

# IV Year Project Report

*Under the guidance of*

*Dr. SANJUKTA DAS*

*Assistant Professor Department of Mathematics*

*Submitted by*

*Sai Mahitha - 17XJ1A0210*

*Surya Teja - 17XJ1A0216*

*Sai Reddy Snehith Reddy-17XJ1A0240*

Department of Mathematics

Department of Electrical and Electronics Engineering May 2021

CONTENTS



1. ACKNOWLEDGEMENT 3
2. ABSTRACT 4
3. INTRODUCTION 5
4. METHODOLOGIES AND RESULTS

4.1 RANDOM WALK SIMULATION 5

4.2 MARKOV PROCESS

4.2.1 INTRODUCTOIN 7

4.2.2 CONSTRUCTION OF MARKOV CHAIN 7

4.2.3 ANALYSIS 8

4.3 HIDDEN MARKOV MODEL(HMM)

4.3.1 WHY HMM? 9

4.3.2 INTRODUCTION 9

4.3.3 WORKING 10

4.3.4 ANALYSIS 12

4.4 OPTIMAL PORTFOLIOS AND EFFICIENT FRONTIER

4.4.1 OPTIMAL PORTFOLIOS 15

4.4.2 EFFICIENT FRONTIER 15

4.4.3 UNDERSTANDING EFFICIENT FRONTIER 16

4.4.4 ASSUMPTIONS AND LIMITATIONS 16

4.4.5 PLOTTIG THE EFFICEINT FRONTIER 17

1. CONCLUSION 17
2. REFERENCES 18

## ACKNOWLEDGEMET

*The achievement and result of this undertaking required a ton of direction and help from numerous individuals and we are extremely privileged to have got this all along the completion of our project. All that we have done is simply because of such management and help and we would not neglect to express gratitude toward them.*

*We respect and thank* ***Dr. Yajulu Medury****, for providing us an opportunity to do the project work in* ***Mahindra University*** *and giving us all support and guidance, which made us complete the project duly. We are incredibly appreciative to him for offering a particularly decent help and direction, although he had busy timetable managing the corporate affairs.*

*We owe my deep gratitude to our project guide* ***Dr. Sanjukta Das****, who took keen interest on our project work and guided us all along, till the completion of our project work by providing all the necessary information for developing a good system.*

*We are thankful to and fortunate enough to get constant encouragement, support and guidance from all Teaching staffs of* ***Electrical and Electronics department*** *as well as* ***Mathematics department*** *which helped us in completing our project work.*

*Many individuals, particularly our classmates and team members itself, have offered significant remark recommendations on this proposition which motivated us to improve our task. We thank all the members for their help directly and indirectly to complete our task.*

## ABSTRACT

Financial exchange is considered as the primary indicator of a nation's economic strength and development. Stock market prices are unstable in nature and are influenced by factors like economic growth, inflation, and so forth. Costs of a share market rely vigorously upon demand and supply. High demanded stocks will increase in price, eventually, vigorously sold stocks will diminish in price. Fluctuating stock costs influences the investor's conviction and subsequently there is a need to foresee the future stock worth.

The target of this undertaking is to anticipate the financial exchange costs to settle on more educated and precise speculation choices. Recent patterns in stock market prediction are overviewed. Different types of approaches and their respective variants. Various approaches and the results of past years are compared based on methodologies and datasets. This report describes various theories and conventional approaches to stock market forecast and analysis.

To overcome the randomness of stock market fluctuation we have used the Random Walk simulation to analyze and forecast upcoming stock prices of the company. Since A Geometric Brownian motion is a constant time stochastic cycle in which the logarithm of the arbitrarily differing amount follows a Brownian with drift we utilize the Geometric Brownian motion equation in the Random Walk algorithm to simulate the desired no of random paths.

There are other stochastic processes like Markov chain which is a probabilistic approach where we deal in states. Further we have also discussed about Hidden Markov model (HMM) where the states are considered to be hidden.

Key words: stock, Random Walk, Geometric Brownian motion, stochastic process, Markov, HMM.

## INTRODUCTION

A stock market is a place where people using their DEMAT and trading accounts buy/sell shares of publicly listed companies which offers a platform to facilitate seamless exchange of shares. NSE (National Stock Exchange) and BSE (Bombay Stock Exchange) are the two major stock exchanges in India. India's stock market is governed by the SEBI (Securities and Exchange Board of India). The market is somehow predictable if it always follows the same price patterns but due to so many factors, there is no repetition of patterns. So, to overcome this problem we are trying to analyze market movements using Random Walk Algorithm and find the best path that meets the highest accuracy among all simulated paths to forecast future prices.

## METHODOLOGIES AND RESULTS

**4.1 Random walk simulation:**

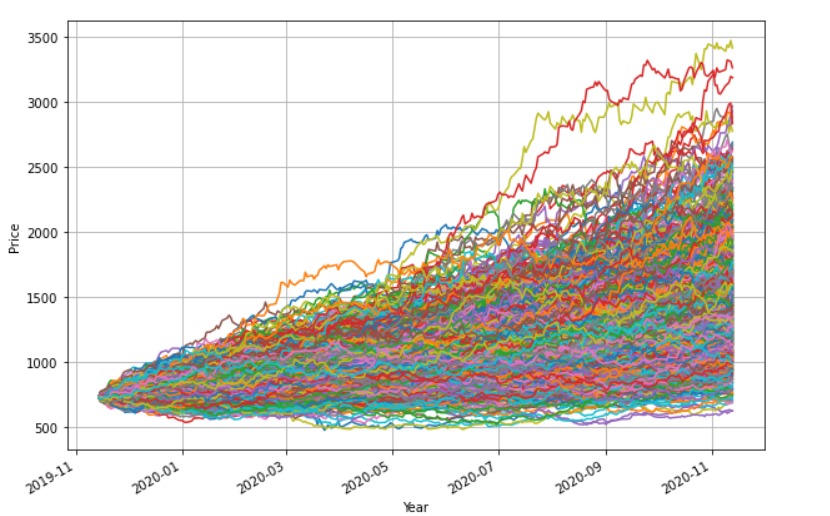
Random walk suggests the direction of stock price is entirely random and un predictable. Technically it is a stochastic process which is seemingly random or unpredictable.

Geometric Brownian movement is a persistent time stochastic cycle in which the logarithm of the randomly changing amount follows a Brownian Motion.

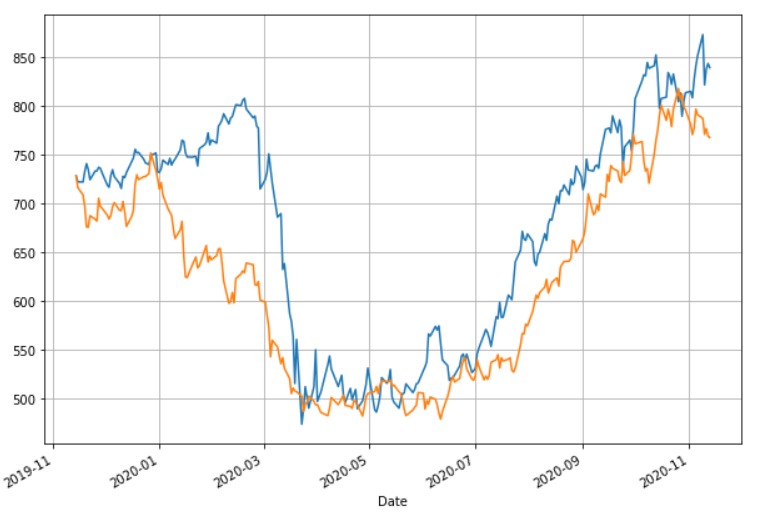
GBM Equation:

St = St-1 \* exp((μ-(σ2 /2))\*t + σWt )

First, we scraped stock data of Tech Mahindra (TECHM.NS) from 17-10-2017 to 17- 11-2020 from https://in.finance.yahoo.com/ and using Jupyter notebook inbuilt libraries we found percentage change of daily closing values using those values we ended up finding mean and standard deviation of percentage change of daily closing values. Using these values in the Geometric Brownian motion equation we simulated 1000 paths and exported data of these 1000 paths. After analyzing we found two paths are almost perfectly matching with the original market trend.



Results: These are the 999 simulated paths of Tech Mahindra company



Blue- Actual Path

Orange- Closest projected path

## Markov Process:

4.2.1 Introduction

In Markov process, rather than predicting the price of the stock itself, we basically try to predict how probable is it for a stock price to finish in a certain state. It is a probabilistic approach towards forecasting a stock in the near future. A state can be classified in many ways. For example, we can refer to the closing price of a stock and define 2 states, S1 if the closing price is greater than the previous day and S2 if it is lower. We can also define sates by giving ranges to each state, for example state S1 from 0-$100 and S2 from $100-$200 and so on.

Markov process is a stochastic process with no after-effect property. This means that the state at time tx+1 is only dependent on the state at time t, not on all the previous states. In other words, the next state is only dependent on the current state, not on all the previous state. This is a very important Markov property which makes Markov process preferable over other predicting models like regression analysis. Another property of Markov process is that it is ergodic, which means that after sufficiently long time, irrespective of the starting state of the system, the probability of finishing in a particular state will converge to a constant.

* + 1. Construction of Markov chain forecasting model

To predict the next state of stock price, we will need to multiply the initial state vector with the transition probability matrix. Because of the ergodic property, after sufficient multiplications we will see that the probability of finishing in each of the states will converge to a constant value. With this we can tell how probable is it for the price to finish in each of the state.

The initial state vector refers to the vector representation of the state of the last observation. For example, if the last observation is in the S1 state and the system has three states in all then the initial state vector would be [1 0 0].

The state transition matrix is made up of the elements Pij  where Pijth element is the probability of the state transitioning from state i to state j. It is explained through the image below.

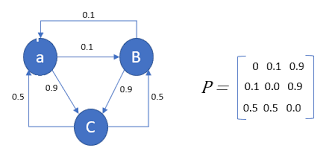


Figure 4.2.1: Example of a transition probability matrix

* + 1. Analysis of a stock price

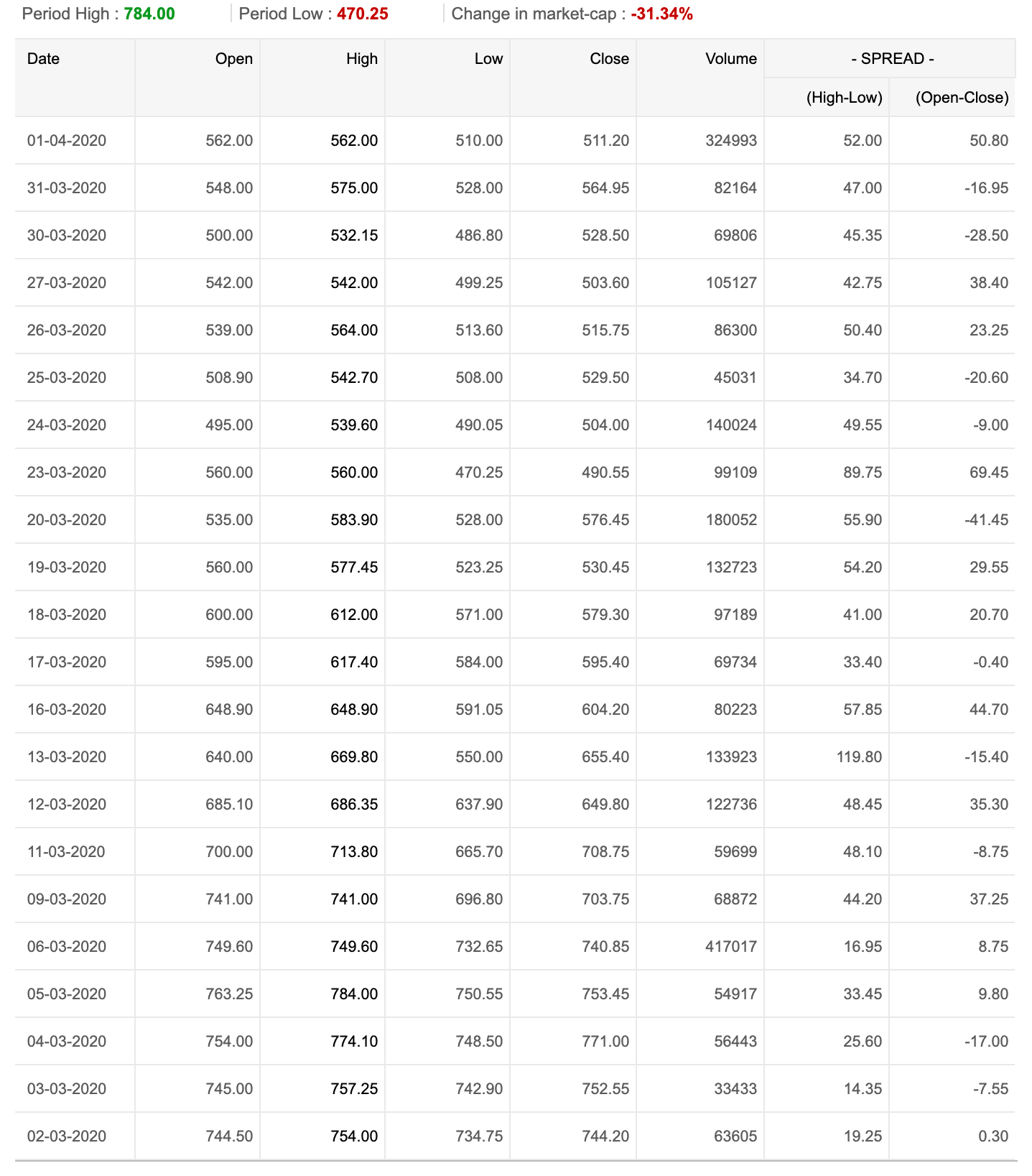


Figure 4.2.2: The stock price of Tech Mahindra for the period 2/3/2020-1/4/2020

Consider the closing values shown in the above figure. For our system, lets assume we operate with three states, S1 S2 and S3.

S1 if the previous day closing value is greater than the present, S2 if it the same and S3 if the previous day closing value is lesser than the present day.

Simple if there is an increase, S1. If the price remains same (less than one rupee difference) S2 and if it decreases, S3.

First, we need to construct a table classifying the states.

Table 4.2.1: Indicating states of observations

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| S.no | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| States | S2 | S3 | S3 | S1 | S1 | S1 | S3 | S1 | S3 | S1 | S2 | S1 | S1 | S3 | S1 | S3 | S3 | S1 | S1 | S3 | S3 | S1 |

We can see from the above table that we have 22 observations. Now we need to build our transition probability matrix. For that we need to count the number of occurrences of each of the states. It is to be noted that we exclude the last observation because we cannot determine its next state. So therefore S1 – 10, S2 – 2, S3 – 9.

For P11 we need to look at all the occurrences where S1 is transitioning to S1 and divide it by 10. So therefore P11 = 4/10 = 0.4

Similarly, all other probabilities are calculated and shown below.

P11 P12 P13 0.4 0.1 0.5

P21 P22 P23  = 0.5 0 0.5

P31 P32  P33 0.67 0 0.33

We now need to multiply this matrix with [1 0 0] because our state in the last observation is S1. This multiplication has to be repeated enough times till we converge to a value.

Finally, we get: S1 - 0.52 S2 – 0.051 S3 – 0.429. These are the probabilities with which they can finish in each of these states.

* 1. **Hidden Markov Model (HMM)**

4.3.1 Why HMM?

The Markov model described previously has limited power in many applications. So the Markov model has been extended to Hidden Markov model (HMM) because there is a good match between the sequential data analysis and HMM as explained later in this section.

* + 1. Introduction

In an HMM, unlike in the Markov model, the states are not directly observable. HMM is a stochastic model where the states are assumed to be hidden but they can be estimated by using observable symbols associated with the hidden states. At each time instance, the HMM emits a symbol and changes the state with certain probability. There is no one to one relation between states and symbols.

A HMM consists of

* A set of hidden states (S)
* A set of output symbols (O)
* State transition probability matrix (A)

Also known as TPM, Probability of going from one state to any of the states

* Emission probability matrix (B)

EPM, the probability of emitting a symbol at a particular state

* Initial state probability distribution vector(π)

HMM is defined as 𝜆 = (S,O,A,B, π) where the following conditions should be met,

where 1 ≤ i ≤ N

where 1 ≤ i ≤ N

where πi ≥0

* + 1. Working

Let’s take an example to better understand the working of HMM. Consider the following figure where we have 5 investment strategies and market movement has been classified into large drop(lr), small drop(sd), no change(nc), small rise(sr) and large rise(ls).

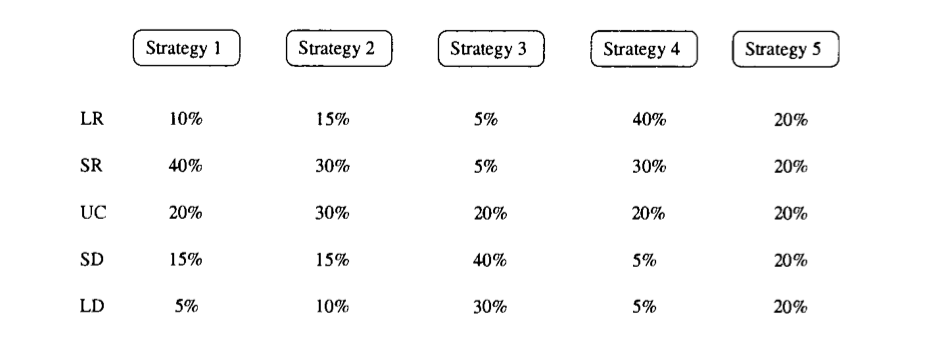


Figure 4.3.1: Strategy and probability associating with price movements

The HMM will next pick a strategy based on the observation probability. The importance here is that it can pick the best overall sequence of strategies based on the observation sequence.

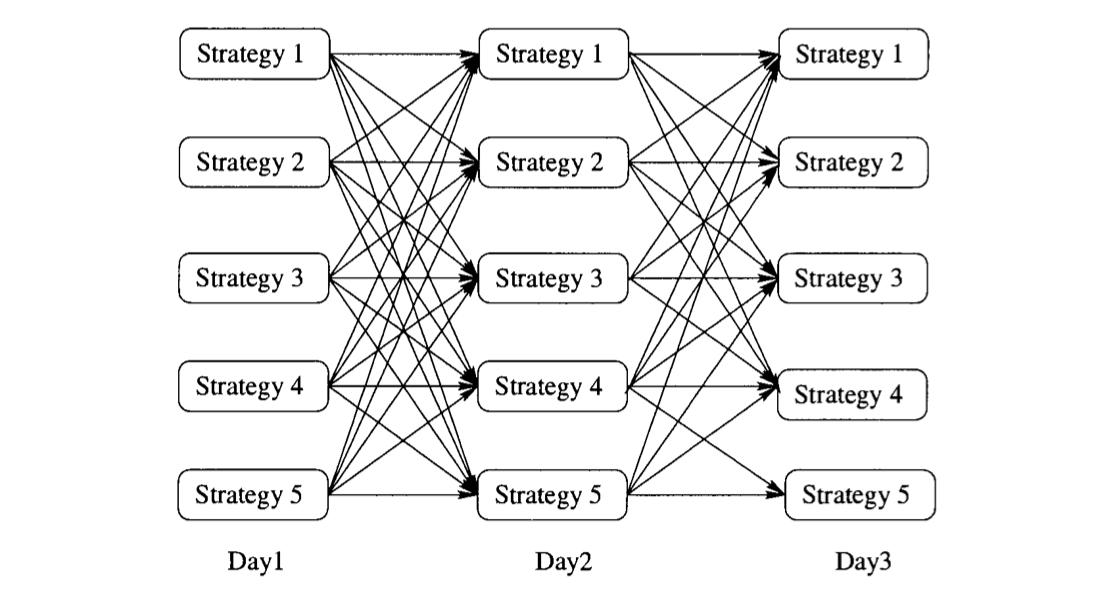


Figure 4.3.2: example of the shifting from one strategy to another representing the states

Now suppose if we wanted the probability of the movement from small drop to large rise, starting from say strategy 2 we can calculate it in the following way.

Table 4.3.1: State transition probability matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Strategy | 1 | 2 | 3 | 4 | 5 |
| 1 | 10% | 10% | 20% | 30% | 30% |
| 2 | 10% | 20% | 30% | 20% | 20% |
| 3 | 20% | 30% | 10% | 20% | 20% |
| 4 | 30% | 20% | 20% | 15% | 15% |
| 5 | 30% | 20% | 20% | 15% | 15% |

For the probability of sd to lr is :

* 1. 0.1\*0.1 + 0.2\*0.15 + 0.3\*0.05 + 0.2\*0.4 + 0.2\*0.2 = 0.325

4.3.4 Analysis on Axis bank data

We have taken data from Yahoo finance and attempted to apply HMM on this data.

Table 4.3.2: The data collected of axis bank

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Date** | **Open** | **High** | **Low** | **Close** | **Adj Close** | **Volume** |
| **2021-02-01** | 663.500000 | 716.849976 | 654.500000 | 710.250000 | 710.250000 | 1192612 |
| **2021-02-02** | 716.000000 | 744.950012 | 704.299988 | 713.700012 | 713.700012 | 915496 |
| **2021-02-03** | 720.900024 | 736.799988 | 715.849976 | 733.599976 | 733.599976 | 501155 |
| **2021-02-04** | 731.700012 | 745.500000 | 720.849976 | 744.049988 | 744.049988 | 1444433 |
| **2021-02-05** | 749.700012 | 766.200012 | 717.049988 | 719.500000 | 719.500000 | 1160592 |
| **2021-02-08** | 725.000000 | 747.799988 | 725.000000 | 736.099976 | 736.099976 | 655455 |
| **2021-02-09** | 735.099976 | 750.200012 | 726.500000 | 742.400024 | 742.400024 | 2596306 |
| **2021-02-10** | 740.599976 | 745.950012 | 726.200012 | 734.500000 | 734.500000 | 4612498 |
| **2021-02-11** | 732.000000 | 744.099976 | 731.500000 | 740.200012 | 740.200012 | 241617 |
| **2021-02-12** | 736.000000 | 754.549988 | 734.650024 | 750.349976 | 750.349976 | 608164 |
| **2021-02-15** | 759.900024 | 798.700012 | 754.049988 | 794.500000 | 794.500000 | 672802 |
| **2021-02-16** | 800.000000 | 800.000000 | 766.799988 | 775.299988 | 775.299988 | 664593 |
| **2021-02-17** | 775.000000 | 784.900024 | 763.799988 | 777.450012 | 777.450012 | 476888 |
| **2021-02-18** | 777.400024 | 781.900024 | 764.250000 | 777.250000 | 777.250000 | 339604 |
| **2021-02-19** | 775.000000 | 777.349976 | 741.400024 | 749.349976 | 749.349976 | 399673 |
| **2021-02-22** | 749.349976 | 750.500000 | 715.000000 | 719.650024 | 719.650024 | 644432 |
| **2021-02-23** | 733.000000 | 734.700012 | 712.700012 | 715.849976 | 715.849976 | 788528 |
| **2021-02-24** | 720.000000 | 754.400024 | 700.150024 | 753.299988 | 753.299988 | 2382990 |
| **2021-02-25** | 766.000000 | 783.250000 | 760.000000 | 770.799988 | 770.799988 | 2082231 |
| **2021-02-26** | 746.900024 | 755.250000 | 718.650024 | 724.700012 | 724.700012 | 884811 |

We have considered two observation symbols;

‘I’ if the present day closing value – previous day closing value > 0

‘D’ if present day – previous day < 0

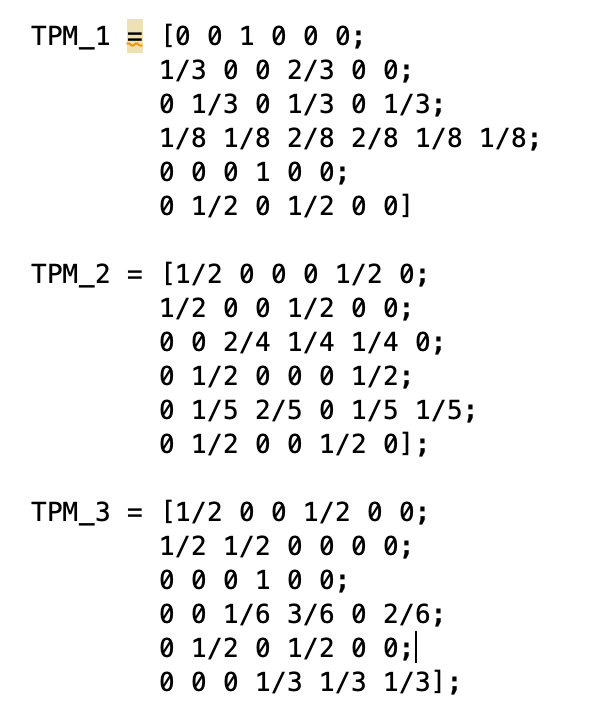
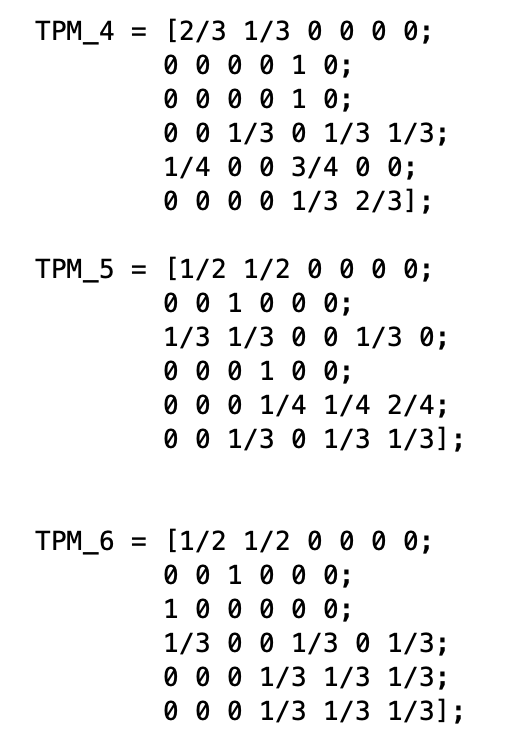
Table 4.3.3: Showing the closing values and difference in 1,2,3,4,5,6 days as D1,D2,D3,D4,D5 D6 respectively.

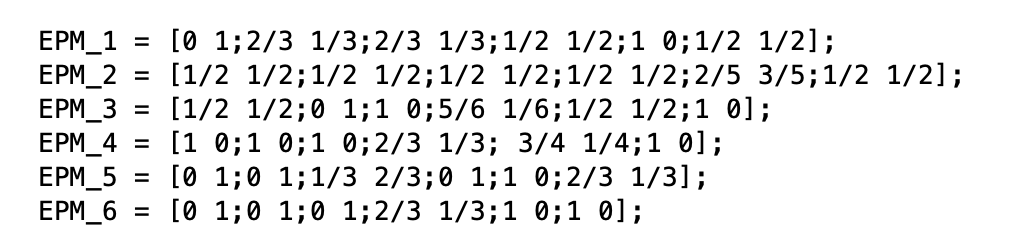
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| S.NO | C.V | D1 | OS1 | D2 | OS2 | D3 | OS3 |
| 1 | 710.25 |  |  |  |  |  |  |
| 2 | 713.700012 | 3.450012 | I |  |  |  |  |
| 3 | 733.599976 | 19.899964 | I | 23.349976 | I |  |  |
| 4 | 744.049988 | 10.450012 | I | 30.349976 | I | 33.799988 | I |
| 5 | 719.5 | -24.549988 | D | -14.099976 | D | 5.799988 | I |
| 6 | 736.099976 | 16.599976 | I | -7.950012 | D | 2.5 | I |
| 7 | 742.400024 | 6.300048 | I | 22.900024 | I | -1.649964 | D |
| 8 | 734.5 | -7.900024 | D | -1.599976 | D | 15 | I |
| 9 | 740.200012 | 5.700012 | I | -2.200012 | D | 4.100036 | I |
| 10 | 750.349976 | 10.149964 | I | 15.849976 | I | 7.949952 | I |
| 11 | 794.5 | 44.150024 | I | 54.299988 | I | 60 | I |
| 12 | 775.299988 | -19.200012 | D | 24.950012 | I | 35.099976 | I |
| 13 | 777.450012 | 2.150024 | I | -17.049988 | D | 27.100036 | I |
| 14 | 777.25 | -0.200012 | D | 1.950012 | I | -17.25 | D |
| 15 | 749.349976 | -27.900024 | D | -28.100036 | D | -25.950012 | D |
| 16 | 719.650024 | -29.699952 | D | -57.599976 | D | -57.799988 | D |
| 17 | 715.849976 | -3.800048 | D | -33.5 | D | -61.400024 | D |
| 18 | 753.299988 | 37.450012 | I | 33.649964 | I | 3.950012 | I |
| 19 | 770.799988 | 17.5 | I | 54.950012 | I | 51.149964 | I |
| 20 | 724.700012 | -46.099976 | D | -28.599976 | D | 8.850036 | I |
|  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| D4 | OS4 | D5 | OS5 | D6 | OS6 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| 9.25 | I |  |  |  |  |
| 22.399964 | I | 25.849976 | I |  |  |
| 8.800048 | I | 28.700012 | I | 32.150024 | I |
| -9.549988 | D | 0.900024 | I | 20.799988 | I |
| 20.700012 | I | -3.849976 | D | 6.600036 | I |
| 14.25 | I | 30.849976 | I | 6.299988 | I |
| 52.099976 | I | 58.400024 | I | 75 | I |
| 40.799988 | I | 32.899964 | I | 39.200012 | I |
| 37.25 | I | 42.950012 | I | 35.049988 | I |
| 26.900024 | I | 37.049988 | I | 42.75 | I |
| -45.150024 | D | -1 | D | 9.149964 | I |
| -55.649964 | D | -74.849976 | D | -30.699952 | D |
| -61.600036 | D | -59.450012 | D | -78.650024 | D |
| -23.950012 | D | -24.150024 | D | -22 | D |
| 21.450012 | I | -6.450012 | D | -6.650024 | D |
| 5.049988 | I | -24.649964 | D | -52.549988 | D |
|  |  |  |  |  |  |

From the above data we have calculated the TPM and EPM for all the cases

Table 4.3.4



Next we have used the ‘hmmgenerate’ function to generate the sequences of emitted symbols and states. The results are shown below.

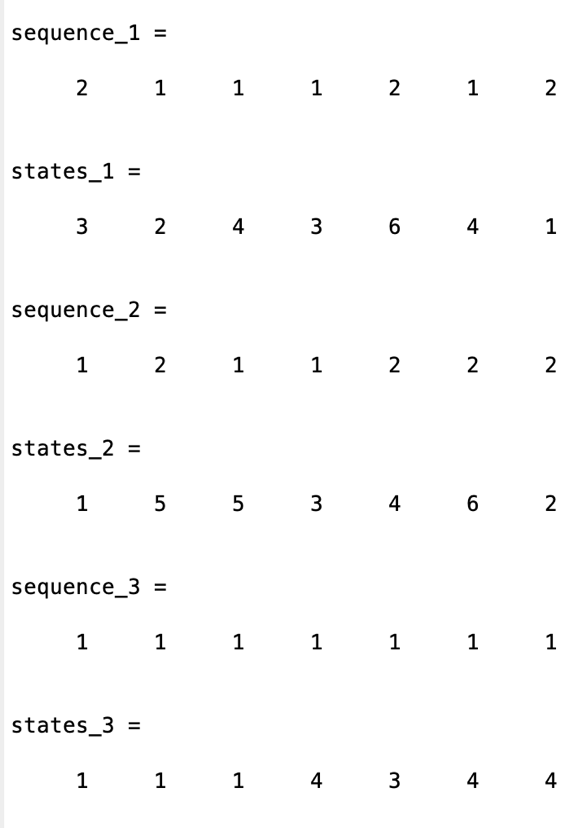
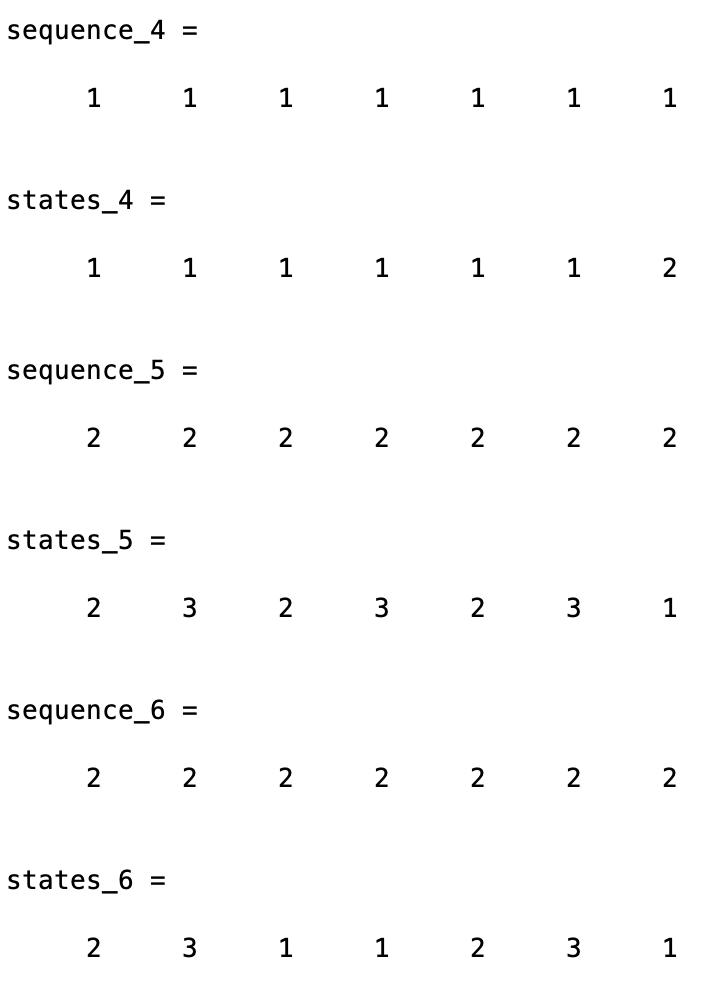
 

Figure 4.3.3: Results - sequence of observation symbols

We can use this sequence and find out the fitness value of all the cases.

4.4 Optimal Portfolios and the Efficient Frontier:

4.4.1 Optimal Portfolios:

An optimal Portfolio is one that occupies the efficient parts of the risk return premium spectrum. It satisfies the requirement that no collection exists with a higher expected return at the same standard deviation of the different combinations of assets produce different levels of return. The optimal Portfolio concept represent the best of these combinations, those that provide maximum possible expected return for a given level of acceptable risk.

The relationship between assets in an essential part of the optimal Portfolio theory. Some prices move in the same direction under similar circumstances, while others going opposite directions. The more out of sync these prices development are, the lower of covariance between two assets is, which translates into lower overall risk.

The optimal Portfolio does not focus on investment with either high expected returns or low risk. Its goal is to balance stocks carrying the best potential returns with acceptable risk. When we plot these, we get the Efficient frontier.

4.4.2 Efficient frontier:

The Efficient Frontier is the set of optimal Portfolios that offer the highest expected return for a defined level of risk or the lowest risk for a given level of expected return. portfolios that lie below the Efficient frontier are sub-optimal because they do not provide return for the level of risk. Portfolios that cluster to the right of efficient frontier are sub-optimal because they have a higher level of risk for that defined rate of return.

4.4.3 Understanding Efficient frontier:

The Efficient frontier rates Portfolios (investments)on a scale of return (y axis) versus risk(x axis) compound annual growth rate (CARG) of an investment is commonly used as the return component while standard Deviation (annualized)depicts the risk metric.

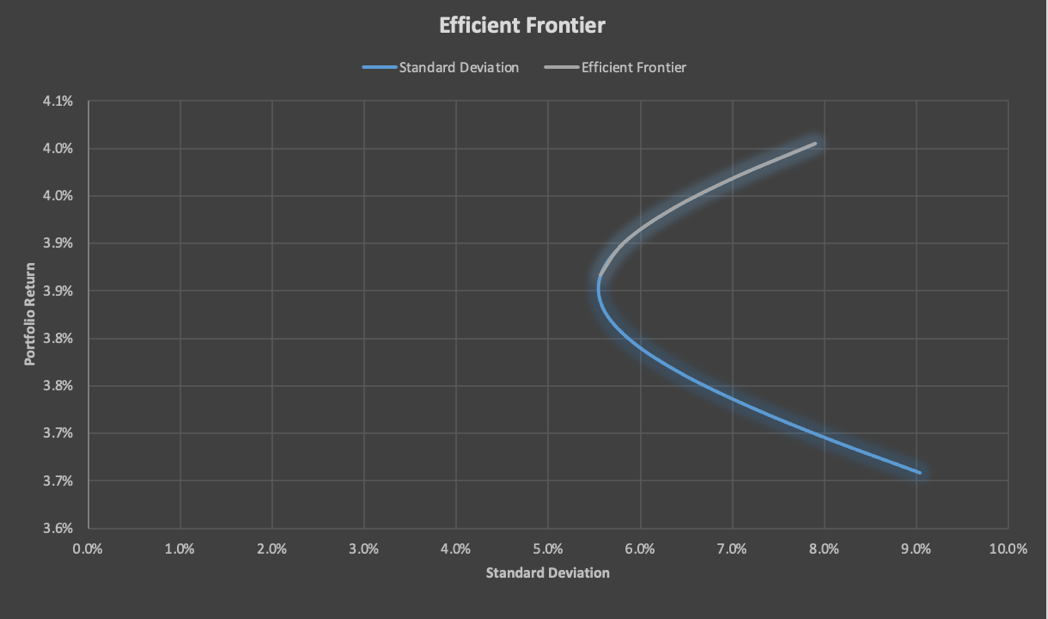


Fig 4.4.1 : Efficient frontier graph for HDFC and Reliance stock close prices from 01/04/20 to 12/03/21.

The Efficient Frontier graphically represent Portfolios that maximize returns for the risk assumed. Returns are dependent on the investment combination that make up the Portfolio. The standard Deviation of a security is synonymous with risk. Ideally, an investors seeks to populate the portfolio with securities offering expectational return but whose combined standard Deviation is lower than the standard deviations of the individual Securitate less synchronized the securities (lower covariance) the lower the standard deviation. If these mix of optimizing the return versus risk paradigm is successful then that portfolio should line up along the Efficient frontier line.

4.4.4 Assumption and limitations:

The theory behind the Efficient frontier relies heavily on same assumptions, not all of which represent reality. The underlying assumptions for the optimal Portfolio focused primarily on the investors.

* . We expect investors to be rational and all have access to the same information.
* They are all risk-averse and share the goal to maximize returns.
* No single investor can influence the market.
* . Investors base all decisions on the market on expected return on standard deviation as a measure of risk .

4.4.5 Plotting the Efficient Frontier:

To form the curve of the Efficient Frontier, we need to keep three main factors in consideration.

* . Expected return of portfolios.
* variance or standard deviation as a measure of the return variability (risk).
* The covariance of the assets in the portfolio.
* For a two assets portfolio, we can calculate the expected return as Er(p)=w\* Er(A)+w\*Er(A)

Where:

* Er(P/A/B)is the expected returns of the portfolio, asset A, and asset B.
* w[A/B]are the weights of assets A and B in the portfolio.

## CONCLUSION

Even though the stock market is random and unpredictable most of the time using random walk we can simulate paths among which few paths follow market trends perfectly and using those perfect paths we can forecast upcoming stock prices of a particular company.

We have also explored other techniques like Markov process in which we deal in probabilities and predict how probable is it to finish in a particular state. Extending this Markov process, we have explored the hidden Markov model where the states are not directly visible and the market movements are considered hidden. Here we have generated sequence of symbols through which we can calculate the fitness and conclude which path is the best.

Finally, the Efficient frontier help us identify and visualize a combination of assets with the optimal level of expected return of any given risk level. Portfolios on the curve are the most efficient. Other collections either have lower expected return for the same risk level or introduce higher risk levels of the same expected returns.

By plotting the efficient frontier, we can perform a more thorough analysis of our investment opportunity and make more informed decision.

1. **REFERENCES**

<https://blog.quantinsti.com/random-walk/>

Zhang, D., & Zhang, X. (2009). Study on forecasting the stock market trend based on stochastic analysis method. *International Journal of Business Management*, *4*(6), 163-170.

Sarsour, Wajeeh & Rijal, Shamsul. (2020). Forecasting the Long-Run Behavior of the Stock Price of Some Selected Companies in the Malaysian Construction Sector: A Markov Chain Approach.

Zhang, Yingjian. (2021). PREDICTION OF FINANCIAL TIME SERIES WITH HIDDEN MARKOV MODELS.

Kavitha, G. & Annamalai, Udhayakumar & Deivanayagampillai, Nagarajan. (2013). Stock Market Trend Analysis Using Hidden Markov Models.

V, Deneshkumar & Kaliyaperumal, Senthamarai & Niraikulathan, Sonai. (2020). Stock Market Trend Prediction Using Hidden Markov Model. 10.5772/intechopen.93988.

<https://magnimetrics.com/>

<https://in.finance.yahoo.com/quote/TECHM.NS?p=TECHM.NS&.tsrc=fin-srch>