

PAPER • OPEN ACCESS

Time series forecasting for the adobe software company's stock prices using ARIMA (BOX-JENKIN') model

To cite this article: R. Vaibhava Lakshmi and S. Radha 2021 *J. Phys.: Conf. Ser.* **2115** 012044

View the [article online](#) for updates and enhancements.

You may also like

- [A new GNSS outlier mitigation method for GNSS/INS integrated system](#)
Jingwen Guo, Yilan Zhou, Shuai Zhao et al.
- [Self-Identification Deep Learning ARIMA](#)
Paisit Khanarsa, Arthorn Luangsodsai and Krung Sinapiromsaran
- [Autoregressive Planet Search: Application to the Kepler Mission](#)
Gabriel A. Caceres, Eric D. Feigelson, G. Jogesh Babu et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Time series forecasting for the adobe software company's stock prices using ARIMA (BOX-JENKIN') model

R. Vaibhava Lakshmi¹, S. Radha^{2,*}

¹ Department of Computer Technology, Madras Institute of Technology, Chennai

² School of Advanced Sciences, Vellore Institute of Technology, Chennai

* radha.s@vit.ac.in

Abstract. The time series forecasting strategy, Auto-Regressive Integrated Moving Average (ARIMA) model, is applied on the time series data consisting of Adobe stock prices, in order to forecast the future prices for a period of one year. ARIMA model is used due to its simple and flexible implementation for short term predictions of future stock prices. In order to achieve stationarity, the time series data requires second-order differencing. The comparison and parameterization of the ARIMA model has been done using auto-correlation plot, partial auto-correlation plot and auto.arima() function provided in R (which automatically finds the best fitting model based on the AIC and BIC values). The ARIMA (0, 2, 1) (0, 0, 2) [12] is chosen as the best fitting model, with a very less MAPE (Mean Absolute Percentage Error) of 3.854958%.

Keywords: Time series forecasting, Stock prices, ARIMA

1. Introduction

A time series is a succession of data values indexed with incrementing time intervals, i.e, it is a sequence taken at successive evenly spaced timespans. A univariate time series refers to a time series that comprises of single (scalar) observations recorded successively over intervals of time . A multivariate time series refers to a time series consists of multiple regressands. Stock prices forecasting is of paramount importance in the realm of business and finance. Predicting the performance of stock market is a herculean task. The entire motive of forecasting stock prices is to acquire immense profits. Stock price prediction using ARIMA model helps to determine the future values of companies' stock and other pecuniary assets dealt on an exchange. Predominantly, stock market analysis encompasses two types - Fundamental and Technical analysis. Fundamental Analysis analyses the company's future profitability depending on the current business environment and financial performance. Technical Analysis identifies the trends in the stock market using charts and statistical figures. Our focus of study is technical analysis. Few renowned algorithms used for time series analysis include Autoregressive (AR) model, Box-Jenkins model (ARIMA), Exponential smoothening, Long Short Term Memory Networks and the N-BEATS algorithm. Generally, Machine learning and Deep learning algorithms work really well for long term prediction, whereas, ARIMA



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

model is preferable over Machine learning and deep learning algorithms for short term predictions , due to its flexible and simple implementation.

Adobe, otherwise known as, Adobe Systems Incorporated, is a multinational software company with its headquarters located in San Jose, California, United States. Adobe has multitude of users globally. It has remarkably flourished in software creation and publication of a broad spectrum of software content, constituting of illustration, animations, multimedia, print, motion pictures and graphics. It's leading products include: The Photoshop, Adobe Illustrator, Adobe Acrobat Reader and PDF and a large number of tools chiefly for audio-visual content creation, modification and publishing. In this paper, Box-Jenkins (ARIMA) model is used to predict the future stock prices of Adobe multinational company for a period of one year. The dataset used provides the history of monthly prices of Adobe Stock from the year 2014 to 2021. It is collected from Yahoo! Finance, a media property which is a part of Yahoo! The sole motive behind selecting ARIMA model for this study is that, this model incorporates, the non-zero autocorrelation between the sequential time series data points. The free statistical software 'R' (4.1.0) having statistical and time series packages such as 'tseries', 'seastests', 'forecast' and 'TSA' are used apart from the other common packages for the purpose of this study.

2. Literature Review

Ayodele A. Adebisi et al [1] used ARIMA model to predict stock price by fitting the model for the published data from New York Stock Exchange (NYSE) and Nigeria Stock Exchange (NSE). From their results it was concluded that ARIMA model is very efficient for short term predictions. Carina Intan Permatasari, Wahyudi Sutopo and Muh. Hisjam [2] used the output of ARIMA models to predict printed newspaper demand accurately to reduce the number of returns and to restrain the oversupply. Jamal Fattah, Latifa Ezzine et al [3] used Box - Jenkins time series method to forecast the demand in a food company. They used the historical demand data to forecast future. García-Cremades, S., Morales-García, J., et al. [4] used an extensive range of models including artificial neural networks such as LSTM and GRU and statistical models like ARIMA for the prognosis of rampant emergence of the pervasive COVID-19 disease to construct a decision support system for policyholders.

Kumar Manoj and Anand Madhu [5] used ARIMA model to study and sugarcane production in India for a period of 5 years by fitting ARIMA (2,1,0) model. Their results forecast that the sugarcane production per year would increase in 2013, then there would be a steep decline in 2014 and it would continuously proliferate with an estimated average growth rate of 3% from 2015 to 2017. Prapanna et al., [6] analysis studied the effectiveness of ARIMA model, on fifty six Indian stocks from different sectors and used Akaike information criterion for comparison and parameterization of the ARIMA model. Rupinder Katoch and Arpit Sidhu [7] used ARIMA method to study and analyse the time dependent surge of COVID-19 in India between 30 January 2020 and 16 September 2020 and predicted the eventual trend and magnitude of the effects of the pandemic after 16 September 2020 with epidemiological details at state and national levels in India. Subrina Noureen, Sharif Atique et al [8] used ARIMA to forecast the seasonal time series data. They also present analysis on testing stationarity and converting non-stationarity into stationarity. Subhash Arun Dwivedi, Amit Attry et al [9] compared SARIMA, CNN and LSTM models for time series forecasting and predicted Nifty-500 indices tendency. The outcomes disclosed the strength of Deep Learning through LSTM and CNN and hence, sanctions the SARIMA model.

3. Basics of ARIMA – 'Box-Jenkins' model

The Box-Jenkins model (ARIMA) is a kind of regression model that computes the puissance of a regressand compared to all the independent variables. The motive of using this model is prognosticating future securities by analysing the differences between data values in lieu of using the

original observations. It was developed in the 1970s by George Box and Gwilym Jenkins as an effort to describe changes on the time series using a mathematical approach.

An ARIMA model has the following components:

- AutoRegression (AR): A model that shows a changing variable that depends on its own lagged values.
- Integrated (I): Refers to the differencing between the observed values and the residual error so as to achieve the stationarity of the time series data.
- Moving average (MA): Includes the dependency between an observed value and a residual error from a moving average model that is applied to lagged observations.

For a p^{th} order autoregressive (AR) model, equation is :

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (1)$$

The term y_t denotes the data on which the ARIMA model will be applied and it implies that the series is power-transformed already and differencing is done in that order. The parameters $\phi_1, \phi_2, \dots, \phi_p$ represent the AR coefficients.

For a q^{th} order moving average (MA) model, equation would be:

$$y_t = C + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_p \varepsilon_{t-p} \quad (2)$$

Where y_t is as defined formerly and $\theta_1, \theta_2, \dots, \theta_p$ denote the MA coefficients.

Equation for an ARMA (p,q) model:

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_p \varepsilon_{t-p} \quad (3)$$

The components in ARIMA has an associated parameter with a standard notation ARIMA (p,d,q) where the parameters p, q and d are integers and they specify the type of ARIMA model used and they are defined as:

p: also called as the lag order denotes the count of lagged observations in the model

d: the number of times the raw observed values are differenced, also called the order of differencing

q: moving average window's size

An ARIMA model with seasonal component has the following additional parameters:

P: Seasonal autoregressive order

D: Seasonal difference order

Q: Seasonal moving average order

In an ARIMA model, differencing is done in order to eliminate trends or seasonal components. If a trend appears without evidence of stationarity, then the computations involved in the process cannot be done efficiently. Autocorrelation and partial autocorrelation plots are massively used in time series analysis and forecasting.

These plots graphically abridge the dependency of data values on lagged values, hence it is called a serial correlation, also known as autocorrelation. This plot is often called a correlogram. A partial autocorrelation depicts the same as the autocorrelation plot with the correlations of interposing observations discarded.

4. Time Series Analysis and Building ARIMA model

In order to build an ARIMA model for forecasting; model identification, estimation of parameters and diagnostic checking prior to forecasting are the steps required.

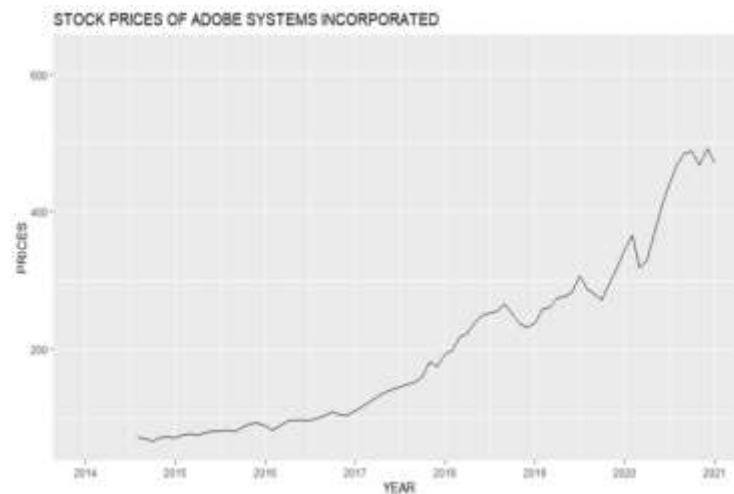


Figure 1 Time Series plot for Adobe Stock prices from 2014 to 2021

5. Model Identification

For building an ARIMA model, the first stage involves checking if the time series exhibits stationarity or not. For a time series data to be stationary, there shouldn't be trends or seasonality in the data. Seasonality can exist in a time series, in which the data experiences regular and predictable changes that recur every calendar year. In other way, Stationary data means the mean of the data doesn't not change with time. The variance of the series should not be time dependent, which is called Homoscedasticity.

With trends, the mean and variance are dependent on past values, so differencing needs to be done to remove the dependency of the data. Also, the covariance of the i^{th} and $(i+m)^{\text{th}}$ term must never be a function of time. The stationarity of the time series data is tested using the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test and it is obvious from the results that it is non-stationary. The p-value should be exceeding 0.05 for a time series to be stationary. The null hypothesis (H_0) in the test states that the time series data must be stationary while the alternative hypothesis (H_1) states that the series is non-stationary. The result of the KPSS test is as follows:

KPSS Level = 2.0679, Truncation lag parameter = 3, p-value = 0.01

Since the p-value is less than 0.05, the time series data is non-stationary. Upon first order differencing, the p-value is 0.02372 which is still lesser than 0.05, as given in the results below:

KPSS Level = 0.5881, Truncation lag parameter = 3, p-value = 0.02372

Only after second order differencing, the p-value is 0.1 which is greater than 0.05, as given in the results below:

KPSS Level = 0.023596, Truncation lag parameter = 3, p-value = 0.1

Hence, second order differencing is required. Also, using the `ndiffs()` function provided in the R-studio, it can be found that the time series requires second order differencing. Hence, the d - parameter value is 2.

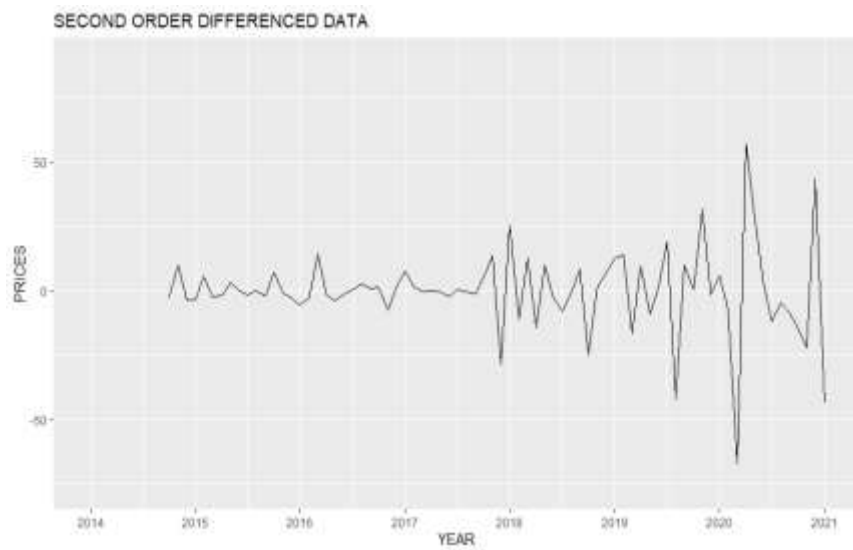


Figure 2 Plot for Second Order Differenced data

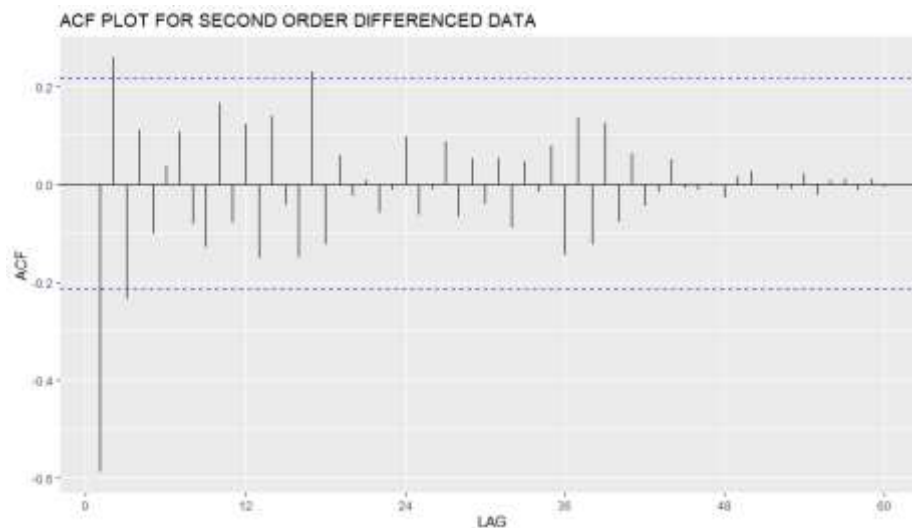


Figure 3 Auto Correlation plot for Second Order Differenced data

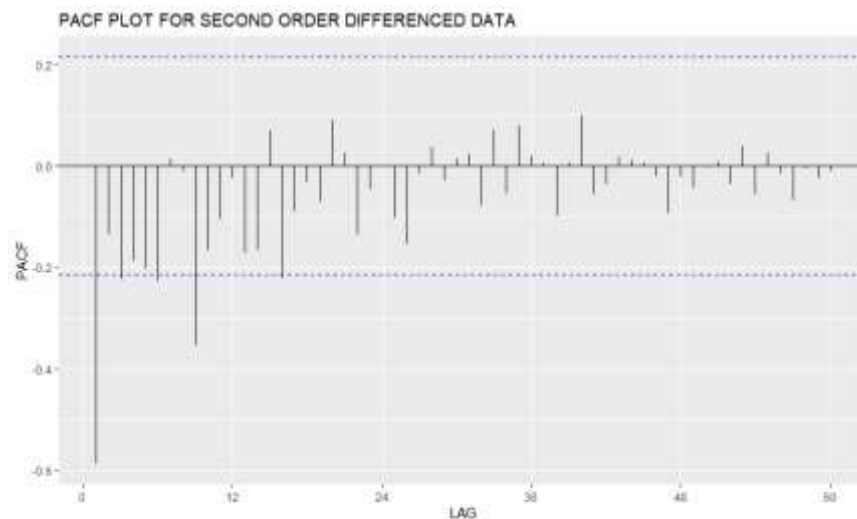


Figure 4 Partial Auto Correlation plot for Second Order Differenced data

Non-seasonal terms: The most significant spike at lag value 1 in the ACF plot suggest a non-seasonal MA(1) term. Also, the significant spike at lag value 1, suggests a non-seasonal AR(0) or AR(1) term. So, using `auto.arima()` function, it is found that the model has a non-seasonal AR(0) term. The order of non-seasonal differencing is already determined to be 2. Seasonal terms: Since, a monthly data is used in this study, the patterns across lag values of 12, 24, 36, . . . and so on are observed (mostly up to 3 terms only) in the ACF and PACF plots. In the PACF plot, no such pattern can be observed. Therefore, it suggests a seasonal AR(0) term. Similarly, from the ACF plot, it can be concluded that the model must have seasonal MA(2) term. We checked if the model parameters determined using the ACF and PACF plots are correct and optimum, using `auto.arima()` function. The summary of the ARIMA model obtained using `auto.arima()` function in R, is as follows:

```
ARIMA(0,2,1)(0,0,2)[12]
Coefficients:
      ma1      sma1      sma2
    -0.9476  0.3228  0.4604
s.e.  0.0372  0.1281  0.1963
sigma^2 estimated as 256.2: log likelihood=-350.43
AIC=708.86 AICc=709.38 BIC=718.54
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
2.134425 15.52877 10.04374 0.7862514 3.854958 0.140186 -0.003778985
```

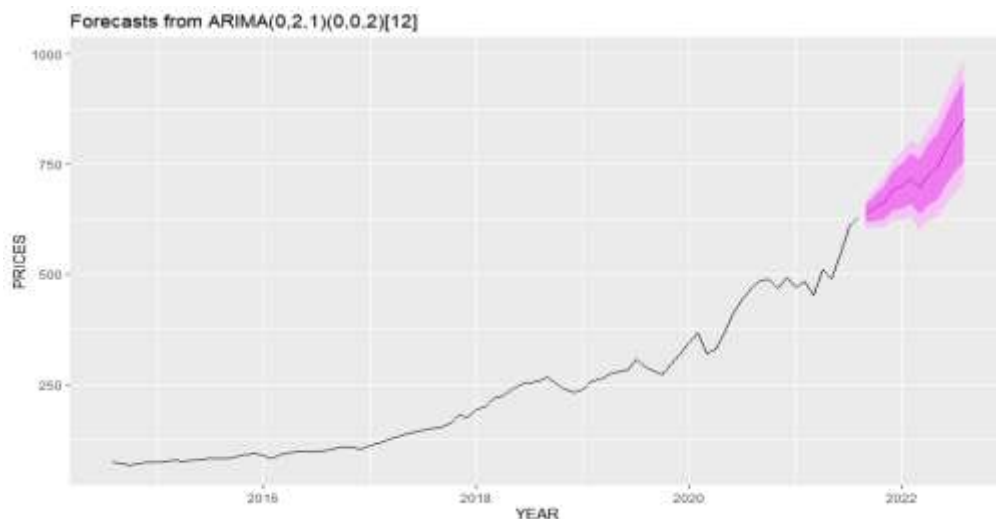
Hence, the best fitting ARIMA model is (0, 2, 1)(0, 0, 2)[12] with a MAPE (Mean Average Percentage Error) value of 3.854958% which indicates that the model is a very good fitting.

6. Results of Forecasting

The chosen ARIMA(0, 2, 1)(0, 0, 2)[12] model is suitable to predict the future values of the time series. The prediction of stock prices for the next one year with 80% and 95% prediction intervals is given in the table below.

Table 1: Forecast for a period of one year from September 2021

Time	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
September 2021	638.7708	618.2446	659.2969	607.3787	670.1628
October 2021	650.2292	620.4283	680.0300	604.6527	695.8056
November 2021	662.8857	625.4342	700.3372	605.6085	720.1629
December 2021	689.6155	645.2622	733.9688	621.7830	757.4480
January 2022	700.3382	649.5032	751.1731	622.5928	778.0835
February 2022	716.5672	659.5070	773.6274	629.3011	803.8333
March 2022	697.8188	634.6955	760.9421	601.2800	794.3575
Apr 2022	726.8696	657.7859	795.9532	621.2152	832.5239
May 2022	742.8417	667.8605	817.8230	628.1678	857.5157
June 2022	784.0769	703.2328	864.9210	660.4365	907.7173
July 2022	819.0913	732.3986	905.7839	686.5063	951.6762
August 2022	849.2691	756.7271	941.8111	707.7383	990.7998

**Figure 5** Forecasts using ARIMA (0, 2, 1)(0, 0, 2)[12]

6.1 Residual Diagnostics

In Residual diagnostics, it is checked if the residuals are white noise or not. To explore distribution of the forecasting errors, the standard residuals and histogram of standard residuals of the fitted

ARIMA(0, 2, 1)(0, 0, 2)[12] model are plotted.

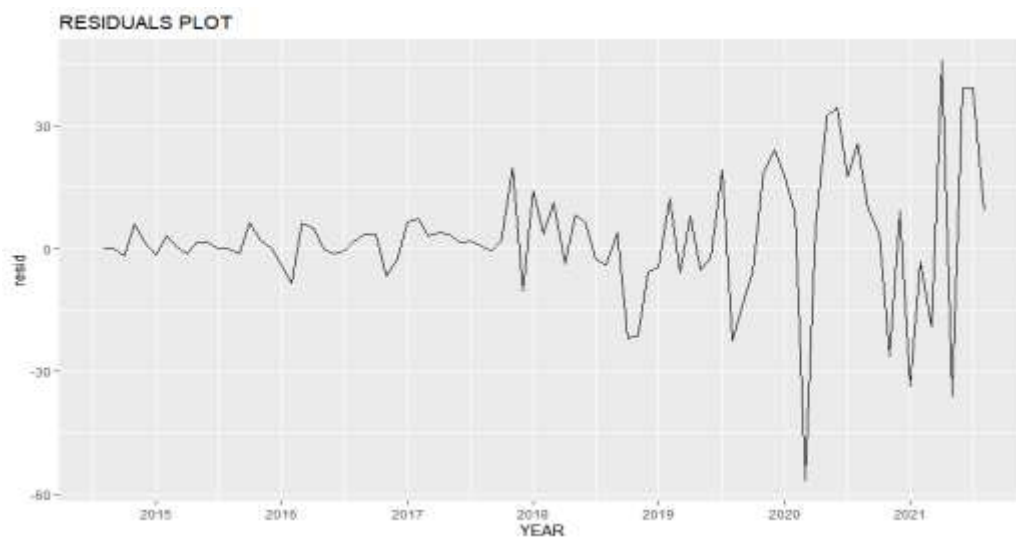


Figure 6 Residual Plot

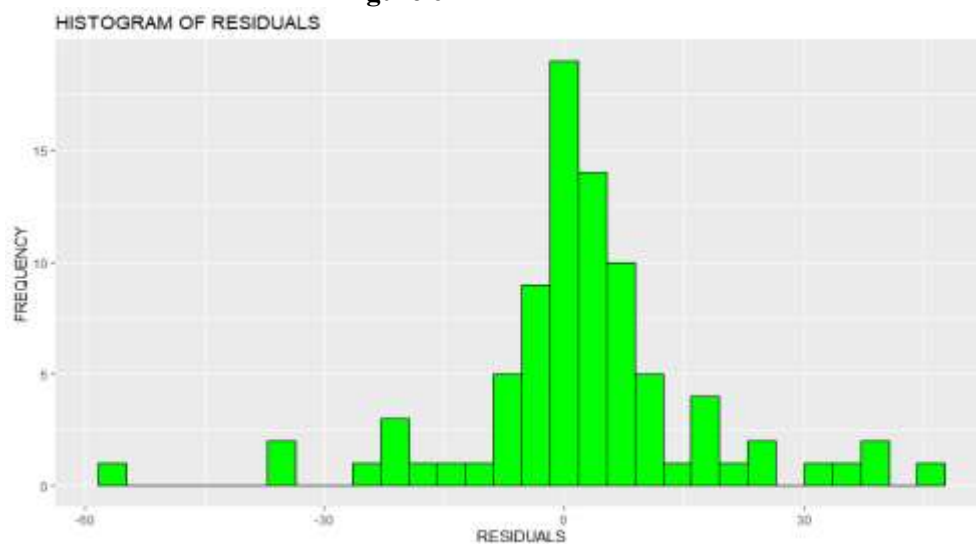


Figure 7 Histogram of Residuals

From the histogram, it can be inferred that the residuals are almost normally distributed and mean of the distribution seems to be zero.

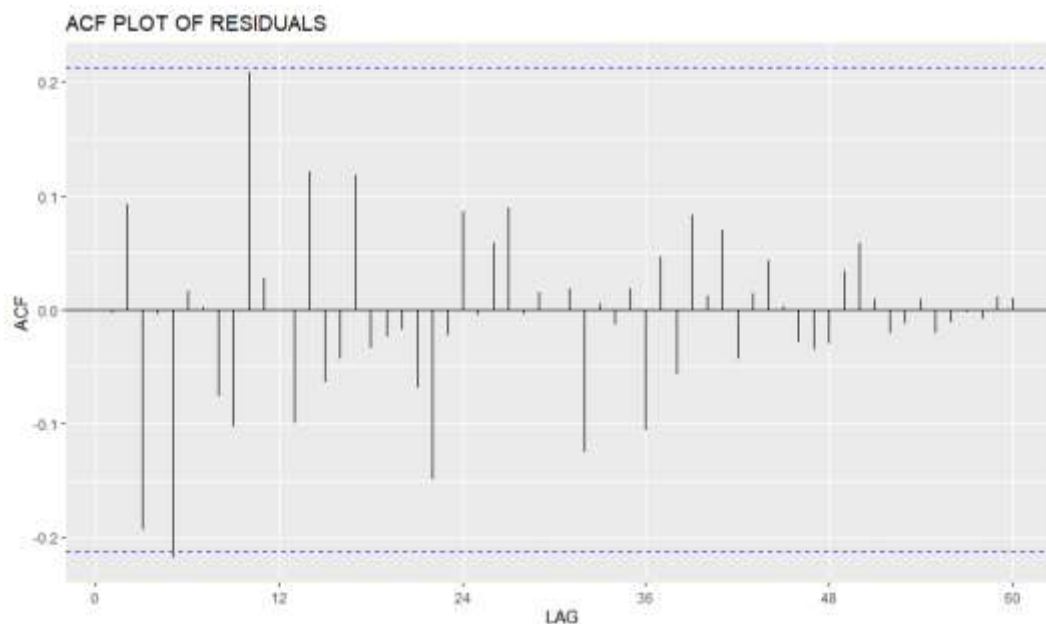


Figure 8 ACF plot for Residuals

From the above ACF plot, it can be observed that, all the correlation value lies within the significance limit, indicating that the residuals are white noise, hence the model is a good fit. The Ljung-Box test is used time and again in ARIMA modeling. The residuals of a fitted ARIMA model, are supposed to pass this test, not the original time series data. If the p value is in excess of 0.05 then the residuals are independent which is required for the model to be correct.

Results of Ljung - Box test:

X-squared = 0.0012572 df = 1 p-value = 0.9717

The model chosen is completely up to the mark and the forecasts are also very meticulous because the p-value is greater than 0.05.

7. Conclusion

In this paper, stock prices prediction of Adobe Systems Incorporated multinational company is done using the ARIMA model. From the results obtained, it can be incurred that there is a 12.179 % increase in stock prices from September 2021 to February 2022. There is a mild decrease of 2.6164 % in prices from February 2022 to March 2022. Again, there is a increase of 21.7034 % in prices from March 2022 to August 2022. The predictions made for the stock prices for next one year are precise and could guide investors in stock market to make pertinent decisions to invest. From the results obtained, it can be incurred that ARIMA model is advantageous and highly efficient for short-term predictions.

References

- [1] Adebiyi A. Ariyo, Adewumi O. Adewumi and Charles K Ayo (2014): "Stock Prediction using the ARIMA model", Proceedings of the 16th International Conference on Computer Modeling and Simulation UK – Sim (2014).
- [2] Carina Intan Permatasari, Wahyudi Sutopo and Muh. Hisjam (2018): "Sales forecasting newspaper with ARIMA: A case study", AIP Conference Proceedings, Vol. 1931.

- [3] Jamal Fattah, Latifa Ezzine, Zineb Aman, Haj El Moussami and Abdeslam Lachhab (2018): “Forecasting of demand using ARIMA model”, International Journal of Engineering Business Management, Vol.10, 1-9.
- [4] García-Cremades, S. Morales-García, J., Hernández-Sanjaime, R. et al.(2021): “Improving prediction of COVID-19 evolution by fusing epidemiological and mobility data”, Scientific Reports 11, 15173.
- [5] Kumar Manoj and Anand Madhu(2014):”An application of the time series forecasting model for predicting sugarcane production in India”, Studies in Business and Economics, Vol. 9, no. 1, pp. 81-94.
- [6] Prapanna Mondal, Labani Shit and Saptarsi Goswami (2014): “Study of Effectiveness of Time series modeling(ARIMA) in forecasting stock prices”, International Journal of Computer Science, Engineering and Applications, Vol. 4, No. 2.
- [7] Rupinder Katoch and Arpit Sidhu (2021):”An Application of ARIMA Model to Forecast the Dynamics of COVID-19 Epidemic in India”, Global Business Review, pp.1-14.
- [8] Subrina Noureen, Sharif Atique ,Vishwajit Roy and Stephen Bayne (2019): “Analysis and application of seasonal ARIMA model in Energy Demand Forecasting: A case study of small scale agricultural load”, IEEE Green Technologies Conference, 2019, pp. 521-524.
- [9] Subhash Arun Dwivedi, Amit Attry, Darshan Parekh and Kanika Singla (2021): “Analysis and forecasting of Time-Series data using S-ARIMA, CNN and LSTM”, 2021 International Conference on Computing, Communication and Intelligent Systems (ICCCIS), 2021, pp.131-136.