iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XI **Month of publication:** November 2021

DOI: https://doi.org/10.22214/ijraset.2021.38942

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

Validating The Assumptions of Residuals in ARIMA Model for Daily Stock Price Data By using R

Akasam Srinivasulu[1], V. Pavan Kumari[2], M. V. Narayana Murthi[3], M. Bhupati Naidu[4]
[1, 2, 3]*Research Scholar, Department of Statistics, Sri Venkateshwara University, Tirupati, Andhra Pradesh, India*
[4]*Professor & Associate Director, DDE, Sri Venkateshwara University, Tirupati, Andhra Pradesh, India*

Abstract: ***Identifying the past data and plannig for future is very important for every organization . Now a days Stock market playes a major role for the development of economy. For the countries economic development, stock market plays a vital role. For this modelling, forecasting is the best way to know the future stock prices based on the past stock prices data. In stock price data, forecasting of closed price plays a major role in financing economic decisions. The Arima model has developed and implemented in many applications .So the researchers utilize arima model in forecasting the closed prices of AMAZON stock price data for future which have been collected from AMAZON 2007-01-03, to 2020-10-12.In this paper the researcher aim is to forecast by using the ARIMA time series model with particular reference to Box and Jenkins approach on daily stock prices of AMAZON With open statistical software R. The validity of ARIMA model is tested by using the standard statistical tests.***
Keywords: Auto Regressive Integrated Moving Average, Auto Correlation Function, Partial Auto Correlation Function, Akaikae Information Criterion, Auto Regressive Conditional Heteroscedasticity

I. INTRODUCTION

A time series is a sequence where a metric is recorded over regular time intervals.
Depending on the frequency, a time series can be of yearly , quarterly , monthly ,weekly , daily , hourly , minutes and even seconds wise. The stock market works a network of exchanges Companies list shares of their stock on an exchange through a process called an initial public offering.
Investors purchase the share which allows the company to raise the money to grow its business. Investors can then buy and sell these shares among themselves and exchange tracks the supply and demand of each listed Stock. To buy the shares of a stock on a stock exchange investors go through brokers who trained in the science of stock trading .Stock prices are not randomly generated values this data is a discrete time series model based on asset of well defined numerical data items collected at successive points at regular intervals of time .To analyse trends of stock prices arima is a better approach it gives better results. The general research associated with the stock market is highly focusing on neither buy nor sell but it fails to address the dimensionality and expectancy of a new investor .The common trend for stock market among the society is that it is risky for investment .To solve this type of problem is the time series analysis is the best tool for forecasting the trend.
The stock market is a market that enables the seamless exchange of buying and selling of company stocks . The stock holders are interested to know the future value of stock prices on basing the past stock price data ,to purchase the shares of a perticular company. Every stock exchange has its own stock index value .The index is the average value that is combining several stocks. This helps representing the stock market and predicting the markets movements over time.In stock price data the closed price forecasting plays a major role in finance and economics. The arima model has developed and implemented in many applications .So the researchers utilize arima model in forecasting the closed prices of AMAZON stock price data for future which have been collected from AMAZON 2007-01-03, to 2020-10-12.The AMAZON stock price data can be downloaded by using the ticker symbol AMZN.

II. LITERATURE REVIEW

Stergiou (1989) in his examination utilized ARIMA model procedure on a 17 years' time series information (from 1964 to 1980 and 204 perceptions) of month to month gets of pilchard (Sardina pilchardus) from Greek waters for anticipating as long as a year ahead and figures were contrasted and real information for 1981 which was not utilized in the assessment of the boundaries. The examination found mean mistake as 14% proposing that ARIMA method was equipped for anticipating the perplexing elements of the Greek pilchard fishery, which, in any case, was hard to foresee on account of the year-to-year changes in oceanographic and organic conditions.

Raymond Y.C.Tse (1997) recommended two inquiries should be addressed to recognize the information series in a period series investigation (1) Whether the data are arbitrary (2) The data have any patterns . This followed by one more three stages of model recognizable proof ,boundary assessment and testing for model validity.If the perceptions of time series are genuinely subject to one another then the arima is suitable for time series investigation.

Meyler et al (1998) drew a framework for ARIMA time series models for forecasting Irish inflation. In their research, they emphasized heavily on optimizing forecast performance while focusing more on minimizing out-of-sample forecast errors rather than maximizing in-sample 'goodness of fit'.

Contreras et.al(2003) in their study using arima model , provides a method to predict next day electricity prices both for spot markets and long term contracts for mainted spain and California markets.

Josni in their study using Arch –LM test the persistence of volatility is more than Indian stock market.

Maheshchandra ,in their study using arfima model ,the absence of long memory in return series of the the Indian stock markets. Strong evidence of long memory in conditional variance of stock indices.

Shalini ,in their study Return of the BSE Sectorial indices exhibit characteristics of normality stationarity and heteroscedasticity.

Forecasting of demand using ARIMA model by Jamal Fattah ,the work presented in this article constitutes a contribution to modelling and forecasting the demand in a food company, by using time series approach. This paper explains how the historical demand data could be utilized to forecast future demand and how these forecasts affect the supply chain.

III. DATA SOURCE, VARIABLES AND METHODOLOGIES

The purpose of the study is to forecast the closing prices of AMAZON stock price data by using the ARIMA time series model .This study mainly concentrates Arima model for this data is satisfying assumptions of residuals (Autocorrelation, Heteroscedasticity) by using standard statistical tests.

A. Variables

In Stock price data there are 6 columns but we are interested to forecast the closing prices of the AMAZON stock prices. These closing prices are converted into Returns. Here Return is the target variable.

IV. METHODOLOGY

The study concentrated daily data, this data is taken from 2007-01-03, to 2020-10-12 . First stage of ARIMA model building is to identify whether the variable, which is being forecasted, is stationary in time series or not. By stationary we mean, the values of variable over time varies around a constant mean and variance. The ARIMA model cannot be built until we makes series stationary. First we have to difference the time series 'd' times to obtain a stationary series in order to have an ARIMA(p,d,q) model with 'd' as the order of differencing used. Caution to be taken in differencing as over differencing will tend to increase in the standard deviation, rather than a reduction. The best idea is to start with differencing with lowest order (of first order, d=1) and test the data for unit root problems. To forecast the daily returns by using the 'R' language. In the process of forecasting first we check the stationarity of the data by using the Augmented Dicky fuller test or by inspecting the plots of ACF and PACF.If the data is stationary then we construct ARIMA model otherwise convert data into stationary. ARIMA stands for autoregressive integrated moving average. The general model is written as

$$Yt = \phi 1 Yt-1 + \phi 2 Yt-2 \ldots \phi p Yt-p + \epsilon t + \theta 1 \epsilon t-1 + \theta 2 \epsilon t-2 + \ldots \theta q \epsilon t-q$$

Where Y_t is the differenced time series value, ϕ and θ are unknown parameters and are independent identically distributed error terms with zero mean. Here Y_t is expressed in terms of its past values and the current and past values of error terms.

A. The ARIMA Model has three Components

1) *Auto Regression (AR):* In auto regression the values of given time series data are regressed on their own lagged values .this is denoted by 'p' in ARIMA model
2) *Differencing (I):* This involves differencing the time series data to remove the trend and convert a non stationary time series to a stationary one . This is denoted by 'd'.
3) *Moving Average (MA):* The moving average in ARIMA model is represented by 'q' which is the number of lagged values of the error term.

This model is called Autoregressive integrated moving average or ARIMA(p,d,q)

B. Testing and Ensuring Stationarity

To model a time series with ARIMA ,the series must be stationary which means the series having a constant mean and variance over time , which makes it easy for predicting values.

We test for stationarity using the Augmented Dicky Fuller unit root test .The P-value from ADF test has to be less than 0.05 then the data is stationary.If the value of p is greater than 0.05 which means the data is non stationary

C. Differencing

To convert a non stationary data to stationary we apply diffencing method . Differencing a time series means finding the difference s between consecutive values of a time series data..We can apply differencing method consecutively more than once.

D. Identification of pand q

We identify the appropriate orders AR and MA process by using the ACF and PACF

For AR models ,the ACF will dampen exponentially and the PACF will be used to identify the order (p) of the AR model. If one significant spike at lag 1 on the PACF ,then we have AR model of order 1. If the significant spikes at lag1,2, on PACF ,then we have an AR model of the order 2 ,that is AR(2).For MA model ,the PACF will dampen exponentially and the ACF plot will be used to identify the order of the MA process. If we have one significant spike at lag 1on the ACF ,then we have an MA model of order 1 ,i.e. MA(1).If we have significant spikes at lag 1,2,3 on the ACF ,then we have an MA model of the order 3,i.e.MA(3).

E. Estimation and Forecasting

Once we building the ARIMA model next check the assumptions of heteroscedasticity and autocorrelation by using Arch-LM test and Ljung-Box test. If the assumptions are satisfied next we forecast the future values by using the ARIMA model

V. REVIEW OF RESULTS

*****By seeing the graph of Histogram we observe the data is more or less normally distributed .

*In Augmented Dicky fuller test the null and alternative hypothesis is

H0: The data is Non-stationary

H1: The data is stationary

The value of ‘P’ is smaller than 0.05 so we reject the null hypothesis .so the data is stationary.

*By observing the graph of ACF and PACF there is no identification of stationarity in the data. But we observe a geometric pattern in PACF. So the model is moving average model ,in ACF 2 spikes are above the border line so here the model is MA(2)

*auto.arima fuction calculates the best arima model for this data ,for which model having the lowest AIC that model is the best model.For this data, the best arima model is ARIMA(0,0,2)

Here the P value is’ 0’, that is there is no AR term and d value is’ 0’ because there is no need of differencing because the data is stationary .The q value is ‘2 ‘that is we need two past MA terms

*From the residuals graph the large spikes are followed by large spikes and small spikes are followed by small spikes, this shows volatility present in the residuals.

*In Ljung-Box test the null and alternative hypothesis is

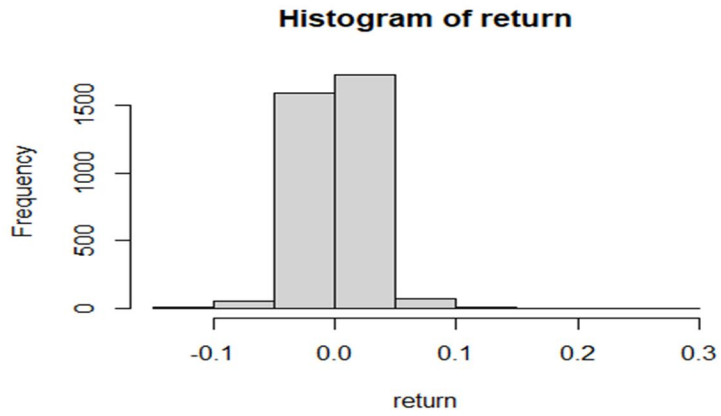
H0: There is no serial correlation

H1:There is serial correlation

The value of P in Ljung-Box test is larger than 0.05 so there no evidence to reject our H0.so there is no serial correlation in residuals

*In Arch LM test the null hypothesis and alternative hypothesis is

H0: There is no ARCH effects
H1: There is ARCH effects
Here the p-value is smaller than 0.05 so we reject our null hypothesis
So the data has an arch effect that is there is presence volatility in the residuals.

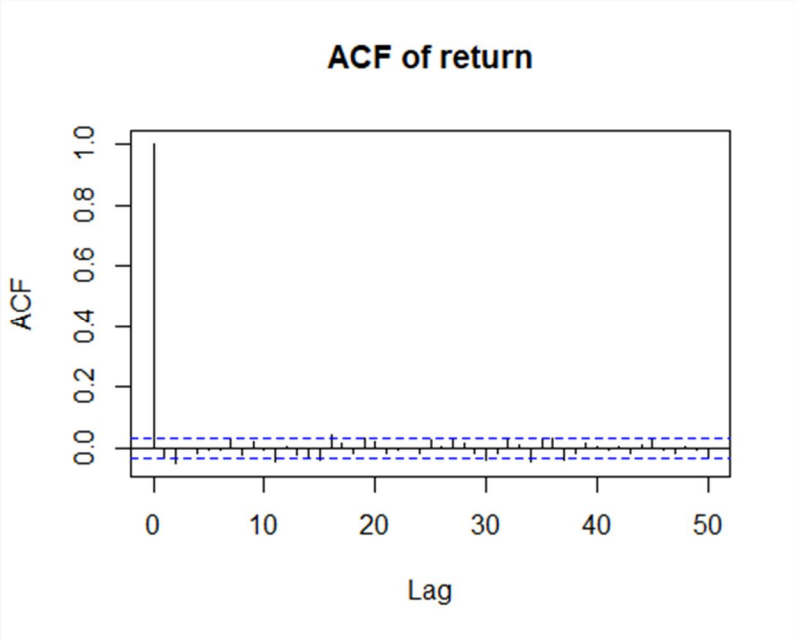


Augmented Dickey Fuller test:

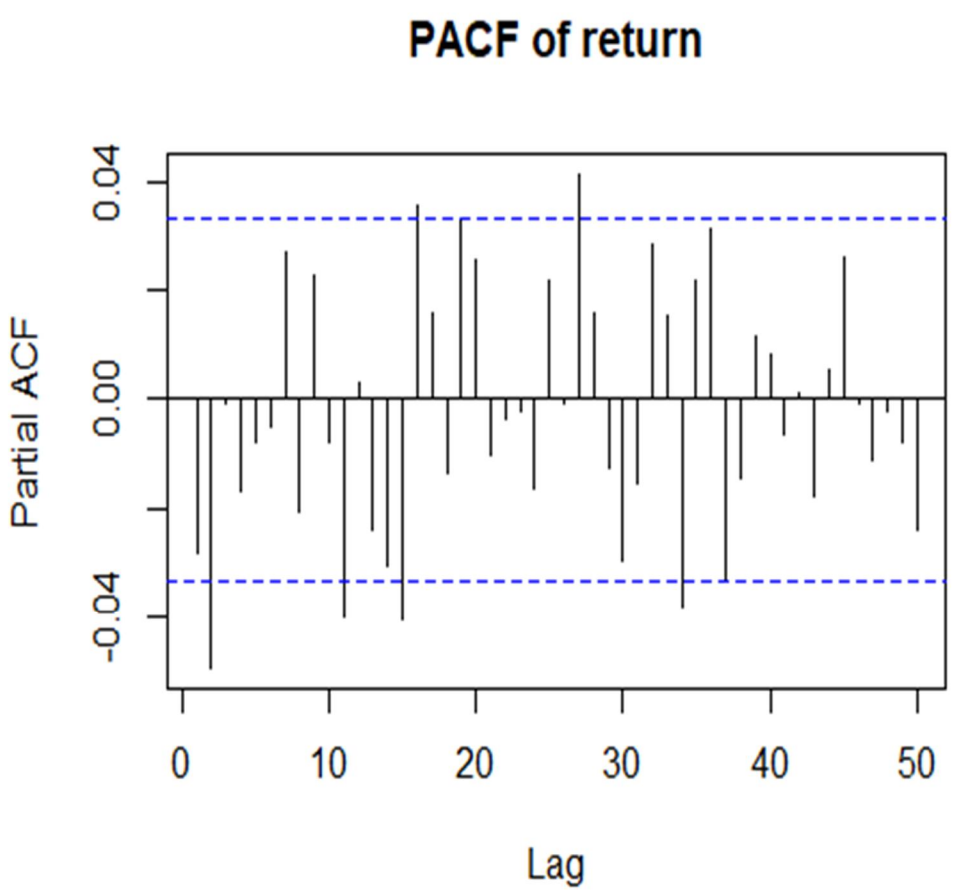
adf.test(return)

Augmented Dickey-Fuller Test:

data: return
Dickey-Fuller = -15.843, Lag order = 15, p-value = 0.01
alternative hypothesis: stationary



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com




```
ARIMA model
fit<- auto.arima(return,seasonal =FALSE)
fit

 Series: return
ARIMA(0,0,2) with non-zero mean

 Coefficients:
          ma1      ma2    mean
      -0.0295  -0.0502  0.0016
s.e.   0.0170   0.0172  0.0004

 sigma^2 estimated as 0.000594:  log likelihood=7959.55
 AIC=-15911.1   AICc=-15911.09   BIC=-15886.5

summary(fit)

 Series: return
ARIMA(0,0,2) with non-zero mean
```

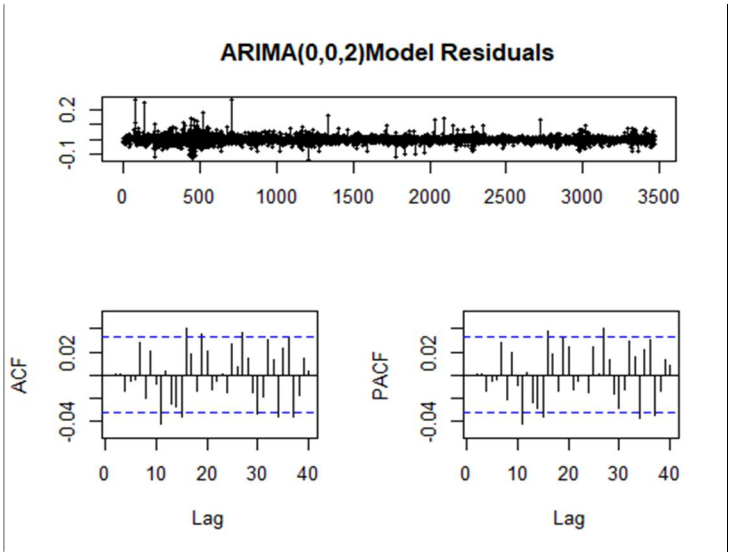
```
Coefficients:
         ma1      ma2    mean
      -0.0295  -0.0502  0.0016
s.e.   0.0170   0.0172  0.0004

sigma^2 estimated as 0.000594:  log likelihood=7959.55
AIC=-15911.1   AICc=-15911.09   BIC=-15886.5

Training set error measures:
                     ME       RMSE        MAE       MASE
Training set -6.334281e-07 0.02436171 0.01597536  0.6750501
                    ACF1
Training set -3.969631e-05
```

Forecast for Future 10 days:

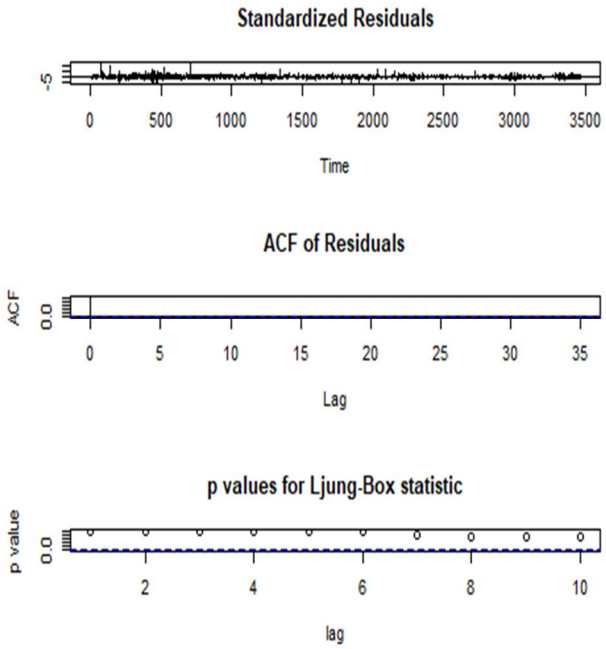
```
forecast_fit <-forecast(fit,h=10)
forecast_fit

##      Point Forecast       Lo 80      Hi 80       Lo 95      Hi 95
## 3468   8.910692e-04 -0.03034323 0.03212537 -0.04687767 0.04865981
## 3469   7.123688e-05 -0.03117669 0.03131916 -0.04771834 0.04786081
## 3470   1.573879e-03 -0.02971343 0.03286118 -0.04627593 0.04942368
## 3471   1.573879e-03 -0.02971343 0.03286118 -0.04627593 0.04942368
## 3472   1.573879e-03 -0.02971343 0.03286118 -0.04627593 0.04942368
## 3473   1.573879e-03 -0.02971343 0.03286118 -0.04627593 0.04942368
## 3474   1.573879e-03 -0.02971343 0.03286118 -0.04627593 0.04942368
## 3475   1.573879e-03 -0.02971343 0.03286118 -0.04627593 0.04942368
## 3476   1.573879e-03 -0.02971343 0.03286118 -0.04627593 0.04942368
## 3477   1.573879e-03 -0.02971343 0.03286118 -0.04627593 0.04942368
```

Checking the assumptions of Residuals:-

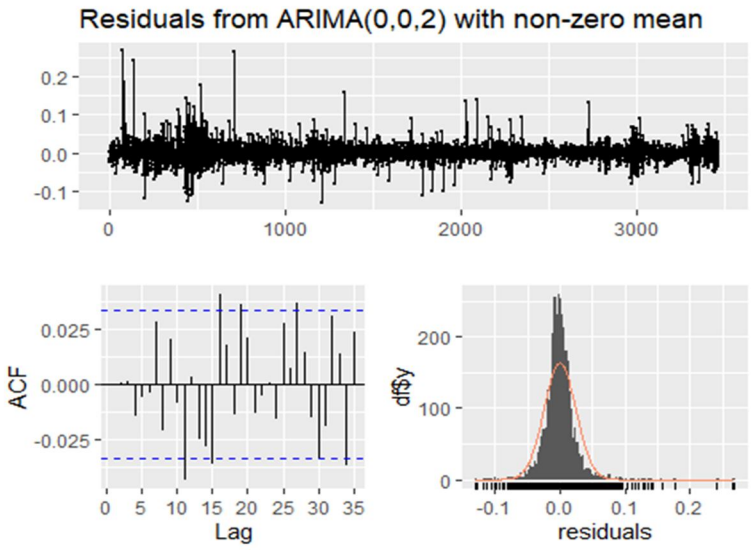
```
tsdisplay(residuals(fit),lag.max=40, main ='ARIMA(0,0,2)Model Residuals')
```



```
arima002 <-arima(return, order=c(0,0,2))
arimar <-arima002$residuals

summary(arimar)

      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-1.289e-01 -1.110e-02 -4.997e-04 -6.300e-07  1.078e-02  2.676e-01



tsdiag(arima002)
```




```
checkresiduals(arima002)
```



Ljung-Box test:

data: Residuals from ARIMA(0,0,2) with non-zero mean
Q* = 6.8322, df = 7, p-value = 0.4466

Model df: 3. Total lags used: 10

ARCH LM-test:

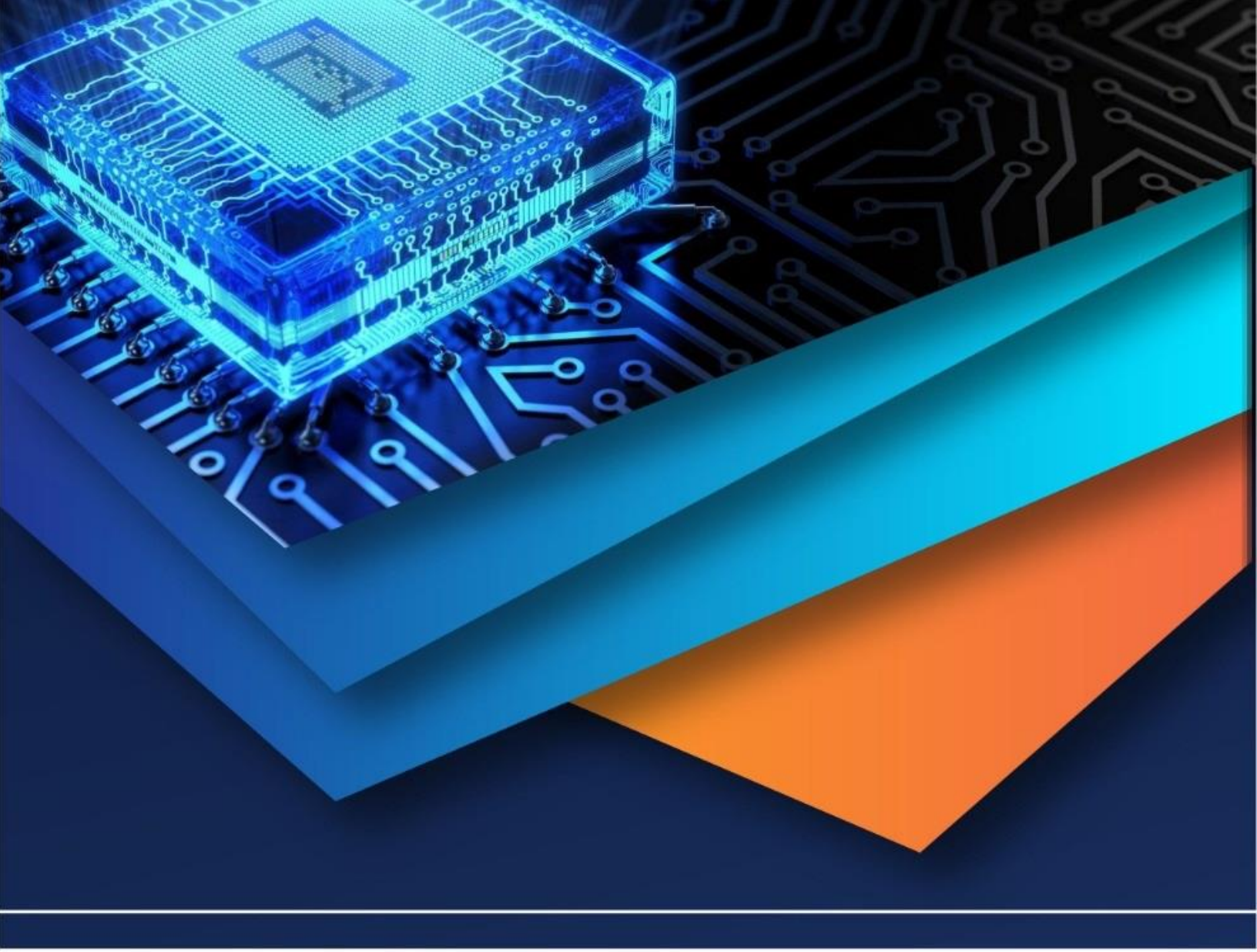
data: arimar
Chi-squared = 41.82, df = 1, p-value = 1.001e-10

VI. CONCLUSION AND FUTURE RESEARCH WORK

The basic aim of the study is to select best model for stock price data.By using the auto.arima function it displays MA(2) is the best model for the data. By observing the graph of residuals there is volatility clustering presence in the data, also in ARCH LM test the data has an arch effect.So in this case ARIMA model is not a good model. So, we go for ARCH and GARCH model for better predictions.ARIMA models are better for only small data.Further reading and findings are recommended in fitting ARCH and GARCH and advanced models for the data and their effectiveness. Volatile and dynamic data values requires proper diagnosis and analysis is recommended.

REFERENCES

[1] Box, G.E.P and Jenkin, G. M (1976). “Time series analysis in forecasting and control Applied statistics”. Holden-Day, San Francisco.
[2] Granger, C. W. J. and Newbold, P. (1976). “Forecasting transformed series, Journal of the Royal Statistical Society B”, 38, 189-203.
[3] Ljung, G.M. and Box, G.E.P. (1978). “On a Measure of Lack of Fit in Time Series Models”, Biometrika, 65, 297-303.
[4] P.Pai and C.Lin, “A hybrid Arima and support vector machines model in stock price prediction” Omega vol .33 pp.497-505,2005.
[5] Quan ,B., & Rasheed ,K.(2007), “Stock market prediction with multiple classifiers” .Applied intelligence,26(1),25-33
[6] Schumaker ,R.P.,& Chen ,H.(2010). “A discrete stock price prediction enginebased on finantial News” .Computer ,43(1),51-56.
[7] J.J. Wang J.Z. Wang , Z.G. Zhang and S.P.Guo. “Stock index forecasting based on a hybrid model” Omega vol.40 pp.758-766,2012.
[8] S.K. Mitra, “Optimal Combination of trading Rules using Neural networks”, International business research, vol.2, no1, pp.86-89.
[9] Malhotra, R. (2015). “A systematic review of machine learning techniques for software fault Prediction” .Applied soft computing , 27, 504-518.
[10] Mosavi ,A., Ozturk ,p., & Chau,K.W.(2018). “Flood prediction using machinelearning models” :Literature review .Walter .10(11),1536.
[11] Zhong ,X., &Enke ,D.(2019) “Predicting the daily return direction of the stock market using hybrid machine learning algorithms”.Financial innovation,5(1),4.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY