Norwegian University of
Science and Technology

The Effects of Crude Oil on Stock Markets with use of Markov Switching Models

Thor August Mediaas Wiese

Master of Science in Physics and Mathematics
Submission date: September 2016
Supervisor: Jacob Laading, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Norwegian Univeristy of Science and Technology

The Effects of Crude Oil on Stock Markets with use of Markov Switching Models

Author:
Thor August Mediaas Wiese

Supervisor:
Jacob Laading

September 21, 2016

NORWEGIAN UNIVERISTY OF SCIENCE AND TECHNOLOGY

Abstract

Faculty of Information Technology, Mathematics and Electrical Engineering

Department of Mathematics

Master of Science

The Effects of Crude Oil on Stock Markets with use of Markov Switching Models

by Thor August Mediaas Wiese

In this thesis, a two regime Markov switching (MS) model is implemented to examine the relationship between crude oil, both brent oil and WTI, and stock markets. In particular, the model is applied to stock markets in both oil importing and exporting countries which include Canada, China, Japan, Germany, Netherlands, Norway, the United Kingdom and the United States. This paper first evaluates the significance of oil parameters in the detected regimes, where the two regimes respond to low mean and high variance (bear state), as well as high mean and low variance (bull state) respectively. We find evidence of stronger significance of oil returns in high volatility regimes. Overall, crude oil plays a significant role in determining stock returns. There is a stronger and more consistent relationship between oil and stock market in oil importing nations, regardless of regimes. The paper also presents an estimation of the correlation between oil and national indices for both regimes. The results provide further evidence of consistently higher correlation in high volatility regimes. The correlation ratio between the regimes are higher for oil importing nations, indicating that such nations are more strongly affected by volatility regimes.

Acknowledgements

The master's thesis was written as a completion of a master degree in Industrial Mathematics at the Norwegian University of Science and Technology. The work has been carried out at the Department of Mathematical Science during the spring semester 2016.

I would like to thank my supervisor Professor Jacob Laading for his patience and guidance, which has been greatly appreciated.

Contents

1. Introduction

1.1 Background of the Study

Crude oil (CO) is currently, and have been since the development of automation, the most important source of fuel in the world, and the worlds most traded commodity. Any news related to oil price are likely to affect the global economy. It is well documented that volatility shocks in crude oil have significant effects on several economic activities. Studies such as Hamilton [1996] have found empirical evidence on the affect of oil price changes on real economic activity, including inflation rates and Gross Domestic Product (GDP). And further Aloui and Jammazi [2009] and Jones and Kaul [1996] argue that stock price movements can be accounted for by the impact of crude oil shocks. As oil price shocks affect the expected net profitability of stock markets, it should be accounted for quickly into the stock markets. These conclusions are supported by many other papers such as Nandha and Faff [2008], Park and Ratti [2008]and Sadorsky [1999] amongst some. The main conclusion is that energy prices and oil prices in particular are likely to have affects on the stock price behaviour.

In more recent years, empirical studies have been focusing on structural breaks and shifts in stock market returns. History reveals that stock markets have experienced periods in which their behaviour change dramatically. Oil price increase in 1973-1974, stock market crash in 1984, the Asian economic crisis in 1997, the dot-com bubble in early 2000´s, 9/11 terrorist attacks, the Iraqi war in 2003 and the financial crises in 2007-2008 can be given as examples. All these events cause change in the dynamic process of financial time series and motivate the use of regime switching models. During the unstable economic time periods stock prices have followed with negative return and high volatility. Larger risk in stocks and changes in monetary polices during unstable periods would reduce the market efficiency, and is likely to affect the consistency in macro economic relationships. Accordingly the affect of oil price changes on stock market returns are likely to vary during time periods of recession or economic growth. The proposed time period from Jan-1996 to June-2016 has experienced several structure changes, and is therefore well suited for models account for structural changes. A Markov switching (MS) model is proposed to capture the regime shifts behaviour. It seeks to identify two distinct economic regimes within the markets. By incorporating such a model one are hopefully able to observe a relationship between oil price and market indices dependent on the economic regime. This regime dependent relationship between market indices and oil is evaluated in both oil-importing and oil-exporting countries.

A Markov switching model was first proposed by Hamilton [1989] and later further developed by Hamilton and Susmel [1994] as MS-ARCH[1] in connection to stock markets behaviour. In more recent time Bauwens et al. [2006], Henry [2009] and Aloui and Jammazi [2009] have all tried to model the behaviour of equity markets with MS-models, and their behavior towards other activa. However, regarding the effects of oil price on stock markets dynamics, the MS models applied to stock markets are very limited, with only a few papers such as Aloui and Jammazi [2009] and Hammoudeh and Choi [2007]. They both reveal that oil markets play an important role in explaining the stock market returns during changes in the time series. In summary it can be stated that the regime shifts are observed in stock market behaviour. This motivates us to check whether these shifts exist in some developed stock markets and to see if they are associated with price shocks and movements of the stock markets.

In this study we investigate the effects of CO on the regime shifts behaviour on nine developed markets: U.S. with NASDAQ and S&P 500, Japan, Germany, Netherlands, U.K., Canada and Shanghai over the sample period of January 1996 to May 2016. This time period consists of regime shifting time periods and should be well suited to establish how the CO affects the markets in high volatile periods, and also establish how the correlation between national indices and CO are depending on their regime. The next section will cover the research questions related to this thesis.

The main objective of this thesis is to investigate how the impact of crude oil returns affect and relate to the stock market returns in volatile time periods. Further along the way there are some sub-questions that need to be answered to make sure our models and estimations are reasonable.

1) Are Markov Switching models an appropriate way estimating financial time series? What strength and weaknesses do these models have?

Subjects related to the implementation and how the time series behave in a MS environment. Evaluation regarding the regime shifting and the distinction between the regimes will be considered.

2) What advantages does the t-distribution have compared to the normal distribution in a regime shifting framework?

The initial distributions will be implemented with both normal and t-distributed innovations, and an evaluation regarding these two distributions to our time series will follow.

[1] Auto regression conditional hetroskedasticity

3) How is the relationship between oil price returns and stock markets returns? Does this relationship vary depending on export/import in the two different regimes?

How does the relationship between markets and crude oil returns vary across regimes. An evaluation of parameter significance and correlation will be made based on the regimes.

The calculations and estimations have been done in MATLAB and R. The code used for the estimations has been based on Perlin [2015] and Visser et al. [2010], with some extended customization.

2. Theory

2.1 Financial Time Series

The analysis of financial time series has attracted the attention of both academics and practitioners for decades. The motivation and interest to describe the movement of financial time series do though vary. Researchers have studied the financial markets and later tried to replicate their movement with different stochastic models. Over the years the financial industry has approached the theoretical models to gain better understanding of the market, in hope of generate profit by forecasting movements. Financial modeling has thus become an important part of financial strategies, most commonly in hedge funds. Making modeling of financial time series a very relevant and interesting field of study. However, due to the complexity of the financial markets, modeling these time series is a very complicated problem. The challenges are mainly due to the statistical empirical findings, often referred to as stylized facts. A stylized fact is often a broad generalization that summarizes some complicated statistical calculations. These empirical findings are often difficult to replicate artificially with the use of stochastic models. Let´s consider a financial asset with price p_t at time t. The return can be calculated as the raw return

$$r_t = \frac{p_{t+1} - p_t}{p_t}. \tag{2.1}$$

Or the more commonly used log-return, which shall be used in this thesis. The use of log returns are for simplicity in calculations. With discrete returns less than 10%, the log return is a good approximation to discrete return, as it is the first order Taylor approximation[1].

$$r_t = \log(\frac{p_{t+1}}{p_t}). \tag{2.2}$$

[1] $r_t = log\frac{P_t}{P_{t-1}} = log(1 + R_t) = log(1) + R_t - \frac{R_t^2}{2!} + \frac{R_t^3}{3!} \approx R_t.$

The price and the return are closely related since the return determines the price and vice versa, but the two time series are very different both visually and with respect to their mathematical properties.

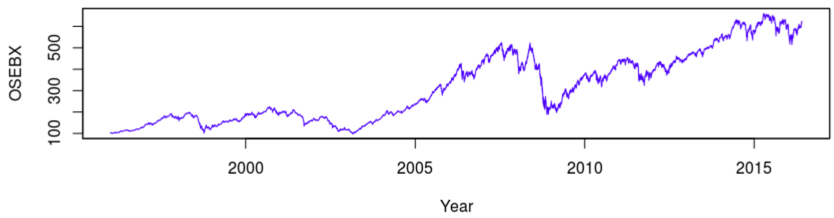


FIGURE 2.1: Value of OSEBX

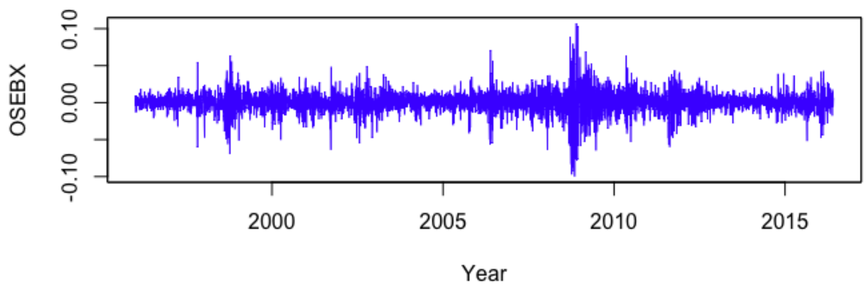


FIGURE 2.2: Daily return of OSEBX

Figure 2.1 and 2.2 display the daily index value and the return for the Norwegian stock market index, OSEBX, from January 1996 to June 2016. The index value draws resemblance to a random walk, on the other hand the returns could be related to a more stationary process with finite variance. The fluctuations in the the return series vary over time, and periods on the time series where there are large fluctuations observed is called clustering. The most obvious clustering is due to the financial crisis in 2008. The returns, regardless of the magnitude of the fluctuations, seem to be concentrated around zero. In one of the classic articles regarding financial data, Mandelbrot [1997] it is stated "...large changes tend to be followed by large changes - of either sign - and small changes tend to be followed by small changes". This statement certainly seems to fit well with the returns of OSEBX. In fact this statement will fit well with all of the time series introduced in this thesis. During this time period, different events will affect the time series with different magnitude. But in general the Asian financial crisis in the late 1990´s, dot-com bubble in the beginning of the millennium and the financial crisis in 2008 are all events that create fluctuations globally. The clustering of the time series makes the assumption of volatility dependency natural. By considering the squared returns and hence only looking at the magnitude of the fluctuations, a measure of the return volatility is the outcome

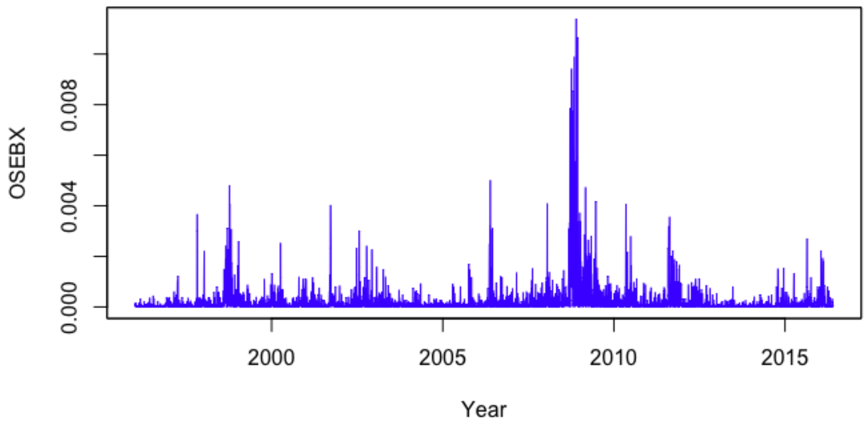


FIGURE 2.3: Daily squared return of OSEBX

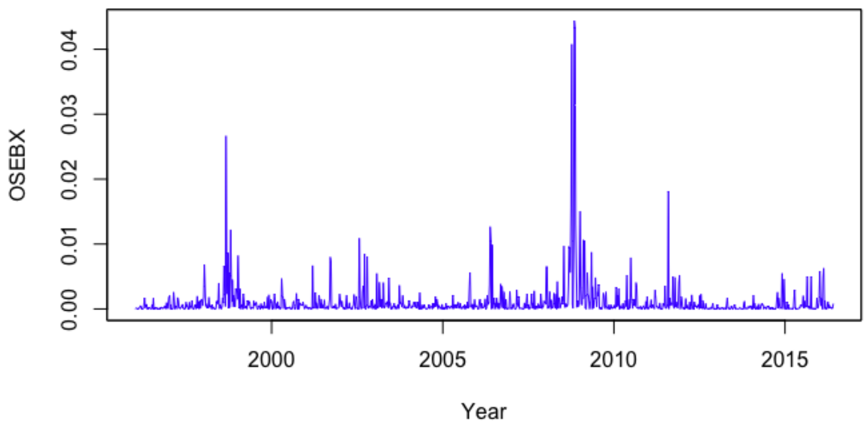


FIGURE 2.4: Weekly squared return of OSEBX

Figure 2.3 displays the squared returns of OSEBX. It is clear that periods with higher volatility reoccur during the time period, however, there is no system in the sudden change of variance. Note that in this thesis time series will be evaluated under weekly returns, and the main reason for this is that high frequency data often reflects a noise problem as stated in Ramchand and Susmel [1998]. This will be elaborated more in the next chapters. For comparison between the frequencies, figure 2.4 shows squared returns of OSEBX, and comparison to figure 2.3 reveals that the weekly plot has removed some of the "lone" spikes and the clustering periods are more distinct. On the other hand, low frequency data such as monthly or quarterly do not offer enough observation and would make analysis during crisis periods worthless as crises tend to live relativity short.

So far the assumption of volatility clustering seems reasonable for time series, and thus incompatible with marginal distribution with constant variance, or known as homoskedastic marginal distributions. A more realistic assumption is that the time series have time varying conditional variance, or known as hetroskedastic marginal distributions. Some of the popular hetroskedasticity models are more commonly known as ARCH (Auto Regressive Conditional Hetroskedasticity) or GARCH (Generalized Auto Regressive Conditional Hetroskedasticity). However, in this thesis we consider a model in between - a model with regime shifting parameters, but constant within its regime. We propose using a Markov Switching (MS) model to capture and predict asset returns, which will switch between a constant mean and variance depending on its regime, that is a high or low volatile period.

Examination of the auto correlation of both the returns and the squared returns will be made for all the time series in the preliminary analysis chapter, and a brief introduction on how they should be interpreted in financial time series will be explained. The expression "volatility clustering" indicates that there is serial dependence in the variance structure. Thus a process which displays volatility clustering should have significant auto correlation in the squared returns. Figure 2.5 and 2.6 display the auto correlation for OSEBX in both the returns and the squared returns. The returns series itself show no obvious sign of auto correlation, and makes it comparable to a white noise process. On the other hand, the squared returns auto correlation function shows very significant auto correlation in the data. As seen in the preliminary analysis chapter, this behaviour is common for all our time series. So, when looking past the direction of the return, and hence only on the magnitude of the returns, a serial dependence occur. This means that when choosing a model to replicate the return data, it must have a flexible variance structure, which makes it capable of capturing the dependence in the return variance.

Arguable the most common distribution in financial modeling is the normal distribution. Its popularity goes back to the beginning of modern financial modeling, and has been used in classic papers such as Fama [1965]. However, studies such as Haas and Pigorsch [2009] have also argued that the financial returns have a distribution with fatter tails and more peaked than the normal distribution. If a distribution inhabit these characteristics, it is called leptokurtic. A distribution is said to be leptokurtic if it has a kurtosis larger than three, which is what the normal distribution has. Figure 2.7 displays OSEBX standardized weekly returns along the density curves of a normal distribution and t-distribution with degrees of freedom $\nu = 5$. The plot illustrates that fatter tails of the t-distribution capture more extreme values compared to the normal distribution. However the fit is arguable, supporting the assumption of the observation to not come from the same distribution over the whole time period.

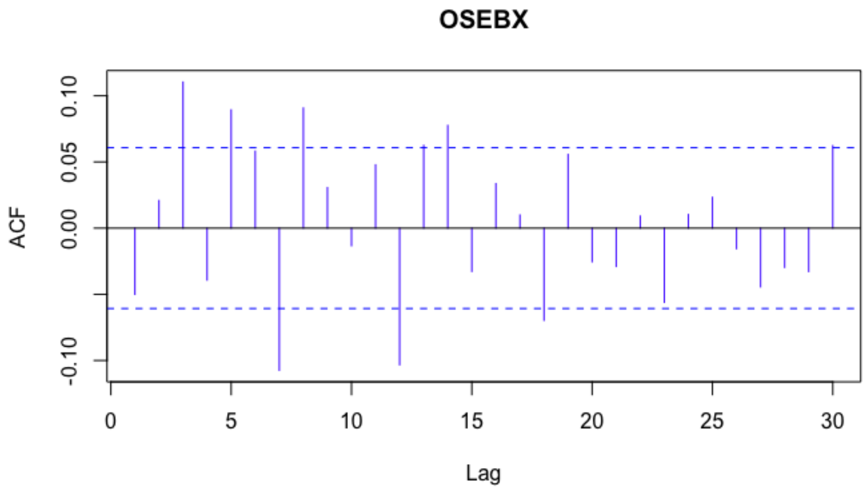


FIGURE 2.5: Auto correlation function of weekly returns of OSEBX

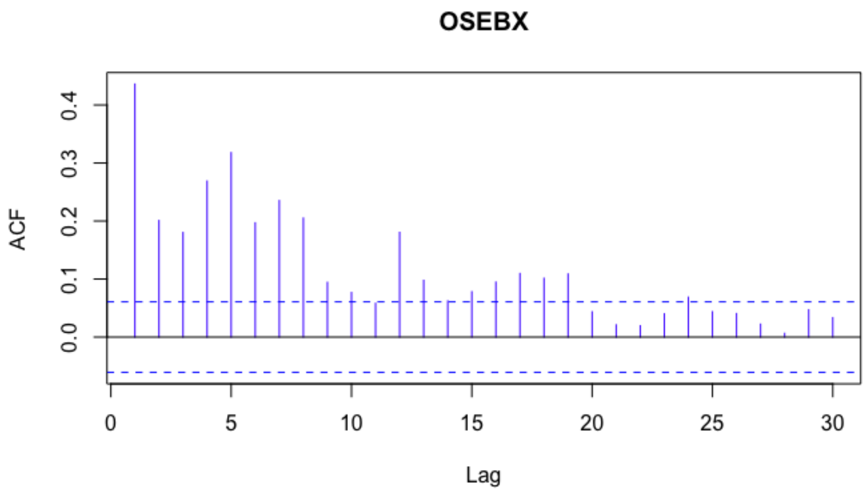


FIGURE 2.6: Auto correlation function of weekly squared returns of OS-EBX

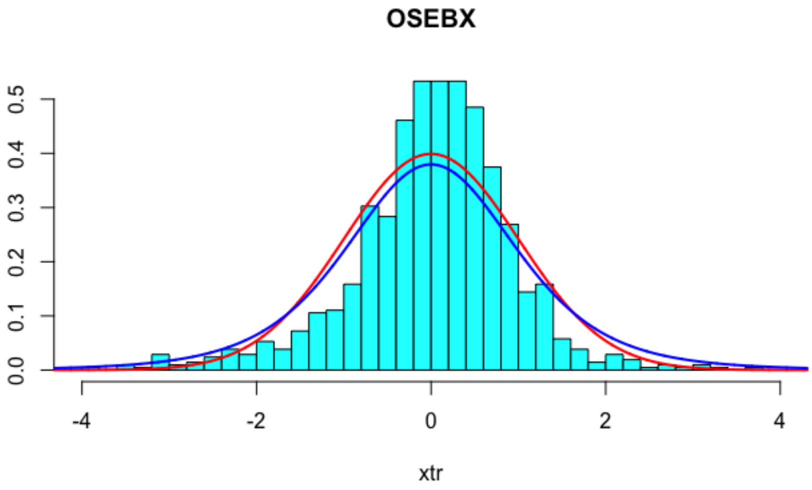


Figure 2.7: Distribution histogram of OSEBX returns from Jan-1996 to June-2016 together with a normal distribution (red line) and t-distribution with 5 degrees of freedom (blue line).

The variance structure and the distribution are two stylized facts about time series that make them very hard to replicate. A model aimed to capture the movement of financial time series must have these aspects in mind. The next chapter will cover the methodology on how to estimate the MS models, but first an introduction to some of the statistical tests done in this thesis.

2.2 Tests and Evaluation

This thesis uses a variation of statistical tests, and this section will go through some of the more classical statistical tests used. Some more specific tests applied will be explained in detail in the results section.

2.2.1 Jarque-Bera: Test for normality

In statistics, the Jarque-Bera test is a goodnes-of-fit [2] test of whether a data sample has the skewness and kurtosis matching a normal distribution. The test statistic JB is defined as

$$JB = \frac{n-k+1}{6}(S^2 + \frac{1}{4}(C-3)^2). \tag{2.3}$$

[2] It describes how well a set of observations fits a model

Where n is the number of observations, S is the sample skewness, C is the sample kurtosis, and k is the number of regressors. If the data comes from a normal distribution, the JB statistic has chi-squared distribution with two degrees of freedom. The null hypothesis is joint hypothesis of skewness being zero and excess kurtosis being zero, as in the normal distribution. Thus rejection of H_0 is rejection of the normality assumption.

2.2.2 Ljung-Box test: Testing auto correlation effects

The Ljung-Box test is a statistical test of whether any group of autocorrelations of time series are different from zero. Instead of testing randomness at each distinct lag, it tests the overall randomness based on a number of lags, and is therefore a portmanteau test [3]. The hypothesis is defined as: H_0 : "The data are independently distributed" and H_1 : "Not independently distributed". The test statistic is defined as the following

$$Q = n(n+1)\sum_{k=1}^{h}\frac{\hat{\rho_k^2}}{n-k}. \tag{2.4}$$

Where n is the sample size, $\hat{\rho}_k^2$ is the sample autocorrelation at lag k and h is the number of lags being tested. Under the H_0 the statistic Q follows a chi-squared distribution with h degrees of freedom. Note that it is applied to the standardized residuals of a fitted model, as in the MS models proposed.

2.2.3 Log-Likelihood ratio test: Comparison of two models

A likelihood ratio test is used to compare the goodness of fit of to models, one of which the null model is a special case of the alternative model. Each of the competing models is separately fitted to the data and log-likelihood recorded. The LLR statistic is defined as follows

$$D = 2 \times [\ln(\text{Likelihood alternative model}) - \ln(\text{Likelihood null model})].$$

The model with the most parameters, here the alternative, will always fit at least as well as the model with fewer parameters, here the null model. Whether the fit is significantly better and should be preferred is determined by its significance level. The test statistic is respectively chi-squared distributed with degrees of freedom equal to the difference of

[3] A test where the null hypothesis is well specified, but the alternative hypothesis is more loosely specified.

free parameters between the alternative and null model. The LRT has some limitations in the comparison of models in this thesis, therefore some alternatives to LRT will be introduced. Also the Akaike information criterion (AIC) and Bayesian information criterion (BIC) will be introduced as a supplement to the likelihood comparison. This will be further explained in the relevant chapters.

3. Methodology

This section introduces how our observed data is defined and further describe the general Markov process and the Hidden Markov Model. The papers Rabiner and Juang [1986] and Rabiner [1989] give a thorough introduction to Markov Processes, Hidden Markov Models and the main challenges with Markov switching models. Also how to create a HMM with use of computation and algorithms will be explained.

This thesis examines the relation between correlation and variance in a conditional time and state varying framework. This is done by evaluating at the correlation in low and high-volatile times of the crude oil returns with use of Hidden Markov Models (HMM). In comparison with a more simpler Markov Model, like a discrete Markov Chain, the states are visible to the observer. However, in HMM the states are not directly visible, but output, dependent on the state, is visible. In the next sections there will be an introduction to the parameters defining a HMM and the procedures to finding them. For simplicity there will first only be considered observations characterized as discrete, and therefore have a discrete probability density within each state of this model. E.g. by letting the observation to either be heads or tail of a coin. However, the data used is assumed to be drawn form a continuous probability density, and HMM estimation with such assumptions will also follow.

3.1 Markov Properties

The Hidden Markov Model gets its defining property from its name. First, the classification of the state is hidden from the observer. Secondly we assume our time series to inhabit the Markov property, that is the prediction of the next state only depends on the current state. Meaning that the state transition probabilities do not depend on the whole history of the past process. A Markov Process for a random discrete Variable X is defined the following way:

$$P(X_{t+1} = s_k \mid X_1, \ldots, X_t) = P(X_{t+1} = s_k \mid X_t). \tag{3.1}$$

Because the state transition is independent of time, the model is expressed in a transition matrix A:

$$a_{ij} = P(X_{t+1} = s_j \mid X_t = s_i) \tag{3.2}$$

Note that a_{ij} is a probability and hence

$$a_{ij} \geq 0, \quad \sum_{j=1}^{N} a_{ij} = 1 \qquad \forall i, j. \tag{3.3}$$

Further one also needs to know the probability to start in a state. In this case this probability is the same as the stationary distribution which tells us the proportion of time one spends in a certain state.

$$\pi_i = P(X = s_i). \tag{3.4}$$

Where $\sum_i^N \pi_i = 1$.

We will use these properties later to calculate the probability of state sequences which is defined the following way:

$$\begin{aligned} P(X \mid A, \pi) &= P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1, X_2) \cdots P(X_T \mid X_1, \ldots, X_{T-1}) && (3.5) \\ &= P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_2) \cdots P(X_T \mid X_{T-1}) && (3.6) \\ &= \prod_{t=1}^{T-1} \pi_1 P(X_{t+1} \mid X_t) && (3.7) \end{aligned}$$

3.1.1 Hidden Markov models

The general Markov Model described in the previous section has limited power in many applications. Therefore we extend it to a model with greater representation power; the Hidden Markov Model. In a HMM one does not know anything about what generates the sequence. The number of states, the transition probabilities, and from which state an observation is generated are all unknown. Instead of combining each state with a deterministic output, in our case high and low volatility, each state of the HMM is associated with a probabilistic function. At time t, and observation o_t is generated by function $b_j(o_t)$, which is associated with state j, with the probability:

$$b_j(o_t) = P(o_t \mid X_j). \tag{3.8}$$

Unlike the Markov Chain, the HMM will pick a strategy to follow also based on the observation probability. The key to the HMM is to pick the best overall sequence of strategies based on an observation sequence. The methods for finding the optimal chain of states connected to the time series will be introduced later in the chapter.

3.1.2 General form of HMM

A HMM is composed of five factors (S, K, Π, A, B).

1) $I = \{1, \ldots, N\}$ the set of states. The state of time t is denoted as s_t. In our HMM problem we have two states, thus $N = 2$.

2) $K = k_1, \ldots, k_M$ is the output alphabet and M is the number of observation choices. In our HMM, the output is continuous, since it is represented as a return rate. And therefore define the output alphabet as $K \in \Re$, where it accept any observation if it is a Real number.

3) Initial state distribution $\Pi = \pi_i, i \in S$ is defined as

$$\pi_i = P(s_1 = i) \tag{3.9}$$

4) State transition probability distribution $A = \{a_{ij}\}, i, j \in S$.

$$a_{ij} = P(s_{t+1} = j \mid s_t = i), 1 \leq i, j \leq N \tag{3.10}$$

5) Observation symbol probability distribution $B = b_j(o_t)$. The probabilistic function for each state j is:

$$b_j(o_t) = P(o_t \mid s_t = j) \tag{3.11}$$

This thesis assumes the time-series to be HMM's, generated by such a model explained here. Thus our data of returns, Y, is our observation sequence $O = \{O_1, O_2, \ldots, O_T\}$. By doing this one is able to calculate the probabilities of the observation sequence, B, and thus calculate the most probable underlying state sequence S'.

3.1.3 Observation probability distribution

In the discrete observation probability distribution there is a fixed number of possible observations o_t. The probabilistic function $b_j(o_t)$ is discrete in each state j. We could for example partition our observations of returns, O, into discrete observations {large drop, small drop, no change, small rise, large rise}- this partitioning is called vector quantization. The observation probability density function will be on the form

$$b_j(t) = P(o_t = k_i \mid s_t = j), \quad 1 \leq i \leq M. \tag{3.12}$$

As briefly mentioned, time-series and observations are assumed to be continuous distributed, this is natural since the rate of return will not take any fixed value. In comparison to a discrete case where the observations for example can be either tail or heads, the rate of return has no fixed value. Thus in continuous observation probability HMM, the function $b_j(o_t)$ is in the form of a continuous probability density function or a mixture of cont. pdf's:

$$b_j(o_t) = \sum_{k=1}^{M} w_{jk} b_{jk}(o_t), \qquad j = 1, \ldots, N \tag{3.13}$$

M is the number of mixtures and w is the weight for each mixture. Recall that N is the number of possible states. The mixture weights have the following constraints:

$$\sum_{k=1}^{M} w_{jk} = 1, \qquad w_{jk} \geq 0, \quad j = 1, \ldots, N \quad k = 1, \ldots, M \tag{3.14}$$

This paper for instance use a Gaussian and t-distributed HMM, so that our two states are described with respect to the chosen distribution. However a generalization of the density function could have been used.

$$b_j(o_t) = \mathcal{N}(o_t; \mu_j, \sigma_j). \tag{3.15}$$

An alternative way, and probably more common in the description of Markov Switching models is the following parametrization:

$$o_t = \mu_j + \epsilon_t \tag{3.16}$$
$$\epsilon_t \sim \mathcal{N}(0, \sigma_j^2). \tag{3.17}$$

When use of pairwise observations in multivariate HMM, the observations are multivariate Gaussian distributed:

$$\vec{b_j}(o_t) = \mathcal{N}(o_t; \vec{\mu}_j, \Sigma_j). \tag{3.18}$$

Or alternatively

$$\vec{o_t} = \vec{\mu_j} + \vec{\epsilon_t} \tag{3.19}$$
$$\vec{\epsilon_t} \sim \mathcal{N}(0, \Sigma_j). \tag{3.20}$$

Where $\vec{\mu}_j$ is a vector of means, and Σ is the covariance matrix.

3.2 Three Fundamental Problems for HMM

There are three fundamental problems related to HMM.

1) Given a HMM model $\lambda = (A, B, \Pi)$ and a observation sequence $O = (o_1, \ldots, o_T)$, how do we efficiently compute the probability of the observation sequence given the model? In other words, how do we find $P(O \mid \lambda)$?

2) Given a HMM model λ and the observation sequence O, what is the underlying state sequence S' that best "explains" the observations?

3) Given an observation sequence O, and a space of model parameters, how do we adjust the model parameters $\lambda = (A, B, \Pi)$ to maximize $P(O \mid \lambda)$.

The first problem can be used to determine which of the trained models is most likely to fit the model when the observation sequence is given. That is which model best accounts for the movement history. The second problem is about finding out the hidden path. And the third problem is about fitting the models with the historical data.

When going through the algorithms in the next sections, we try to keep the procedures simplified and consider our observations to be discretely distributed. The transition to model HMM with continuous observation will be explained later.

3.2.1 Problem 1: Finding the probability of a observation

Given the observation sequence $O = (o_1, \ldots, o_T)$ and a HMM $\lambda = (A, B, \Pi)$, we want to find the probability of the sequence $P(O \mid \lambda)$. Since the observations are independent of each other, the probability of a state sequence $S = (s_1, \ldots, s_T)$ generating the observation sequence can be calculated as:

$$
\begin{aligned}
p(O \mid S, \lambda) &= \prod_{t=1}^{T} P(o_t \mid s_t, s_{t+1}, \lambda) && (3.21)\\
&= b_{s1}(o_1) \cdot b_{s_2}(o_2) \cdots b_{s_T}(o_T). && (3.22)
\end{aligned}
$$

And the state transition probability,

$$
P(S \mid \lambda) = \pi_{s_1} \cdot a_{s_1 s_2} \cdot a_{s_2 s_3} \cdots a_{s_{T-1} s_T}. \tag{3.23}
$$

And the joint probability of O and S is given by:

$$
P(O, S \mid \lambda) = P(O \mid S, \lambda) P(S \mid \lambda). \tag{3.24}
$$

Which further imply:

$$
\begin{aligned}
P(O \mid \lambda) &= \sum_S P(O, S \mid \lambda) = \sum_S P(O \mid S, \lambda) P(S \mid \lambda) && (3.25)\\
&= \sum_S \pi_{s_1} \prod_{t=1}^{T} a_{s_{T-1} s_T} b_{s_T}(o_T). && (3.26)
\end{aligned}
$$

The computation is done by summing up the observation probabilities for each of the possible state sequences. Note here that $b_{s_t}(o_t)$ is the probability to generate value o_t given our current state s_t. The probability to transit from state i to state j is given by a_{ij}. However, the problem with this approach is to enumerate each state sequence. The computation grows exponentially and requires $(2T-1) \cdot N^{T+1}$ multiplications and $N^T - 1$ additions. To solve this problem one uses the forward-backward algorithm. It is an inference algorithm which computes the posterior marginals of all hidden state variables given a sequence of observations. The algorithm involves three steps:

1) Computing forward probabilities

2) Computing backward probabilities

3) Computing smoothed values(We will do this in problem 2).

Forward probabilities

The forward variable $\alpha_i(t)$ is defined as:

$$\alpha_i(t) = P(o_1, o_2, \ldots, o_{t-1}, s_t = i \mid \lambda). \tag{3.27}$$

Here $\alpha(t)$ describes the total probability of ending up in state i at time t given the observation sequence $o_1, \ldots, o_{t-1}$. The next steps of the algorithm can be solved inductively:

1)Initialization:

$$\alpha_i(1) = \pi_i b_i(o_1) \quad 1 \leq i \leq N \tag{3.28}$$

2) Induction

$$\alpha_j(t+1) = \left[\sum_{i=1}^{N} \alpha_j(t) a_{ij} \right] b_j(o_t) \quad 1 \leq j \leq N \quad 1 \leq t \leq T-1 \tag{3.29}$$

3)Update time, set $t = t + 1$, and return to step 2 if $t \leq T$

4) Termination

$$P(O \mid \lambda) = \sum_{i}^{N} \alpha_i(T). \tag{3.30}$$

The complexity of the forward algorithm is $O(TN^2)$.

Backward probabilities

The induction computation done for the forward probabilities can also be made in reverse order. The backward procedure calculates the probabilities of the partial observations from $t+1$ to the end, given the model mu and state s_t at time t. The backward algorithm is defined as:

$$\beta_i(t) = P(o_{t+1}, o_{t+1}, \ldots, o_T \mid s_t = i, \lambda) \tag{3.31}$$

This algorithm moves from right to left, if we consider our observations along an x-axis.

1)Initialization:

$$\beta_i(1) = 1 \quad 1 \leq i \leq N \tag{3.32}$$

2) Induction

$$\beta_j(t) = \left[\sum_{i=1}^{N} \beta_j(t+1) a_{ij}\right] b_j(o_{t+1}) \quad 1 \leq j \leq N \quad t = T-1, T-2, \ldots, 1 \tag{3.33}$$

3)Update time Set $t = t - 1$, and return to step 2 if $t \leq T$

4) Termination

$$P(O \mid \lambda) = \sum_{i}^{N} \beta_i(1). \tag{3.34}$$

This result will be used to solve problem 2.

3.2.2 Problem 2: Finding the best state sequence

The second problem is to find the best state sequence given a model and the observation sequence. There are several possible ways to solve this problem. The difficulty is that there may be several optimal criterias. One method is to choose the state that is individually most probable at time t. For each time t, $1 \leq t \leq T$, we find the related probability variable from our results from the Forward-Backward algorithm.

$$\gamma_i(t) = P(s_t = i \mid O, \lambda) \tag{3.35}$$

$$= \frac{P(s_t = i, O \mid \lambda)}{P(O \mid \lambda)} \tag{3.36}$$

$$= \frac{\alpha_i(t)\beta_i(t)}{\sum_{j=1}^{N} \alpha_j(t)\beta_j(t)} \tag{3.37}$$

The normalization factor $P(O \mid \lambda) = \sum_{j=1}^{N} \alpha_j(t)\beta_j(t)$ makes $\gamma_i(t)$ a probability measure so that

$$\sum_{i=1}^{N} \gamma_i(t) = 1 \tag{3.38}$$

These probabilities are known as the "smoothed" probabilities, where the forward and backwards probabilities are combined.

Further, the individually most likely sequence S' can be found as:

$$S' = \underset{1 \le i \le N}{\operatorname{argmax}}\, \gamma_i(t), \quad 1 \le t \le T, \quad 1 \le i \le N \tag{3.39}$$

However this approach may generate an unlikely state sequence. This is because it doesn't take transition probabilities into consideration, and may result in invalid state sequences, for example if a transition probability is equal to zero, $a_{ij} = 0$ at some point. Therefore, a more efficient algorithm is the Viterbi algorithm. It is based on dynamic programming and is used to find the best state sequence.

The Viterbi algorithm

The Viterbi Algorithm is designed to find the most likely state sequence, $S' = (s_1, s_2, \ldots, s_T)$ for the observation sequence $O = (o_1, o_2 \ldots, o_T)$:

$$\underset{S'}{\operatorname{argmax}}\, P(S' \mid O, \lambda). \tag{3.40}$$

To solve this problem one needs to define the quantity

$$\delta_j(t) = \max_{s_1 s_2 \cdots s_{t-1}} P(s_1 s_2 \cdots s_{t-1}, o_1 \ldots o_2, s_t = j \mid \lambda). \tag{3.41}$$

This variable stores the probability of observing $o_1 o_2 \cdots o_t$ using the most likely path that ends in state i at time t, given model λ. To retrieve the state sequence one needs to keep track of the argument which maximized (3.41) for each t and j. For this we use the array $\psi_i(t)$, which stores the node/state that leads to the most probable path. The calculation is done by recursion, it is similar to the one in Forward-Backward algorithm, except that the forward-backward algorithm uses summing over previous states, while the Viterbi algorithm uses maximization. The recursion comes from the call in the previous expression within our max operator, and so it goes all the way back to the initial condition. The complete procedure for finding the best state sequence can now be stated as follows:

Initialization:

$$\delta_i(t) = \pi_i b_i(o_1), \quad 1 \leq i \leq N \tag{3.42}$$

$$\psi_i(1) = 0 \tag{3.43}$$

Recursion:

$$\delta_j(t) = \max_{1 \leq i \leq N} \left[\delta_i(t-1) a_{ij}\right] b_j(o_t), \qquad 2 \leq t \leq T \tag{3.44}$$

$$1 \leq j \leq N \tag{3.45}$$

$$\psi_j(t) = \operatorname*{argmax}_{1 \leq i \leq N} \left[\delta_i(t-1) a_{ij}\right], \qquad 2 \leq t \leq T \tag{3.46}$$

$$1 \leq j \leq N \tag{3.47}$$

Termination:

$$P^* = \max_{1 \leq i \leq N} \left[\delta_i(T)\right] \tag{3.48}$$

$$s_T^* = \operatorname*{argmax}_{1 \leq i \leq N} \left[\delta_i(T)\right] \tag{3.49}$$

Path backtracking:

$$s_t^* = \psi_{t+1}(s_{t+1}), \quad t = T-1, T-2, \ldots, 1 \tag{3.50}$$

Thus our optimal sequence is $S^* = \{s_1^*, \ldots, s_T^*\}$.

3.2.3 Problem 3: Parameter estimation

The last and most difficult problem about HMM's is to determine a method to adjust the model parameters (A, B, Π) to maximize the probability of the observation sequence given the model. The problem can be reformulated as to find the parameters that maximize the following probability:

$$\underset{\lambda}{\operatorname{argmax}} P(O|\lambda).$$

There is no known analytic method to choose to maximize $P(O|\lambda)$. In fact, for any finite observation set, there is no optimal way of estimating the model parameters. The goal is to maximize the observation probability, $P(O|\lambda)$, by choosing appropriate λ. This paper has evaluated two methods to estimate the parameters. First method used is called the Baum-Welch, which is a local maximization algorithm to obtain the maximum likelihood estimation of the observed data, without computation of the likelihood. This is a special case of Expectation Maximisation (EM) method. It works iteratively to improve the likelihood of $P(O \mid \lambda)$. Secondly the parameters have been estimated by maximum likelihood. Both methods have been applied for comparison, however, only the estimations of the maximum likelihood have been used in this paper. Nevertheless both estimations methods are described in the following sections.

3.2.4 Baum-Welch algorithm

In order to describe the procedure for re-estimation (iterative update and improvement) of HMM parameters, one first defines $\xi_{i,j}(t)$, the probability of being in state i at time t and j at time $t+1$, given the model and observation sequence.

$$\xi_{i,j}(t) = P(s_t = i, s_{t+1} = j \mid O, \lambda). \tag{3.51}$$

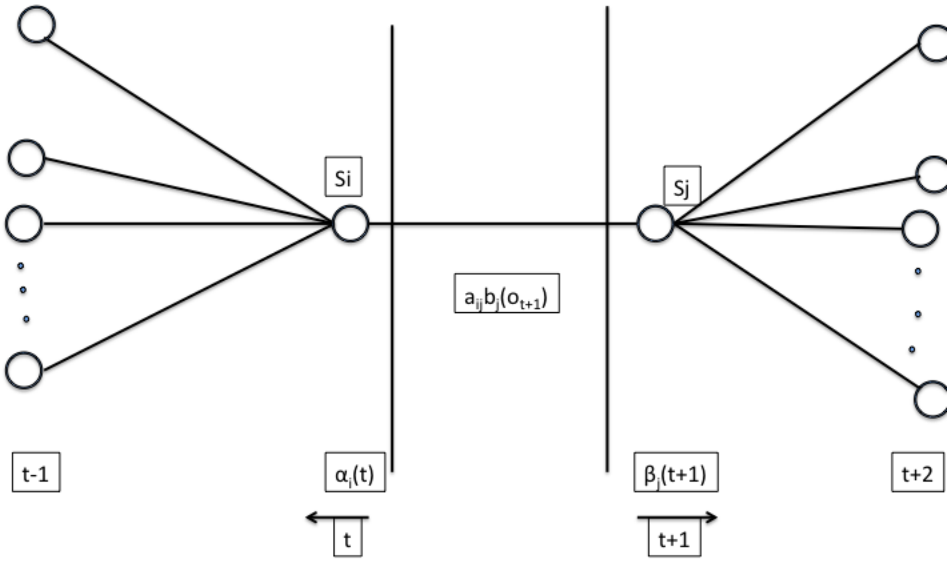


FIGURE 3.1: Illustration of the sequence of operations required for the computation of the joint event that the system is in state i at time t abd state j at time $t+1$

The sequence of events leading to the conditions required by (3.51) is illustrated in figure 3.1. From the definition of the forward and backward variables we write $\xi_{i,j}(t)$ in the form

$$\xi_{i,j}(t) = P(s_t = i, s_{t+1} = j \mid O, \lambda) \tag{3.52}$$

$$= \frac{P(s_t = i, s_{t+1} = j \mid O, \lambda)}{P(O \mid \lambda)} \tag{3.53}$$

$$= \frac{\alpha_i(t) a_{ij} b_j(o_t) \beta_j(t+1)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i(t) a_{ij} b_j(o_{t+1}) \beta_j(t+1)}. \tag{3.54}$$

In the previous section $\gamma_i(t)$ was defined as the probability of being in state i at time t, given observation sequence and the model. And thereby can relate $\gamma_i(t)$ to $\xi_{ij}(t)$ by summing over j, giving

$$\gamma_i(t) = \frac{\alpha_i(t)\beta_i(t)}{\sum_{j=1}^{N}\alpha_j(t)\beta_j(t)} = \sum_{j=1}^{N}\xi_{ij}(t) \tag{3.55}$$

$$= \frac{\sum_{j=1}^{N}\alpha_i(t)a_{ij}b_j(o_t)\beta_j(t+1)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i(t)a_{ij}b_j(o_{t+1})\beta_j(t+1)}. \tag{3.56}$$

By summing $\gamma_i(t)$ over the time index t, this value can be interpreted as the expected number of times state i is visited, or alternatively, the expected number of transitions from state i if one considers the grid from $t = 1$ to $t = T - 1$. Also, the summation of $\xi_{ij}(t)$ from $t = 1$ to $t = T - 1$ can be interpreted as the expected number of transitions from state i to state j. That is

$$\sum_{t}^{T-1}\gamma_i(t) = \text{Expected number of transitions from state } i \tag{3.57}$$

$$\sum_{t}^{T-1}\xi_i(t) = \text{Expected number of transitions from state } i \text{ to } j \tag{3.58}$$

Using the equations above one can create a method of re-estimation of the parameter of the HMM. The re-estimations formulas for Π, A and B are

$$\pi_i' = \text{Expected number of times in state } i \text{ at time } t = 1 = \gamma_i(1) \tag{3.59}$$

$$\begin{aligned} a_{ij}' &= \frac{\text{Expected number of transitions from state } i \text{ to state } j}{\text{Expected number of transitions from state } i} \\ &= \frac{\sum_{t=1}^{T-1}\xi_i(t)}{\sum_{t}^{T-1}\gamma_i(t)} \end{aligned} \tag{3.60}$$

$$\begin{aligned} b_j'(o_t) &= \frac{\text{Expected number of times in state } j \text{ observing } o_t = k}{\text{Expected number og times in state } j} \\ &= \frac{\sum_{t}^{T-1}{}_{o_t=k}\gamma_j(t)}{\sum_{t}^{T-1}\gamma_j(t)} \end{aligned} \tag{3.61}$$

Given the definitions above one begins with an initial model $\lambda = (A, B, \Pi)$ and runs the data O through the current model to estimate the expectation of each model parameter $\lambda' = (A', B', \Pi')$. As mentioned the Baum-Welch method is indeed an implementation of the EM method. The equations (3.57)-(3.58) can be used to calculate the expected values of transition properties of the Markov Chain, the expectation step of EM-algorithm. Further the maximum likelihood estimation of the model is computed through recursion of equations (3.59)-(3.61), the maximum step of the EM-algorithm. By iteratively use λ' instead of λ, and repeat re-estimation, we can improve the probability of O being observed from the model until some limiting point is reached. The final result of this procedure is called a maximum likelihood estimate of the HMM.

3.2.5 Maximum likelihood estimation

The likelihood is of high importance, since the model is estimated on the idea of maximizing the likelihood. Both the Baum-Welch and the MLE have the intention to maximize the likelihood function, however with two different approaches. Before one can use the MLE method, one needs to derive the likelihood function. This section derives the likelihood function, and its dynamics. The method of deriving the likelihood function is similar to the methods used by Gray [1996].

Recall the regime switching model from equation (3.17)

$$o_t = \mu_{S_t} + \epsilon_t \tag{3.62}$$

$$\epsilon_{S_t} \sim \mathcal{N}(0, \sigma^2_{S_t}). \tag{3.63}$$

Thus the log likelihood of this model, given the normality assumption is given by

$$\ln L = \sum_{t=1}^{T} \ln\Big(\frac{1}{\sqrt{2\pi\sigma^2_{S_t}}} \exp\big(-\frac{o_t - \mu_{S_t}}{2\sigma^2_{S_t}}\big)^2\Big). \tag{3.64}$$

Here S_t is the state at time t, and μ_{S_t} and σ_{S_t} are the respective parameters belinging to state S_t. If one knew the state sequence, then one would maximize equation 3.64 as a function of $\mu_1.\mu_2, \sigma_1$ and σ_2. However, it should be clear by now that this is not the case for Markov switching models, where the state is in fact unknown. Thus in order estimate regime switching models, it is necessary to change the notation of the likelihood function. Before the construction of the log-likelihood function there are some measurements that need to be clarified, which are the conditional mean and variance.

The conditional mean is the constant mean depending on t and the regime. The mean can follow an Auto regression(AR) process or and Auto regression mean average (ARMA) process as seen in Gray [1996]. The conditional mean will be extended to where it includes oil returns and an associated coefficient, such as $\mu_{S_t} = a + \lambda_{S_t} r_{t\text{oil}}$, where a is the intercept value and$r_{t\text{oil}}$ is the oil return at time t . Here λ_{S_t} is a regime shifting parameter which also needs to be estimated.

The next element is the important conditional variance. The variance expression can be made a lot more complex by adding ARCH parameters for example, however in this paper the variance expression takes a simpler form where hetroskedasticity is not accounted for in the variance, but assumed to captured by the regime switching. The conditional variance is thus simply $\sigma^2_{S_t} = \sigma^2_{i,t}$, which is the variance at time t and $S_t = i$ is the responding regime at the current time. Note that the conditional mean and variance, take a value based on the current regime. However, the estimation process do not classify a regime, but estimates the probability of being in a regime. Thus the conditional mean and variance plotted will be a weighted average between the regimes.

The density of the returns conditioned on all observable history, $\tilde{o}_{t-1}$, is defined as $f(o_t \mid \tilde{o}_{t-1})$. And since o_t is set to switch between regimes it can be split up as

$$f(o_t \mid \tilde{o}_{t-1}) = \sum_{i=1}^{2} f(o_t \mid \tilde{o}_{t-1}, S_t = i) \cdot Pr(S_t = i \mid \tilde{o}_{t-1}) \quad = \sum_{i=1}^{2} f(o_t \mid \tilde{o}_{t-1}, S_t = i) \cdot p_{i,t}. \tag{3.65}$$

The notation $p_{i,t} = Pr(S_t = i \mid \tilde{o}_{t-1})$ is introduced for simplicity. Equation (5.16) is the weighted average of the two regime dependent density functions, weighted with the probability of being in regime i at time t, which is $p_{i,t}$. Further the equation of the conditional distribution of the returns can can written as follows

$$f(o_t \mid \tilde{o}_{t-1}) = \begin{cases} f(o_t \mid \tilde{o}_{t-1}, S_t = 1) & \text{w.p.} p_{1,t} \\ f(o_t \mid \tilde{o}_{t-1}, S_t = 2) & \text{w.p.} p_{2,t}. \end{cases} \tag{3.66}$$

Where the normality assumption yields

$$f(o_t \mid \tilde{o}_{t-1}, S_t = i) = \frac{1}{\sqrt{2\pi\sigma^2_{i,t}}} \exp\left(-\frac{o_t - \mu_{i,t}}{2\sigma^2_{i,t}}\right)^2. \tag{3.67}$$

Due to the Markov properties, $p_{i,t}$, is only dependent on the regime switching process from S_{t-1} to S_t at time $t-1$. Conditioning on the regime at time $t-1$ gives

$$p_{1,t} = Pr(S_t = 1 \mid \tilde{o}_{t-1}) = \sum_{i=1}^{2} Pr(S_t = 1 \mid S_{t-1} = i, \tilde{o}_{t-1}) Pr(S_{t-1} = i \mid tildeo_{t-1}). \tag{3.68}$$

The Markov structure defines switching probabilities, introduced as a_{ij} in previous sections, as

$$\begin{aligned} Pr[S_t = 1 \mid S_{t-1} = 1] &= p_{11} \\ Pr[S_t = 2 \mid S_{t-1} = 1] &= p_{12} = 1 - p_{11} \\ Pr[S_t = 2 \mid S_{t-1} = 2] &= p_{22} \\ Pr[S_t = 1 \mid S_{t-1} = 2] &= p_{21} = 1 - p_{22}. \end{aligned} \tag{3.69}$$

Thus the Markov property gives

$$Pr(S_t = 1 \mid S_{t-1} = i, \tilde{o}_{t-1}) = Pr(S_t = 1 \mid S_{t-1} = i). \tag{3.70}$$

Then by substituting equation (3.69) and (3.70) into (3.68), the following expression for regime probabilities yield

$$\begin{aligned} p_{1,t} = Pr(S_t = 1 \mid \tilde{o}_{t-1}) &= \sum_{i=1}^{2} Pr(S_t = 1 \mid S_{t-1} = i, \tilde{o}_{t-1}) Pr(S_{t-1} = i \mid \tilde{o}_{t-1}) \\ &= Pr(S_t = 1 \mid S_{t-1} = 1) Pr(S_{t-1} = 1 \mid \tilde{o}_{t-1}) \\ &\quad + Pr(S_t = 1 \mid S_{t-1} = 2) Pr(S_{t-1} = 2 \mid \tilde{o}_{t-1}) \\ &= p_{11} Pr(S_{t-1} = 1 \mid \tilde{o}_{t-1}) + (1 - p_{22}) Pr(S_{t-1} = 2 \mid \tilde{o}_{t-1}) \\ &= p_{11} Pr(S_{t-1} = 1 \mid \tilde{o}_{t-1}) + (1 - p_{22}) Pr(S_{t-1} = 2 \mid \tilde{o}_{t-1}) \\ &= p_{11} Pr(S_{t-1} = 1 \mid \tilde{o}_{t-1}) + (1 - p_{22})(1 - Pr(S_{t-1} = 1 \mid \tilde{o}_{t-1})). \end{aligned} \tag{3.71}$$

Here p_{11} and p_{22} are probability parameters to be estimated, and the last equality in (3.71) has been made due to the two state assumption made in the model. Equation (3.71) describes the probability of being in regime 1 at time t, given all observable information until time $t - 1$.

Next step in deriving the likelihood function is the implementation of Bayes Rule and the law of total probability which together can be written as the following

$$Pr(A \mid B) \underset{\text{(Bayes)}}{=} \frac{Pr(B \mid A \cdot Pr(A)}{Pr(B)} \underset{\text{(Tot. prob.)}}{=} \frac{Pr(B \mid A) \cdot Pr(A)}{Pr(B \mid A) \cdot Pr(A) + Pr(B \mid \bar{A}) \cdot Pr(\bar{A})}. \tag{3.72}$$

The first equation states Bayes Rule, and second equation displays the law of total probability. Here $\bar{A}$ is the complementary element of A, in other words, the event of being in regime 1 or 2. With the implementation of Bayes Rule, the regime probability $Pr(S_{t-1} = 1 \mid \tilde{o}_{t-1})$ in equation (3.71) can be reformulated to a function of $Pr(S_{t-1} = 1 \mid \tilde{o}_{t-2})$, and written as the following

$$\begin{aligned} p_{1,t} = Pr(S_t = 1 \mid \tilde{o}_{t-1}) &= Pr(S_t = 1 \mid \tilde{o}_{t-1}, \tilde{o}_{t-2}) \\ &= \frac{f(\tilde{o}_{t-1} \mid S_{t-1} = 1, \tilde{o}_{t-2}) Pr(S_{t-1} = 1 \mid \tilde{o}_{t-2}}{\sum_{i=1}^{1} f(\tilde{o}_{t-1} \mid S_{t-1} = i, \tilde{o}_{t-2}) Pr(S_{t-1} = i \mid \tilde{o}_{t-2})} \end{aligned} \tag{3.73}$$

Further by substituting equation (3.73) into (3.71), the probability of being in regime 1 at time t can be written as

$$\begin{aligned} p_{1,t} &= Pr(S_t = 1 \mid \tilde{o}_{t-1}) \\ &= p_{11} \left[\frac{f_{1,t} p_{1,t-1}}{f_{1,t} p_{1,t-1} + f_{2,t}(1 - p_{1,t-1})} \right] + (1 - p_{22}) \left[\frac{f_{2,t}(1 - p_{1,t-1})}{f_{1,t} p_{1,t-1} + f_{2,t}(1 - p_{1,t-1})} \right] \end{aligned} \tag{3.74}$$

Where $f_{1,t} = f(o_t \mid S_t = 2)$ and $f_{2,t} = f(o_t \mid S_t = 2)$. Note that the variables need to calculate $p_{1,t}$, are previous values of $p_{1,t-1}$, the constants p_{11} and p_{22} and previous densities $f_{1,t-1}$ and $f_{2,t-1}$. Since the regime probability only depends on constants and last periods regime probabilities and densities, the regime probability is called a first order recursive process.

The specification of the conditional means and variances were clarified earlier in this section, and can be modified based on the model specification. Thus the sample log likelihood function can be calculated as the weighted sum of the logarithm of the two regime specific densities:

$$L = \sum_{t=1}^{T} \log \left[f(o_t \mid \tilde{S}_t, \tilde{o}_{t-1}) \right] = \sum_{t=1}^{T} \log \left[p_{1,t} f_{1,t} + (1 - p_{1,t}) f_{2,t} \right]. \tag{3.75}$$

Or written out completely

$$\ln L = \sum_{t=1}^{T} \ln \left[p_{1,t} \Big(\frac{1}{\sqrt{2\pi\sigma_{1,t}^2}} \exp\big(-\frac{o_t - \mu_{1,t}}{2\sigma_{1,t}}\big)^2 \Big) + (1 - p_{1,t}) \Big(\frac{1}{\sqrt{2\pi\sigma_{2,t}^2}} \exp\big(-\frac{o_t - \mu_{2,t}}{2\sigma_{2,t}}\big)^2 \Big) \right]. \tag{3.76}$$

It is important to note that this estimation is based on a two regime model. The transition probabilities are not variation free, since they are probabilities they need to sum up to 1. This is the case for both the transition matrix, where each row sum up to one, but also for the $p_{i,t}$ probabilities. The estimation of several regimes $k > 2$ can be made, however the estimation process will differ a bit.

Lastly, to summarize this section, a description of the likelihood dynamics is provided. As mentioned, this likelihood estimation is a one lag recursive structure, and it has the following form: **(1)** The regime dependent normal densities can be calculated from equation (3.67). **(2)** The regime probability $P(S_t = 1 \mid \tilde{o}_t)$ is achieved from equation (3.74), as a function of regime dependent densities and the constant transition probabilities. **(3)** The likelihood contribution is calculated as the sum of the densities, weighted with the regime probabilities. **(4)** The algorithm is restated by setting $t = t + 1$ and repeating the steps.

The estimation is complete when the model is obtained by finding the set of parameters that maximize equation (3.76).

3.3 Transformation to t-distributed MS-models

As argued earlier in the paper, a normal distribution may not be the best distribution for financial time series, due to the heavy tails often observed in such series. More common in modeling of financial time series is a more heavy tailed distribution assumed. An alternative is the t-distribution, which in its most basic form only has the degree of freedom parameter ν.

$$f(x \mid \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \cdot \sqrt{\pi\nu}}(1 + \frac{x^2}{\nu})^{-\frac{\nu+1}{2}}. \tag{3.77}$$

However the t-distribution can be expanded to include both mean and variance parameters, which is requirement for this kind of modeling. With the expansion the t-distribution is called the scaled t-distribution and takes the following form

$$f(x \mid \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \cdot \sqrt{\pi\nu\sigma^2}}(1 + \frac{1}{\nu}(\frac{x-\mu}{\Sigma})^2 - \frac{\nu+1}{2}. \tag{3.78}$$

In this MS model ϵ_t follows a conditional t-distribution with degrees of freedom ν_i, t, conditional mean μ_i, t and conditional variance σ_i, t, where $i \in \{1, 2\}$.

$$\epsilon_t \mid \tilde{S}_t, \tilde{y}_{t-1} \sim \begin{cases} t(\nu_1, 0, \sigma_{i,t}) & \text{w.p.} p_{1,t} \\ t(\nu_2, 0, \sigma_{i,t}) & \text{w.p.} p_{2,t}. \end{cases} \tag{3.79}$$

In summary the model MS model with t-distributed innovation takes the form

$$\begin{aligned} o_t &= \mu_{i,t} + \epsilon_t \\ \mu_{i,t} &= a + \lambda \cdot r_{\text{oil}_t} \\ o_t \mid \tilde{S}_t, \tilde{y}_{t-1} &\sim \begin{cases} t(\nu_1, \mu_{1,t}, \sigma_{1,t}) & \text{w.p.} p_{1,t} \\ t(\nu_2, \mu_{2,t}, \sigma_{2,t}) & \text{w.p.} p_{2,t}. \end{cases} \end{aligned} \tag{3.80}$$

Under the assumption of t-distributed innovation the likelihood function looks similar to equation (3.75), just modified with the density of the returns.

4. Data and Preliminary Evidence

This chapter introduces the time series and gives a preliminary analysis of their most common statistical features.

4.1 Index Data

The featured Indices are selected with the purpose of seeing how leading national indices, with varying relation to oil, are affected by oil price returns. The indices represent some of the leading oil importing and exporting countries stock indices: Norway (OSEBX), U.S. (SPX and NASDAQ), Japan (Nikkei 225), Germany (DAX), Netherlands (AEX), UK (FTSE), Canada (TSX), China (SHCOMP). The time series used go back 20 years, from the beginning of 1996 to May 2016, and are weekly values. This time period includes several regime shifts for all time series and is thus suitable for regime switching models. As in Ramchand and Susmel [1998] this thesis uses weekly returns as opposed to daily, due to the noise created by higher frequency returns. On the other hand, lower frequency returns such as monthly observations or quarterly may be misguiding with respect to regime observations, where empirical evidence state that recessions or high volatile periods often are not very durable. Weekly observations is therefore the preferred middle ground for observing regimes.

4.1.1 Norway, OSEBX

The OSEBX, also known as the benchmark Index, helds a representative selection of all stocks on Oslo Stock Exchange, and is considered the most prominent Index on the Norwegian Market. Oslo Stock Exchange is world leading among the energy, shipping and seafood industry. The energy sector is the largest with oil companies such as Statoil, as the biggest company in Norway. Norway is one of the worlds leading oil exporting countries, and thus has a strong weight in the energy sector.

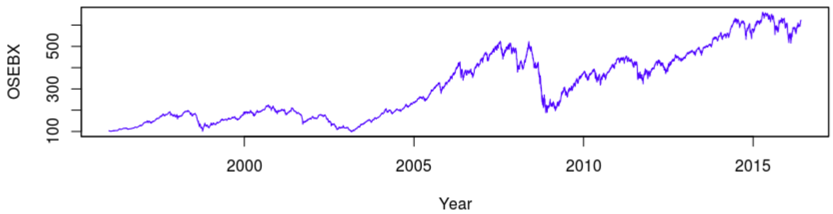


FIGURE 4.1: Development of OSEBX

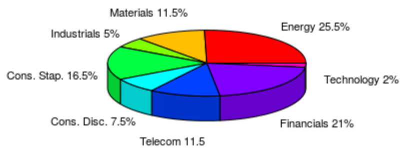


FIGURE 4.2: The weight of the GSCI sectors in OSEBX at the end of May 2016

4.1.2 U.S., SPX

The S&P 500 is widely regarded as the best single gauge of large-cap U.S. equities. There are over USD 7.8 trillion benchmarked to the index, with index assets comprising approximately USD 2.2 trillion of this total. The index includes 500 leading companies and captures approximately 80% coverage of available market capitalization. It differs from other U.S. stock market indices, such as Dow Jones and NASDAQ because of its diversity and weighting methodology. It is considered to be one of the best representations of the U.S. stock market and indicator of U.S. economy. The U.S. is primarily considered as a oil importing nation at present time, however the nation´s oil production is still an important part of the country´s industry.

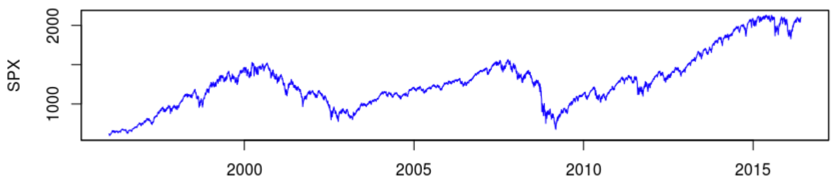


FIGURE 4.3: Development of SPX

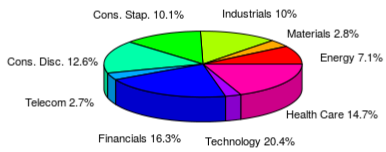


FIGURE 4.4: The weight of the GSCI sectors in SPX at the end of May 2016

4.1.3 U.S., NASDAQ

Nasdaq is a global electronic marketplace for buying and selling securities, as well as the benchmark index for U.S. technology stocks. Nasdaq was created by the National Association of Securities Dealers (NASD) to enable investors to trade securities on a computerized, speedy and transparent system, and commenced operations on February 8, 1971. The term "Nasdaq" is also used to refer to the Nasdaq Composite, an index of more than 3,000 stocks listed on the Nasdaq exchange that includes the world's foremost technology and biotech giants such as Apple, Google, Microsoft, Oracle, Amazon, Intel and Amgen. The Nasdaq has very low weight in the energy sector, and will provide a good indication how a non-energy index is affected by oil returns in a very oil dependent nation.

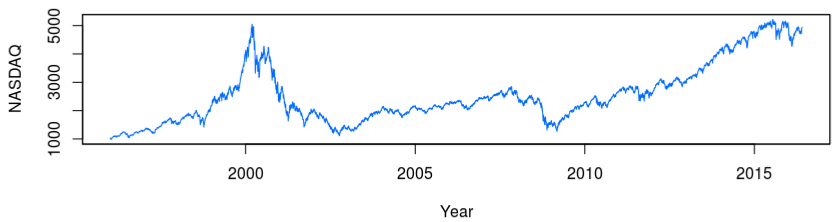


FIGURE 4.5: Development of NASDAQ composite index

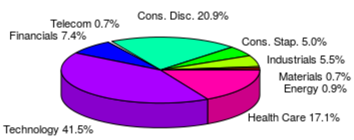


FIGURE 4.6: The weight of the GSCI sectors in NASDAQ at the end of May 2016

4.1.4 Japan, Nikkei 225

Nikkei is short for Japan's Nikkei 225 Stock Average, the leading and most-respected index of Japanese stocks. It is a price-weighted index comprised of Japan's top 225 blue-chip companies traded on the Tokyo Stock Exchange. The Nikkei is equivalent to the Dow Jones Industrial Average Index in the United States.The index has been calculated since September 1950, retroactive to May 1949. Among the best known companies included in the Nikkei index are Canon Inc., Sony Corporation and Toyota Motor Corporation, and it is the oldest stock index in Asia. Japan has been very dependent on importing oil to meet the nation´s energy demand, and has been one of the worlds largest oil importers for several decades.

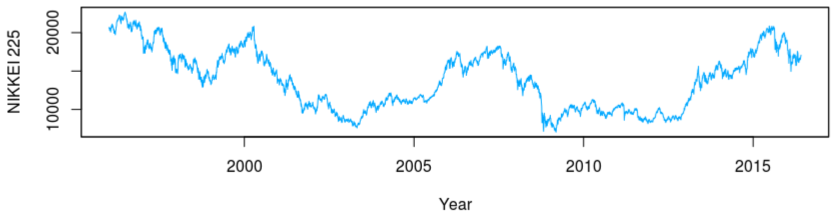


FIGURE 4.7: Development of Nikkei 225 composite index

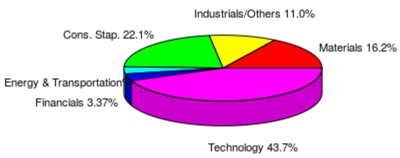


FIGURE 4.8: The weight of the GSCI sectors in Nikkei 225 at the end of May 2016

4.1.5 Germany, DAX

DAX is a stock index that represents 30 of the largest and most liquid German companies that trade on the Frankfurt Exchange. The DAX was created in 1988 with a base index value of 1,000. DAX member companies represent roughly 75% of the aggregate market cap that trades on the Frankfurt Exchange. Germany is considered amongst the world´s largest oil importers, and import over 95% of its domestic use.

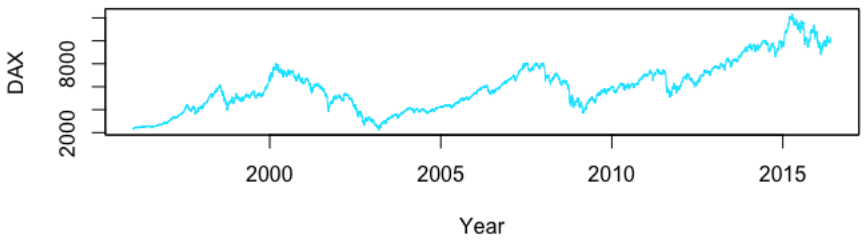


FIGURE 4.9: Development of DAX index

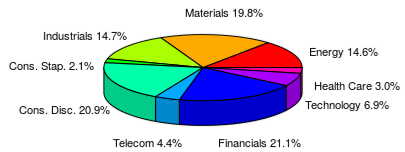


FIGURE 4.10: The weight of the GSCI sectors in DAX at the end of May 2016

4.1.6 Netherlands, AEX

The AEX index, derived from Amsterdam Exchange index, is a stock market index composed of Dutch companies that trade on NYSE Euronext Amsterdam, formerly known as the Amsterdam Stock Exchange. Started in 1983, the index is composed of a maximum of 25 of the most actively traded securities on the exchange. It is one of the main national indices of the stock exchange group NYSE Euronext alongside Brussels' BEL20, Paris's CAC 40 and Lisbon's PSI-20. Netherlands is also considered to be one of Europe's greatest oil importers alongside Germany.

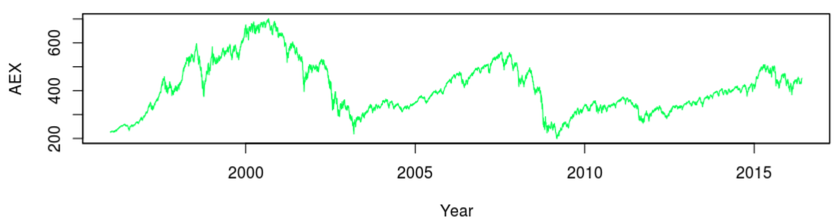


FIGURE 4.11: Development of AEX index

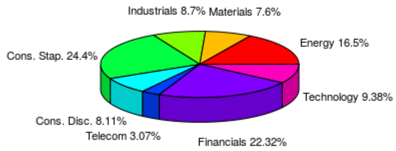


FIGURE 4.12: The weight of the GSCI sectors in AEX at the end of May 2016

4.1.7 U.K., FTSE

This index, more colloquially known as the Footsie, is arguably the most popular and widely used stock market index around the globe. This index is representative of approximately 80% of the market capitalization of the LSE (London Stock Exchange) in its entirety. Larger companies comprise a greater portion of the index because it is weighted by market capitalization. The FTSE 100 is often regarded as an indicator of prosperity among qualifying United Kingdom companies and the economy in general. However, one should note that many of the companies listed in FTSE 100 are based in other countries around the world. U.K. are one of Europe's largest oil exporters, only second to Norway.

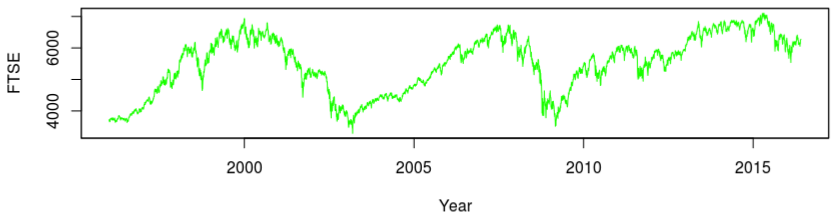


FIGURE 4.13: Development of FTSE 100 index

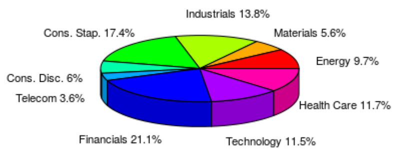


FIGURE 4.14: The weight of the GSCI sectors in FTSE 100 at the end of May 2016

4.1.8 Canada, S&P/TSX

The S&P/TSX composite index is the Canadian equivalent to the S&P 500 market index in the United States. The S&P/TSX Composite Index contains stocks of the largest companies on the Toronto Stock Exchange (TSX). The index is calculated by Standard and Poor's, and contains both common stocks and income trust units. Additions to the index are generally based on quarterly reviews. The Toronto Stock Exchange is dominated by a lot of commodity stocks, most notably crude oil, due to the concentration of natural resources in Canada. Thus, the S&P/TSX Composite Index is more correlated in commodity prices than its counterparts in the U.S. Canada has the third largest oil reserves in the world and is the world's fifth largest oil producer and fourth largest oil exporter.

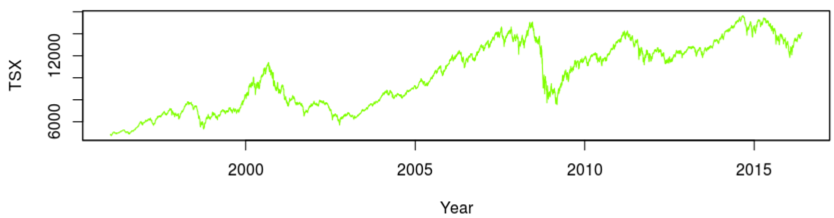


FIGURE 4.15: Development of S&P/TSX 100 index

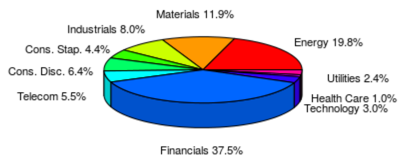


FIGURE 4.16: The weight of the GSCI sectors in S&P/TSX 100 at the end of May 2016

4.1.9 China, SHCOMP

The SSE Composite is a good way to gain a broad overview of the performance of companies listed on the Shanghai exchange. More selective indexes, such as the SSE 50 Index and SSE 180 Index, show market leaders by market capitalization. Over time, it is likely that the SSE Composite will closely resemble the overall economy of China; there are still many large, state-run companies that have yet to go public in sectors such as banking, energy and health-care. China is one of the world´s largest oil importers, and has had a steady increase in its demand for the last decades.

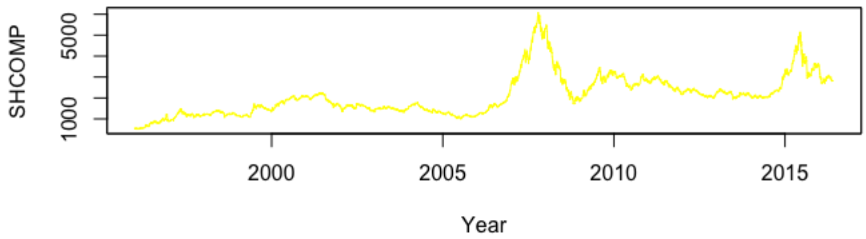


FIGURE 4.17: Development of SHCOMP 100 index

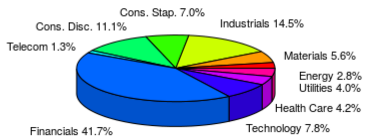


FIGURE 4.18: The weight of the GSCI sectors in SHCOMP at the end of May 2016

4.1.10 Europe brent oil

Roughly two-thirds of all crude contracts around the world refere to brent blend, making it the most widely used oil of all. These days, "Brent" actually refers to oil from four different fields in the North Sea: Brent, Forties, Oseberg and Ekofisk. Crude from this region is light and sweet, making them ideal for the refining of diesel fuel, gasoline and other high-demand products. The supply is water-borne, which makes it easy to transport to distant locations.

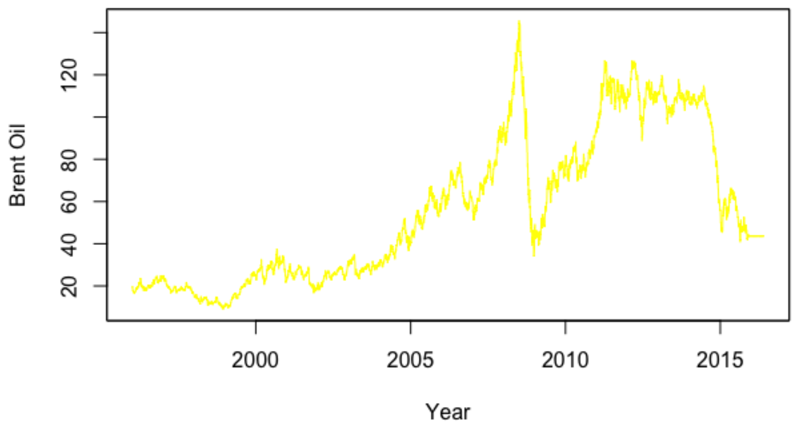


Figure 4.19: Development of brent oil

4.1.11 WTI

WTI refers to oil extracted from wells in the U.S. and sent via pipeline to Cushing, Oklahoma. The fact that supplies are land-locked is one of the drawbacks to West Texas crude – it's relatively expensive to ship to certain parts of the globe. The product itself is very light and very sweet, making it ideal for gasoline refining, in particular. WTI continues to be the main benchmark for oil consumed in the United States

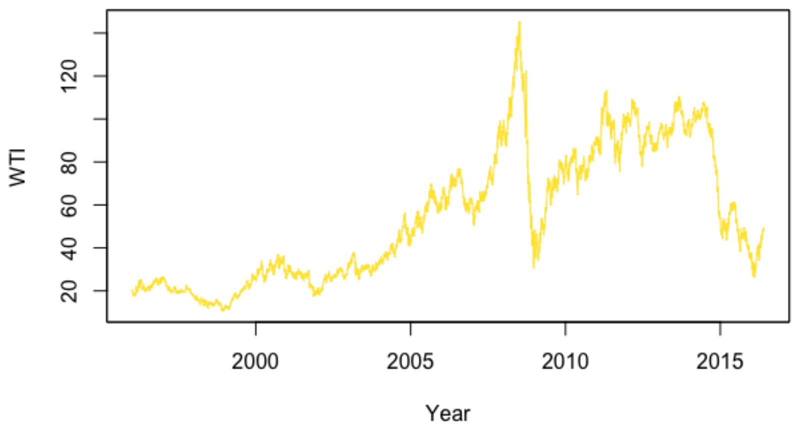


Figure 4.20: Development of WTI

4.2 Descriptive Statistics

The data used is weekly log-returns for the indices described in the previous section. The data covers the period from January 1996 to June 2016, and consists of 1041 observations. As a preliminary look at the data, table 4.1 reports summary statistics on index and commodity returns. Including mean, standard deviation, skewness, excess kurtosis and Jarque-Bera normality test statistic.

TABLE 4.1: Summary Statistics on weekly Market Log-Returns

Market	Mean	SD	SK	Excess κ	JB	$Q(4)$	$Q^2(4)$
OSEBX	0.0017	0.0327	-0.7884	6.1935	1771.1**	17.418**	354.94**
SPX	0.0012	0.0249	-0.5647	3.7673	670.92**	8.901	119.42**
NASDAQ	0.0015	0.0345	-0.5969	3.1637	495.96**	16.046**	186.12**
NKY	-0.0002	0.0314	-0.4756	3.1946	481.91**	6.805	67.167**
DAX	0.0014	0.0327	-0.6741	3.2780	544.92**	4.9466	82.828**
AEX	0.0007	0.0307	-0.6294	3.2569	528.83**	5.5447	151.56**
FTSE	0.0005	0.0243	-0.5300	2.8176	393.07**	13.85**	83.642**
TSX	0.0010	0.0243	-0.8315	4.5699	1025.8**	5.9393	127.10**
SH	0.0016	0.0395	-0.3624	4.6459	959.02**	15.59**	63.63**
Brent	0.0009	0.0534	-0.1812	2.6458	309.33**	7.766	67.29**
WTI	0.0009	0.0551	-0.1198	4.0076	699.15**	27.189**	243.1**
Gold	0.0008	0.0430	0.4905	14.1497	8726**	73.591**	117.78**

SK: Skewness coefficient.
κ: Excess kurtosis coefficient.
JB: Jarque-Bera Normality test. H0: Is Normal.
Q(4): Ljung-Box test statistics, testing for auto-correlation of the returns for up to four lags. H0: No Auto-Correlation
ARCH(4): Ljung-Box test statistics, testing for auto-correlation in the variance of the returns for up to four lags. H0: No Auto-Correlation.
*= Significance at 5% level
**= Significance at 1% level

The average weekly return is positive for all indices except NIKKEI 225, and the index standard deviation is ranging from 0.0243 to 0.0395, and higher for the commodities with a span from 0.0430 to 0.0551. Skewness is a measure of the distribution´s asymmetry, and a negative skewness value indicates that crashes are more likely to happen than booms, e.g. that the majority of the extreme values observed are negative. Further, the excess kurtosis measures the heaviness of the tails in the time series compared to a normal distribution. With positive excess kurtosis for all the time series, the empirical distributions have fatter tails than a normal distribution. Skewness and Excess kurtosis are used to evaluate the normality of observations through a Jarque-Bera test. Test is overwhelmingly rejected for all time series. In summary the financial time series used in this thesis confirm some widespread results on stock returns: Positive return, negative skewness and fat tails, with the exception of Nikkei 225. The Ljung-Box Q-statistics tests the auto-correlation between the weekly log-returns. The test shows significance at 1% for Norway, NASDAQ, UK and China.

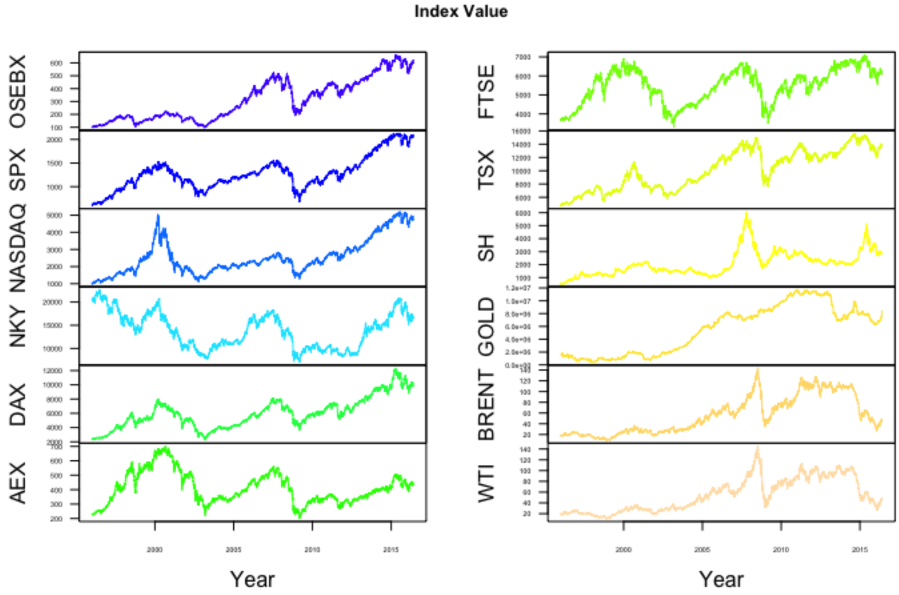


FIGURE 4.21: Value of all indices

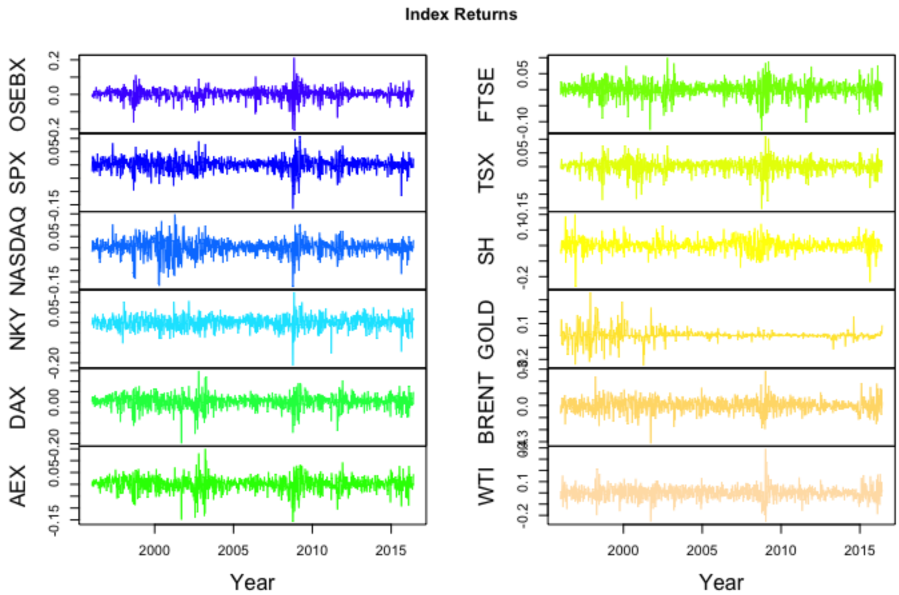


FIGURE 4.22: Weekly log returns

The hetroskedasticity is considered by regressing over past squared residuals with use of Ljung-Box test. The test is overwhelmingly rejected for all indices, indicating the variance of the returns to have significant serial correlation. The ACF-plot for the return are shown in figure 4.23 and 4.24. There is little to none consistency in the auto-correlation of the returns. As shown in figure 4.23 and 4.24 serial correlation of the returns has little to no affect for most indices. The ACF of the squared returns are shown in figure 4.25 and 4.26. The dashed lines make up a 95% confidence interval under the null hypothesis of independence. The ACFs for the squared returns show significant auto-correlation that persists for about ten lags for all indices and commodities. Except for Nikkei 225, which has significance for up to about five lags. The long memory of squared returns is closely related to the volatility clustering noted in the figure 4.22.

The returns shown in figure 4.22 are seen to be mean stationary for all indices and commodities, since they fluctuate around a constant mean level close to zero. The volatility related to the returns vary across the time line, creating clusters. Further, this explains the high amount of hetroskedasticity in the time series, as mentioned in Chapter 2.

The correlations between the national indices and WTI and brent appear to be far from constant. This appears from figure 4.27 and 4.28 where a one-year rolling correlation of the indices and oil are shown. The rolling correlation seems to be moving in the same patterns for the indices. Norway and Canada seem to have the overall highest correlation with oil. China, with the Shanghai composite index, have the lowest overall oil correlation.

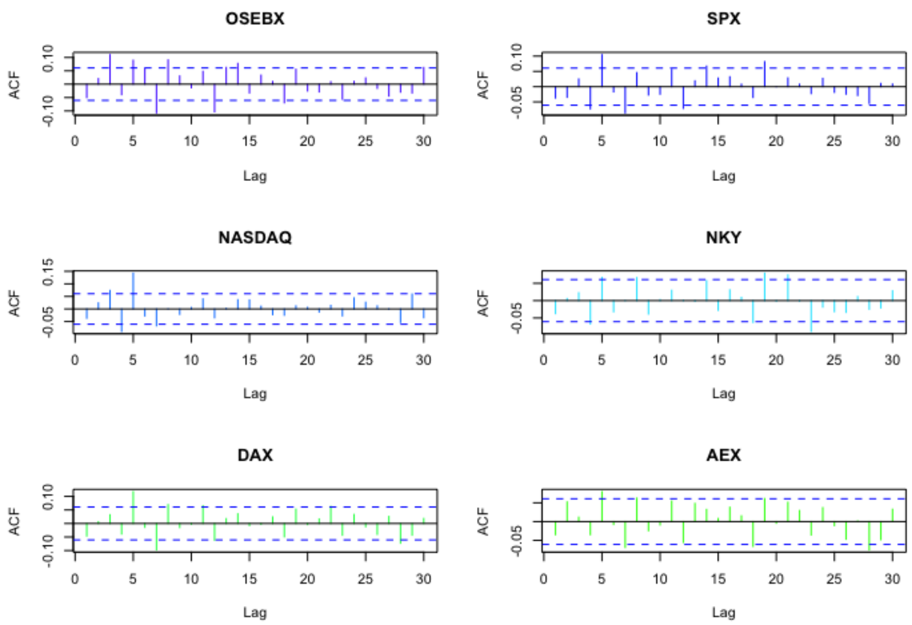


FIGURE 4.23: Auto-correlations for the returns.

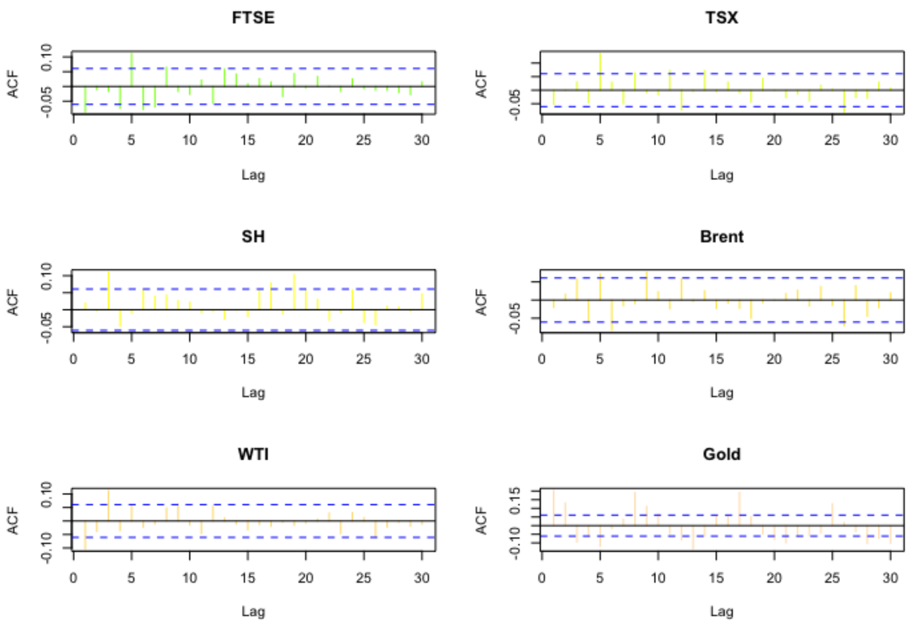


FIGURE 4.24: Auto-correlations for the returns.

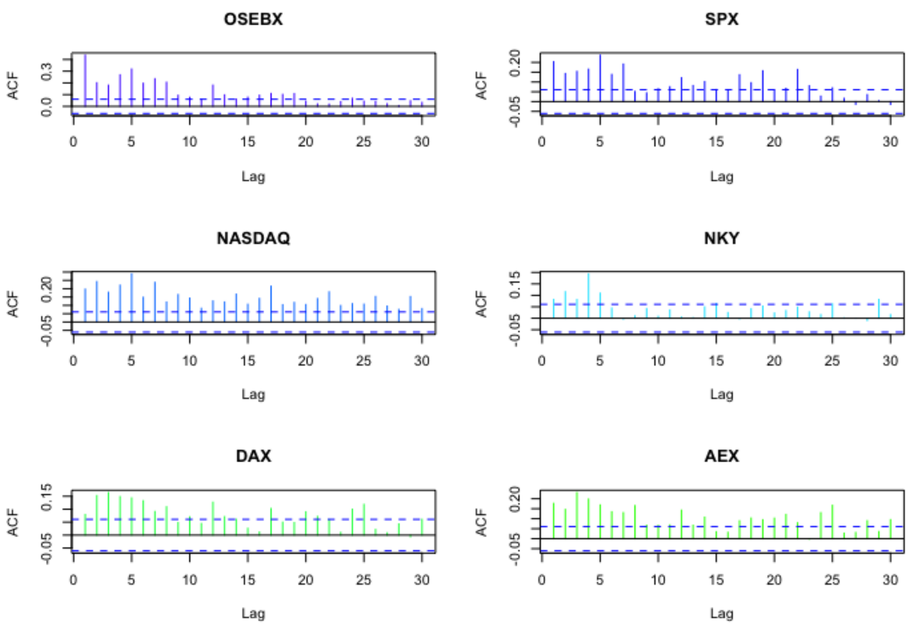


FIGURE 4.25: Auto-correlations for squared returns.

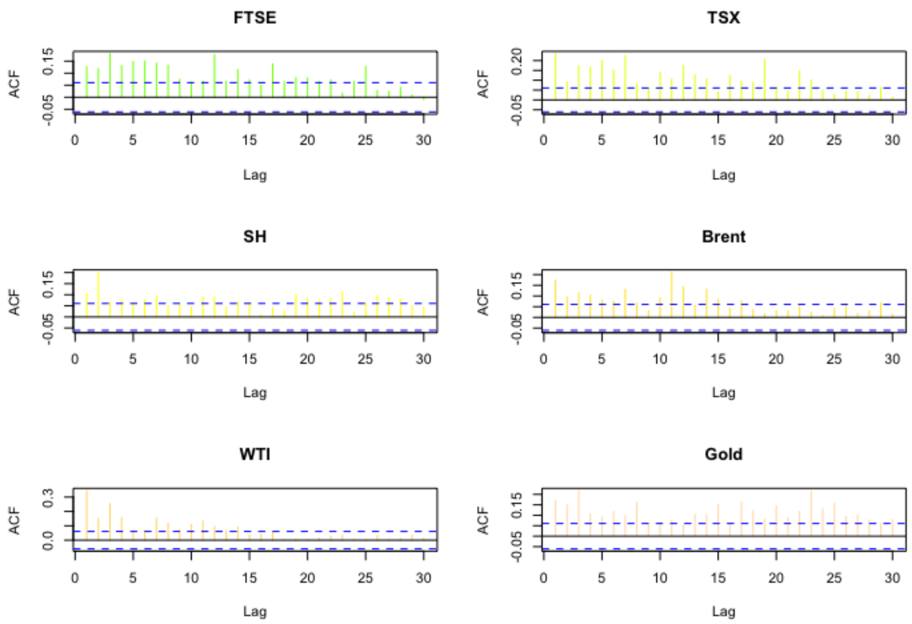


FIGURE 4.26: Auto-correlations for the squared returns

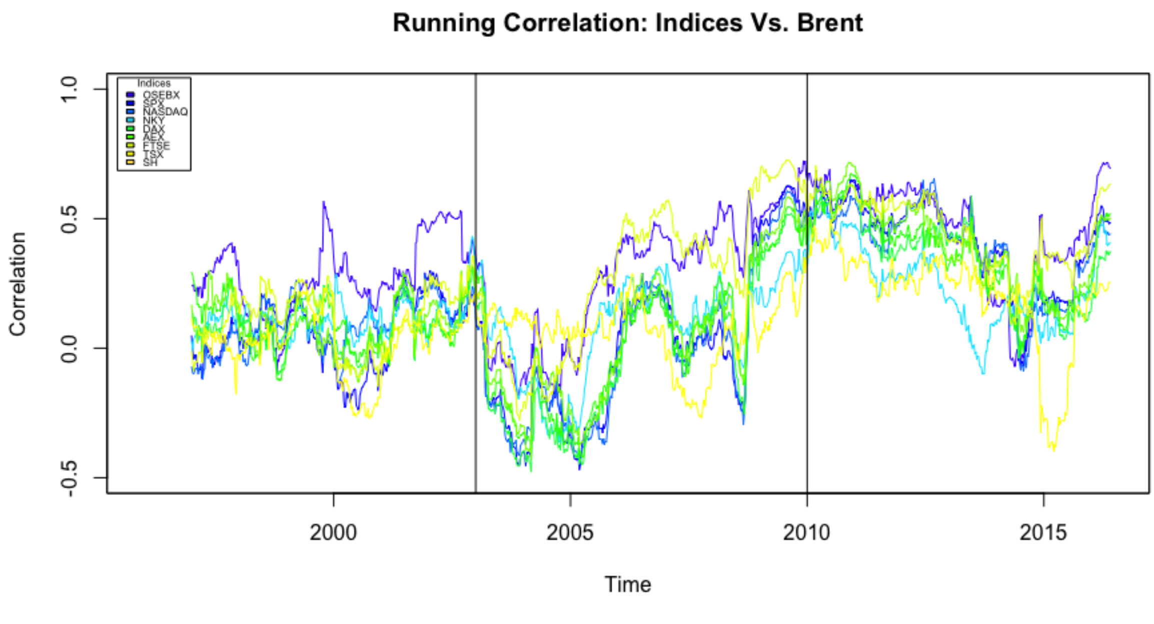


FIGURE 4.27: Running correlation for national indices and Euro brent oil

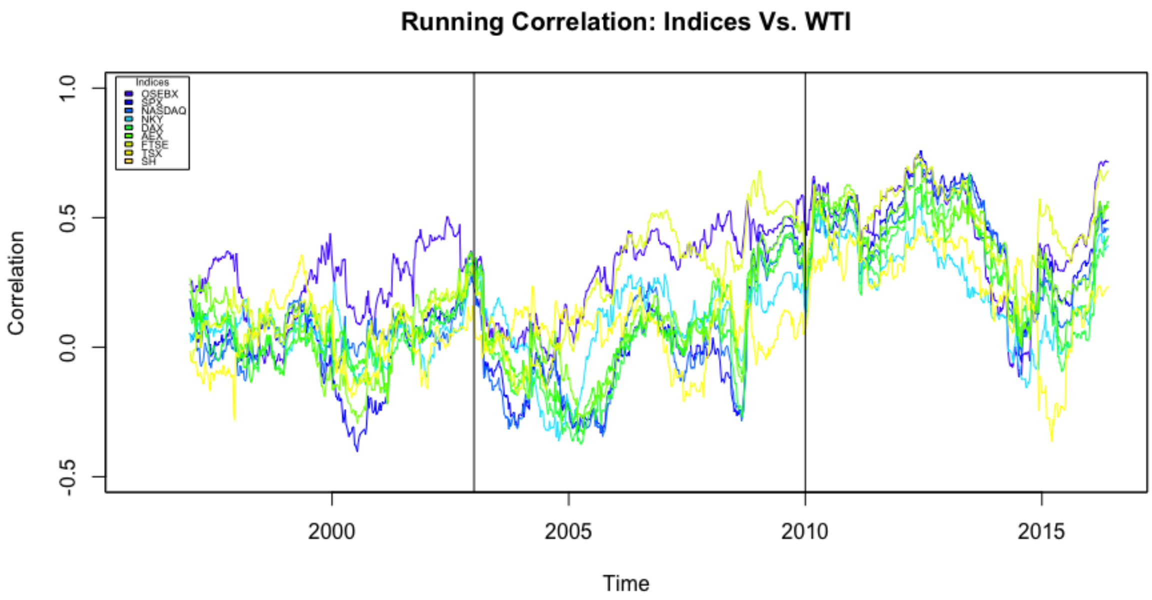


FIGURE 4.28: Running correlation for national indices and WTI oil

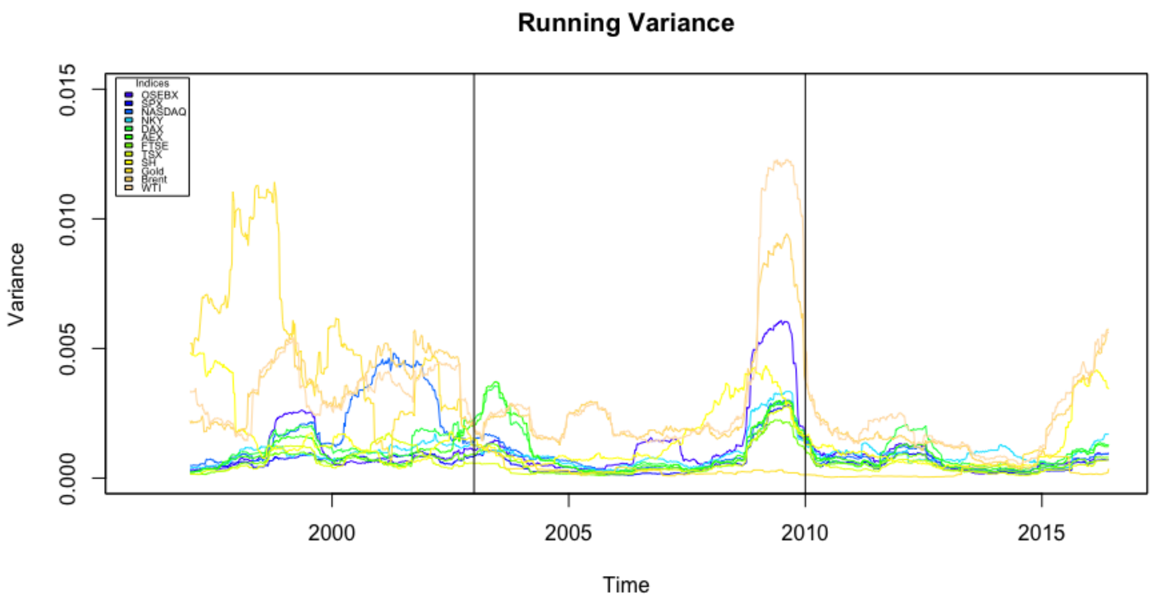


FIGURE 4.29: Running variance for all indices and commodities

This thesis investigates the relation between variance and correlation, and the rolling variance for the indices are shown in figure 4.29. The variance movement between all indices are more correlated at the times of high market volatility, where the volatility movement are more alike during turbulent periods. However, by comparing the running correlation of the oil prices against the running variance, an increase in correlation cannot be systematically related to an increase in variance in our data sample. This can be illustrated by the high correlation in the none volatile periods after the financial crisis in 2008. On the other hand, the oil drop in mid 2014 did not cause markets shocks, and the oil-stock market correlation dropped. Which might be an indicator that the cause of the volatility shocks, with respect to both the indices and the oil, have different affects on the correlation.

The plots and discussion so far only partially confirm the relationship between correlation and stock-market turbulence. The same goes for by evaluating the correlation matrices of the stock indices against the oil prices over various sub-periods. The observed correlations responding to each sub periods are seen i tables 4.2, 4.3 and 4.4. A formal test for constant unconditional correlation can be performed using Jennrich [1970] test of equality of two correlations matrices computed over independent sub-samples. Tables 4.5 and 4.5 report the results of the Jennerich test over a set of sub-periods, for the correlations between the national indices and crude oil. For a $(n \times n)$-dimensional correlation matrix, the statistic is distributed as χ^2 with $(n(n-1)/2$ degrees of freedom. The correlation matrices are between the oil prices and the stock indices. The sub-periods are from 1996-2002,2003-2009 and 2010-present (June 2016).The null hypothesis is equal correlation matrices. Considering brent oil and table 4.5, OSEBX, SPX, FTSE have low p-values between both transitions between the time periods, indicating a change, and in fact, increasing correlation over a long time period. Further TSX has a significant increase in correlation between the first and second time period, which remains at high level for the next period. NASDAQ, DAX, AEX have a significant increase in correlation between the second and third time period. Both the Shanghai and Nikkei 225 index have no significance in the change in correlation between the time periods. However, mutual for all indices is a consistent increase in correlation over the time periods. This coincide with the evaluation of the running correlation in figure 4.27, where the correlation for all indices is at a top during the third period.

Further the evaluation of WTI against the national indices coincide with the brent oil in most areas. By evaluating table 4.6, OSEBX, SPX, FTSE and TSX have significant increase with p-values below 5% between both time periods. NASDAQ, DAX, AEX have significant increase between the second and third time period and lastly Shanghai and Nikkei 225 have no significance between the time periods. However, there is a consistent increase in correlation between the time periods. The running correlation in figure 4.28 shows an increasing correlation in the third period.

By evaluating the running variance and correlation in figures 4.27, 4.28 and 4.29 there tend to be clear evidence of the indices moving closer together, and indicating a growing integration, not only towards the oil price but amongst the indices themselves as well. Also by dividing the indices into oil exporting and importing countries the exporting countries like Norway (OSEBX), Canada (TSX) and UK (FTSE) have a more significant increase in correlation than the remaining oil importing countries over the sub periods.

Several papers including Loretan and English [2000] and Boyer et al. [1997], have argued that testing of unconditional correlation consistency for sub-periods may be misleading. This is because the unconditional correlation estimate is biased in case of variance shifts. Therefore, even though one might be aware of volatility shocks, unconditional correlations estimates have to be corrected before any testing. As stated by Boyer, when a period cannot be related to a clear regime (low or high volatility), changes in correlation over time or across regimes cannot be detected reliably by splitting the sample according to the realized value of the data. This is a consequence of the selection bias that occurs when sub-samples are chosen a priori, according to the data. To evaluate correlation it is therefore necessary to do the following:

1) Use a data generating process allowing for the possibility for structural changes.

2) Estimate the models parameters.

3) Evaluate correlation with respect to regimes.

These issues are covered in the next chapter. We test the null hypothesis of constant conditional correlation, where the regimes are determined endogenously.

Table 4.2: Correlation and Variance in the period 1996-2002

	Correlation		
	Brent	WTI	Variance
OSEBX	0.250	0.208	9.48
SPX	0.064	0.021	7.99
NASDAQ	0.109	0.063	19.96
NKY	0.123	0.057	9.55
DAX	0.080	0.040	12.98
AEX	0.084	0.057	10.14
FTSE	0.091	0.068	6.54
TSX	0.115	0.084	7.31
SH	0.020	0.020	17.28

Variance is scaled with 10^4.

Table 4.3: Correlation and Variance in the period 2003-2009

	Correlation		
	Brent	WTI	Variance
OSEBX	0.412	0.388	15.45
SPX	0.244	0.194	6.65
NASDAQ	0.181	0.130	9.15
NKY	0.167	0.152	10.38
DAX	0.120	0.140	10.31
AEX	0.150	0.164	10.96
FTSE	0.233	0.232	6.14
TSX	0.469	0.406	6.36
SH	0.108	0.074	17.41

Variance is scaled with 10^4.

Table 4.4: Correlation and Variance in the period 2003-present.

	Correlation		
	Brent	WTI	Variance
OSEBX	0.525	0.524	6.83
SPX	0.424	0.448	4.83
NASDAQ	0.393	0.405	6.34
NKY	0.267	0.267	9.49
DAX	0.338	0.338	8.77
AEX	0.402	0.402	7.08
FTSE	0.420	0.450	5.05
TSX	0.505	0.553	3.94
SH	0.151	0.170	11.97

Variance is scaled with 10^4.

Table 4.5: Jennrich Test of Equality of Correlation Matrices for Brent Vs. National Indices over various Subperiods

height	Degrees of Freedom	1996-2002 vs. 2003-2010		2003-2010 vs.,2011-2015	
		Statistics	p-value	Statistics	p-value
Brent-OSEBX	1	5.391	0.0202	3.173	0.0749
Brent-SPX	1	5.865	0.0154	6.711	0.00958
Brent-NASDAQ	1	0.932	0.3341	8.805	0.0030
Brent-NKY	1	0.353	0.5521	1.892	0.1689
Brent-DAX	1	0.2936	0.5879	7.641	0.0057
Brent-AEX	1	0.7651	0.3817	11.464	0.000709
Brent-FTSE	1	3.723	0.0536	7.069	0.007839
Brent-TSX	1	24.830	0.0000	0.325	0.5685
Brent-SH	1	0.874	0.3499	0.331	0.5652

Table 4.6: Jennrich Test of Equality of Correlation Matrices for WTI Vs. National Indices over various Subperiods

height	Degrees of Freedom	1996-2002 vs. 2003-2010		2003-2010 vs.,2011-2015	
		Statistics	p-value	Statistics	p-value
WTI-OSEBX	1	6.351	0.0117	4.554	0.0328
WTI-SPX	1	5.381	0.0203	13.109	0.0003
WTI-NASDAQ	1	0.781	0.3769	14.655	0.0001
WTI-NKY	1	1.607	0.2049	2.495	0.1142
WTI-DAX	1	1.805	0.1790	7.408	0.00086
WTI-AEX	1	2.056	0.1515	11.104	0.00086
WTI-FTSE	1	4.911	0.0266	9.804	0.00174
WTI-TSX	1	19.666	0.0000	5.589	0.0181
WTI-SH	1	0.527	0.4675	1.668	0.1965

5. Effects of Oil Returns on Markets

This chapter includes the presentation of the estimation results. It provides argumentation of the model choice with its advantages and downfalls. An analysis of the results and their statistical significance is evaluated with respect to the research questions proposed earlier in the thesis.

5.1 Model Selection

This thesis wants to evaluate national indices against oil price in a regime shifting environment. The regime shifting is intended to capture the different volatility regimes. However, while working with Markov switching models, it is important to discuss which parameters one should let be regime depending. Let´s first consider a standard ARCH(q) model.

$$r_t = a_0 + \sum_{i=1}^{q} a_i r_{t-i} + \epsilon_t \tag{5.1}$$

$$\epsilon_t^2 = b_0 + \sum_{i=1}^{q} b_i \epsilon_{t-i}^2 \tag{5.2}$$

As seen from a simple ARCH(q) model, there are no limitations to how many parameters one can wish to add. Let´s assume a MS model that switches between two ARCH(q) models, and all the parameters shifts. One of the problems that may occur is while one might want to identify regimes as "high volatility" and "low volatility", if everything is changing, it could turn out that it fits better with "high mean" vs. "low mean", or that it even shifts on one of a´s, the autoregression coefficients, or the b´s, the hetroskedasticity coefficients. This means that by letting more parameters be allowed shifts, one is more likely to produce a model estimating something else than what was intentional. However, this issue can be sorted out either by constraining some of the parameters or by removing them. Both result in less "moving parts" that might ease the possible confusion in the model. In regard to MS models, often more complex models need more restriction than the simpler models.

As seen from the descriptive statistics chapter, there is a fairly high amount of hetroskedasticity in the time series. So if one was to model the time series in a single regime model, then variants of ARCH or GARCH would probably be a natural choice. However, this thesis wants to classify volatility regimes in time series, and thus the application Markov switching models. How to choose an appropriate MS model is not given, and further investigation motivate us to evaluate the ARCH effects in regime shifting models. Let´s consider a univariate MS-ARCH(k=2,q=1) model, with two regimes and one autoregressive coefficient in the mean and variance equation[1]. Here k indicates number of regimes and q the number of lags. As done in Ramchand and Susmel [1998] and due to the argumentation of restricting the shifting parameters, this model only allows for changes in the variance equation. The suggested model will look like the following

$$r_t = a + a_1 r_{t-1} + \epsilon_t \qquad \epsilon_t \sim N(0, h_t) \tag{5.3}$$

$$\frac{h_t}{g_{s_t}} = b_0 + b_1 \frac{\epsilon_{t-1}^2}{g_{s_{t-1}}} \tag{5.4}$$

$$g_1 = 1. \tag{5.5}$$

The g´s are a scale parameters that capture the change in variance. One of the g´s are unidentified and hence g_1 is set equal to 1. The g´s are the only regime shifting parameter, and allow us to capture the possible high and low volatile regimes. More details about this model can be seen in Hamilton and Susmel [1994]. By evaluating all the time series under this model, there is little significance in the ARCH estimated coefficient b_1, in any of the estimated time series except for Shanghai and Nikkei 225. However, for both Shanghai and Nikkei 225, the transition matrix P, reveals that there was very little persistence of staying in a regime. Low probability of staying in a regime is often result of the model not fitting well with the time series. These results coincide with Ramchand and Susmel [1998] and Chesnay and Jondeau [2001], whom also were unable to find significant ARCH effect in national indices. Thus the hetroskedasticity, seen in the ACF plot of the residuals in figure 4.25 and 4.26, is assumed to be captured by the switching nature of the model. Any further hetroskedasticity within each regime is assumed negligible. This leads to the fact that the constant mean and variance model is the most appropriate MS model, especially for classifying persistent regimes. The preferred model used in this thesis is the following MS model as seen in Chapter 2

[1]Other variants of MS-ARCH have been tried with no qualitative change in result

$$o_t = \mu_{S_t} + \epsilon_t \tag{5.6}$$

$$\epsilon_{S_t} \sim \mathcal{N}(0, \sigma^2_{S_t}). \tag{5.7}$$

Variants of single regime ARCH and GARCH are in way regime switching in the way the variance adapts through the time series. Modeling proves that MS models with ARCH effects are not well suited for the objective of this thesis where the more complex switching models provide higher transition probabilities between regimes, thus low persistence. The μ_{S_t} will further be extended to include oil returns as well, $\mu_{S_t} = a + \lambda_{S_t} r_{t\text{oil}}$.

MS modeling so far has been done with the assumption of residuals being normally distributed. As seen from descriptive statistics, the normal distribution assumption is rejected for all indices. Nevertheless, with the assumption of regimes, where each regime has its own mean and variance, the observations are assumed to belong to a mixture distribution. Thus testing for normality one needs to take into consideration the parameters belonging to each regime. Before the analysis of oil returns with respect to national indices, there are some assumptions that need to be clarified. First, what is the significance of regime shifting, is this an appropriate model to capture volatility periods? Secondly, justifying the assumption of letting our observations to either be normal or t-distributed in their respective regime.

5.2 The Significance of Regime shifting

This thesis discusses the option of time series having two states, first and foremost depending on the variance. An important feature is to test the significance of two regimes against single regimes. Under the same argumentation as above, more regimes lead to more parameters shifting. Thus the number of regimes added should be well argued and constrained to the point that there is a clear distinction between the regimes. As mentioned before, this paper emphasizes on the regimes classifying and distinction between high and low volatility regimes.

Note that a regular Likelihood Ratio (LR) test is an informal test of regime shifting. This is because parameters of the second regime are not identified under the hypothesis of no regime shifting. Therefore the regularity conditions justifying the χ^2 approximation to LR test do not hold. Nevertheless, a regular LR test gives a good indication of how well a model perform next to the other. Hansen et al. [1996] proposed a LR test procedure to overcome this problem, but even for simple models it is very computational demanding. Therefore the preferred test is proposed by Ang and Bekaert [1999], which is based on Monte Carlo simulation to obtain the small sample distribution of the LR test statistic. Let´s consider a baseline MS model, in which both variance and mean are assumed to be regime dependent. Preliminary testing suggests that both mean and variance are regime dependent, and will provide the alternative hypothesis of regime shifting against the null hypothesis of no regime shifting.The mean and variance are all constant over time in their respective regime. The small-sample distribution of the associated LR test statistics is obtained as followed:

1) Estimate a model assuming one regime.

2) Create N samples, using the one regime model parameters of length T, where $T = 1041$ is the length of our data samples.

3)Estimate a two regime model for each sample, and compute the Likelihood ratio of the estimated sample, one regime vs. two regime model.

4) Sample distribution of LR is computed for the N samples.

By comparing the LR from from our time series, one regime vs. two regime, against the simulated small sample LR distribution the significance of the regime shifting can be evaluated. The results can be seen in table 5.1.

Table 5.1: Test for Regime Shifting of National Indices

	OSEBX	SPX	NASDAQ	NKY	DAX	AEX	FTSE	TSX	SH
Sample LR statistic	316.318	223.095	290.50	90.306	210.577	230.316	218.27	282.511	254.906
Monte Carlo sample distribution									
Number of samples	200	200	200	200	200	200	200	200	200
Mean	2.702	2.878	2.878	2.844	2.929	2.831	2.233	2.722	3.029
Standard Deviation	3.024	3.457	3.458	3.311	3.273	3.126	2.571	3.196	3.355
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q3	4.838	3.996	4.452	4.452	4.384	4.270	4.350	4.350	4.326
Max	11.180	11.790	17.970	20.400	21.230	16.56	15.710	15.71	16.82

Table 5.2: Test for Regime Shifting of brent and WTI

	Brent	WTI
Sample LR statistic	142.763	155.635
Monte Carlo sample distribution		
Number of samples	200	200
Mean	3.017	3.196
Standard Deviation	3.304	3.5437
Min	0.000	0.000
Q1	0.106	0.000
Q3	4.7350	5.234
Max	16.780	16.560

Table 5.1 and 5.2 report descriptive statistics for the small sample distribution of LR test statistic for the national indices and the oil. The null hypothesis of no regime switching is overwhelmingly rejected for all time series, since all LR test statistics obtained with simulated samples are smaller than the sample LR test statistic from our time series. Not any of the maximum values of the small sample distribution is even close to the sample LR statistics. This strongly enhance our assumption for MS models.

5.3 Significance of Parameters

It is of high importance to test whether the estimated parameters are statistically significant. A student t-test is applied to test this, and the null hypothesis and the alternative hypothesis is the following

$$H_0 : \hat{\theta} = 0 \qquad\qquad H_1 : \hat{\theta} \neq 0 \tag{5.8}$$

Where $\hat{\theta}$ is the vector of all parameter estimates. Under H_0 the student t-test is t-distributed with $n - k$ degrees of freedom, where n is number of observations and k is estimated parameters. Due to the high degree of freedom this test is closely related to Wald´s test where the statistic is approximately $N(0, 1)$. The test statistic has the following form

$$T = \frac{\hat{\theta} - \theta_0}{SD(\hat{\theta}}. \tag{5.9}$$

Where θ_0 is the value of the true parameter, which when testing H_0 is 0. The test evaluates if the estimated parameter can be assumed to be zero and therefore redundant in the model. The significance is chosen at a 5% and 1% level.

5.4 Model Estimates

This section will include the results of MS model with constant mean and variance in their regime for both normal and t-distributed innovations. Also will the estimations be divided into estimations including oil variables in the mean expression.

5.4.1 MS-Model without oil variables

With this model, both the mean and the variance are assumed to be constant for the whole estimation period within a single regime. The model used is the same proposed earlier in the thesis

$$o_t = \mu_{S_t} + \epsilon_t \tag{5.10}$$

$$\epsilon_{S_t} \sim \mathcal{N}(0, \sigma^2_{S_t}). \tag{5.11}$$

Previous results and testing imply strongly that a one-regime model not accounting for hetroskedasticity is a poor model and that our time series is very likely to consist of regimes. As emphasised earlier, this paper wants to establish two clearly distinct volatility regimes, and therefore do not apply more sophisticated models, as switching ARCH and GARCH models, because this is more likely to create regimes less depending on the volatility and create higher transition probabilities. This model assumes that any conditional hetroskedasticity is to be controlled by the switching between regimes. It will also be shown later that the hetroskedasticity is significantly reduced by the effect of regime switching itself. Also recall that weekly data is used, opposed to monthly data. This gives enough observations to be able to estimate the different states, but without the noise of daily data.

The regimes are not observable, but have been identified through the estimation process described in the methodology chapter. From the results in table 5.3 and 5.4 two regimes have been identified with clear economic interpretation. First, two regimes are identified: One captures the behaviour of the stock market in a recession state with low expected return and high variance. The other regime (regime 1) captures the market in an expansion state with high expected return and low variance. This identification is consistent for all the national indices and the oil time series as well, except for the Shanghai index. This is the case for both the normally and t-distributed estimations.

The standard deviation parameter σ is significant for the two regimes in all time series, for both normal and t-distributed innovations. For the normally distributed innovation model the mean or the intercept parameter, μ, in regime 1 is significant at a 5% or 1% level for all indices except the Shanghai index. The same go for the ones with t-distributed innovations. It reveals that the Shanghai index has experienced high volatile periods in both "bull" and "bear" markets, opposite to the other national indices where high volatile periods are consistently related to "bear" markets. This can also be seen from figure 4.17 where the Shanghai index has steeper positive spikes than the other indices. In contradiction to regime 2 with high variance and low mean, the μ value is not significant for most indices in regime 1, expect for OSEBX, DAX and AEX which have a significance at a 5% level. For the t-distributed models only OSEBX and Shanghai have significant mean parameters in regime 1.

As expected, the main driver for regime classification is the variance for both models. The ratios between the standard deviation in the two regimes is approximately two for all the time series. The oil indices are clearly to be the most volatile ones with low volatile regime standard deviation, σ_1, close to 0.040 for WTI and brent, and high volatile regime standard deviation, σ_2, at 0.075 and 0.079 for the normally distributed innovations. For the t-distributed models, the standard deviation is consistently lower than its equivalent parameters in the normal distribution. This is because of the t-distributions larger tails compared to the normal distribution, thus more of the observation falls under the t-distribution model with smaller standard deviation. The national indices have σ_1 from

0.0146 to 0.0252 with respectively FTSE 100 and Nikkei 225, and σ_2 span from 0.0342 to 0.0584 with FTSE 100 and SH, from the normal distributed model. From the t-distributed models the σ_1 span from 0.0128 to 0.0223 from with TSX and SH, and σ_2 span from 0.0299 to 0.0503 with TSX and DAX. The consistency of the volatility ratio within the indices being around two indicates that the relative volatility, between high and low, is the same for the national indices. Regardless of some indices being more volatile than others. This observation is consistent for both the normal and t-distributed models.

In order to identify the persistence of the regimes one needs to evaluate the probability estimates p_{11} and p_{22}. Considering a Markov chain, they are the probabilities for staying in current state. The transition probability estimations are crucial in the evaluation of Markov switching models. Models with low transition probability and therefore strong persistence, indicating that there is a significant difference between the two regimes. Low probability values often indicate that the estimation process has been having difficulties sorting out the difference between the regimes, and keep jumping between the regimes quickly, thus low persistence. If this is the case, comparisons of the two regimes are often of little value. However, as seen in tables 5.3 and 5.4 all the probabilities are high and significant. The probability plots with smoothed probabilities can be seen in Appendix A. The probability of staying in regime 1, low volatile market, is higher for all the time series, thus having higher persistence. The expected duration of being in an expansion phase or in the "bull"-market is longer than in recession phase, which implies that only an extreme event can switch the market from "bull" to "bear". On the other hand, recession periods often do occur over a long time span, thus it is natural to have a lower probability of staying in its regime. In general are the probabilities p_{11} and p_{22} consistently higher in the t-distributed models than their normally distributed equivalents. One reason for this might be the t-distributions larger tails that require even more extreme observations to make the jump from one regime to the other. The degrees of freedom, ν, for the t-distributed models have been set to be non-switching parameter, and the degrees of freedom are significant for all time series except DAX, AEX and FTSE.

Considering table 5.3 and 5.4, the initial expectation of persistent regimes is confirmed with constant probabilities p_{11} and p_{22} all over 0.9 for both models and all time series. Some of the upcoming analysis in this paper are based on how each observation is classified to its respective regime. There are several ways to find the state sequence by using the smoothed probabilities. Ramchand and Susmel [1998] have chosen to classify the state sequence S' by letting the highest probability of one observed state be the classified state. In other words, the smoothed probability for each observation is evaluated individually as seen in equation (5.12).

$$S' = \underset{1 \leq i \leq N}{\operatorname{argmax}}\, \gamma_i(t), \qquad 1 \leq t \leq T, \quad 1 \leq i \leq N \tag{5.12}$$

This paper propose the use of the Viterbi algorithm, which allows to classify the most probable state sequence considering the transition probabilities as well, and evaluate the expression in equation (5.13). In other words would the first method classify each observation individually, but the Viterbi algorithm allow us to classify the observations considering the whole time series. Thus, by the use of the Viterbi algorithm, some observations would be classified to a regime contradicting its value. The algorithm can be revised in chapter 2. The plot of regime classification and smoothed probabilities with conditional standard deviation for all models and time series can be seen in appendix A.

$$\underset{S'}{\operatorname{argmax}}\, P(S' \mid O, \lambda). \tag{5.13}$$

Mandelbrot‘s quote about financial markets [2] in the introduction, gave the initial expectation of estimated regimes to be persistent. Evaluating the plots of regime classification and smoothed probabilities in Appendix A, and the probability values calculated in table 5.3 and 5.4, the persistence of the regimes is strongly argued for. As mentioned in the preliminary analysis the time period from 1996 to present consists of several economic crises and thus economic shocks, allowing MS modeling to be well suited. The most consistent high volatile periods seen from the probability plots in Appendix A are the dot.com bubble in early 2000’s and later the financial crisis in 2008. Regarding our oil indices the Asian economic crises in 1997 and the 9/11 terror attacks seem to provide a longer streak of high volatile period for the brent oil, and the WTI is switching a bit more rapidly between the regimes. Later the movement for brent oil and WTI seems to be more correlated, and both switch regimes in the 2008 financial crises and the oil drop in 2014.

The reason for choosing Markov Switching models was its ability to capture structural changes unbiased, and by doing so also capture the hetroskedasticity, the serial dependence of the squared returns. It is important to have in mind that the structural breaks in our model have no middle ground, it is either in high or low volatile regimes. However, we may evaluate the probability of being in either of the states. When regime classification is applied to the observations it may be considered a little too binary. In contradiction one has GARCH models which capture the serial dependence of the squared returns based on the variance and residuals of previous observation, and thus perhaps a more

[2] "...large changes tend to be followed by large changes - of either sign - and small changes tend to be followed by small changes

accurate way of modeling these time series would be a MS-GARCH model combining these properties. Nevertheless, this paper wants to classify volatility regimes, which our models so far have been doing satisfactory. Further, more complex MS models require more parameter estimations, thus a further evaluation of which parameters to allow shifting or not. As argued earlier in this chapter; allowing more parameters to shifts also changes the "reason" of the shift. Even though the MS model with constant mean and variance do not inhabit any ARCH parameter in the variance expression, this model was assumed to capture most of the hetroskedasticity in our time series. To see if the MS models have been able to capture this effect, the standardized squared model residuals are defined as

$$\hat{\epsilon}_t^2 = \left(\frac{r_t - \hat{\mu_t}}{\hat{\sigma_t}}\right)^2. \tag{5.14}$$

Where $\hat{\sigma_t}$ and $\hat{\sigma_t}$ are defined from the state sequence defined by the Viterbi algorithm. An alternative method to calculate the standardized squared model residuals would be to use the conditional standard deviation and mean, which captures the nuance between the regimes based on the probability of being in a state, and thus have stronger capability to capture the ARCH effects. However, by classifying each observation based on the Viterbi, there is a stronger assumption of distinct regimes. In the preliminary analysis a Ljung-Box test of the squared residuals in the time series were performed, and the hypothesis of homoskedasticity and no auto-correlation of the squared residuals were overwhelmingly rejected. From the results in table 5.3 and 5.4 the $Q(4)^2$ statistics are drastically reduced, and in some cases even insignificant. If the models have been able to capture all the systematic volatility found in the time series, the estimated standardized squared residuals would be random uncorrelated volatility. From the ACF plot for the standardized squared residuals in Appendix A, the systematic volatility is clearly reduced compared to the ACF plot in the preliminary analysis. The t-distributed models have overall more struggle capturing the systematic volatility due to its higher $Q^2(4)$ value. However due to the t-distribution models low probability of leaving its state and the use of the Viterbi algorithm, lead us to believe that some of the observations will be "miss classified" in the sense that the observation do not fit with its state. Consequently the estimate squared residuals being affected of ARCH effects. As a result the ARCH effect is not fully captured for either models with respect to their regimes. Therefore volatility is shown to have two sources: Volatility captured by regime shifting, which is modeled with MS model, and volatility clustering within regime, a feature not captured in the proposed model. Figure 5.1 displays the standardized residuals estimated from the MS model with normal innovations, and acquired from the Viterbi algorithm. It can be viewed as opposed to histogram plot of OSEBX in the Theory chapter, and the effect of the regime shifting is obvious. The use of histograms is good for visuals, however a

QQ-plot is preferred with the use of confidence bands, which allow for a more thorough analysis and will be covered in the next paragraph.

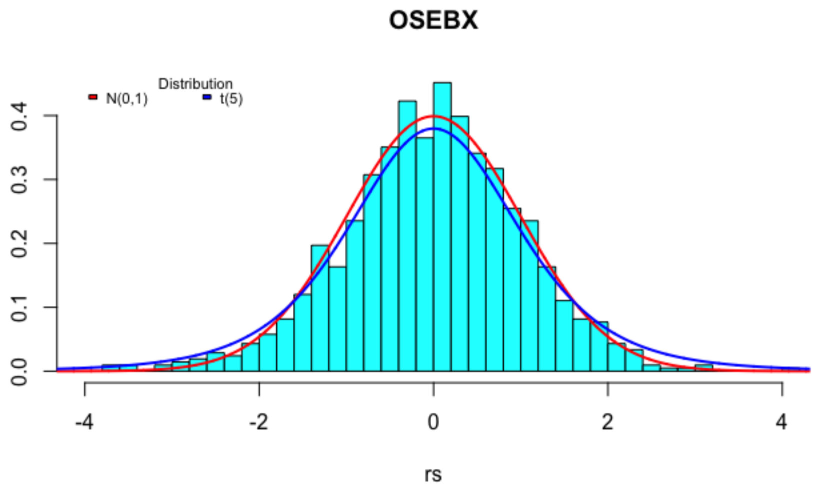


FIGURE 5.1: Histogram of residuals calculated from MS model with normally distributed innovations and Viterbi algorithm for OSEBX

.

One other feature discussed in Chapter 2 was the the distribution of the returns. Fama [1965] stated that the tails of the returns were fatter than what a normal distribution with kurtosis of 3 was able to capture. Thus a more leptokurtic distribution might be able to match the actual distribution better. Recall that excess kurtosis is defined as kurtosis minus 3, and from the preliminary analysis leading to the t-distribution to be a good alternative. Assuming that the conditional distribution and the Markov switching is a good fit for the time series, then the estimated standardized residuals will be distributed as the conditional distribution proposed. To evaluate the fit of the normal and t-distribution qq-plots have been estimated based on the standardized residuals provided from parameter estimation and Viterbi. The qq-plots can be seen for all time series and for both estimation models in Appendix A. Given that the conditional distribution of the innovations and the estimated distribution of the standardized residuals are the same, the estimated residuals will imitate a straight line in qq-plot. The tail sensitive confidence bands are provided by simulation, a method proposed by Aldor-Noiman et al. [2013]. The tail-sensitive bands reject the conditional normal distribution for the residuals at a 5% level of significance in a great proportion of the observations for OSEBX, SPX, NASDAQ, DAX, AEX and TSX. However, the conditional t-distribution for the residuals is only rejected for a few of the observations in SPX, NASDAQ, TSX and WTI. Overall the conditional t-distribution seems to be a better match, especially in the lower end of the tail. Have in mind that the use of Viterbi might "miss-classify" some observations and create outliers. Thus the use of conditional means and standard deviation in calculation of both residuals and squared residuals are more likely to lead to a better fit for the conditional distribution or less hetroskedasticity. Nevertheless, the approach of using the state sequence as reference to the standardized residuals, allow us to confirm two things: First that t-distribution is more likely to fit persistent regimes, and secondly that there are volatility clustering within classified regimes.

The use of t-distribution and Viterbi algorithm have been proven useful to classify persistent volatility regimes, but both the qq-plots and $Q(4)^4$ statistics argue that volatility clustering within regimes still are present. To consider which of the MS models fit the time series best, based on the conditional distribution, evaluation of the Akaike information criterion (AIC) and Bayesian information criterion (BIC) have been performed. Values of the estimations are in table 5.5. Note that the preference of the model is here turned into a maximization problem and not a minimization, which is most common. Kass and Raftery [1995] and Burnham and Anderson [2004] explain that if a difference in AIC being greater than 4, then a model is substantially less preferred. Thus in cases for DAX and FTSE, where the AIC in the normal model is larger than the AIC in the t-distributed model, the difference is insignificant. Thus the models with t-distributed innovations are overall preferred from the AIC, and the normal distribution is strongly rejected for all cases except DAX, AEX, FTSE and brent. The BIC is closely related to the AIC, however the penalty term is larger in BIC than in AIC. If the difference in the BIC's are greater than 6, there is a strong preference in one of the models. The result from the BIC values indicate DAX, AEX, FTSE and brent would prefer a normal-distributed model, however the difference in value is not significant in any case. On the other hand t-distribution is strongly preferred over normal distribution for the other indices. In general the conditional t-distribution is preferred for modeling financial time series from these results, which coincide with previous studies.

TABLE 5.3: Estimated parameters of MS model with constant mean and variance, and normal distributed innovations

	OSEBX	SPX	NASDAQ	NKY	DAX	AEX	FTSE	TSX	SH	BRENT	WTI
μ_1	0.0055**	0.0031**	0.0037**	0.0015	0.0041**	0.0034**	0.0024**	0.0034**	-0.0001	0.0037*	0.0036*
μ_2	-0.0103*	-0.0026	-0.0031	-0.0064	-0.0093*	-0.0062*	-0.0025	-0.0039	0.0047	-0.0051	-0.0052
σ_1	0.0206**	0.0159**	0.0217**	0.0252**	0.0240**	0.0204**	0.0146**	0.0148**	0.0239**	0.0391	0.0398
σ_2	0.0543**	0.0364**	0.0522**	0.0470**	0.0541**	0.0469**	0.0342**	0.0366**	0.0584**	0.0755	0.0794
p_{11}	0.96**	0.97**	0.98**	0.98**	0.98**	0.97**	0.96**	0.97**	0.97**	0.98	0.98
p_{22}	0.89**	0.94**	0.97**	0.91**	0.92**	0.93**	0.94**	0.93**	0.94**	0.97	0.96
$Q(4)$	3.393	7.858	6.235	2.281	4.462	8.046	15.505**	6.521	11.762*	5.325	7.891
$Q^2(4)$	26.902**	13.453**	6.915	4.025	10.04*	13.735**	16.765**	14.483**	1.7136	8.226	28.582**

TABLE 5.4: Estimated parameters of MS model with constant mean and variance, and t-distributed innovations

	OSEBX	SPX	NASDAQ	NKY	DAX	AEX	FTSE	TSX	SH	BRENT	WTI
μ_1	0.0058**	0.0037**	0.0040**	0.0035**	0.0043**	0.0037**	0.0024**	0.0041**	-0.0002	0.0037*	0.0040*
μ_2	-0.0070*	-0.0015	-0.0015	-0.0041	-0.0083	-0.0052	-0.0019	-0.0023	0.0074*	-0.0030	-0.0049
σ_1	0.0183**	0.0141**	0.0201**	0.0217**	0.0230**	0.0191**	0.0142**	0.0128**	0.0223**	0.0369	0.0371
σ_2	0.0435**	0.0306**	0.0465**	0.0333**	0.0503**	0.0425**	0.0322**	0.0299**	0.0486**	0.0663	0.0691
p_{11}	0.98**	0.97**	0.99**	0.99**	0.98**	0.97**	0.96**	0.97**	0.99**	0.99	0.98
p_{22}	0.93**	0.96**	0.98**	0.98**	0.94**	0.94**	0.95**	0.95**	0.97**	0.98	0.97
ν	7.72**	10.57**	9.23**	9.11**	22.08	16.08	24.69	8.81**	7.29**	14.40*	14.22**
$Q(4)$	2.123	7.832	4.936	3.465	7.311	9.853*	14.591**	6.932*	13.073*	4.912	10.918*
$Q^2(4)$	40.803**	18.100**	4.025	15.984**	11.813*	17.128**	15.466**	20.515	10.332*	10.095*	45.398**

$Q(4)$: Ljung-Box test statistics, testing for auto-correlation of the returns for up to four lags. H0: No Auto-Correlation
$Q^2(4)$: Ljung-Box test statistics, testing for auto-correlation in the variance of the returns for up to four lags. H0: No Auto-Correlation.
*= Significance at 5% level
**= Significance at 1% level

TABLE 5.5: Estimation of Likelihood, AIC, BIC for the normal and t-distributed models

	Normal			T-distribution		
	Log-liklihood(L)	AIC	BIC	Log-liklihood	AIC	BIC
OSEBX	2238.937	4465.874	4436.186	2247.267	**4480.534**	**4445.899**
SPX	2477.729	4943.457	4913.770	2482.746	**4951.491**	**4916.856**
NASDAQ	2169.778	4327.557	4297.869	2175.960	**4337.920**	**4303.285**
NKY	2169.937	4327.874	4298.187	2176.867	**4339.735**	**4305.099**
DAX	2184.985	**4357.970**	**4328.282**	2185.890	4357.780	4323.145
AEX	2260.2260	4509.304	**4479.616**	2262.460	**4510.920**	4476.284
FTSE	2498.168	**4985.336**	**4954.648**	2498.988	4983.975	4949.340
TSX	2530.649	5049.229	5019.612	2537.820	**5061.639**	**5027.003**
SH	2011.132	4010.264	3980.576	2021.800	**4029.598**	**3994.963**
Brent	1643.447	3274.889	**3245.202**	1645.920	**3277.839**	3243.204
WTI	1617.661	3223.323	3193.635	1621.882	**3229.764**	**3195.128**

AIC: The AIC are computed as $2L - 2k$, where k are the number of parameters estimated, which is 6 for normal and 7 for t-distribution.
BIC: The BIC are computed as $2L - kln(n)$ where $n = 1041$ are the number of weekly observations.
The bold results are the largest of the criterion and the preferred model for the given time series.

To summarize the results presented in this section; two distinct regimes have been observed with low mean/high variance and high mean/low variance, and the variance was observed as the main driver for the regime shifting with strong significance. There are also evidence that under the assumption of t-distributed innovations there is a stronger persistence in its regime, which is due to the fatter tails. Further, low volatile periods have a stronger regime persistence. Regime classification done by the Viterbi algorithm provides evidence that most of the hetroskedasticity is captured by regime shifting, however there is still serial correlation within regimes. In general the t-distributed models are preferred over the normally distributed models form the AIC and BIC criterion.

5.4.2 MS model with oil variables

This section evaluates the MS models in the previous section with extension adding the oil price return of brent and WTI to the model. The proposed Markov switching model is the following

$$o_t = \mu_{S_t} + \epsilon_t \tag{5.15}$$

$$\epsilon_{S_t} \sim \mathcal{N}(0, \sigma^2_{S_t}) \qquad \mu_{S_t} = a + \lambda_{S_t} r_{\text{toil}} \tag{5.16}$$

Obviously this section is only for the national indices with oil parameters. The modeling of the oil returns in the previous section is a preliminary analysis of further evaluation of correlation with respect to oil regimes. Papers such as Aloui and Jammazi [2009] add the oil variables to the variance term, and evaluate how oil price affects the volatility. This paper, on the other hand, adds the oil returns in mean expression,μ_{S_t}, to evaluate how the oil price affects the overall return of the national indices. The results in table 5.6 and 5.7 indicate that the distribution properties observed in the previous section are consistent with the extended model. The regimes are associated with low mean and high variance, "bear" market, or high mean and low variance, "bull" market. Also the t-distributed models have consistent higher probabilities of staying in a regime. As the previous section focused on a thorough analysis of the assumptions related to the models, and an analysis of estimation parameters, this section will focus on how the oil parameter,λ_{S_t}, relates to the national indices. To see whether the oil price is statistically linked to real indices returns and whether they can explain behaviour shifts in stock market. Recall that regime 1 is associated with high mean/low variance and regime 2 with low mean/high variance, respectively "bear" and "bull" market.

Methodologically, in order to find out whether oil price is significantly correlated to real stock returns, comparison of the log-likelihood with and without the oil variables are made. The LR statistic is approximately χ^2-distributed with two degrees of freedom. This test reveals that all indices with either brent or WTI incorporated in the model have larger log-likelihood, and H_0 without oil returns are overwhelmingly rejected at 1% significance level for all indices. These findings provide evidence that oil price are statistically correlated to the national indices. This result coincides with Maghyereh et al. [2007] and Park and Ratti [2008], along with previous mentioned Aloui and Jammazi [2009].

Table 5.6, with normally distributed innovations, reports that the coefficients relative to the brent and WTI, λ_1 and λ_2, are statistically significant different from zero, thus statistically significant on their impact on the returns. The normal distributed model has significant λ_2 at 1% level for all indices except Shanghai, indicating that shocks in national indices are strongly correlated with oil returns. Also regime 1 is strongly correlated with oil prices, only DAX has insignificant coefficient with respect to WTI. However, some of the indices like SPX, NKY, and AEX have some of the coefficients "only" significant to a 5% level. Overall the oil impacts the indices more strongly in high volatile periods. Moreover the relative ratio between λ_1 and λ_2 lays between approximately 2 to 5, indicating stronger affect on the return in high volatile periods for significant coefficients.

Table 5.7, with t-distributed innovations, reports much of the same as the normal distributed model. The parameter estimation, and overall behaviour, for the two regimes are consistent with the results without oil variables. In regime 2, only Shanghai has insignificant coefficient, λ_2. For the other indices λ_2 is significant a 1% level. As well as with the results in table 5.4, regime 1 and bull period have strong correlation with oil returns for most indices. However, there are some different results between the models: The brent coefficient proves to be insignificant with respect to SPX, NKY and AEX, and the WTI coefficient insignificant with respect to NKY, DAX and AEX. Likewise as with normal innovations, the ratio between λ_1 and λ_2 approximately lays between 2 to 4.

Overall oil exporting nations such as Norway, Canada and U.K., and their corresponding indices, have larger dependence of oil price in both high an low volatile regimes. Also the LR statistic is notably larger for these indices.

TABLE 5.6: Estimated parameters of MS model with constant mean and variance and oil variables, and normal distributed innovations

	OSEBX		SPX		NASDAQ		NKY		DAX		AEX		FTSE		TSX		SH	
	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI
μ_1	0.0048**	0.0050**	0.0030**	0.0030**	0.0035**	0.0035**	0.0016	0.0017	0.0039**	0.0042**	0.0033**	0.0035**	0.0022**	0.0022**	0.0029**	0.0030**	-0.0002	-0.0002
μ_2	-0.0067*	-0.0068*	-0.0022	-0.0023	-0.0029	-0.0030	-0.006	-0.006*	-0.0059	-0.0072	-0.0053	-0.0046	-0.0019	-0.0019	-0.0025	-0.0026	0.0047	0.0047
σ_1	0.0193**	0.0193**	0.0158**	0.0157**	0.0217**	0.0217**	0.0247**	0.0247**	0.0236**	0.0235**	0.0200**	0.0199**	0.0144**	0.0142**	0.0141**	0.0142**	0.0238**	0.0239
σ_2	0.0482**	0.0480**	0.0346**	0.0352**	0.0508**	0.0516**	0.0442**	0.0443**	0.0516**	0.0503**	0.0452**	0.0445**	0.0330**	0.0327**	0.0336**	0.0339**	0.0581**	0.0580
λ_1	0.1515**	0.1357**	0.0369*	0.0450**	0.0560**	0.0560**	0.0592*	0.0465*	0.0421*	0.0086	0.0485**	0.0371*	0.0523**	0.0523**	0.0904**	0.0823**	0.0519**	0.0597**
λ_2	0.3356**	0.3071**	0.1586**	0.1266**	0.1865**	0.1406**	0.1699**	0.1379**	0.1818**	0.2102**	0.1663**	0.1713**	0.1385**	0.1375**	0.2025**	0.1889**	0.0899	0.0522
p_{11}	0.97**	0.97**	0.97**	0.97**	0.98**	0.98**	0.97**	0.98**	0.98**	0.98**	0.97**	0.97**	0.96**	0.96*	0.97**	0.97**	0.97	0.97
p_{22}	0.90**	0.91**	0.94**	0.94**	0.97**	0.97**	0.92**	0.92**	0.92**	0.92**	0.92**	0.93**	0.94**	0.94**	0.94**	0.94**	0.95	0.95
LR	139.07**	125.07**	37.38**	30.99**	28.08**	20.61**	26.07**	18.20	16.42**	18.24**	25.85**	25.67**	43.57**	46.58**	101.65**	92.35**	10.35**	10.178**

TABLE 5.7: Estimated parameters of MS model with constant mean and variance and oil variables, and t-distributed innovations

	OSEBX		SPX		NASDAQ		NKY		DAX		AEX		FTSE		TSX		SH	
	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI
μ_1	0.0052**	0.00540**	0.0035**	0.0035**	0.0040**	0.0040**	0.0003	0.0035**	0.0045**	0.0047**	0.0038**	0.0039**	0.0022**	0.0022**	0.0037**	0.0037**	-0.0003	-0.0003
μ_2	-0.0043	-0.0048	-0.0018	-0.0016	-0.0019	-0.0017	-0.002	-0.0036	-0.0060	-0.0057	-0.0037	-0.0037	-0.0013	-0.0014	-0.0017	-0.0017	0.0073*	0.0073*
σ_1	0.0167**	0.0169**	0.0141**	0.0141**	0.0197**	0.0199**	0.0206**	0.0217**	0.0221**	0.0219**	0.0181**	0.0182**	0.0138**	0.0138**	0.0124**	0.0126**	0.0221**	0.0220**
σ_2	0.0368**	0.0376**	0.0296**	0.0300**	0.0442**	0.0453**	0.0308**	0.0329**	0.0454**	0.0442**	0.0376**	0.0382**	0.0299**	0.0303**	0.0276**	0.0285**	0.0484**	0.0484**
λ_1	0.1380**	0.1354**	0.0317	0.0393*	0.0557**	0.0536**	0.0403	0.0419	0.0388*	0.0030	0.0359	0.0300	0.0492**	0.0524**	0.0809**	0.0736**	0.0583**	0.0675**
λ_2	0.3560**	0.2963**	0.1491**	0.121**	0.2032**	0.1506**	0.1164**	0.0963**	0.1800**	0.2073**	0.1897**	0.1812**	0.1451**	0.1385**	0.2119**	0.1920**	0.0657	0.0349
p_{11}	0.98**	0.98**	0.97**	0.97**	0.99**	0.99**	0.99**	0.99**	0.98**	0.98**	0.97**	0.97**	0.960**	0.97*	0.97**	0.97**	0.99	0.99
p_{22}	0.95**	0.95**	0.96**	0.96**	0.98**	0.98**	0.98**	0.98**	0.94**	0.94**	0.94**	0.94**	0.97**	0.95**	0.95**	0.95**	0.97	0.97
ν	6.66**	6.88**	10.84**	11.1**	8.27**	8.76**	9.32**	9.71**	15.22	15.92	10.77**	11.92*	17.67	20.29	8.71**	9.93**	7.31**	7.32**
LR	149.71**	133.34**	36.57**	29.84**	32.45**	22.42**	16.18**	20.30**	19.25**	20.28**	30.56**	29.28**	45.60**	47.68**	102.06**	89.68**	11.00**	12.43**

LR: The likelihood ratio (LR) are computed as follows: $2 \times$ (log-liklihood of$H_1 -$ log-liklihood ofH_1), where H_0 is the MS model without oil variables, and H_1 is the MS model with oil variable. The LR test is $\chi^2(2)$-distributed with 2 degrees of freedom.

*= Significance at 5% level

**= Significance at 1% level

6. Correlation across Oil and Stock Markets

This section provides analysis of how the volatility structure of oil affects the national market indices. As the previous section stated that there is a strong correlation between oil and national indices in volatile periods, and this section estimates the actual correlation across regimes. The time series is modeled in state varying framework with two regimes for high volatile and low volatile periods. This allows to examine the relation between correlation and variance in a conditional time and state varying framework. Correlation will be evaluated under a univariate Markov Switching model, which describes the regime shifting and parameters of a single time series. High and low volatile observations of the oil series will be linked pairwise with the national indices.

6.1 Univariate Correlation

The previous chapter incorporated oil variables in the MS model of national indices. The results stated that the oil returns were strongly correlated with the national indices, especially in the high volatile periods. This section evaluates the correlation of the national indices and oil returns with respect to regime classification. The estimations will consider the regime classification provided by both the national indices and the oil returns for the correlation estimations. This allow us to evaluate how the volatility shocks of markets and oil price affect each other. The time series deciding the regime classification is called the home reference.

6.1.1 Correlation when national indices are home reference

A main concern of this paper is how national indices respond to regime classification, and their relation to oil returns during these periods. This first section evaluates the correlation between national indices and oil returns, based on the regime classification of the national indices [1]. This will be interpreted as to how strongly oil and national indices correlate depending on the movements of the national indices. The next section will cover how the correlation depends on the movement of the oil returns.

[1] Classified by Viterbi algorithm.

The results are shown in table 6.1 and 6.2, where regimes are classified with respect to either normal or t-distribution. The state sequence plots can be seen in Appendix A. The overall increase in correlation lays between $50 - 300\%$, only with exception with WTI-DAX, where it is 7 times larger in the t-distribution. In general is the ratio for the correlation observed in the regimes lower for the t-distribution. This is probably because of the strong persistence in the regimes, which allows for the miss classification explained earlier. The inclusion of these miss classified observations in each regime leave the observations in each observation sequence to be less distinct, thus a correlation more close to each other. The normal distributed MS model is more sensitive to regime shifts, enabling shifts due to smaller corrections in the market, thus less likely to achieve miss classification, on the cost of regime persistence.

The oil exporting countries, OSEBX, FTSE and TSX, have consistently smaller correlation ratios than the oil importing nations for brent oil, and OSEBX, SPX, NASDAQ and TSX with the lowest ratio with respect to WTI. Leaving us to believe that the correlation relatively variates less than oil importing countries, or in other words: The correlation is higher in low volatile regimes for oil exporting countries. Nevertheless, except for Shanghai, the high volatile regimes have consistently higher correlations than in the low volatile regimes, for all indices and oil returns. The findings state that price shocks have the same affect on oil returns, regardless if it is exporting or importing oil. These results coincide with Filis et al. [2011] and Ramchand and Susmel [1998], where national indices and oil markets tend to move closer together during high volatile periods.

Table 6.1: Correlation between oil returns and national indices in their regime estimated from the normal distribution and the Viterbi algorithm, with national indices as reference

	Brent				WTI			
Indices	High	Low	Ratio	Hedging	High	Low	Ratio	Hedging %
OSEBX	0.48	0.29	1.66	19.82%	0.44	0.29	1.52	21.66%
SPX	0.30	0.10	3.00	23.75%	0.24	0.13	1.85	24.69%
NASDAQ	0.25	0.11	2.27	23.13%	0.19	0.12	1.58	23.75%
NKY	0.28	0.11	2.55	25.31%	0.23	0.09	2.56	25.94%
DAX	0.23	0.08	2.88	18.75%	0.28	0.04	7.00	17.50%
AEX	0.26	0.09	2.89	18.12%	0.27	0.08	3.38	16.25%
FTSE	0.26	0.15	1.73	20.25%	0.27	0.15	1.93	19.25%
TSX	0.38	0.25	1.52	18.93%	0.35	0.26	1.35	20.50%
SH	0.06	0.11	0.55	19.43%	0.09	0.12	0.75	19.43%

The hedging percentage indicates how often brent and WTI return a positive return when the home index (the national indices) hedge a negative return in the high volatile period. The hedging value seems to be relatively constant for all the markets against brent and WTI; around 25%. Leaving us with none obvious hedging opportunities.

Table 6.2: Correlation between oil returns and national indices in their regime estimated from the t-distribution and the Viterbi algorithm, with national indices as reference

	Brent				WTI			
Indices	High	Low	Ratio	Hedging	High	Low	Ratio	Hedging %
OSEBX	0.44	0.32	1.38	20.15%	0.41	0.31	1.32	21.27%
SPX	0.27	0.11	2.45	22.67%	0.22	0.16	1.37	23.63%
NASDAQ	0.24	0.12	2.00	23.10%	0.18	0.13	1.38	24.42%
NKY	0.23	0.07	3.29	23.72%	0.18	0.10	1.80	23.72%
DAX	0.22	0.09	2.44	20.60%	0.28	0.04	7.00	20.00%
AEX	0.25	0.10	2.50	18.22%	0.26	0.09	2.88	18.62%
FTSE	0.25	0.17	1.47	17.81%	0.26	0.17	1.53	17.00%
TSX	0.37	0.25	1.48	18.88%	0.36	0.24	1.50	19.94%
SH	0.10	0.09	1.11	26.05%	0.05	0.12	0.42	29.81%

6.1.2 Correlation when oil is home reference

In contradiction to the previous section, this section uses the regime shifting of the oil returns to evaluate how correlation changes between the regimes. The results are shown in table 6.3 and 6.4. The regime classification for the oil returns reveal some different observations than national indices. First, let´s evaluate the results in table 6.3, where the state sequence is based on the normal distributed MS model of brent and WTI. Similar to the results for correlation based on index regime shifting, high volatile regimes for oil returns provide higher correlation for most indices. However, with regime shifts with respect to brent oil DAX, AEX and FTSE have lower correlation in the high volatile regime. The ratio is close to 1, implying that oil shocks have little affect on correlation based on this model. Further, regime shifting with respect to WTI also provides results where the correlation is lower in the high volatile regime, in this case for OSEBX and NKY. Overall the normal distributed MS model of the oil returns provides a state sequence where the correlation ratio is low compared to the results form the national indices. Implying that oil shocks have smaller affect on the correlation than the fluctuations caused by global business cycles, as seen in the previous section. The correlation ratio from the brent lays between 0.94 to 1.56, and the WTI ratio between 0.88 to 1.5. Recall that the normal distributed MS model allows to capture smaller and less significant shocks than the t-distributed model, and soon to be seen, this will have great affect on the results.

The state sequence provided from the t-distributed model have some different results than the normally distributed. First, the high volatile regime for both WTI and brent provide consistently higher correlations between the national indices than the normally distributed state sequence. Consequently, the low volatile regime achieved from the t-distributed model consistently produces lower correlation than the corresponding normal distributed state sequence. Thus the correlation ratio is higher for the state sequence gained from the t-distribution. Nevertheless, in general the correlation ratio is lower than for the state sequences provided by the national indices, laying approximately between 1.5 to 3 for brent oil, and 1 to 2 for WTI.

By comparing the results from the table 6.3 and 6.4, the differences are notable. Especially in the cases of DAX, AEX and FTSE with the brent oil home reference. The tables are contradicting, and leading to the state sequence to create these differences. As seen from Appendix A, the state sequence for brent oil gained from both the normal and t-distributed MS model are quite persistent and identical, nevertheless, the correlation estimations is quite different. Implying that the smaller oil shocks that the normally distributed model are able to obtain, either have low or even negative correlation to the national indices. Further investigation lead these oil shocks to be related to the Iraq war and natural disasters, which Filis et al. [2011] have stated to have negative correlation on national markets. The high volatile periods also consist of fewer observations, thus just a few changes in the state sequence may leave quite numerical changes. The same goes for the WTI state sequence seen in Appendix A, where the difference between the sequences gained from the normal and t-distributed models only seem to be the Iraqi war in 2003, leaving a measurable difference. In general brent seems to correlate stronger with national indices than WTI.

Table 6.3: Correlation Between oil returns and national indices in their regime estimated from the normal distribution and the Viterbi algorithm, with brent and WTI as regime reference

	brent				WTI			
Indices	High	Low	Ratio	Hedging	High	Low	Ratio	Hedging %
OSEBX	0.38	0.36	1.06	19.84%	0.35	0.36	0.97	22.40%
SPX	0.25	0.16	1.56	24.40%	0.22	0.16	1.38	25.42%
NASDAQ	0.21	0.14	1.50	24.93%	0.18	0.12	1.50	26.76%
NKY	0.20	0.15	1.33	22.52%	0.14	0.16	0.88	25.42%
DAX	0.15	0.16	0.94	24.66%	0.16	0.14	1.14	28.43%
AEX	0.17	0.20	0.85	25.20%	0.19	0.18	1.06	28.09%
FTSE	0.22	0.23	0.96	23.32%	0.24	0.21	1.14	24.08%
TSX	0.34	0.31	1.10	22.52%	0.31	0.31	1.00	24.41%
SH	0.08	0.11	0.72	26.54%	0.06	0.10	0.60	28.76%

Table 6.4: Correlation between oil returns and national indices in their regime estimated from the t-distributiom and the Viterbi algorithm, with brent and WTI as regime reference

	Brent				WTI			
Indices	High	Low	Ratio	Hedging	High	Low	Ratio	Hedging %
OSEBX	0.42	0.31	1.36	18.92%	0.36	0.35	1.03	21.71%
SPX	0.30	0.10	3.00	23.16%	0.23	0.15	1.53	24.67%
NASDAQ	0.25	0.07	3.57	24.01%	0.19	0.10	1.90	25.98%
NKY	0.22	0.13	1.69	21.75%	0.15	0.14	1.07	25.00%
DAX	0.21	0.08	2.63	22.88%	0.17	0.13	1.31	27.30%
AEX	0.24	0.11	2.18	24.01%	0.19	0.17	1.12	27.30%
FTSE	0.27	0.16	1.69	22.60%	0.25	0.20	1.25	23.02%
TSX	0.36	0.27	1.33	21.75%	0.32	0.29	1.10	24.01%
SH	0.09	0.09	1.00	26.83%	0.06	0.10	0.60	29.60%

7. Conclusion

The use of Markov Switching models (MS) is based on the assumption that interconnection and dependency of economic variables can vary at different states of economic growth. The MS models capture structural breaks and create a flexibility that enables to capture characteristics of the data in the long run. Empirically two regimes have been detected for the time series of the stock markets and oil returns. One regime represents high mean and low volatility, and coincides with economic growth. The other regime with low mean and high volatility, relates to periods of economic downfall and recessions. Stylized facts are often very difficult to model, and serial correlation and distributions with fat tail have been two of these properties in the time series. However, throughout the theoretical analysis, it was shown that the MS models were capable of capturing these characteristics. The MS models were, however, not able to capture all of the hetroskedasticity, even though it was significantly reduced. In addition were t-distributed innovations proposed to increase the kurtosis and the heavy tails.

MS-models with GARCH parameters were fitted in the preliminary analysis. The results revealed that these variables had little impact in a regime shifting environment. However, a problem with standard GARCH models is that they, just like our model, have high and low volatility periods. This is a result of the models recursive structure, and a shift dummy in the variance model will, even in a fixed model, create its own high and low states. Thus, our ambitions to observe persistent and consistent regimes led to a Markov switching model where two parameters are allowed to switch; the mean and variance parameters.

In Chapter 5 there is an estimation of the national markets with and without oil variables. The estimation without the oil variables was first and foremost to see how well the time series respond to the Markov switching models. The same patterns were discovered in all time series; two states with high mean/low variance and low mean/high variance. The bear state has overall lower probability of staying in current regime. This coincides with empirical events, where financial downturns often occur in shorter time spaces. Also t-distributed innovations were proposed, since the excess kurtosis for all the time series were positive, and thus have a larger kurtosis than the normal distribution. Consequently, the MS models with t-distributed innovations have lower variance, and stronger regime persistence. The heavy tailed t-distribution has a better ability to capture extreme observations than the normal distribution. This means that where the MS-normal model would be forced to switch regime in order to capture the extreme observation, the MS t-distributed model has the fat tails to capture an extreme observation due to its high leptokurtic. More extreme values is needed to make the model leave its current state. In general the MS-models with t-distributed innovations are preferred. This assumption is supported by AIC and BIC estimations, but also from the qq-plots based on the standardized residuals presented in appendix A.

Further in Chapter 5, the national markets were modeled with oil variables included in the mean parameter μ. The estimates for the intercept and variance were very similar to the estimates without oil parameters. In the bull regime, oil prices seem to have smaller impact on the national indices, especially the oil importing nations such as Japan, Germany, Netherlands, China and U.S. However, the U.S. do oil exporting as well, but they are considered one of the world´s top oil importers. On the other hand, oil exporting nations such as Norway, Canada and U.K. provide more significant oil parameters in the "bull" period, indicating these nations to be more consistently interconnected with oil than the importing nations. Both the models with normal and t-distributed innovations provide much of the same evidence, but the t-distributed models provide even clearer evidence that the oil importing nations are less correlated during low volatile periods. WTI and brent have more or less the same relationship to the respective market, with the exceptions of a slightly stronger correlation between WTI and U.S, and European countries and brent. It is also worth noting that LR tests for models without oil are strongly rejected for the alternative with oil variables in the MS models.

Correlation estimation based on the regime classification was done in Chapter 6. The regime classification was done with use of the Viterbi algorithm, which calculates the most probable state sequence. The correlation between oil and national indices was estimated with respect to both oil and market regimes. With the national indices as home reference to the regimes, the correlation ratio was overall lower for the oil exporting countries, but with an overall higher and consistent correlation. In addition, the U.S. and Canada have a stronger consistent correlation with WTI, and likewise with the European indices and brent oil. The consistent higher correlation in high volatile periods suggests that national indices and oil markets tend to be more correlated during high volatile periods in national markets.

The volatility regimes provided from the oil returns provide very similar results as the regimes for the markets. However, the high volatile periods for oil do not have the same consistent high correlation between markets and oil. Also, the differences between normal and t-distributed innovations seem to play a significant role, indicating that the cause of the regime shifts for oil do not consistently provide the same regime shifts in the markets. This coincides with the preliminary analysis. Overall, the oil shocks result in higher correlation with the national markets. However, this result implies that the root cause of the oil shocks play an important part in the ways the national markets respond.

In general, the oil exporting nations are more affected by the oil price, regardless of high or low volatile periods. Oil importing nations are, on the other hand, less influenced by the oil price, especially in bull-periods.

Extension of the research could be to incorporate more complex MS models such a multivariate and ARCH models. Further investigation on asset allocation and portfolio management based on regime switching would be a natural step in further research.

8. Appendix A: Plots

8.1 QQ-Plot

8.1.1 Normal distribution

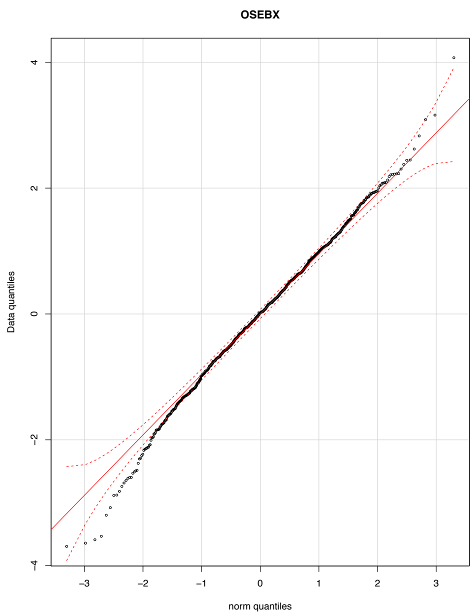


FIGURE 8.1: qq-plot of residuals calculated from Markov switching model with normally distributed innovations and Viterbi algorithm for OSEBX

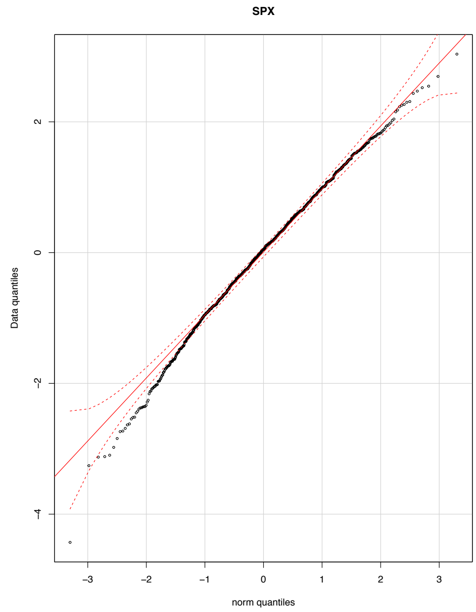


FIGURE 8.2: qq-plot of residuals calculated from Markov switching model with normally distributed innovations and Viterbi algorithm for SPX

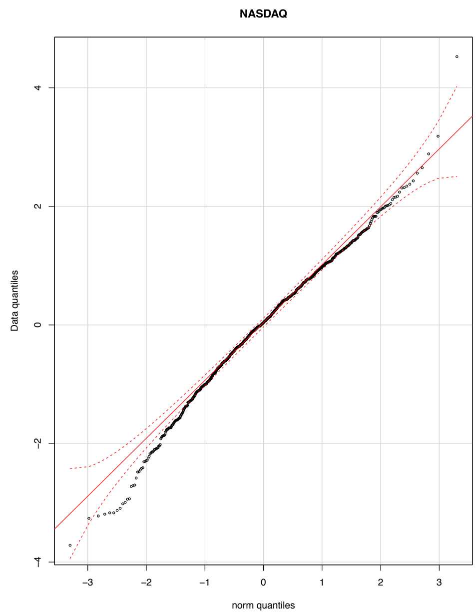


FIGURE 8.3: qq-plot of residuals calculated from Markov switching model with normally distributed innovations and Viterbi algorithm for NASDAQ

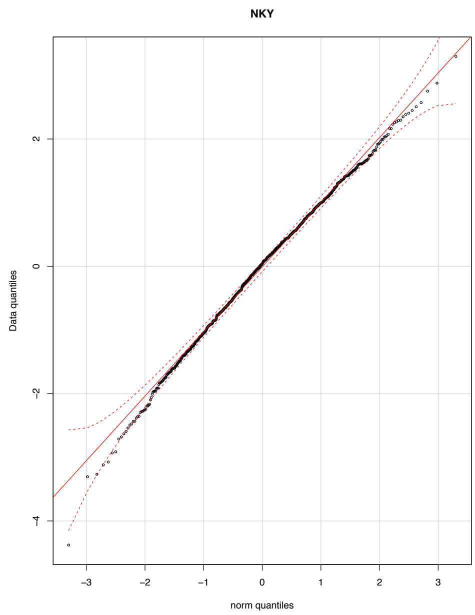


FIGURE 8.4: qq-plot of residuals calculated from Markov switching model with normally distributed innovations and Viterbi algorithm for NKY

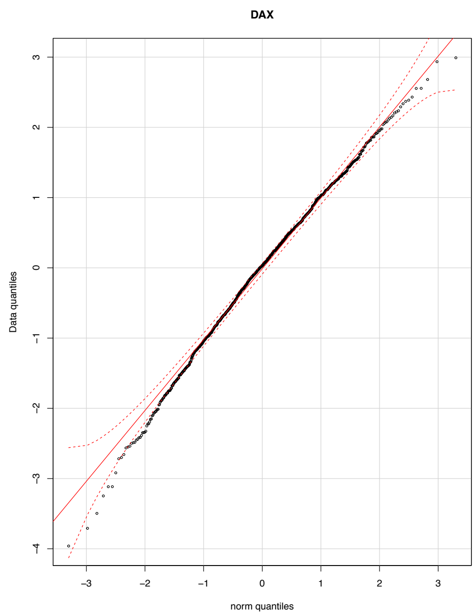


FIGURE 8.5: qq-plot of residuals calculated from Markov switching model with normally distributed innovations and Viterbi algorithm for DAX

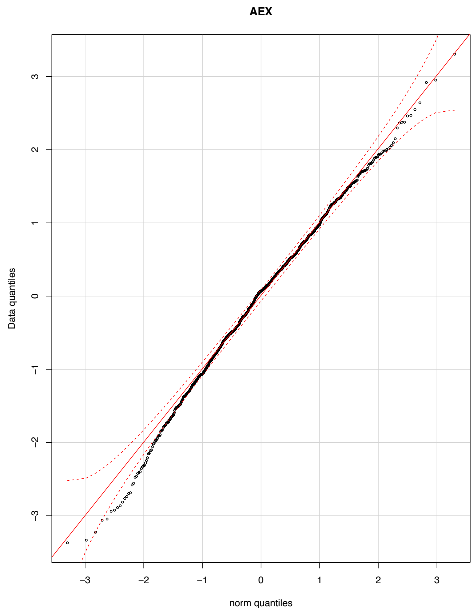


FIGURE 8.6: qq-plot of residuals calculated from Markov switching model with normally distributed innovations and Viterbi algorithm for AEX

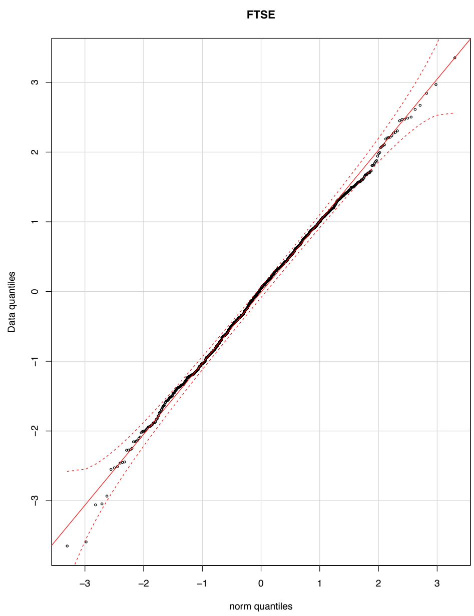


FIGURE 8.7: qq-plot of residuals calculated from Markov switching model with normally distributed innovations and Viterbi algorithm for FTSE

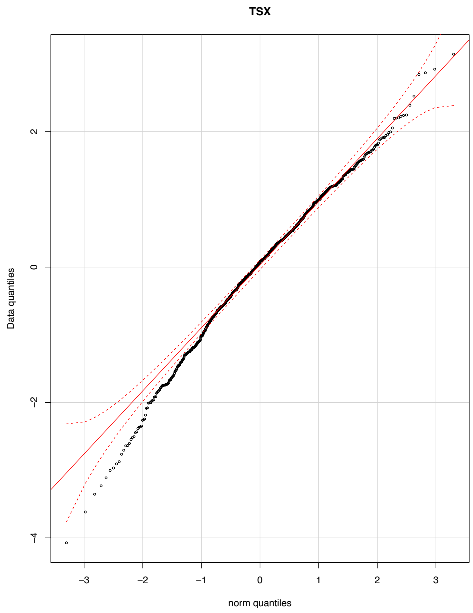


FIGURE 8.8: qq-plot of residuals calculated from Markov switching model with normally distributed innovations and Viterbi algorithm for TSX

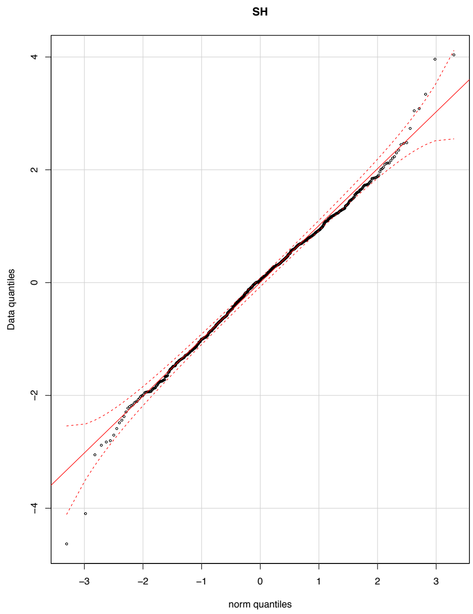


FIGURE 8.9: qq-plot of residuals calculated from Markov switching model with normally distributed innovations and Viterbi algorithm for SH

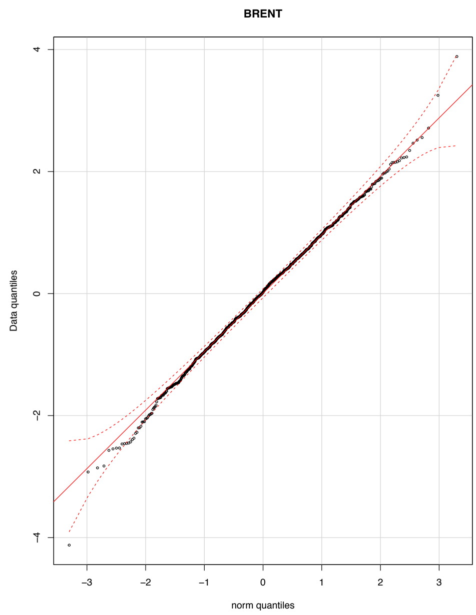


FIGURE 8.10: qq-plot of residuals calculated from Markov switching model with normally distributed innovations and Viterbi algorithm for Brent

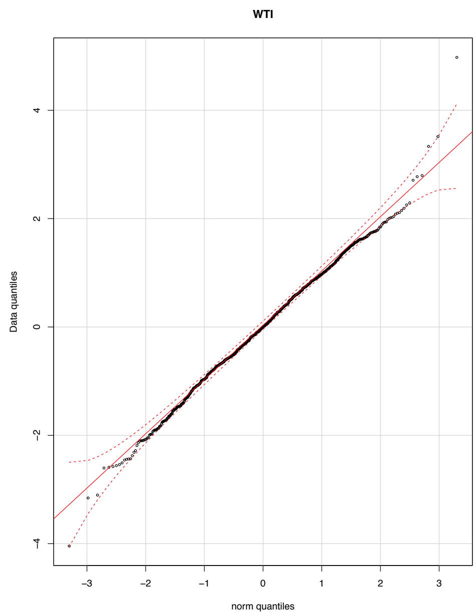


FIGURE 8.11: qq-plot of residuals calculated from Markov switching model with normally distributed innovations and Viterbi algorithm for WTI

8.1.2 T-distribution

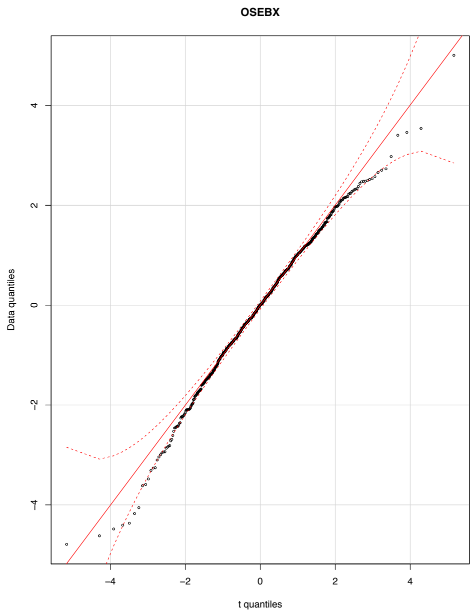


FIGURE 8.12: qq-plot of residuals calculated from Markov switching model with t-distributed innovations and Viterbi algorithm for OSEBX

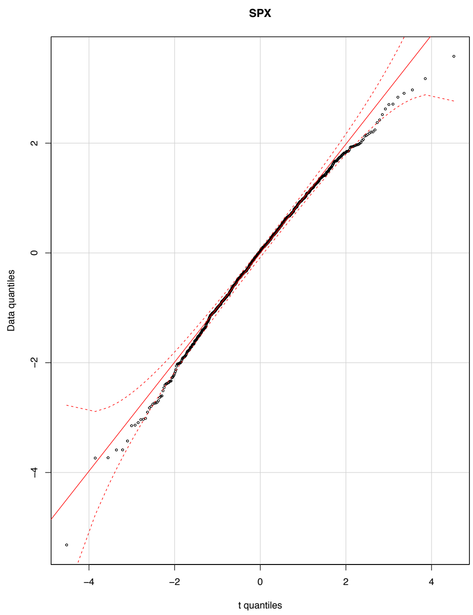


FIGURE 8.13: qq-plot of residuals calculated from Markov switching model with t-distributed innovations and Viterbi algorithm for SPX

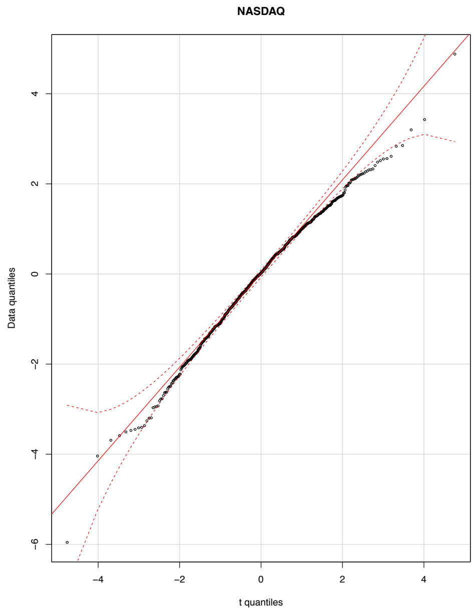


FIGURE 8.14: qq-plot of residuals calculated from Markov switching model with t-distributed innovations and Viterbi algorithm for NASDAQ

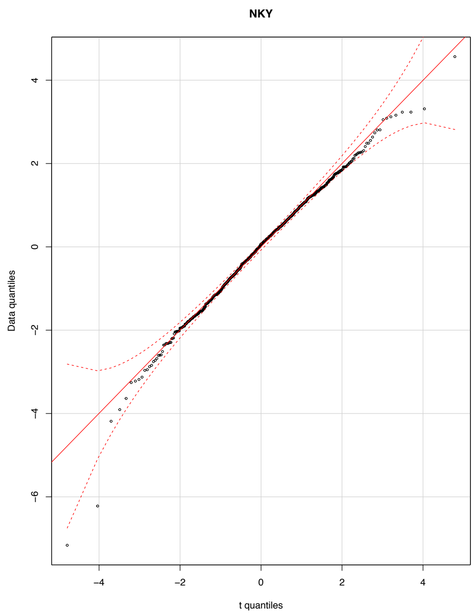


FIGURE 8.15: qq-plot of residuals calculated from Markov switching model with t-distributed innovations and Viterbi algorithm for NKY

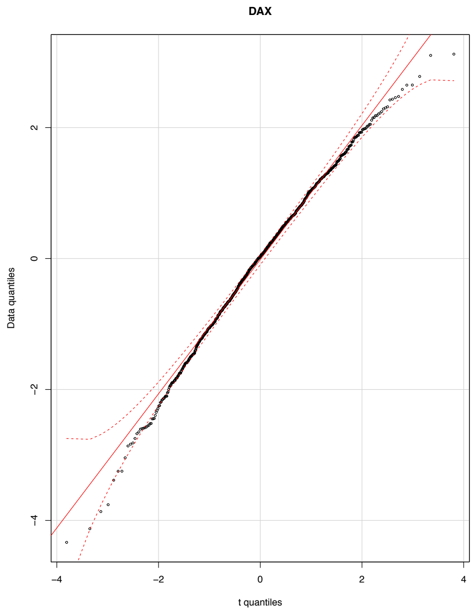


FIGURE 8.16: qq-plot of residuals calculated from Markov switching model with t-distributed innovations and Viterbi algorithm for DAX

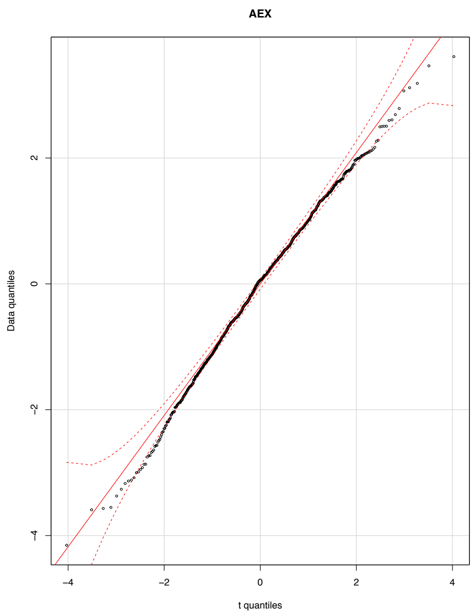


FIGURE 8.17: qq-plot of residuals calculated from Markov switching model with t-distributed innovations and Viterbi algorithm for AEX

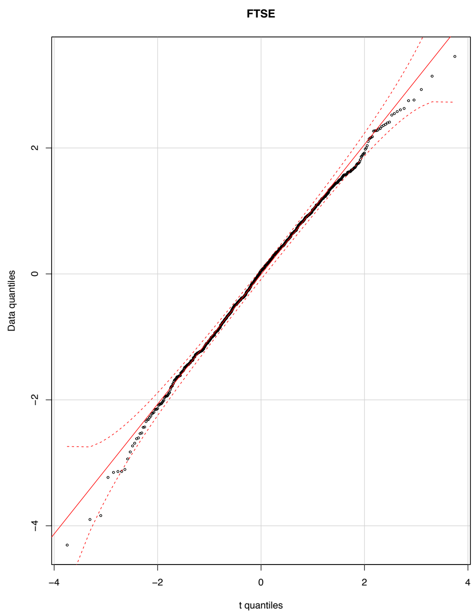


FIGURE 8.18: qq-plot of residuals calculated from Markov switching model with t-distributed innovations and Viterbi algorithm for FTSE

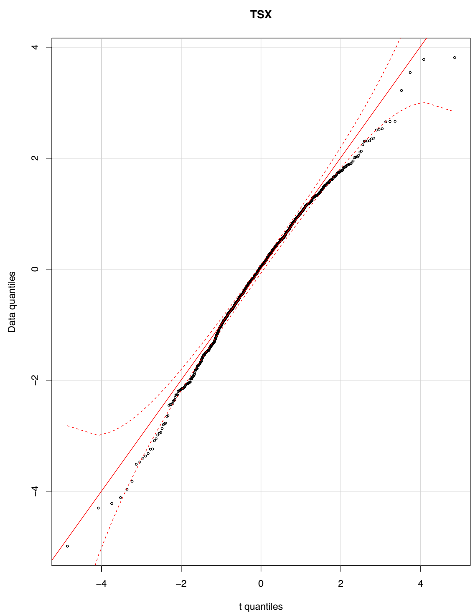


FIGURE 8.19: qq-plot of residuals calculated from Markov switching model with t-distributed innovations and Viterbi algorithm for TSX

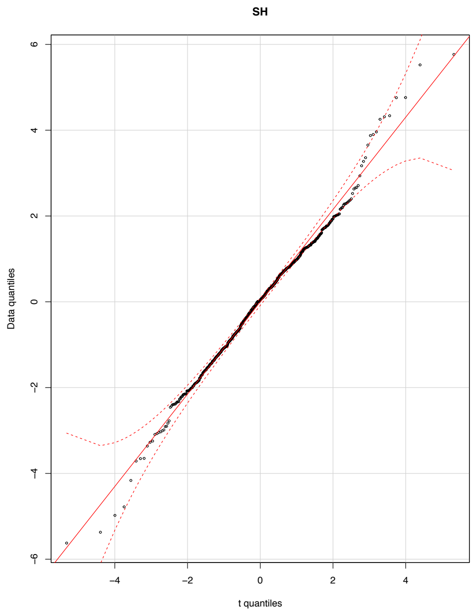


FIGURE 8.20: qq-plot of residuals calculated from Markov switching model with t-distributed innovations and Viterbi algorithm for SH

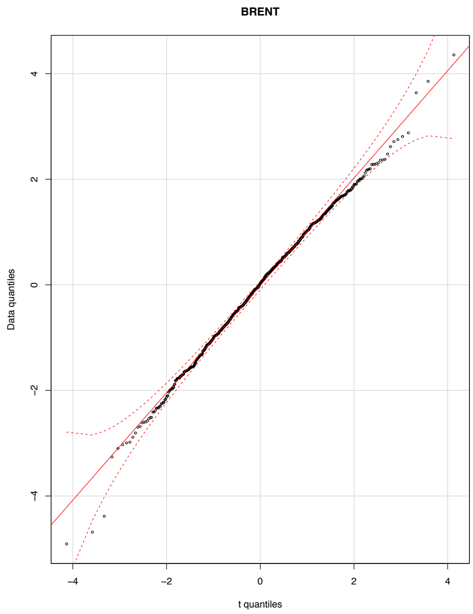


FIGURE 8.21: qq-plot of residuals calculated from Markov switching model with t-distributed innovations and Viterbi algorithm for Brent

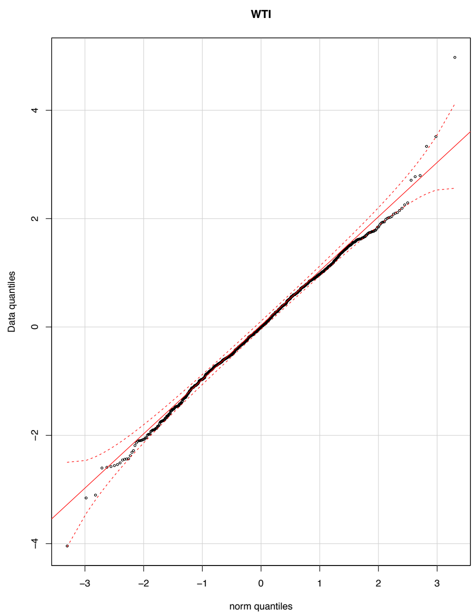


FIGURE 8.22: qq-plot of residuals calculated from Markov switching model with t-distributed innovations and Viterbi algorithm for WTI

8.2 ACF plot

8.2.1 Normal distribution

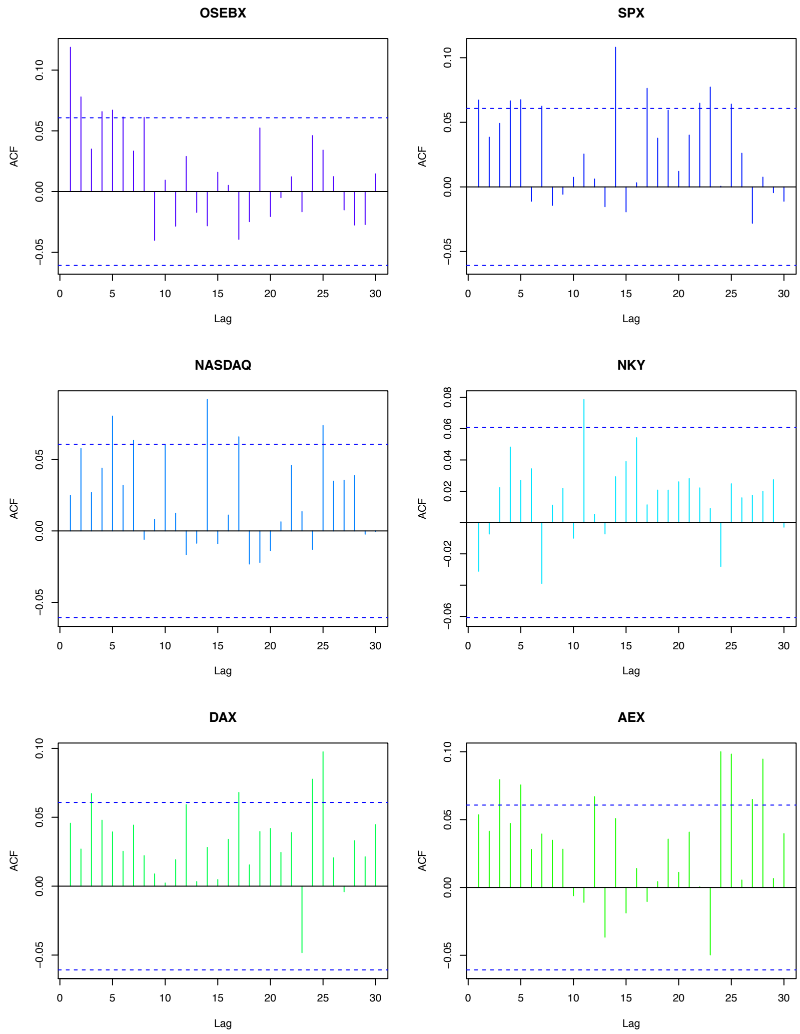


FIGURE 8.23: ACF plot of squared residuals calculated from Markov switching model with normal distributed innovations and Viterbi algorithm

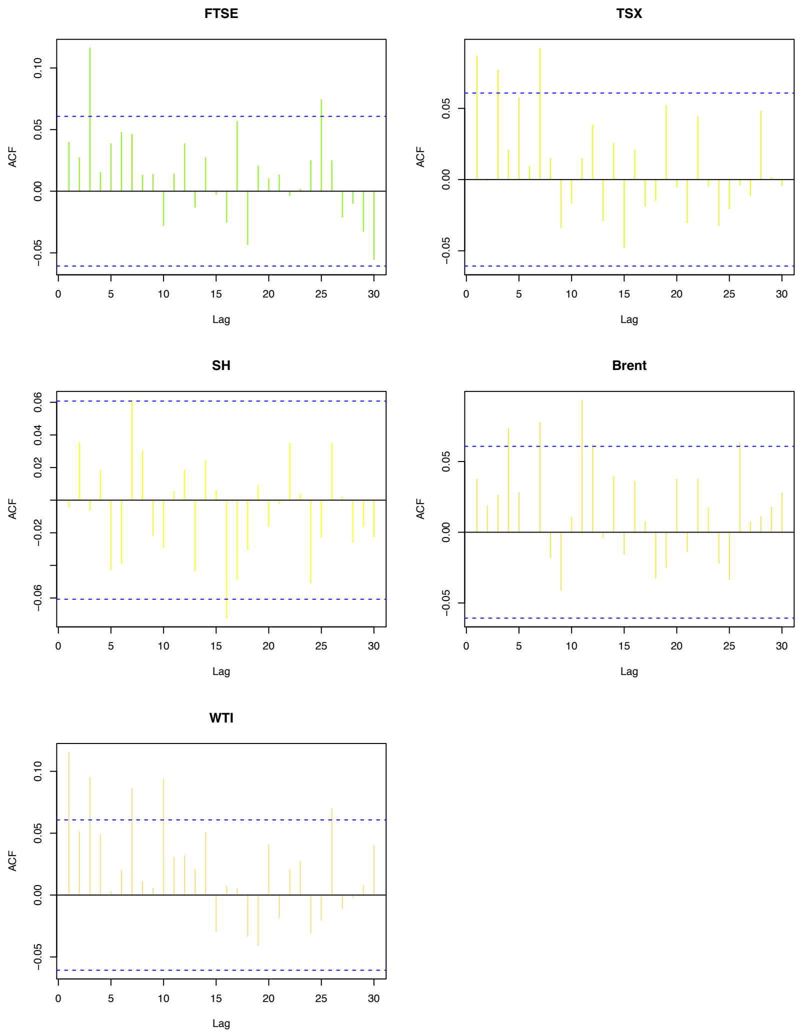


FIGURE 8.24: ACF plot of squared residuals calculated from Markov switching model with normal distributed innovations and Viterbi algorithm

8.2.2 T-distribution

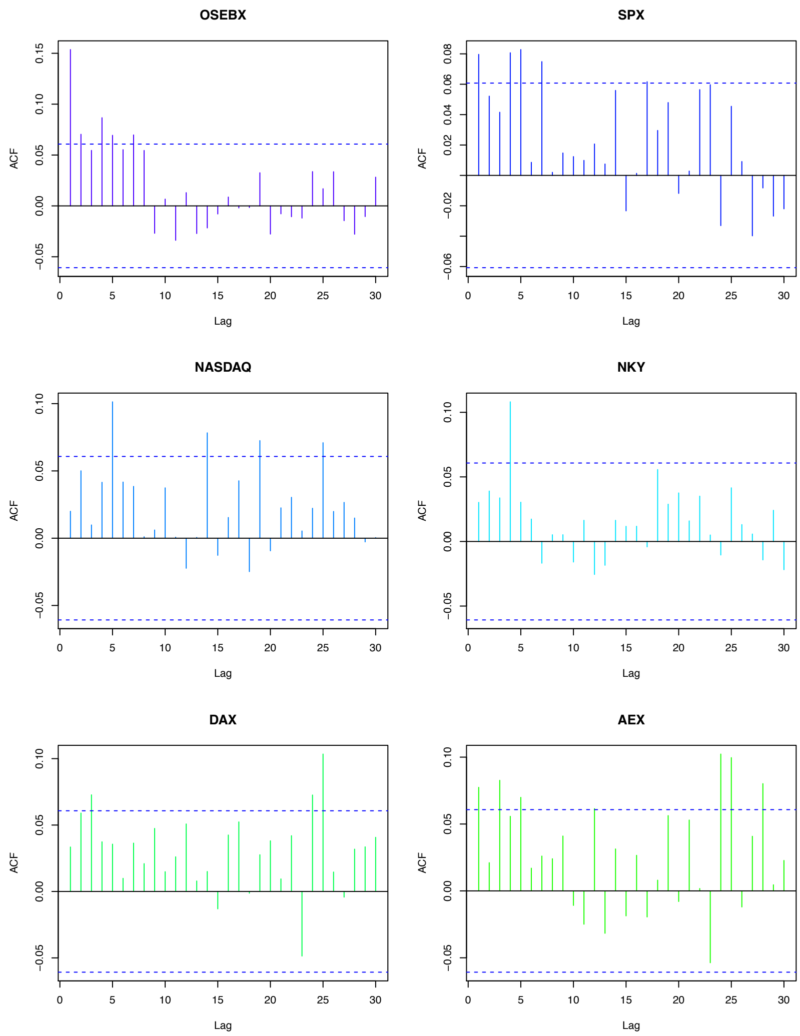


FIGURE 8.25: ACF plot of squared residuals calculated from Markov switching model with t-distributed innovations and Viterbi algorithm

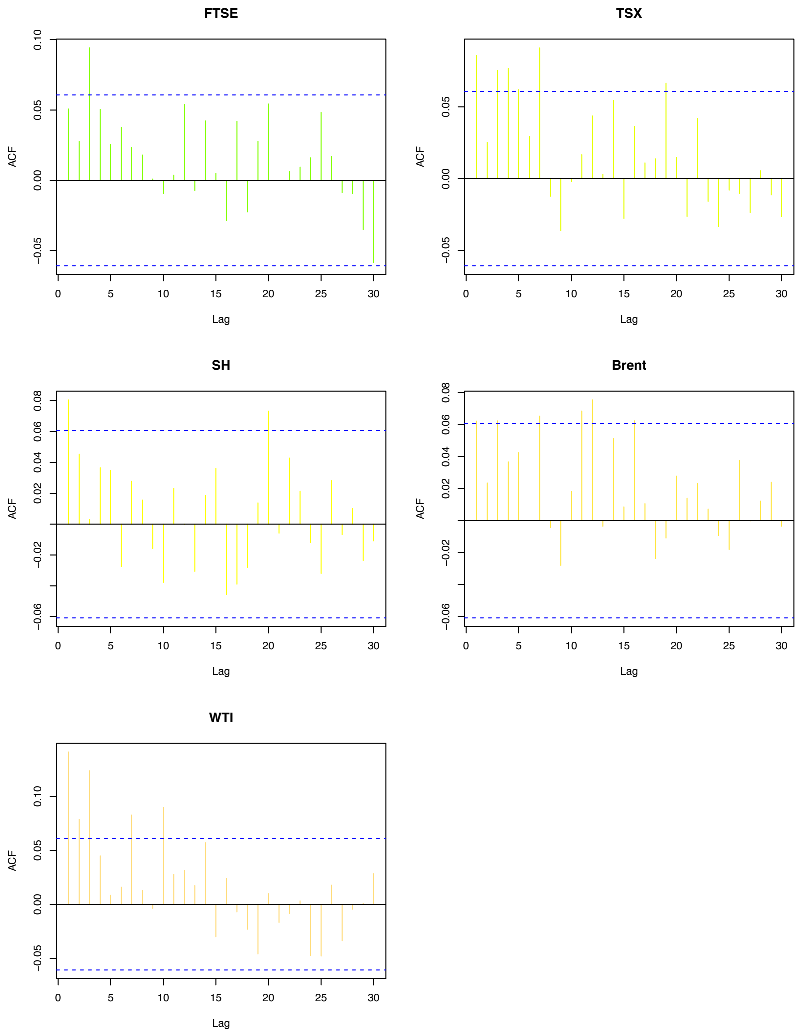


FIGURE 8.26: ACF plot of squared residuals calculated from Markov switching model with t-distributed innovations and Viterbi algorithm

8.3 Smoothed probabilities and conditional standard deviation

8.3.1 Normal distribution

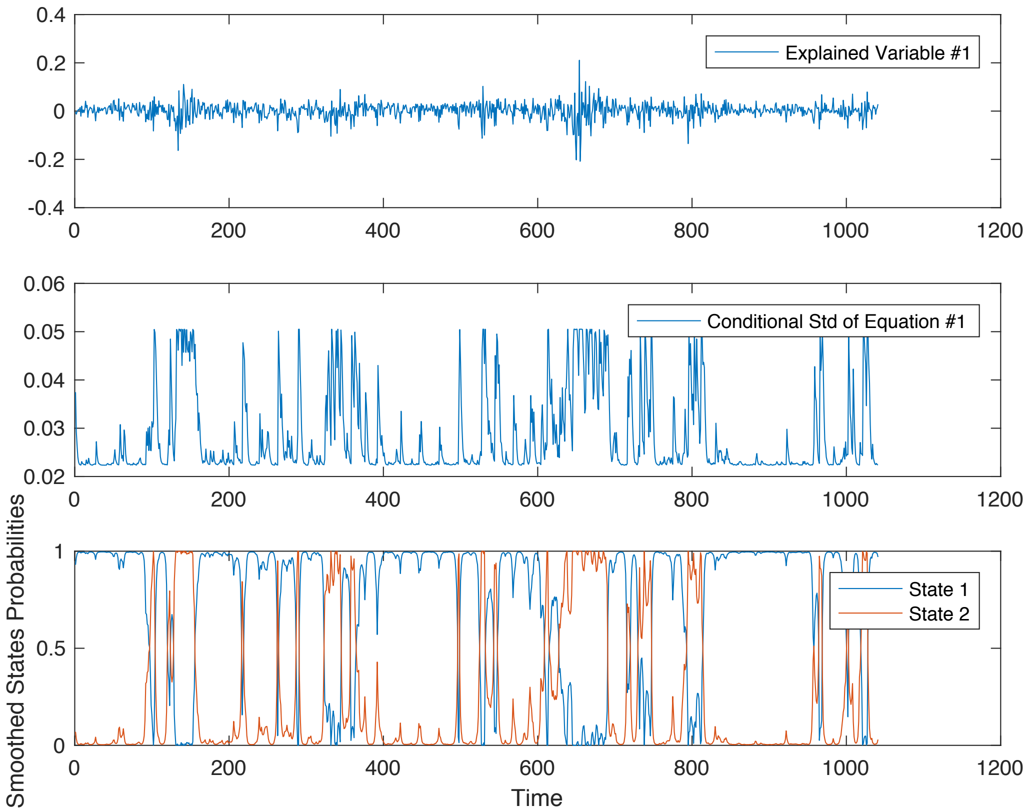


FIGURE 8.27: Plot of observed time series OSEBX with conditional standard deviation and smoothed probability modeled with normal innovations.

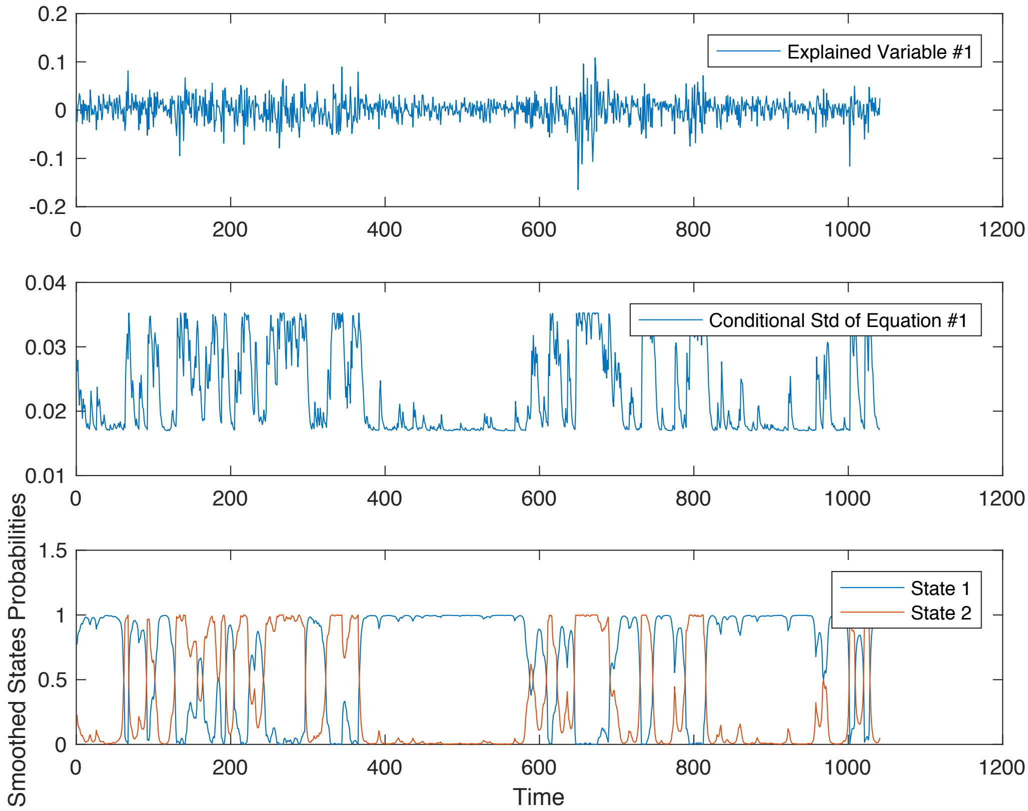


FIGURE 8.28: Plot of observed time series SPX with conditional standard deviation and smoothed probability modeled with normal innovations.

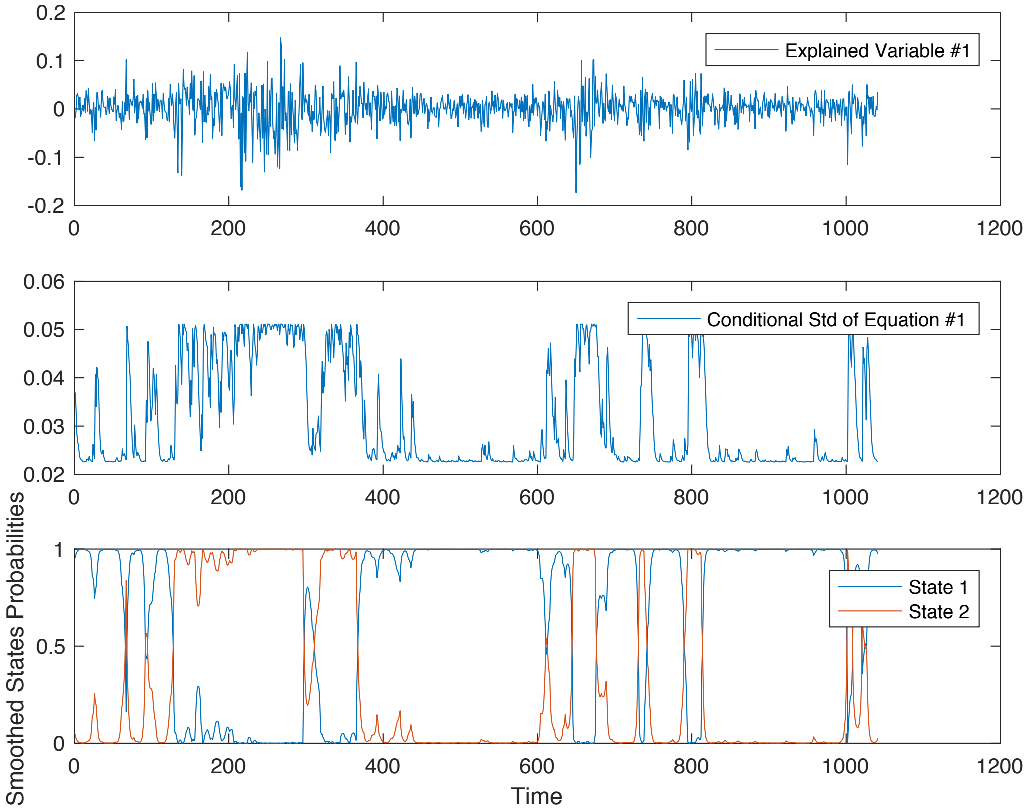


FIGURE 8.29: Plot of observed time series NASDAQ with conditional standard deviation and smoothed probability modeled with normal innovations.

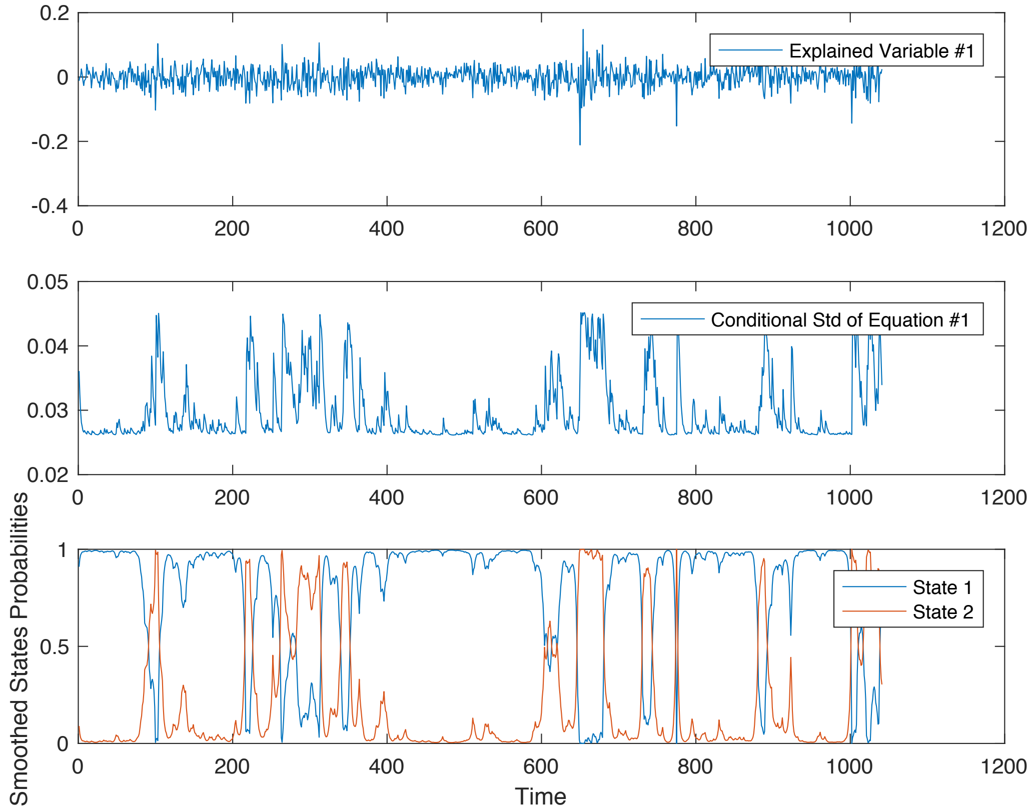


FIGURE 8.30: Plot of observed time series NKY with conditional standard deviation and smoothed probability modeled with normal innovations.

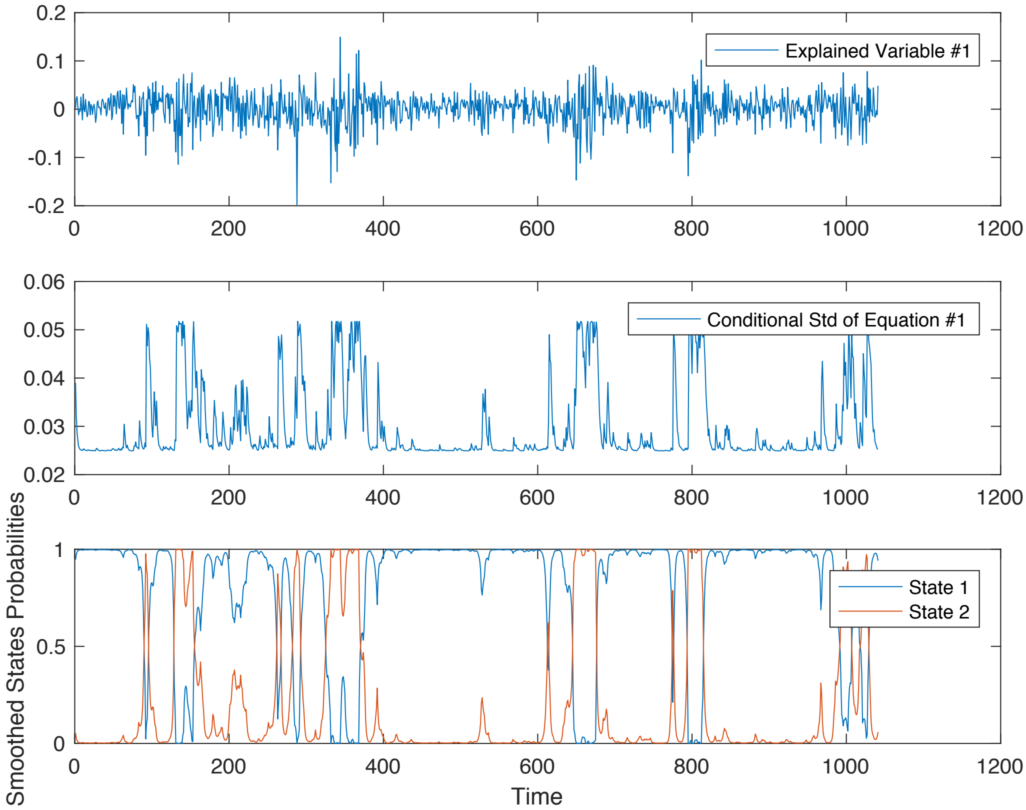


FIGURE 8.31: Plot of observed time series DAX with conditional standard deviation and smoothed probability modeled with normal innovations.

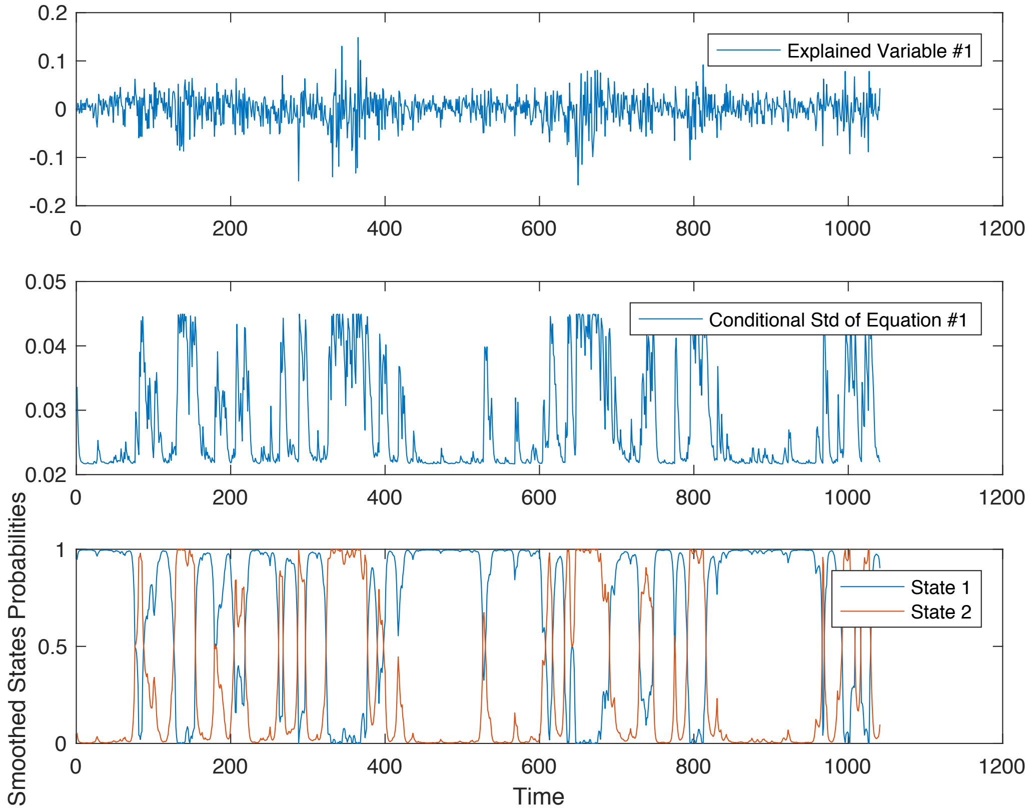


FIGURE 8.32: Plot of observed time series AEX with conditional standard deviation and smoothed probability modeled with normal innovations.

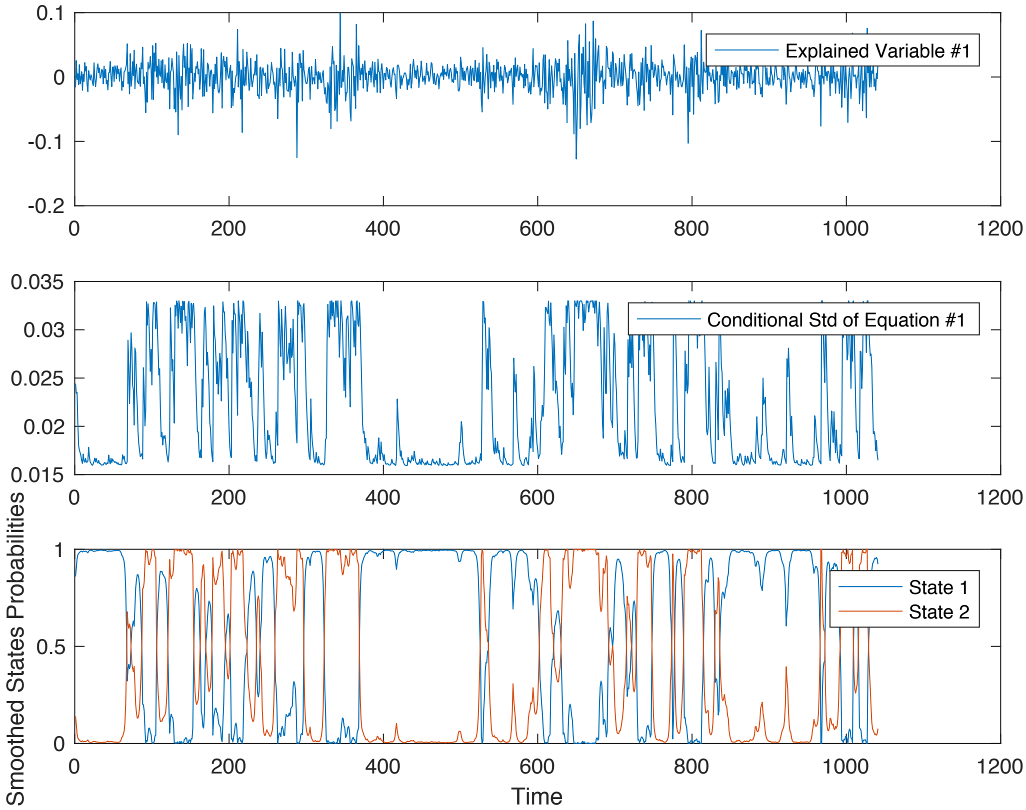


FIGURE 8.33: Plot of observed time series FTSE with conditional standard deviation and smoothed probability modeled with normal innovations.

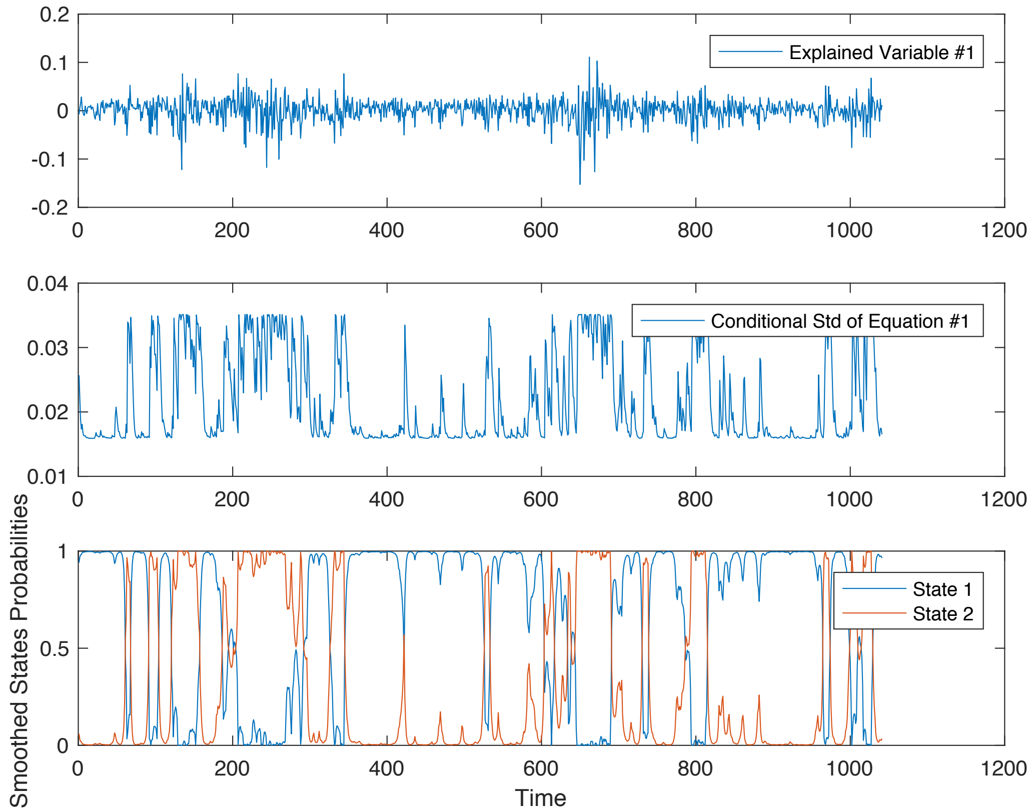


FIGURE 8.34: Plot of observed time series TSX with conditional standard deviation and smoothed probability modeled with normal innovations.

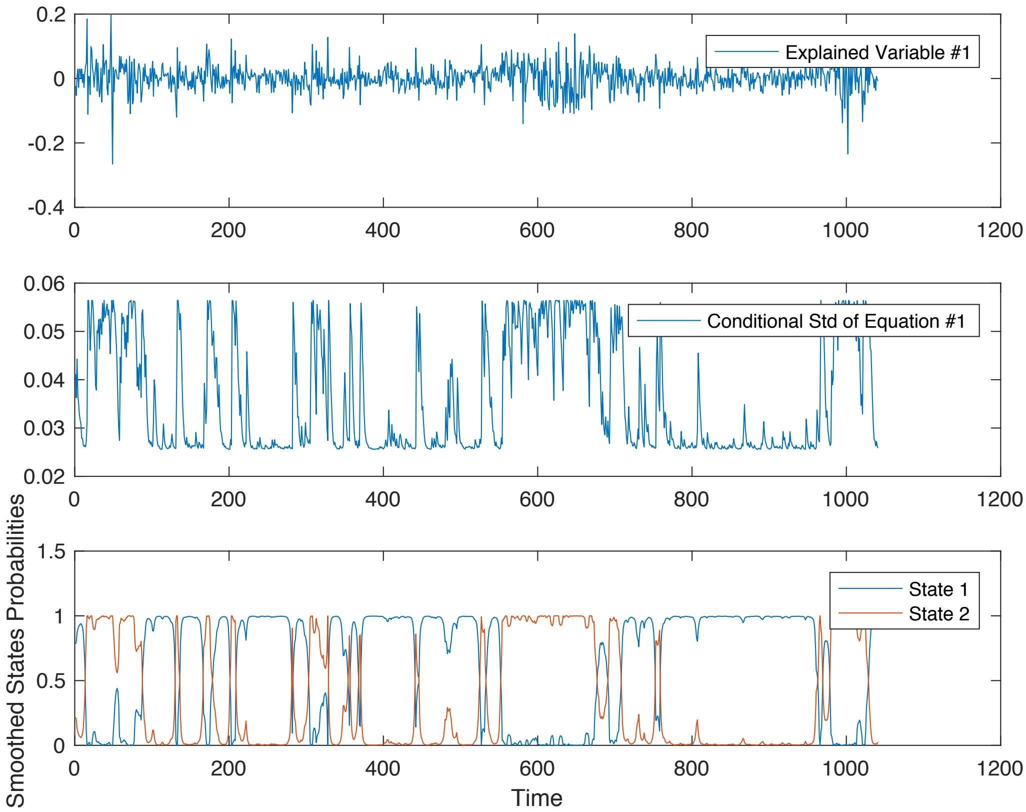


FIGURE 8.35: Plot of observed time series SH with conditional standard deviation and smoothed probability modeled with normal innovations.

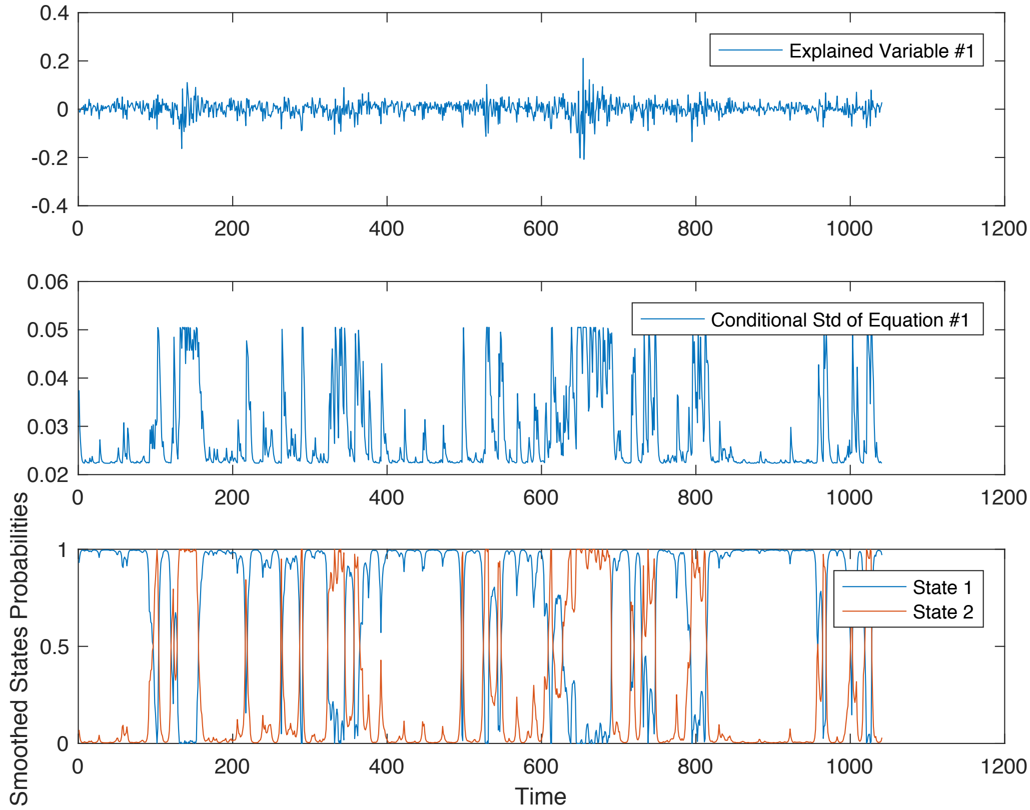


FIGURE 8.36: Plot of observed time series OSEBX with conditional standard deviation and smoothed probability modeled with normal innovations.

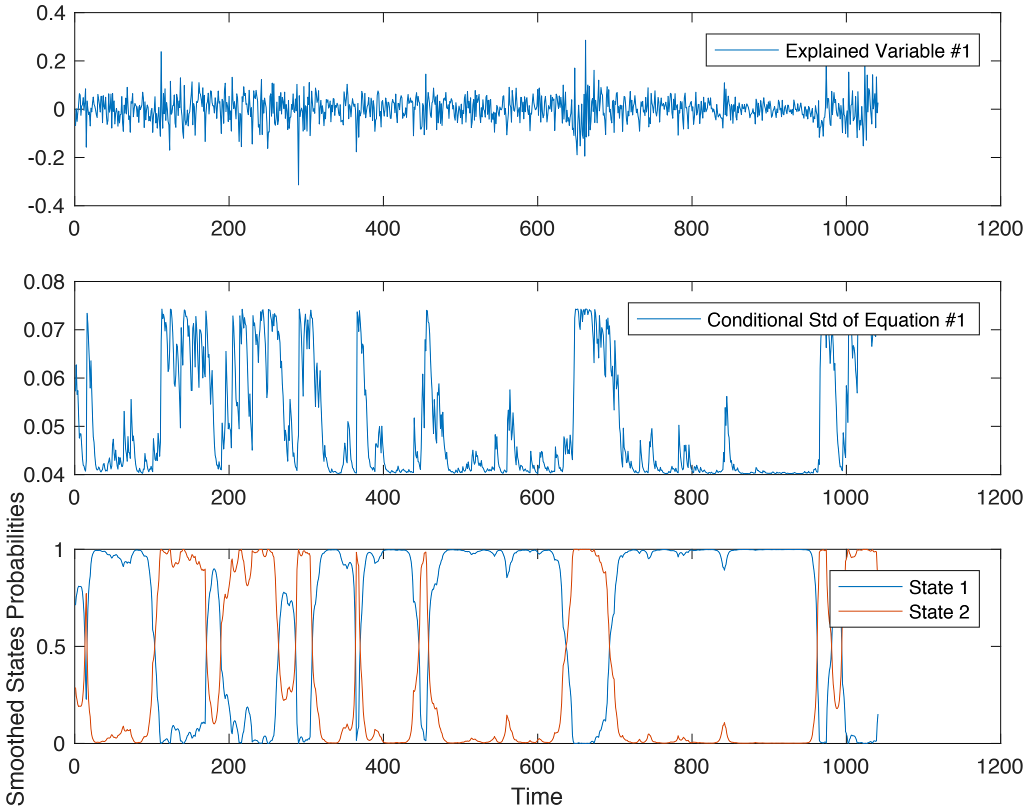


FIGURE 8.37: Plot of observed time series OSEBX with conditional standard deviation and smoothed probability modeled with normal innovations.

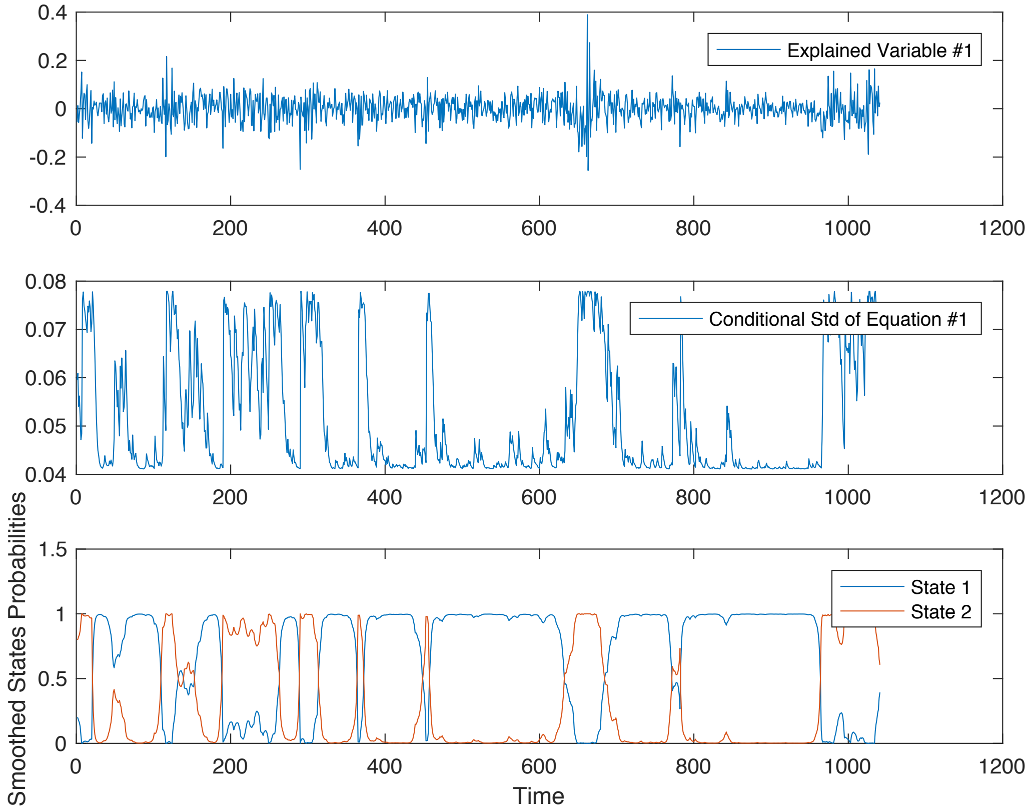


FIGURE 8.38: Plot of observed time series WTI with conditional standard deviation and smoothed probability modeled with normal innovations.

8.3.2 T-distribution

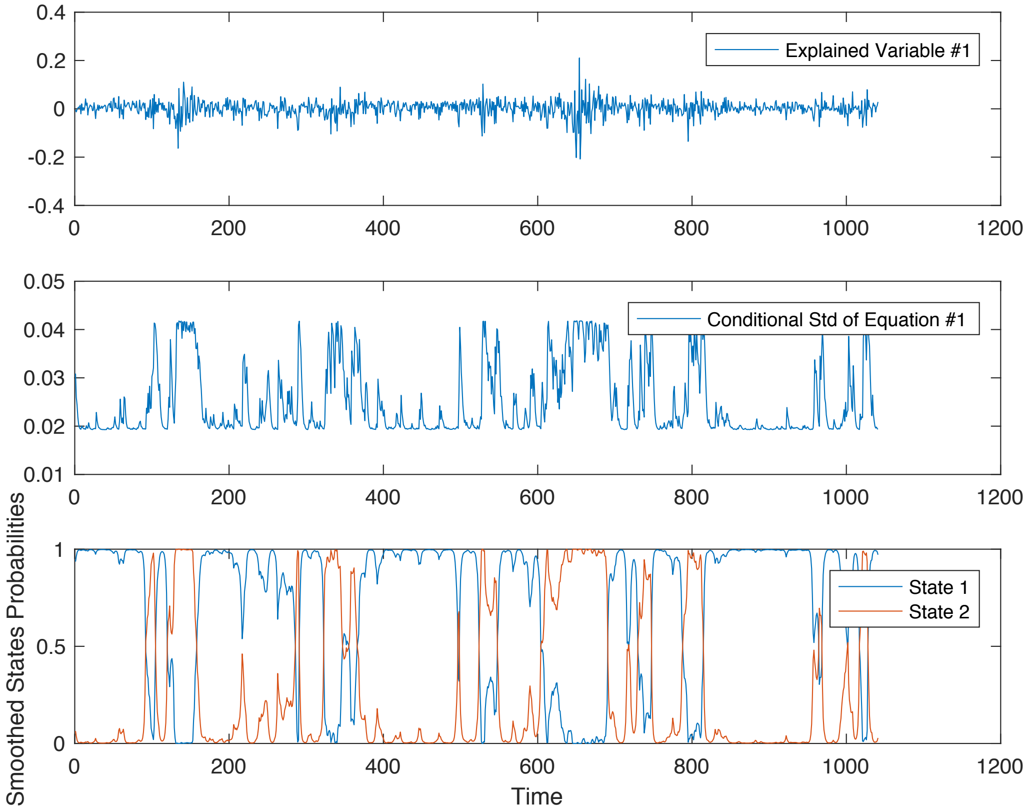


FIGURE 8.39: Plot of observed time series OSEBX with conditional standard deviation and smoothed probability modeled with t-distributed innovations.

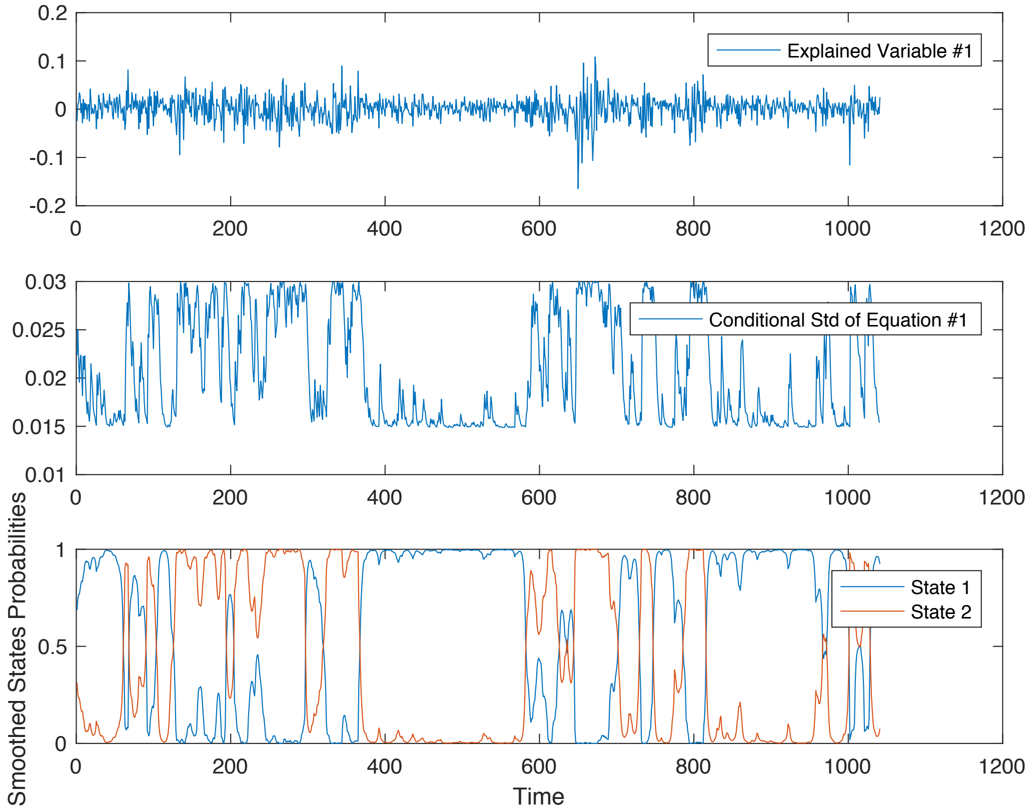


FIGURE 8.40: Plot of observed time series SPX with conditional standard deviation and smoothed probability modeled with t-distributed innovations.

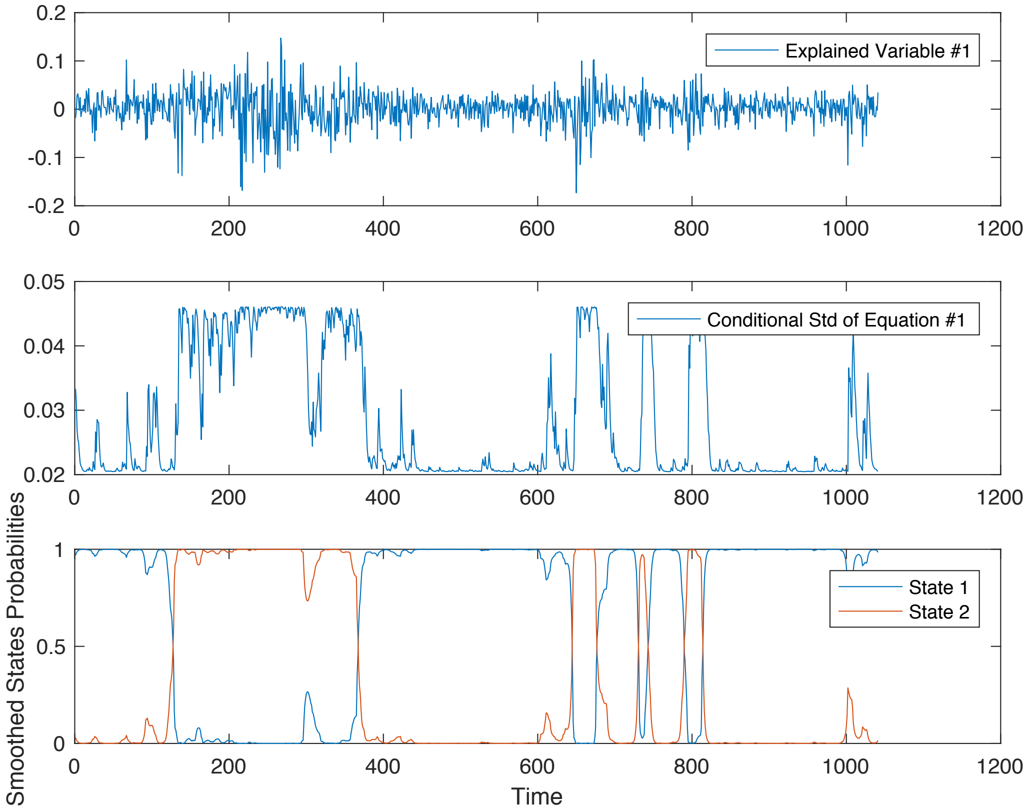


Figure 8.41: Plot of observed time series NASDAQ with conditional standard deviation and smoothed probability modeled with t-distributed innovations.

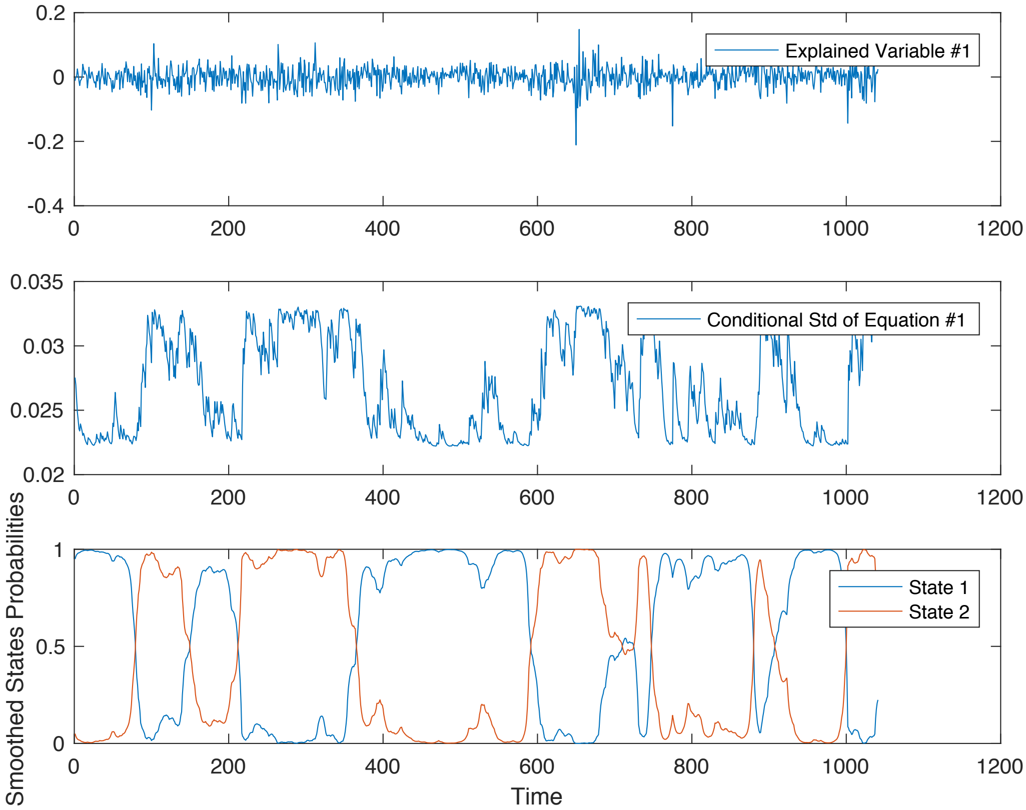


FIGURE 8.42: Plot of observed time series NKY with conditional standard deviation and smoothed probability modeled with t-distributed innovations.

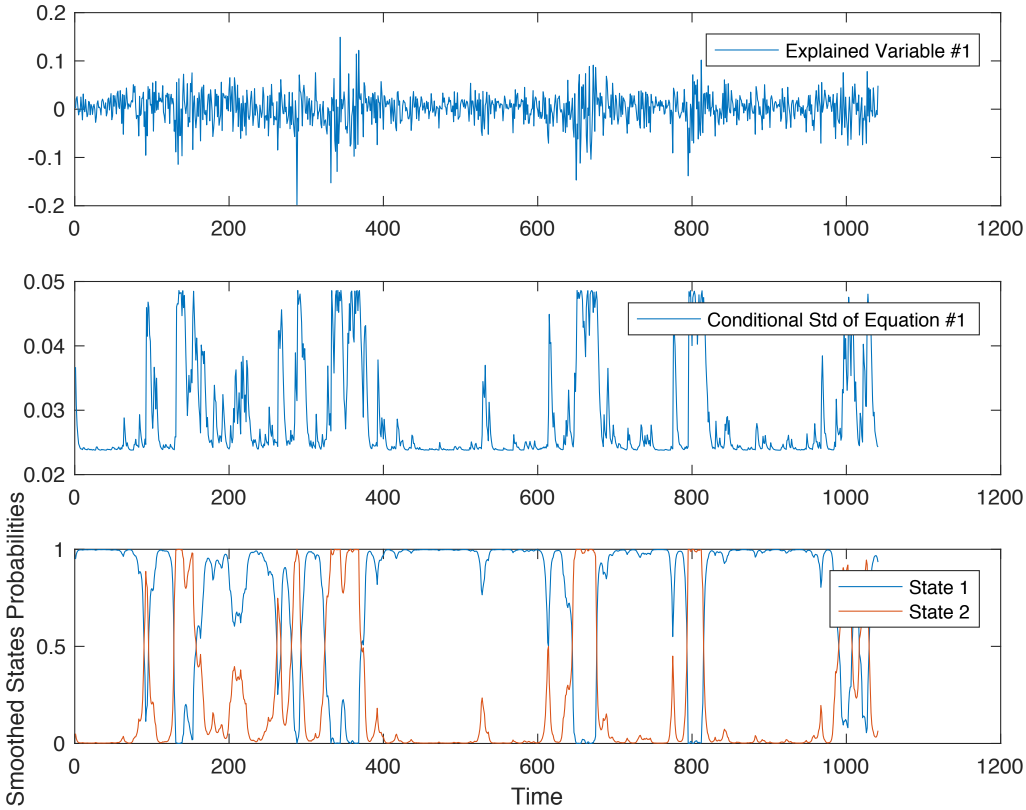


FIGURE 8.43: Plot of observed time series DAX with conditional standard deviation and smoothed probability modeled with t-distributed innovations.

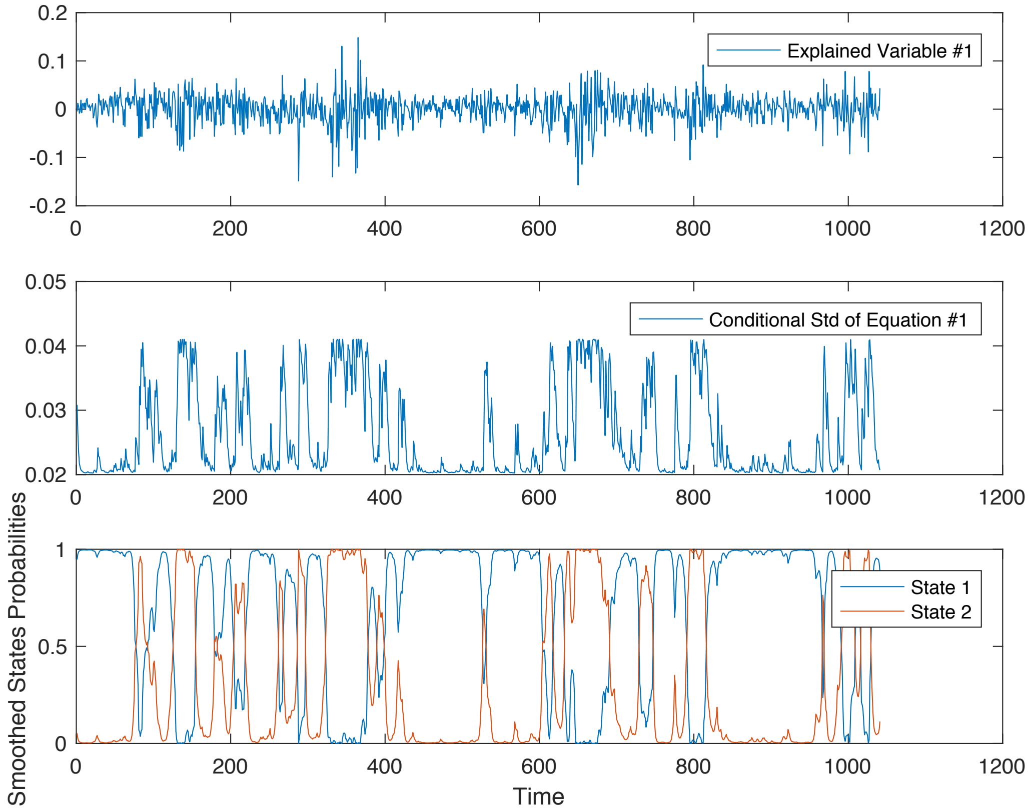


FIGURE 8.44: Plot of observed time series AEX with conditional standard deviation and smoothed probability modeled with t-distributed innovations.

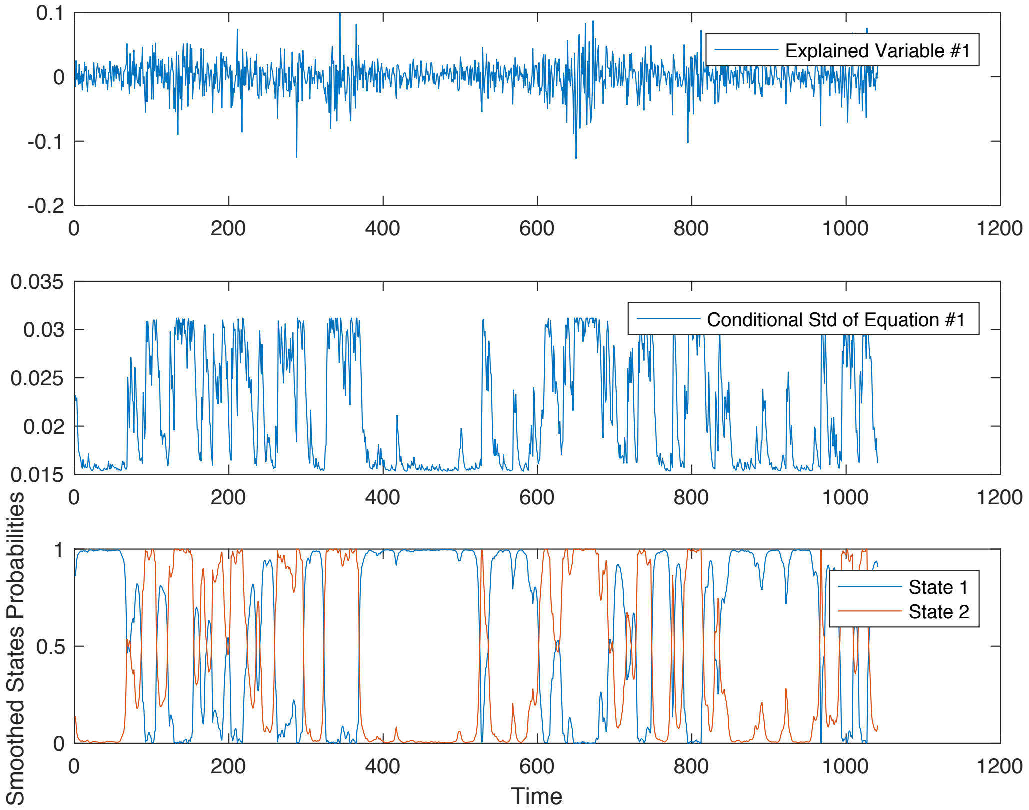


FIGURE 8.45: Plot of observed time series FTSE with conditional standard deviation and smoothed probability modeled with t-distributed innovations.

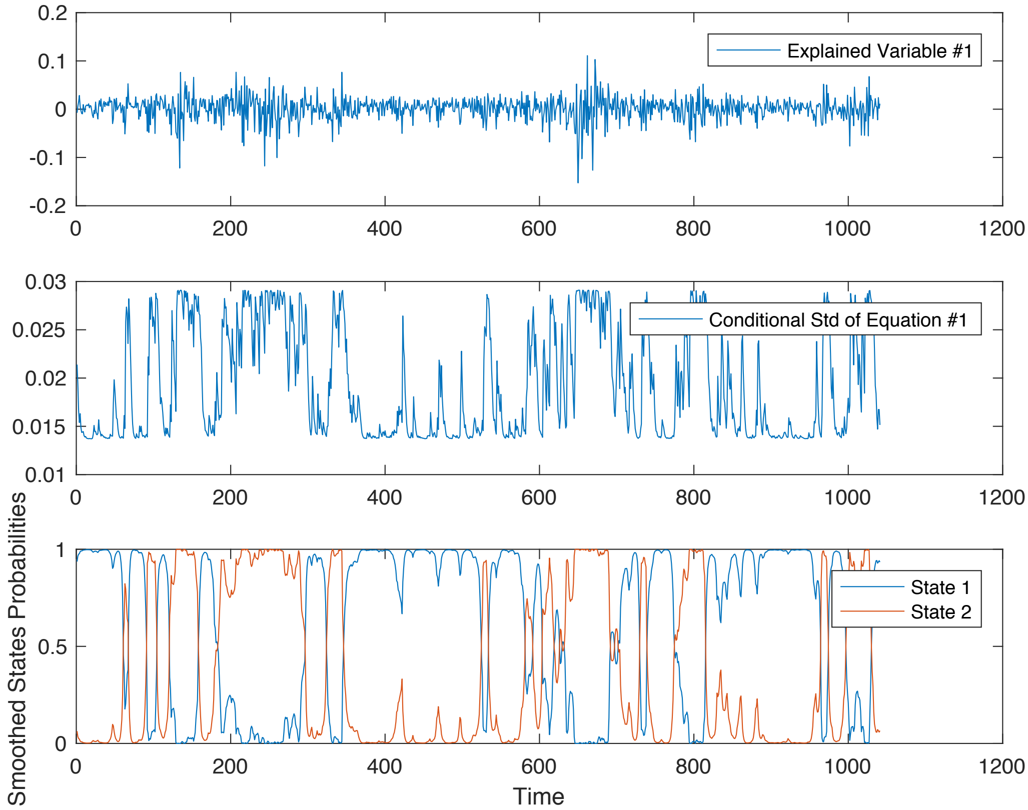


FIGURE 8.46: Plot of observed time series TSX with conditional standard deviation and smoothed probability modeled with t-distributed innovations.

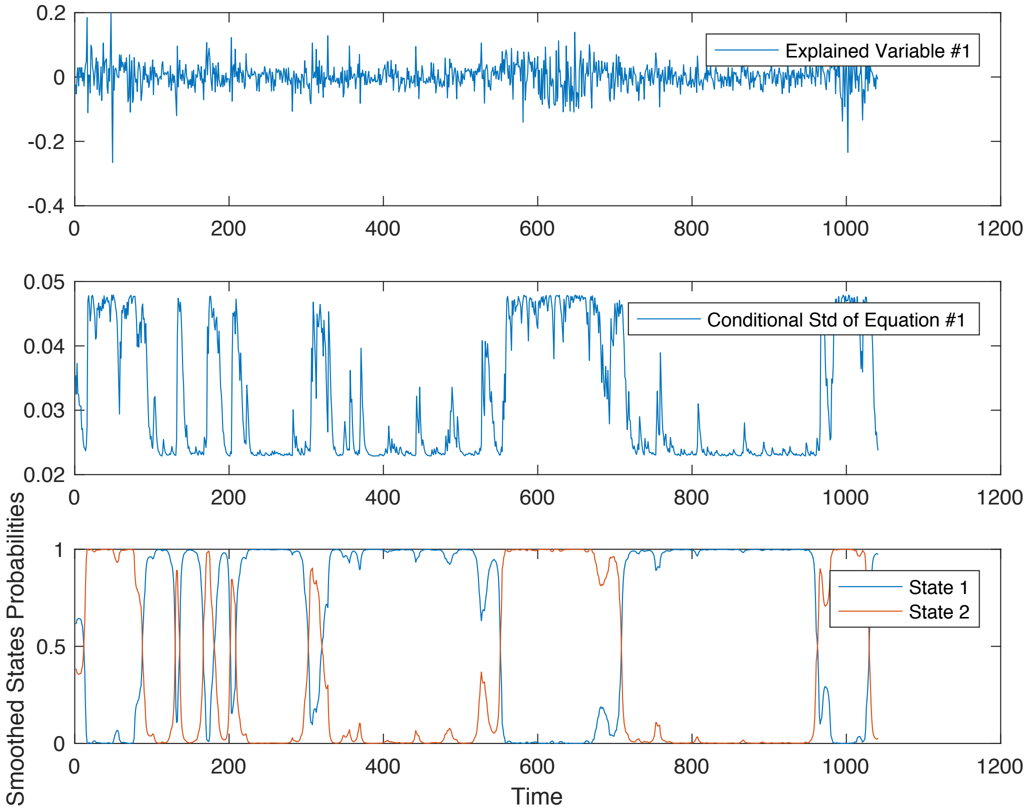


FIGURE 8.47: Plot of observed time series SH with conditional standard deviation and smoothed probability modeled with t-distributed innovations.

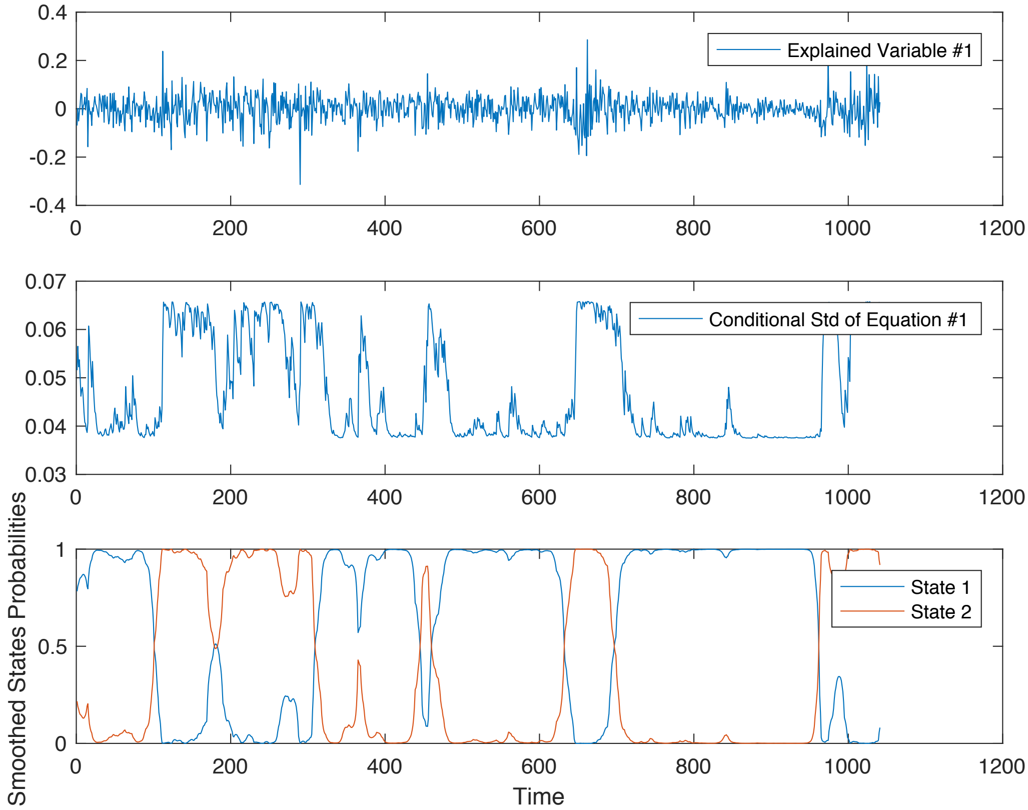


FIGURE 8.48: Plot of observed time series Brent with conditional standard deviation and smoothed probability modeled with t-distributed innovations.

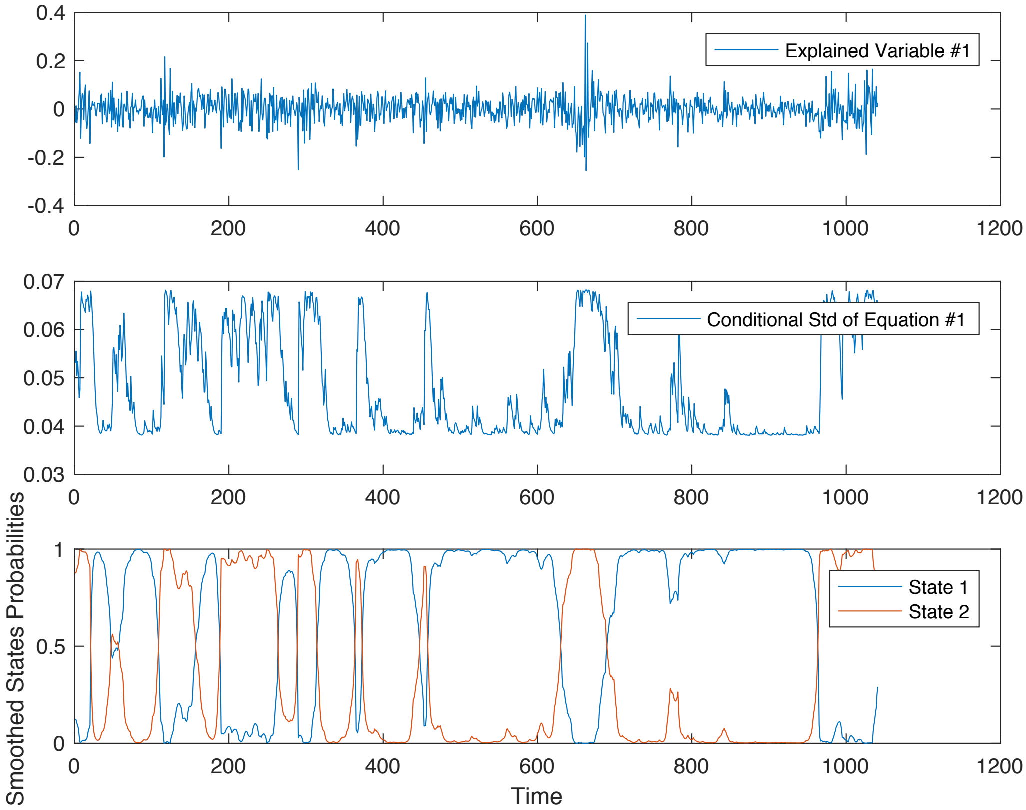
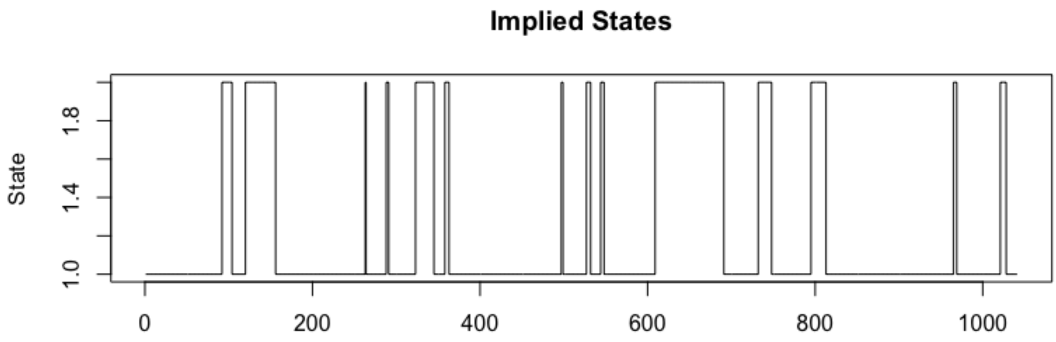
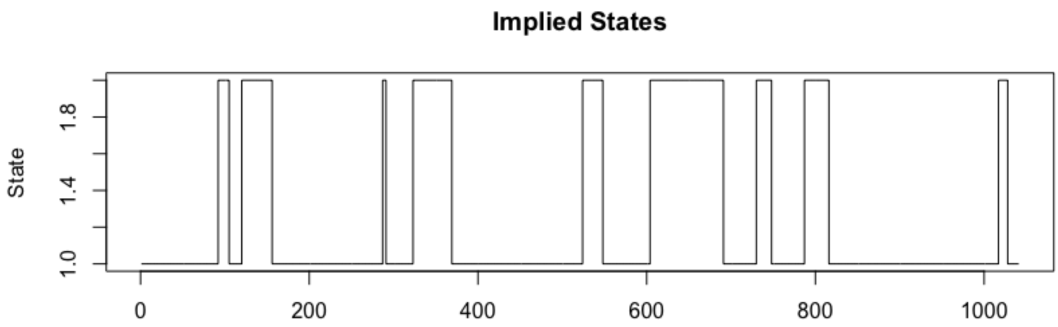


FIGURE 8.49: Plot of observed time series WTI with conditional standard deviation and smoothed probability modeled with t-distributed innovations.

8.4 Estimated state sequence from Viterbi algorithm

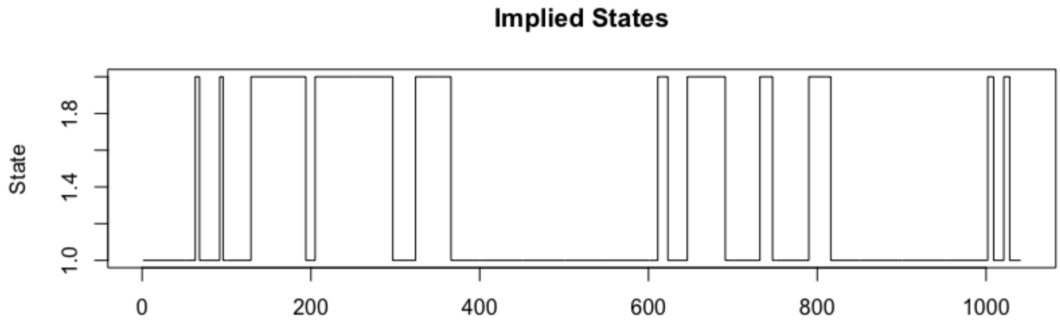


(A) Normal distribution

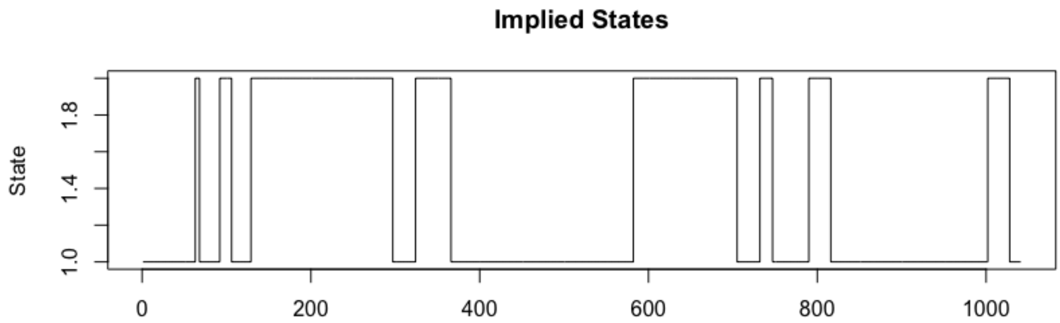


(B) T-distribution

FIGURE 8.50: Estimated state sequence from the Viterbi algorithm of OSEBX

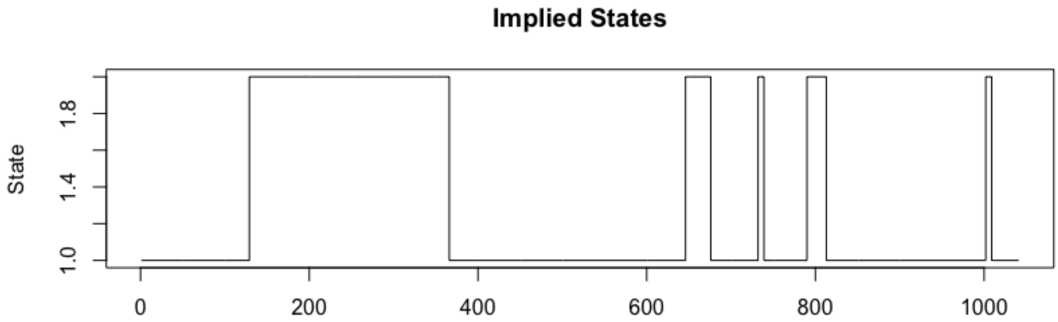


(A) Normal distribution

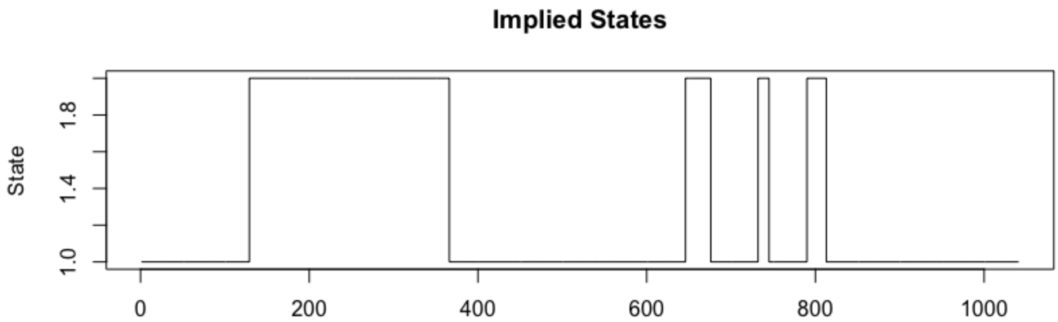


(B) T-distribution

FIGURE 8.51: Estimated state sequence from the Viterbi algorithm of SPX

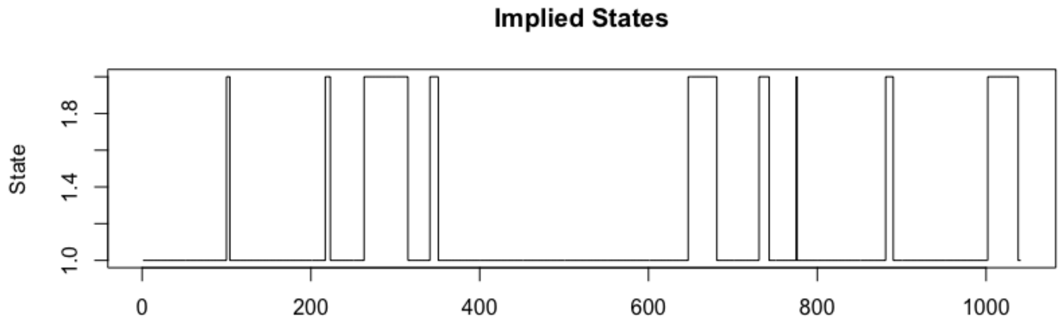


(A) Normal distribution

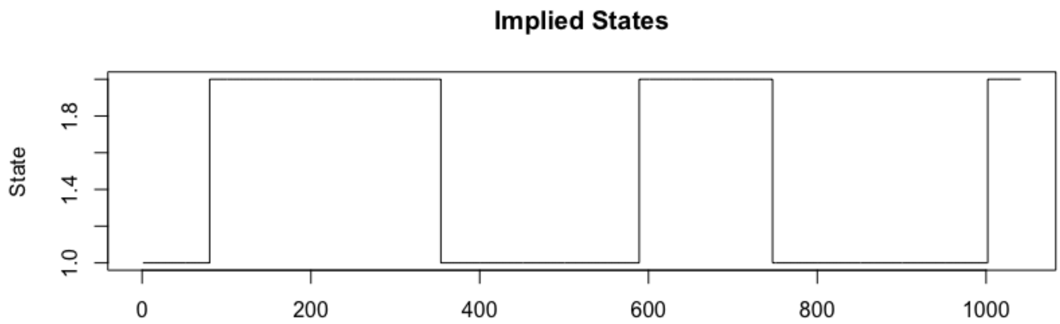


(B) T-distribution

FIGURE 8.52: Estimated state sequence from the Viterbi algorithm of NASDAQ

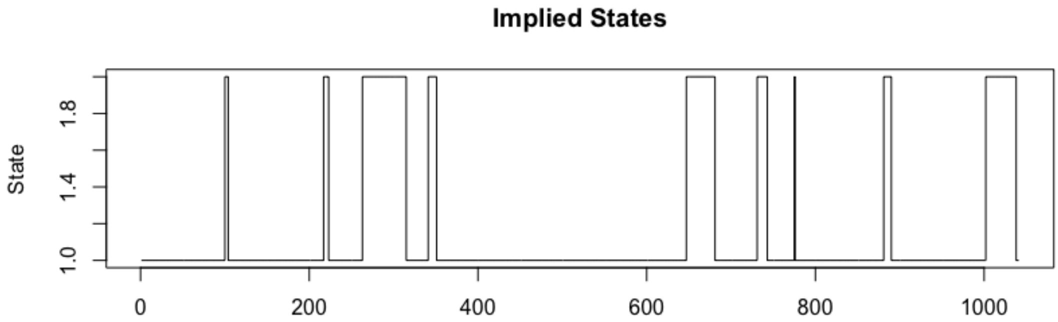


(A) Normal distribution

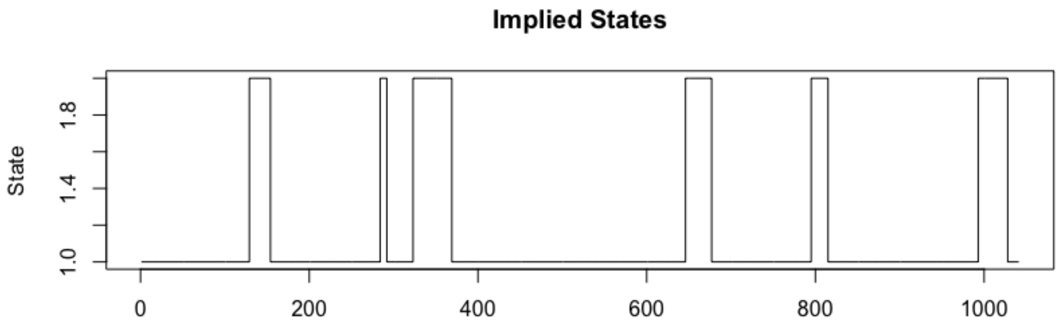


(B) T-distribution

FIGURE 8.53: Estimated state sequence from the Viterbi algorithm of NKY

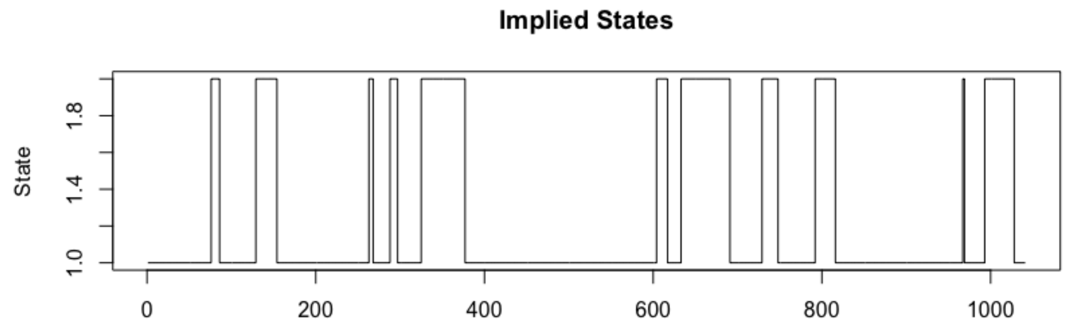


(A) Normal distribution

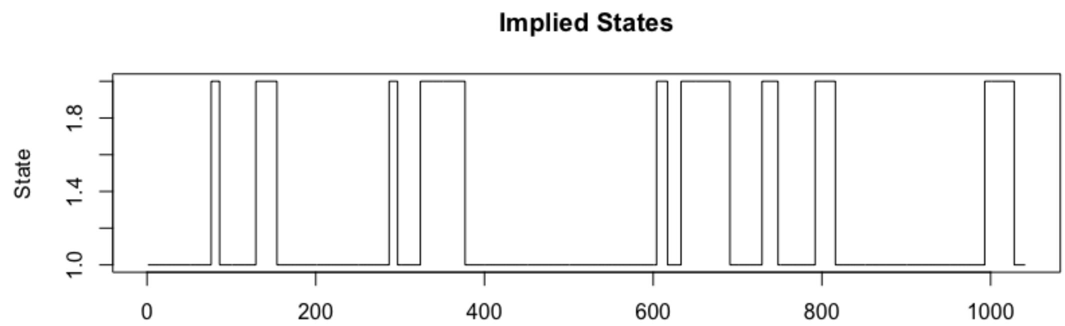


(B) T-distribution

Figure 8.54: Estimated state sequence from the Viterbi algorithm of DAX

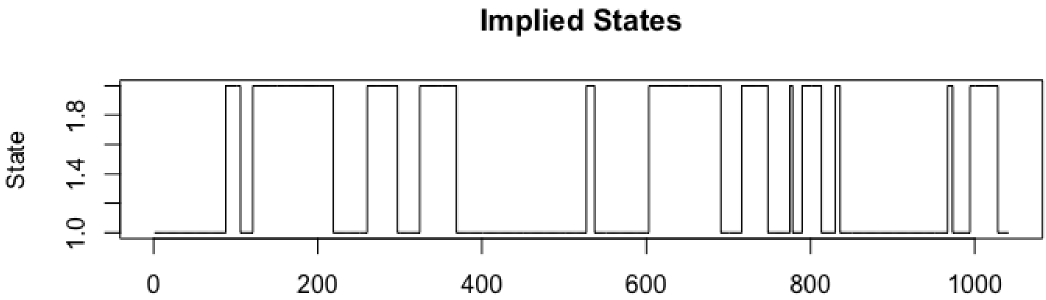


(A) Normal distribution

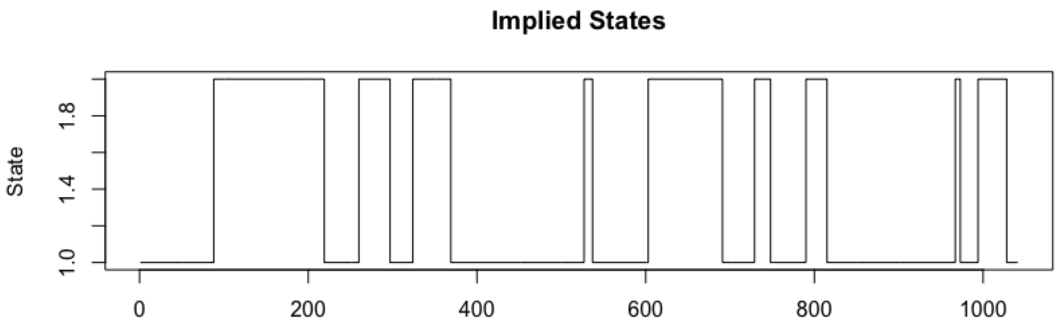


(B) T-distribution

FIGURE 8.55: Estimated state sequence from the Viterbi algorithm of AEX

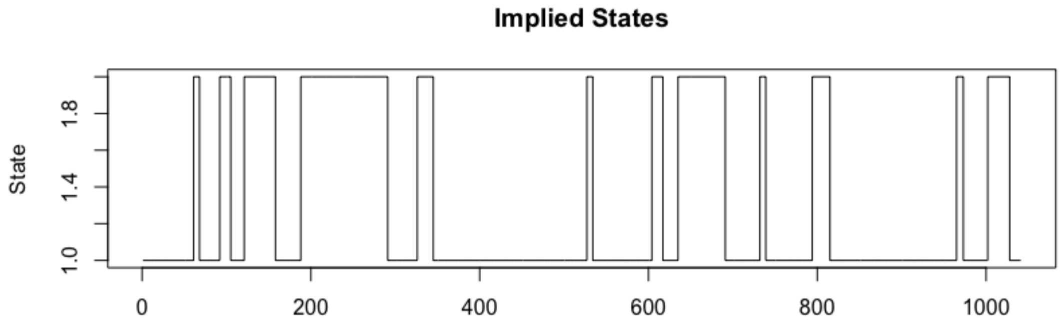


(A) Normal distribution

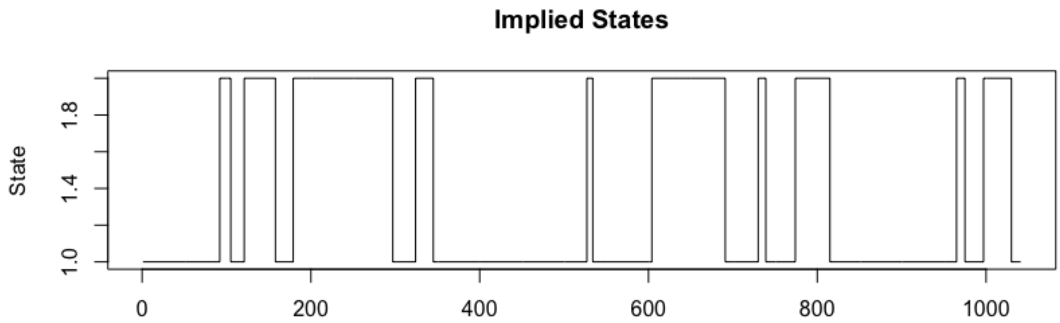


(B) T-distribution

Figure 8.56: Estimated state sequence from the Viterbi algorithm of FTSE

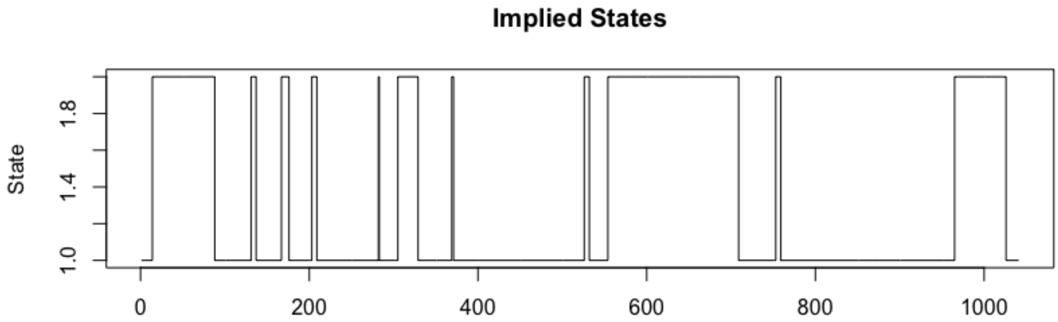


(A) Normal distribution

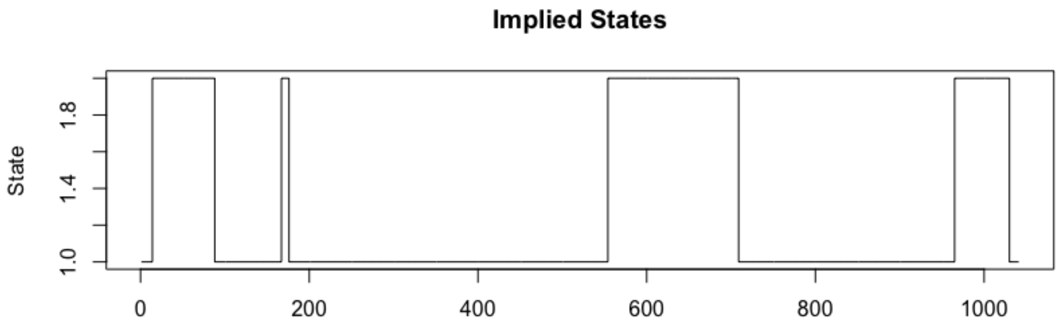


(B) T-distribution

FIGURE 8.57: Estimated state sequence from the Viterbi algorithm of TSX

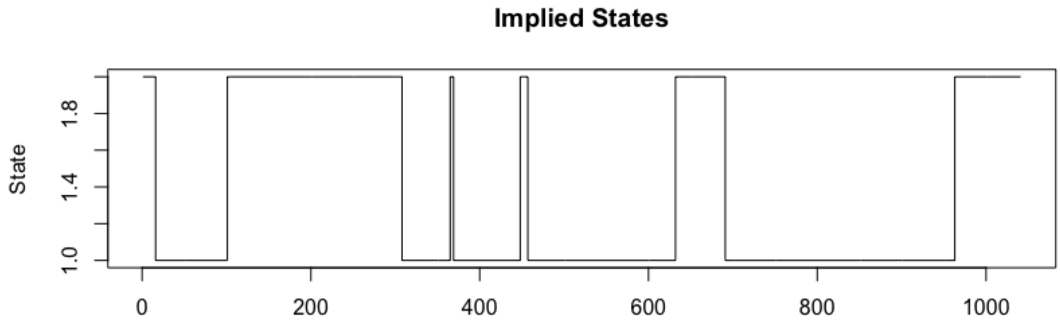


(A) Normal distribution

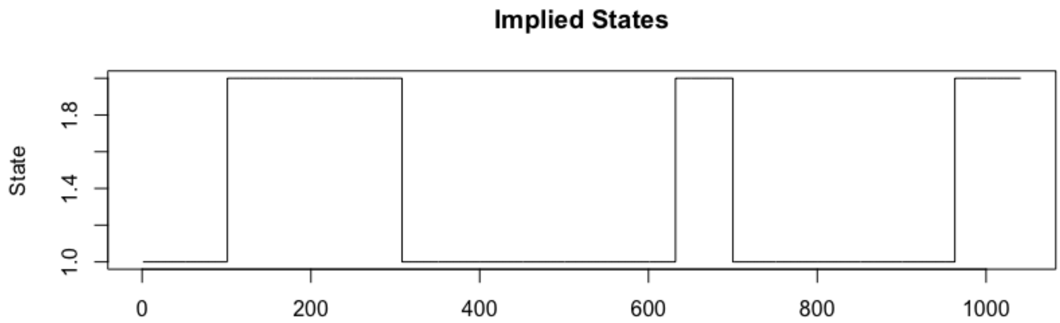


(B) T-distribution

FIGURE 8.58: Estimated state sequence from the Viterbi algorithm of SH

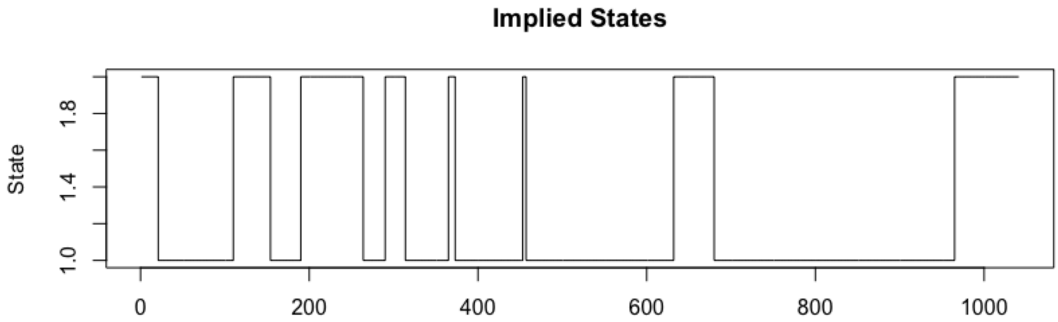


(A) Normal distribution

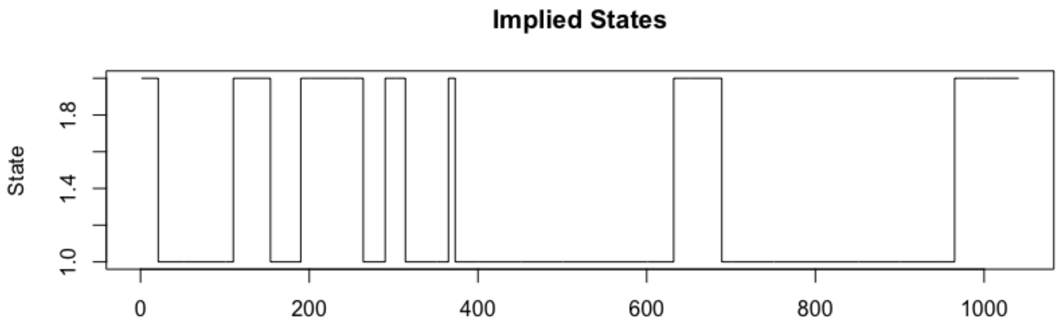


(B) T-distribution

FIGURE 8.59: Estimated state sequence from the Viterbi algorithm of Brent



(A) Normal distribution



(B) T-distribution

FIGURE 8.60: Estimated state sequence from the Viterbi algorithm of WTI

Bibliography

Sivan Aldor-Noiman, Lawrence D Brown, Andreas Buja, Wolfgang Rolke, and Robert A Stine. The power to see: A new graphical test of normality. *The American Statistician*, 67(4):249–260, 2013.

Chaker Aloui and Rania Jammazi. The effects of crude oil shocks on stock market shifts behaviour: A regime switching approach. *Energy Economics*, 31(5):789–799, 2009.

Andrew Ang and Geert Bekaert. International asset allocation with time-varying correlations. Technical report, National Bureau of Economic Research, 1999.

Luc Bauwens, Arie Preminger, and Jeroen VK Rombouts. Regime switching garch models. *Available at SSRN 914144*, 2006.

Brian H Boyer, Michael S Gibson, Mico Loretan, et al. *Pitfalls in tests for changes in correlations*, volume 597. Board of Governors of the Federal Reserve System, 1997.

Kenneth P Burnham and David R Anderson. Multimodel inference understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304, 2004.

Francois Chesnay and Eric Jondeau. Does correlation between stock returns really increase during turbulent periods? *Economic Notes*, 30(1):53–80, 2001.

Eugene F Fama. The behavior of stock-market prices. *The journal of Business*, 38(1): 34–105, 1965.

George Filis, Stavros Degiannakis, and Christos Floros. Dynamic correlation between stock market and oil prices: The case of oil-importing and oil-exporting countries. *International Review of Financial Analysis*, 20(3):152–164, 2011.

Stephen F Gray. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62, 1996.

Markus Haas and Christian Pigorsch. Financial economics, fat-tailed distributions. In *Complex systems in finance and econometrics*, pages 308–339. Springer, 2009.

James D Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384, 1989.

James D Hamilton. This is what happened to the oil price-macroeconomy relationship. *Journal of Monetary Economics*, 38(2):215–220, 1996.

James D Hamilton and Raul Susmel. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1):307–333, 1994.

Shawkat Hammoudeh and Kyongwook Choi. Characteristics of permanent and transitory returns in oil-sensitive emerging stock markets: The case of gcc countries. *Journal of International Financial Markets, Institutions and Money*, 17(3):231–245, 2007.

Bruce E Hansen et al. Erratum: the likelihood ratio test under nonstandard conditions: testing the markov switching model of gnp. *Journal of Applied econometrics*, 11(2): 195–198, 1996.

Olan T Henry. Regime switching in the relationship between equity returns and short-term interest rates in the uk. *Journal of Banking & Finance*, 33(2):405–414, 2009.

Robert I Jennrich. An asymptotic $\chi 2$ test for the equality of two correlation matrices. *Journal of the American Statistical Association*, 65(330):904–912, 1970.

Charles M Jones and Gautam Kaul. Oil and the stock markets. *The Journal of Finance*, 51(2):463–491, 1996.

Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

Mico Loretan and William B English. Evaluating correlation breakdowns during periods of market volatility. *Board of Governors of the Federal Reserve System International Finance Working Paper*, (658), 2000.

Aktham Maghyereh, Aktham Maghyereh, and Ahmad Al-Kandari. Oil prices and stock markets in gcc countries: new evidence from nonlinear cointegration analysis. *Managerial Finance*, 33(7):449–460, 2007.

Benoit B Mandelbrot. The variation of certain speculative prices. In *Fractals and Scaling in Finance*, pages 371–418. Springer, 1997.

Mohan Nandha and Robert Faff. Does oil move equity prices? a global view. *Energy Economics*, 30(3):986–997, 2008.

Jungwook Park and Ronald A Ratti. Oil price shocks and stock markets in the us and 13 european countries. *Energy economics*, 30(5):2587–2608, 2008.

Marcelo Perlin. Ms_regress-the matlab package for markov regime switching models. *Available at SSRN 1714016*, 2015.

Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

Lawrence R Rabiner and Biing-Hwang Juang. An introduction to hidden markov models. *ASSP Magazine, IEEE*, 3(1):4–16, 1986.

Latha Ramchand and Raul Susmel. Volatility and cross correlation across major stock markets. *Journal of Empirical Finance*, 5(4):397–416, 1998.

Perry Sadorsky. Oil price shocks and stock market activity. *Energy Economics*, 21(5): 449–469, 1999.

Ingmar Visser, Maarten Speekenbrink, et al. depmixs4: An r-package for hidden markov models. *Journal of Statistical Software*, 36(7):1–21, 2010.