FORECASTING NIFTY 50 INDEX USING ARIMA TIME SERIES MODEL

A Term Paper on Econometric Methods

Submitted to Mahatma Gandhi University Kottayam as part of the Second Semester Course on *Term Paper on Econometric Methods* for the Degree of:

Master of Arts in Econometrics

By:

MELVIN MATHEW

(Register Number: 200011008396)

Under the guidance and supervision of

Mr. Ansel Kurian, MA

Guest Lecturer

Department of Economics

T.M. JACOB MEMORIAL GOVERNMENT COLLEGE MANIMALAKUNNU, KOOTHATTUKULAM, ERNAKULAM (DT), KERALA

MARCH 31, 2022

DEPARTMENT OF ECONOMICS
TM JACOB MEMORIAL GOVT.COLLEGE MANIMALAKUNNU
KOOTHATTUKULAM, ERNAKULAM, KERALA

March 2022

CERTIFICATE

This is to certify that this Term Paper on Econometric Methods, titled "**Forecasting NIFTY 50 Index Using ARIMA Time Series Model"**, submitted by Mr. MELVIN MATHEW, as part of the Second Semester MA Econometrics Programme for the Course of *Term Paper on Econometrics Methods (EM010206)* and is in partial fulfilment for the Degree of Master of Arts in Econometrics (2020-2022), is a record of work done by the candidate. Certified further that to the best of my knowledge the term paper represents an independent work done by the candidate and does not form part of any other thesis or dissertation.

PRINCIPAL **HEAD OF THE DEPARTMENT** **SUPERVISOR**

PLACE: KOOTHATTUKULAM
DATE: 31-03-2022

DECLARATION

I, MELVIN MATHEW, Second Semester, M.A. Econometrics student of Department of Economics, T.M Jacob Memorial Govt. College Manimalakunnu, Koothattukulam hereby declare that the *Term Paper on Econometric Methods* titled “ **Forecasting NIFTY 50 Index Using ARIMA Time Series Model ”** which is submitted as part of the Second Semester MA Econometrics Programme for the Course of *Term Paper on Econometrics Methods (EM010206)* and is in partial fulfilment for the Degree of Master of Arts in Econometrics (2020-2022) has been completed by me under guidance and supervision of MR. ANSEL KURIAN as part of the Second Semester MA Econometric Programme and is in partial fulfilment of the Degree of Masters of Arts in Econometrics (2020-2022).

MELVIN MATHEW

PLACE: KOOTHATTUKULAM
DATE : 31-03-2022

ACKNOWLEDGEMENT

This project has taken sufficient amount of time and effort from my side but its implementation would not have been possible without the help of many generous personalities. I would like to extend my sincere gratitude towards all of them.

I thank Mr. Ansel Kurian, guest lecturer, my project guide, for his continuous support and supervision. I also thank Dr. Tojo Jose, Assistant Professor and Head of the department and other faculties of Department of Economics, T.M Jacob Memorial Govt. College Manimalakkunnu, Koothattukulam for their support and help. I am also grateful for the constant support of my classmates. The encouragement that I received from my parents was also important at every stage and I am thankful to them. I acknowledge every single person who has helped in the successful completion of this project.

ABSTRACT

This paper forecasts the movement of the Nifty 50 index using ARIMA model. The prediction of this index is most common and vital in the perspective of economics and business. Many research works have been carried out over the years to develop predictive models. The paper presents the process of building stock price predictive model using ARIMA Model. NIFTY 50 can be used for a variety of purposes such as benchmarking fund portfolios, launching of index funds, exchange traded funds (ETFs) and structured products. The index tracks the behaviour of a portfolio of "Blue-chip" companies, the largest and most liquid Indian securities and can be regarded as a true reflection of the Indian stock market. Published stock data obtained from yahoo finance is used with stock predictive model developed. Therefore, weekly data from 27th March 2017 up to 28th March 2022 (263 observations) is used for this study. The results obtained revealed that ARIMA model has high potential in short run prediction and will be helpful to investors in stock market.

CONTENTS

1. INTRODUCTION

Any investor will try to depend on a forecasting method that could guarantee easy profiting and minimize investment risk from the stock market. This stands as major motivating factor for researchers in evolving and developing new predictive models. The Nifty 50 is an indicator of the top 50 major companies on the NSE. Stock market price may be of opening price, lowest price, highest price, adjusted closing price and volume. A large number of methods have been used for NSE including AR (Autoregressive model), ARMA (Autoregressive Moving Average Model), ARIMA (Autoregressive Integrated Moving Average Model) and so on. But ARIMA is most widely used on among them.\

1.1. NIFTY 50

The NIFTY 50 index is National Stock Exchange of India's benchmark broad based stock market index for the Indian equity market. Full form of NIFTY is National Stock Exchange Fifty. It represents the weighted average of 50 Indian company stocks in 13 sectors and is one of the two main stock indices used in India, the other being the BSE Sensex. NIFTY 50 is also one of the world's most actively traded contract. In the United States, the term Nifty Fifty was an informal designation for fifty popular large-cap stocks on the New York Stock Exchange in the 1960s and 1970s that were widely regarded as solid buy and hold growth stocks, or “Blue-chip” stocks.

1.2. OBJECTIVE OF THE STUDY

The objective of our study is to forecast the weekly closing price of NSE NIFTY 50 index using time series ARIMA Model for 2022-2023.

2. REVIEW OF LITERATURE

The major works using ARIMA model in the study of stock market data are reviewed. Box et al. (1970) proposed the Autoregressive Integrated Moving Average model using stationary concept for the forecast purpose Pankratz (2008) studied are helpful to the proper construction of univariate cases for Box-Jenkins forecasting method in various fields. Mohamed Ashik et al (2017) applied ARIMA model to forecast the upcoming daily closing stock price of Nifty50.

Rachlin Gil et al. (2007) "ADMIRAL: A data mining based financial trading system." Computational Intelligence and Data Mining. CIDM 2007 specified that Machine learning methodology (Regression, RBFN, SOM, BN, and SVM) is better but also an old way to Forecasting about uncertain market stock prices with the use of monotonous patterns. Author underwent a detailed examination of each and every methodology of prediction about crime, hospital patient arrival etc. They found pros and cons with different methods and finally proposed Back Propagation Neural Network as a technique to forecast the future values as well as to have an utmost financial gain for the holder.

Somani, Poonam, Shreyas Talele, and Suraj Sawant (2014) "Stock market prediction using Hidden Markov Model". "Information Technology and Artificial Intelligence Conference (ITAIC), 2014 IEEE, 7th Joint International, predicts the stock with Markov chain model and found that it is more efficient than neural network and support vectors until the dataset is too large. In our case, we have to take a sufficiently large dataset to analyse the share market.

Banerjee D. (2014) applied ARIMA model to forecast in Indian Stock exchange the future stock indices. Paulo Rotela Ju-nior et al. (2014) described ARIMA model to obtain short-term forecasts to minimize prediction errors for the Bovespa Stock Index. Renhao Jin et al. (2015) used ARIMA model to predict in Shanghai Composite Stock Price Index. All the studies were based on closing stock price.

3. DATA & METHODOLOGY

3.1. DATA AND VARIABLES USED IN THE STUDY

Here we have taken the data from yahoo finance(https://finance.yahoo.com/). Our data consist of Nifty 50 index historical data about five years (27th March 2017- 28st March 2022). We have taken only date and close value for our work. The summary statistics of the NSE Nifty Fifty data for the analysis period is tabled below. Summary statistics, using the observations from 27th March 2017 – 28th March2019 for the variable 'Close' (263 valid observations). Table (1) shows the summary statistics of the data and table is given below

Nifty 50 weekly closing price
Min. :8084
1st Qu. :10475
Median :11282
Mean :12211
3rd Qu. :13758
Max. :18339

Table (1): Summary Statistics of Nifty 50 weekly closing price

Figure (1) shows the graph of the raw data that we had taken and have general overview whether the time series is stationary or not and it can be seen that time series is not stationary.

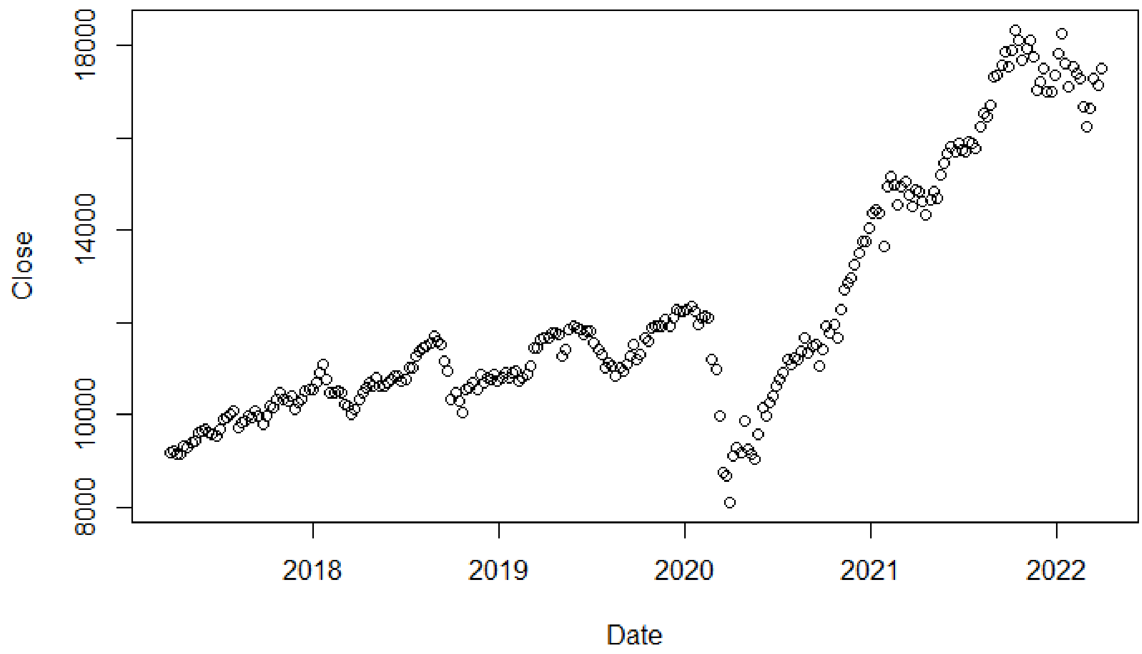

Figure (1): Graph of the Nifty 50 Price Index,

3.2. ARIMA MODEL- AN OVERVIEW

3.2.1. ARIMA

ARIMA (autoregressive integrated moving average) is a commonly used technique utilized to fit time series data and forecasting. It is a generalized version of ARMA (autoregressive moving average) process, where the ARMA process is applied for a differenced version of the data rather than original.

Three numbers p, d and q specify ARIMA model and the ARIMA model is said to be of order (p,d,q). Here p, d and q are the orders of AR part, difference and the MA part respectively. An ARIMA model is characterized by 3 terms: p, d, q

- p is the order of the AR term
- d is the number of differencing required to make the time series stationary
- q is the order of the MA term

AR and MA both are different techniques to for stationary time series data. ARMA (and ARIMA) is a combination of these two methods for better fit of the model.

The steps of building an ARIMA model will be explained. Finally, a demonstration using R will be presented on annexure.

3.2.2.AUTO REGRESSION (AR)

AR is a class of linear model where the variable of interest is regressed on its own lagged values. If $y_t$is modeled via AR process, it is written as

$$y_t = \delta + \emptyset_1 \, y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \epsilon_t$$

The above has a form similar to simple linear regression. It has an intercept like term (δ), regressors (y_{t-i}), parameters (ϕ_{t-i}) and an error term(ϵ). The only special thing is the regressors are the dependent variable's own lagged terms. If lag up to p is included in the model like above, the AR process is said to be of order p.

3.2.3.MOVING AVERAGE (MA)

MA is another class of linear model. In MA, the output or the variable of interest is modeled via its own imperfectly predicted values of current and previous times. It can be written in terms of error terms:

$$y_t = \mu + \theta_1 \, \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_p \varepsilon_{t-q} + \epsilon_t$$

Again, it has a form similar to classic linear regression. The regressors are the imperfections (errors) in predicting previous terms. Here the model is specified with positive sign for the parameters. It is not uncommon where we have negative sign for the parameters. The model above included errors for q lags and said to have an order of q.

3.2.4. STEPS TO MAKE ARIMA MODEL FOR TIME SERIES

If a process is ARIMA(p,d,q) then the differenced data is ARMA(p,q) process. The ARMA(p,q) process has the following mathematical form:

$$y_t = \delta + \{\emptyset_1 \, y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p}\} + \{\theta_1 \, \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_p \varepsilon_{t-q}\} + \varepsilon_t$$

$$y_t = \delta + \sum_{i=1}^{p} \emptyset_1 \, y_{t-i} + \sum_{j=1}^{q} \theta_1 \, \varepsilon_{t-j} + \varepsilon_t$$

This methodology comprises the following steps

3.2.5. IDENTIFICATION OF MODEL

It involves identifying three numbers p, d, q which specifies ARIMA model.

- First, we determine the value of d. We first check the stationarity of the data. If the data is stationary, then d=0. If the data is trendy then we take the first difference and check for stationarity again. We keep taking differences until we get a stationary output. The number of differences required is the number d.
- There are different tools to detect stationarity. Visual observation of the data plot, auto correlation, variogram are the most common.
- After determining d, we can utilize sample PACF (partial auto correlation) to get the AR order p. If the PACF cuts off after some lags, that number is the order of AR.
- We can determine the MA order q by looking at the sample ACF (auto correlation) of the differenced data. If the ACF cuts off after some lags, that number is the order of MA.

Theoretically, ACF of PACF cut off if one of the AR or MA orders is zero. If both of them are non-zero, then both ACF and PACF may exhibit damped sinusoid and/or exponential decay. More works to do in these cases to identify the model specifications.

3.2.6. DIVISION OF DATA

This stage involves dividing data into two parts i.e., train and test. Train data consist of 80% of our observation in the data frame and other 20% is test data. Separating data into training and testing sets is an important part of evaluating data mining models. Typically, when you separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing. By using similar data for training and testing, you can minimize the effects of data discrepancies and better understand the characteristics of the model. The procedure is appropriate when there is a sufficiently large dataset available. The train-test split procedure is used to estimate the performance of econometric models when they are used to make forecasts on existing data and thereby validating the model using various parameters such as RMSE (Root Mean Square Error), MAE (Mean Absolute Error), Mean Absolute Percentage Error (MAPE), etc. Model validation refers to the process of confirming that the model actually achieves its intended purpose. In most situations, this will involve confirmation that the model is predictive under the conditions of its intended use. A model has been processed by using the training set, you test the model by making predictions against the test set. Because the data in the testing set already contains known values for the attribute that you want to predict, it is easy to determine whether the model's guesses are correct

3.2.7. FORECASTING MODEL

It means prediction of values of a variable based on identified past values of that variable or other associated variables. Forecasting may also be based on expert judgments, which in turn are based on chronological data and experience. When model selected is found satisfactory during the analysis, it can be used for forecasting purpose. ARIMA model uses the historic data and decomposes it into AR (Auto Regressive) – indicates weighted moving average over past observations, integrated (I) –indicates linear trends or polynomial trend and moving average (MA) –Indicates weighted moving average over past errors. As such it has three model parameters AR (p), I(d) and MA(q) all combined to forming ARIMA (p,d,q) model where p represents order of auto correlation, d represents order of integration (differencing) and q represents order of moving averages.

3.2.8. ERROR RATE

Error rate is used to measure of forecast precision technique in statistics. Forecast error is a measure of how accurate our forecast was in a given time period. In this paper, consider about the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE); Y_t – Actual Value; F_t – Forecast Value.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Yt - Ft)^2} \qquad \text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Yt-Ft}{Yt}\right| \times 100$$

3.3. AUGMENTED DICKY FULLER TEST (ADF)

Augmented Dickey-Fuller test (ADF) is a common statistical test used to test whether a given time series is stationary or not. It is one of the most commonly used statistical test when it comes to analysing the stationary of a series.

In general, a p-value of less than 5% means you can reject the null hypothesis that there is a unit root. You can also compare the calculated Dickey-Fuller test statistic with a tabulated critical value. If the Dickey-Fuller test statistic is more negative than the table value, reject the null hypothesis of a unit root.

In Statistics and Econometrics, an augmented Dickey–Fuller test (ADF) the null hypothesis that a unit root is present in a time series sample. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity. It is an augmented version of the Dickey-Fuller test for a larger and more complicated set of time series models.

The augmented Dickey–Fuller (ADF) statistic, used in the test, is a negative number. The more negative it is, the stronger the rejection of the hypothesis that there is a unit root at some level of confidence.

An ARIMA Model is a generalisation of an ARMA model in time series analysis. These models are set in time series data to predict future points in the series. Such models are applied in cases where data is non-stationery where in differencing can be done to reduce the non-stationarity. Non-seasonal ARIMA models are generally denoted ARIMA (p, d, q) where parameters are non- negative integers then p, d, q refer to the autoregressive, differencing, and moving average terms for the component of the ARIMA model.

4. EMPIRICAL ANALYSIS AND INTERPRETATION

Here we had plotted the time series data which we have converted during the process. Both figure (1) and figure (2), X axis represents trading years and Y axis represents stock index price. Most of our graphs represents the same parameter in the X-axis and Y-axis.

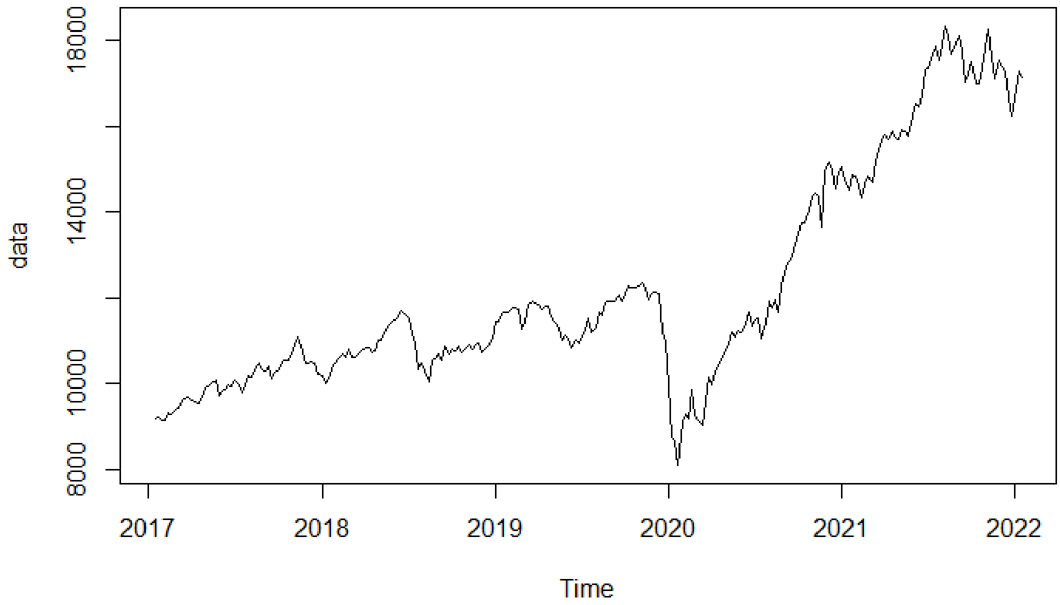

Figure (2): Graph of the time series data

Then we had checked Augmented Dicky Fuller test (ADF) to the time series data which we had converted and result of the test given below:

ADF Test variables	**Result**		
	Dickey-Fuller	**Lag order**	**p-value**
	-1.6439	**6**	**0.7258**

Table (2): RESULTS OF ADF TEST VARIABLES

Here p-value is greater than 0.5, so the time series data is non-stationary. The ACF the graph, it is seen that ACF dies down slowly which simply means that the time series is non-stationary to stationary. Figure (3) shows the ACF and figure (4) shows the PACF diagrams of our time series data.

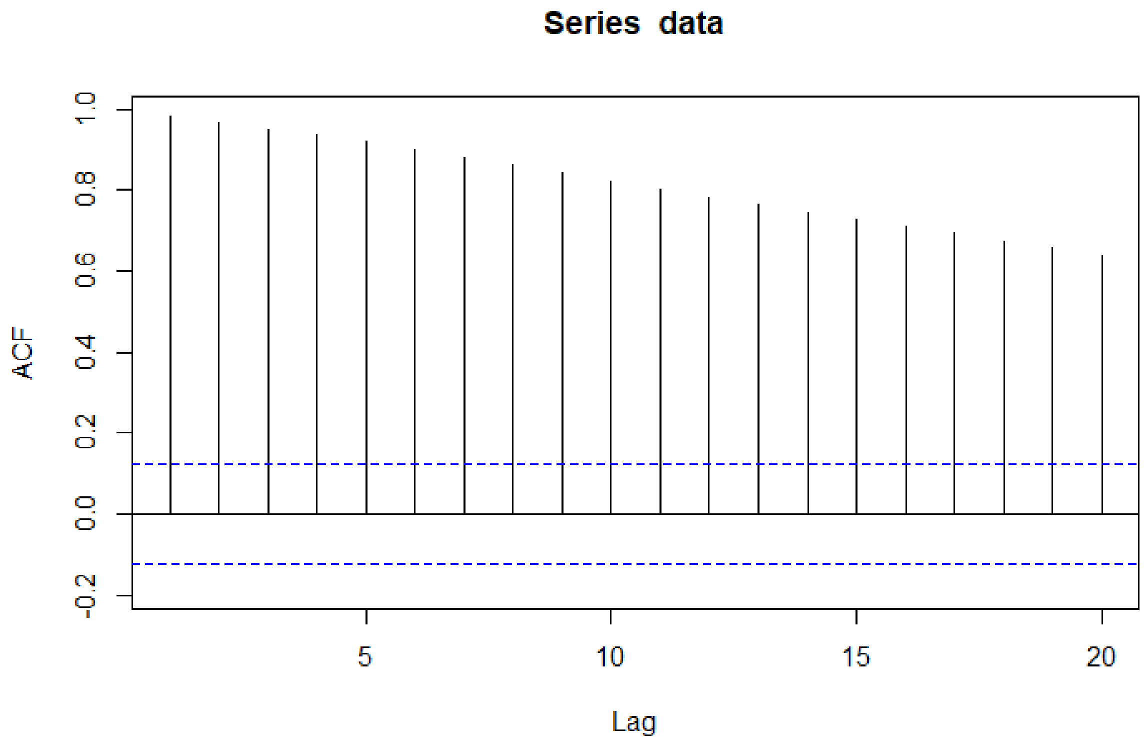

Figure (3): The ACF diagram of Nifty 50 Stock Price Index

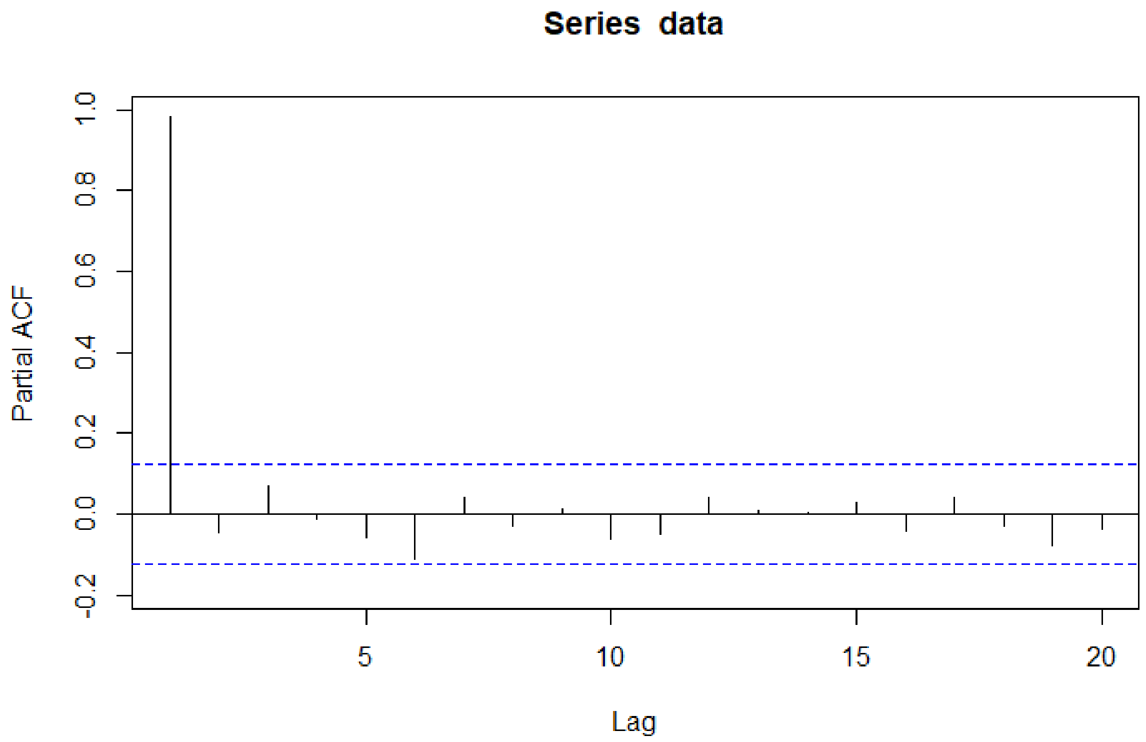

Figure (4): The PACF diagram of Nifty 50 Stock Price Index

When series is non-stationary, it is converted to a stationary series by taking the first difference. Then we had ADF test after taking the first difference and result of ADF test is given below:

ADF Test variables	**Result**		
	Dickey-Fuller	**Lag order**	**p-value**
	-6.042	**6**	**0.01**

Table (3): Results of ADF test variables

Here p-value is less than 0.5, so it is now stationary time series data. Figure (5) represent the stationary time series data.

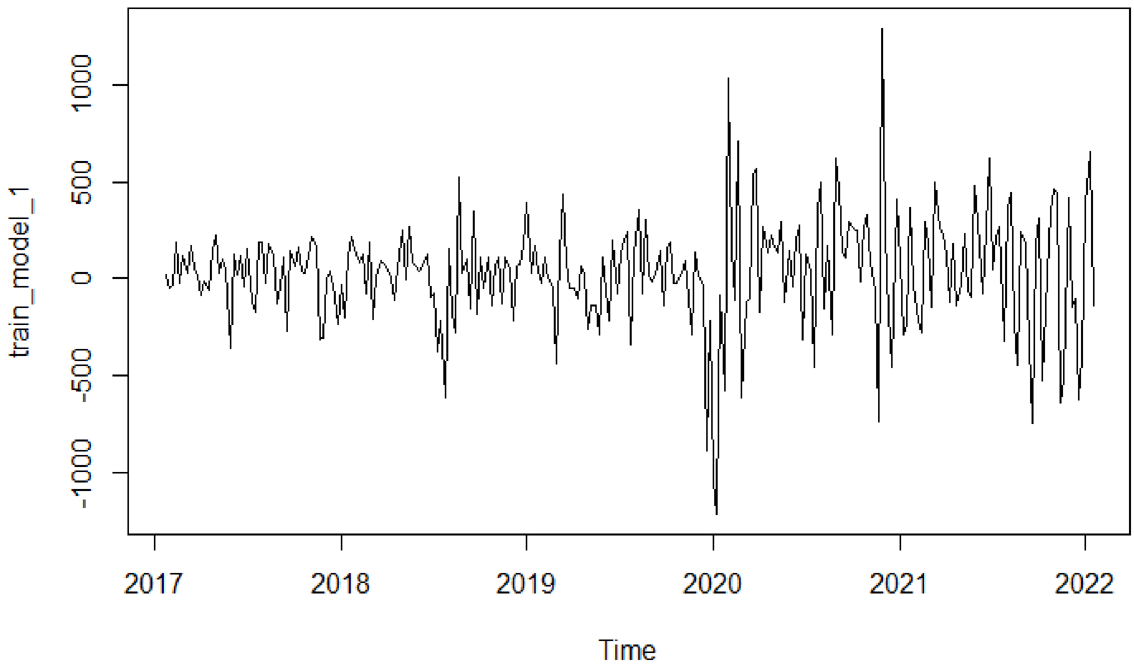

Figure (5): Graph of the NIFTY 50 stock price index after first differencing

The basic idea of ARIMA model is to view the data sequence as formed by a Stochastic Process on time. When the model has been identified, model can be used to estimate the future value based on the past and present value of the time series. Based on the identification rules on time series, the corresponding model can be established. If a partial correlation function of a stationary sequence is truncated, and auto-correlation function is tailed, it can be concluded the sequences for AR model; if partial correlation function of a stationary sequence is tailed, and the auto-correlation function is truncated, it can be strong that the MA model can be fitted for the sequence. If the partial correlation function of a stationary sequence and the autocorrelation function are tailed, then the ARMA model is appropriate for the sequence.

We had found out the p and q value from the PACF and ACF graphs respectively. Graphs are given below, from that p = 6 and q = 6. Both p and q value are from the X-axis as the lag value cut off on 6. Already know that d = 1, because we had done first difference to make the time series data stationary.

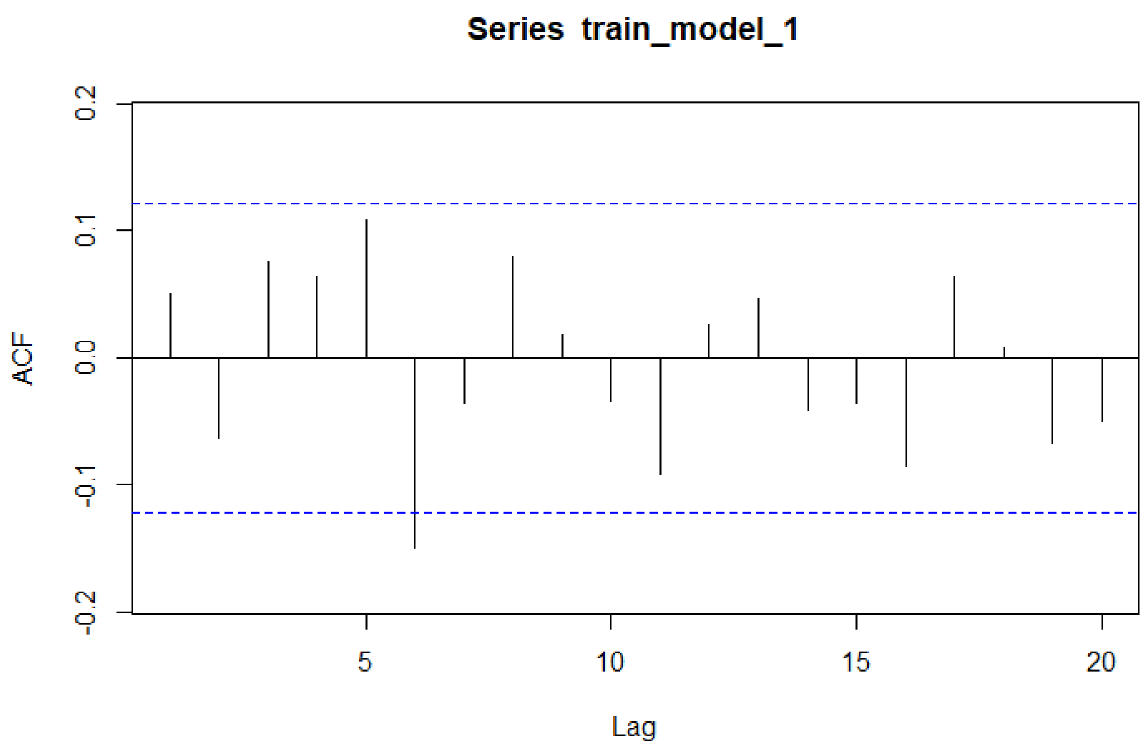

Figure (6): The ACF of Nifty 50 stock price index after first differencing

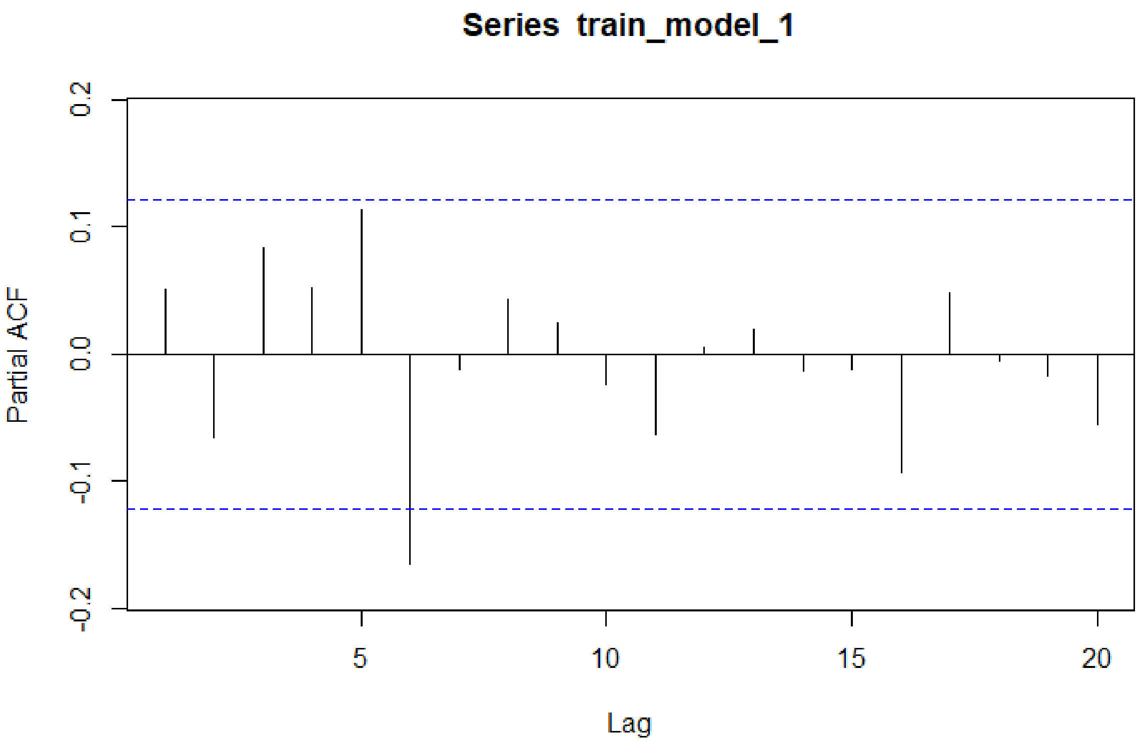

Figure (7): The PACF diagram of first order differencing of Nifty 50 closing stock price

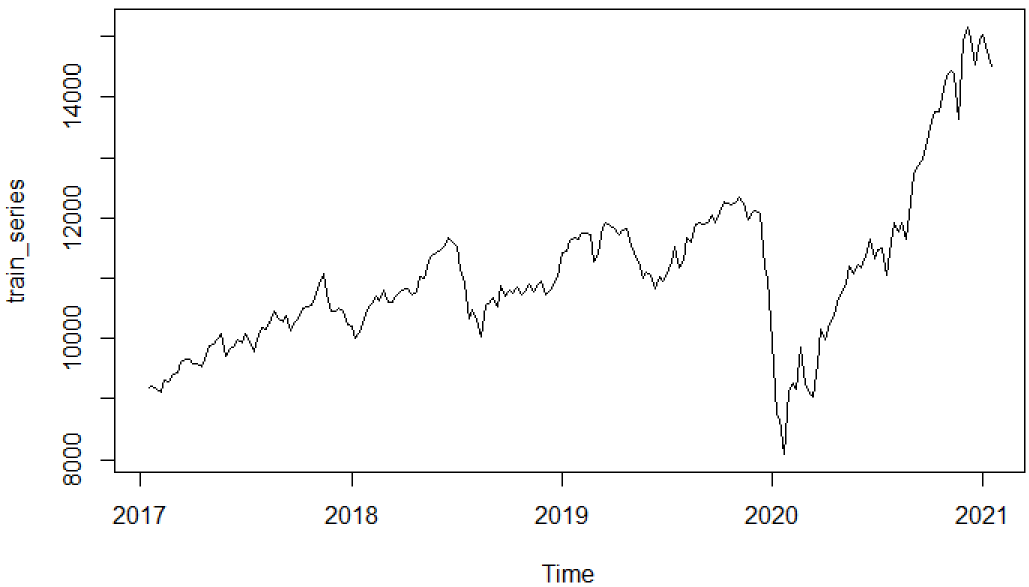

Figure (8): Graph of the train series data

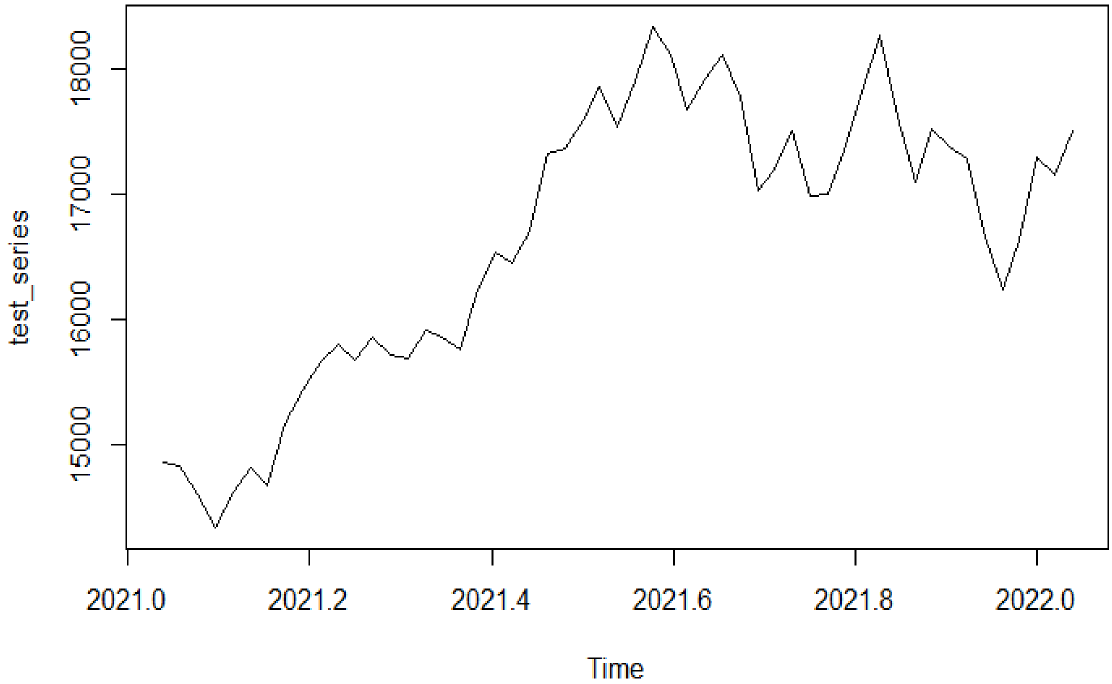

Figure (9): Graph of the test series data

The model verification is done by checking the residuals of the model to observe whether they contain any systematic pattern which still can be removed to get better on the chosen ARIMA (6,1,6). This is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders. After model verification is done then ARIMA (6,1,6) is forecasted using the train data and then with whole time series data. Figure (10.1) represent the ARIMA (6,1,6) model which is the forecasted. Magnified view of the figure (10.1) is given in the figure (10.2) in which X-axis ranges from 2021-2022. Both figures had been forecasted using train series data. After that we made ARIMA (6,1,6) model using time series data where forecast is given in figure (11.1) and (11.2). Figure (11.1) represent the ARIMA (6,1,6) model which is the forecasted. Magnified view of the figure (11.2) is given in the figure (11.2) in which X-axis ranges from 2022-2023.

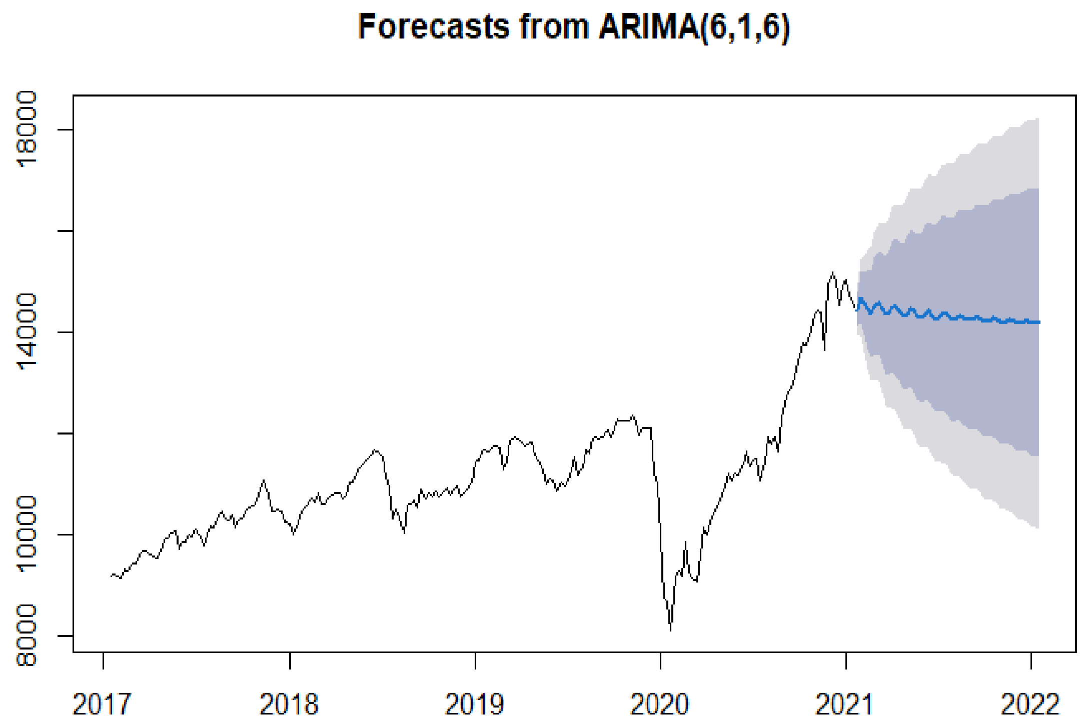

Figure (10.1): Forecast graph of ARIMA (6,1,6) using train model data

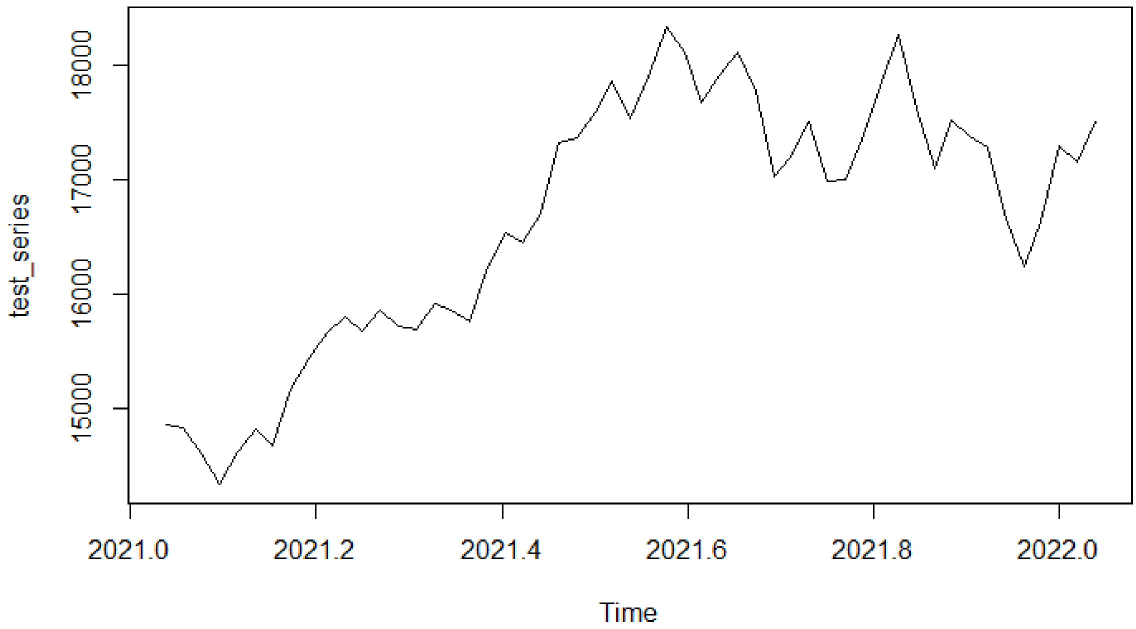

Figure (10.2): Forecast graph of ARIMA (6,1,6) using train model data which is magnified

MAE	2.374991e+03
RMSE	2.642072e+03
MSE	6.980545e+06
MAPE	1.382992e+01

Table (4): Error rates

Here the error rates such as RMSE and MAPE is very less so we can surely take the ARIMA (6,1,6) model. Then we had forecasted the ARIMA (6,1,6) model using the time series data and given below.

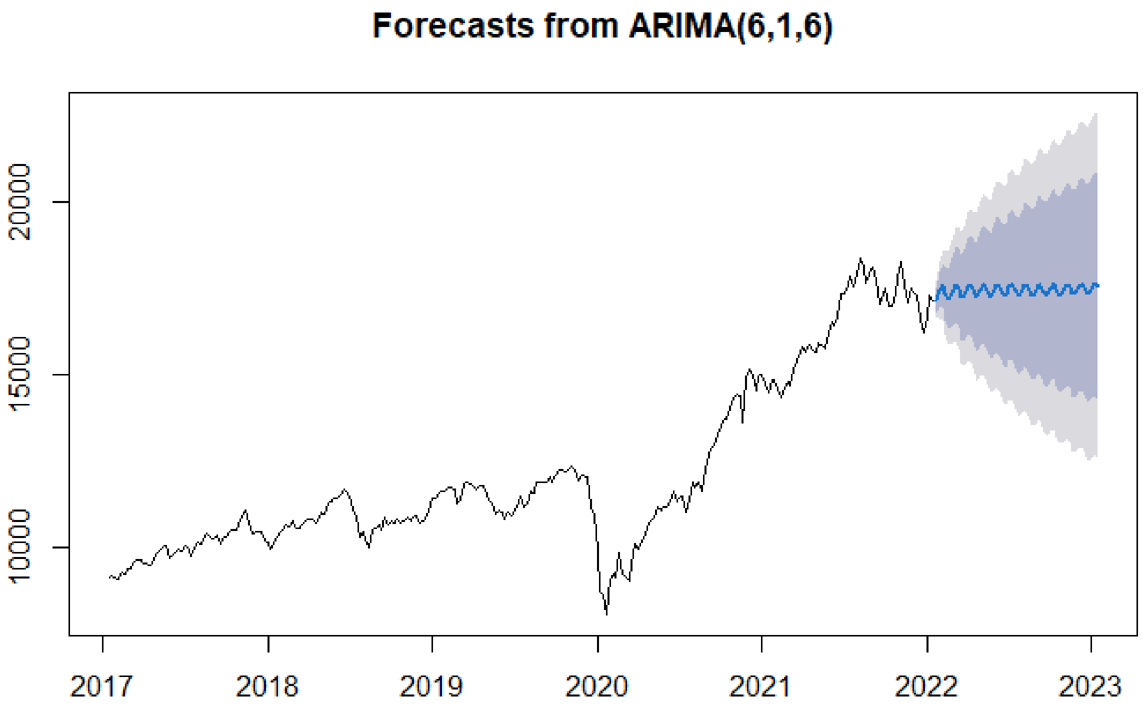

Figure (11.1): Forecast graph of ARIMA (6,1,6) using time series data

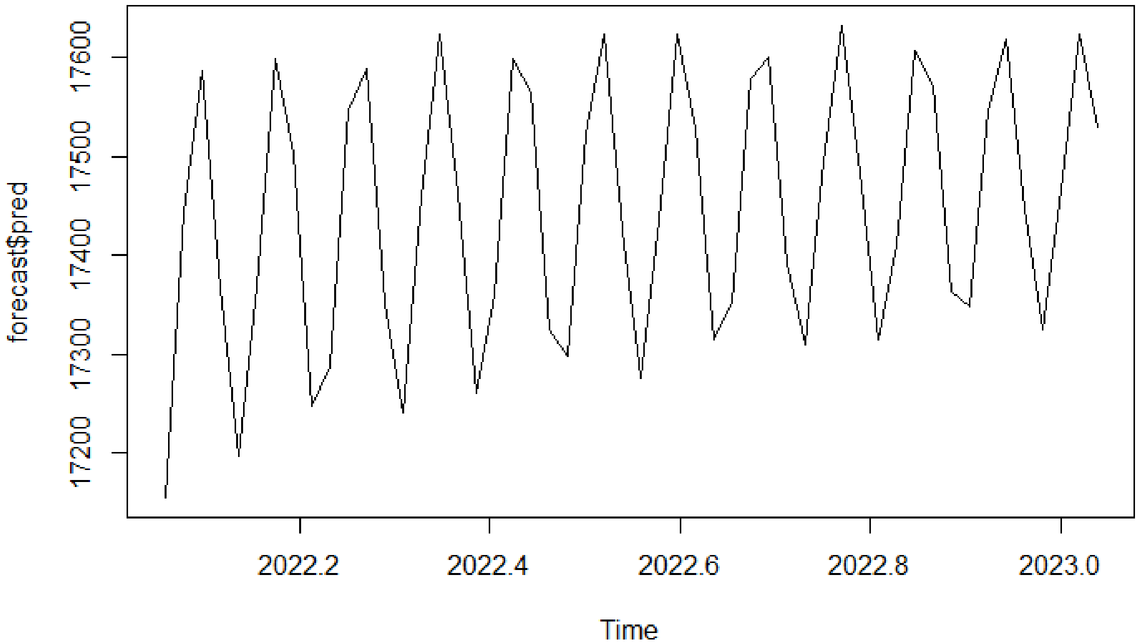

Figure (11.2): Forecast graph of ARIMA (6,1,6) using whole time series data which is magnified

5. CONCLUSION

The study focuses on forecasting future values for NIFTY 50 using its historical 5 years data (27th March 2017- 28th March 2022). Various statistical models are used to forecast future values in advance. ADF testing techniques are applied to obtain the p-values and check for stationarity of data. Differenced data set values have been taken to make future predictions as they have more stationarity than original data. All the trainings have been performed by dividing into train and test to ensure that data fits into the models and predict future values with accuracy. The best values of forecasting are achieved from interpreting the PACF and ACF after making the data stationary. After that, ARIMA (6, 1, 6) model is developed for analysing and forecasting of closing stock prices. Both RMSE and MAPE value is very small for this ARIMA (6,1,6) model. Further it can be concluded that this stock forecasting model can help the investors or retirees to know the performance trends of a particular share and helping them to invest in right shares with less loss and more profit. As individual share prices build the NIFTY 50 index, predicting the individual share prices of different organizations will be a great help for building a good portfolio for the investors.

6. REFERENCES

1. https://finance.yahoo.com/
2. https://en.wikipedia.org/wiki/Main_Page
3. https://www.nseindia.com/
4. D. Banerjee, “Forecasting of Indian stock market using time-series ARIMA model,” in Proc. Conference Paper, ICBIM-14, 2014.
5. A A Adebiyi and Charles Ayo (2014), Stock Price prediction using the ARIMA Model,2014 UK Sim-AMSS 16th International Conference on Computer Modelling and Simulation Research Gate
6. Somani, Poonam, Shreyas Talele, and Suraj Sawant (2014) "Stock market prediction using Hidden Markov Model". "Information Technology and Artificial Intelligence Conference (ITAIC), 2014 IEEE, 7th Joint International
7. Rachlin Gil et al. (2007) "ADMIRAL: A data mining based financial trading system." Computational Intelligence and Data Mining. CIDM 2007
8. Mohamed Ashik et al (2017) applied ARIMA model to forecast the upcoming daily closing stock price of Nifty50
9. Box et al. (1970) proposed the Autoregressive Integrated Moving Average model using stationary concept for the forecast purpose Pankratz (2008)
10. Deepak Kumar Gupta, Raj Mohan Singh, Vikalp Ravi Jain (2018) Analysis and Forecasting of Individual Stock Prices of Various Constituents in NIFTY 50
11. Riaz Khan (2017), ARIMA model for forecasting- Example in R

ANNEXURE 1

Time Series Data

Call:

Arima (x = data, order = c (6, 1, 6))

	AR1	AR2	AR3	AR4	AR5	AR6	MA1	MA2	MA3	MA4	MA5	MA6
Coefficients	0.1754	-0.1102	0.3079	0.1988	0.7645	-0.4509	-0.1138	0.1251	-0.2248	-0.1865	0.7398	0.2946
SE	0.4869	0.0893	NaN	0.1003	0.1890	0.5050	0.5187	0.0836	NaN	0.0788	0.1766	0.5118

sigma^2 estimated as 76632: log likelihood = -1832.09 aic = 3690.17

FORECASTED VALUES

Start = c (2022, 4)

End = c (2023, 3)

Frequency = 52

[1] 17154.76 17444.21 17587.11 17366.46 17197.25 17369.11 17598.86 17502.99 17247.03 17287.72 17546.90 17588.16

[13] 17351.61 17241.01 17458.41 17623.89 17462.57 17261.55 17360.33 17599.10 17564.62 17324.78 17298.72 17523.58

[25] 17623.20 17426.17 17276.88 17435.99 17623.14 17528.08 17314.56 17351.53 17578.43 17600.85 17391.66 17309.34

[37] 17494.49 17632.88 17484.65 17314.02 17411.23 17607.61 17571.66 17364.01 17347.73 17546.96 17618.93 17447.28

[49] 17324.81 17465.19 17624.41 17530.23

SUMMARY STATISTICS OF FORECASTED VALUES

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
17155	17342	17453	17446	17573	17633

ANNEXURE 2

- library(readr)
- library(dplyr)
- library(lubridate)
- library(reshape)
- library(reshape2)
- library(tidyr)
- library(tidyverse)
- library(bizdays)
- library(timeDate)
- library(ggplot2)
- library(plotly)
- library(timeSeries)
- library(fpp2)
- library(forecast)
- library(tseries)
- library(MASS)
- library(xts)
- library(data.table)
- library(tsibble)
- library(ggfortify)
- library(fable)
- library(zoo)
- library(bizdays)
- library(ggeasy)
- library(harrypotter)
- library(bizdays)
- library(timeDate)
- library(astsa)
- library(Metrics)
- library(quantmod)
- library(tseries)
- library(TSstudio)
- library(DMwR2)
- library(rminer)
- library(WDI)
- library(RJSONIO)

##To import Data set##

- df <- read.csv("C:/Users/user/Desktop/NIFTY_WEEKLY.csv")
- head(df)
- str(df)
- summary(df)
- class(df)
- df$Date<-as.Date(df$Date,format = "%d-%m-%Y")
- plot(df)

##Converting into time series data##

- data<- ts(df$Close, start =c(2017,03,27), end=c(2022,03,28),frequency= 52)
- plot(data)

##Stationarity test or ADF test##

- adf.test(data)
- Pacf(data,lag.max = 20)
- Acf(data,lag.max=20)
- train_model_1<- diff(data,difference=1,)
- train_model_1
- plot(train_model_1)
-

##Stationarity test or ADF test after taking first difference##

- adf.test(train_model_1)
- Pacf(train_model_1,lag.max = 20)
- Acf(train_model_1,lag.max=20)
- str(train_model_1)
- ts_info(train_model_1)
- ts_plot(train_model_1)

##Dividing time series data into train and test data##

- train_series = ts(data[1:209], start =c(2017,03,27), end=c(2021,03,22),frequency= 52)
- head(train_series)

- tail(train_series)
- plot(train_series)
- test_series = ts(df$Close[210:262], start =c(2021,03,29), end=c(2022,03,28),frequency= 52)
- head(test_series)
- df$Close[210]
- tail(test_series)
- tail(df)
- plot(test_series)
- str(train_series)

##Forecasting Arima model using train series data##

- arimaModel=arima(train_series, order=c(6,1,6))
- print(arimaModel)
- forecast=predict(arimaModel, 52)
- str(forecast)
- plot(forecast$pred)
- ts_info(forecast$pred)
- plot(forecast(arimaModel, h = 52))
- mmetric(test_series, forecast$pred, c("MAE","RMSE","MSE","MAPE"))

##Forecasting Arima model using time series data##

- arimaModel=arima(data, order=c(6,1,6))
- print(arimaModel)
- forecast=predict(arimaModel, 52)
- str(forecast)
- plot(forecast$pred)
- ts_info(forecast$pred)
- plot(forecast(arimaModel, h = 52))