Market Regime Identification Using Hidden Markov Model

Yuan YUAN

Department of Computer Science, University College London, United Kingdom

Summer intern at OptiRisk Systems

Supervised by: DR. CHRISTINA ERLWEIN-SAYER, DR TILMAN SAYER and PROF. GAUTAM MITRA
OptiRisk Systems

ABSTRACT

Market conditions change over time leading to up-beat (bullish) or down-beat (bearish) market sentiments. The concept of bull and bear markets, also known as market regimes, is introduced to describe market status. Since regimes of the total market are not observable and the return can be calculated directly, the modelling paradigm of hidden Markov model is introduced to capture the tendency of financial markets which change their behavior abruptly.

In this project we analyze the FTSE 100 and the Euro Stoxx 50 data series via the well-known Hidden Markov Model (HMM). Using this model, we are able to better capture the stylized factors such as fat tails and volatility clustering compared with the Geometric Brownian motion (GBM), and find the market signal to forecast the future market conditions.

Keywords: Hidden Markov model · Parameter estimation

1 Introduction

1.1 Overview

The terms “bear” and “bull” markets are used to describe the general mood or sentiment of the stock market participants, mainly the investors, whether the market is depreciating or appreciating in value. Before financial market participants start to look for a trading opportunity, it is important to look at the current market situation and forecast the subsequent market conditions so that they can decide on the appropriate trading strategies. From that perspective, the terms of bear and bull market also reflect the total investors’ attitudes towards the market and ensuing trend.

The reason for studying the market status is that the trading strategies of financial market participants are highly connected to the market conditions. For example, the financial market participant tends to make aggressive trading strategies in the bull market, which describes a time where the overall economic signals are good and the market is in the upward market trend. However, in the bear market, which describes times such as the recent economic recession, it indicates that the market trend is not good, it is a downward market trend. The participant prefers to choose the conservative strategies. The choice the financial participant makes is based on the market signal and the expectation of future market conditions. After that, the participant also takes each assets and their market performances into consideration.

The Geometric Brownian motion is widely used as a standard method to model financial time series, especially asset returns. However, one main criticism of this method is that it assumes the parameters constant over time. Moreover, the market status is not taken into account. We utilize a hidden Markov model to introduce the market state and incorporate it in the modelling. Since the mean and variance of the return in different market states are different, the forecast of return should be better in theory if the regime or the current economic situation can be figured out compared with the model using the constant mean and variance.

The aim of this project is to construct the regime switching model, and figure out how many states should be used in the hidden Markov model in order to calibrate. Specifically, different numbers of market states should be considered, and analyzed by an information criterion like AIC, log likelihood and BIC. A second country’s index is used to set up a further regime switching model, with the information contained as the total economic situation. After that, we compare the new regime switching model with the model established before, which using the FTSE 100 market data. These two regime switching models tend to figure out the synergistic effect of different financial markets. After utilizing the stock regime switching model, it can be used for the single stock or portfolio performance forecasting.

1.2 Literature review

Hidden Markov models have been used for at least three decades in several areas such as automatic speech recognition. A review of hidden Markov theory and solutions for speech recognition is given in Rabiner (1989). An application of face recognition and segmentation is in detail in the dissertation written by Samaria and Young (1994).

Specifically, as return is an important concept of finance, many researchers focus on the return using hidden Markov models with the optimized algorithms in order to capture the stylized facts such as volatility and average return. In a pioneering work by Elliott, Aggoun and Moore (1995), the estimation as well as control of hidden Markov models is introduced, which refers to signal filtering, identification of model parameter, estimation of state, signal filtering and signal prediction. In Elliott and van der Hoek (1997), the application of hidden Markov models is used to solve the discrete time asset allocation problems with filtering and parameter estimation techniques. In Erlwein, Mamon and Davison (2009), investment strategies for asset allocation relying on hidden Markov models are examined. The study of stylized facts of daily return time series and hidden Markov model is provided by Rydén, Teräsvirta and Åsbrink (1998). In Hamilton (1989), a discrete-state Markov process is introduced to the economic analysis of non-stationary time series and business cycle. In Hassan and Nath and Kirley (2007), forecasting stock price for interrelated market using hidden Markov model approach is presented. A regime switching model of long term stock return with log-normal distributions is proposed in Hardy (2001). In order to describe the possible evolution, Messina and Toscani (2008) use autoregressive HMM for modelling univariate financial time series and for generating scenarios. The multivariate normal distribution and HMMs are used to model the multivariate financial time series in Roman et al. (2010).

In hidden Markov models, the EM (expectation-maximization) algorithm and Viterbi algorithm are widely used. For the parameter estimation part, the standard approach is to find a maximum likelihood. In Baum, Petrie and Weiss (1970), the Baum-Welch algorithm, also called EM algorithm is developed to estimate parameters in HMMs. A gentle tutorial of the EM algorithm and its application to parameter estimation is provided by Bilmes (1998). In Kushary, McLachlan and Krishnan (1998), the EM algorithm as well as its extension is described in detail. Also known as Baum-Welch algorithm, Baggenstoss (2001) provides the modification of Baum-Welch algorithm for hidden Markov models considering multiple observation spaces. In Dempster, Laird and Rubin (1997), an approach of calculating maximum likelihood from incomplete data is presented using the EM algorithm.

In addition to EM algorithm, the recognition systems relying on HMMs use the Viterbi algorithm for decoding. Named after its inventor, the Viterbi algorithm is a dynamic programming algorithm which aims to find the most likely sequence of hidden states. In other words, it is a recursive optimal solution which aims to solve the problem of estimating the state sequence. In Viterbi (1967), the full description of this algorithm is provided. In Vintsyuk (1972), the Viterbi algorithm is first applied to the field of speech and language processing. In Forney (1973), a tutorial exposition of the Viterbi algorithm, the implementation, the application and the analysis is given, as well as a review of the origin of it in the field of information and communications theory of Viterbi algorithm.

1.3 Guided tour

The structure of the rest of the paper is organized as follows. In section 2, we describe the financial data used for modelling. In section 3, the hidden Markov model is introduced as well as the algorithms such as Baum Welch algorithm and Viterbi algorithm. After that, the model selection criteria e.g. AIC, BIC, likelihood are explained and compared. In section 4, we estimate the parameters with different numbers of market states using FTSE 100 index data. After that, the best model is chosen based on the model criteria provided. Subsequently, Euro Stoxx 50 index data is used to establish another regime switching model to make the comparison with the above model which using the FTSE 100 index data. Also, we show the transition probabilities of the chosen model, using these probabilities to show how to predict the future market conditions. Then analysis of the regime switching model is based on the data of single assets accompanied by the comparison with the entire economic environment. In section 5, we summarize the findings, as well as the comparisons.

2 Data

2.1 Data description

The project selects two streams of financial time series data: one is the FTSE 100 index data, another one is the Euro Stoxx 50. The sources of these two index data are supplied by Thomson Reuters. The basic statistic analysis of data is as follows.

Table 2-1 Descriptive statistics of FTSE 100 and Euro Stoxx 50 (daily data)

Data	FTSE 100	EURO STOXX 50
Frequency	Daily	Daily
Start Date	5/31/11	5/31/11
End Date	6/1/16	6/1/16
Actual number of days (length)	1264	1282
Mean of log return	0.000027	0.000047
Max. of log return	0.037710	0.058980
Min. of log return	-0.049500	-0.063180
Variance of log return	0.000104	0.000196
Skewness of log return	-0.307979	-0.159735
Kurtosis of log return	5.202590	4.805026

As shown in table 2-1, the descriptive statistics of the daily FTSE 100 index as well as the Euro stoxx 50 index are given. Both skewness of these two data series are negative, which indicates they have long left tails. Meanwhile, the kurtosis is larger than 3, which indicates the existence of fat tails.

2.2 Data selection

To analyze the single asset regime switching, we pick up some targeted stocks in the FTSE 100 as the component of the portfolio and analyze the behavior as well as performance in order to see if there exist synchronous market movements. As it can be seen from table below, the top 5 constituents of FTSE 100 index are given, which are ranked by market capital. For example, Royal Dutch Shell accounts for the biggest weight among 100 companies in FTSE 100.

Table 2-2 General information of top 5 constituents in FTSE 100 index

Rank	Name	EPIC	Mkt cap(m)
1	Royal Dutch Shell	RDSA+RDSB	151977.2
2	HSBC Holdings	HSBA	106798.5
3	British American Tobacco	BATS	89781.3
4	GlaxoSmithKline	GSK	80507.3
5	BP	BP	79859

3 Models and methods

3.1 Modelling approach

A number of inter-related models and algorithms are used in our study. We use HMM for regime identification; HMM in turn considers and uses three canonical problems:

1. One problem is how to calculate the probability of the sequence of output when the parameters are given and this problem is solved by the Forward and Backward algorithm.

2. The second problem is called the decoding problem, which aims to find the most likely sequence of latent states. This is solved by the Viterbi algorithm and Posterior decoding.

3. The last problem is about how to find the most likely set of state transition and output probabilities when the sequence of output is provided. Solved by the Expectation-Maximization algorithm.

Finally, we evaluate how well our method performs by two well known evaluation criterion, namely, (i) Akaike information criteria (AIC) and (ii) Bayesian information criterion (BIC). These are described in section 3.5.

3.2 Hidden Markov model

The standard method for analyzing the financial time series data is the Geometric Brownian motion. However, typically models (as e.g. the famous Black-Scholes model) take the assumption that the mean and standard deviation are both constants. This simplifies the model while it leads to some deviation and differences from the real market, making the calculated results of the model inaccurate.

Compared with other traditional finance theories and modelling focusing too much on the data characteristics, the hidden Markov models care more about the real market environment and the dynamics of the market conditions. Not only because of the general mathematical tractability, the experts try to use the hidden Markov model, but also because of the simplicity and the fact that the likelihood is relatively straightforward to compute.

We choose an HMM in discrete time to model the return series. In a certain regime, the log return series of the asset prices are stationary and have volatility clustering. The number of states is part of the model selection and aims to pick up the best model to fit the data and to analyse the market states. The hidden Markov model has finite sets of states, which are associated with probability distributions. Transitions among the states are governed by a set of probabilities called transition probabilities, which can be visualized as finite states with directed edges. The mathematics of HMM is presented with three main problems, followed by the Markov chain and two algorithms, which specifically are the Expectation-Maximization Algorithm and the Viterbi Algorithm. We follow the notation by Rabiner (1989) to formulate these algorithms.

In order to define the hidden Markov model completely, following elements are needed:

•Latent states $\mathcal{Q} = \{x_i\}, i = 1,,,, N$.

•Observations series $\mathcal{O} = \{o_k\}, k = 1,,, M$.

•Transition probabilities $A = \left\{a_{ij} = P\left\{x_j \text{ at } t+1 \middle| x_i \text{ at } t\right\}\right\}$, where $P(a|b)$ is the conditional probability of a given b, $t = 1, \dots, T$ is time, and x_i in Q. Informally, a_{ij} is the probability that the next state is x_j given that the current state is x_i.

•Output (Emission) probabilities $B = \{b_{ik} = b_i(o_k) = P(o_k|x_i)\}$, where o_k in $\mathcal{O}$.

Informally, B is the probability that the output is o_k given that the current state is x_i.

•Initial state probabilities $\Pi = \{p_i = P(x_i \text{ at } t = 1)\}$.

The model is characterized by the complete set of parameters: $\Lambda = \{A, B, \Pi\}$, and the finite time interval of length T has been considered in the behaviours of hidden Markov models.

Hidden variables: $x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \cdots \rightarrow x_N$

Observed variables: $o_1, o_2, o_3, \cdots, o_M$

3.3 The Forward-Backward Algorithm

3.3.1 Forward algorithm

Firstly, the forward probability alpha is introduced.

$$\alpha_t(j) = P(x_t = j; o_1, o_2, \ldots, o_t) \tag{3.1}$$

Where $x_t = j$ means the tth Markov chain is in state j.

The value of $\alpha_t(j)$ is computed as follows:

Where $\alpha_{t-1}(i)$ is the previous forward path probabilities from the previous time step, α_{ij} is the transition probability from the previous state x_i to the state j, that is x_j and $b_j(o_t)$ is the state observation likelihood of the observation o_t given the current state x_j.

$$\alpha_{t+1}(j) = P(x_{t+1} = j; o_0, \ldots, o_{t+1}) = \sum_i P(x_{t+1} = j, x_t = i; o_0, \ldots, o_{t+1}) \tag{3.2}$$

$$= \sum_i P(o_{t+1}|x_{t+1} = j, x_t = i, o_0, \ldots, o_t)P(x_{t+1} = j|o_0, \ldots, o_t)P(x_t$$

$$= i; o_0, \ldots, o_t) = \sum_i a_{ij} b_j(o_{t+1}) \alpha_t(i)$$

So $\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i) \alpha_{ij} b_j(o_t)$.

The forward algorithm can be presented in three steps.

(1) Initialization step: $\alpha_1(j) = \alpha_{0j} b_j(o_1),\ 1 \le j \le N$

(2) Recursion step: $\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i) \alpha_{ij} b_j(o_t),\ 1 \le j \le N, 1 < t \le T$

(3) Termination step: $\alpha_T(x_F) = \sum_{i=1}^{N} \alpha_T(i)\, \alpha_{iF}$

3.3.2 Backward-algorithm

The backward probability is denoted as:

$$\beta_t(j) = P(o_{t+1}, \dots, o_T | x_t = j) \tag{3.3}$$

The steps of the backward algorithm are similar with the forward algorithm. There are three steps in general, and we present these steps as follows:

(1) Initialization step: $\beta_T(j) = \alpha_{jF}, 1 \leq j \leq N$

(2) Recursion step: $\beta_t(j) = \sum_{i=1}^{N} \alpha_{ji} b_i(o_{t+1}) \beta_{t+1}(j), 1 \leq j \leq N, 1 \leq t < T$

(3) Termination: $\alpha_T(x_F) = \beta_1(x_0) = \sum_{i=1}^{N} \alpha_0 b_i\,(o_1) \beta_1(i)$

3.4 Viterbi algorithm

The solution lies in finding the sequence of the market states represented by the market state vector that maximizes the posterior probability. For instance, the probability of observing the vector containing the sequence of market states over chosen days given that we have observed the sequence of market conditions done.

Mathematically, we write this:

$$\hat{x} = argmaxP(x|o) \tag{3.4}$$

$P(x|o)$ is denoted as the posterior probability.

Generally, solving the above equation involves searching all possible market states sequence while it becomes impossible for all possible sequences if the vector length is long. From Bayes' theorem that the posterior possibility is expressed in terms of the observation likelihood the observation likelihood $P(o|x)$, the prior distribution of states $P(x)$ and the marginal distribution of observation $P(o)$ as

$$P(x|o) = \frac{P(o|x)P(x)}{P(o)} \tag{3.5}$$

where $P(o|x) = P(o_N, o_{N-1}, \dots, r_2, r_1, r_0 | x_N, x_{N-1}, \dots, x_2, x_1, x_0) = P(o_N|x_N) \dots P(o_2|x_2) P(o_1|x_1) P(o_0|x_0) = \prod_{n=0}^{N} P(o_n| \ x \ _n)$ as we have assumed that each observation o is conditionally dependent only on the present model state x. Similarly, the probability of observing the vector of state sequence is given by considering the Markov property, that is:

$$P(x) = P(x_N, \ldots x_2, x_1, x_0) \tag{3.6}$$
$$= P(x_N | x_{N-1}, \ldots x_2, x_1, x_0) P(x_{N-1} | x_{N-2}, \ldots x_2, x_1, x_0) \ldots P(x_1 | x_0) P(x_0)$$
$$= P(x_N | x_{N-1}) P(x_{N-1} | x_{N-2}) \ldots P(x_1 | x_0) P(x_0) = \prod_{n=1}^{N} P(x_n | x_{n-1}) P(x_0)$$

In fact, $P(x|o)$ is difficult to compute, we could maximize the joint probability $P(x, R)$ rather than the posterior probability.

$$\hat{x} = \arg maxP(x|o) = argmaxP(x, o) \tag{3.7}$$
$$= argmax \prod_{n=1}^{N} P(o_N | x_n) P(x_n | x_{n-1}) P(x_0)$$

From a practical point of view, it is easier to calculate the above function while it is the log probability rather than the probability.

The procedure of the Viterbi algorithm is as follows:

The first step is to initialize the log likelihood sum and the state. Then we do forward iterations and try to find the most probable state transitions.

3.5 Expectation-Maximization Algorithm

There are two steps repeated until the convergence criterion has been satisfied. The first step relates to the computation of the conditional expectation of the missing data (market states), when the observations (log return) are given. The next step is to maximize the log likelihood with the functions of the missing data that appear in the log likelihood.

The first thing to do is to find the parameter set which aims to maximize the posterior likelihood, the procedure is as follows:

$$\pi_i = P(X_0 = i) \tag{3.8}$$

$$a_{ij} = P(x_t = j | x_{t-1} = i) \tag{3.9}$$

$$N\left(\mu_{X_t}, \sigma_{X_t}^2\right) = f_{X_t}(o_t) \tag{3.10}$$

where f is the probability distribution.

In order to do the following steps, the log likelihood function (the E step) is computed, θ is denoted as parameters:

$$\mathcal{L}_c(\theta) = P\{o_o, \dots, o_N, x_0, \dots, x_N; \theta\} \quad (3.11)$$
$$= P\{X_0 = x_0; \theta\} P\{o_o; \theta\} \dots P\{X_N = x_\mathrm{N}; \theta\} P\{o_N; \theta\}$$
$$= P\{x_0; \theta\} P\{o_0; \theta\} \prod_{t=1}^{N} P(x_t | x_{t-1}; \theta) P(o_t)$$

$$log\mathcal{L}_c(\theta) = log\pi_{x_0} + \sum_{t=1}^{N} log\alpha_{x_t x_{t-1}} + \sum_{t=0}^{N} (-\frac{1}{2}\log\left(2\pi\sigma_{x_t}^2\right) - \frac{(o_t - \mu_{x_t})^2}{2\sigma_{x_t}^2}) \quad (3.12)$$

As it can be seen from the last equation, the computation of likelihood function is based on the Gaussian distribution. Then, the next step is trying to maximize the log likelihood function with the expectation.

$$E\left[log\mathcal{L}_c(\theta) \middle| o_1, \dots o_N; \tilde{\theta}\right] \quad (3.13)$$
$$= \sum_i P\left(x_0 = i \middle| o_0, \dots, o_N; \tilde{\theta}\right) log\pi_i$$
$$+ \sum_{i,j} \sum_{t=1}^{N} P(x_t = j, x_{t-1} = i |\, o_0, \dots, o_N; \tilde{\theta}) log\alpha_{ij}$$
$$+ \sum_i \sum_{t=0}^{N} P(x_t = i |\, o_0, \dots, o_N; \tilde{\theta})(-\frac{1}{2}\log(2\pi\sigma_i^2)$$
$$- \frac{\left(o_t - \mu_{x_t}\right)^2}{2\sigma_{x_t}^2})$$

The maximization step is as follows:

$$max_\theta E\left[log\mathcal{L}_c(\theta) \middle| o_1, \dots o_N; \tilde{\theta}\right] \quad (3.14)$$
$$= max_\theta \sum_i P\left(x_0 = i \middle| o_0, \dots, o_N; \tilde{\theta}\right) log\pi_i$$
$$+ \sum_{i,j} \sum_{t=1}^{N} P(x_t = j, x_{t-1} = i |\, o_0, \dots, o_N; \tilde{\theta}) log\alpha_{ij}$$
$$+ \sum_i \sum_{t=0}^{N} P(x_t = i |\, o_0, \dots, o_N; \tilde{\theta})(-\frac{1}{2}\log(2\pi\sigma_i^2)$$
$$- \frac{\left(o_t - \mu_{x_t}\right)^2}{2\sigma_{x_t}^2})$$

$s.t. \sum_i \pi_i = 1, \sum_j a_{ij} = 1$

3.6 Model evaluation

The goodness-of fit tests aims to assess the performance of the model with respect to how well the model explains the data to reflect the real world situations. When given the same set of the data, the purpose is to decide which of the candidate models best approximates the reality. In this project, we mainly choose the criterion AIC and BIC to select the hidden Markov models with different number of states as these two are derived from the information-theoretic approach and are commonly used.

3.6.1 Criteria description

Akaike (1973) showed that the maximized log-likelihood is biased upward and developed the Akaike information criteria (AIC), which is a combined measure that trades off the number of parameters required to obtain the improvement of the fit. Based on information theory and Kullback-Leibler distance, AIC rewards the goodness of fit in a given set of candidate models for the financial data by the likelihood function. The criterion is calculated as

$$AIC = -2logL + 2k \tag{3.15}$$

where $logL$ is the log-likelihood of the fitted model, and k denotes the number of parameters of the model.

BIC, also called Bayesian information criterion or Schwarz information criterion, was introduced in 1978 by Schwarz. As it is also based on the information theory but set within a Bayesian context, BIC is a model selection tool as well as the competitor to the Akaike information criterion. BIC is given by the following formula:

$$BIC = -2logL + p \times logT \tag{3.16}$$

where $logL$ and p are as for AIC, and T is the number of observations.

3.6.2 Criteria comparison

Assume that the generating model is of a finite dimension and is represented in the candidate collection, a consistent criterion will asymptotically select the fitted model having the correct structure with probability 1. While for the generating model of the infinite dimension, an asymptotically efficient criterion is selected which minimize the mean square error of the prediction.

Compared with AIC and BIC, both of them feature the same goodness-of-fit term while the penalty is different in terms of the number of parameters. As for the model choice, the BIC tends to choose the simpler models compared to Akaike Information Criterion because the penalty term of BIC is more stringent than the penalty term of AIC (For $T \geq 8, plnT$ exceeds $2k$).

4 Results

4.1 Parameter estimation

For the parameter estimation part, we use the RHmm package in R, which aims to solve the latent Markov models problems. It mainly fits the 2-state HMM and 3-state HMM models with the Baum-Welch algorithm for the estimation of parameters.

The numbers of states to be considered for establishing the hidden Markov model are successively chosen as 2, 3, 4. The 2-state hidden Markov model is divided by the bull market and bear market, while 3-state hidden Markov model add another market state – neutral market into the bull and bear market. The 4-state HMM taking both the mean and the volatility into the consideration, it may give a more elaborate description of the market conditions compared with the above two models. Specifically, the 4-state HMM is dividing the market states into 4 states, which are the bull with low volatility state, bull with high volatility state, bear with high volatility state and bear with low volatility. The model selection is based on the 3 main criteria, and we will describe it after parameter estimation.

Table 4-1 Parameter estimation of 2-state HMM

Parameter estimation of 2-state HMM		
Initial state probabilities	$P_1 = 0$	$P_2 = 1$
Transition probabilities	**to** S_1	**to** S_2
From S1	**0.96970**	**0.03030**
From S2	**0.01060**	**0.98940**
State mean and Variance	**Mean**	**Variance**
St1	**-0.0009167**	**0.0003553**
St2	**0.0002496**	**0.00005221**

As it can be seen from the parameter estimation of the 2-state HMM. The initial state probabilities suggest that the initial state is in state 2, which is considered as bull market state and it assumes the state 1 is the bear state with the mean of the log return is negative. For the transition matrix of this 2-state HMM model, it shows that if we know the today's state probabilities, then we can predict the probability of staying in the state as before or moving to different state. For example, we know that today's probability on bull state is expressed as P_{bull} and probability on bear state is represented as P_{bear}. These two probabilities can be obtained from the gamma of Forward and Backward algorithm.

Use the transition matrix, if today is in bear state, then tomorrow's probability in the same state is $P_{bear} \times P_{11}$ while the probability of being in the bull state tomorrow is $P_{bear} \times P_{12}$,. If today is in the bull state, the probability of being in the bull state will be $P_{bull} \times P_{22}$ and the probability of being in the bear state will be $P_{bull} \times P_{21}$.

$$\begin{bmatrix} 0.9697 & 0.0303 \\ 0.01060 & 0.9894 \end{bmatrix}$$

Where $P_{11} = 0.9697, P_{12} = 0.0303, P_{21} = 0.01060, P_{22} = 0.9894$. P_{ij} means the probability of transferring from state i to state j.

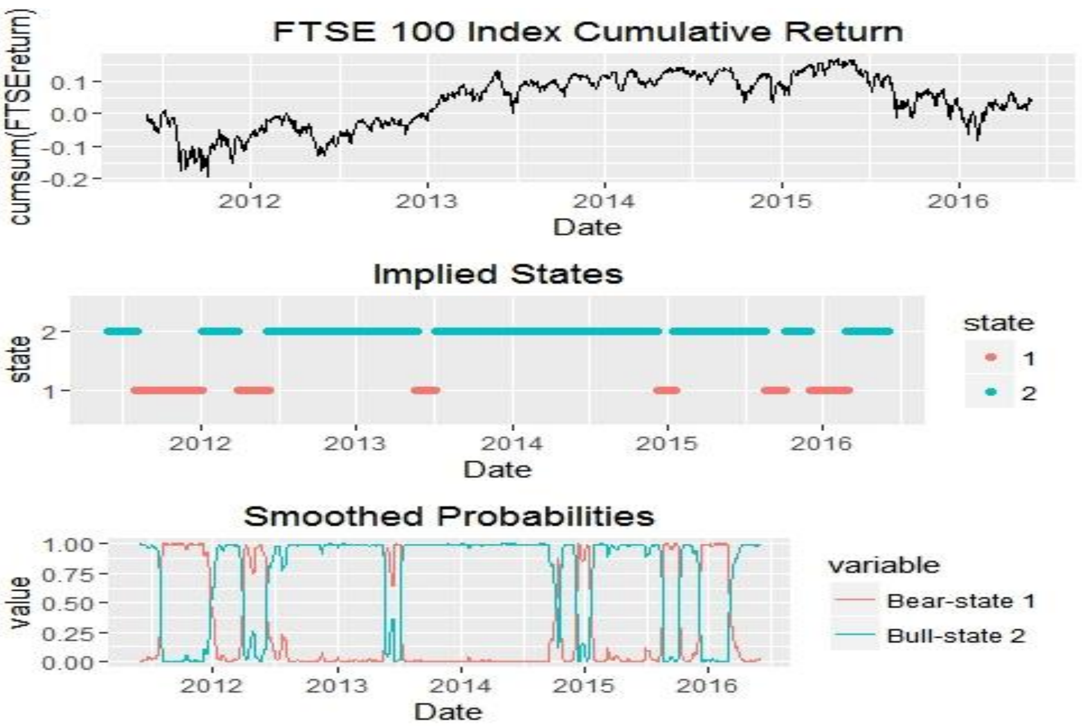

Figure 4-1 Cumulative log return, implied states and smoothed probabilities of 2-state HMM model

4.2 model selection

For a model selection part, it is usually assumed that there exists a single correct or at least best model in the set of models based on the same data and information. Basically, there are three main types of criterion for the model selection. Specifically, these are the AIC and BIC. Owing to the fact that the BIC and AIC use the information theory and maximize the log likelihood, we mainly considered the AIC and BIC rather than the log likelihood.

For all three HMMs, the same algorithm is performed to estimate the model parameters. As shown in table 4-2, the following information criteria are calculated for each model set-up.

Table 4-2 Model selection criteria of HMM models

Number of states	2	3	4
Log likelihood	4135.269	4155.077	4171.121
AIC	-8256.538	-8282.154	-8296.241
BIC	-8220.549	-8210.176	-8177.993

From the statistical data of the model selection criteria table, the AIC criteria tends to choose the 4-state univariate HMM model for the FTSE 100 index data while the BIC prefers the 2-state univariate HMM model among the three models.

In order to visualize the state distribution with the real world distribution, we plot the distribution comparison of these three different number of states univariate HMM models (Appendix). We can see that the 2-state univariate HMM fits well to the real-world data while for the 3-state univariate HMM model, the other two states fit well except the state 3. For the 4-state HMM model, only the distribution of state-2 is a little skewing of the distribution of real market data.

As for the BIC, the criterion favours the 2-state hidden Markov model as the BIC is -8220.549, which is the smallest among all three HMMs. However, the AIC prefers the 4-state HMMs among these candidate models. There exists a different selection of the candidate model in these two criteria. Consider the number of estimate parameters, the 2-state HMM is significantly less than the 4-state HMM. Because of the existence of error in each estimate parameters, we choose the 2-state HMM to describe the stock index returns.

Figure 4-2 illustrate return in different states of 2-state HMM, 3-state HMM and 4-state HMM separately.

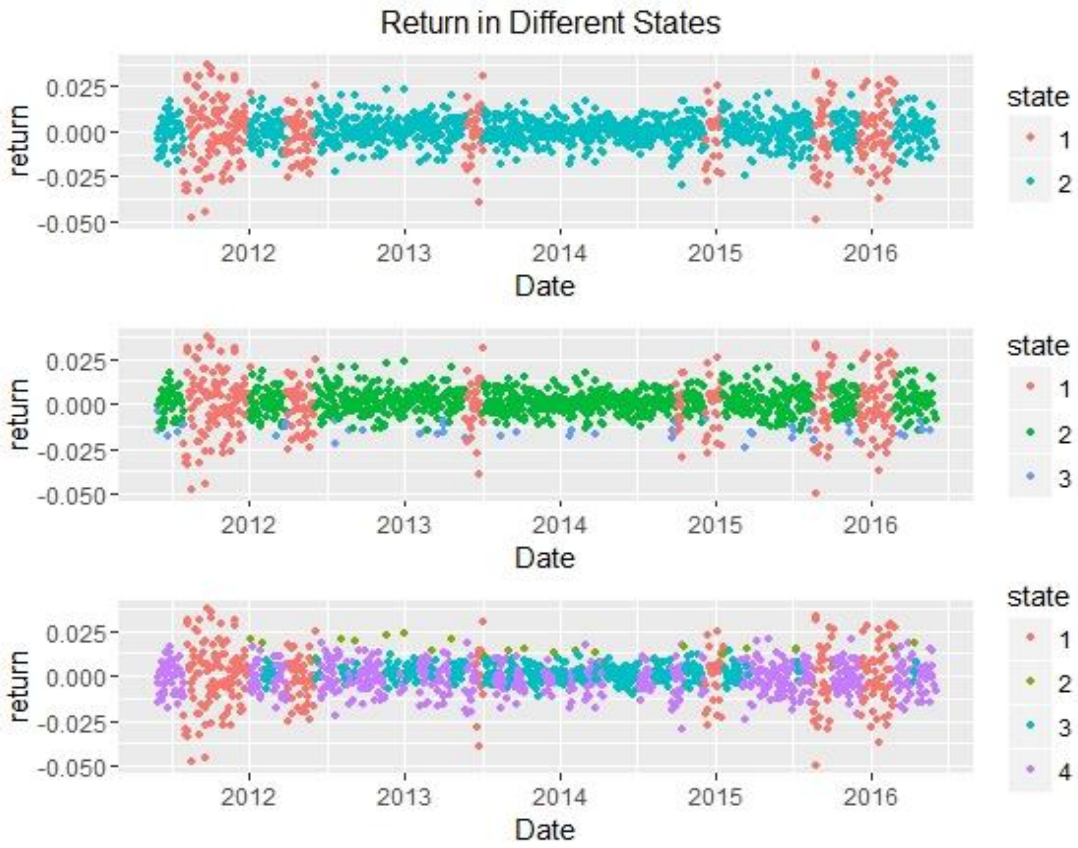

Figure 4-2 Return in Different States (HMMs)

4.3 Performance comparison

With 2-state HMM chosen, we focus on the market performance of indices in different areas. In this project, we use the FTSE 100 index and Euro Stoxx 50 index in order to make the comparison.

Table 4-3 2-state HMM models of FTSE 100 and Euro Stoxx 50

	FTSE 100	EURO STOXX 50
Initial probability p_1	0.000000	0.000000
Initial probabilityp_2	1.000000	1.000000
Transition probability P_{11}	0.969900	0.962016
Transition probability P_{12}	0.030100	0.037984
Transition probability P_{21}	0.010520	0.016172

Transition probability P_{22}	0.989500	0.983828
Mean of state 1	-0.000916	-0.001481
Variance of state 1	0.000355	0.000411
Mean of state 2	0.000250	0.000695
Variance of state 2	0.000005	0.000102
Log likelihood	4135	3752
AIC	-8221	-7454
BIC	-8257	-7490

Consider the model fit of these two data series with the same selected time period. The 2-state hidden Markov model fits FTSE 100 index better than Euro Stoxx 50 index as the criteria such as log likelihood, AIC and BIC regarding the former perform better than the latter.

After the comparison of parameter estimation, we use the Viterbi Algorithm to find the Viterbi path. Below are the implied states of these two data series calculated by the Viterbi algorithm. As it can be seen from the figure 4-9, the general tendency of these two series of implied states is the same while the time length and the time point of regime switching differ.

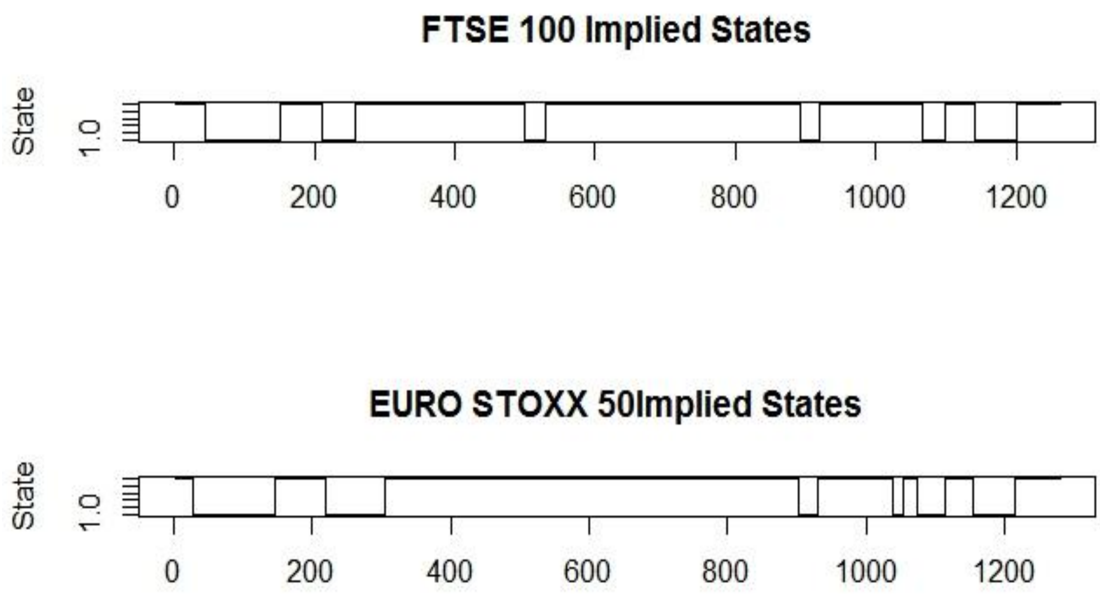

Figure 4-3 Comparison between FTSE 100 index and Euro Stoxx 50

4.4 Single asset and Index results

In this project, we pick up some components of the FTSE 100 index to figure out if the regime switching synchronize. Then the analysis focuses on the link of the estimated states and the correlation both on the whole time period considered and the rolling window approach for a time period of one year.

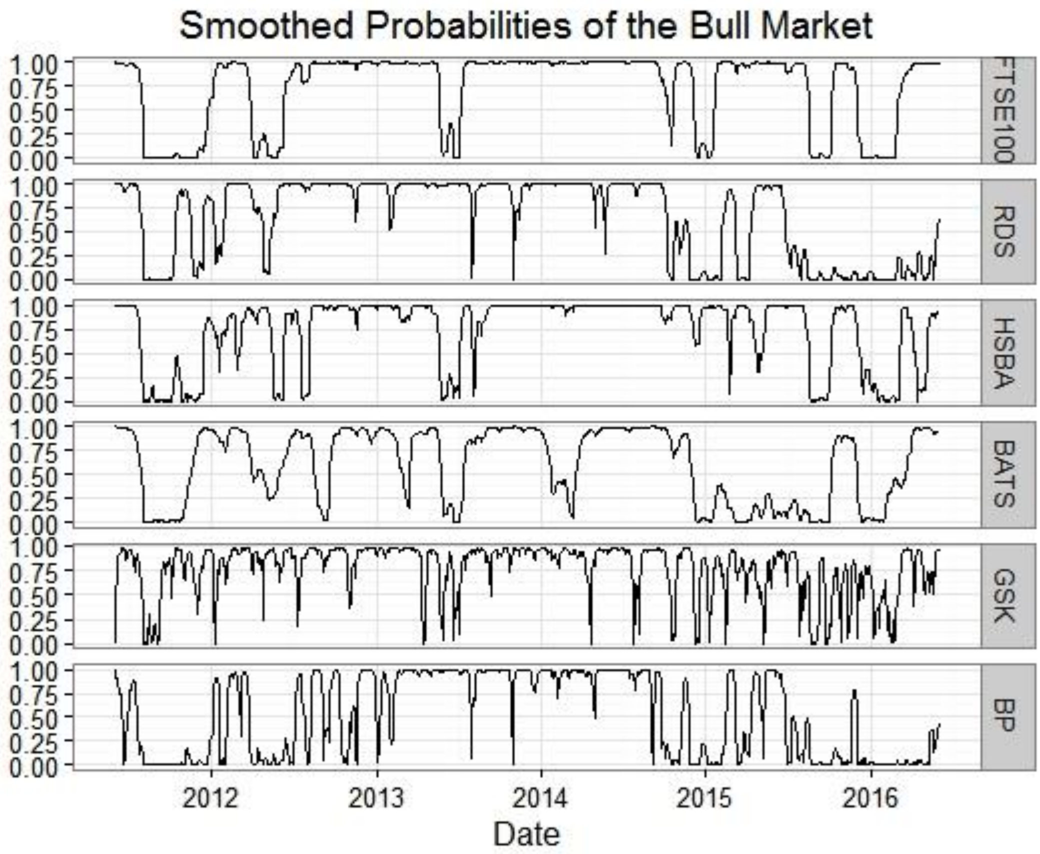

Figure 4-4 Smoothed probabilities of the bull market

Shown in figure 4-4, most components of the FTSE 100 index emerge the synchronous trend of the FTSE 100 index except GlaxoSmithKline. The difference in market states between the index and GlaxoSmithKline stock implies that the correlation between them may be small as correlation is the term referring the link between them.

In statistics, the Pearson product-moment correlation coefficient is introduced to describe the linear dependence calculated by the covariance renormalized by the standard deviation. If the Pearson correlation is zero, it suggests that these two events are independent and therefore are uncorrelated; if the Pearson correlation is positive, it indicates that these two are positive linearly dependent; if the above correlation is negative, it means that these two are anti-correlated.

As we establish the HMMs of the FTSE 100 index and the top 5 components of index, we can figure out the mean and variance of each state in order to decide which state is the bull state and which is the bear state. Only the initial state of GlaxoSmithKline's stock data is different from the other five analyzed data. Moreover, after knowing the initial states of each time series, the smoothed probabilities of bull states in regards of these six data series can be plotted.

In order to find out the correlations between the FTSE 100 index and the component stocks, we use two approaches in terms of the time period. One approach is using the whole time period data to calculate the correlation matrix, and another one is using a rolling time period for a time period of one year and calculating the correlations in a year. The second approach is dynamic as it calculates a correlation every year.

Table 4-4 Pearson correlation for the whole selected period

Pearson correlation for the whole selected period						
	FTSE 100	RDS	HSBC	BATS	GSK	BP
FTSE 100	1.0000	0.7008	0.7882	0.6413	0.6158	0.7114
RDS	0.7008	1.0000	0.5349	0.4393	0.4595	0.7357
HSBC	0.7882	0.5349	1.0000	0.4712	0.4632	0.5503
BATS	0.6413	0.4393	0.4712	1.0000	0.5375	0.3791
GSK	0.6158	0.4595	0.4632	0.5375	1.0000	0.4256
BP	0.7114	0.7357	0.5503	0.3791	0.4256	1.0000

The second approach we mentioned is a rolling window approach with the size of the rolling window is one year. Because the selected time period is five years, there are five correlation matrix which are calculated. These correlation matrixes are available in the appendix.

Taking the rolling approach into consideration, Figure 4-16 illustrates the time-varying correlation between the FTSE 100 index and the top five components. The Pearson correlation between the FTSE 100 index and HSBC is the largest in general and the smallest is the correlation between the FTSE 100 index and GlaxoSmithKline. Between 2011 and 2012, all correlations became smaller and smaller, with British American Tobacco and GlaxoSmithKline reached the bottom.

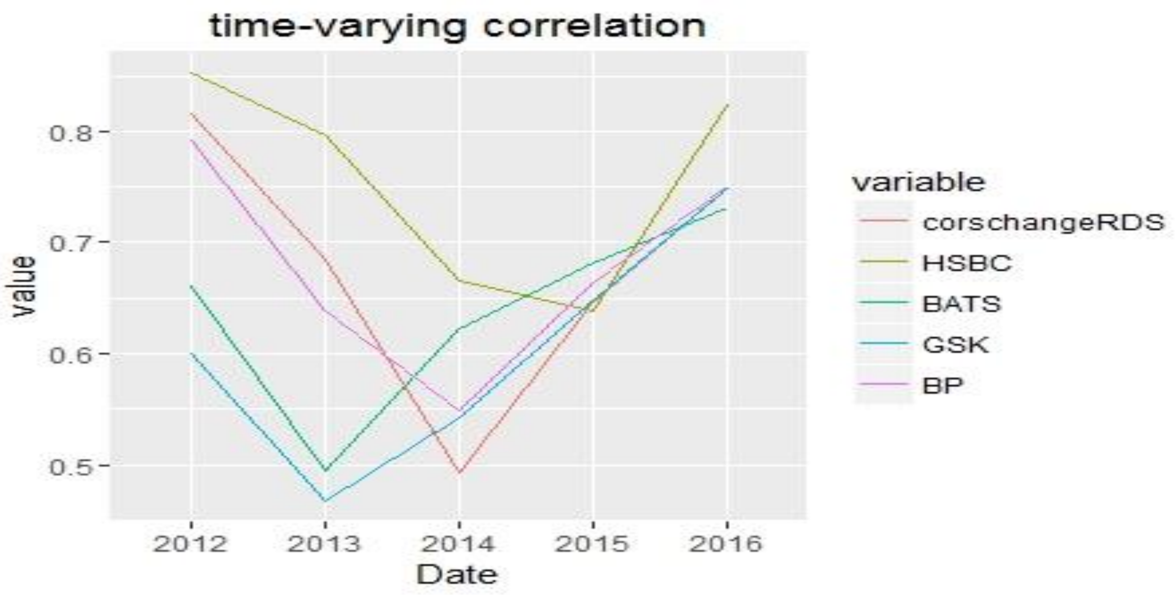

Figure 4-5 Correlation matrix based on a rolling window approach for a time period of one year

5 Discussion

5.1 Conclusions

In this report, we have reported the results of our analysis of two financial time series using the hidden Markov model, whereby the parameters of the observation process can switch in different market states, which are not observable. Given the main shortcoming of the Geometric Brownian Motion (the assumption of the constant parameters), the hidden Markov model is more general mathematical tractability in terms of the algorithms used (Viterbi algorithm and Baum-Welch algorithm) and accurate to fit the financial data compared with GBM.

The estimation parameters part is focused on the selection of different number of states hidden Markov models and the 2-state HMM is chosen as the best model among the three models because of the fitting criteria. Once the parameter estimation and transition matrix are computed, the model is used to forecast the tomorrow's market state; the details are discussed in section 4.

We compare the model fit of two markets, namely FTSE 100 index and the Euro Stoxx 50 index and find there exist the Synchronous trend of the market conditions. Besides, the data of the top 5 components of FTSE 100 index are given to analyse the market states in the same time period. The correlation is calculated to explain connection between the stocks market condition and the FTSE 100 index's market states.

5.2 Further directions

The project has put forward an efficient and flexible model (hidden Markov model), which can estimate the financial time series with different states. As the number of states is decided regarding the HMM, further works focuses on forecasting the market states and the volatility of the return using hidden Markov model.

6 References

Baggenstoss, P. (2001). A modified Baum-Welch algorithm for hidden Markov models with multiple observation spaces. IEEE Transactions on Speech and Audio Processing, 9(4), pp.411-416.

Baum, L., Petrie, T., Soules, G. and Weiss, N. (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. Ann. Math. Statist., 41(1), pp.164-171.

Buffington, J. and Elliott, R. (2002). American Options with Regime Switching. Int. J. Theor. Appl. Finan., 05(05), pp.497-514.

Dempster. N. Laird and D. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, J. Royal stat.soc, vol.39, No.1. (1977), pp. 1-38.

Elliott, R. and van der Hoek, J. (1997). An application of hidden Markov models to asset allocation problems. Finance and Stochastics, 1(3), pp.229-238.

Elliott, R., Aggoun, L. and Moore, J. (1995). Hidden Markov models. New York: Springer-Verlag.

Erlwein, C., Mamon, R. and Davison, M. (2009). An examination of HMM-based investment strategies for asset allocation. Appl. Stochastic Models Bus. Ind., 27(3), pp.204-221.

Forney, G. (1973). The viterbi algorithm. Proceedings of the IEEE, 61(3), pp.268-278.

Hamilton, J. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. Econometrica, 57(2), p.357.

Hardy, M. (2001). A Regime-Switching Model of Long-Term Stock Returns. North American Actuarial Journal, 5(2), pp.41-53.

Hassan, M., Nath, B. and Kirley, M. (2007). A fusion model of HMM, ANN and GA for stock market forecasting. Expert Systems with Applications, 33(1), pp.171-180.

Jeff. A. Bilmes, A Gentle Tutorial of the EM Algorithm and Its Application to Parameter Estimation for Gauss, Technical Report, U.C. Berkeley, TR-97-021, 1998.

Kushary, D., McLachlan, G. and Krishnan, T. (1998). The EM Algorithm and Extensions. Technometrics, 40(3), p.260.

Messina, E. and Toscani, D. (2007). Hidden Markov models for scenario generation. IMA Journal of Management Mathematics, 19(4), pp.379-401.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), pp.257-286.

Roman, D., Mitra, G. and Spagnolo, N. (2010). Hidden Markov models for financial optimization problems. IMA Journal of Management Mathematics, 21(2), pp.111-129.

Rydén, T., Teräsvirta, T. and Åsbrink, S. (1998). Stylized facts of daily return series and the hidden Markov model. J. Appl. Econ., 13(3), pp.217-244.

Samaria, F. and Young, S. (1994). HMM-based architecture for face identification. Image and Vision Computing, 12(8), pp.537-543.

Vintsyuk, T. (1972). Speech discrimination by dynamic programming. CybernSyst Anal, 4(1), pp.52-57.

Viterbi, A. (1967). Orthogonal tree codes for communication in the presence of white gaussian noise. Pasadena, Calif.: Jet Propulsion Laboratory, California Institute of Technology.

7 Appendix

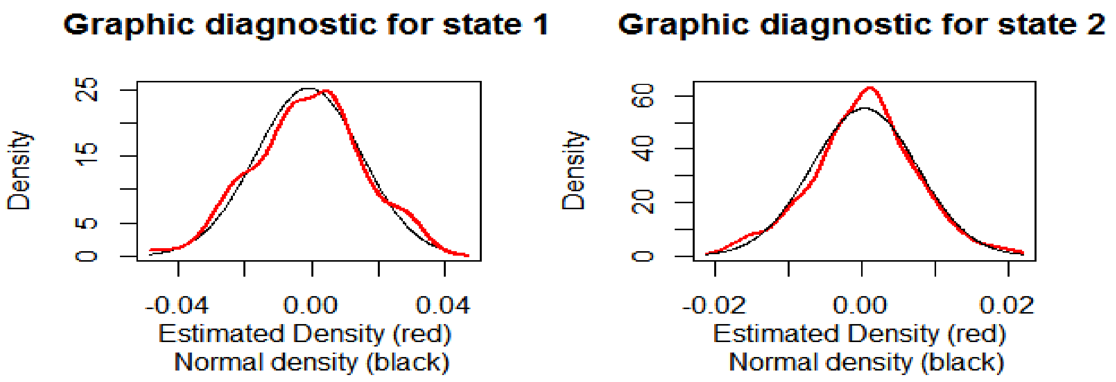

Figure 7-1 Density distribution of 2-state HMM

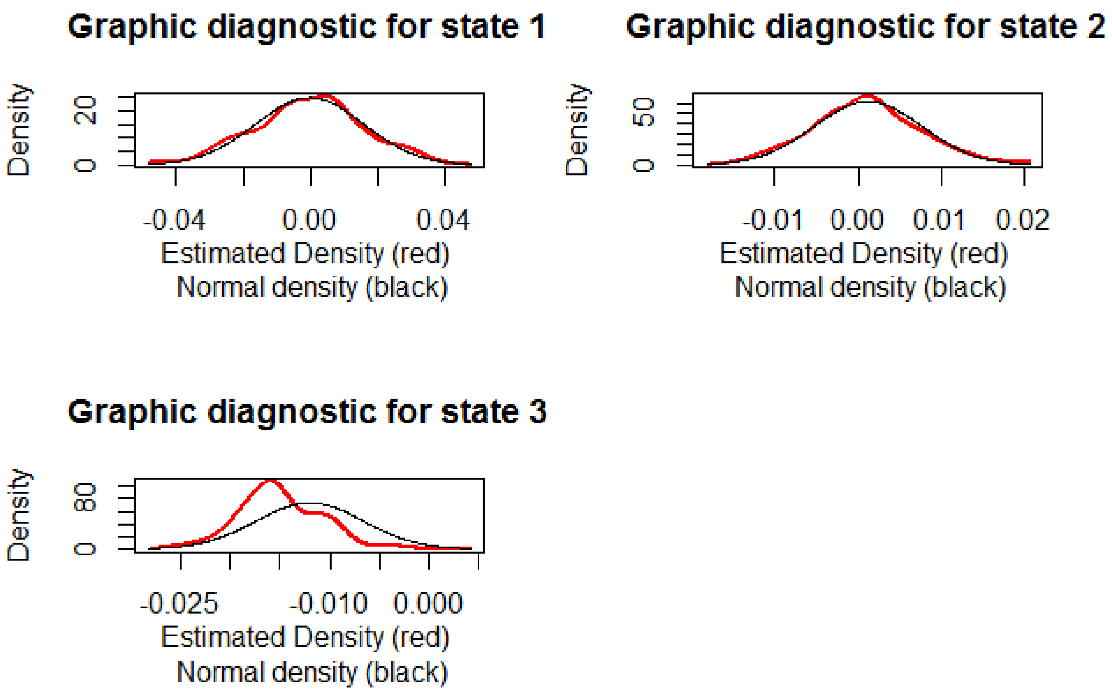

Figure 7-2 Density distribution of 3-state HMM

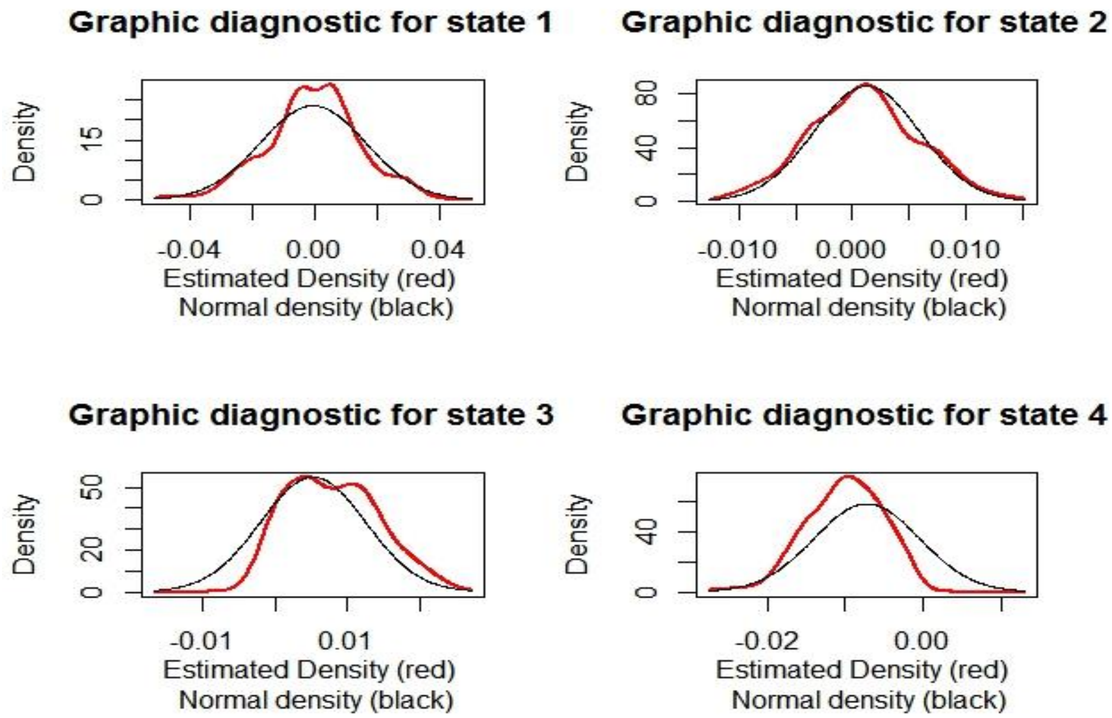

Figure 7-3 Density distribution of 4-state HMM