

## Statistical Properties of Stock Prices

**Company:** Ultratech Cement Limited

Investors and traders make buying and selling decisions of stocks based on the current and past data. We will follow the technical analysis approach where we'll be using the historical data. Here, we'll be checking whether the data is well-modeled by a normal distribution. The reason behind doing this is that many statistical tests have an underlying assumption of data being normally distributed. Hence, before applying statistical tests and models on the data, one should have an idea about their data's distribution. Along with the actual raw prices, we will also analyze the returns because it gives us scale-free assessment of the data and has more attractive statistical properties. The normality of the log-returns for the price of the stocks is one of the most important assumptions in mathematical finance.

**Process of data analysis (used python):**

### (I) Data Preparation:

1. Get the data : The data is taken from Yahoo! Finance, it has daily stock price data of Ultratech Cement Limited company for the period of one year from October 1, 2019 to September 30, 2020.
  2. Import the required libraries
  3. Load the data into a pandas DataFrame
  4. Preprocess the data
    - Extract only the columns that is required for the analysis. Here, 'Adj Close' and 'Date' are the desired column.
    - Detect and treat the null values
  5. Calculate simple returns and log returns of holding an asset for one period
- Simple returns :  $R_t = (P_t / P_{t-1}) - 1$
- Log returns :  $R_t = \ln(P_t / P_{t-1})$

### (II) Data Analysis:

1. Summarize the data using functions like describe, skew and kurtosis
2. Time series plot
  - Calculate the trend line using polyfit function(degree 1) which can be called from the numpy library
3. Frequency plot
  - Created 4 subplots with different number of bins.
  - Calculated the number of bins using the Sturge's rule for the first plot
  - The values for the number of bins for the other 3 plots are chosen randomly.
4. Test the normality:
  - Using graphs (histogram, q-q plot and box plot)

- Using statistical tests (Jarque-Bera Test, Shapiro-Wilk Test, Kolmogorov-Smirnov test and Lilliefors Test)

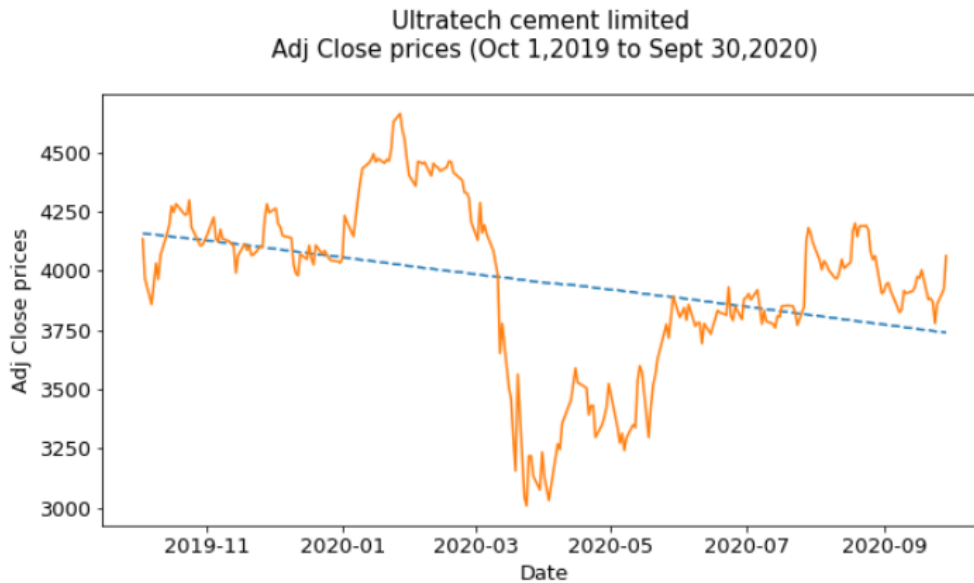
Results of the data analysis:

a. Descriptive Statistics

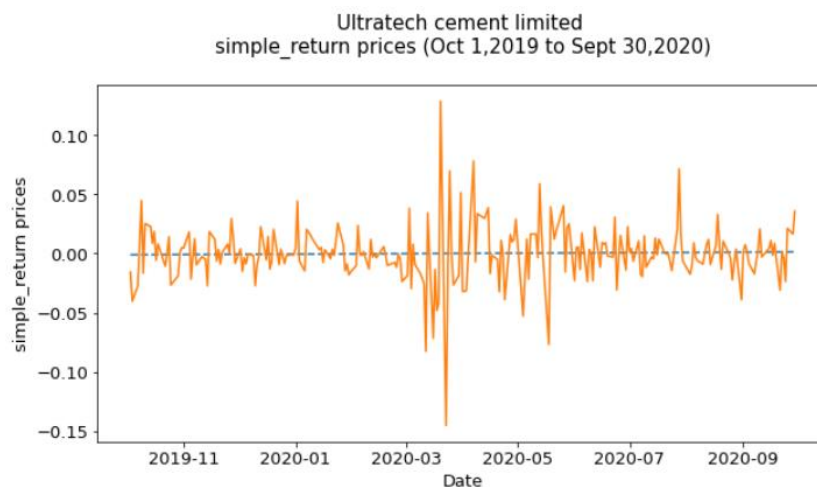
	<b>Adj Close</b>	<b>simple_return</b>	<b>log_return</b>
<b>count</b>	247.000000	246.000000	246.000000
<b>mean</b>	3949.255639	0.000167	-0.000135
<b>std</b>	353.241746	0.024548	0.024697
<b>min</b>	3008.614502	-0.145222	-0.156913
<b>25%</b>	3788.480713	-0.010162	-0.010214
<b>50%</b>	4004.250000	-0.000578	-0.000578
<b>75%</b>	4177.594726	0.011907	0.011837
<b>max</b>	4662.250977	0.128804	0.121159

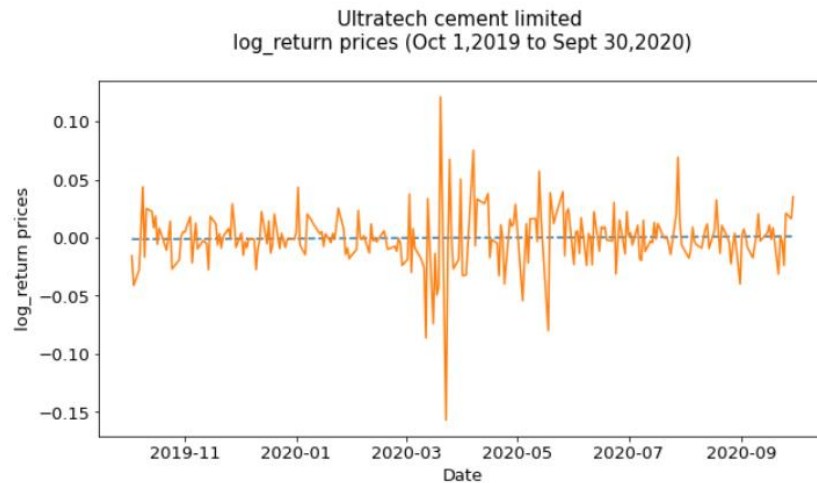
	<b>Skewness</b>	<b>Kurtosis</b>
<b>Adj Close</b>	-0.570128	-0.068119
<b>simple_return</b>	-0.254471	8.515679
<b>log_return</b>	-0.646713	9.429893

b. Time Series Plot:



The above graph is a time series plot of the Adjusted Close prices of Ultratech Cement Limited company for the period of one year from October 1, 2019 to September 30, 2020. The impact of the Coronavirus Crash that began on 20 February 2020, and ended on 7 April is clearly distinguished in the graph. Hence, we have a general decreasing trend over time ,though the company is trying to recover.



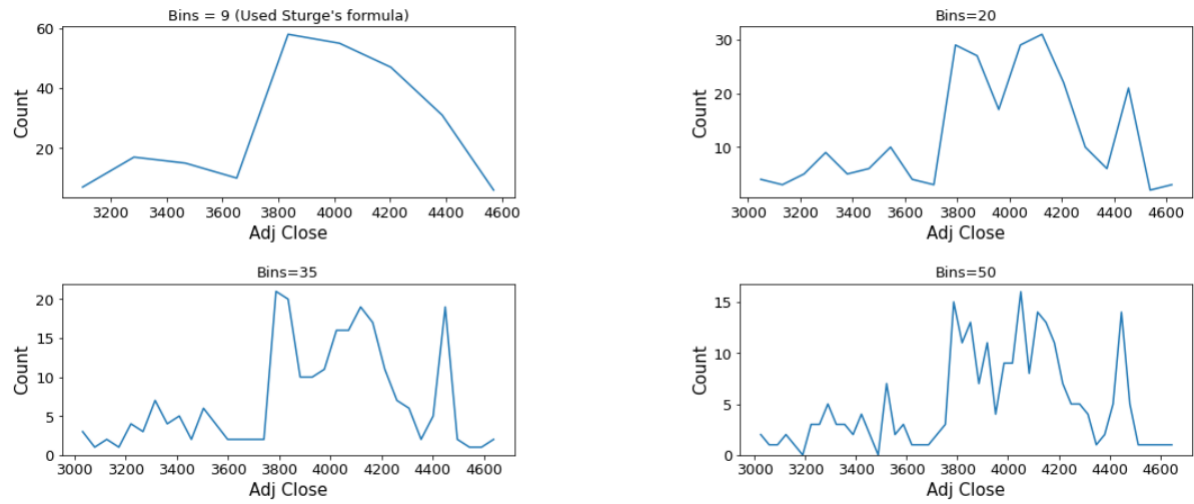


The above two graphs are very close to each other at each time point. We can see that the mean is almost zero, but there are significant fluctuations i.e. the returns are quite volatile.

### c. Frequency Plots

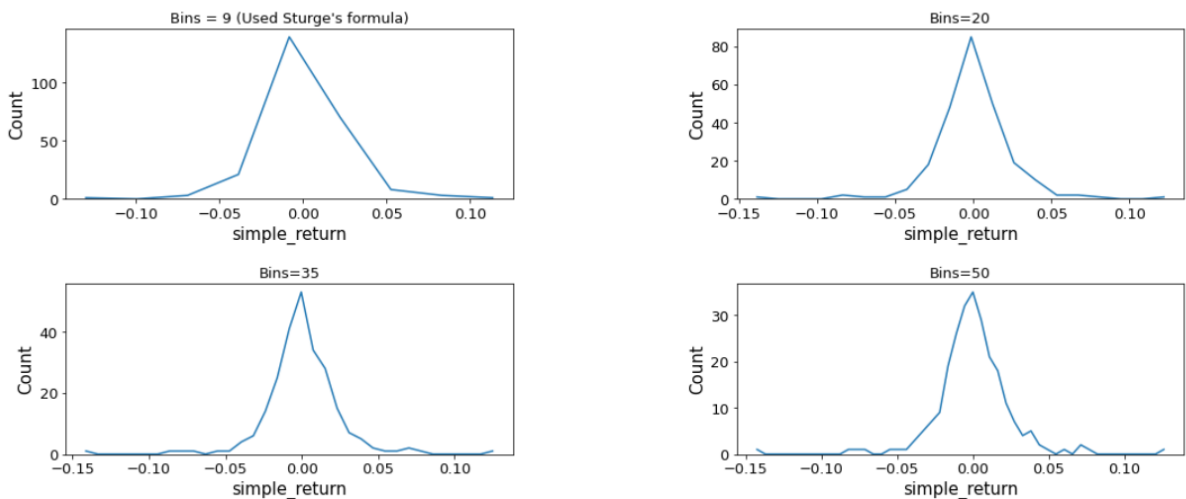
(Tried different bin sizes which helps us to reveal different features of the data)

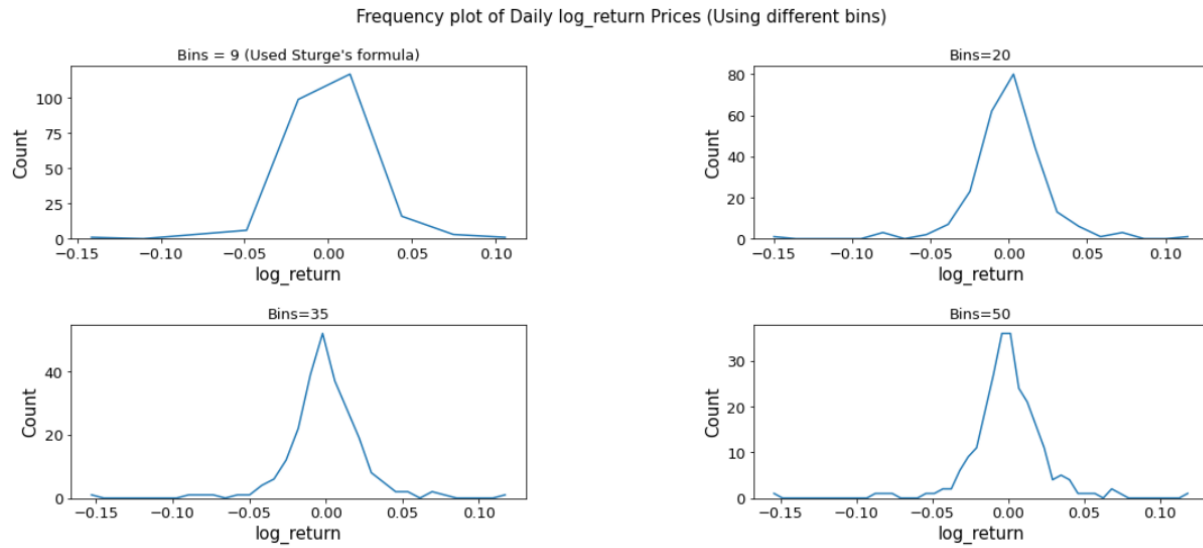
Frequency plot of Daily Adj Close Prices (Using different bins)



By looking at the graphs we get an intuition that the frequency plots of the Adjusted Close prices do not form a bell curve. Also, most of the prices are between 3800 and 4500. The data is skewed towards left.

Frequency plot of Daily simple\_return Prices (Using different bins)

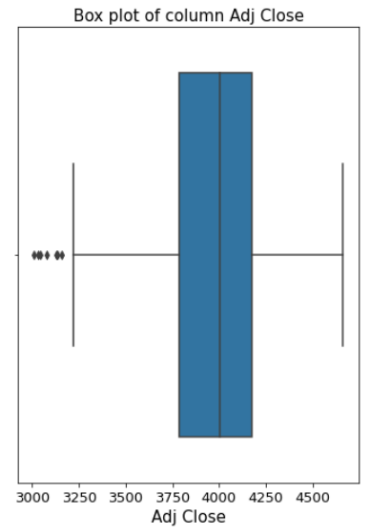
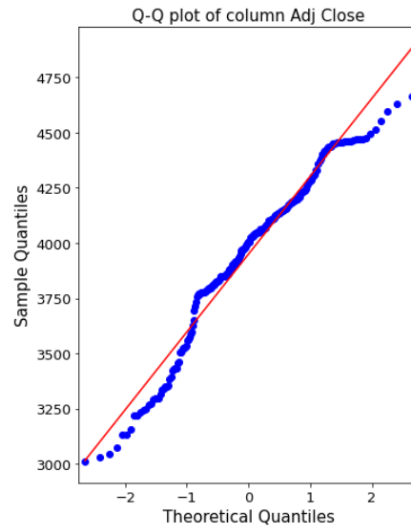
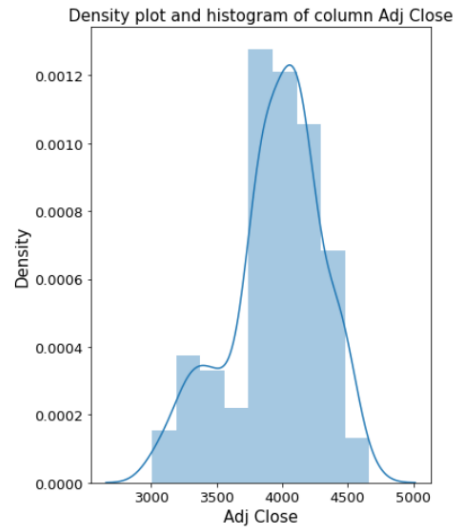




Frequency plot for the both the simple returns and log returns have heavier tails than a normal distribution. Most financial returns are leptokurtic.

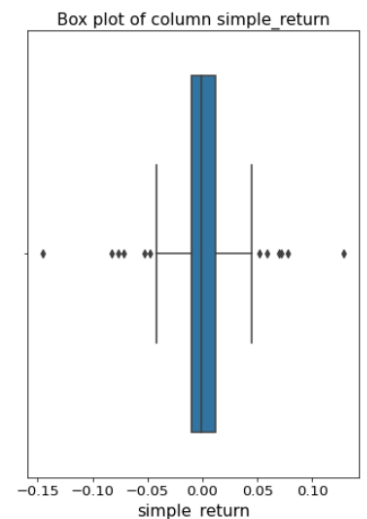
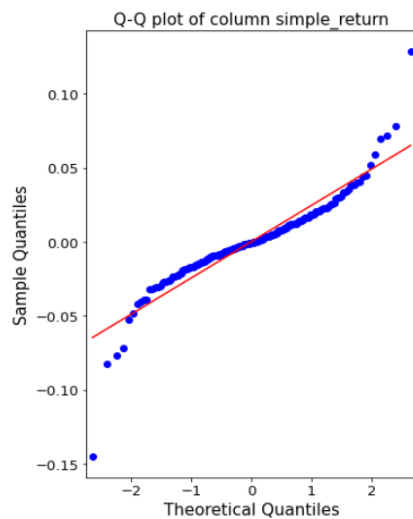
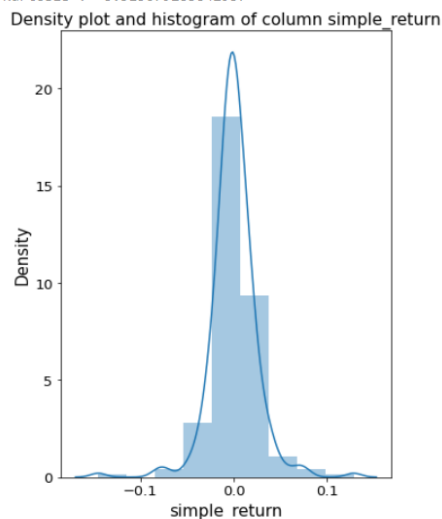
#### d. Plots for normality tests

Skewness : -0.5701283349544228  
Kurtosis : -0.06811853766815101



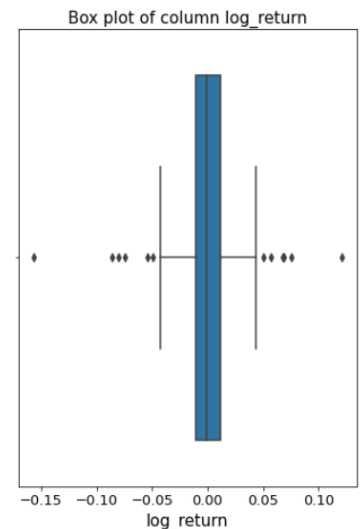
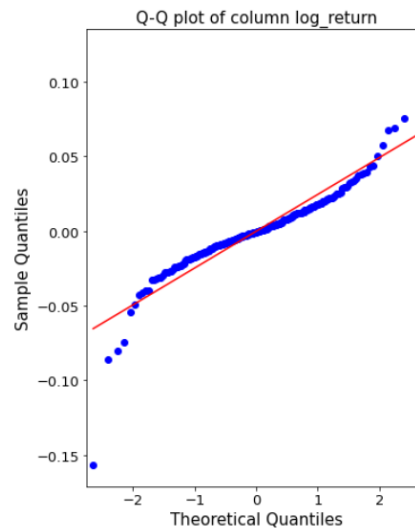
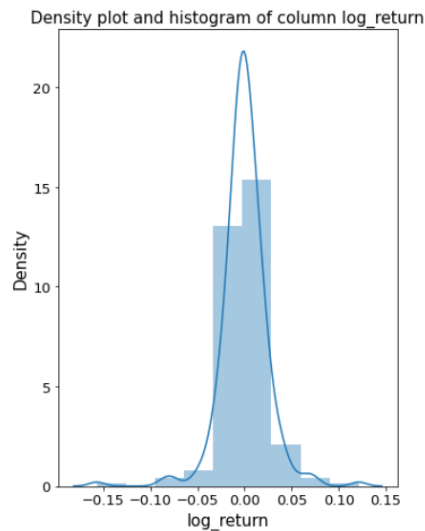
- We get a negative skewed value, which is not very far away from 0. Data is moderately skewed.
- Kurtosis is less than 3, so this is Platykurtic distribution(lighter tails than a normal distribution)
- In the first plot, the curve shows the density plot which is essentially a smooth version of the histogram. The curve doesn't really look like a bell curve.
- In the second graph, we can say that many observations are deviating from the straight line. The points form a curve rather than a straight line, which usually is an indication of skewness in the data.
- In the third graph, outliers are clearly visible. The median cuts the box into two unequal pieces, the longer part is to the left or below the median, hence, the data is skewed left.

Skewness : -0.25447137727537267  
Kurtosis : 8.515679283842937



- Data is moderately skewed, we get a negative skewed value which is not very far away from 0.
- Kurtosis is greater than 3, so this is Leptokurtic distribution(heavier tails than a normal distribution).
- In the first graph, the peak of the curve is high. More of the data is concentrated towards the mean.
- In the second graph, the points form a curve that deviates markedly from a straight line. Possible outliers are points at the ends of the line, distanced from the bulk of the observations.
- In the third graph, outliers are clearly visible on the both sides. Also, few of them are very extreme.

Skewness : -0.6467128261272341  
Kurtosis : 9.42989287261117



- The data has negative skewed value. Kurtosis is greater than 3, so this is Leptokurtic distribution(heavier tails than a normal distribution).
- In the first graph, the peak of the curve is high. More of the data is concentrated towards the mean.
- In the second graph, the points form a curve that deviates markedly from a straight line. Possible outliers are points at the ends of the line, distanced from the bulk of the observations.
- In the third graph, outliers are clearly visible on the both sides. Also, few of them are very extreme.

#### e. Test of significance



**Null hypothesis ( $H_0$ ):** The data is normally distributed.

**Alternate hypothesis ( $H_1$ ):** The data is not normally distributed.

**Jarque–Bera test:**

The Jarque-Bera test tests whether the sample data has the skewness and kurtosis matching a normal distribution.

**Shapiro-Wilk test:**

The Shapiro-Wilk test is a way to tell if a random sample comes from a normal distribution. This test may be suitable for smaller samples of data.

**Kolmogorov-Smirnov test:**

Kolmogorov-Smirnov test performs the one sample or two samples tests for goodness of fit. The one-sample test performs a test of the distribution  $F(x)$  of an observed random variable against a given distribution  $G(x)$ .

**Lilliefors test:**

The Lilliefors test is a normality test based on the Kolmogorov-Smirnov test. As all the above methods, this test is used to check if the data comes from a normal distribution.

**For all the above tests:**

If the  $p\text{-value} \leq 0.05$ , then we reject the null hypothesis i.e. we assume the distribution of our variable is not normal.

If the  $p\text{-value} > 0.05$ , then we fail to reject the null hypothesis i.e. we assume the distribution of our variable is normal.

--Normality tests for the data in the column Adj Close--

----Jarque-Bera Test-----

Test-statistic : 13.30388524293268

p-value : 0.00129151075044176

Reject the null hypothesis.

----Shapiro-Wilk Test-----

Test-statistic : 0.9628831148147583

p-value : 5.190190222492674e-06

Reject the null hypothesis.

----Kolmogorov-Smirnov Goodness of Fit-----

Test-statistic : 1.0

p-value : 0.0

Reject the null hypothesis.

----Lilliefors Test-----

Test-statistic : 0.09984840195904779

p-value : 3.0051799820451355e-06

Reject the null hypothesis.

We reject the null hypothesis. Thus, we have sufficient evidence to say that the data in the column 'Adj Close' has skewness and kurtosis that is significantly different from a normal distribution.

--Normality tests for the data in the column simple\_return--

----Jarque-Bera Test-----

Test-statistic : 712.0306543766475

p-value : 0.0

Reject the null hypothesis.

----Shapiro-Wilk Test-----

Test-statistic : 0.8914567828178406

p-value : 2.7014738102765223e-12

Reject the null hypothesis.

----Kolmogorov-Smirnov Goodness of Fit-----

Test-statistic : 0.4647633392450605

p-value : 3.4112218659096916e-49

Reject the null hypothesis.

----Lilliefors Test-----

Test-statistic : 0.09271120547527775

p-value : 2.4939661362494747e-05

Reject the null hypothesis.

---

--Normality tests for the data in the column log\_return--

----Jarque-Bera Test-----

Test-statistic : 887.3348387513382

p-value : 0.0

Reject the null hypothesis.

----Shapiro-Wilk Test-----

Test-statistic : 0.8848373293876648

p-value : 1.0166122501112373e-12

Reject the null hypothesis.

----Kolmogorov-Smirnov Goodness of Fit-----

Test-statistic : 0.46592010129989614

p-value : 1.8870022740268145e-49

Reject the null hypothesis.

----Lilliefors Test-----

Test-statistic : 0.09574798316815436

p-value : 1.0624886914303124e-05

Reject the null hypothesis.

We reject the null hypothesis. Thus, we have sufficient evidence to say that the data in the columns 'Simple return' and 'Log returns' have skewness and kurtosis that is significantly different from a normal distribution.

**Conclusion :**

The impact of the Coronavirus Crash that began on 20 February 2020, and ended on 7 April is clearly distinguished in the time series plot. The stock prices are negatively skewed and platykurtic in nature, but most of the prices are between 3800 and 4500. Returns are moderately skewed and leptokurtic in nature. More of the returns data are concentrated towards the mean. Returns have comparatively more extreme outlier values when compared with the prices. Graphical method helps us to qualify that the skewness and kurtosis in the prices and the returns are significantly different from a normal distribution. Failure of all the Statistical normality tests used, gives us enough evidence and helps us to quantify that both the prices and the returns are not normal.