109	0,6707	0,4172	10,6982
MOM	0,0197	0,9981	-0,7921	11,3873

*Table 4: Data analysis of the different factors.*

Last, figure 3 displays the cumulative returns of the factors over the whole time frame, 1 January 1998 to 31 January 2017. We can for instance note the huge draw down of the momentum factor associated with the financial crisis in 2008.

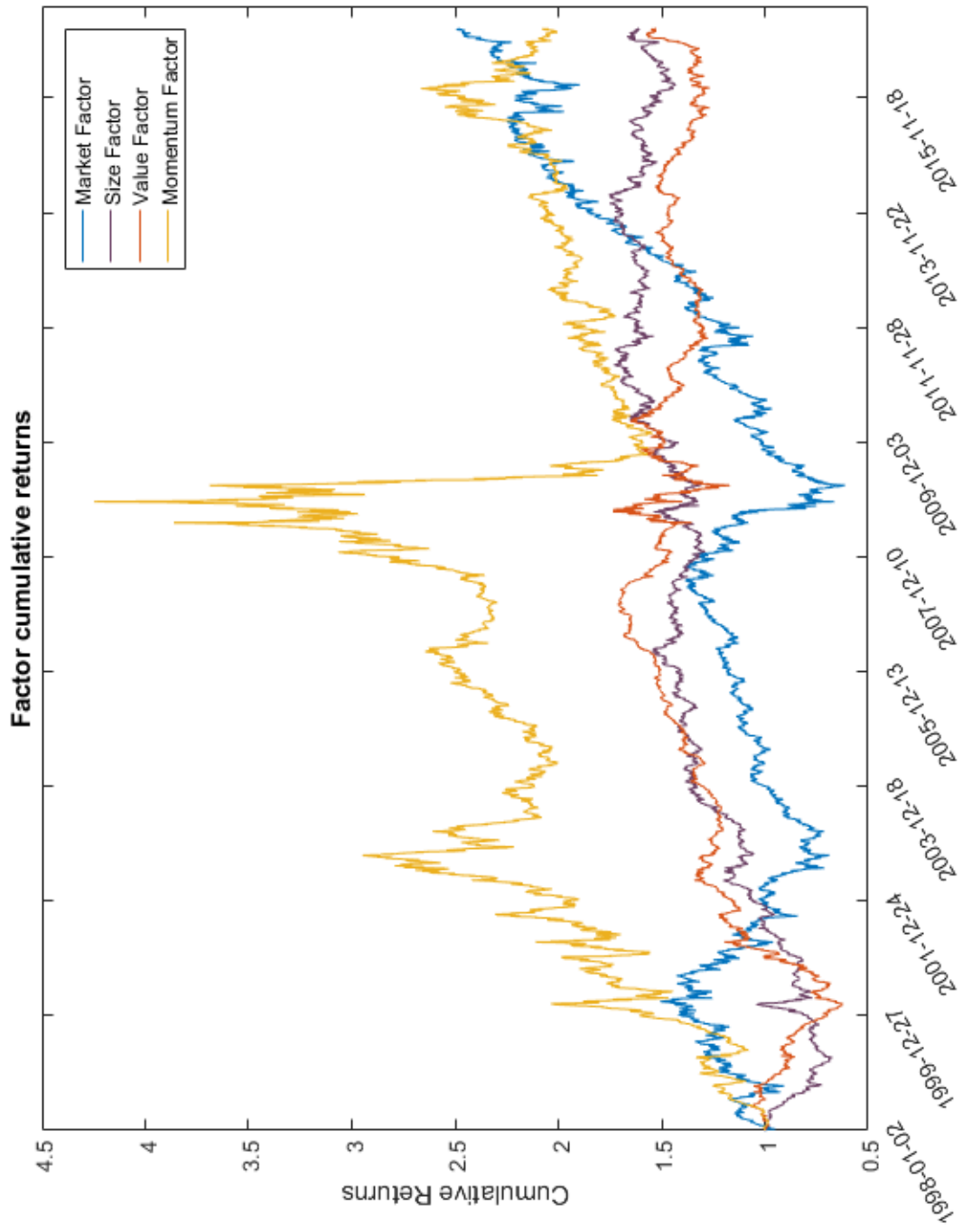


Figure 3: The cumulative returns for the 4 factors over the entire data set.



## 4 Modeling

A typical problem with modeling time series data is how to estimate and evaluate the model over the available data set. Within many fields it is common to split the data set into one modeling section and one evaluation section. For such a model approach to be suitable the system that is to be modeled needs to be time consistent. Financial data is typically difficult to evaluate as time consistent over long time spans. In this thesis the Markov chain will be considered to be time consistent, and therefore the transition probabilities will be static. The parameter for the normal distribution of the observed log-returns will however change over time in the online application of the model.

The model approach is a two regime hidden Markov model, with the observed factors assumed to be generated by a multidimensional normal distribution. As in section 2 the Markov process is denoted  $Q_t$  with  $t \in [1 : T]$  and the observed process will have the probability density function  $p_i(\mathbf{x}_t) = p(\mathbf{x}_t | Q_t = i)$  with  $i \in [1, 2]$ . Regime 1 will denote the low volatility regime, and 2 the high volatility regime.

The first part of the modeling section is to find a suitable model. In order to do so the whole set of data is considered. The Baum-Welch algorithm is used to estimate the parameters, and the Viterbi's path is calculated. Figure 4 shows the probability,  $\mathbb{P}(Q_t = 1 | \mathbf{x}_1, \dots, \mathbf{x}_T, \Theta)$ , and the Viterbi's path when all 4 factors are considered and when only the market factor is considered.

We can see that the 4 factor model is very unstable and volatile in its regime decision, especially if we compare it with the Viterbi's path associated with the market factor model. The 4 factor model and the market model has the same regime in 3239 trading days. The number of regime shifts are 60 for the market model, while it is 444 for the 4 factor model. In total the market model spent 3614 trading days in the low volatility regime and 1187 trading days in the high volatility regime, and on average the duration in the high volatility regime is 79 trading days. The 4 factor model tend to leave the high volatility regime more rapidly, on average the duration for the 4 factor model is 11.8 trading days in the high volatility regime, and in total the 4 factor spent 1305 days in the high volatility regime.

The conclusion from the 4 factor model is that it is not regime stable, since it changes regime too often and tend to have unsatisfactory average duration in the high volatility regime. The comparison with the model only based on the market factor emphasize these features.

In order to address the problems we decide to reduce the model to a 3 factor model without the size factor, and we use the shrinkage approach from section 2.2.4 for the covariance matrices. The weighting is done by empirical testing, and the result is that the weight for the high volatility

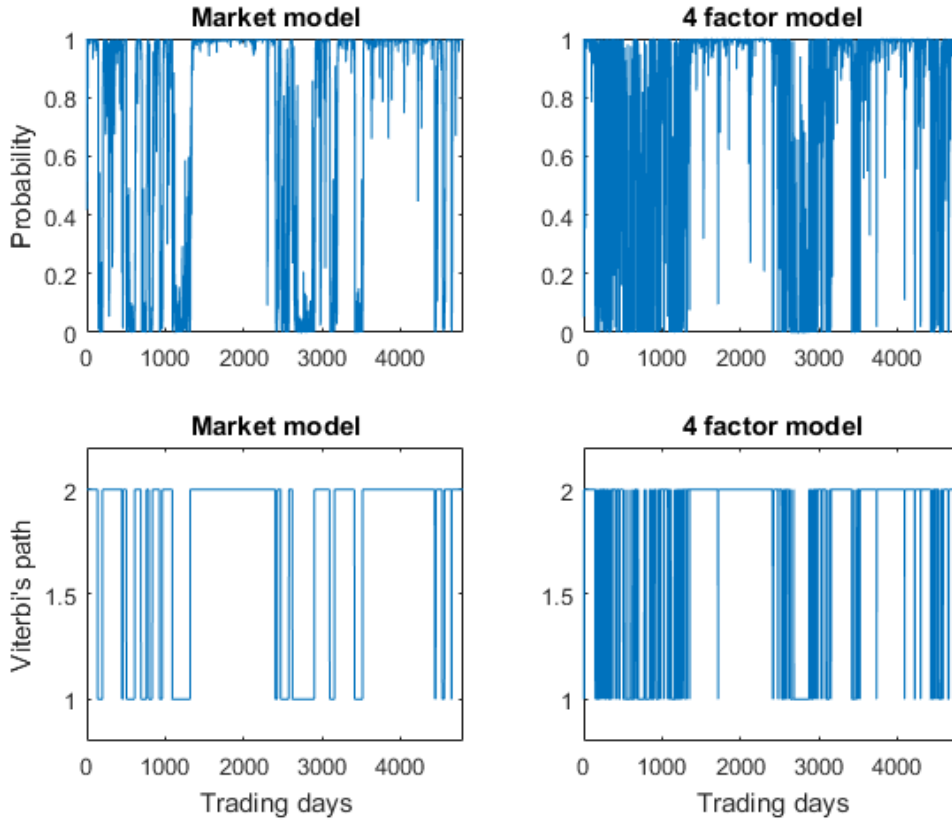


Figure 4: First the probability for market model and the 4 factor model to be in the low volatility regime,  $\mathbb{P}(Q_t = 1|x_1, \dots, x_T, \Theta)$ . Below is the corresponding Viterbi's path, i.e. the most likely path for the state process.

covariance matrix is set to 0, i.e. no shrinkage, while the weight for the low volatility covariance matrix is set to 0.25. In Figure 20 in section 8.2 the average duration in the high volatility regime is displayed as a function of different shrinkage weights. Figure 5 displays the shrinkage 3 factor model over the entire time set and the probability  $\mathbb{P}(Q_t = 1|x_1, \dots, x_T, \Theta)$  for  $t \in [1 : T]$ . In total the 3 factor shrinkage model spends 1237 trading days in the high volatility regime, and the average duration is 22,3 days, a significant improvement from the more volatile 4 factor model.

#### 4.1 3 factor shrinkage model

In the low volatility regime the mean value for the market factor is 0.0721%, for the value factor it is -0.0014% and for the momentum factor it is 0.0612%. For the high volatility regime the mean value is -0.1044% for the market factor, 0.0466% for the value factor and -0.0999% for the momentum factor. We can see that the momentum and market factor has significantly higher expected returns in the low volatility regime than value, but in the high volatility regime the expected returns for market and momentum are very low while the value factor has a higher

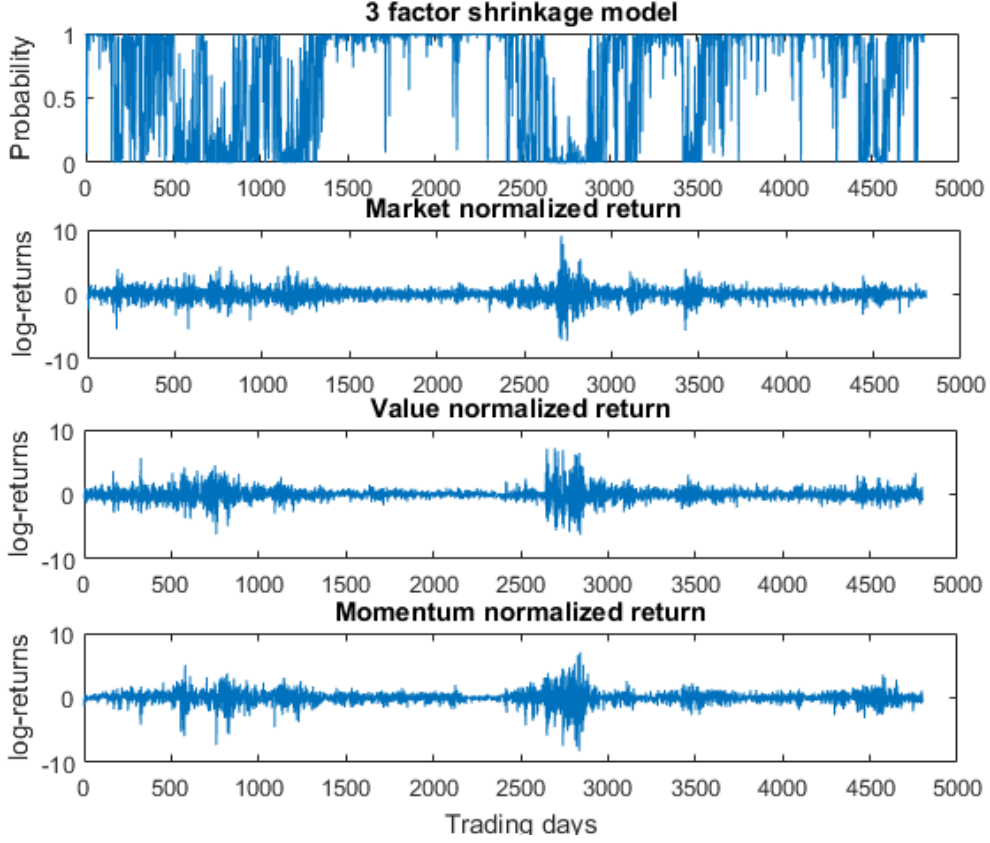


Figure 5: The probability  $\mathbb{P}(Q_t = 1|x_1, \dots, x_T)$  from the 3 factor model, and the corresponding return time series for the 3 factors.

expected return than in the low volatility regime. Hence, we can conclude that there is significant different behaviour from the factors in the different regimes.

Furthermore, if we consider the covariance matrices estimated by the 3 factor shrinkage model, and the correlation matrices estimated from the sorted return time series, we can see further indication on distinct behaviour for the factors in different regimes. We can deduce that the momentum factor and the market factor have significant increased correlation in the high volatility regime which is to be expected based on previous research such as Erb, Harvey, and Viskanta (1994) et al. Also the value factor displays increased correlation with momentum in the high volatility regime while the correlation between the market factor and the value factor stays low. Overall the variances for the factor returns are significantly different in the high and the low volatility regime. Matrices (28) show the low volatility regime and matrices (29) show the high volatility regime.

$$\Sigma_1 = \begin{pmatrix} 0.5822 & -0.0314 & 0.0366 \\ -0.0314 & 0.2141 & -0.0231 \\ 0.0366 & -0.0231 & 0.3323 \end{pmatrix}, \quad \rho_1 = \begin{pmatrix} 1 & -0.1493 & 0.1090 \\ -0.1493 & 1 & -0.1460 \\ 0.1090 & -0.1460 & 1 \end{pmatrix} \quad (28)$$

$$\Sigma_2 = \begin{pmatrix} 4.1474 & 0.1298 & -1.5198 \\ 0.1298 & 1.3122 & -0.7852 \\ -1.5198 & -0.7398 & 2.9802 \end{pmatrix}, \quad \rho_2 = \begin{pmatrix} 1 & 0.0605 & -0.4370 \\ 0.0605 & 1 & -0.4007 \\ -0.4370 & -0.4007 & 1 \end{pmatrix} \quad (29)$$

Considering the distribution of the factor returns we can also observe the skewness and kurtosis for the two regimes. Figure 6 shows the skewness and kurtosis, and its confidence interval. As we can see it does not perfectly satisfy the assumption of normally distributed log-returns, and especially it still displays the feature of fat tails.

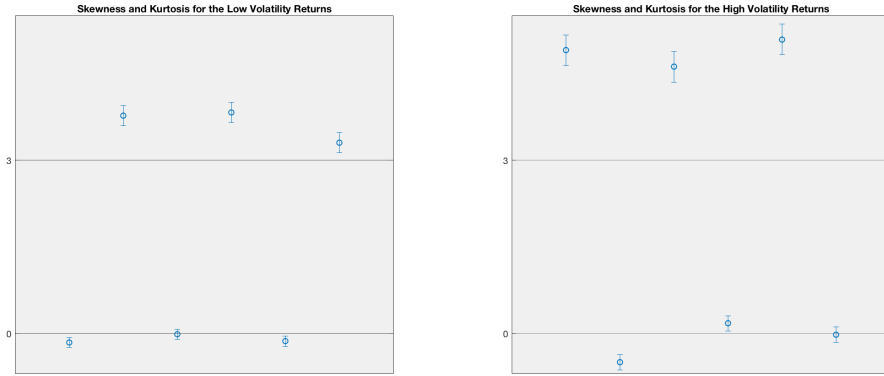


Figure 6: The kurtosis and skewness for the low volatility and high volatility regime with 95% confidence interval for a normal distribution. In theory the skewness should be 0 and the kurtosis should be equal to 3.

Last, we examine the Markov chain and the transition matrix estimated. In the online application of the model this transition matrix will stay fixed and not be updated.

$$\mathbf{A} = \begin{pmatrix} 0.9746 & 0.0254 \\ 0.0723 & 0.9277 \end{pmatrix} \quad (30)$$

We can see that the model is more probable to stay in low volatility regime than in the high volatility regime, but that the probability of staying in the high volatility regime still is very high, which further states the regime stability of the model. By the estimated transition probability we can also estimate the steady state probabilities defined in equation (6) from section 2.2.1. The Markov process converges to,

$$\mathbf{T} = \begin{pmatrix} 0.74 & 0.26 \\ 0.74 & 0.26 \end{pmatrix}, \quad (31)$$

which means we expect the process to spend 74% of the time in the low volatility regime, and 26% of the time in the high volatility regime.

## 4.2 Efficient Frontier Static Model

Based on the parameters estimated over the whole data set we investigate the efficient frontier in the different regimes. In the first approach we assume that we are certain of the regime the model is in at time  $t$ . This yields one static efficient frontier when we know that we are in the high volatility regime and one static frontier when we are in the low volatility regime. The result can be seen in figure 7. The calculation is done with the constraint to short selling not being accepted, and the weighting is summed up to 1.

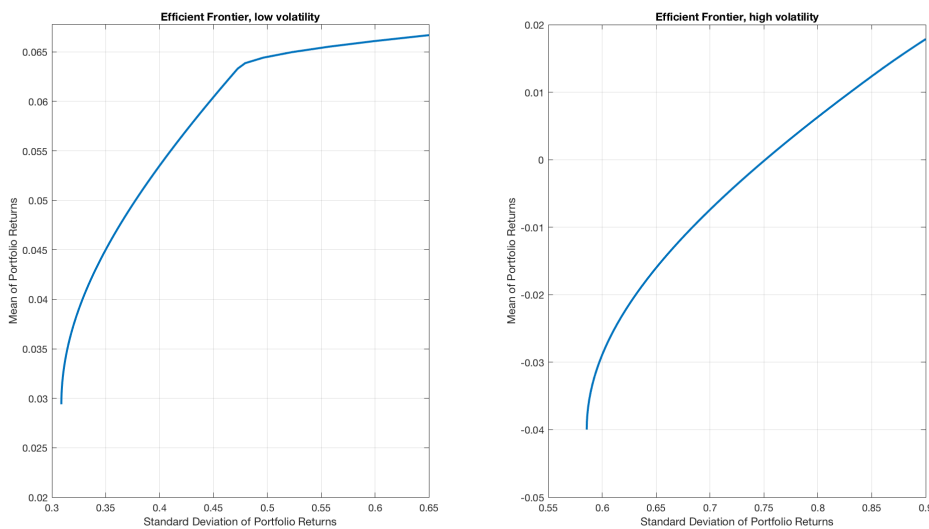


Figure 7: The efficient frontier in the high and the low volatility regime based on the parameters estimated by the 3 factor shrinkage model when the whole time set is considered.

We also calculate the portfolio construction for the highest Sharpe ratio. The optimal portfolio construction in the high volatility regime is a 100% allocation to the value factor, which yields an expected return of 0.037% per day and with a standard deviation of 1.115. In the low volatility regime the optimal portfolio construction is 34.92% invested in the market factor, 7.50% in the value factor and 57.58% in the momentum factor. The expected return is 0.0585% with a standard deviation of 0.436. Important to remember that these calculations and numbers are based on the data set of log-transformed returns.

To have a comparable benchmark in evaluating the efficient frontiers of the 3 factor shrinkage model we use the maximum likelihood approach and estimates a multidimensional normal distribution over the whole data set with no regime approach. The estimated expected values and covariance

matrix are used to calculate the optimal portfolio when we assume there exists only one regime. The result is 22.87% in the market factor, 41.72% in the value factor, and 35.40% in the momentum factor.

## 5 Time varying parameters and regime decision

Section 4 has led to the determination of a shrinkage 3 factor model based on the parameter estimations of the whole data set and the corresponding Viterbi's path. As previously mentioned we assume the Markov chain process to be static over time, and therefore the transition probabilities will be fixed. However, to make the model applicable in an online scenario, and used as a basis for asset allocation decisions in different time periods we will allow the parameter to change over time. We use a rolling window with fixed length  $n$ , and re-estimate the parameters daily. At time  $t$  the time frame from  $t - 1$  to  $t - n$  is used to update the parameter in the conditional multidimensional normal distributions. The window length is set to 250 trading days. The choice of window length is based on a heuristic approach, but also have the nice feature of accounting for a calendar year of trading days. Figure 8 shows the estimated parameters for  $\mu$  in the low volatility and the high volatility regime and the Frobenius norm of the corresponding covariance matrices. Note that the data set now has decreased to 4551 trading day since the first 250 days are used to estimate the first set of parameters.

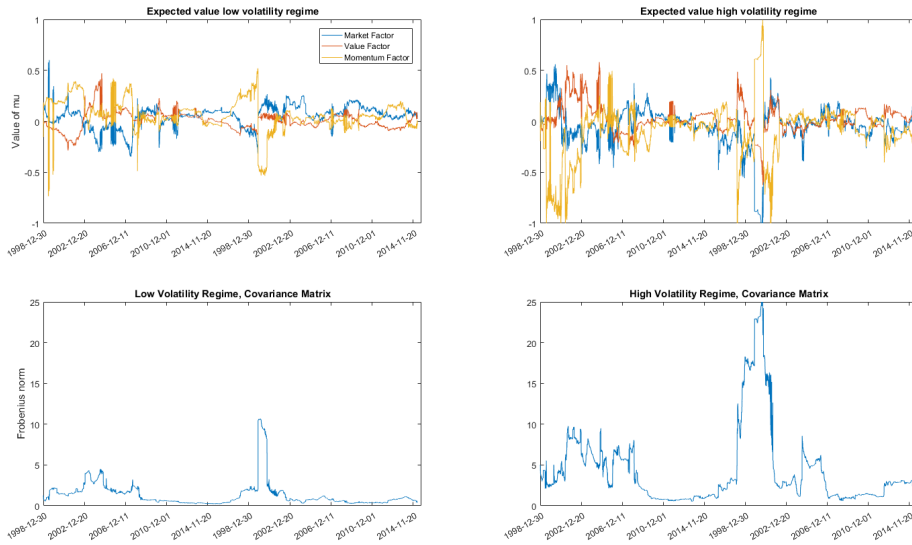


Figure 8: The  $\mu$  parameters over time when a fixed window of 250 observations are used in estimating the model parameters, and the Frobenius norm of the corresponding covariance matrix.

As we can see the parameters are quite volatile over time, especially in the high volatility regime. Considering that we evaluating almost 20 years of returns, this may not be too surprising. We denote the set of parameters estimated at time  $t$  with  $\Theta_t$

In order to use the model online we also have to take regime decision for time  $t$  at time  $t - 1$ . In section 4.2 we used the Viterbi's path to decide which regime the process was in, which assumes that we know all existing future observations. For the online application we will use the so called

filter probability,  $\mathbb{P}(Q_t = q_t | x_{t-1}, \dots, x_{t-n})$  with  $q_t \in [1, 2]$ . However, to solely base regime decision on the most likely regime makes the process regime unstable, and therefore we only change regime if the model has a high conviction of a regime change. The regime decision we use can be expressed as,

$$Q_t = \begin{cases} 2 & \text{if } \mathbb{P}(Q_t = 2 | x_{t-1}, \dots, x_{t-n}, \Theta_t, Q_{t-1} = 1) > w_2 \\ 1 & \text{if } \mathbb{P}(Q_t = 1 | x_{t-1}, \dots, x_{t-n}, \Theta_t, Q_{t-1} = 2) > w_1 \\ Q_{t-1} & \text{else} \end{cases} \quad (32)$$

The limits  $w_1$  and  $w_2$  are set to  $w_1 = 0.98$  and  $w_2 = 0.95$  by heuristically evaluating of the one step regime predictions and the Viterbi's path from section 4.

We now have a 3 factor shrinkage model with time dependent parameters and a scheme for online regime decision. Since the model only uses the past 250 trading days, and no future values, the performance of the model should in theory be applicable to any time set of factor returns. The one step regime prediction can be seen in figure 9. In some shorter time periods we can observe that the model seems to be a bit unstable in its regimes decisions and falls back and forth between regimes, but overall the outcome seems satisfactory stable. The total number of regime shifts are 120, almost exactly as many as when the parameters were fixed over the whole time interval. Per year that is around 6.6, with an average duration in the high volatility regime being 30.8 trading days. In total the model predicts 1848 trading days in the high volatility regime which accounts for almost 41% of the time period. Compared to the stable distribution from equation (31) this figure is higher than expected.



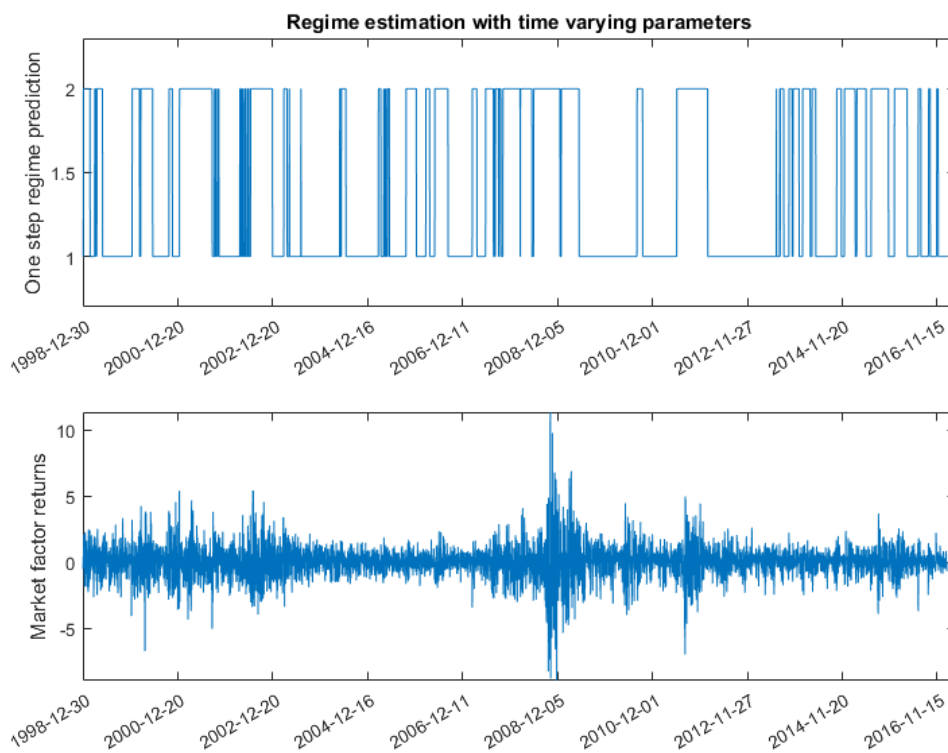


Figure 9: The one step predicted regimes based on the filter probabilities from the time varying parameters and the decision scheme.

## 6 Asset Allocation

Based on the online approach of the shrinkage 3 factor model presented in section 5 we will evaluate the investment performance of the model. A few different strategies will be evaluated and compared. The first strategy is a market/cash strategy similar to the strategy evaluated in Nystrup, Hansen, et al. (2015). If the one step regime prediction predicts that the market is in the low volatility regime the strategy invest 100% in the market factor, if the one step regime prediction believes the market will be in the high volatility regime 100% is invested in cash which is equal to 0 return. This strategies would result in a total of 120 re-allocation over the whole time period, which makes it highly feasible in practice. The main benchmark for this strategy is to static allocate 100% in the market factor, or a portfolio weighted between cash and the market factor.

The remaining strategies are more theoretical since they demand reallocation each day. These are the optimal portfolios deduced in section 4.2, since both the optimal portfolio with respect to regimes and with no respect to regimes assumes a constant portfolio weighting, the portfolios must be re-weighted every day in order to satisfy the strategies' weighting schemes. For the regime based optimal portfolio strategy 100% is allocated to the value factor in the low volatility regime, and in the high volatility regime the allocation is 57.58% in the momentum factor, 34.92% in the market factor and 7.5% in the value factor. For the no regime optimal portfolio strategy the allocation is 22.87% in the market factor, 41.72% in the value factor and 35.40% in the momentum factor.

### 6.1 Time Varying Efficient Frontier

We will also consider an efficient frontier that takes advantage of the daily updates of the parameters,  $\Theta_t$ . It will not use the regime decision scheme, but rather the filter probabilities to estimate the expected return and expected risk for every portfolio construction. If we consider the three factors and denote their returns at day  $t$  as  $X_t$ . We do not know  $X_t$ , but we know the filter probabilities for the two regimes for day  $t$ . Then  $X_t$  can be expressed as,

$$\mathbf{X}_t = \lambda_t \mathbf{X}_{1,t} + (1 - \lambda_t) \mathbf{X}_{2,t}. \quad (33)$$

$\mathbf{X}_{1,t}$  is the returns in the low volatility regime, and  $\mathbf{X}_{2,t}$  is the returns in the high volatility regime.  $\lambda_t$  is the filter probability of being in regime 1. The expected value of  $\mathbf{X}_t$  is simply given by

$$\boldsymbol{\mu}_t = \lambda_t \boldsymbol{\mu}_{1,t} + (1 - \lambda_t) \boldsymbol{\mu}_{2,t} \quad (34)$$

The covariance can be calculated as,

$$\begin{aligned} \boldsymbol{\Sigma}_t &= \mathbb{E}(X_t X_t^*) - \boldsymbol{\mu}_t \boldsymbol{\mu}_t^* \\ &= \lambda_t^2 \mathbb{E}(X_{1,t} X_{1,t}^*) + (1 - \lambda_t)^2 \mathbb{E}(X_{2,t} X_{2,t}^*) + 2\lambda_t(1 - \lambda_t) \boldsymbol{\mu}_{1,t} \boldsymbol{\mu}_{2,t}^* - \boldsymbol{\mu}_t \boldsymbol{\mu}_t^*. \end{aligned} \quad (35)$$

Here we have used the expression (33) and (34). We have also used the fact that  $\boldsymbol{\mu}_{1,t} \boldsymbol{\mu}_{2,t}^* = \boldsymbol{\mu}_{2,t} \boldsymbol{\mu}_{1,t}^*$ . Since we know the covariance of  $\mathbf{X}_{1,t}$  and  $\mathbf{X}_{2,t}$  from the estimation of  $\boldsymbol{\Theta}_t$  and can express  $\boldsymbol{\mu}_t$  by equation (34) we can calculate the covariance matrix of  $\mathbf{X}_t$  by,

$$\begin{aligned} \boldsymbol{\Sigma}_t &= \lambda_t^2 (\boldsymbol{\Sigma}_{1,t} + \boldsymbol{\mu}_{1,t} \boldsymbol{\mu}_{1,t}^*) + (1 - \lambda_t)^2 (\boldsymbol{\Sigma}_{2,t} + \boldsymbol{\mu}_{2,t} \boldsymbol{\mu}_{2,t}^*) + 2\lambda_t(1 - \lambda_t) \boldsymbol{\mu}_{1,t} \boldsymbol{\mu}_{2,t}^* \\ &\quad - 2\lambda_t(1 - \lambda_t) \boldsymbol{\mu}_{1,t} \boldsymbol{\mu}_{2,t}^* - \lambda_t^2 \boldsymbol{\mu}_{2,t} \boldsymbol{\mu}_{2,t}^* - (1 - \lambda_t)^2 \boldsymbol{\mu}_{1,t} \boldsymbol{\mu}_{1,t}^* \\ &= \lambda_t^2 \boldsymbol{\Sigma}_{1,t} + (1 - \lambda_t)^2 \boldsymbol{\Sigma}_{2,t}. \end{aligned} \quad (36)$$

By these equation we update  $\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t$  at each time step based on the estimation of  $\boldsymbol{\Theta}_t$ , and the filter probability of being in regime 1, we can calculate the optimal portfolio for each day.

## 6.2 Results

Table 5 presents the results from the different strategies. First we consider the two efficient frontier described in the beginning of this section, we denote the efficient frontier with two regimes as *Frontier2*, and the efficient frontier with no respect to regimes as *Frontier1*. Then we consider the cash/market allocation based on regime, denoted *CashMarket4*, and its most logical benchmark which is simply being static the market factor, denoted *Market5*. We will also consider a static portfolios with 100% Value and 100% Momentum, denoted *Value6* and *Momentum7*. Last we have the efficient frontier which is calculated daily with respect to the updated parameter estimations, we denote the strategy *Frontier3*.

As we can see the strategy that performance best is *Frontier3*, both in absolute return and when we consider risk adjusted returns. The *CashMarket4* strategy significantly outperform static allocation in any of the factors. Figure 10 shows the cumulative returns for the *CashMarket4* and *Market5*, together with the estimated regime states. We can see that the model helps *CashMarket4* to leave the market early in to many draw down periods.

Figure 11 shows the cumulative returns of the three different strategies based on the efficient frontier. The two strategies taking advantage of the regimes both have considerably higher volatility

	Total return	Average return	S.D.	Sharpe Ratio
<i>Frontier1</i>	103.6%	0.0165%	0.4254	0.0388%
<i>Frontier2</i>	147.5%	0.0222%	0.6733	0.0330%
<i>Frontier3</i>	333.2%	0.0352%	0.7753	<b>0.0454%</b>
<i>CashMarket4</i>	263.8%	0.0310%	0.7271	0.0426%
<i>Market5</i>	109.9%	0.0240%	1.2462	0.0193%
<i>Value6</i>	65.3%	0.0134%	0.6774	0.0198%
<i>Momentum7</i>	57.0%	0.0149%	1.0171	0.0146%

Table 5: The results for the different strategies considered, we can see that the best performing strategy is the online updated efficient frontier which uses the time varying parameters, and second best is the Cash/Market based strategy.

than the *Frontier1* strategy that does not take regimes into consideration. However, both *Frontier2* and *Frontier3* offers better absolute returns, and *Frontier3* has the overall best returns. This is a great performance of our model since *Frontier3* is the only strategy of the three that does not depend on future data.

Figure 12 shows how strategy *Frontier3* weights the efficient portfolio over time. As we can see it changes quite rapidly, at least in periods. The main explanation to this is probably due to volatility in the the filter probability time series, displayed in figure 13. Figure 14 shows how the model predicts the risk exposure at each time step. As one would expect there seems to exist a high correlation between the risk exposure and the model's conviction of a high volatility regime, and we can also find a significant spike around the great financial crisis which started at the end of 2008.

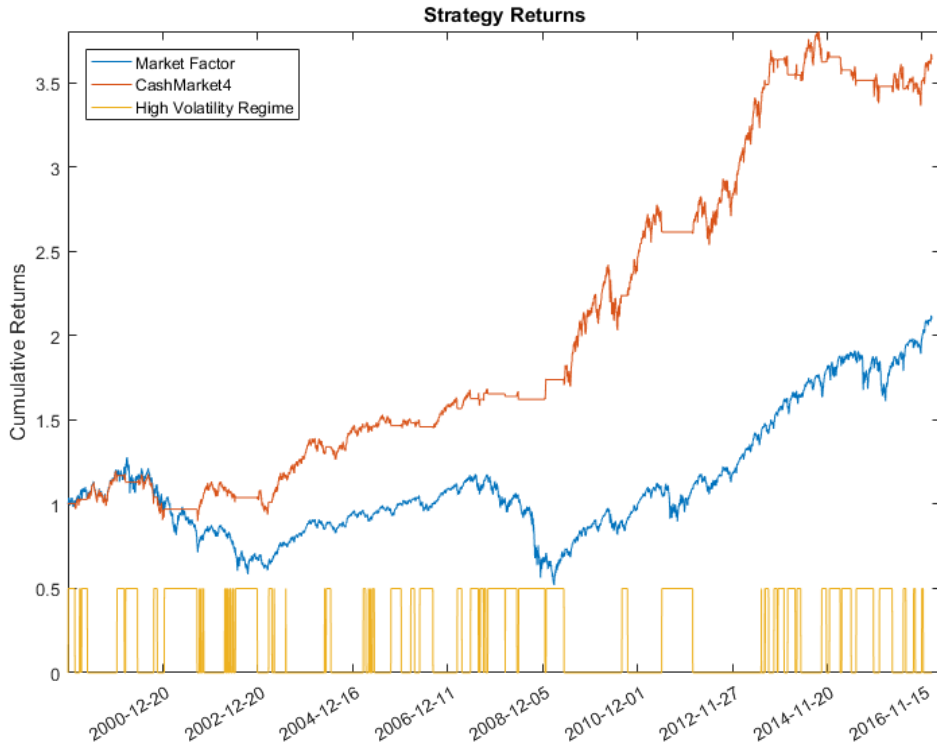


Figure 10: Cumulative returns over time for the 2 strategies and the corresponding predicted regime.

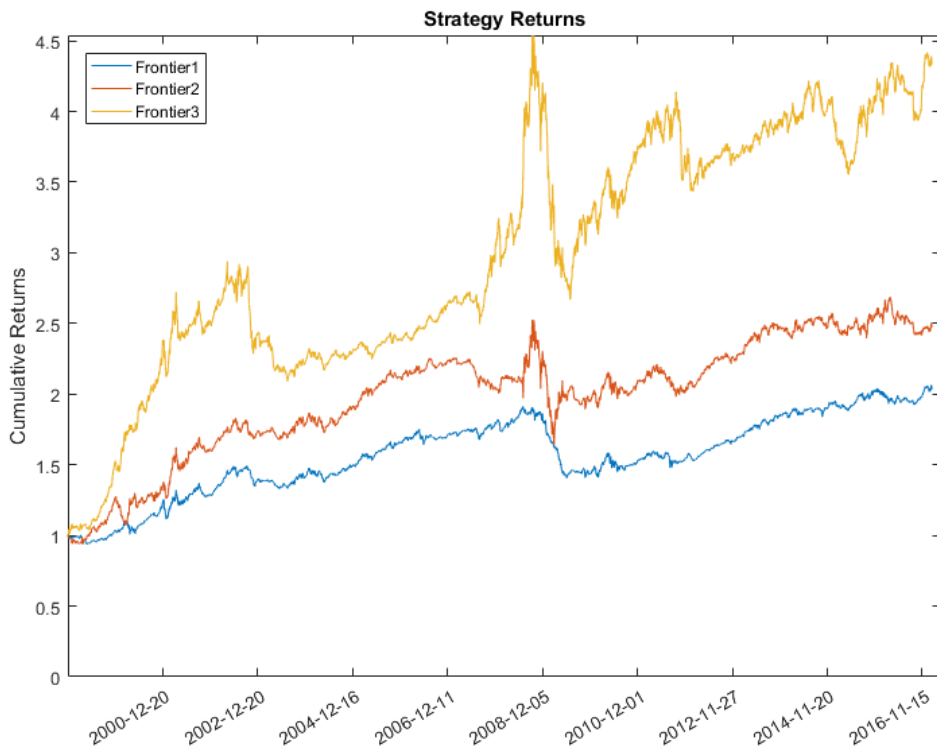


Figure 11: Cumulative returns over time for the 3 efficient frontier strategies.

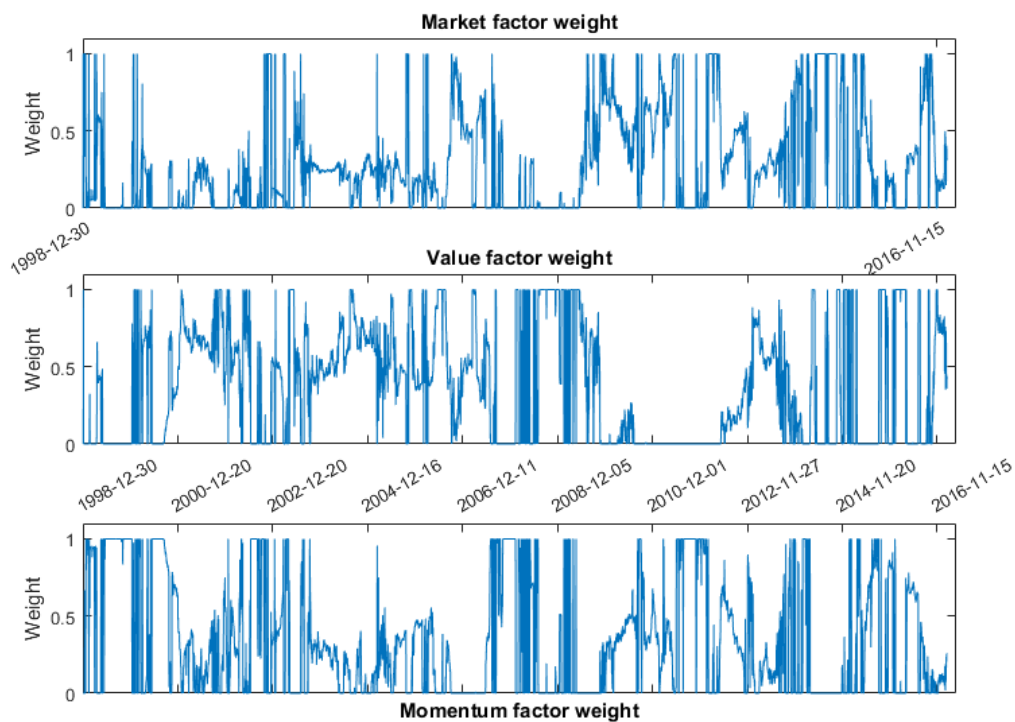


Figure 12: The weights over time for the efficient frontier denoted Frontier3.

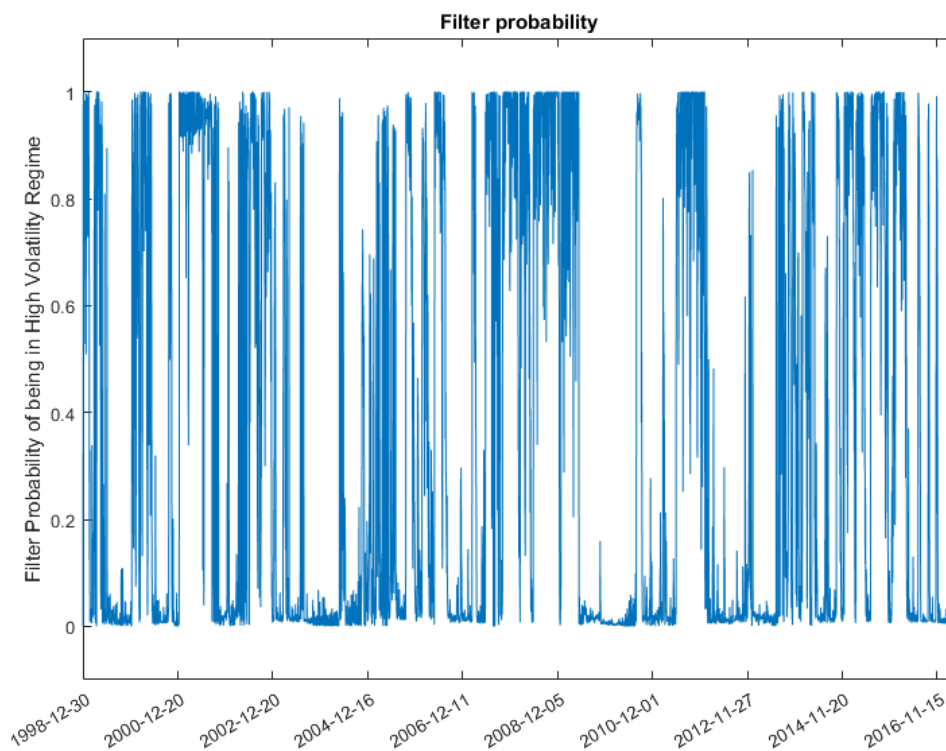


Figure 13: The filtered probability of being in the high volatility regime.

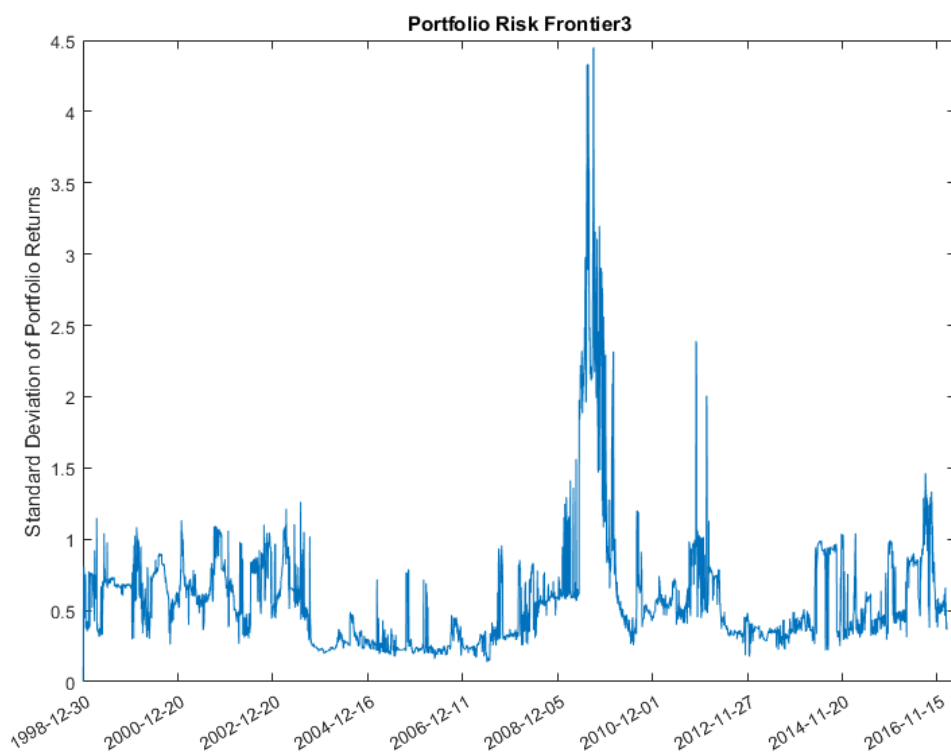


Figure 14: The estimated portfolio risk predicted by the strategy Frontier3.

## 7 Discussion and Conclusions

### 7.1 Model approach

The results in this thesis is inline with the promising results from previous papers. The results from section 4 indicates that there exists two clear market regimes where the risk factors have significantly different behaviour, both in terms of variance and expected returns but also in their correlation. Especially promising from an investor's point of view is the expected return of value in the high volatility regime where it offers almost as high expected return as the expected return for the market and the momentum factor in the low volatility regime. This result should imply great potential for investors to increase their returns in both low and high volatility markets.

However, section 4 also highlighted a difficulty when considering multidimensional normally distributed log-returns in the *HMM* setting, which was regime instability. Even though the model only was a 4 factor model we had to reduce it in order to get stable results. This unfortunately indicates a great difficulty if one wants to incorporate more factors or indices in a regime model. However, we found improved stability when the shrinkage approach was applied in the 3 factor model, and this should be a strong tool that can be applied in future multidimensional *HMM* approaches.

### 7.2 Online application

For the model to fulfill its purpose it must be applicable in an online application. We found that the model could outperform static strategies and non regime based strategies. The highly applicable strategy of investing in cash or the market indicated a significantly better performance than a static strategy with a constant exposure towards the market factor.

In the more theoretical application of the efficient frontier which demands daily re-weighting we also saw a better performance of the online application of the model even though the non-regime based efficient frontier was estimated with all future values known. However, the relatively poor performance of the regime based efficient frontier with weights calculated over the whole set of data (denoted *Frontier2*) indicates that our online approach fails to capture all the strengths found in the regime evaluation in section 4.

Jet another indications of the online application failing to capture all the regime benefits from the modeling of the whole data set is that the model predicted almost as much time in the high volatility regime as in the low volatility regime, while the steady state probabilities from equation (31) indicates that it should predict the high volatility regime for 26% of the time.



Possibly this performance can be increased with optimization of the regime shifting weights,  $w_1$  and  $w_2$  from the regime decision scheme presented in section 5 and expression (32). Also different window lengths for the time varying parameters or exponential decreasing weights of the observations may improve the performance in the online application.

### 7.3 Student's $t$ -distribution

To use normally distributed log-returns is a common approach which simplify and enables efficient estimation of the model parameters. However, before modeling we saw in section 3.2 that the data we considered had considerably high kurtosis. The problem was improved after the model approach, but as Figure 6 shows the kurtosis are sill higher than the normal distribution indicates. This discrepancy from the assumed distribution could possibly be of great importance in the online application of the model.

We highlight this problem by simulating Student's  $t$ -distributed data from a two state  $HMM$ . We then estimate a 2 state  $HMM$  which assumes normally distributed observations on the simulated data set, and let the filtered probabilities estimate the states. Figure 15 shows the results. The implications of high kurtosis is simply and excessive model reaction to outliers which instantly makes the model to change regime.

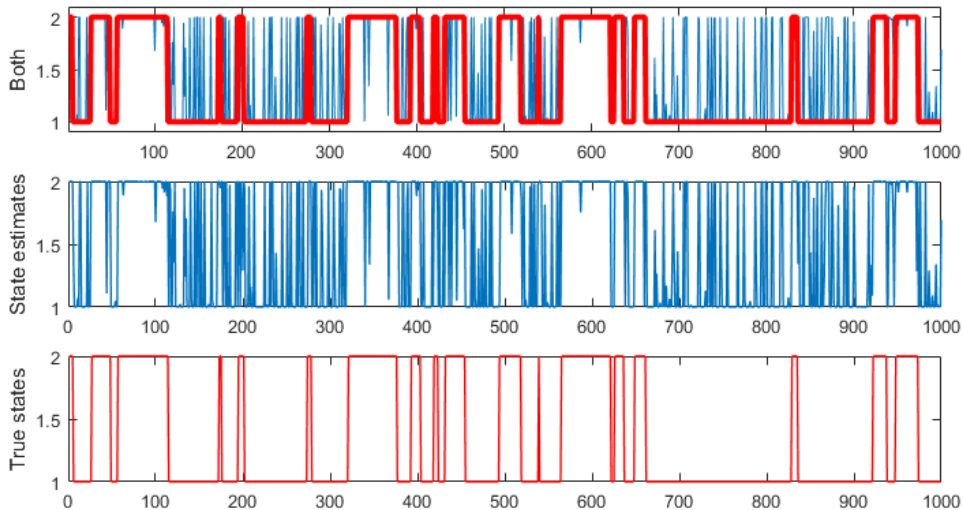


Figure 15: Simulated example of the Viterbi's path when a  $HMM$  which assumes normally distributed observations is fitted to  $t$ -distributed data.

The simulated data is done with a higher dimension number than the one used in this thesis, but there seems to be clear resemblance with both Figure 4 and Figure 13.

## 7.4 t-distribution in filter probabilities

Instead of using the pdf of the normal density function from equation (9) when the filter probabilities are calculated we will try to use a density function of a t-distribution.

$$p_i(\mathbf{x}_t) = \frac{\Gamma((v+d)/2)}{\Gamma(v/2) \cdot (v\pi)^{d/2} |\boldsymbol{\Sigma}_{i,t}|} \left( 1 + \frac{1}{v} (\mathbf{x}_t - \boldsymbol{\mu}_{i,t})^* \boldsymbol{\Sigma}^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_{i,t}) \right)^{-(v+d)/2} \quad (37)$$

$d$  is the dimension of the observations, i.e. 3. The parameters  $\boldsymbol{\Sigma}_{i,t}$  and  $\boldsymbol{\mu}_{i,t}$  are the same as the one estimated from the online application when the returns were assumed to be log-normal. That means that we do not estimate the parameters for the t-distribution but rather assess it to be sufficient to use the estimations from the normal distribution. However, the t-distribution also have a degree of freedom, parameter  $v$ . By an heuristic approach we set  $v = 13$ . We also estimate  $\boldsymbol{\Sigma}_{i,t}$  without any shrinkage.

By follow a similar procedure as in section 5 we get the weights to be  $w_1 = 0.8$  and  $w_2 = 0.8$ , and the model predicts 34% of the time in the high volatility regime, which is more in line with what we would expect from section 4. It also reduces the average trades per year slightly to 5.6. (that is around target from the Viterbi's path). However it does not enhance performance, it seems to have the same effect as the shrinkage. As Figure 16 shows the outcome is very similar when comparing the CashMarket Strategy.

## 7.5 In Summary

Our model approach in predicting factor behaviour has shown some promising results that is easy to apply in practice by investors how wants to manage their factor risk exposure in an optimal way. It also shows that the regime based model approach to market returns have significant impact on the efficient frontier and the value of risk over time.

We also found some problems in the online application of the model which possibly degraded the results partly and left some possible room for improvement in the application of the model. Never the less using our online 3 factor shrinkage model shows significant better results compared to following a static approach.

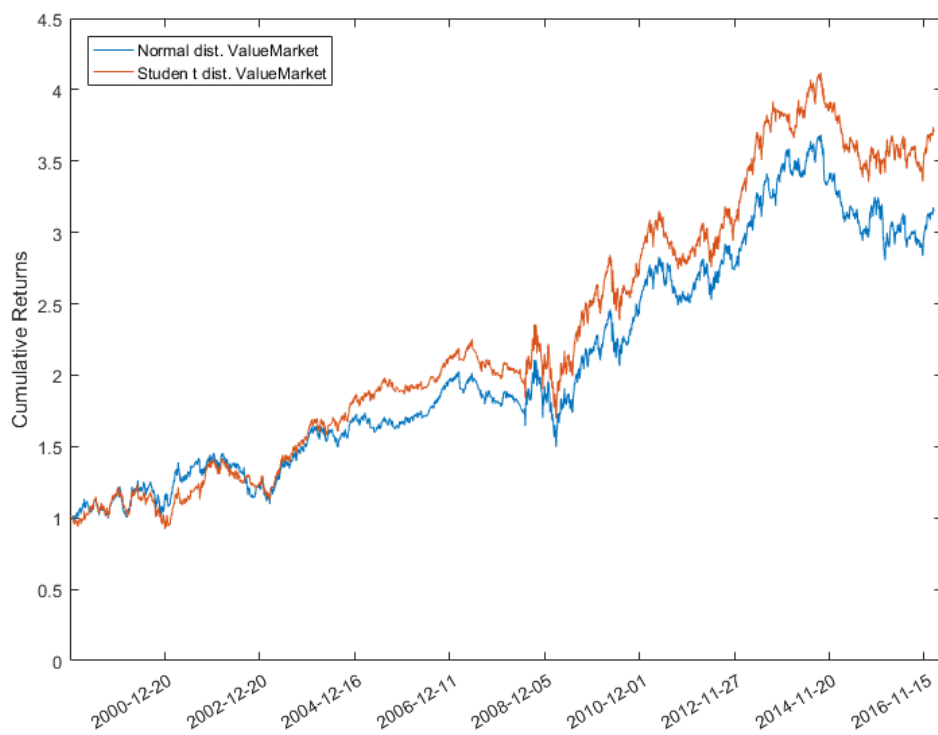


Figure 16: Comparing the strategy CashMarket14 when the returns are assumed to be log-normally distributed respective  $t$ -distributed.

## References

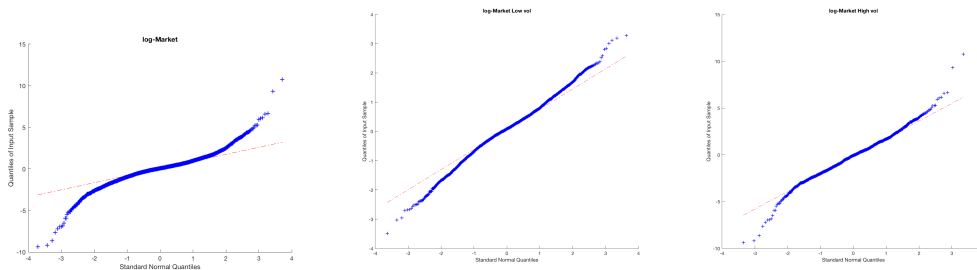
- [ ] *Kenneth R. French*. [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). Accessed: 2017-04-30.
- [AB02] Andrew Ang and Geert Bekaert. “International Asset Allocation With Regime shifts”. In: *The Review of Financial Studies* 15.4 ((2002)), pp. 1137–1187.
- [AB04] Andrew Ang and Geert Bekaert. “How do Regimes Affect Asset Allocation?” In: *Financial Analysts Journal* 60.2 (Mar. (2004)), pp. 86–99.
- [AG13] Noël Amenc and Felix Goltz. “Smart Beta 2.0”. In: *The Journal of Index Investing* 4.3 ((2013)), pp. 15–23.
- [AMP13] Clifford S. Asness, Tobias J. Moskowitz, and Lasse Heje Pedersen. “Value and Momentum Everywhere”. In: *The Journal of Finance* 3.69 (Sept. (2013)), pp. 929–985.
- [AT12] Andrew Ang and Allan Timmermann. “Regime Changes and Financial Markets”. In: *Annual Review of Financial Economics* 4.1 ((2012)), pp. 313–337.
- [AV06] Manuel Ammann and Michael Verhofen. “The Effect of Market Regimes on Style Allocation”. In: *Financial Markets Portfolio Management* 20.3 (Sept. (2006)), pp. 309–337.
- [Ban81] Rolf W. Banz. “The relationship between return and market value of common stocks”. In: *Journal of Financial Economics* 9.1 (Mar. (1981)), pp. 3–18.
- [Bha88] Laxmi Chand Bhandari. “Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence”. In: *The Journal of Finance* 43.2 (June (1988)), pp. 507–528.
- [Bil+98] Jeff A Bilmes et al. “A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models”. In: *International Computer Science Institute* 4.510 ((1998)), p. 126.
- [Bul+11] Jan Bulla et al. “Markov-switching Asset Allocation: Do Profitable Strategies Exist?” In: *The Journal of Asset Management* 12.5 ((2011)), pp. 310–321.
- [Car97] Mark M. Carhart. “On Persistence in Mutual Fund Performance”. In: *Journal of Finance* 52.1 ((1997)), pp. 57–82.
- [DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. “Maximum Likelihood from Incomplete Data via the EM Algorithm”. In: *Journal of the Royal Statistical Society* 39.1 ((1977)), pp. 1–38.
- [EG97] Edwin J Elton and Martin J Gruber. “Modern portfolio theory, 1950 to date”. In: *Journal of Banking & Finance* 21.11 ((1997)), pp. 1743–1759.
- [EHV94] Claude B. Erb, Campbell R. Harvey, and Tadas E. Viskanta. “Forecasting International Equity Correlations”. In: *Financial Analysts Journal* 50.6 (Nov. (1994)), pp. 32–45.

- [FF04] Eugene Fama and Kenneth French. “The Capital Asset Pricing Model: Theory and Evidence”. In: *Journal of Economical Perspectives* 18.3 ((2004)), pp. 25–46.
- [FF92] Eugene Fama and Kenneth French. “The Cross-Section of Expected Stock Returns”. In: *The Journal of Finance* 47.2 (June (1992)), pp. 427–465.
- [FF93] Eugene Fama and Kenneth French. “Common risk factors in the returns on stocks and bonds”. In: *Journal of financial economics* 33.1 ((1993)), pp. 3–56.
- [FF98] Eugene Fama and Kenneth French. “Value versus growth: The international evidence”. In: *The journal of finance* 53.6 ((1998)), pp. 1975–1999.
- [Fie+17] Mark Fiecas et al. “Shrinkage estimation for multivariate hidden Markov models”. In: *Journal of the American Statistical Association* 112.517 ((2017)), pp. 424–435.
- [Ham89] James D. Hamilton. “A New Approach to the Economic Analysis of non Stationary Time Series and the Business Cycle”. In: *Econometrica* 57.2 (Mar. (1989)), pp. 357–384.
- [Lin65] John Lintner. “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets”. In: *The Review of Economics and Statistics* 47.1 (Feb. (1965)), pp. 13–37.
- [LW04] Olivier Ledoit and Michael Wolf. “A well-conditioned estimator for large-dimensional covariance matrices”. In: *Journal of multivariate analysis* 88.2 ((2004)), pp. 365–411.
- [Mar52] Harry Markowitz. “Portfolio selection”. In: *The journal of finance* 7.1 ((1952)), pp. 77–91.
- [Mar59] Harry Markowitz. *Portfolio Selection, efficient diversification of investments*. Cowles Foundation for Research in Economics at Yale University, (1959).
- [NML16] Peter Nystrup, Henrik Madsen, and Erik Lindström. “Long Memory of Financial Time Series and Hidden Markov Models with Time-Varying Parameters”. In: *Journal of Forecasting* (2016).
- [Nys+15] Peter Nystrup, Bo William Hansen, et al. “Regime-Based Versus Static Asset Allocation: Letting the Data Speak”. In: *The Journal of Portfolio Management* 42.1 ((2015)), pp. 103–109.
- [Sha64] William F. Sharpe. “Asset Prices: A Theory of Market Equilibrium under Conditions of Risk”. In: *The Journal of Finance* 19.3 (Sept. (1964)), pp. 425–442.
- [Sha66] William F Sharpe. “Mutual fund performance”. In: *The Journal of business* 39.1 ((1966)), pp. 119–138.
- [Sud77] Basu Suddhasatwa. “Investment Performance of Common Stocks in Relation to Their Price-Earning Ratios: A test of the Efficient Market Hypothesis”. In: *The Journal of Finance* 32.3 (June (1977)), pp. 663–682.

- [Vit67] Andrew Viterbi. “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm”. In: *IEEE transactions on Information Theory* 13.2 ((1967)), pp. 260–269.

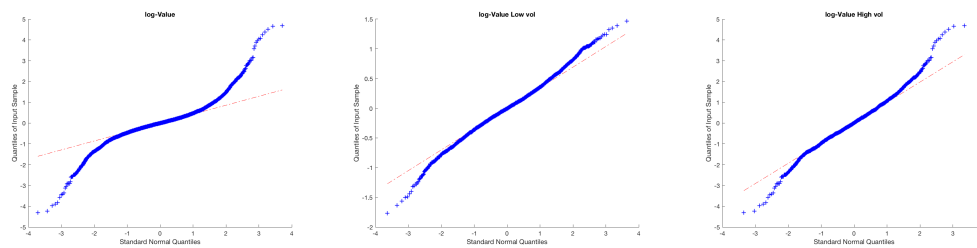
## 8 Appendices

### 8.1 QQ-plots



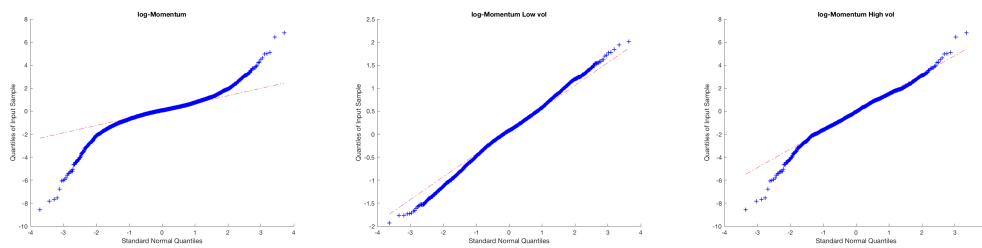
(a) The market factor without any model. (b) The market factor in the low volatility regime. (c) The market factor in the high volatility regime.

Figure 17: QQ-plots, for the market factor, quite satisfactory results after the model is applied



(a) The value factor without any model. (b) The value factor in the low volatility regime. (c) The value factor in the high volatility regime.

Figure 18: QQ-plots, for the value factor, quite satisfactory results after the model is applied.



(a) The momentum factor with- (b) The momentum factor in (c) The momentum factor in  
 out any model. the low volatility regime. the high volatility regime.

Figure 19: QQ-plots, for the momentum factor, quite satisfactory results after the model is applied.



## 8.2 Shrinkage weights

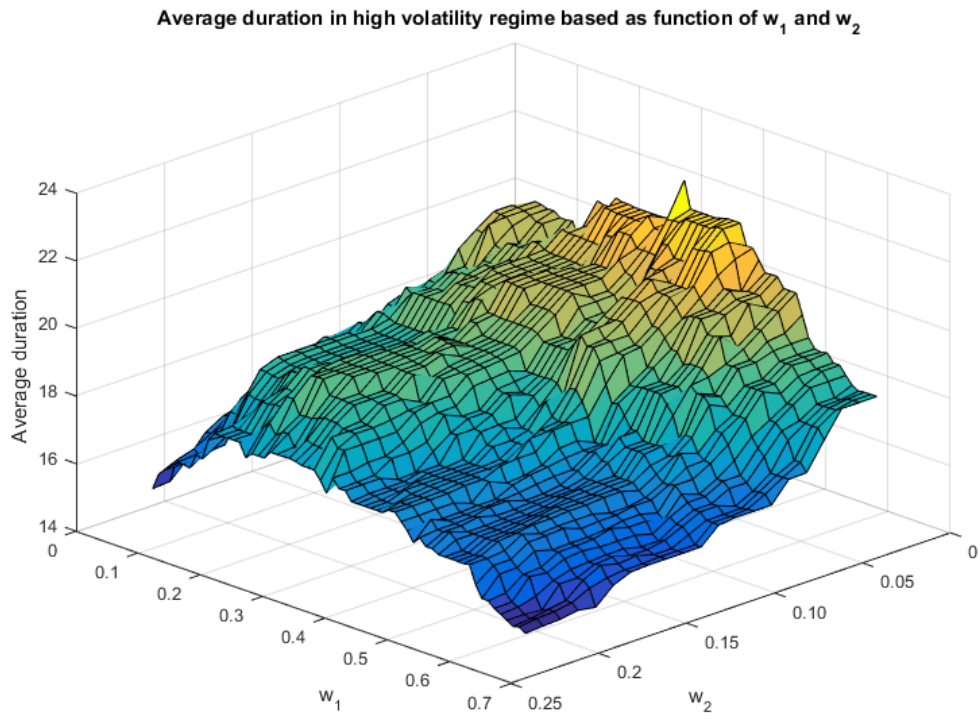


Figure 20: Different shrinkage weights and the corresponding average duration time in the high volatility regime.

### 8.3 Regime Performance

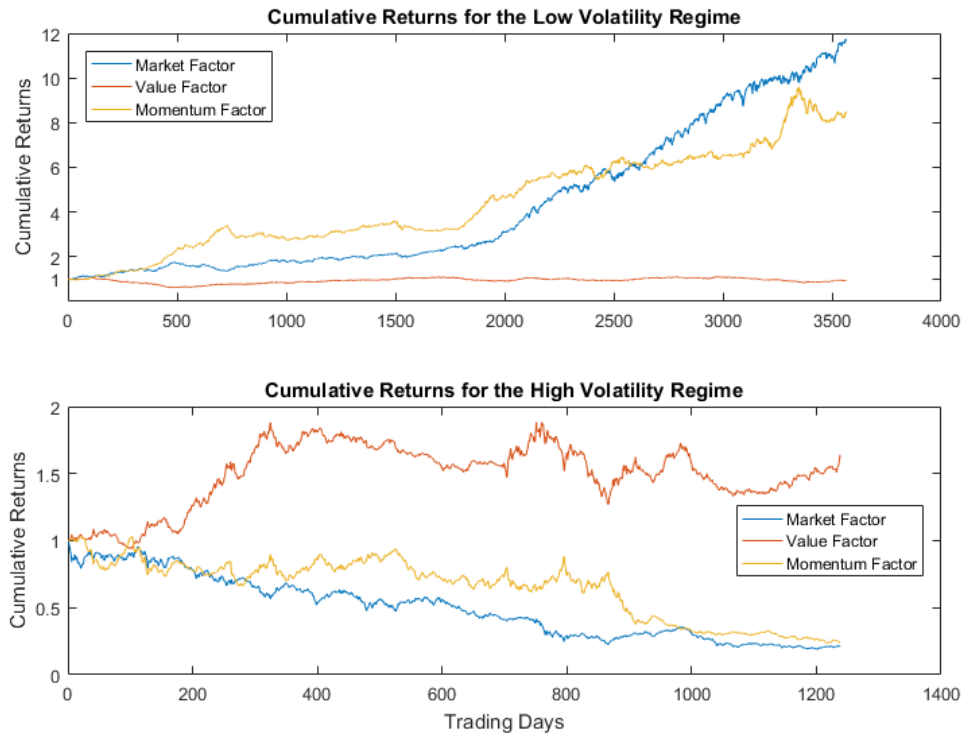


Figure 21: The performance of the different factors when the whole data set and the corresponding Viterbi's path is considered. The strong performance of the market and momentum factor in the low regime and the resistance of value in the high volatility regime indicates great potential for investment strategies.