

**COURSE: Random Process and Estimation Techniques (B.Tech AI)**

**Name:** Maitri Shah

**Year:** 2021

**Program:** B.Tech Artificial Intelligence

## **Stock Analysis**

**Software/IDE/Interactive notebook:** Google Colaboratory

### **Algorithm and Program:**

Project Link:

<https://colab.research.google.com/drive/1lqS5BWPg21Hjcs1FDQhmCOnqU1ytvvKT?usp=sharing>

Stock Analysis for 3 different stock indices – FTSE, BSE SENSEX and DJIA.

The Financial Times Stock Exchange is index of the companies listed on the London Stock Exchange. The BSE SENSEX is a market-weighted stock listed on the Bombay Stock Exchange. The Dow Jones Industrial Average is index of companies listed on stock exchanges in the US.

We used Matplotlib, Pandas and NumPy library for all the operations on the dataset. Using the yfinance library we imported the datasets and used data on the 3 indices from Yahoo Finance. mpl\_finance was used to call candlesticks\_ohlc specially to show the profit and loss i.e., returns for each index.

Since the stock prices of different countries are in different currencies, we had to import currency-exchange datasets to convert GBP and INR to USD for FTSE and BSE SENSEX respectively. This was needed as we were plotting the values together and, in this case, different currencies would end up generating an erroneous graph.

Pandas head() method is used to return top n (5 by default) rows of a data frame or series.

Then graphs of Stock opening and Volume Traded are plotted from 2003 to 2021. From the graph it can be clearly inferred that stock price for BSE SENSEX is far much lower as compared to FTSE and DJIA. This is due to the low opening and closing price of BSE SENSEX as compared to the other two and the difference in value of the Indian Rupee compared to the British Pound or the US Dollar

**COURSE: Random Process and Estimation Techniques (B.Tech AI)**

Pandas `iloc()` function enables us to select a particular cell of the dataset, that is, it helps us select a value that belongs to a particular row or column from a set of values of a data frame or dataset whereas `argmax()` function returns the indices of the maximum value present in the input index.

To find the total trade for the particular date, volume and opening price are multiplied.

In statistics, a moving average is a calculation used to analyze data points by creating a series of averages of different subsets of the full data set. In finance, a moving average (MA) is a stock indicator that is commonly used in technical analysis. The reason for calculating the moving average of a stock is to help smooth out the price data by creating a constantly updated average price.

By calculating the moving average, the impacts of random, short-term fluctuations on the price of a stock over a specified time frame are mitigated. The 50-day and 200-day moving average figures for stocks are widely followed by investors and traders and are considered to be important trading signals. It shows the trend and range of price movement. Secondly, the points of resistance and support that lie along the 50-day line are often respected by the daily trades. These points do not break easily, and prices bounce back from the support levels or pull back from the resistance levels aligned on the moving average line.

Correlation, in the finance and investment industries, is a statistic that measures the degree to which two securities move in relation to each other. The correlation can measure the movement of a stock with that of a benchmark index. Correlations play an important role in finance because they are used to forecast future trends and to manage the risks within a portfolio. We find a stronger correlation in the Western markets like DJI and FTSE when compared to BSE.

A scatter matrix is an estimation of covariance matrix when covariance cannot be calculated or costly to calculate. The scatter matrix is also used in a lot of dimensionality reduction exercises. Scatter plots show how much one variable is affected by another or the relationship between them with the help of dots in two dimensions. Scatter plots are very much like line graphs in the concept that they use horizontal and vertical axes to plot data points.

Candlestick charts are also known to have been derived from Japanese rice traders. These are widely used for technical analysis in trading as they visualize the price size within a period. They have four points: Open, High, Low, Close (OHLC). The box's filling represents the price's direction. Usually, a filled or red box means the price went down (Bearish market), so the open price is the rectangle's top part. An empty or green box means the opposite (Bullish market), and the top part of the box is the closing price.

We have used the `candlestick_ohlc` of `mpl_finance` method to plot the matplotlib candlestick chart in Python. We can pass arguments, like width, colorup, colordown, alpha etc. to this method. Thereafter, we have set the labels and title to our Python Subplot.

Volatility is a statistical measure of the dispersion of returns for a given security or market index. It is a rate at which the price of a security increases or decreases for a given set of returns. In

**COURSE: Random Process and Estimation Techniques (B.Tech AI)**

most cases, the higher the volatility, the riskier the security. Volatility represents how large an asset's prices swing around the mean price—it is a statistical measure of its dispersion of returns.

The cumulative return is the total change in the investment price over a set time—an aggregate return, not an annualized one. The cumulative return is equal to your gain (or loss!) as a percentage of your original investment. Thus, the formula for cumulative return is:

$R_c = (P_{\text{current}} - P_{\text{initial}}) / P_{\text{initial}}$       which can also be written:

$R_c = (P_{\text{current}} / P_{\text{initial}}) - 1$

The `cumprod()` method returns a DataFrame with the cumulative product for each row. The `cumprod()` method goes through the values in the DataFrame, from the top, row by row, multiplying the values with the value from the previous row, ending up with a DataFrame where the last row contains the product of all values for each column.

**Applications:**

To find correlation and profit/loss between different stock indices using the scatter matrix.

**Simulation results:**

As we infer from the plot for total trade FTSE has high total trade for the year span from 2004 to 2017 but later on from 2017 to 2021, DJIA takes over the total trade with an increase in a few months in 2017. BSE SENSEX due to its low volume and low open price have low trade as inferred from the plot.

While finding the correlation using the scatter matrix we find that the Western markets have a better linear correlation when compared to BSE. From the volatility graph we thus see that BSE is most volatile then FTSE and lastly the least volatile is DJI. We again get the same result using the box graph that BSE is most volatile then FTSE and lastly the least volatile is DJI.

Using the KDE, we estimate the deviation of the stock prices which helps us to understand that BSE, having the lowest peak in kde, is the most volatile stock amongst the three followed by FTSE and DJI.

The graph of cumulative returns shows that BSE will give the most returns in the long run. We can also observe that all the three stocks' cumulative returns fell drastically during the Covid-19 period. This also shows that FTSE if used for long term investments would not give good returns.

**Conclusion:**

This project thus shows the comparison of a stock from the stock markets of the USA, UK and India. We observe the activity of these stocks in the time span of 2003 to 2021. The comparison is done in USD due to which the activity of BSE seems negligible when compared to other stocks. This statement gets verified when we see that BSE has given better cumulative return for the past years. The moving average helps us in detecting the general trends that we observe in the

**COURSE: Random Process and Estimation Techniques (B.Tech AI)**

activity of a stock. While finding the correlation using the scatter matrix, we find that the Western markets have a better linear correlation when compared to BSE. Using the KDE, we estimate the deviation of the stock prices which helps us to understand that BSE, having the lowest peak in kde, is the most volatile stock amongst the three followed by FTSE and DJI. Same is also seen using the Box Plots. Finally the cumulative return does show that BSE has given the most cumulative return per stock to its investors in the period between 2003 to 2021.