Stock Market Prediction Using Probabilistic Models

By

Irtaza Siddiqui

205874 NS

Supervisor

Col (R) Imran Touqir, PhD

Department of Electrical Engineering

Military College of Signals(MCS)

National University of Sciences and Technology (NUST)

Islamabad, Pakistan

May 2020

Stock Market Prediction Using Probabilistic Models

By

Irtaza Siddiqui

205874 NS

Supervisor

Col (R) Imran Touqir, PhD

A thesis submitted in conformity with the requirements for the degree of *Master of Science* in Electrical (Telecommunication) Engineering

Department of Electrical Engineering

Military College Of Signals(MCS)

National University of Sciences and Technology (NUST)

Islamabad, Pakistan

May 2020

Declaration

I, *Irtaza Siddiqui* declare that this thesis titled "Stock Prediction Using Probabilistic Models" and the work presented in it are my own and has been generated by me as a result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a Master of Science degree at NUST
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at NUST or any other institution, this has been clearly stated
3. Where I have consulted the published work of others, this is always clearly attributed
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work
5. I have acknowledged all main sources of help
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself

Irtaza Siddiqui,

205874 NS

This thesis is dedicated to *my beloved parents*

Abstract

Stochastic processes are frequently used in the assessment and follow-up of activities depending on restricted and unstable data. Various kinds of activities, including prediction, speech and gestures recognition and involvement even in social interaction can be simulated using stochastical processes. The reasons for using Hidden Markov Model to predict stocks are HMM have a solid statistical background and can easily handle highly volatile and time-invariant data. The proposed mechanism is supposed to be a Markov process with numerous hidden states. This study refers HMM to time series data for economic purposes to investigate the parameters which the model can predict.

Most of the machine learning techniques typically suffer two types of problems. First, these techniques are too complex to handle and secondly, these techniques could not adjust their initial parameters to get the desired results. To overcome these problems, HMM is used in which model train its data through the Baum-Welch Algorithm and Viterbi Algorithm adjust the initial parameters of the model to obtain the desired output. HMM has been using four visible states (Open, High, Low and Close) to predict the next day's stock prices. The predicted stocks prices of the next day are used as the present-day parameters for the second forecast in the range of predictions.

Stocks of different companies are predicted and the Mean Absolute Percentage Error is calculated by using actual and predicted prices. These results are compared with Artificial Neural Networks (ANN), Auto-Regressive Integrated Moving Average (ARIMA) and Simple Vector Machine (SVM) to evaluate the performance of the model. This analysis provides stock broker to decide based on the probability and transition pattern of each hidden state that cannot be identified from business information. In other words, HMM is the future model for stock trading as it reflects the stock price trends well.

Acknowledgments

Acknowledgments: First of all, I would like to express my gratitude to my supervisor, Dr. Imran Tauqir. He has been an exceptional mentor who gave detailed and timely responses to my queries. Throughout my dissertation he convincingly motivated me. Without his ultimate support and relentless help, this dissertation would have been next to impossible. I really appreciate the emotional support from my parents and siblings throughout the year. I can't forget to thank my friend Fahd Maqsood, who has been very kind in supporting and encouraging me throughout my time.

Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations

HMM	Hidden Markov Model
AIC	Akaike's Information Criteria
BIC	Bayesian Information Criteria
ANN	Artificial Neural Networks
SVM	Support Vector Machines
AI	Artificial Intelligence
ML	Maximum Likelihood
ARIMA	Auto-Regression Integrated Moving Average
MAPE	Mean Absolute Percentage Error
EMA	Exponential Moving Average

Symbols

$\boldsymbol{X_t}$	State Of the System
$\boldsymbol{x_i}$	Previous State
$\boldsymbol{x_j}$	Next State

$\boldsymbol{a_{ij}}$	Transition Probability
$\boldsymbol{\pi}$	Initial Probability
$\boldsymbol{v_k}$	Visible states
$\boldsymbol{b_j(k)}$	Observation Output Probability
$\boldsymbol{S_N}$	Total Number Of States
$\mathbf{M}$	Number Of Observable Symbols Per State
$\boldsymbol{q_1}$	First State
$\boldsymbol{\lambda}$	Hidden Markov Models
$\boldsymbol{W^T}$	Hidden state Sequence
$\boldsymbol{V^T}$	Sequence of Visible Symbols
$\boldsymbol{O_1{}^0}$	Present Day Vector Of Four Parameters
$\boldsymbol{O_1{}^1}$	Predicted Day Sequence
$\boldsymbol{O_2{}^0}$	Chosen Day Log-likelihood Vector
$\boldsymbol{O_1{}^{-1}}$	Day Before the Present Day
$\boldsymbol{O_2{}^1}$	Day After The Chosen Day
$\boldsymbol{Y}$	Actual Values
$\boldsymbol{Y'}$	Predicted Values
$\boldsymbol{K}$	Previous Observations Of Data Except Present Day Data
$\boldsymbol{T}$	Total Number Of Observations
$\boldsymbol{\gamma}$	Change between chosen day and the day after chosen day sequence

Chapter 1

Introduction

In simple words stock market is a place for collective markets where people can do business. Buying, selling, and issuance of shares publicly under defined set of rules and regulations falls into the category of stock market. Stock market is governed by rules defined by the government itself and the companies are bound to follow these rules to maintain the steady market. There can be variety of markets in the country where people will do the trading and make the money by playing the right cards at the right moment. Economic stability of any country is determined by the stability of stock market. Stable stock market means the country is flourishing like America or China. On the other hand fluctuating stock market means the economy of any country is in its worst condition and currency of that country would be devalued in comparison to other countries. Stock market and stock exchanges are often considered as one but stock exchange is the stock market that does the trading in shares[1]. The stock market brings hundreds and thousands of buyers to come to single market place and do the business. This means that the place need to be completely secure and prices are usually fair and practical as people gather from different places means the business will bloom so there is no need to increase the price or disturb the market. All the transactions are done in secure and controlled environment which means that operational risks are next to zero. All the transactions are done in the form of shares and rules are enforced by the regulators. Initial public offering or the IPO allows the common people to get their hands for the first time on the shares of any company that is willing to sell the shares to the open market. The main reason that was noted of selling the shares like this is to raise the necessary capital for the further business investment. To sell the shares publicly any company would need a common marketplace and at this point stock market takes over and provides the marketplace for the company to sell its share. By buying the shares the common people becomes the silent partners of the company and if the company manages to do the profitable business then the share price of that company will increase and this means the common people who are holding the shares can now sell them at higher price and make the profit of their own.[2] To get in touch with the business of the company or how the company is performing, stock exchanges keep the close eye and track each and every move of the company. This helpful information is the information upon which the shareholders act and do the business by accessing the situation of the company and the stock market. Stock market is a fair business place and people could make millions. It all boils down to the critical decision making strategies and reading the situation and state of the company before going all guns blazing and buying the stocks.

As of now you would have understood what stock market or stock exchange means. Jumping a little further, I will now discuss market timing. Market timing is basically a financial strategy that allows the person to move in and out of the stock market based on the prediction made on stock behavior. These tools help the person to time the market and increase the chances of making the money without having to fear as most of the time you will know how the market is going to move. There is a lot of debate that it is impossible to time the market as and many investors believe that all the predictive tools are nothing but bullshit. However, there are many investors who are making a lot of money proving that the market can be timed and it is not just a myth. Depending on the facts, it is not impossible to time the market. Before market timing, people were using buy and hold strategy. So, we can say that the behavior or trend of the market can be analyzed and recorded and then used to the maximum advantage. Buy and hold strategy was a long term investment. However, many investors have made success through the market timing strategy and they are choosing it over the buy and hold strategy every day. The stock market is a vast platform for economic transactions. ‘Market timing’ is the act of investing in or taking out your investment at the right time. It can be based upon forecasting methods like technical signs or economic data set. Technical indicators may include mathematical calculations based upon current prices, stock volume, etc., by analyzing previous data analysts may predict the future of the market.

Knowledge of stock market beforehand could be very productive, even a fraction of a second can result in enormous gains. Hence researchers in academia and industry took a special interest in forecasting techniques. Artificial intelligence(AI) and machine learning(ML) is extensively used to forecast the stock market[3].

1.1 Thesis Contributions

In this thesis, Hidden Markov model will be used to forecast stock value. In recent years, Hidden Markov models have been applied to this problem, and they have produced promising results. Hidden Markov models have been widely used in many applications like speech recognition, ECG analysis, image processing, etc. I shall be using five different stocks to evaluate this approach- Ferozsons Pharmaceuticals,Sui Northen Gas Pakistan, Pakistan Telecommunication Company Limited, DG Cement, and Fauji Fertilizers. A separate model is trained for each stock. Large chunks of data have been collected so that we achieve suitable variability in the data. The accuracy of the system also depends on the size of dataset. If the dataset is bigger, training of HMM model will be rigorous and hence accuracy of the model will be enhanced. The input of this thesis is to suggest a fresh definition of HMMs that combines stocks data from various sources. From stocks databases such as yahoofinance, pkfinance and eoddata, this mutual information can be utilized. Chapter 3 explains the data about these databases. In addition, our suggested model requires into consideration the stock market's volatile nature. This volatile nature can be depicted as the variability in the conduct and time-related responses of the market depending on certain evolving circumstances; i.e. the conduct of the stock market in comparable circumstances shifts over moment. In many realistic applications other than the stock market, this volatile nature remains, so this job can be helpful for reuse in other areas. Our proposed model uses time-varying transition probabilities to achieve this goal; these dynamic transitions reflect the fact that the movements of stock prices differ from the time spent on the market.

1.2 Stock Market Prediction

A stock market is a location of trading of a company's shares or stocks. It can be divided into two parts, that are, main market and secondary market. Primary market is where, through initial public offerings, fresh problems are brought to the industry. Secondary market is where shareholders are already in possession of securities. Stock market data from time series are extremely fluctuating and non-linear. A time series is a collection of information that is evaluated over time to obtain some activity's position. For stock market forecasting, linear models such as AR, ARMA, ARIMA were used[4]. The only issue with these designs is that they only operate for a specific time series data, i.e. the model defined for a specific business will not function well for another. Because of the equivocal and unpredictable nature of the inventory market, inventory market forecasting increases the danger relative to other industries. It is one of the main reasons for the stock market prediction trouble.

Stock market prediction is a difficult task; due to nonlinear time dependent nature of the market. Stock market prediction depends on multitude of factors. Factors like economic crisis, government policies, political instability etc. These factors cannot be measured or estimated, which makes stock prediction very difficult task. This makes it a field of interest for researchers to come up with a system robust and intelligent enough to cope with the non-linearity of stock market and produce positive results.

There are many applications for machine learning, one of which is to predict time series. One of the most exciting or possibly the most lucrative time sequence to forecast is inventory prices. We have carried out our research using Hidden Markov Model (HMM) which will be discussed in next chapter.

1.3 Research Probabilistic Model Used For Prediction Analysis

Stock market prediction has been an active research for a long period of time. In past, numerous machine learning techniques were used for prediction. Because it can produce important earnings, stock price forecast has been focused for years. Predicting the stock market is not a easy job, primarily as a result of a inventory time series' near-random-walk conduct. The first two techniques used to predict inventory prices were fundamental and technical analyzes. The most frequently used method is Artificial Neural Networks (ANNs). In most cases, ANNs suffer from overfitting problem due to the large number of parameters to be fixed and the little knowledge of the previous user about the relevance of the inputs in the problem being analyzed. Due to unpredictable and changing nature of stock market those machine learning techniques were not highly accurate. Machine learning techniques that were used for prediction purposes involved neural networks (ANN), support vector machine (SVM), and finally fuzzy logic (FL)[5]. All these machine learning techniques were extensively used and tested. The testing and comparison showed that artificial neural networks (ANN) has been the most successful. ANN showed promising results but still did not cope well with the volatile and unpredictable nature of the stock market.

1.3.1 Artificial Neural Network

An artificial neuron network (ANN) is a computing model centered on biological neural network design and features. Information flowing through the network influences the ANN's composition as a neural network shifts-or in a way learns-dependent on that input and output. ANNs are regarded to be nonlinear statistical information modeling instruments where modeling or models are discovered for the complicated interactions between inputs and outputs.It is also referred to as a neural network is ANN [6].An ANN has several benefits, but the reality that it can effectively benefit from watching information sets is one of the most known of these. ANN is thus used as an estimation instrument for random functions. These kinds of instruments assist to predict the most cost-effective and perfect techniques to reach alternatives while defining computing or distribution functions. ANN requires information samples to arrive at alternatives rather than complete information sets, saving time and money. ANNs are regarded to be mathematical models that are relatively easy to improve current techniques for data analysis.

1.3.2 Support Vector Machine

A Support Vector Machine (SVM) is officially described by a separate hyperplane as a discriminatory classifier. In other words, considering the marked training data (controlled learning), an ideal hyperplane is produced by the algorithm that categorizes fresh instances. This hyperplane is a row separating a plane into two sections in two dimensional spaces where it lies on either hand in each category[7]. There are four parameters of SVM, which are as follows:

i. Kernel

ii. Gamma

iii. Regularization

iv. Margin

1.4 Proposed Method

The primary methods used to cope with up-trends are technical indices; and sometimes they are coupled with the methods of machine learning. There is a difference in learning the trend conduct of the stock market using HMM, according to our literature review, and according to our finest understanding there is no HMM model that sees the trend forecast issue of stocks. The present HMM model dealing with stock market forecasting is aimed at predicting the precise stock price values.

As said before prediction has been a field of interest for many years now, but HMM has not been extensively used for prediction purposes. In the past HMM has been successful used for DNA sequencing and ECG analysis. HMM has also been productive in image de-noising and other digital image processing related researches. In the recent past HMM was used successfully forecast changes in financial time series data trajectories.

1.5 Thesis Organization

The remainder of this thesis is as follows structured. In Chapter 2, previous techniques and the already implemented HMM techniques to predict stocks are discussed. Chapter 3 provides the context needed for understanding the HMMs. It also presents the various HMM applications to incorporate data from various sources as well as the use of HMMs in stocks prediction. In addition, the section also presents datasets of different companies stocks that provide limitations on understanding. Next, a fresh definition of HMMs is proposed in Chapter 4 that requires into consideration information from stocks datasets. This section also discusses new algorithms to train and operate the HMM. Chapter 5 discusses a comprehensive overview of the synthetic data and the outcomes of the experiments. Finally, Chapter 6 discusses concluding comments and potential work.

Chapter 2

Literature Review

Hidden Markov Model has been tried and tested by many researchers. The main purpose of these researchers has been to evaluate and compare among many different methods. A comparison among neural networks and its parameters was carried out to determine the stock prices of IBM Company [6]. Another research was carried out where neural networks and external factors were determined [8]. External factors include many different things including political and financial stability of the country, investment, and market growth, etc. Another researcher tried data mining technique where they tried to find hidden patterns from the data set, from that data they tried to find any influence on the investment mind set [9]. The moving average was used by the researcher where they managed to extract the important variables that will help predict future patterns [10]. In [11] the researcher used Hidden Markov Models to predict the future stick values. They analyzed the data set and managed to extract the key points which helped to correctly forecast the future value of the stock. The researcher claimed that the value obtained by the Hidden Markov Model was nearly error-free. This research pronounced that to develop a vigorous system, which will be able to sustain and maintain precision on its own. In [12] Markov Chain combined with fuzzy logic was used, the results obtained by that signified some improvements in forecasting trends over the previously used technique.

Now in [12], the author used multiple filters and filter banks to collect and combine a bunch of data and apply Hidden Markov forecasting technique to it. the author also compared multiple methods and their accuracies with the one of his own.

Market forecasting has been a field of interest for quite a long time. Investors and researchers both have an interest in this field. Different types and techniques were used and evaluated to obtain the best possible results. At the end of the day, Neural Networks (ANN), fuzzy logic and machine learning were used for forecasting purposes. Regardless of all, that market prediction is still a very fragile topic. Due to the unpredictable and volatile nature of the stock market, it varies quickly and a lot of factors are involved. Factors that can contribute to effecting the market are multiple in nature, for example, political and financial instability can directly affect the stock market either in a positive or negative way[13]- [14]. Another big factor to take into account is to forecast using previously known data. When you deal with the previously known data, you are unaware of multiple factors like economic condition and stability of the company or country. A multitude of techniques was used for forecasting. These techniques include Artificial Neural Networks [3], which showed reasonable results at the end of the day. Fuzzy logic [7] was also used for forecasting purpose but did not show very promising results due to its inability to handle unpredictable nature. Then (SVMs) support vector machines were used for this purpose and to overcome the short comings of Fuzzy logic. At the end of the day, the Artificial Neural Network was the most successful to overcome and cope with the volatile nature of the stock market. In [4], the author used feature extraction technique to replicate and reproduce the features, (ARIMA) produced good results and reasonable outcome [4].

Nature if the market is extremely unpredictable. Which makes it time-dependent and other factors also play their role as discussed above. Due to this fact, Hidden Markov Models were rarely used for this purpose. Hidden Markov Models were used extensively used in speech recognition analysis, EGC analysis and comprehension and many other medical related techniques which involves data collection [15]. In [16] author used Hidden Markov Models to successfully predict the change in the stock market to a great extent. In [5], the author successfully combined fuzzy logic with Hidden Markov Model. This resulted in the successful representation of non-stationary and volatile stock market nature.

A huge amount of research has been carried out over time to accurately predict the volatile patterns of stock. This unprecedented nature of the market has made prediction extremely inaccurate.

Statistical time series analysis has been used extensively to take into account all the volatile factors involved in the unpredictable nature of the stock market. These techniques include as discussed above; auto regression moving average [17]. A lot of machine learning techniques has been deployed as we discussed earlier, including fuzzy logic [18], neural network [19][6][20] and hybrid techniques involving artificial neural network and fuzzy logic [31,32,33]. In all the machine learning techniques discussed above artificial neural networks has been the most accurate to predict the stock market. Statistical analysis and non-stationary nature of the stock market make it suitable for artificial neural networks. Some researchers have carried out extensive research to grasp the volatile nature of the stock market. They used fuzzy logic combined with artificial neural networks. They called it a Hybrid model. One must keep in mind that to operate fuzzy logic you should have extensive knowledge on the subject. Unlike Hidden Markov Models, fuzzy logic requires expert background learning on the subject.

In [21] Feed forward neural networks were used to train the system. Feed forward neural networks has two layers, you can call them input layer and hidden/output layer. You feed input data to the system in one layer, and get output from the other layer. In first layer data is taken as input and weights are added and summed up that are received from the previous layer. Then that weighted data is sent to the output layer. The error calculated at the output layer is fed back at the input, this makes this system self-learning. This is the working of feed forward neural networks self-learning for any data set. In [22] the author emphasized on using Hidden Markov Models for modeling a system which can compensate and comprehend the dynamic and non-stationary nature of the stock market. The author also explained and experimented on usingthe Baum-Welch algorithm for that purpose. Hidden Markov Models were modeled on Baum-welch and Progressive algorithm. Baum-Welch algorithm was used to model Hidden Markov where the data set was evaluated at each set and after each increment, the mean error probability was calculated. This sequence made sure that our system has a minimalistic error at all times. This research showed promising results with 81 percent accuracy.

In [23], the researcher used Hidden Markov Models to see and figure out the past dataset values and analyze the similarities. Baum-Welch was used to train the system to see the similarities. Their system worked on four input values, which involves Opening value, High Value, Low value, and finally the Closing value. Once the statistical analysis was done, the system will take all these four values and compare these values to the previously obtained dataset. Now the system compares these four values to the closest matching from the data set. That value that was obtained from the dataset is then taken as a reference value. Now that reference value is matched and computed with the present values of today. This computation will give you four values. Those four values will be the resultant predicated values. Once you successfully compute these four predicted values, you will then compute the error in your system. For that purpose, they have used (MAPE) mean absolute probability error. That will define the percentage error in your system.

A system that we are purposing is similar to that of [9]. In the Methodology section, we will discuss in detail what similarities and changes are made to enhance the accuracy of the system. Just to give an over view of what's about to come I will summaries the working of our system. We will take four values; Opening value, High Value, Low value, and finally the Closing value. Now we compare these values with the data set. The closest match will be taken into account, let's call it reference value. That reference value and our current value is then computed. The main difference in [2] and our system is the computation process. We will be computing fractional values. This will give us more accurate matching; hence we will get less error.

In [16], a unique and novel approach was taken by the researcher. The author implemented neural networks technique called Long Short Term Memory. This system integrates all the necessary elements of prediction and marketing factors (market fluctuations like the company's financial situation and countries financial and political stability). Bayes method is implemented in such a way to achieve maximum accuracy. The author classified the system in three categories i.e. positive value, a negative value, and the neutral value. Now the researcher created a second segment where market investor index was created to comprehend the trends of the market. In the third segment, long short term memory algorithm is created and the stock market prediction was carried out with good accuracy. Then the investor index was incorporated into the third segment to improve the accuracy further.

Chapter 3

Hidden Markov Model

3.1 Markov Property

We will be discussing Hidden Markov Models in this chapter and these are some of the most common methods of reasoning about sequences and time series. Hidden Markov Model is the statistical model generating models based on input sequence sets. If this model has an entry sequence collection, it will give you fresh transition states. To understand the model of Hidden Markov, we should first know what is the method of Markov. Thus, the Markov process is a straightforward stochastic process in which the allocation of potential variables relies solely on the current state and not on how it has reached the current state. If its distribution is determined exclusively by its present state, a random sequence has the Markov property. Any random process that has this property is called the Markov Random Process and in the case of observable state sequence, known as the Markov Chain Model, whereas in the case of non-observable states, this leads to the Hidden Markov Model.

3.1.1 Markov Property:

The state of the system at time 't+1' depends only on the state of the system at previous time instant 't'.

$$P[X_{t+1} = x_{t+1} | X_1 ... X_t = x_1 ... x_t] = P[X_{t+1} = x_{t+1}]$$

Where X_t is the state of the system. It is a memoryless property, this means that future is independent of the past given present. So given this we can predict the next state without looking at the previous states. There three basic information to define a Markov Model:

i. Parameter Space

ii. State Space

iii. State Transition Probability

3.1.2 Stationary Assumption:

In general, a process is called stationary if transition probabilities are independent of time 't', namely

$$P[X_{t+1} = x_j | X_t = x_i] = a_{ij}$$

This means that if the system is in state 'i', the probability that the system will next move to state 'j' is $\boldsymbol{a_{ij}}$, no matter what is the value of 't'. This probability is independent of time.

3.2 Hidden Markov Model

A Hidden Markov Model is a Markov model expansion in which the numbers of entry are not the same as the countries. This implies that we do not understand the state in which we are. Also recognized as probabilistic finite state automata is the Hidden Markov model. We often face certain scenarios where states can not be observed immediately, so we need an expansion, that is, HMM.

Parameters of Hidden Markov Model:

In HMM, we generally define the parameters to observe the sequences. So following are the elements of HMM:

i. Number of states, $\boldsymbol{N}$. Where, we represent each state as

$$S = S_1, S_2, ..., S_N$$

and the state at time t is q_t.

ii. Number of observable symbols per state, $\boldsymbol{M}$. We represent these observable states

Input	Word	The	man	went	to	the	park
Symbol	Tag	Determiner	Noun	Verb	Preposition	Determiner	Noun

Table 3.1: Example of speech recognition in hidden markov model where words are the input symbols and speech part is the state

as:

$$V = v_1, v_2, v_3, ..., v_M$$

To define markov model following probabilities have to be specified:
iii. Transition probability, $\boldsymbol{A}$, where

$$A = a_{ij} = P[q_{t+1} = S_j | q_t = S_i]$$

This is mainly defines as a_{ij}, this expression means that given S_j we can find the probability of S_i.
iv. Initial Probability, $\boldsymbol{\pi}$, where

$$\pi_i = P[q_1 = S_i]$$

This is the starting probability and this decides which state is likely to be the starting state.
v. Observation output probability, **B**, where

$$B = b_j(k) = P[v_k \;\; at \;\; t | q_t = S_j]$$

The range of j is from $1 \leq N$ and for k, the range is $1 \leq M$. As we described above that N is the number of state and M is the observation symbol per state.

So the overall HMM can be denoted as $\boldsymbol{\lambda}$, where

$$\lambda = (A, B, \pi)$$

Example 1:

HMM is mostly used is prediction and in speech recognition. Following is the example of speech recognition in HMM.

In the above example, the words are the input symbols, and the state is the speech part.

This is a typical instance of what phrases mean and what marks mean. The phrases

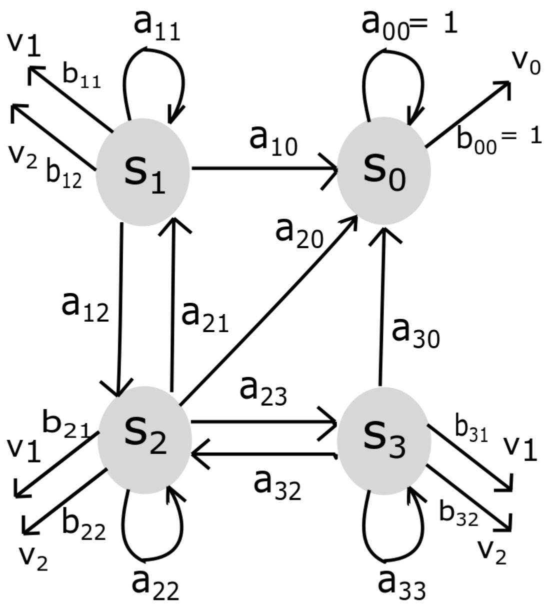

Figure 3.1: Hidden markov model used for temporal pattern recognition

here are the symbols of entry recorded and the countries are the labels.

Example 2:

You have **N** urns that contain balls of color and we have separate colours for **M**. Each

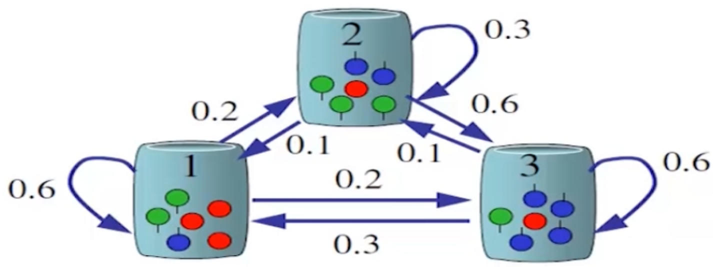

Figure 3.2: Finding State Transition Using Markov Assumption, Output-Independent Assumption And Hidden Markov Model Through Three Urn Example

urn contains different color balls numbers. Therefore, the total number of balls is the same, but we don't understand how many of them are in each urn. As in the figure above, the probability of change is obviously provided, so we need to locate the algorithm producing series. To do this, we need to do some method of generation.

Step 1: Choose original urn based on a random process.

Step 2: Choose a ball from the urn at random and then substitute it.

Step 3: Choose another urn by random selection method.

Step 4: Repeat steps 2 and 3.

Number of states = N = 3.

Number of observation symbols = M =3, $V = G, R, B$

Initial state probability = $\pi = P[q_1 = i]$ = [1,0,0]

State transition probability distribution =

$$A = a_{ij} = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.1 & 0.3 & 0.6 \\ 0.3 & 0.1 & 0.6 \end{bmatrix}$$

Observation symbol probability distribution =

$$A = b_i v(k) = \begin{bmatrix} \frac{3}{6} & \frac{2}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{3}{6} & \frac{2}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{4}{6} \end{bmatrix}$$

Here, the Markov Process is $q(t)$ and the output process is $f(x|q)$. What's concealed now is the primary issue? We can see the selected balls first, but we don't understand which urn is selected at the same moment. So the information about urn selection (state transition) is hidden. So HMM uses some assumptions which are as follows:

***i*. Markov Assumption:**

The state transition only depends on the origin and destination.

***ii*. Output-Independent Assumption:**

All observation frames are dependent on the states that generated them, not on neighbouring observation frames.

***iii*. HMM Structure:** The first Markov assumption of transition states that the current state depends only on the previous state.

$$P[q_t|q_1, q_2, L, q_{t-1}] = P[q_t|q_{t-1}]$$

So the conditional independency of observation parameters are

$$P[X_t|q_t, X_1, L, X_{t-1}, q_1, L, q_{t-1}] = P[X_t|q_t]$$

3.3 HMM Advantages and Applications

Advantages of HMM on Sequencial Data:

i. Natural Model Structure: doubly stochastic process *ii*. Transition parameters model: temporal variability

iii. Output distribution model: spatial variability

iv. Efficient and good modeling tool for real world complex process

v. Sequences with temporal constraints

vi. Spatial variability along the sequence

vii. Volatile and non-stationary datasets

viii. Efficient evaluation, decoding and training algorithms

ix. Mathematically strong *x*. Computationally efficient

Application Areas of HMM:

i. Stock market prediction

ii. Speech recognition

iii. Gesture recognition

iv. On-line handwriting prediction

v. Weather prediction

vi. Modeling of coding/non-coding regions in DNA

vii. Optical character recognition

3.4 Three Problems of HMM

In this section, we will discuss the three main issues associated with Hidden Markov Model and their solution algorithms.

Problem 1: Evaluation - Compute the probability of a given observation sequence.

Problem 2: Decoding - Given an observation sequence, compute the most likely hidden state sequence.

Problem 3: Learning - Given an observation sequence, and set of possible models, which model closely fits the data?

Before discussing these problems, first we will discuss how to build an observation sequence. Given the HMM parameters, we can build an observation sequence:

$$O = O_1 O_2 ... O_T$$

Where **T** is the number of observations - This observation sequence is build as following:

i. Set t=1.

ii. Choose an initial state $q_1 = S_i$, according to initial distribution.

iii. Choose $O_t = V_k$, according to symbol probability distribution.

iv. Transit a new state $q_{t+1} = S_j$, according to state transition probability distribution.

v. Set t=t+1 and return to step 3 while $t \leq T$.[15]

3.4.1 Evaluation Problem

So the first problem is known as the **Evaluation Problem**. In this problem given the observation sequence $O = O_1O_2...O_T$ and an HMM model $\lambda = (A, B, \pi)$, how do we compute the probability of O given the model? Solution to this problem is **Baum Welch Algorithm**.

In evaluation problem you have access to the HMM "λ" and you know the sequence of visible symbols V_T and you have to find out $P[V_T|\lambda]$?

$$\lambda = \pi, V, a_{ij}, b_{jk}$$

3.4.2 Baum-Welch Algorithm

This is also recognized as the dynamic programming algorithm forward-backward. The probability is evaluated using any series of duration T states in this technique. This is also recognized as a technique called "Any Path." So, we just need to figure out the sequence of action and the probability of the series of action whatever the route used may be, and this is called any type of route. λ, V_T is given or known and we have to find $P[V_T|\lambda]$?

We can find out the all possible sequence of W^T hidden state sequence that have generated given V_T.

$$P[V_T|\lambda] = \sum_{r=1}^{r_{max}} P[V_T|{W_r}^T]P[{W_r}^T] \tag{3.4.1}$$

where **r** in general means one of the possible hidden sequence W^T of length "T" that has generated V_T.

$${W_r}^T = W(1), W(2), W(3), ...W(T)$$

If N is the number of hidden states then $r_{max} = N^T$

$$P[{W_r}^T] = \prod_{t=1}^{T} P[W(t)|W(t-1)]$$

Where $P[{W_r}^T]$ is the probability of particular hidden state and w(t) is just the transition probability.

$$P[V_T|{W_r}^T] = \prod_{t=1}^{T} P[V(t)|W(t)]$$

Where $P[V_T|{W_r}^T]$ means the given machine is in hidden state one at the time instant "t" and it has generated first visible symbol.By putting the values in equation 1,

$$P[V_T|\lambda] = \sum_{r=1}^{r_{max}} \prod_{t=1}^{T} P[V(t)|W(t)] \cdot P[W(t)|W(t-1)] \tag{3.4.2}$$

Where r_{max} is the number of possible paths of hidden states through which transition occur while generating V_T. But the complexity of the final equation is $(\boldsymbol{N^T \cdot N})$. We can use **recursive algorithm** to reduce this complexity. In recursive algorithm, we know that in which state the machine is, lets say at $t = 0$ machine is in state W_1. At $t = 1$, we have to find out the transition probabilities, that is, what is the probabilities that machine will go from W_1 to W_0 at $t = 1$ or W_1 to W_1 at $t = 1$ or W_1 to W_i at $t = 1$?

If $V(1) = V_0$, this means that if visible symbol is at V_0 at time instant $t = 1$, then it is sure that W_0 can emit this symbol. So this assures that at $t = 1$ machine will surely make a transition from W_1 to W_0. If $V(1) \neq V_0$, then machine will not make a transition from W_1 to W_0. It will make transition to W_1 or W_0 or W_i etc.

Backward Probabilities:

Let at $t = t$ machine is in hidden state W_2, we have to find out what is the probability that machine was in state W_1 at $t = t - 1$ or in state W_2 at $t = t - 1$ or in W_3 at $t = t - 1$? We don't consider W_0 at $t = t - 1$ because if the machine enters in the final state W_0 then it cannot make a transition from the final state to any of the other state. If we somehow able to find the probability of the machine being in state W_1 at $t = t - 1$ and we know all the transition probabilities from all states to W_2 from transition table which is given in HMM model "λ". Let,
The probability of being in state W_1 at $t = t - 1 = \alpha_1(t - 1)$
The probability of being in state W_2 at $t = t - 1 = \alpha_2(t - 1)$
Then, the probability that machine will make a transition from W_1 to $W_2 = \alpha_1(t-1) \times a_{12}$ and the probability that machine will make a transition from W_2 to $W_2 = \alpha_2(t-1) \times a_{22}$

So,

$$\beta_t(i) = P[O_{t+1} \ldots O_T|q_t = S_i, \lambda]$$

Forward Probabilities:

in this algorithm we find that what is the probability that, given an HMM λ, at time t, the state is i and partial observation $O_1O_2...O_t$ has been generated?

$$\alpha_t(i) = P[O_1O_2...O_t, q_t = S_i|\lambda]$$

$$\alpha_t(j) = \left[\sum_{t=1}^{N} \alpha_{t-1}a_{ij}\right] b_j(O_t)$$

So initialization occurs at

$$\alpha_1(i) = \pi_i b_1(O_1)$$

Forward recursion at

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i)a_{ij}\right] b_j(O_{t+1})$$

Termination at

$$P[O|M] = \sum_{i=1}^{N} P[O_1, O_2, ..., O_T, q_T = S_i|M] = \sum_{i=1}^{N} \alpha_T(i)$$

As the model uses both forward and backward algorithms for training machine. So it is called as forward-backward algorithm.

Example:

For the given emission and transition probabilities, find $P[V_T|\lambda]$?

Transition probabilities =

$$A = a_{ij} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.2 & 0.3 & 0.1 & 0.4 \\ 0.2 & 0.5 & 0.2 & 0.1 \\ 0.7 & 0.1 & 0.2 & 0.1 \end{bmatrix}$$

Visible probabilities =

$$B = b_{jk} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0.4 & 0.1 & 0.2 \\ 0 & 0.1 & 0.1 & 0.7 & 0.1 \\ 0 & 0.5 & 0.2 & 0.1 & 0.2 \end{bmatrix}$$

So let say visible sequence is given, that is, $V^4 = V_1, V_3, V_2, V_0$ and we also know that initially machine is in hidden state W_1 at $t = 0$. So,$\alpha_1(0) = 1$ at $t = 0$.

By calculating from trellis diagram, the summation of the product of $\alpha_j(1) \times \alpha_i(1)$ and also this summation is multiplied by the emission of of visible symbol at time instant

$t = 2$ in the given visible symbol which is V_3 in the given V_4.
$\alpha_0(1) = 0$, this means that probability of being in state W_0 at $t = 1$ while emitting first visible symbol V_1 is zero as given by emission probability table.

$$\alpha_1(2) = [0.09 \times 0.3 + 0.01 \times 0.5 + 0.2 \times 0.1] \times 0.1$$

$$\alpha_1(2) = [0.027 + 0.005 + 0.02] \times 0.1$$

$$\alpha_1(2) = 0.0052$$

As we know that, equation of evaluation says $P[V_T|\lambda] = \alpha_0(T)$. So,

$$P[V_4|\lambda] = \alpha_0(4)$$

$$\alpha_0(4) = 0.00138$$

3.4.3 Decoding Problem

In this problem, given the observation sequence $O = O_1O_2...O_T$ and an HMM model $\lambda = (A, B, \pi)$, how do we find the most probable sequence of states, given a state of observations?

We find the most likelihood sequence of the hidden states through which the machine passes to generate the visible sequence V_T. Most probable sequence of the hidden states through which the machine has been made transition while making/generating visible sequence V_T. In each state effects of all other state probability is included, that is, at each time step, the state probability us calculated by summation formula of $\alpha_j(t)$, i.e., effect of probabilities from all possible states.
This problem can be solved by **Viterbi dynamic programming algorithm**.

3.4.4 Viterbi Dynamic Programming Algorithm

We can choose an HMM by the probability generated using the best possible sequence of states. We'll refer to this problem as "Best Path" method. In other words, we have to find the state sequence $Q = q_1....q_T$ which maximizes $P[Q|O_1, O_2, ..., O_T]$.

$$Q = \arg\max P[Q^{'}|O, \lambda]$$

If we know the identity of Q_t, then the most probable sequence on $t+1, ..., t$ doesn't depend on observations before time t. The purpose of using viterbi algorithm is its a dynamic programming technique and it align the observation and state transitions. It is used for internal segmentation, to find the most likely state sequence and for the analysis for internal processing results. Similar to computing the forward probabilities, but instead of summing over transitions from incoming states, compute the maximum.

$$\alpha_t(j) = \left[\sum_{t=1}^{N} \alpha_{t-1}(i) a_{ij}\right] b_j(O_t)$$

$$\sigma_t(j) = \left[\max_{1 \le i \le N} \sigma_{t-1}(i) a_{ij}\right] b_j(O_t)$$

$$\psi_t(j) = \left[\arg\max_{1 \le i \le N} \psi_{t-1}(i) a_{ij}\right] \quad 2 \ \le \ T, \ 1 \ \le \ j \ \le \ N$$

$$p^* = \max_{1 \le i \le N} \sigma_T(i)$$

$$q_T^* = \arg\max_{1 \le i \le N} \sigma_T(i)$$

If the best path ending in $q_t = S_j$ goes through $q_{t-1} = S_i$ then it should coincide with best path ending in $q_{t-1} = S_i$.

$$\sigma_j(t) = max_{x_1,...,x_{t-1}} P[x_1...x_{t-1}, O_1...O_{t-1}, x_t = j, O_t]$$

The state sequence which maximizes the probability of seeing observations to time $t-1$, landing in state j, and seeing the observation at time t.

3.4.5 Learning Problem

How do we adjust model parameters, $\lambda = (A, B, \pi)$, to maximize $P[O|\lambda]$? In this problem rough structure of Hidden Markov Model is given, means we are provided by number of hidden states and number of visible states but we don't know the probabilities associated with this model. In Learning problem through the set of training sequences we have to find out the transition probabilities. For this purpose we have to use a large number of training sequence to estimate transition probabilities a_{ij} and b_{jk}.
The solution of this problem is **Expectation Maximization hueristic**.

3.4.6 Expectation Maximization Hueristic

There are two types of learning:

i. Unsupervised Learning

ii. Supervised Learning

We will use distinct sets of visible symbols with recognized series of recognizable symbols in the HMM teaching phase to train the information. So it is supervised learning to train or learn HMM. We use the Expectation Maximization algorithm for the teaching phase. You can infer the values of the parameters by the EM algorithm. If someone informs us which points originated from which origin we could estimate the parameters, but no one informs us which point originated from where. Now you have the parameters, from which distribution we could find out which point originated.

EM algorithm is an iterative estimation algorithm that can derive the maximum likelihood estimates in the presence of incomplete data.

Example:

The classical convenience is the Guassian mix where we have an unspecified allocation of Guassian. It's a lot to one mapping basically. It is used to fill missing information in a sample information set, to detect the importance of latent variables and to estimate cluster parameters, finite mixtures, Guassian mixture models and HMMs.

There are two steps in the EM algorithm:

***i*. Expectation Step:**

Assume an established information comes from a particular model and use it to predict the missing data-formulate some parameters for that model, use this to imagine the missing data/values in information sets.

***ii*. Maximization Step:**

We use full information in this phase to update the missing information parameters and discover the most probable parameters.

Just like in the case of Stock prediction we find the most likelihood sequence from the HMM trained data.

$$\theta_{ML} = \arg\max_{\theta\epsilon\Omega} L(\Theta)$$

$$\theta_{ML} = \arg\max_{\theta\epsilon\Omega} \log P[X, Y|\Theta]$$

Where, X = Observed data, Y = Missing data, Θ is the parameter vector and $L(\Theta)$ is the likelihood function.

Chapter 4

Methodology

4.1 Proposed Model

The proposed model shows how stock prices can be predicted using HMM. The complete working of the proposed Hidden Markov Model is shown in Figure 4.1.

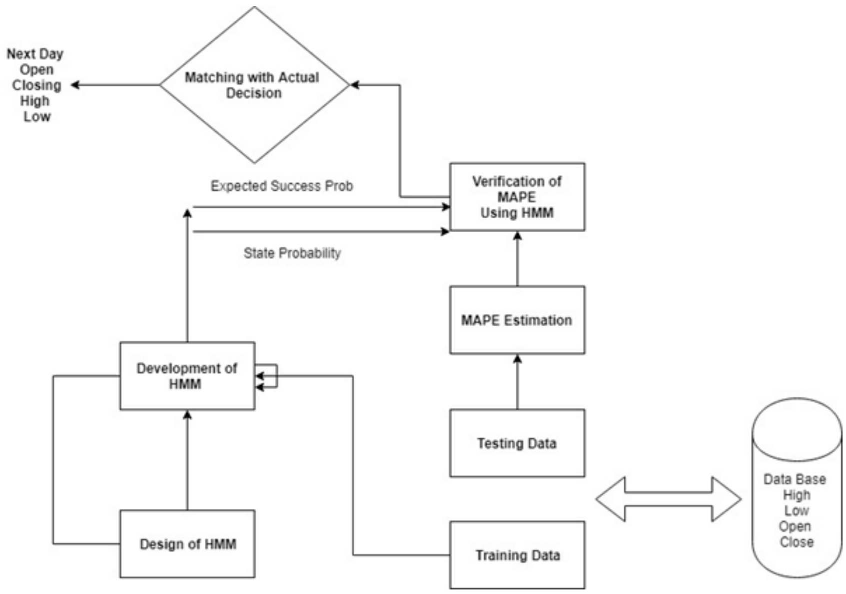

Figure 4.1: Complete flow chart from training data to predicting next day stocks parameters through Hidden Markov Model

It is possible to divide the forecast method by three measures. Step one: calibrate the parameters of HMM and calculate the model's probability. Step two: second in the previous day with a comparable probability to the last day. Finally, step three: in the past, to forecast future stock rates, use the stock price distinction on the "comparable" day. This strategy to forecast is focused on Hassan and Nath's work [11]. The proposed procedure, however, is distinct from their technique in that it bring the sign of the latest day's distinction in probability of the "comparable" day in the past into calculating future stock rates. Futhermore, they use one observation data that is, close price to predict close price of next day but I use multiple observations data to predict all the parameters (close, low, high and open) of the next day stock price. There are four visible and five hidden states in our model.

Visible states $= v_i = [v_0, v_1, v_2, v_3]$

Hidden States $= w_i = [w_0, w_2, w_3, w_4, w_5]$

where, $v_0 =$ Open price , $v_1 =$ High price

$v_2 =$ Low price , $v_3 =$ Close price

$w_0 =$ Predicted open price , $w_1 =$ Predicted high price

$w_2 =$ Predicted low price , $w_3 =$ Predicted close price

$w_5 =$ Predicted trade volume

According to the first step, find all the parameters of our model. So there are five parameters of Hidden Markov Model (HMM) as discussed in section 3. Each parameter are determined by the trained model, so there is no initialization process. Parameters are different for every output to achieve the desired result. Following are the parameters of the models:

Number of hidden states $= N_{states} = 5$

Number of symbols per state = M = 252

Transition Probability Matrix =

$$A = a_{ij} = \begin{bmatrix} 0.80366 & 0 & 0 & 0.062969 & 0.01336 \\ 0 & 0.97950 & 0 & 0 & 0.020489 \\ 0 & 0 & 0.829100 & 0.17081 & 0 \\ 0.328618 & 0 & 0.291004 & 0.26265 & 0.11722 \\ 0.2649892 & 0 & 0 & 0.10575 & 0.62925 \end{bmatrix}$$

Emission Probability Matrix =

$$B = b_j(k) = \begin{bmatrix} 0.142112 & 0.342559 & 0.2275503 & 0.287784 \\ 0.287809 & 0.316539 & 0.067863 & 0.327787 \\ 0.2078820 & 0.614720 & 0.160583 & 0.010814 \\ 0.276889 & 0.189812 & 0.453832 & 0.079465 \\ 0.508565 & 0.1599016 & 0.233789 & 0.098628 \end{bmatrix}$$

Initial Probability =

$$\pi = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

All the probabilities of the model are also represented in the form of state diagram which is shown in figure 4.2 and the hidden sequence of the model is shown in figure 4.3. The number of symbols per state (M) means the number of observations in each visible state. For example, the dataset contains visible state (open price) from November 2018 to July 2019 has precisely 252 observations which means our model execute backward algorithm to 252 observations to find the likelihood sequence.

Probability of transition from state w_1 to w_1 with the emission probability v_0 is given by:

$$w_1 \xrightarrow{v_0} w_1 = a_{ij} \times b_{j1} = 0.9795 \times 0.317 = 0.3105$$

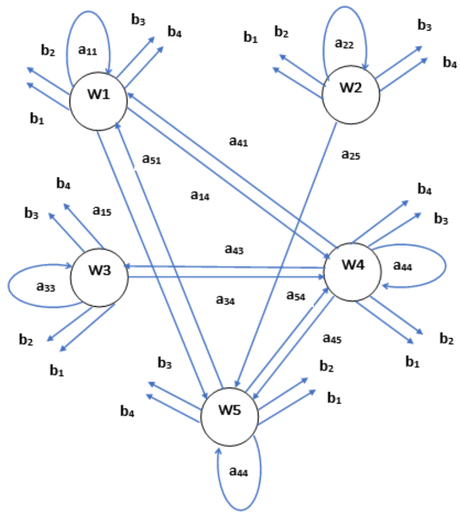

Figure 4.2: State diagram of the proposed HMM with transition probabilities a_{ij} and emission probabilities $b_j(k)$

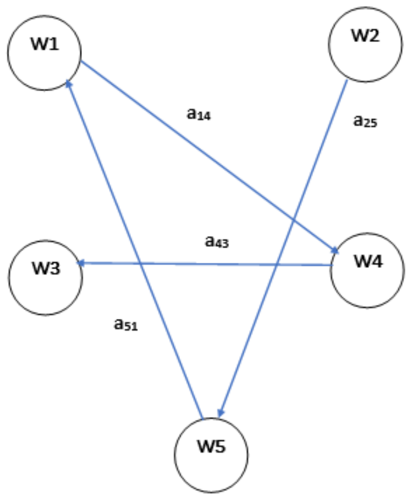

Figure 4.3: Hidden sequence of the states with initial probability of $\pi = [0\ 1\ 0\ 0\ 0]$ is $w_T = w_2 w_5 w_1 w_4 w_3$

After the first forecast of stock's value parameters for the next day, model update observation sequence ${O_1}^1$ by shifting it forward one day to estimate inventory cost for day ${O_1}^2$ in which $O_1 = {O_1}^0, {O_1}^1, {O_1}^2, ..., {O_1}^100$ where ${O_1}^2$ is the next day to the predicted day . The calibrated parameters of the HMM were used as the original parameters for the second forecast in the range of predictions. For the second forecast, I repeat the three-step forecast method and so on. We use ${O_1}^1$ as the present day in case of predicting second day values.

4.2 Model Selection

Two common criteria are used in the section: the Akaike's Information Criteria (AIC) and the Bayesian Information Criteria (BIC) to assess HMM's performance. These requirements are appropriate for HMM because the EM technique has been used to improve the model's log-likelihood in the model learning matrix, the Baum–Welch Algorithm is to maintain the model straightforward and viable to inventory forecast, we restrict the number of hidden states from four to eleven. The following equations are used to calculate these requirements:

$$AIC = -2\ln(L) + 2k$$

$$BIC = -2\ln(L) + k\ln(M)$$

Where L is the maximized value of likelihood, M= observation points and $\boldsymbol{k = N^2 + 2N - 1}$ where N is the number of states. In the above performance measures model with lowest BIC is preferred and the criteria for model selection for finite set of model. I describe the maximum value of likelihood,$\boldsymbol{L}$ as $L = P[x|\hat{\theta}, \lambda]$, where x is the observed data, λ is the model and $\hat{\theta}$ is the model parameters.

4.3 Flow Chart Of Predicting Stock Values

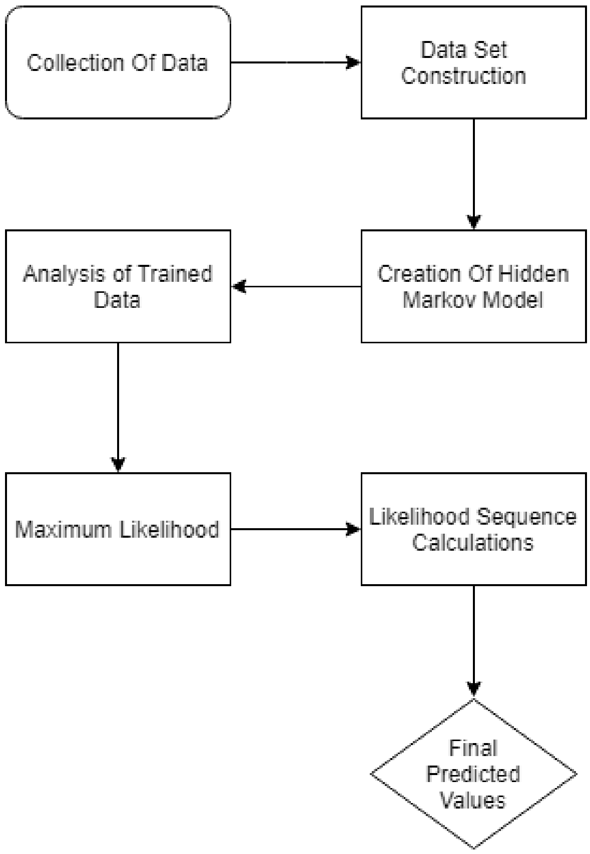

Figure 4.4: Flow chart of training data through Hidden Markov Model

In the region of artificial intelligence and finance, the Hidden Markov model has been commonly used to forecast economic systems or predict stocks prices. HMM is the new method to anticipating stocks prices and trading stocks which will be explored in this setion.The flow chart shows all the steps involved in my thesis work. Working of my thesis compromises of these following steps.

i. First step in prediction of any type of entity is that you obtain or get your hands on a good and legitimate data set. To execute our model, we select historical stocks data of different companies. From July 2018 to June 2019, most of the data was drawn from finance.yahoo.com and pkfinance.com. Once you get a legitimate data set you then need to check it for any irregularity. Irregularity can be of many forms, including incomplete data set or flaws in dataset.

ii. Then once you get your hands on some authentic data set you need to prepare your system to cope with large datasets at once. You might have trained your system on smaller data sets while testing, but once the whole dataset is compiled at once, there is good chance that your system might crash.

iii. To prevent your system from crashing you will have to do extensive training. This makes your system immune to extreme testing.

iv. First thing that I did was, to start test with smaller data sets and check the accuracy of our system. Then two errors, that are, mean absolute probability error (MAPE) and Root Mean Square Error (RMSE) are then calculated and checked for model prediction accuracy. After calculating the accuracy of the model then it will be compared with different previously used AI techniques for comparison.

v. Then you need Hidden Markov Model to start prediction process.

vi. For Hidden Markov Modeling we first need to get the over view of the process that takes into account all the working of Hidden Markov Model, that are, the parameters of Hidden Markov Models.

vii. We need to observe the sequence of today's stock values i.e. Opening Value, High Value, Low Value and Closing Value. Table 4.1 shows the entire data set of one company which I chose for prediction.

viii. Matching values is an interesting concept, what really happens is that your system matches todays stock values with the values in the data set. After matching system chooses the closest matching value. Consider O_1^0 be a vector of four parameters - daily close, open, high and low. The mathematical expression for computation is as under:

$$j = argmin_i(|P[{O_1}^0, {O_1}^{-1}, {O_1}^{-2}, ..., {O_1}^{-K}|\lambda] \\ - P[{O_1}^{-i}, {O_1}^{-i-1}, {O_1}^{-i-2}, ..., {O_1}^{-i-K}|\lambda] \\ where, i = 1, 2, ..., \frac{T}{K} \quad (4.3.1)$$

ix. To make sure the results are accurate, we find differentiate price change from the targeted day to the coming day. This change is the factor that is added to get the next day's prices.

x. Now we subtract both values:

$$\gamma = ({O_2}^1 - {O_2}^0)$$

Where, ${O_1}^0$ = present day vector of four elements
${O_1}^1$ = predicted day
${O_2}^0$ = chosen day sequence
${O_1}^{-1}$ = day before the chosen day
${O_2}^1$ = day after the chosen day.

xi. Now that we have obtained γ, we will add γ to the current four values of stock market.

opening value + γ = **predicted opening value**

highest value + γ = **predicted highest value**

lowest value + γ = **predicted lowest value**

closing value + γ = **predicted closing value**

xii. This is how the Hidden Markov Models work. This process shows that prediction using Hidden Markov Models is accurate.

xiii. After the calculation of all the predicted values the biggest and the foremost important component is the accuracy and error. Then I will calculate the Mean Absolute Probability error MAPE and Root Mean Square Error RMSE to calculate the probability of error and accuracy of my model .

$$MAPE = \frac{1}{T}\sum_{i=1}^{m}\frac{|Y'(i) - Y(i)|}{|Y(i)|} \times 100\% \tag{4.3.2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m}(Y'(i) - Y(i))^2}{T}} \tag{4.3.3}$$

Where Y'(i) = predicted values, Y(i) = actual values and T = total number of observations.

4.4 Making Prediction with HMM

I build an HMM as an observer density function with guassian mixtures. The HMM has five hidden states. Each state is connected with the Guassian mixture having four distributions of Guassian probability. The HMM is the fully linked trellis, each with a random transition probability. The probability of allocation of initial state probability is a collection of random variables with initial mixture weights w. For each of the Guassian probability distribution functions, the initial covariance and mean values are determined. But the question arises we do not know how many times this algorithm runs? to solve this problem we assigned number of iteration value to 10,000. this algorithm will then run 10,000 times and get the convergence.

The amount of states is intuitively the amount of approaches that we have to predict. This amount is not supposed to be too big or too low. Too many approaches will create little difference between themselves and boost the computing costs. On the other hand, if the number of states is too small, there will be an increased risk of misclassification. Using financial theory, this can also be clarified intuitively: the more diverse the approaches, the less danger of investment and the less yield.

The forecast method is actually conducted simultaneously with the training. Upon completion of the training and updating of HMM parameters, we conducted the Viterbi algorithm to discover the most accurate Gaussian mixtures connected with the most likely state. The weighted average of the Gaussian mixture mean will be given as tomorrow's forecast. The training set can be of any magnitude, but as most economic experts think, it would be beneficial to set a five day or twenty day set because five days are trading days in a week and one month is about twenty trading days. There is a balance between the training set size and the model's sensitivity.The larger the training set, the more precise the model will be in discovering a long-term pattern. The qualified model may not, however, be susceptible to whipsaws.

4.5 EM Maximization or γ Index

For correct understanding of Forecasting System deployed in this research, you must have handy working knowledge of Baum-Welch algorithm. Now Baum-Welch algorithm uses a technique called Expectation Maximization to achieve maximum stability with Hidden Markov Models. Realizing the change in importance over time and motivated by the concept of the financial time series ' Exponential Moving Average (EMA), I present the derivation of the EM algorithm. I demonstrate that the ultimate re-estimation of the HMM and Gaussian densities parameters can be represented as a mixture of the intermediate state variables' estimated moving averages (EMAs). To discover stronger equilibrium between short-term and long-term trends, and I further create an EM algorithm that is sensitivity-adjustable by including data on volatility.

4.5.1 Log-Likelihood

I have a $P[X|\theta]$ distribution matrix that reflects the probability of the $X = X_i, .., X_N$ observational sequence considering the θ parameter. The HMM model θ involves the collection of means and covariances connected with each state of the Gaussian mixtures as well as the transition probability matrix and the likelihood of each element of the mixture. Assuming the information is divided independently and identically (i.i.d.) with distribution p, the probability of observing the entire information set is:

$$L(\theta; t) = f_t(t; \theta) = \prod_{i=1}^{N} p(X_i|\theta) \tag{4.5.1}$$

Where, L is referred to as the likelihood function of the parameters given the data and t is the number of successive events occurs. Similarly, the log-likelihood function l can be written as:

$$l(\theta) = \sum_{i=1}^{n} \log f_i(t_i; \theta) \tag{4.5.2}$$

Since the log function is monotonous, the logarithm shape of the likelihood function is often maximized because it decreases multiplications to sums, making the classification analytically easier. We will use 4-dimension vector; this is due to multitude of inputs that we are operating with.

$$\gamma = ln({O_2}^1 - {O_2}^0)$$

where ${O_2}^0$ is the chosen day log-likelihood vector. This vector is calculated through Estimated Maximkzation. Now to explain the above given equation, we have to remember what we discussed the 4 dimensional matrix that we have used contains 4 values that we discussed about:

$i.$ Opening Value

$ii.$ High Value

$iii.$ Low Value

$iv.$ Closing Value

Now we take this step by step and discuss it in detail:

This shows that we have found out fractional change $\boldsymbol{\gamma}$ by taking the difference of all four parameters of the sequence of chosen day with the day the after the chosen day sequence. this fractional change has some vector value which is basically our γ index. Only because of this index we are able to predict the next day price of the stocks. This way we can compute the fractional change in the system.

Chapter 5

Results

The aim of the research is to predict the values of stocks parameters for the next 100 days by the predicting value as an initial value and check the accuracy of the trained model. These parameters have estimated by EM-algorithm. I use the Hidden Markov Models for stock prices prediction because if we do not suppose any shift in future stock prices for trading reasons, then there is no trade. I observed that the predicted values for Open, Close, High, Low closely followed the trends shown in the HMM as well as the other techniques implementations and the MAPE values by their corresponding true values were found to be similar. It was not discovered that the projections produced using the other models were influenced by dramatic stock price changes. However, it was discovered that the model applied using HMM is extremely susceptible to stock price changes.

5.1 Comparison With Other Techniques

The findings show that the stocks trends are well captured by the HMM. The other prediction techniques had some enormous mistakes in anticipating stocks prices. The other systems also reveal its vulnerability when Google stock rates are far from the actual rates. In the following table 5.1, comparison of different MAPE values using different techniques are shown. The calculation of MAPE signifies that the system has that much percentage of error. For example, the highest accuracy achieved in prediction so far is 81%. The other techniques Mape values are taken from the sources (Nguyet Nguyen and Ayush Jain)[24] [25]. According to our research the MAPE value for different stock is as under:

COMPARISON OF MAPE VALUES				
Company Name	**HMM**	**ARIMA**	**ANN**	**SVR**
Apple	0.0113	0.0810	0.0730	0.0811
Dell	0.0111	0.0665	0.0675	0.0773
Google	0.0107	0.0884	0.0798	0.0914

Table 5.1: Comparison of MAPE of three stock price prediction of Apple, Dell and Google between Hidden Markov Model and other Artificial Intelligence techniques

5.2 Proposed Method Results

In this section, we will see the predicted parameter (close, high, low and open) of different stocks. From investor point of view, by viewing this predicted values, investor can invest in the stock price. For example, if HMM predicts that the stock price of certain company will increase tomorrow, then he can buy this stock today and sell it tomorrow by observing the close prices of the stocks. On the other hand, if the HMM predicts that the stock price will not increase tomorrow, then investor will do nothing. To check the accuracy of the proposed model, I take the actual and the predictedd values of one particular day which is shown in Table 5.2.

Date	**Actual Prices in USD**				**Predicted Prices in USD**				**Accuracy (%)**
	Open	**High**	**Low**	**Close**	**Open**	**High**	**Low**	**Close**	
01/06/2019	5.88	5.91	5.83	5.84	5.33	5.4	5.35	5.34	90.6

Table 5.2: Accuracy of Proposed Hidden Markov Model by observing the actual and predicted prices of one day i-e; 1 June 2019

Based on the AIC and BIC results, we only use HMM with five states for the stock prediciton. Again, we will use a block of 253 days as an observation data, from 2 October 2018 to 1 June 2019 for model testing. We present the results of AIC and BIC of different companies in the Table 5.3.

Mean Absolute Percentage Error		
Company Name	AIC	BIC
NOKIA CORPORATION	561.94	583.121
DG CEMENTS	-22.444	19.908
PTCL	-1293.11	-1222.52
SNGP	-1507.64	-1402.760
GOOGLE	783.821	772.408
APPLE	787.861	792.728
HUAWEI	-370.308	-369.366

Table 5.3: AIC and BIC of Different Companies for performance measurement of the model

After training the model and get the results, the performance of HMM is also checked by calculating Mean Absolute Percentage Error (MAPE) and RMSE (Root Mean Square Error) of the estimations. Table 5.4 and 5.5 shows the Performance of the proposed models implemented on the stocks of few companies.

Mean Absolute Percentage Error				
Company Name	Open	High	Low	Close
FEROZSONS	0.03345	0.029876	0.037892	0.037463
DG CEMENTS	0.02343	0.024896	0.027312	0.023931
PTCL	0.0302022	0.025759	0.033014	0.029641
SNGP	0.0445	0.04769	0.04792	0.04163
GOOGLE	0.02767	0.02860	0.02092	0.027431
APPLE	0.02110	0.01881	0.01949	0.02033
HUAWEI	0.03174	0.03976	0.03589	0.03743

Table 5.4: Mean Absolute Percentage Error of stock price prediction of different companies using HMM with same number of observation data and hidden states

Mean Absolute Percentage Error				
Company Name	Open	High	Low	Close
FEROZSONS	0.13211	0.11207	0.12104	0.12833
DG CEMENTS	0.2343	0.24896	0.27312	0.23931
PTCL	0.302022	0.25759	0.33014	0.29641
SNGP	0.3345	0.29876	0.37892	0.37463
GOOGLE	0.1127	0.1503	0.1379	0.1564
APPLE	0.2408	0.2716	0.3158	0.2763
HUAWEI	0.5228	0.4575	0.4450	0.4720

Table 5.5: Root Mean Square Error of stock price prediction of different companies using HMM with same number of observation data and hidden states

The model's hypothesis is that observation sequences were extracted from a hidden state series that is separate information and satisfies a Markov process's first order. HMM has been created from a single observation model to a multi-observation model. HMM's applications have also been extensively extended in many areas, such as speech recognition, biomathematics and financial mathematics.According to the analysis, the number of states used in the HMM model also affects the prediction's accuracy but I keep the number of hidden states and observation data same and maximum to predict the stock prices for more accurate results. However, AIC and BIC measure the calibrations of the HMMs parameters while MAPE and RMSE evaluate the prediction of model stock price. Following figures shows the graphical representation of Hidden Markov Model prediction between actual and predicted stock market prices.

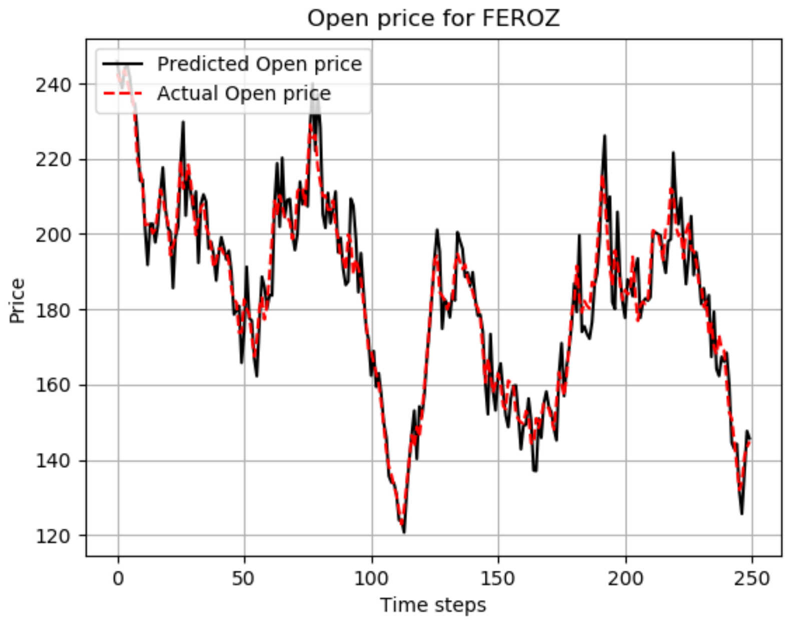

Figure 5.1: Comparison of HMM prediction of Ferozsons open stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

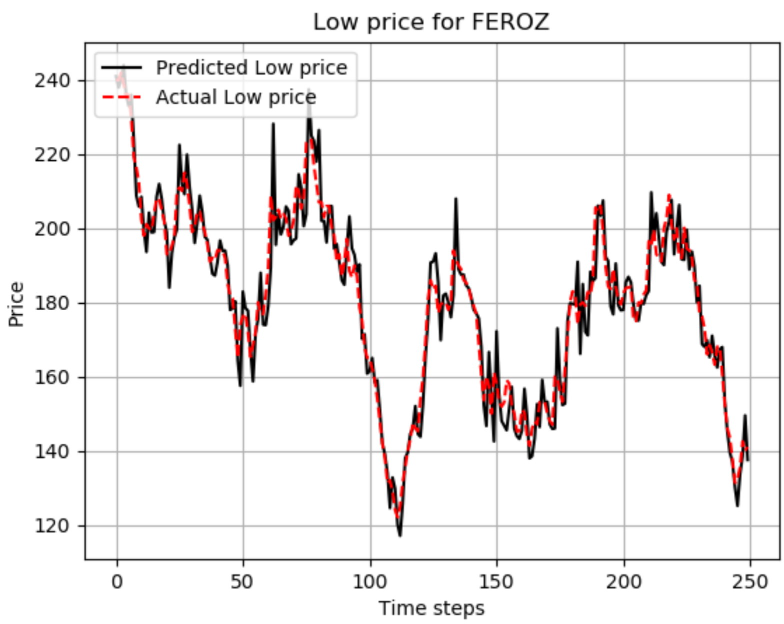

Figure 5.2: Comparison of HMM prediction of Ferozsons low stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

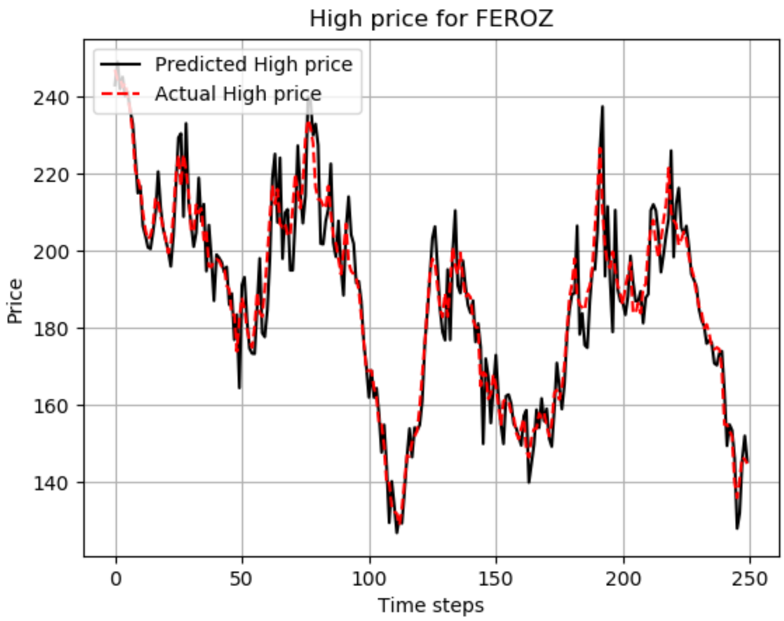

Figure 5.3: Comparison of HMM prediction of Ferozsons high stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

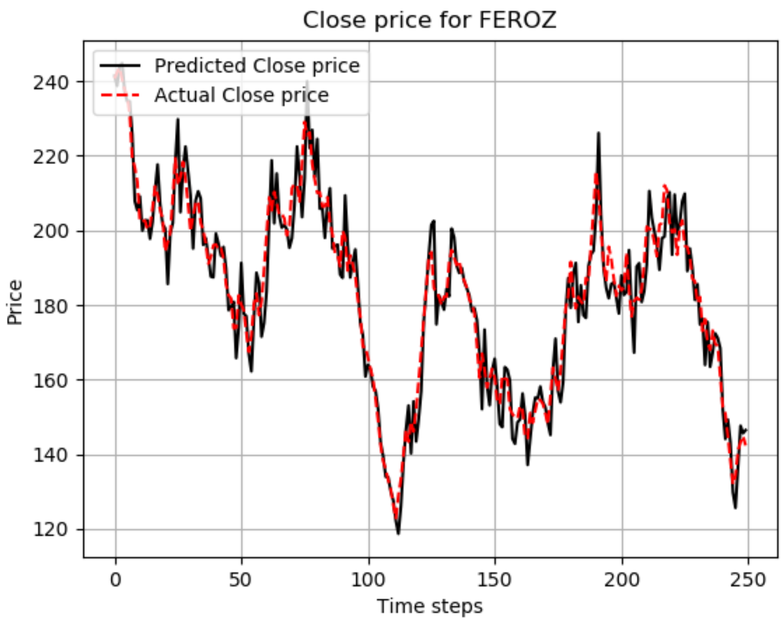

Figure 5.4: Comparison of HMM prediction of Ferozsons close stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

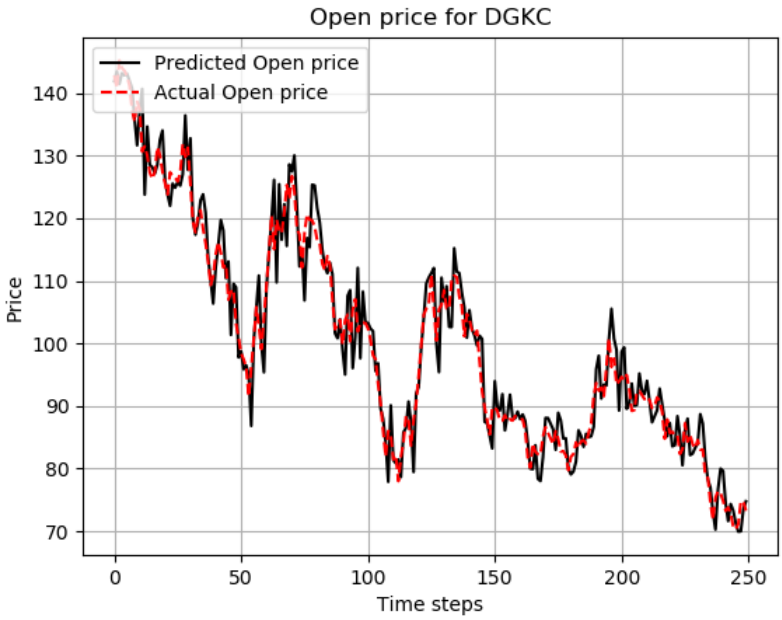

Figure 5.5: Comparison of HMM prediction of DG Cements open stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

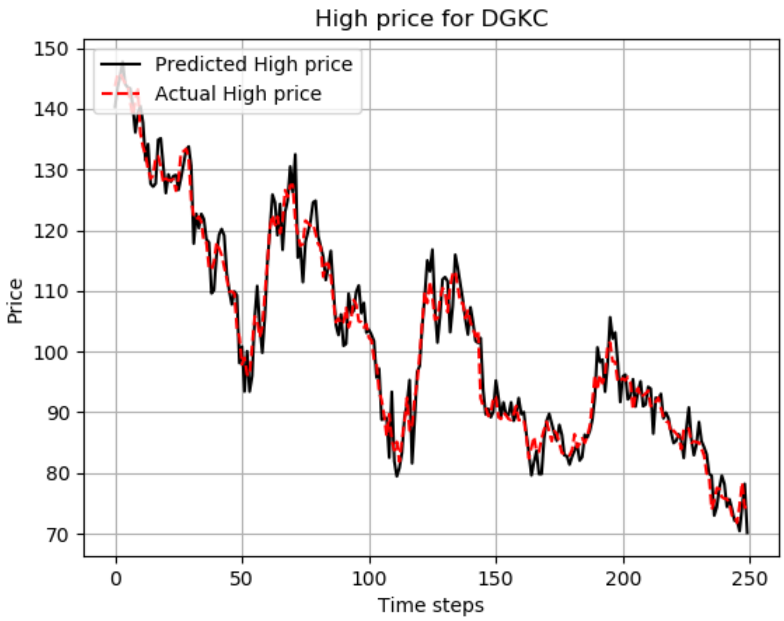

Figure 5.6: Comparison of HMM prediction of DG Cements high stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

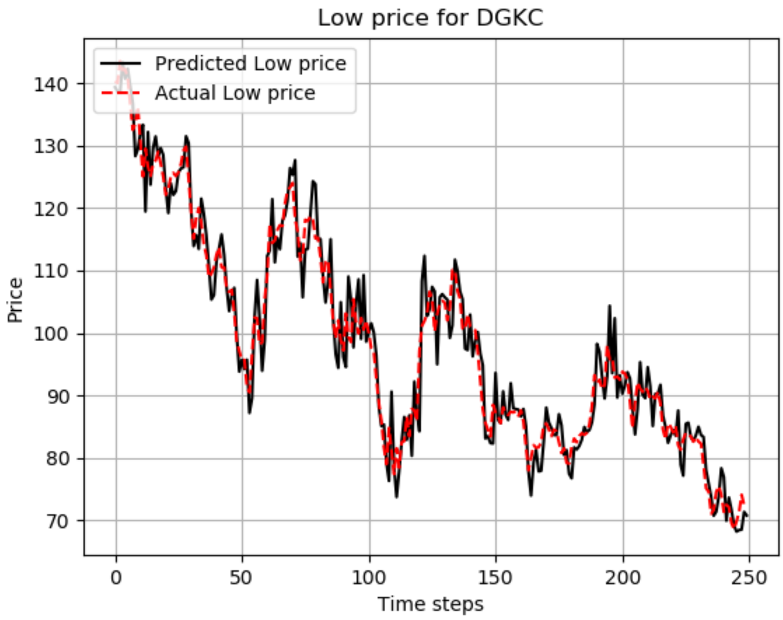

Figure 5.7: Comparison of HMM prediction of DG Cements low stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

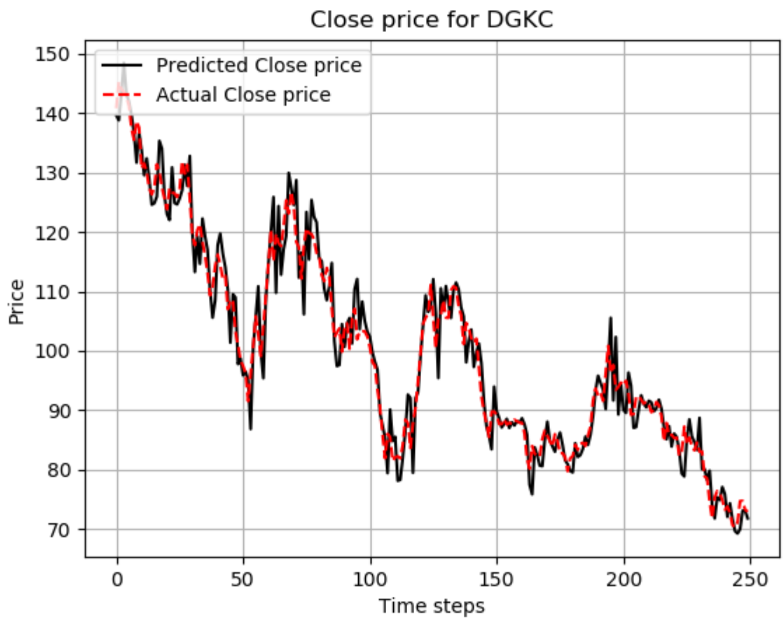

Figure 5.8: Comparison of HMM prediction of DG Cements close stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

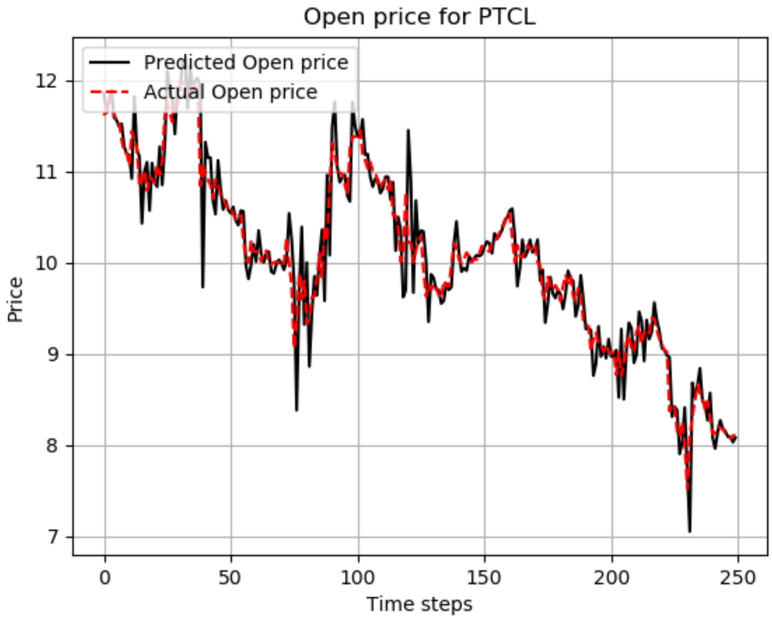

Figure 5.9: Comparison of HMM prediction of Pakistan Telecommunication Company Limited open stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

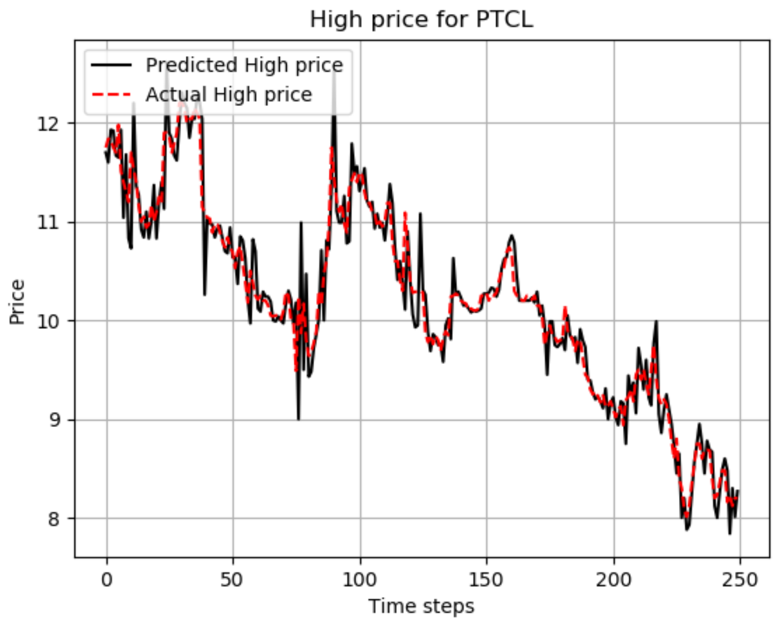

Figure 5.10: Comparison of HMM prediction of Pakistan Telecommunication Company Limited high stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

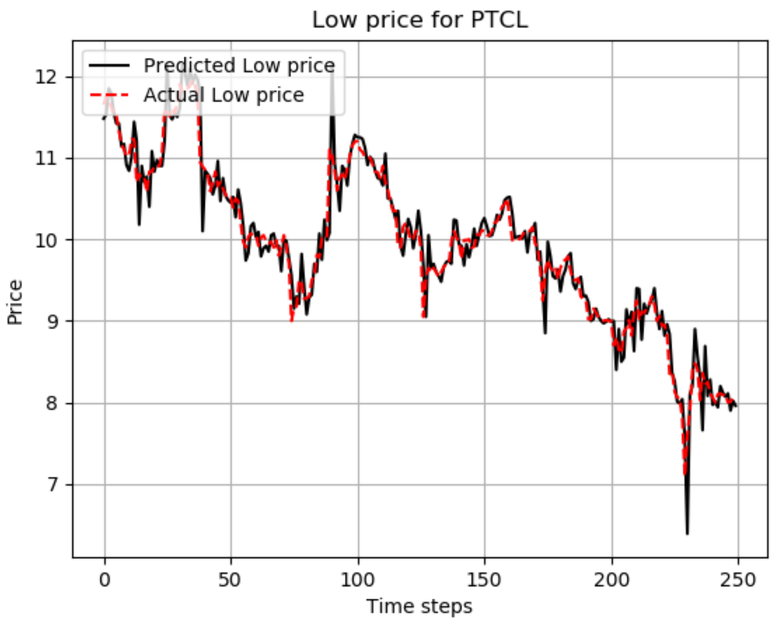

Figure 5.11: Comparison of HMM prediction of Pakistan Telecommunication Company Limited low stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

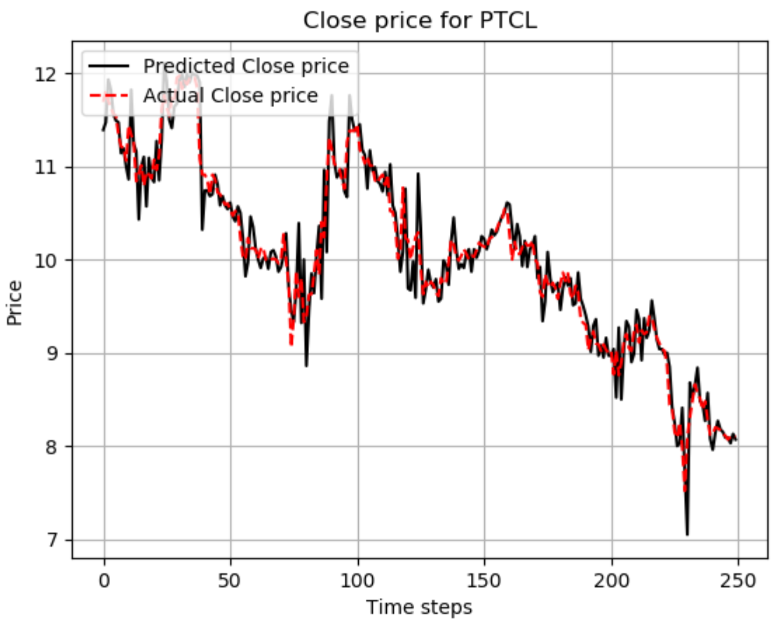

Figure 5.12: Comparison of HMM prediction of Pakistan Telecommunication Company Limited close stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

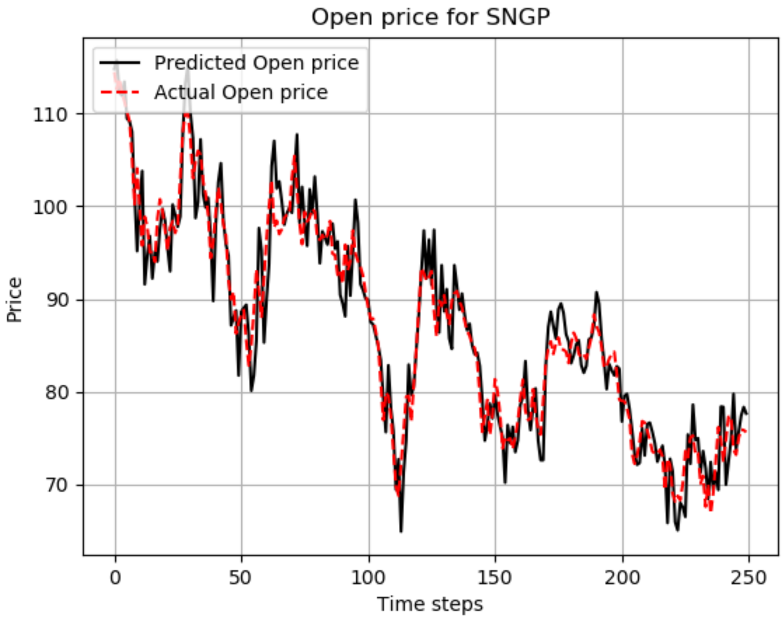

Figure 5.13: Comparison of HMM prediction of Sui Northern Gas Pakistan Limited open stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

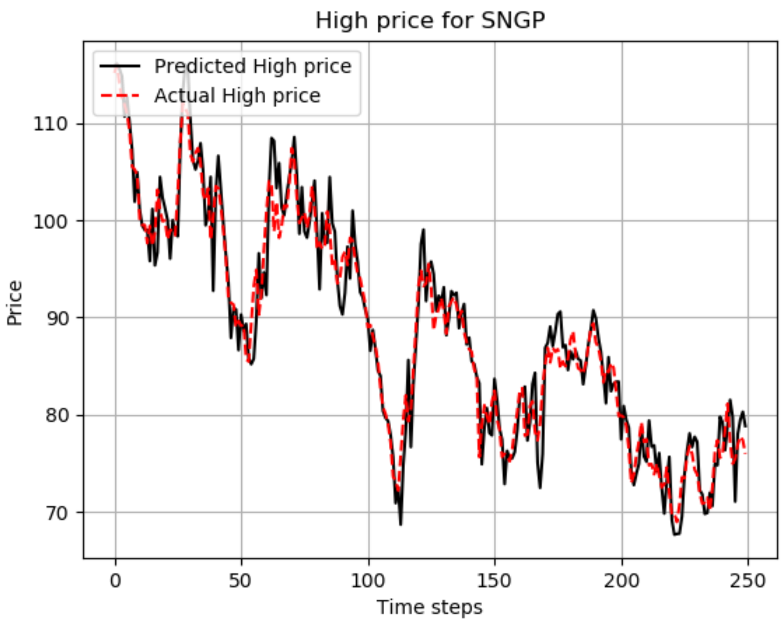

Figure 5.14: Comparison of HMM prediction of Sui Northern Gas Pakistan Limited high stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

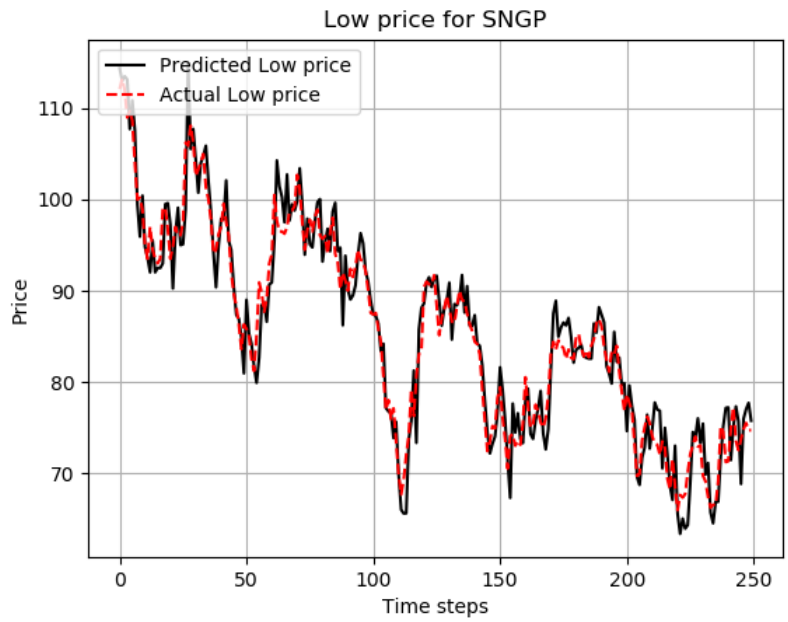

Figure 5.15: Comparison of HMM prediction of Sui Northern Gas Pakistan Limited low stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

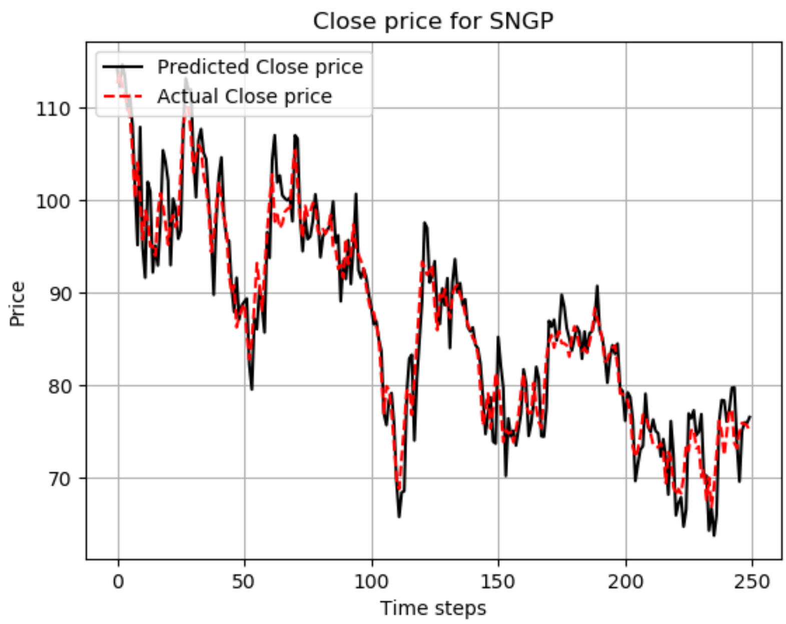

Figure 5.16: Comparison of HMM prediction of Sui Northern Gas Pakistan Limited close stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

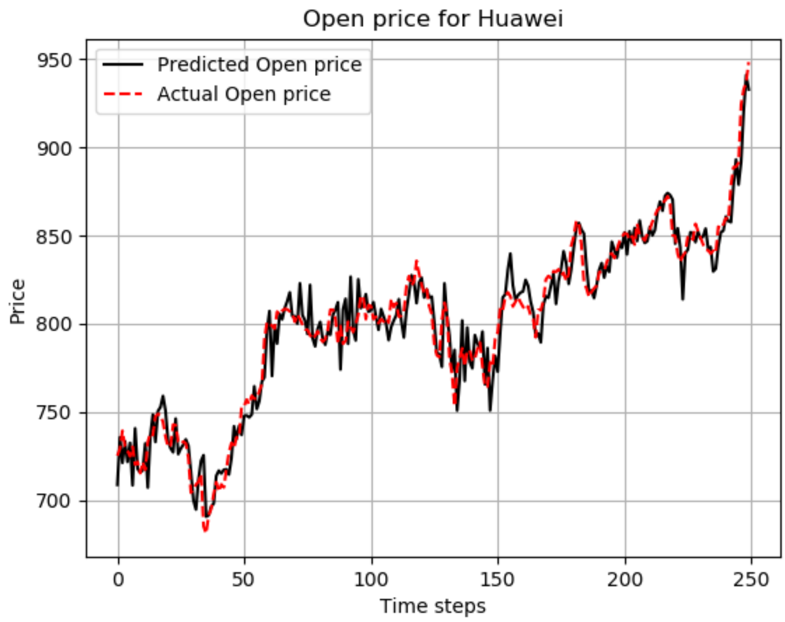

Figure 5.17: Comparison of HMM prediction of Huawei Technologies Company Limited open stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

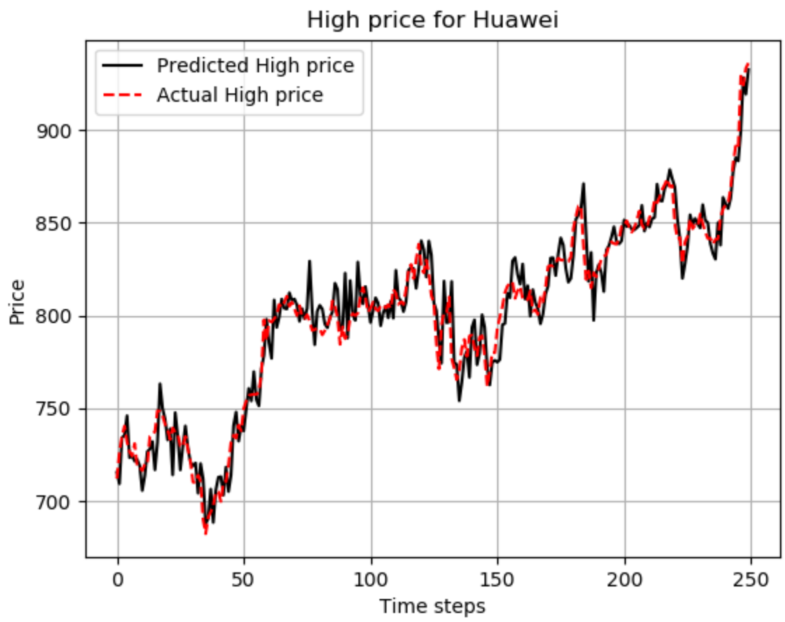

Figure 5.18: Comparison of HMM prediction of Huawei Technologies Company Limited high stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

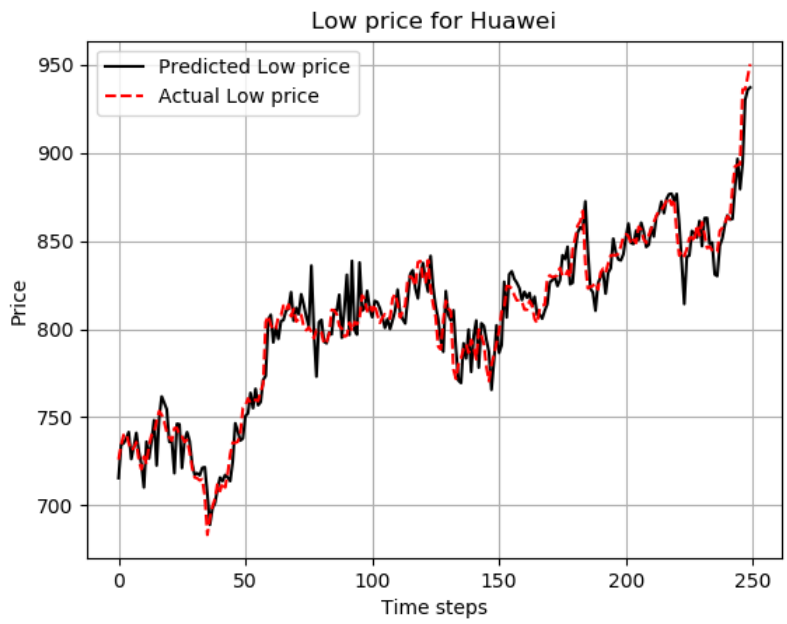

Figure 5.19: Comparison of HMM prediction of Huawei Technologies Company Limited low stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

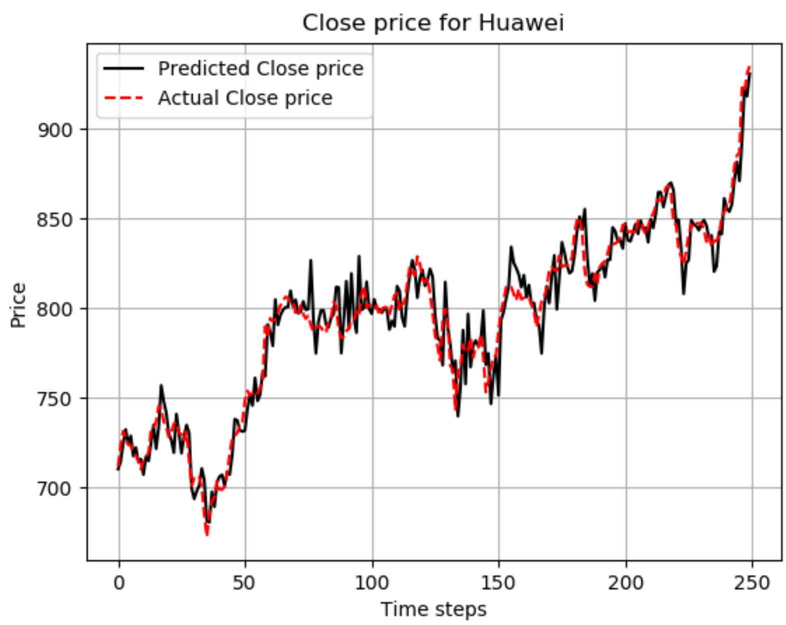

Figure 5.20: Comparison of HMM prediction of Huawei Technologies Company Limited close stock parameter values and actual open value of the observation data from 2^{nd} September 2018 to 1^{st} June 2019

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Stock performance is an important measure of stock corporation's strengths and weaknesses and the economy in particular. There are many variables that drive up or down stock price. We use Hidden Markov Model, HMM, in this thesis to forecast various firms ' open, low, high and close prices. We first use the criteria of AIC and BIC to examine model's performance. The findings indicate that HMM with same number of states yielded very comparable AICs or BICs. Model performance alternated with simulations and shares. While it will create more errors at moments when the economic time series is unstable, HMM may eventually catch up with the trend. By observing the historical data of these companies and calculate the Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) for each parameter. Errors calculated by using Hidden Markov Model as training model is excellent for volatile and unstable data. The resultant output shows the number of states used by the training data in order to predict the closing, opening, high and low prices of the particular company. The numerical findings indicate that the HMM is the greatest option for stock prices to forecast. The results show that the HMM is the future model for stock trading as it reflects the stock price trends well.

6.2 Future Work

Forecasting can play a vital role in boosting Pakistan Stock Exchange, with a maximum number of people taking interest once they get hold of our proposed system. Stock market booms when a maximum number of investors invest their money in PSX.

The proposed system can have multiple applications – like Gold prices prediction, Fuel prices prediction, Dollar price prediction, Flood prediction. Hence, the proposed system will serve our nations with a Forecasting ahead of time to sidestep an unpredicted event. It will assist to make stronger projections by finding helpful side data for forecast, such as creating new indicators. Then, rather than simply measurements, we can expand the forecast to a density function type. In this manner, we can deliver better market movements data. This can be a worthwhile job to do. When modeling random or unknown procedures, Gaussian distribution in Hidden Markov Model is the most common distribution. But it was noted that the Gaussian distribution may not follow many time series. By using other distributions, we can make more accurate predictions.

6.3 Advantages:

i. A new research dimension in Pakistan for prediction of stock market.

ii. More gains and profits to investors and stock brokers.

iii. Robust system trained to compute any dynamic sequences.

iv. 24/7 prediction system, that can compute any complex sequence that can be forecasted accurately.

References