ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/335567761

Stock Market Prediction Using Various Statistical Methods Volume I

Preprint · July 2019

DOI: 10.13140/RG.2.2.13235.17446

CITATIONS
0

READS
6,394

1 author:

Karim Elbahloul
New Jersey City University
8 PUBLICATIONS **3** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Automatic Stock Trading Using Machine Learning Approaches View project

Statistical and Machine Learning Analysis View project

All content following this page was uploaded by Karim Elbahloul on 02 September 2019.

The user has requested enhancement of the downloaded file.

Stock Market Prediction Using Various Statistical Methods Volume I

By: Karim Elbahloul

Abstract- ***The use of statistical analysis has played a major role in assisting stock market prediction. This research paper aims at using various statistical methods to analyze historical data of the S&P 500 ETF (SPY) as well as forecast its direction for the next 12 months. Specifically, ARIMA Modeling, Exponential Smoothing, and Mean Squared Error analysis are the foci of interest.***

INTRODUCTION

Stock market prediction has been quite a desired research topic for years. Although there is the Efficient Market Hypothesis which proposes that stock prices follow a random walk and, thus, cannot be predicted [1-2], there is also the Adaptive Market Hypothesis which states that prices of stocks can be predicted [3] and the latter is what this paper would like to focus on.

When it comes to making trading decisions and stock price predictions, financial experts mainly make use of Technical Analysis, Fundamental Analysis [4-6], or both. Fundamental Analysis deals more with the business performance of a company while Technical Analysis makes use of the corporation's historical stock data and a few statistical tools to interpret those figures to come up with better predictions.

This paper is concentrated on using several statistical methods for stock market analysis and prediction. Specifically, the following methods would be discussed: Exponential Smoothing, Mean Square Error (MSE), and AutoRegressive Integrated Moving Average (ARIMA).

BACKGROUND

Using the statistical computational approach, one way to predict stock prices is to utilize ARIMA like Banerjee's Stock Prediction Model [7]. AutoRegressive Integrated Moving Average (ARIMA), is a form of regression analysis that measures the strength of one dependent variable relative to other changing variables. The model's goal is to predict future stock prices by examining the differences between values in the series.

Exponential smoothing is commonly used for analysis of time-series data. It is a method for making some determination based on prior assumptions by the user, such as seasonality [8].

The Mean Squared Error (MSE) is a measure of the quality of a predictor. It is calculated by getting the average of the square of the difference between the predicted values and what is forecasted. [9]

Employing the above-mentioned statistical tools to the dataset, this research will aid in the analysis and forecast of stock market prices. The tasks will include extracting meaningful statistics and characteristics about the data at monthly intervals over a period of one year to recognize trends, cycles, and seasonal variances.

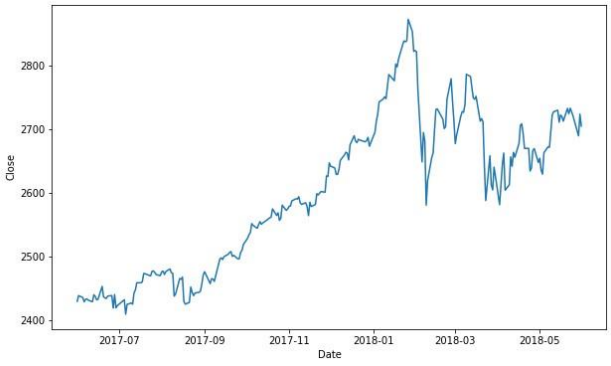

Figure 1: Historical S&P500 (SPY) Dataset Graph from 06/01/2017 – 05/31/2018. X axis is the dates, Y axis is the closing prices.

DATASET DESCRIPTION

For this paper, the dataset from an S&P 500 Exchange-traded fund called SPY will be analyzed. The S&P 500 ETF Trust (SPY) (Formally Standard & Poor's 500 Index) is a market-capitalization-weighted index of the 500 biggest U.S. publicly-traded components by market value. The index is widely regarded as the best single measure of large-cap U.S.

equities. Therefore, it is an excellent gauge of the overall performance of the stock market.

S&P 500 dataset is publicly available on Yahoo Finance. It consists of historical daily opening and closing stock prices along with a few other additional market parameters. Fig. 1 shows the daily closing price of S&P 500 ETF (SPY) from June 1, 2017 to May 31, 2018 [10].

EXPERIMENTAL ENVIRONMENT

The following libraries were used in conducting the statistical analysis:

- Matplotlib [11] was used for plotting the graphs for the actual time-series as well as the predicted trends.
- Pandas [12] was used for reading value from csv files as DataFrames.
- Numpy [13] was used to perform matrix operations like flip, reshape, and create random matrices.

METHODOLOGY

The statistical methods used in this study were Auto-Regressive Integrated Moving Average (ARIMA), Exponential Smoothing, and Mean Squared Error (MSE).

First, for ARIMA, tests were performed to check stationarity, which is particularly useful in determining whether the time-series data was stationary or not. First, rolling statistics were examined, which included rolling mean and rolling standard deviation to plot a moving average. Second, the Augmented Dickey–Fuller (ADCF) test was conducted. Furthermore, a statistical trend was also calculated by taking the logarithmic approach. Non-stationarity was also proven with moving average calculation derived from successive segments of the series values.

Second, for Exponential Smoothing, the chosen dataset was imported and analyzed utilizing the built-in data analysis tool in Excel. The purpose of using Exponential Smoothing, is to level out the data. This method is very useful to show a prediction about the market's behavior after one year. Exponential smoothing has been shown to help stock brokers see whether a bull or bear market is present at the time period. As well as to predict future stock price values.

Lastly, Mean Squared Error (MSE) was used to check how close the points are to the line of best fit. This was carried out using Python with the above-mentioned libraries. In general, the lower the MSE value, the better.

RESULTS AND DISCUSSION

AUTO-REGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) –

The results for the rolling statistics (i.e. rolling mean and rolling standard deviation) show stationarity (See Figure 2).

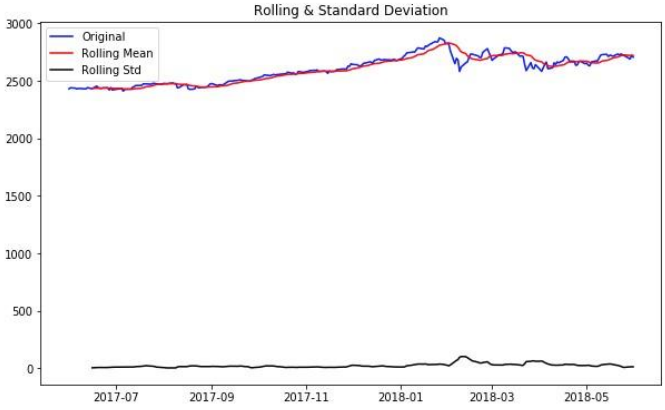

Figure 2: Rolling Mean & Standard Deviation for SPY from 06/01/2017 – 05/31/2018

As regards Augmented Dicky Fuller Test (ADCF) test, the results showed a test statistic of -1.323 and a critical value of -3.45, which means that the data is stationary. Since the test statistic is greater than the critical value.

This brought various results including a Critical Value of -3.45 and a Test Statistic of 0.061. This also concluded that data is non-stationary.

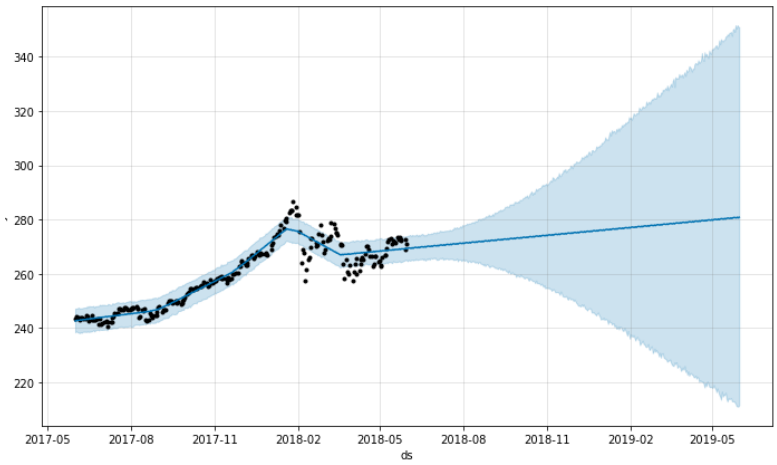

Figure 3: ARIMA 1-year forecast results from SPY data from 06/01/2017 – 05/31/2018

The next goal was to choose the parameters for the ARIMA models using auto-correlation function (ACF) and partial auto-correlation function (PACF). After several models were created and tested, the best result for a 1-year forecast of the stock closing price for SPY showed a directional trend that was going upwards (See Figure 2). And with respect to the Autocorrelation component, the Residual Sum of Squares (RSS) was 0.0140 and the Moving Average RSS was 0.0146. The lower the RSS, the better the ARIMA model.

EXPONENTIAL SMOOTHING –

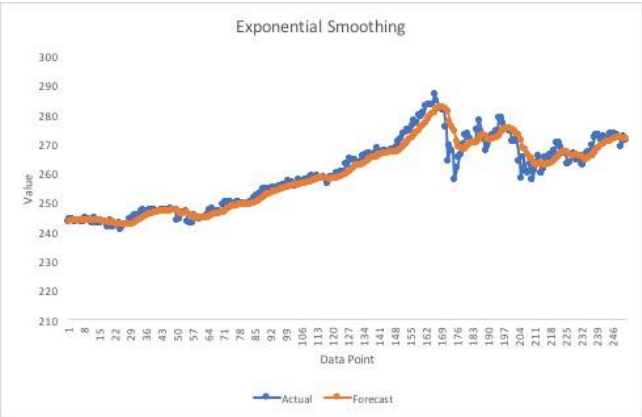

Figure 4: Exponential Smoothing (Orange Line) of the closing price data points for a 12-month period from 6/01/2017 – 5/31/2018

According to Figure 4, there is an upward trend beginning at day 64 until it reached its peak at day 155. Exponential Smoothing helps in looking at the bigger picture more clearly to see where the stock direction is moving.

MEAN SQUARE ERROR (MSE) –

The Linear Regression & Scatter Plot map that was generated is shown below (See Figure 5).

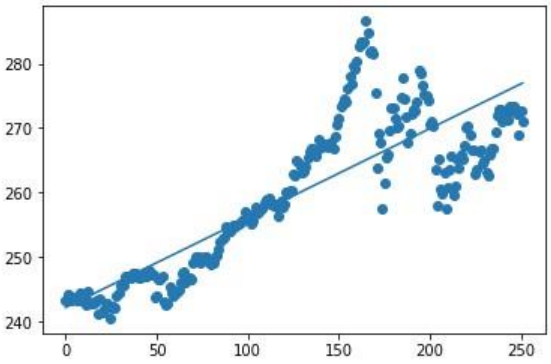

Figure 5: Linear regression & scatter plot of SPY beginning 06/01/2017 – 05/31/2018.

The results of the MSE analysis gave us a value of 0.719. Further improvement of the ARIMA model can give us a better predictor and a much lesser MSE numerical value.

CONCLUSION

Using various statistical methods, this analytical study has demonstrated how the SPY can be analyzed and predicted. By means of ARIMA modeling, the 1-year forecast showed an upward trend. Furthermore, Exponential Smoothing aided in visualizing the same trend. Finally, the Mean Squared Error gave us an indication of how the points are close to the line of best fit.

REFERENCES

[1]

F. Campanella, M. Mustilli and E. D'Angelo, "Efficient Market Hypothesis and Fundamental Analysis: An Empirical Test in the European Securities Market," Review of Economics & Finance, pp. 27- 42, 2016.

[2]

B. G. Malkiel, A Random Walk Down Wall Street: The Time-Tested Strategy for Successful Investing, 11th revised edition. W. W. Norton & Company, 2016.

[3]

A. W. Lo, Adaptive Markets Financial Evolution at the Speed of Thought. Press Princeton, 2017.

[4]

M. C. Thomsett, Getting Started Getting Started in Stock Analysis, Illustrated Edition. Hoboken, Singapore: John Wiley Publishing, 2015.

[5]

M. Krantz, Fundamental Analysis For Dummies. Indianapolis, Indiana, USA: Wiley Publishing, 2016.

[6]

B. Rockefeller, Technical Analysis For Dummies. Wiley Publishing, Indianapolis, 2011.

[7]

D. Banerjee, "Forecasting of Indian Stock Market Using Time-Series ARIMA Model", In: Proceedings of the 2nd International Conference on Business and Information Management, IEEE 2014, pp.131-135.

[8]

NIST/SEMATECH e-Handbook of Statistical Methods,http://www.itl.nist.gov/div898/handbook/, July 11, 2019

[9]

Lehmann, E.L.; Casella, George (1998). *Theory of Point Estimation* (2nd ed.). New York: Springer.

[10]

"Novel Deep Learning Model with CNN and Bi-Directional LSTM for Improved Stock Market Index Prediction" https://vermaabhi23.github.io/publication/2019CCWC1.pdf [Accessed July 11, 2019]

[11]

The Matplotlib development team (2018, Nov. 28). *matplotlib Version 3.0.2* [Online]. Available: https://matplotlib.org/. [Accessed July 11, 2019].

[12]

T. Mester (2018, Jul. 10). *Pandas Tutorial 1: Pandas Basics* [Online]. Available: https://data36.com/pandas-tutorial-1basics-reading-data- files-dataframes-data-selection/. [Accessed July 11, 2019].

[13]

NumPy [Online], (2018). Available: http://www.numpy.org/. [Accessed July 11, 2019].

Karim Elbahloul

View publication stats