Stochastic Modelling of Stock Price Data

Soumyajit Roy

## Introduction

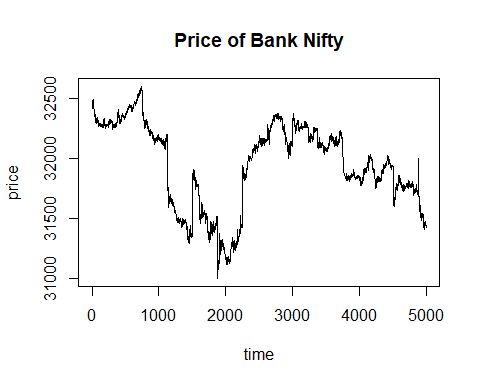
Stock Price Modelling is a very well celebrated as well as challenging job for the people in industry as well as in academics. There are several approaches. In this paper we will implement stochastic modelling based on mainly a dataset of Bank NIFTY starting from 1st Jan,2020 and in each interval of 1 min. We have 5000 observations there. For the entire study, we will try to model Open Price throughout.

## Data

Lets take a look over the data,

'data.frame': 5000 obs. of 8 variables:  
 $ DATE : Date, format: "2020-01-01" "2020-01-01" ...  
 $ TIME : POSIXct, format: "2024-11-07 09:15:00" "2024-11-07 09:16:00" ...  
 $ OPEN : num 32417 32457 32446 32445 32461 ...  
 $ HIGH : num 32465 32466 32452 32464 32474 ...  
 $ LOW : num 32354 32445 32445 32435 32457 ...  
 $ CLOSE : num 32456 32445 32445 32464 32465 ...  
 $ VOLUME: int 36240 18780 14900 18120 11540 9640 8480 8140 4840 10680 ...  
 $ OI : int 1304980 1304980 1316960 1316960 1316960 1328860 1328860 1328860 1337200 1337200 ...

The scatter plot of the data ,

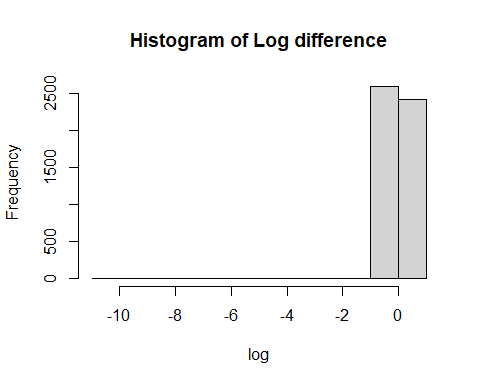


## Modelling

1. **Geometric Brownian Motion:**

* The SDE is,
* , where ~Weiner Process.
* **Diagnostic**:
* Now our stock data to follow the GBM, we must have ,  
    
  \$ logS\_t-logS\\_{t-1}\$ follows normal distribution.
* **Histogram**:

From Histogram it is evident that the data is not coming from a Normal population . Still we fit a GBM to see how it work.

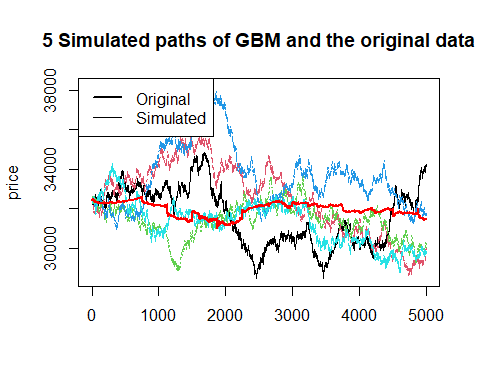


* **Model fitting:**
* **Method of Moment estimates:**

Here we estimated the parametres of GBM using MOM estimates. We found,

Parametre | Estimate | |————————–|————–| | | -0.002083886 | | | 0.14688 |

: Parametre

* **Fit:**
* 
* **Goodness of fit test:**

Here we are calculating the Chi- Square statistic based on a simulated path.

::: {.cell}  
::: {.cell-output .cell-output-stdout}  
```  
[1] 472276.6  
```  
:::  
  
::: {.cell-output .cell-output-stdout}  
```  
P-Value: 0  
```  
:::  
:::

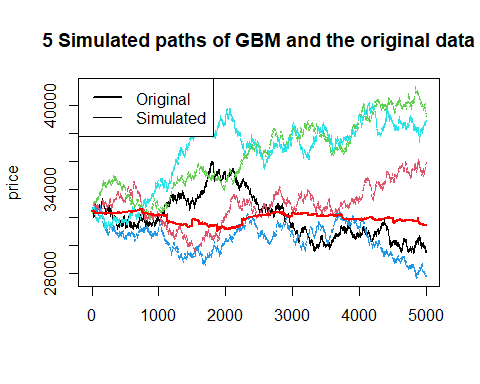
Here the p-value for the test is coming as 0. That means the data is not fitting well in this case.

* **Maximum Likelihood Estimates:**

Similarly as above but using MLE estimates of the GBM we fit it with our data. The estimates are,

Parametre | Estimates | |————————-|————–| | | -0.002083886 | | | 0.14687 |

* **Fit:**



* **Goodness of fit** **:**

Similarly as above,

[1] 812843.3

P-Value: 0

The p-value is coming to be zero. Hence the fit is not good which is also evident from the plot itself.

1. **Jump Diffusion Model:**

* From the plot of the data , we can see that there is several jumps in the data. So for modelling this kind of data, Jump Diffusion Model is very famous in literature.
* Here the SDE is,
* , where