Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018)
IEEE Xplore Compliant Part Number: CFP18K74-ART; ISBN:978-1-5386-2842-3

# *Stock Price Forecasting Using Data From Yahoo Finance and Analysing Seasonal and Nonseasonal Trend*

Jai Jagwani
Dept. of Computer Science and Engineering
Inderprastha Engineering College
Ghaziabad, India
jai.jagwani94@gmail.com

Manav Gupta
Dept. of Computer Science and Engineering
Inderprastha Engineering College
Ghaziabad, India
gupta.manav96@hotmail.com

Hardik Sachdeva
Dept. of Computer Science and Engineering
Inderprastha Engineering College
Ghaziabad, India
sachdeva.hardik@gmail.com

Alka Singhal
Dept. of Computer Science and Engineering
Inderprastha Engineering College
Ghaziabad, India
alka.singhal@ipec.org.in

***Abstract*—The stock market is an aggregation of individuals and corporations engaged in a loose network of buying and selling of shares of companies, called stocks. This buying and selling of shares have high risk associated with it, and therefore several predictions are made to avoid losses and incur profits. Time series analysis is the most basic and fundamental technique to perform this task. This paper depicts the broader view of stock price prediction by combining the results of the different time series analysis model, to find a range of stock prices for the buyers in which no losses will be incurred. The proposed work aims to find the relationship between the stock prices and different existing time series algorithms namely ARIMA and Holt Winter, to explore a safe range of stock prices for investments and thus improving the accuracy of prediction. The results of this work show that these predictions will be equally effective if done for a long or short time intervals.**

***Keywords—stock price forecasting; time series analysis;ARIMA; holt-winters method.***

## I. INTRODUCTION

Generally, the prices of stocks are considered to be fluid and unstable in nature. Still, a huge population invests in stocks daily aiming to yield profits from the market. These investors are mostly unaware of stock market behaviour. Due to this many are confused about which stock to buy or sell to incur more profits. All they know is prices of the stocks of previous dates; which they can use to roughly predict the current and future prices.

Although there are other factors as well such as news about the market, industry or a particular company etc. that can affect the prices of stock; but these issues generally increase the dependency of stock price prediction on them rather than focusing primarily on the actual prices of stocks. Therefore, Time series analysis is the most optimal solution to the problem.

Time series analysis is a fascinating area with practical and relevant impact. It works on numerical data collected over a considerable period of time and then used to generate future values of the series. It makes use of previous data points to predict new data points further. Hence, it eliminates the dependency of one factor of prediction on the other in the other factors [1]. This technique can be used in several areas such as economics, e-commerce, finance, and banking.

In this paper, the analysis is conducted on the data of Apple Inc (appl). This choice has been made because it is an old and popular company and a large amount of information is available online about Apple Inc that is relevant to our research work. Our fundamental data is in the form of old stock prices [2].

There are several techniques under Time series analysis to predict stock prices. In this paper, we have used few of them and combined their results to generate a range of prices between which the buyer will stuck profits.

## II. LITERATURE REVIEW

Analysing and forecasting time series has been a great research endeavour for decades. In 1919, Persons decomposed time series and came up with the idea of decomposition model with seasonal adjustment in time series [3]. Holt(1957) and Winters(1960) developed a method to analyse seasonal time series through weighted exponential smoothening [4]. Later in 1976, Box Jenkins proposed his work of analysing time series using ARIMA model [5]. This model is still used in many real-world applications. Ruey S. Tsay in 2001 described the analysis of any economic data through time series in his book *Analysis of financial time series* [6]. In 2005, Suhartono S. worked on comparing the results of different time series analytical models and finding that if a complex model yields better results [7].

## III. VARIATION IN TRENDS

Generally, stock prices rise at the beginning of while they fall during August and September; issues like these are not coincidental but this is due to the seasonal nature of stock prices. There are several reasons behind seasonal trends; stock funds try to improve at the end of the year resulting in an increase in stock prices. An additional reason can be a partial flow of year-end bond coupon payments into the stock market which is due to the interest payments made in the month of December. The taste of buyers also depends on the holiday season, generally, their moods are positive about investments. Different markets have different reasons for seasonality in the stock prices and it is difficult to set the exact criteria but few of them could be payments in a specific season(i.e, at year end), emotional quotient of investors at a certain time(i.e., holiday season), aberrations in finance reporting etc.

At the same time, nonseasonality in the data of stock prices can be seen in cases such as extreme move during a year, like the crash of 1987; this may affect the seasonal graph more than a normal year would. In addition, it can also happen that similar moves occur in consecutive years forming a series which has no true seasonal basis [8].

## IV. TIME SERIES ANALYSIS

Various analytical methods are described under time series analysis. One of them is Holt-Winters method which is used for forecasting and also it has three exponential smoothening methods to capture level, trend, and seasonality of the data. In mathematical notation these are represented as follows:

$$l_x = \alpha(y_x - s_x - L) + (1-\alpha)(l_x - 1 + b_x - 1) \quad (1)$$

$$b_x = \beta(l_x - l_{x-1}) + (1-\beta)b_{x-1} \quad (2)$$

$$s_x = \gamma(y_x - l_x) + (1-\gamma)s_x - L \quad (3)$$

$$y_{x+m} = l_{x+m} + b_x + s_x - L + 1 + (m-1)modL \quad (4)$$

Here, (1) shows the sum of level and trend where $l_x$ denotes the level, alpha is the smoothening factor, $y_x$ is the first observation in series. In (2) $b_x$ denotes a trend, beta is the trend factor and first trend value is treated as first expected. In (3) gamma is the smoothening factor of seasonal component, in (4) y represents the forecasted value, the expected value index is x+m where *m* can be any integer used to forecast as many values as needed. In (4) $s_x$-L+1+(m-1)modL is the index of seasonal component of the forecast which is actually seasonal component of previous data [9].

Nonseasonal ARIMA(autoregressive integrated moving average) is another analytical method; it is one of the most popular methods to capture the nonseasonal trend in time series. In mathematical notation it can be represented as:

$$y'_t = c + \varphi_1 y'_{t-1} + \cdots + \varphi_p y'_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t \quad (5)$$

Here, in (5) $y'_t$ is a differenced series, *e* defines the error terms and phi and theta are random weights; the right-hand side of the equation includes the lagged values of $y_t$ and error both [10]. This model is called ARIMA(p,d,q) where p, d and q denote the order of autoregression, the degree of first differencing, and order of moving average respectively [11].

The values of p, d, and q are comprehended by calculating ACF(autocorrelation function) and PACF(partial autocorrelation function) described below.

The correlation between two variables $y_1$ and $y_2$ is defined as:

$$\rho = \frac{E[(y_1 - \mu_1)(y_2 - \mu_2)]}{\sigma_1 \sigma_2} \quad (6)$$

In (6), E is the expectation operator,$\mu_1$ and $\mu_2$ are the means respectively for $y_1$ and $y_2$ and $\sigma_1$,$\sigma_2$ are their standard deviations.

In the context of auto-correlation $y_1$ is the original series and $y_2$ is a lagged version or it. Using the above-mentioned formula, sample autocorrelations of order *k*=0,1,2,... can be found by performing simple computations over the observed series $y_t$, t=1,2,...,n shown in (7):

$$\rho(k) = \frac{\frac{1}{n-k}\sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})}\sqrt{\frac{1}{n-k}\sum_{t=k+1}^{n}(y_{t-k} - \bar{y})}} \quad (7)$$

where $\bar{y}$ is mean of the data.

The partial auto-correlation measures the linear dependency of one variable or factor after removing the other thus eliminating its effect on both the variables. Its mathematical representation is:

$$\bar{y_t} = \varphi_{21}\bar{y}_{t-1} + \varphi_{22}\bar{y}_{t-2} + e_t \quad (8)$$

In (8) $\bar{y_t}$ is the original series minus mean of the sample, $\bar{y_t} - \bar{y}$. The estimate of $\varphi_{22}$ will give the value of the partial autocorrelation of order 2. $e_t$ is the error term added to the expression [12].

## V. PROPOSED WORK

This paper proposes to apply two different forecasting models; Holt Winter exponential smoothening and ARIMA(autoregressive integrated moving average ) on the data(stock prices) of Apple Inc and later combining their results to find a range of stock prices in which a buyer can secure its profits.

## VI. WORKING

In this paper, analysis has been conducted on monthly stock prices as it provides many stable results rather than quarterly or weekly data. Monthly historical stock prices from January 2000 to January 2018 are extracted from Yahoo finance website; this dataset contains the open, high, low, close and adjusted close prices of Apple Inc stock on the 1st date of every month throughout these eighteen years. It also contains trading volume values of these months. To achieve consistency, close prices are used as a general measure for the purpose of analysis in this work.

Fig. 1 represents the data which is separated into two parts, training data comprising the stock prices from January 2000 to January 2016 and test data comprising the stock prices from February 2016 to January 2018. This fragmentation is necessary to validate the outputs of predictions through test data.

To perform time series analysis on the historical stock prices, we first plot the training dataset call it an actual trend.

In Fig. 2 the plot shows increasing and decreasing prices of stocks during past eighteen years. Although there is no apparent pattern in the graph yet it is evident that variance has increased slightly and prices are volatile at the end of 2015 and hence logarithmic transformation function needs to be applied on the data for the stability of variance. The rest of the plot after 2015 is the curve of test data which is not considered while taking above mentioned observations.

To capture the seasonal trend in data, Holt-Winters method needs to be applied over the actual trend along with its triple exponential smoothening.

In Fig. 3 the red coloured curve shows the actual trend which turns to blue after prediction and the green coloured curves are the exponential smoothening of Holt-Winters algorithm.

Since the above doesn't consider the nonseasonal trend of stock prices hence ARIMA model is applied over the actual trend to study the behaviour of prices when seasonality in data is disrupted.

In Fig. 4 the graph shows the red coloured curve of an actual trend which later on turns blue after prediction and the purple coloured curve shows the predicted value of stocks through ARIMA(2,1,0). Since there was no linear or discernable pattern in data hence first degree differencing was performed over the series and also the graphs of ACF(auto-correlation function) and PACF(partial auto-correlation function) of data provides the information to perform autoregression twice on the series but no moving average, therefore ARIMA(2,1,0) was preferred.

It is evident from the above two graphs(Fig. 3 & Fig. 4) that a better output can be extracted if results of the two methods Holt-Winter and ARIMA are combined together to find a range of prices in which the investor can incur its profits. Hence, we combine the outputs of the two curves to find the range of stock prices, which our final result as mentioned above.

## VII. RESULTS & OBSERVATIONS

The red coloured curve in Fig. 5 is the actual trend which turned blue in the grey portion after prediction. The green coloured curves are Holt-Winters predictions through exponential smoothening and purple coloured curve shows the prediction through ARIMA model. The dark grey portion in the above curve is the range of prices amongst which the investor will incur no losses and as we move away from the dark grey portion on either side(top or bottom) on the vertical scale the risk of investing on stocks increases.

## VIII. CONCLUSION

This paper simply depicts the combination of two different time series analysis models to find precise results for the investment in the stock market by producing a range of prices to the buyer of stocks. These two models have certain advantages because of their simplicity in nature and also because these predict values purely on the basis of historical stock prices taking into account the seasonal or nonseasonal behaviour of data which is an additional benefit. Yet there are a number of limitations and shortcomings in this experiment as this work never takes into account other factors such as news about any market policy or press release relevant to any company or industry which may affect the prices of stocks. Also, potential improvements can be made in our data collection and analysis method. Further researches can be done which may possibly improve the results of this experiment.

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | Date | Open | High | Low | Close | Adj Close | Volume | |
| 2 | 01-01-2000 | 3.745536 | 4.339286 | 3.089286 | 3.705357 | 2.508365 | 3138794400 | |
| 3 | 01-02-2000 | 3.714286 | 4.283482 | 3.464286 | 4.09375 | 2.771291 | 1829945600 | |
| 4 | 01-03-2000 | 4.234375 | 5.370536 | 4.071429 | 4.850446 | 3.283542 | 2174589200 | |
| 5 | 01-04-2000 | 4.839286 | 4.982143 | 3.745536 | 4.430804 | 2.999461 | 2165601200 | |
| 6 | 01-05-2000 | 4.459821 | 4.508929 | 2.919643 | 3 | 2.03087 | 2451937600 | |
| 7 | 01-06-2000 | 2.919643 | 4.116071 | 2.870536 | 3.741071 | 2.532543 | 2026301200 | |
| 8 | 01-07-2000 | 3.723214 | 4.330357 | 3.348214 | 3.629464 | 2.456989 | 1436692600 | |
| 9 | 01-08-2000 | 3.59375 | 4.392857 | 3.160714 | 4.352679 | 2.946575 | 1409021600 | |
| 10 | 01-09-2000 | 4.379464 | 4.580357 | 1.8125 | 1.839286 | 1.245116 | 3629232600 | |
| 11 | 01-10-2000 | 1.90625 | 1.910714 | 1.25 | 1.397321 | 0.945926 | 5476447200 | |

Fig. 1. Data collected from Yahoo finance presented on excel sheet

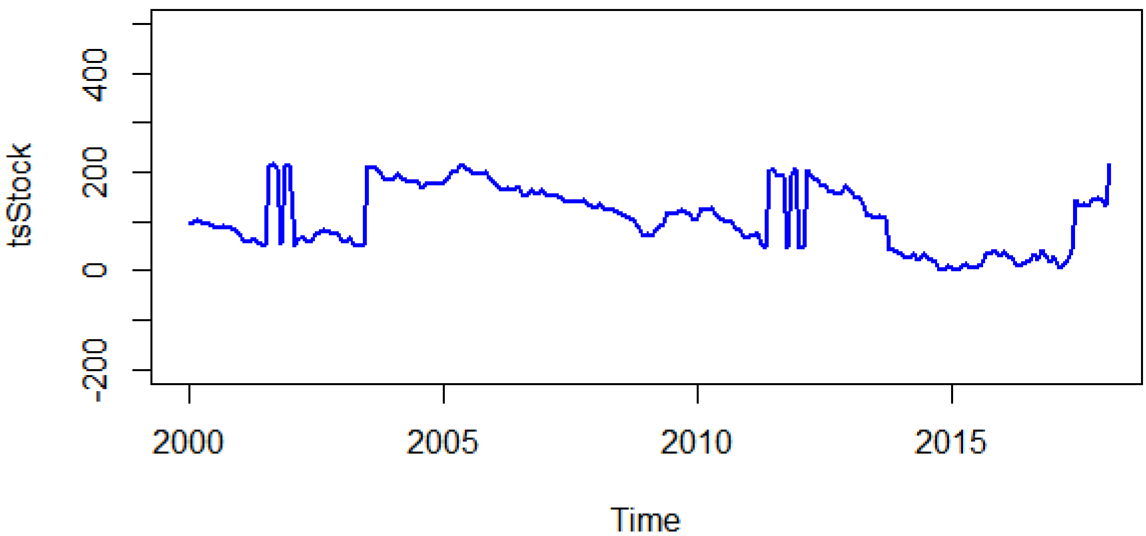


Fig. 2. Monthly close prices of Apple Inc. (Actual Trend)

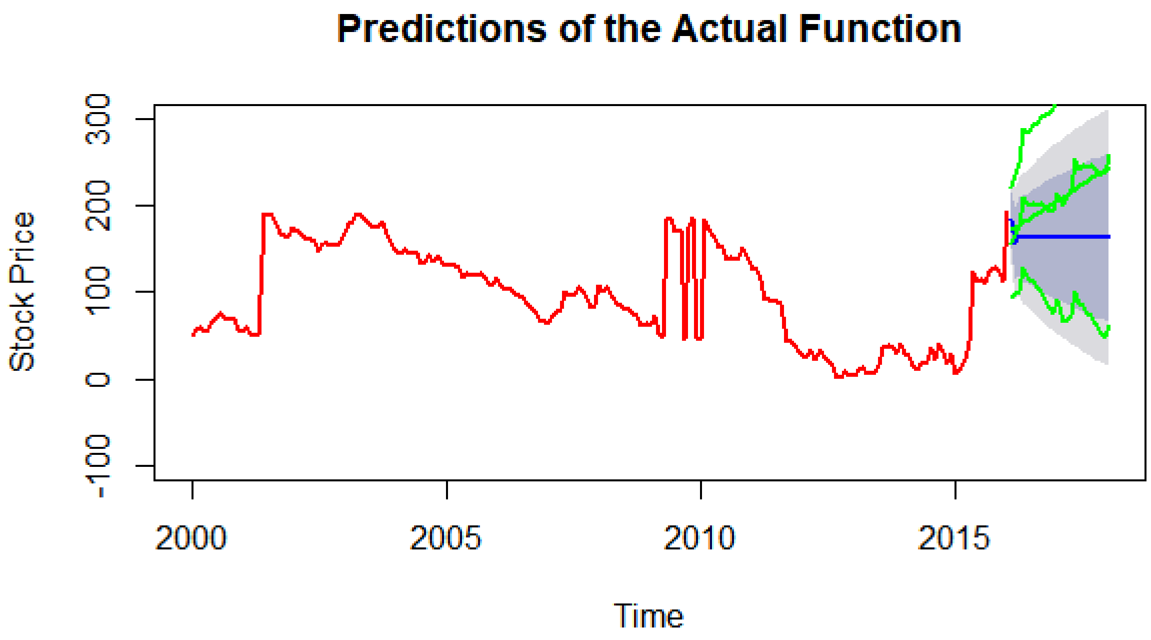


Fig. 3. Application of Holt-Winters method over the actual trend

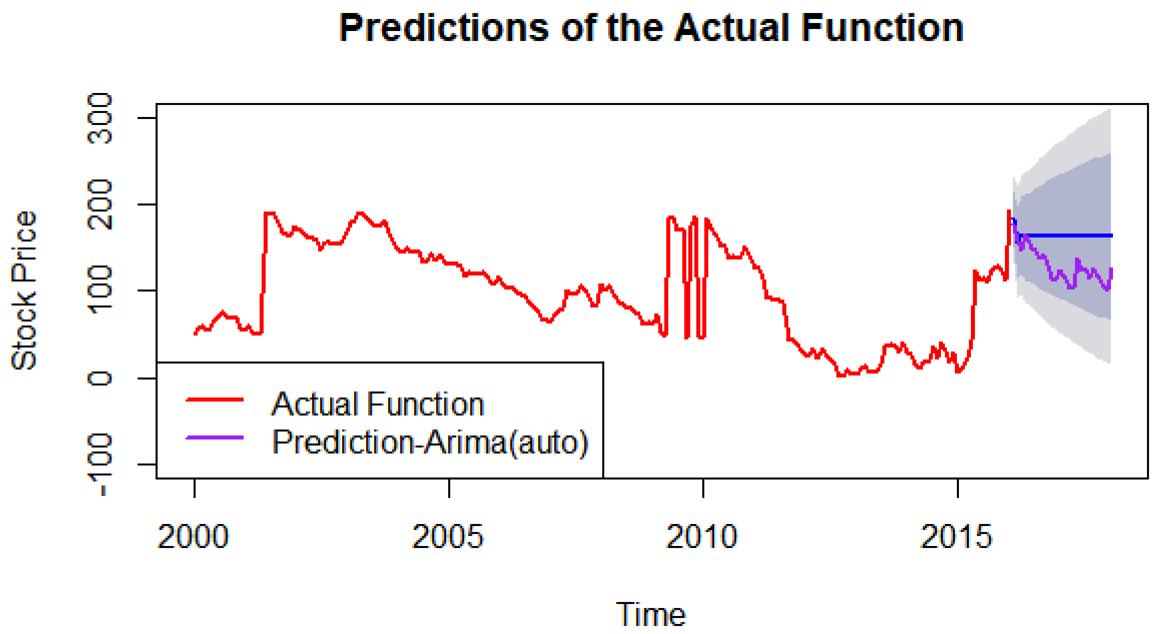


Fig. 4. Application of ARIMA model over the actual trend

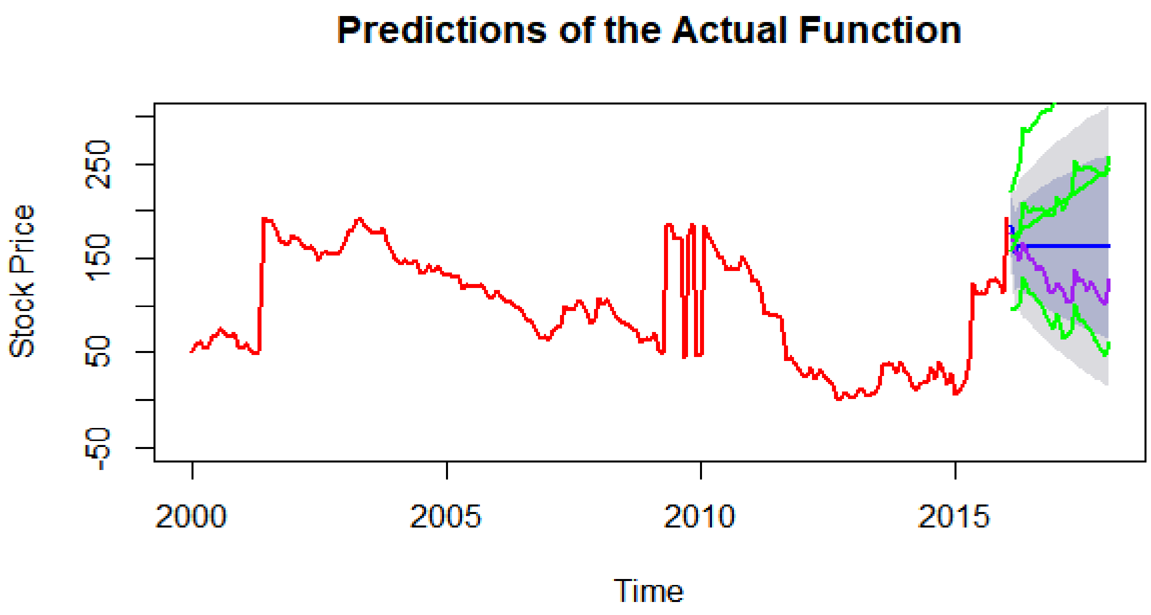


Fig. 5. Combined results of Holt-Winters method and ARIMA model

## REFERENCES

[1] Hamilton, James Douglas, Time series analysis, Princeton university press Princeton, 1994.

[2] Xu, Selene Yue; Berkely, UC, "Stock price forecasting using information from Yahoo finance and Google trend," *UC Brekley*, 2014.

[3] D. and E. Bee, "Time series modeling and decomposition," *Statistica*, vol. 70, p. 433, 2010.

[4] Kalekar; Prajakta S, "Time series forecasting using holt-winters exponential smoothing," *Kanwal Rekhi School of Information Technology*, vol. 4329008, p. 12, 2004.

[5] G. E. Box, G. M. Jenkins and R. , Time series analysis: forecasting and control, John Wiley & Sons, 2015.

[6] Tsay, Ruey S, Analysis of financial time series, John Wiley & Sons, 2005.

[7] Suhartono, Suhartono, "A comparative study of forecasting models for trend and seasonal time series does complex model always yield better forecast than simple models," *Jurnal Teknik Industri*, vol. 7, p. 22, 2005.

[8] Speck, Dimitri, "home," seasonalcharts.com, 2000. [Online]. Available: http://www.seasonalcharts.com/saisonalitaet.html. [Accessed 11 may 2018].

[9] Chawla, Santoshi, December 2015. [Online]. Available: https://www.quora.com/How-do-I-calculate-the-ideal-values-for-Alpha-Beta-and-Gamma-parameters-in-Holt-Winters-method/answer/Santoshi-Chawla. [Accessed 11 May 2018].

[10] Insel, Aysu, 23 May 2011. [Online]. Available: https://www.scribd.com/document/56090364/VAR-and-Coint#. [Accessed 11 May 2018].

[11] Ramasubramanian, V, "Time series analysis," *IASRI, Library Avenue, New Delhi*, 2007.

[12] (https://stats.stackexchange.com/users/48766/javlacalle), javlacalle, *ACF and PACF Formula*, Cross Validated, 2014.