Price Prediction and Trend Analysis of BSE Sensex using Markov Chain and Geometric Brownian Motion (GBM)

Project by:

Priyanka Khatwani - https://www.linkedin.com/in/priyanka-khatwani-a7a61516b/

Ekta Thakur - https://www.linkedin.com/in/ekta-thakur-372b0a212/

Chiranjeev Sharma - https://www.linkedin.com/in/chiranjeev-sharma-0286301ab/

Suraj Goundar - https://www.linkedin.com/in/suraj-goundar-169ba61ab/

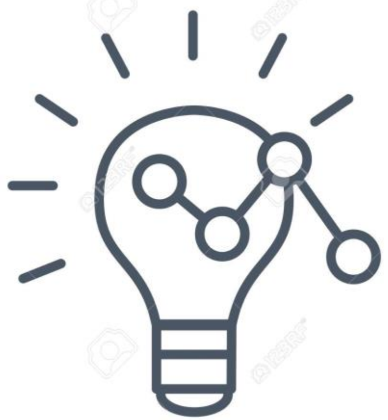
Chhavi Agarwal - https://www.linkedin.com/in/chhavi-agarwal-0a9009209/

Pratik Daga - https://www.linkedin.com/in/pratikdaga12/

Mentor: Dr. Leena Kulkarni

https://www.linkedin.com/in/dr-leena-kulkarni-09959426/

Why?



Generally, for retail investors, stock market is seen as more of a gamble than an informed decision. Studying the previous data and recognizing trends and patterns enabling us to make data-driven decision on investing is the main motive of our project.

With the help of statistical tools and methods available, we intend to make calculated predictions and analysis that can be useful to retail investors as well as giant investment organization in real world.

Objectives

- To analyse the long run behaviour of BSE Sensex using Markov Chain.
- To predict whether the index increases, decreases or remains constant for future trading days using Markov Chain.
- To construct a GBM model to forecast the future closing price of BSE Sensex and to check efficiency of the model.
- To determine the accurate model.

Methodology

Data Collection

- The data used in the project is available on https://www.bseindia.com/Indices/IndexArchiveData.html
- The data consists of date, opening price, closing price, day high and day low for each trading day for the period of January 2016 to September 2021.
- For Markov Chain, from Jan 2020 to Dec 2020 data is used.
- For GBM, from Jan 2016 to Sept 2020 data is used to create the model.

DataSet Information

1	Date	Open	High	Low	Close
1097	01-Jan-16	26101.5	26197.27	26008.2	26160.9
1098	04-Jan-16	26116.52	26116.52	25596.57	25623.35
1099	05-Jan-16	25744.7	25766.76	25513.75	25580.34
1100	06-Jan-16	25628.23	25632.57	25357.7	25406.33
1101	07-Jan-16	25224.7	25230.35	24825.7	24851.83
1102	08-Jan-16	24969.02	25083.55	24887.22	24934.33
1103	11-Jan-16	24787.11	24961.88	24598.9	24825.04
1104	12-Jan-16	24862.93	24882.3	24597.11	24682.03
1105	13-Jan-16	24804.64	24956.54	24387.69	24854.11
1106	14-Jan-16	24606.2	25018.46	24473.22	24772.97
1107	15-Jan-16	24881.76	24912.64	24421.53	24455.04
1108	18-Jan-16	24400.78	24524.85	24141.99	24188.37
1109	19-Jan-16	24257.28	24563.34	24247.23	24479.84
1110	20-Jan-16	24325.77	24325.77	23839.76	24062.04
1111	21-Jan-16	24194.75	24351.83	23862	23962.21
1112	22-Jan-16	24122.06	24472.88	24120.04	24435.66
1113	25-Jan-16	24540.97	24650.57	24433.67	24485.95
1114	27-Jan-16	24643.13	24645.7	24458.13	24492.39
1115	28-Jan-16	24481.86	24587.2	24400.52	24469.57
1116	29-Jan-16	24347.31	24911.9	24340.06	24870.69
1117	01-Feb-16	24982.22	25002.32	24788.58	24824.83
1118	02-Feb-16	24868.21	24928.75	24460.53	24539
1119	03-Feb-16	24393.59	24409.26	24187.54	24223.32
1120	04-Feb-16	24386.45	24514.01	24224.74	24338.43
1121	05-Feb-16	24360.36	24672.9	24345.79	24616.97
1122	08-Feb-16	24637.41	24698.95	24196.84	24287.42
1123	09-Feb-16	24076.85	24111.19	23919.47	24020.98

Markov Chain for Trend Analysis

What are Markov Chains?

- The sequence $\{X_n, n \geq 0\}$ is said to be a Markov chain if

$$\boxed{P\{ X_{n+1} = i_{n+1} / X_0 = i_0, X_1 = i_1, \ldots\ldots\ldots\ldots\ldots, X_n = i\} = P\{X_{n+1} = i_{n+1} / X_n = i_n\}}$$

for all state values $i_0, i_1, i_2, \ldots\ldots, i_n \in I$,where ‘ I ’ is the state space .

- This indicate that regardless of its history prior to time n, the probability that it will make a transition to another state j depends only on state ‘i’.

What is Order of Markov Chain?

- A Markov chain is said to be of order k if the following equation relating the conditional probabilities is satisfied.

'k' is the smallest integer such that

$$\boxed{P\{ X_{n+1} \mid X_{n}, X_{n-1}, X_{n-2}, \ldots\ldots\ldots\ldots\} = P\{X_{n+1} \mid X_{n}, X_{n-1}, X_{n-2}, \ldots X_{n-k}\}}$$

- In simpler words, order of Markov chain tells us the amount of memory Markov chain holds.

Terminology used

- In this project, Markov chain consists of 3 states {-1,0,1}

  Where,

  -1 : decrease

  0 : no change

  1 : increase

  in closing price of index with respect to previous day’s closing price. Similarly, price of index respectively.

- The changes in prices are identified using the interval (previous_closing_price ± 0.001*previous_closing_price)

Frequency Matrices

	-1	0	1
-1	68	10	74
0	13	1	20
1	71	23	113

First Order frequency matrix

		-1	0	1
1	1	24	5	49
-1	-1	18	0	21
1	-1	19	3	28
-1	1	24	6	24
0	0	0	0	1
0	1	2	1	5
0	-1	2	0	4
1	0	5	1	5
-1	0	1	0	2

Second order frequency matrix

Transition Probability Matrices

Second order transition probability matrix

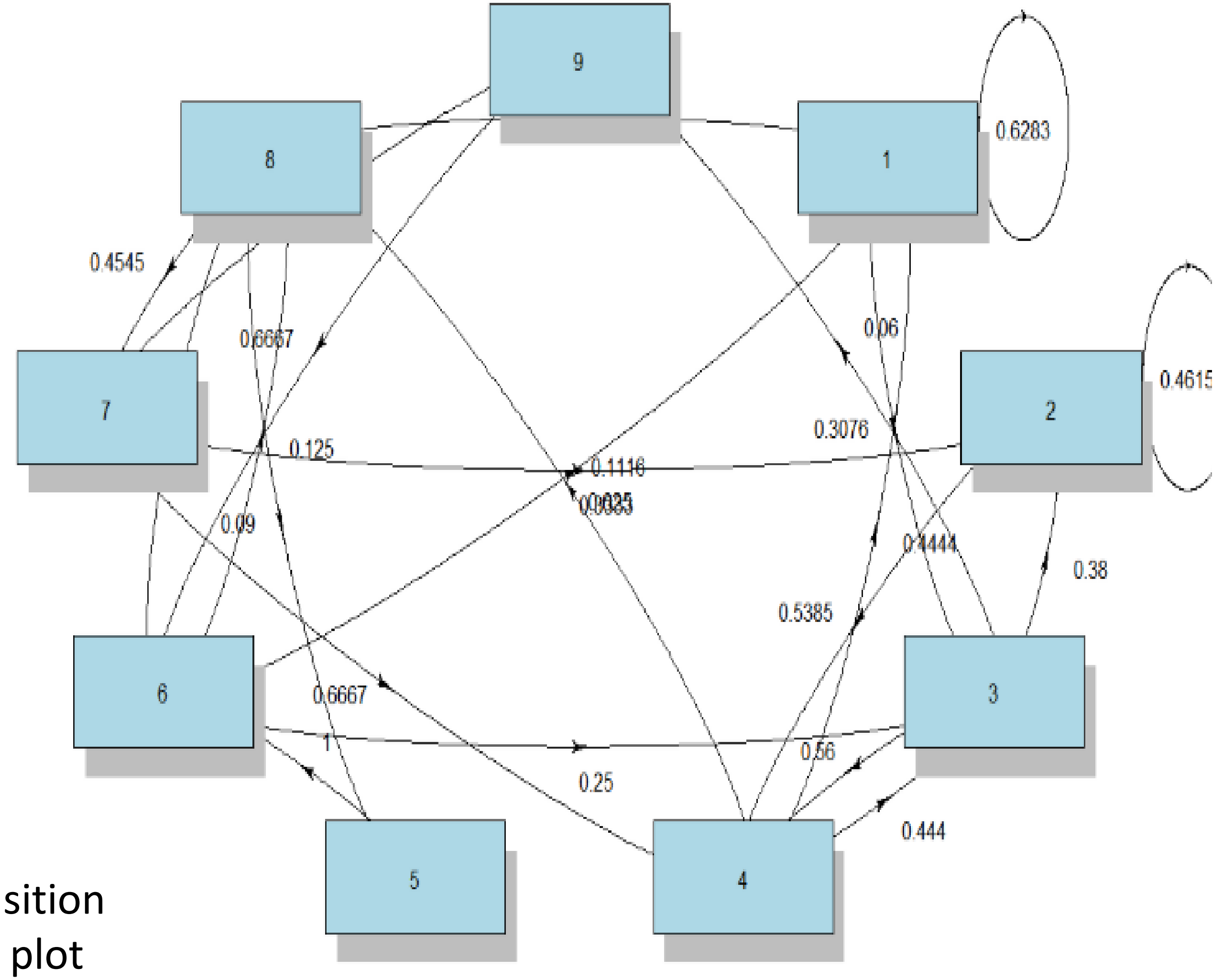
		(1,1)	(-1,-1)	(1,-1)	(-1,1)	(0,0)	(0,1)	(0,-1)	(1,0)	(-1,0)
1	1	0.628205	0	0.307692	0	0	0	0	0.064103	0
-1	-1	0	0.461538	0	0.538462	0	0	0	0	0
1	-1	0	0.38	0	0.56	0	0	0	0	0.06
-1	1	0.444444	0	0.444444	0	0	0	0	0.111111	0
0	0	0	0	0	0	0	1	0	0	0
0	1	0.625	0	0.25	0	0	0	0	0.125	0
0	-1	0	0.333333	0	0.666667	0	0	0	0	0
1	0	0	0	0	0	0.090909	0.454545	0.454545	0	0
-1	0	0	0	0	0	0	0.666667	0.333333	0	0

	-1	0	1
-1	0.447368	0.065789	0.486842
0	0.382353	0.029412	0.588235
1	0.342995	0.111111	0.545894

First order transition probability matrix

Visualization of Transition Probability Matrices

First order transition probability matrix plot



Second order transition probability matrix plot

Likelihood for the transition matrix

Let p_{ij} = P[X_{n+1}=j/X_n=i]. We find the likelihood function of such a matrix.

$\{X_i\}_{i=1}^{T}$ be the path of Markov chain. The conditional probability is given by

P$(X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_T = x_T)$ =P$(X_T|X_{T-1}, X_{T-2}, \ldots\ldots, X_1)$*P$(X_1, X_2 \ldots \ldots, X_{T-1})$

As it is a Markov chain it will follow the Markov property.

=P$(X_T|X_{T-1})P(X_1, X_2 \ldots \ldots, X_{T-1})$ $----------\rightarrow I$

= P$(X_T|X_{T-1})$ P$(X_{T-1}|X_{T-2})P(X_1, X_2 \ldots \ldots, X_{T-2})$ $----------\rightarrow II$

Similarly by repeating this step n-times we'll put all the values of equation in eq I,

P$(X_1, X_2 \ldots \ldots, X_T)$ = L $=\prod_{i=2}^{T} P(X_i|X_{i-1}) * P(X_1)$

Taking log on both sides

Log (L)= $\sum_{\mathrm{i}=2}^{\mathrm{T}}$ Log($P(X_i|X_{i-1})$) +log($P(X_1)$)

Finding the maximum log likelihood function by differentiating

L' = $\sum_{\mathrm{i}=2}^{\mathrm{T}}$ Log($P(X_i|X_{i-1})$)

Define the transition counts n_{ij} ≡ number of times i is followed by j in $\{X_i\}_{i=1}^{T}$, and re-write the likelihood in terms of the

L' $=\sum_{\mathrm{i},j} n_{ij}$ Log(P_{ij})

Testing for optimum order of Markov Chain

H_0: Markov Chain is of order k *v/s* H_1: Markov chain is of order k+1

The test is carried out by using Akaike's Information Criterion.

AIC for k^{th} order Markov chain is calculated using

$$\boxed{AIC(k) = \eta_{(k,\ m)} - 2(s^m - s^k)\ (s - 1)}$$

Where s: total number of states

k: order under null hypothesis

m: order under alternate hypothesis

$\eta_{(k,\ k+1)}$: $2 \ln \lambda_{k+m}$

$$\lambda_{(k+m)} = \frac{M_k(X_1,\ldots\ldots,Xn)}{M_m(X_1,\ldots\ldots,Xn)}$$

The optimum order is the one for which the value of AIC is minimum

In our project, AIC(1) = -21.127987

AIC(2) = -75.041412

AIC(3) = -11.743489

Therefore, We conclude that the order of the Markov chain is 2.

Stationary Distribution

- Stationary distribution is the probability distribution of the states with the property that they no longer change as time progresses. In other words, as time progresses, the probability of achieving a particular state remains constant in each transition and this probability no longer depends on any past transitions.
- For our first order Markov chain, we found that the chain became stationary after 5 transitions.
- The stationary distribution is given below:

-1	0	1
0.3784662	0.06388455	0.5576493

- Interpretation: In the long run we will observe that the state -1 occurs 37.85% of times, state 0 occurs 6.39% of times and state 1 occurs 55.8% of times.

Prediction of future states

- With the help of random number generator, we were able to predict the states for each trading day from 1st January 2021 to 30th July 2021 after creating the second order frequency matrix using the data from 1st January 2020 to 31st December 2020

State	No. of correct predictions	No. of Wrong predictions
-1	21	36
0	1	18
1	26	41
Total	48	95

	-1	0	1
Actual	57	19	67
Forecasted	62	26	55

- Due to large error, we conclude that Markov chain model is not a good fit to predict the increase or decrease of closing price of index.

Geometric Brownian motion (GBM) for Price Prediction of BSE SENSEX

Why Geometric Brownian Motion?

When we talk about prices, assets or any such financial variable, the economists prefer using geometric Brownian motion over Brownian motion because

- The value of the variables are non-negative.
- They exhibit random fluctuations in its path as we observe in real stock prices.
- These properties are possessed by GBM and not BM.

What is Geometric Brownian Motion?

A stochastic process $\{Z(t) : t \geq 0\}$ is called a geometric Brownian motion with drift parameter μ .Let $X(t)$ be a Brownian motion with drift μ and variance parameter σ^2. Then, process $\{Z(t) : t \geq 0\}$ is s.t.b is a GBM with drift μ and variance .

$$Z(t)=e^{X(t)} \quad \Rightarrow \quad \log z(t) = X(t)$$

The GBM Model is:

$$\boxed{Z(t) = ze^{(\alpha - \frac{1}{2}\sigma^2)\, t + \sigma X(t)}}$$

Where,

X(t): Brownian Motion at time t

α: Mean of Average Returns

σ: Variance of Average Returns

Z: Initial Closing Stock Prices

(α – 0.5 *σ2): Drift

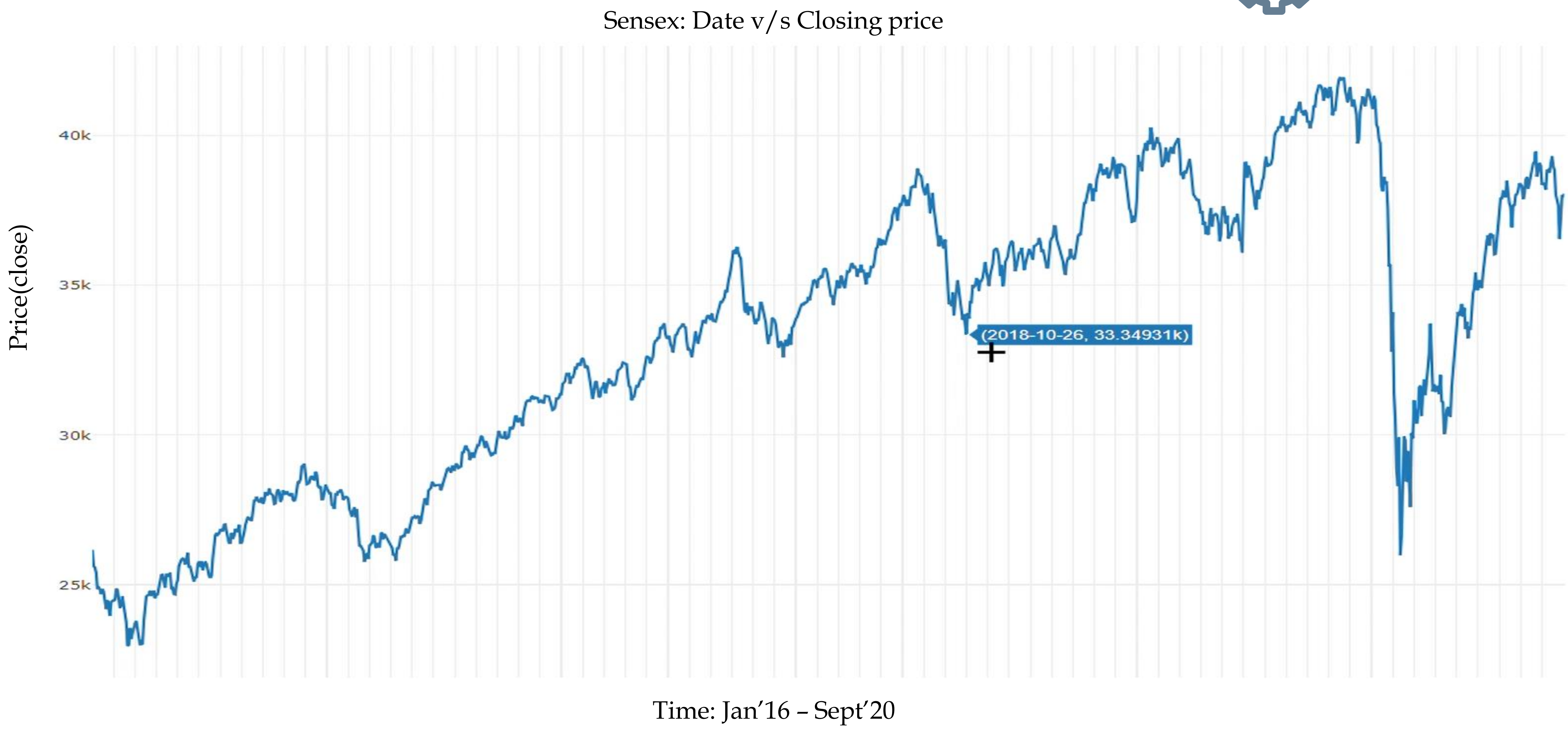
σ X(t): Volatility

Terminology used

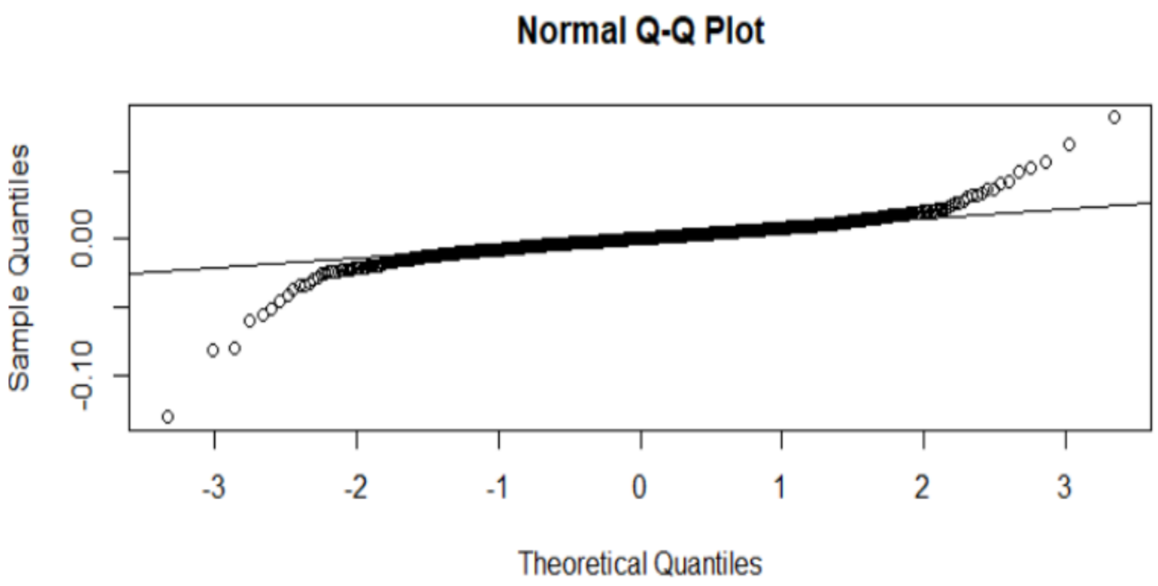
- **Drift** – reflects trend or growth rate. It is denoted by ‘μ’. If the drift is positive, the trend is going up over time. If the drift is negative, the trend is going down.
- **Volatility** - reflects variation or spread of the distribution. It is denoted by σ^2. The value of volatility is always positive (or zero) because it is actually related to standard deviation of the distribution.
- **Returns** – In this project, it is calculated as $\frac{\text{Increment in price}}{\text{Previous day's price}}$

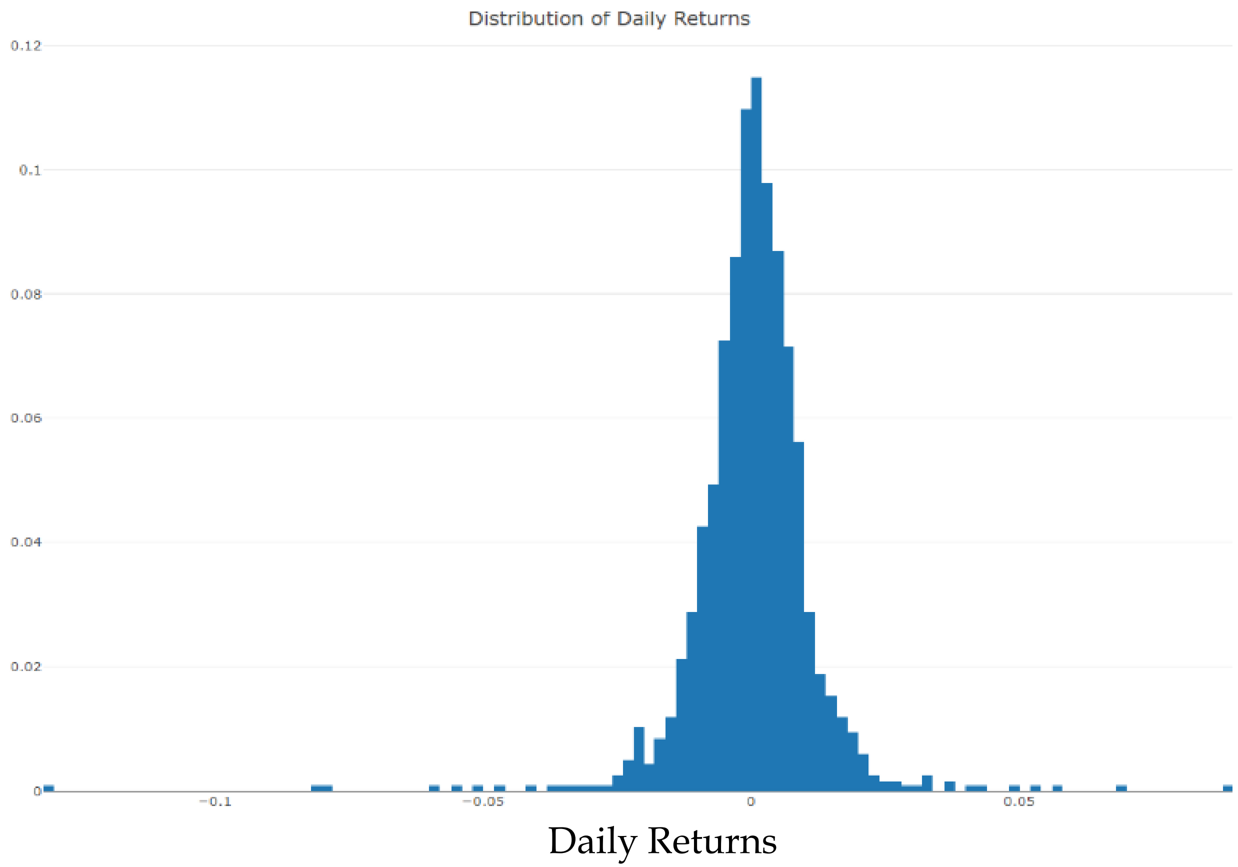
The returns are log-normally distributed.

Visualization of Historical data



Verifying normality assumption for distribution of daily returns

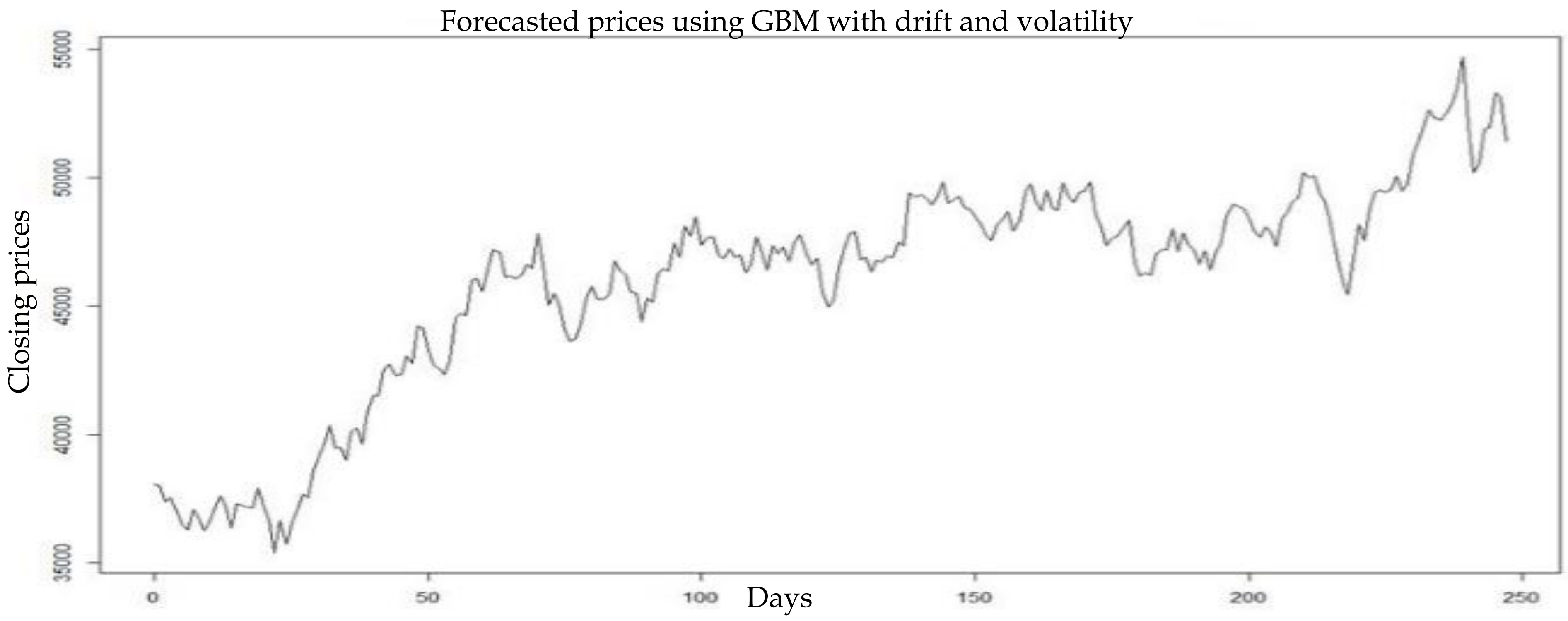




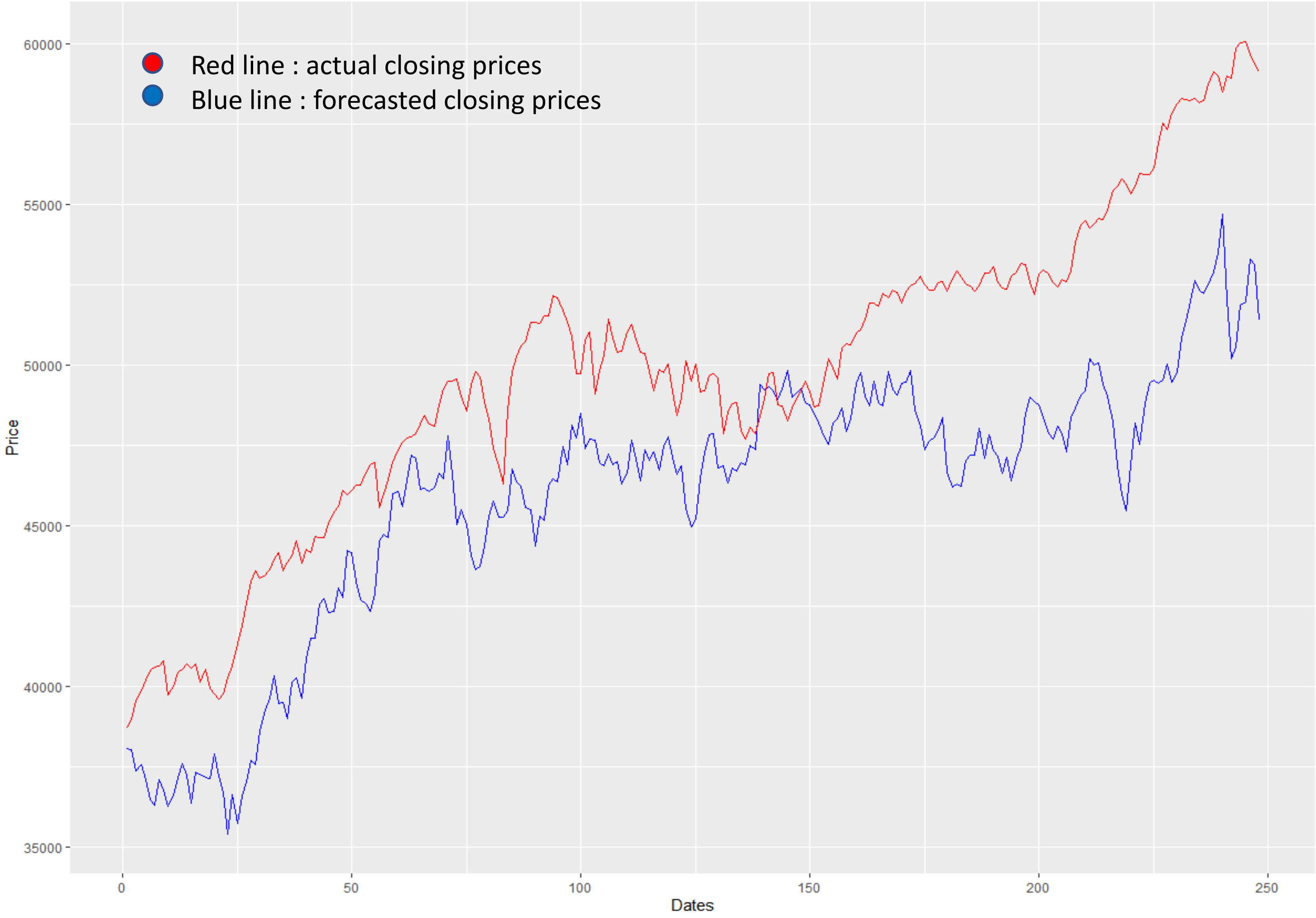
Forecasting index prices using GBM model

- No. of days forecasted : 247 trading days

Estimation	Value
Drift	0.0007765858
Volatility	0.01658694



Comparison between Actual closing prices and forecasted closing prices using GBM



- Visually we can see that the forecasted values follow more or less the same trend as the actual values.
- The Correlation coefficient between the actual values and the forecasted values has a positive correlation with $\rho = \mathbf{0.90051}$ which represents a strong relationship between them.

Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error (MAPE) is a measure of prediction accuracy of a forecasting method in statistics. It usually expresses the accuracy as a ratio defined by the formula:

$$\boldsymbol{MAPE} = \frac{1}{N}\sum_{k=1}^{N}\left|\frac{A_t - F_t}{A_t}\right| * 100$$

where ,

A_t : actual value

F_t : forecast value

Accuracy measurement	Value
MAPE	**7.61%**

According to the scale of judgement of forecast accuracy, a MAPE <10% is said to be highly accurate forecast.

Due to random behavior of stock market, GBM model is highly suitable for short term forecasting.

Conclusion

Method	Error
Markov Chain	**64.2%**
Geometric Brownian Motion (GBM)	**7.61%**

- Therefore, it is evident that GBM model is much better fit for predicting closing prices of BSE Sensex.
- As a form of recommendation, the predictive power of Geometric Brownian Motion model should be used to forecast daily stock prices over short period as it gives a highly accurate result.

Limitations of GBM

- Volatility changes over time, but in GBM it is assumed to be constant .
- Stock prices often show jumps caused by unpredictable events but in GBM the path is continuous.

Future Scope of GBM

- In order to make GBM more realistic/practical for modelling stock prices we can drop the assumption that volatility is constant.
- We can create a jump diffusion model, to predict stock price behavior exhibiting jumps.

References

1. K.Rahul,Dr. Bidhyadhara Bishi (2020) - Forecasting Short Term Return Distribution of S&P BSE Stock Index Using Geometric Brownian Motion: An Evidence from Bombay Stock Exchange, *International Journal of Statistics and Systems ISSN 0973-2675 Volume 15, Number 1 , pp. 29-45*
2. Madhav Kumar Bhusal (Oct 2017) - Application of Markov Chain Model in the Stock Market Trend Analysis of Nepal, *International Journal of Scientific & Engineering Research Volume 8, Issue 10, ISSN 2229-5518*
3. H. Tong (Sept 1975) - Determination of the Order of a Markov Chain by Akaike's Information Criterion, *Journal of Applied Probability , Vol. 12, No. 3, pp. 488-497*
4. Richard W. Katz (Aug 1981) - On Some Criteria for Estimating the Order of a Markov Chain, *TECHNOMETRICS , VOL. 23, NO. 3*

Thank you