FORECASTING THE STOCK PRICE OF AN AUTOMOBILE COMPANY USING ARIMA MODEL

Amirta. V[1], Swetha. C[2]
Department of statistics, S.D.N.B Vaishnav College for Women
Email: ***[1]amritavenkatraman8@gmail.com***, ***[2]nithyaswetha.c@gmail.com***

ABSTRACT

Interest in the prediction of stock prices has tremendously increased during recent times. This study deals with the prediction of an automobile company stock price using the ARIMA model. The opening stock price, high price, low price, closing price, adjusted closing price and the volume of shares traded was available. For the purpose of this analysis only the closing stock price was considered. The data is quantitative in nature and it is time series data. Exploratory data analysis and statistical tools were utilised for analysing the data. By applying ARIMA model of order (0,1,1) for the data, this study predicts the stock prices from 1st April 2021- 26th August 2021

KEYWORDS: *Time series analysis, ARIMA model, Stock price, Prediction*

INTRODUCTION

Stock market is a platform which helps the company to raise funds. The entire capital of a company is broken down into small units known as stocks. Stock market provides an interaction between buyers and sellers of stocks of a company. The movement of stock price is of utmost importance for both the company as well as the investors. It acts as an indicator of investor's confidence in the company. From the company's point of view it acts as a barometer of the financial health of the company. Proper analysis of stock prices helps the investor to make informed decisions before investing and also helps them in early identification of opportunities in the market **(Murphy, 2021)** [1]. Stock market is also an indicator of economic growth **(Ganesh and Bureau, 2009)** [2]. Sector wise analysis of the stock market helps to identify the key sectors and also sectors that require attention. The sector under consideration in this study is automobiles. In India, the automobile industry is one of the major sectors contributing to the growth of the country's Gross Domestic Product(GDP). In 2019 India was the fifth largest manufacturer across the world in the automobile industry and in 2026 it is expected that Indian automobile sector might reach Rs. 16.16-18.18 trillion. Automobile industries is one of the key sectors which accounts for 7.1% share in India's GDP and thereby has a significant contribution in the growth of the economy **(Automobile)** [3]. TVS motors is one among the top 10 automobile companies in India **(Top 10 Automobile Companies in India 2021, 2021)**[4] and is the third largest two wheeler company in India. It is the second largest exporter of automobiles in India **(TVS, 2021)** [5]. Thus, the performance of TVS motors is important. Study of the historical data of the company helps in forecasting the stock prices. In this study the historical data (April 2016 – March 2021) have been considered for analysis and the stock price from April 2021-August 2021 are forecasted.

LITERATURE REVIEW

In the following study ARIMA model is used for forecasting the stock prices. The evidence from literature shows the importances of forecasting and application on ARIMA in the field of finance. In today's time the word forecasting has become more significant as it is one of the keys for planning the future. Several techniques have been discovered to acquire an accurate forecast. One among them is the ARIMA model. The ARIMA model was developed in the 1930's - 1940's by Norbert Wiener et al . Later by developing systematic methods this was applied to business and economic data in the 1970's by Statisticians George Box and Gwilym Jenkins.

Various authors studied modelling and forecasting the financial sector. **(Almasarweh and Alwadi, 2018)** [6] focuses on predicting the banking stock market, **(Fattah, Ezzine, Aman, Moussami and Lachhab, 2018)** [7] worked on forecasting the demand in food companies. In the same way this paper focuses on forecasting the stock prices of an automobile company using the ARIMA model.

These are the authors who focussed on financial forecasts of the automobile sector using the ARIMA model. (**Shakti, Hassan, Zhenning and Caytiles, 2017) [8]** this paper forecasts the tractor sales of Mahindra tractor company and **(Edward and Manoj, 2016) [9**] worked on the forecast of close stock price for four automobile companies.

In this paper the Autoregressive Integrated Moving average [ARIMA(p,d,q)] model was used for modeling the movement of the stocks. This is because for short term forecasts ARIMA yields better results than any other complex models. ARIMA model is useful for predicting the future value of time series based on past values, lags and forecasted errors.

METHODOLOGY

The major objective of the study was to forecast the stock price of TVS motors data. The stock prices were collected over a period of time in chronological order. Hence the data is a time series data.

The models that can be used to analyse time series data are Autoregressive model, Moving average model, Autoregressive Integrated moving average models.

Autoregressive model:

Autoregressive models are used when the current value of the series can be expressed as a linear combination of 'p' past values and random error component in the same series. Mathematically, AR model can be written as:

$$Y_t = C + X_{t-1} + X_{t-2} + \cdots \dots . + X_{t-p} + e_t$$

If the series is dependent on past ‘p’ values, then the model is said to be of order ‘p’ and is denoted as AR(p). The simplest AR model is AR(0) which is white noise process. PACF is useful to identify the order of AR process

Moving Average model:

Moving average models can be used when the current value of the series can be expressed in the form of past forecast errors of the series. Mathematically, the Moving average model can be expressed as:

$$Y_t = C + e_t + e_{t-1} + e_{t-2} + \cdots \ldots \ldots + e_{t-q}$$

If the series is dependent on past ‘q’ error terms, then the model is said to be of order ‘q’ and is denoted as MA(q). ACF provides an idea about the order of the model.

ARIMA model

ARIMA model is useful for predicting the future value of time series based on past values, lags and forecasted errors. In the order of the model, ‘p’ represents the order of AR term, ‘q’ represents the order of MA term and ‘d’ represents the minimum number of differencing that is required to make the data stationary.

Mathematically the ARIMA (p,d,q) can be written as

$$Y_t = \mu + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots \ldots .. + \beta_p Y_{t-p} + \epsilon_t + \phi_1 \epsilon_{t-1} + \cdots \ldots .. + \phi_p \epsilon_{t-q}$$

Here, Yt represents the actual values, et is the random error component, mu, betai and phij are the coefficients, p and q are the order of autoregressive and moving averages respectively. In the order of the model, ‘p’ represents the order of AR term, ‘q’ represents the order of MA term and ‘d’ represents the minimum number of differencing that is required to make the data stationary.

DATASET DESCRIPTION:

For the purpose of this study the historical data of stock price was obtained from Yahoo! Finance **(TVSMOTOR.BO)** [10]**.** The dataset used for analysis contains date, opening price, high price, low price, closing price, adjusted closing price and volume of TVS stock price from 1st April 2016 to 31st March 2021.

Here, opening price is the price at the beginning of the day, high price represents the highest price of the stock during the day, low price is the lowest price of that stock during the day, closing price is the price of the stock at the end of the day, adjusted closing price is the price of the stock after amending for any corporate actions made by the company and volume is the number of stock traded during the day.

For the analysis, the closing stock price was considered as it can be used as a reference point for assessing the performance of the stock over a period of time. It also summarizes the movement of the stock during a particular day.

The data was pre-processed to check for missing observations. Graphs were used for visualizing the data. The period for which the forecast should be made was 100 days which is a very short period of time. Hence, the Autoregressive Integrated Moving average [ARIMA(p,d,q)] model was used for modeling the movement of the stocks.

The presence of auto-correlation in time series is observed by Auto correlation and Partial autocorrelation functions.The testing of stationarity of the time series is conducted using Phillips-Perron test and Augmented Dickey Fuller test.

R studio 3.6.1 version is used for analyzing the data and the packages used for the analysis are tseries, deseasonalize, seasonal and forecast.

ANALYSIS

The stock price of TVS motors was observed for 5 years. Here, the data is a univariate time series. The movement of the stock price was observed using basic data visualization (Fig 1)

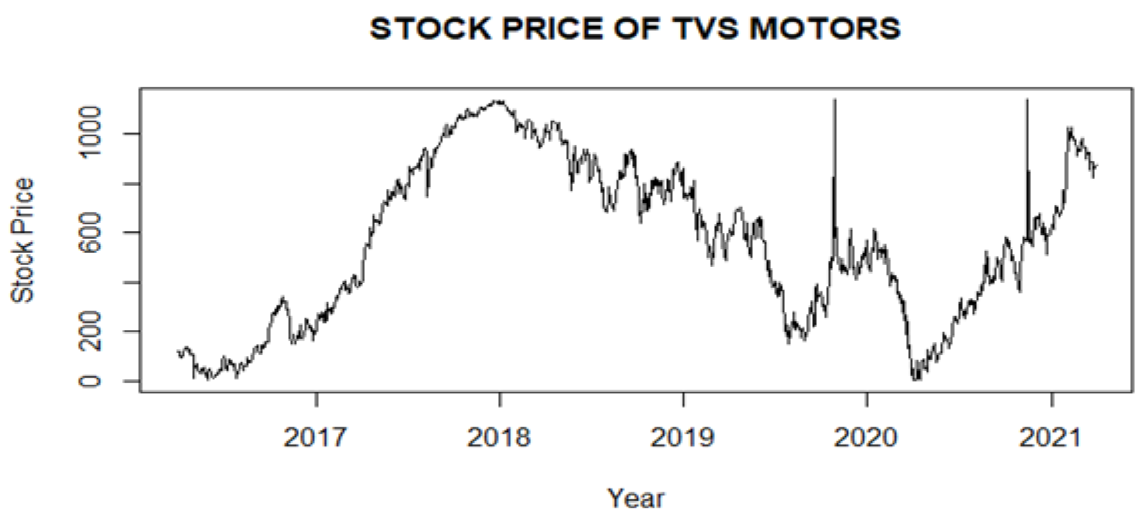

FIGURE 1 – STOCK PRICE OF TVS MOTORS FROM APRIL 2016-July 2021

Here the data is multiplicative in nature as the amplitude of increase or decrease in the value of the stock price is not constant. So, log transformation is applied on the data so that the data becomes additive in nature. Before fitting a model, exploratory data analysis must be performed in order to understand the general structure of the time series data. Decomposing the data separates the components of time series namely, seasonal, trend, cyclical and random error components from the data [Fig 2].

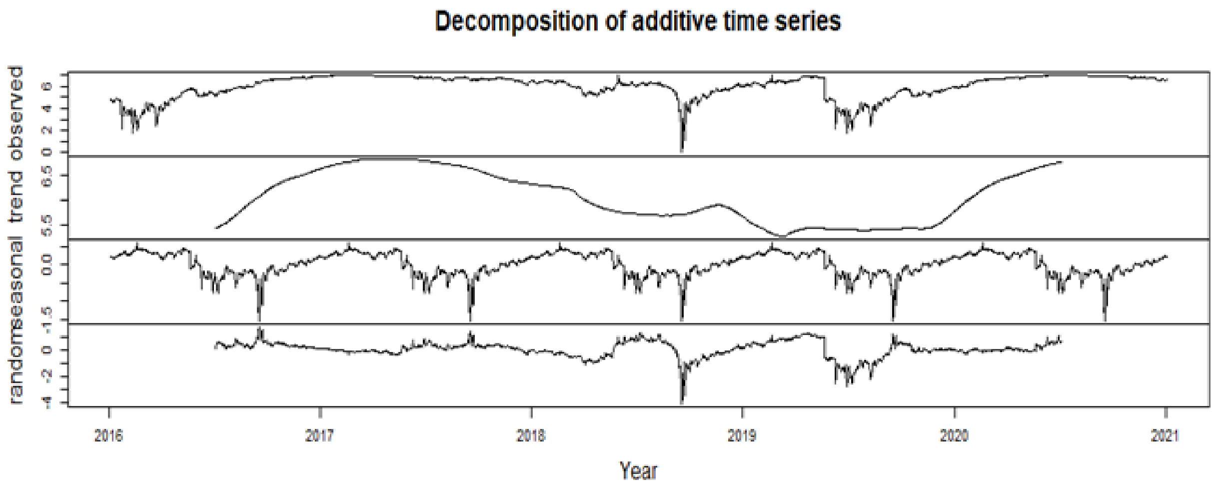

FIGURE 2 – DECOMPOSITION OF TIME SERIES DATA

Visualization of the components of time series is useful to identify the stationarity condition of time series. Scientifically, the stationarity assumption can be verified using ACF graph [Fig 3], Phillips-Perron test (PP test) and Augmented Dickey Fuller test (ADF test) [Table 1].

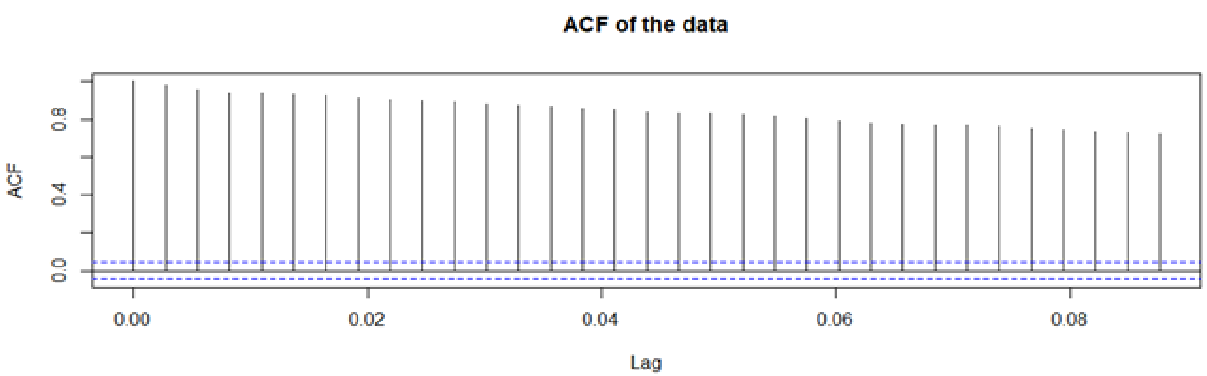

FIGURE 3 – ACF OF THE DATA

TEST	TEST STATISTIC	P VALUE	ALTERNATIVE	DECISION
PP test	-3.5417	0.03828	Data is stionary	Data is not stationary
ADF test	-2.8666	0.2114	Data is stionary	Data is not stationary

TABLE 1- OUTPUT OF PP TEST AND ADF TEST

The ACF does not exponentially converge to 0. Also, the p value is greater than 0.05. So, the data is not stationary.The seasonality can be removed from the data by subtracting the seasonal component from the original data. After subtracting and again stationarity reveals that the data is not stationary. So, the trend component of the time series is removed by differencing the data obtained after the seasonality component is removed. Observing the ACF and PACF of the data after differencing gives the conclusion that the data is stationary [Fig 4].

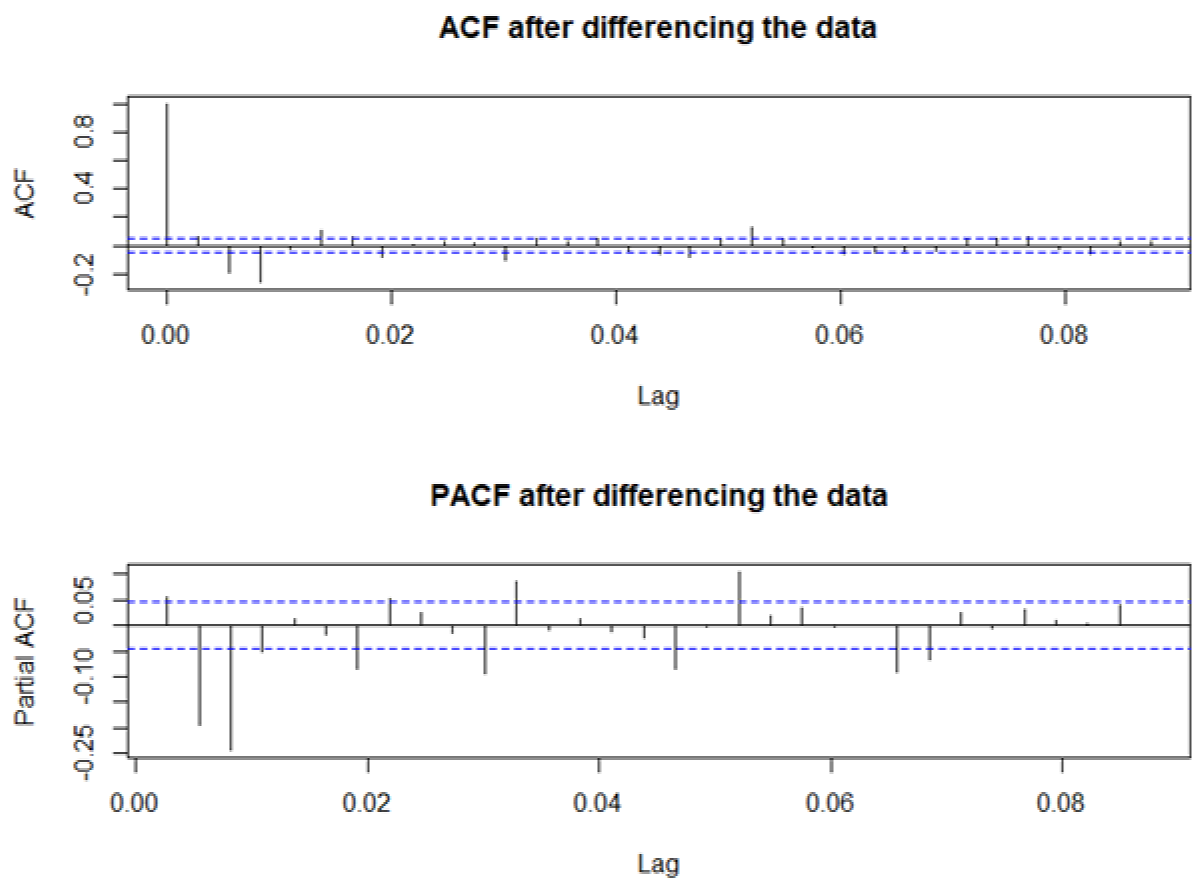

FIGURE 4 – ACF AND PACF OF THE DATA AFTER DIFFERENCING

Phillips Perron test and Augmented Dickey Fuller test gives the same conclusion.Since the data was stationary after differencing once, ARIMA model can be fitted for the data. The parameter p , d, q can be decided based on ACF, PACF and the minimum number of differencing required. Since the minimum number of differencing is 1, ‘d’ is 1. By observing the number of lags outside the control limits before converging to 0, ‘p’ and ‘q’ are approximately 2.

The optimal ARIMA model was estimated using **auto-arima()** function in R. ARIMA(0,1,1) was the model selected for further analysis [Fig 5].

```
Coefficients:
          ma1
      -0.2543
s.e.   0.0224

sigma^2 estimated as 1662:  log likelihood=-9351.07
AIC=18706.13   AICc=18706.14   BIC=18717.15
```

FIGURE 5 - OUTPUT OF AUTO ARIMA FUNCTION

The model equation can be written as**,**

$$Y_t = e_t - 0.2543e_{t-1}$$

<u>Diagnostic study</u>:

In order to check the efficiency of the fitted model, residual analysis was carried out. By carrying out residual analysis it is ensured that the residuals satisfy the assumptions. The assumptions made about the residual are there is no correlation between one and other observations, residuals follow normal distribution and residuals are homoscedastic. The residual plot of the forecast was plotted. The ACF and PACF of the residuals were plotted to ensure that the residuals of the fitted model do not show correlation and are independent of each other. Along with these plots the Quantile-Quantile (qq plot) plot was plotted for the residuals to ensure that these residuals follow normal distribution.

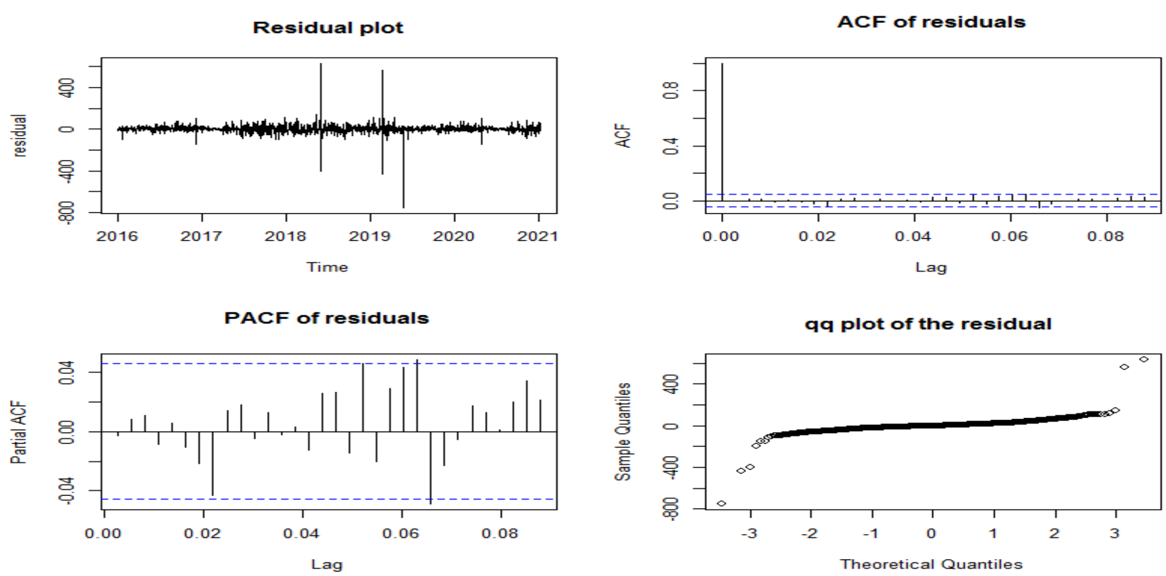

FIGURE 6- Residual analysis

From the above plots it can be observed that most of the points in the residual plot are near zero, this shows that the residuals of the fitted model are homoscedastic. In the ACF and PACF of residuals the lines at different lags do not cross the confidence interval which shows that the residuals are uncorrelated. From the qq plot of residuals it can be observed that the residual follows a normal distribution.

By summing up all the observations of the residual analysis it can be concluded that the model is a good fit. To further investigate the goodness of fitted model Box-Ljung test was carried out. According to the test the null hypothesis states that the model fitted doesn not show a lack of fit and the statement of alternative hypothesis is vice versa.

TEST	TEST STATISTIC	DEGREES OF FREEDOM	P VALUE	DECISION
Box - Ljung test	0.01687	1	0.8967	The model does not show a lack

TABLE 2 : OUTPUT OF Box-Ljung test

The above table 2 shows the output of the Box-Ljung test. The p value in the output is higher than 0.05 so we do not reject the null hypothesis. Thereby we can conclude that the fitted model does not show a lack of fit.

The forecasting of the best fitted model was done by using a function called ‘forecast’. The forecasting was done for the next 100 days from 01 April 2021 (Appendix A, Table A.1)
The forecasted data along with the historical data can be visualized as given in Fig 7.

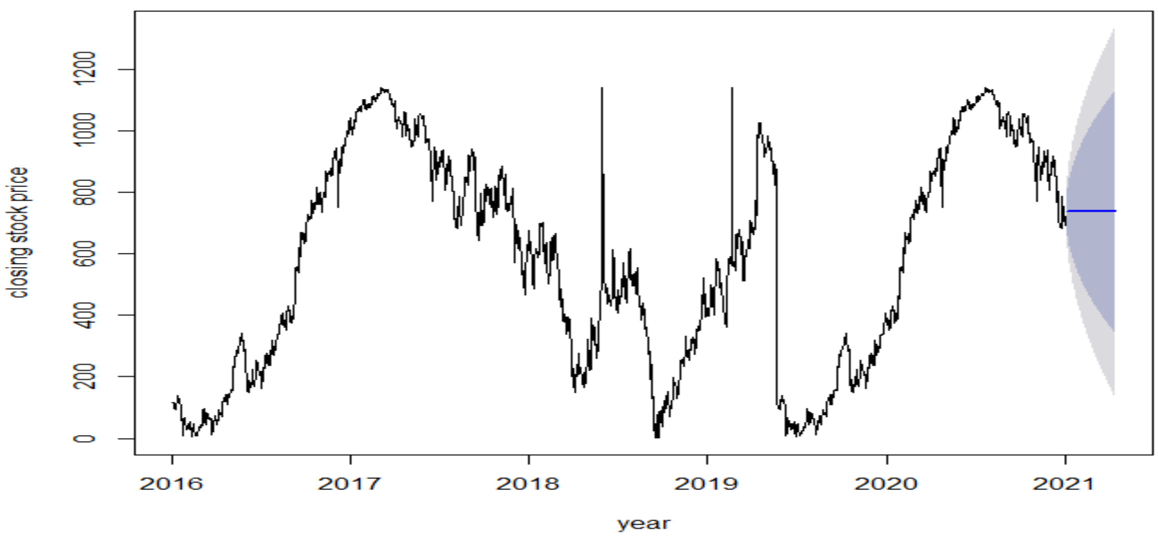

FIGURE 7 - FORECASTED DATA ALONG WITH ORIGINAL DATA

The reliability of the model can be assessed only by observing the biasness of the data. The difference in the actual stock price and the forecasted stock price from 1st April 2021 to 31st May 2021 was considered to observe the extent of bias (Appendix B). Since the model used in this study was very basic the extent of bias is high. A more advanced model using sophisticated techniques can predict the model in a better manner.

RESULTS AND CONCLUSIONS

From the point of view of investors and the company, forecasting the movement of stock is very vital for taking investment and management decisions respectively. Using the ARIMA model for forecasting came with a set of advantages. The major advantage is that it can be used to model and forecast large data. It can also be used when data is not stationary but can be transformed into stationary data. These were the reasons behind using the ARIMA model in this study. Before fitting the model and using it for forecasting the data, steps were taken to make the data stationary. Visualization of data gave the conclusion that the amount of increment present in the data is not constant. So the multiplicative model of time series data was converted into an additive model by taking the log transformation of the data. Decomposition of data helped to distinguish the trend, seasonal and random error component present in the data. As anticipated, the ACF graph, Phillips-Perron test and Augmented Dickey-Fuller test gave away the conclusion that the data is not stationary. Upon using 'deseasonalize' to remove seasonal components from the data and 'differencing' to remove trends in the data, the data was stationary. This was proved using the ACF graph, Phillips-Perron test and Augmented Dickey Fuller test. The Arima model of order (0,1,1) was the best fitted model for the data which was confirmed by the '**auto.arima()**' function. Residual analysis was carried out to check the credibility of the model and also to ensure that the model satisfies all the assumptions. Finally, the fitted model was used for forecasting. The stock price for the next 100 days was forecasted and was plotted in the graph along with the previous data.

ACKNOWLEDGEMENTS

We would like to express our gratitude to Dr. (Mrs) G. Vijayasree, Assistant Professor and Mrs.S. Deepa, Assistant Professor, Department of Statistics for their guidance and constant supervision.

REFERENCES

1. **Murphy, CB. (2021). Why do companies care about their stock prices, Investopedia, viewed 20 May 2021**

   https://www.investopedia.com/investing/why-do-companies-care-about-their-stock-prices/

2. **Ganesh, S. and Bureau, ET. (2009). Is the stock market an indicator? , The Economic Times, [online]. viewed 20 May 2021**

   https://economictimes.indiatimes.com/is-the-stock-market-an-indicator/articleshow/4541481.cms?from=mdr

3. **Automobile, MAKE IN INDIA, viewed 20 May 2021**

https://www.makeinindia.com/sector/automobiles

4. **Top 10 Automobile Companies in India 2021, INDIAN companies.in, viewed 20 May 2021**
https://indiancompanies.in/top-10-companies-in-india-automobile/#7_TVS_Motor_Company_Ltd

5. **TVS, (2021). TVS Overview, TVS Motor Company, viewed 20 May 2021**
https://www.tvsmotor.com/en/About-Us/Overview

6. Almasarweh, MS. and Alwadi, S. (2018). ARIMA Model in Predicting Banking Stock Market Data. ResearchGate, [online]. Available at:
https://www.researchgate.net/publication/328620139_ARIMA_Model_in_Predicting_Banking_Stock_Market_Data

7. Fattah, J., Ezzine, L., Aman, Z., Moussami, HE. and Lachhab, A. (2018). forecasting of demand using ARIMA model. SAGE journals, [online]. Available at:
https://journals.sagepub.com/doi/10.1177/1847979018808673

8. Shakti, SP., Hassan, MK., Zhenning, Caytiles, RD. and Iyengar, N.Ch.S.N.(2017). Annual Automobile Sales Prediction Using ARIMA Model. International Journal of Hybrid Information Technology Vol. 10, No. 6 (2017), pp.13-22, [online]. Available at:
https://gvpress.com/journals/IJHIT/vol10_no6/2.pdf

9. Edward, A. and Manoj, J,(2016). FORECAST MODEL USING ARIMA FOR STOCK PRICES OF AUTOMOBILE SECTOR. International Journal of Research in Finance and Marketing Volume 6, Issue 4 (April, 2016) , [online]. Available at:
http://euroasiapub.org/wp-content/uploads/2016/10/1FMApril-3374-1.pdf

10. **TVSMOTOR.BO. TVS Motor Company Limited. Yahoo finance, viewed 20 May 2021**

https://in.finance.yahoo.com/quote/TVSMOTOR.BO/history/

Appendix A

Table A. 1. Forecasted values

Date	Estimated , Closing stock price	Lo 80	Hi 80	Lo 95	Hi 95	Actual, Closing stock price	Bias
01-04-21	735.8305	683.5932	788.0678	655.9405	815.7205	589.1	-146.731
05-04-21	735.8305	670.6676	800.9934	636.1725	835.4885	584.2	-151.63
06-04-21	735.8305	659.9117	811.7493	619.7227	851.9383	570.25	-165.581
07-04-21	735.8305	650.501	821.16	605.3302	866.3308	569.45	-166.38
08-04-21	735.8305	642.0297	829.6313	592.3745	879.2865	564.25	-171.581
09-04-21	735.8305	634.2625	837.3985	580.4956	891.1654	568	-167.831
12-04-21	735.8305	627.0485	844.6125	569.4628	902.1982	540.15	-195.68
13-04-21	735.8305	620.284	851.377	559.1174	912.5436	547.95	-187.88
15-04-21	735.8305	613.8942	857.7668	549.345	922.316	542.35	-193.481
16-04-21	735.8305	607.823	863.838	540.0599	931.6011	550.35	-185.481
19-04-21	735.8305	602.0269	869.634	531.1956	940.4654	537.35	-198.481
20-04-21	735.8305	596.4718	875.1892	522.6997	948.9613	536.55	-199.281
22-04-21	735.8305	591.1297	880.5313	514.5297	957.1313	533.7	-202.13
23-04-21	735.8305	585.9779	885.6831	506.6508	965.0102	533.55	-202.281
26-04-21	735.8305	580.9975	890.6635	499.0339	972.6271	552.85	-182.981
27-04-21	735.8305	576.1724	895.4886	491.6545	980.0065	566.3	-169.531
28-04-21	735.8305	571.4888	900.1722	484.4916	987.1694	645.9	-89.9305
29-04-21	735.8305	566.9351	904.7259	477.5273	994.1337	633.5	-102.331
30-04-21	735.8305	562.501	909.16	470.7459	1000.915	630.95	-104.88
03-05-21	735.8305	558.1775	913.4835	464.1337	1007.527	616.4	-119.43
04-05-21	735.8305	553.9568	917.7042	457.6786	1013.982	608.55	-127.281
05-05-21	735.8305	549.8318	921.8292	451.37	1020.291	604.05	-131.781
06-05-21	735.8305	545.7963	925.8647	445.1983	1026.463	626.15	-109.68
07-05-21	735.8305	541.8448	929.8162	439.155	1032.506	618	-117.831
10-05-21	735.8305	537.9722	933.6888	433.2323	1038.429	611.75	-124.081
11-05-21	735.8305	534.1739	937.4871	427.4234	1044.238	624.7	-111.13
12-05-21	735.8305	530.4459	941.2151	421.7218	1049.939	614.25	-121.581
14-05-21	735.8305	526.7843	944.8767	416.1219	1055.539	603.15	-132.68

17-05-21	735.8305	523.1858	948.4752	410.6185	1061.043	617.1	-118.731
18-05-21	735.8305	519.6471	952.0139	405.2066	1066.454	640.1	-95.7305
19-05-21	735.8305	516.1655	955.4955	399.8819	1071.779	637.95	-97.8805
20-05-21	735.8305	512.7382	958.9228	394.6403	1077.021	627.45	-108.38
21-05-21	735.8305	509.3628	962.2982	389.478	1082.183	625.05	-110.781
24-05-21	735.8305	506.0369	965.6241	384.3915	1087.27	620.7	-115.13
25-05-21	735.8305	502.7585	968.9025	379.3776	1092.283	634.1	-101.731
26-05-21	735.8305	499.5255	972.1355	374.4332	1097.228	652	-83.8305
27-05-21	735.8305	496.3362	975.3248	369.5556	1102.105	650.3	-85.5305
28-05-21	735.8305	493.1888	978.4721	364.7421	1106.919	619.4	-116.43
31-05-21	735.8305	490.0818	981.5792	359.9903	1111.671	615.05	-120.781