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.2021.38943

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

Comparing Two ARIMA Models for Daily Stock Price Data

M. V. Narayana Murthi[1], Akasam Srinivasulu[2], V. Pavan Kumari[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: Analyzing the past data and planning for future is very important for every public and private organizational decisions. Now a days individuals also using forecasting methods to invest in Stock market. Investments in mutual funds and in registered companies in companies in stock market is the order of the day. In this paper, advanced forecasting methods are fitted to the time related stock price data to study its effectiveness in forecasting future events. Auto correlation and standard models have been analyzed before fitting this model to the above data. The forecasting can be done by using the ARIMA time series(using auto. arima) model. A particular reference have been made to Box and Jenkins approach for day to day stock price data values of Exxon Mobile Corporation from '1995-01-01 to 2020-03-01. With usual statistical software R. Here, ARIMA(1,1,1,) is fitted to this data, These results are compared with the model ARIMA(1,1,1,) by using accuracy measures.

Keywords:
ARIMA: Auto Regressive Integrated Moving Average
ACF: Auto Correlation Function
PACF: Partial Auto Correlation Function
AIC: Akaikae Information Criterion
RMSE: Root mean square error
XOM: Exxon Mobil Corporation

I. INTRODUCTION

A time series is a sequence where a metric is recorded over regular time intervals. Forecasting is the next step where you want to predict the future values the series is going to take. Forecasting a time series is often of tremendous commercial value. In most manufacturing companies, it drives the fundamental business planning, procurement and production activities. Any errors in the forecasts will ripple down throughout the supply chain or any business context for that matter. So it’s important to get the forecasts accurate in order to save on costs and is critical to success.

Not just in manufacturing, the techniques and concepts behind time series forecasting are applicable in any business. Forecasting a time series can be broadly divided into two types.

we use only the previous values of the time series to predict its future values, it is called Univariate Time Series Forecasting, and we use predictors other than the series (exogenous variables) to forecast it is called Multi Variate Time Series Forecasting.

Auto Regressive Integrated Moving Average, is a forecasting algorithm based on the idea that the information in the past values of the time series can alone be used to predict the future values.

II. REVIEW OF LITERATURE

Raymond Y.C. Tse (1997) suggested two questions must be answered to identify the data series in a time series analysis (1) Whether the data are random (2) The data have any trends . This followed by another three steps of model identification ,parameter estimation and testing for model validity. If the observations of time series are statistically dependent on each other then the arima is appropriate for time series analysis.

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’.

An Arima model based approach to seasonal adjustment by S.C.Hillmer, in this paper the researcher proposes a model-based procedure to decompose a time series uniquely into mutually independent additive seasonal, trend, and irregular noise components. The series is assumed to follow the Gaussian ARIMA model.

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

Forecasting Cotton Exports in India using the ARIMA model by Sudeshna Ghosh ,this study utilized Autoregressive Integrated Moving Average model to obtain the predictions of Indian cotton exports in the short run. The fttings are also compared with other two forecasting exercises. The results show that ARIMA has better forecasting results than other two models. The ARIMA models generate good forecast for recent future. Moreover the ARIMA model gives importance to the observations which are not very far off in the past. This study proposes the ARIMA (1,1,0) as the best model.

III. DATASOURCE, VARIABLES &METHODOLOGY

The purpose of the study is first check the Stationarity of the stock price data and forecast the closing prices of Exxon Mobil Corporation stock prices by using the fundamental time series model .

A. Variables

In Stock price data there are 6 columns but we are interested to forecast the closing prices of the Exxon Mobil Corporation stock prices..Here Closing prices are the target variable.

IV. METHODOLOGY

Auto Regressive Integrated Moving Average' is actually a class of models that 'explains' a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so this equation can be used to forecast future values. Any 'non-seasonal' time series that exhibits patterns and is not a random white noise can be modeled with ARIMA models.An ARIMA model is characterized by 3 terms: p, d, qwhere,p is the order of the AR term,q is the order of the MA term,d is the number of differencing required to make the time series stationary.If a time series, has seasonal patterns, then you need to add seasonal terms and it becomes SARIMA, short for 'Seasonal ARIMA'. More on that once we finish ARIMA.The first step to build an ARIMA model is to make the time series stationary.Because, term 'Auto Regressive' in ARIMA means it is linear regression model that uses its own lags as predictors. Linear regression models, work best when the predictors are not correlated and are independent of each other. The most common approach is to difference it. That is, subtract the previous value from the current value. Sometimes, depending on the complexity of the series, more than one differencing may be needed.The value of d, therefore, is the minimum number of differencing needed to make the series stationary. And if the time series is already stationary, then d = 0.'p' is the order of the 'Auto Regressive' (AR) term. It refers to the number of lags of Y to be used as predictors. And 'q' is the order of the 'Moving Average' (MA) term. It refers to the number of lagged forecast errors that should go into the ARIMA Model.

A pure AR model is one where Yt depends only on its own lags. That is, Yt is a function of the 'lags of Yt'.

$$Y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots\ldots\ldots\ldots\ldots + \beta_p\, y_{t-p} + €_t$$

PACF can be used for building AR models.
A pure **MA** is one where Yt depends only on the lagged forecast errors.

$$Y_t = \alpha + €_t + \phi_1 €_{t-1} + \phi_2 €_{t-2} + \ldots\ldots + \phi_q €_{t-q}$$

These are the AR and MA models respectively.
An ARIMA model is one where the time series was differenced at least once to make it stationary and combine the AR and the MA terms. So the equation is

$$Y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots\ldots + \beta_p y_{t-p} + €_t + \phi_1 €_{t-1} + \phi_2 €_{t-2} + \ldots.\phi_q €_{t-q}$$

A. ARIMA Model in Words

Predicted Yt = Constant + Linear combination Lags of Y (upto p lags) + Linear Combination of Lagged forecast errors (upto q lags)
The objective, therefore, is to identify the values of p, d and q. The purpose of differencing it to make the time series stationary. But you need to be careful to not over-difference the series. Because, an over differenced series may still be stationary, which in turn will affect the model parameters.The right order of differencing is the minimum differencing required to get a near-stationary series which roams around a defined mean and the ACF plot reaches to zero fairly quick. If the autocorrelations are positive for many number of lags (10 or more), then the series needs further differencing. On the other hand, if the lag 1 autocorrelation itself is too negative, then the series is probably over-differenced. In the event, you can't really decide between two orders of differencing, then go with the order that gives the least standard deviation in the differenced series.

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

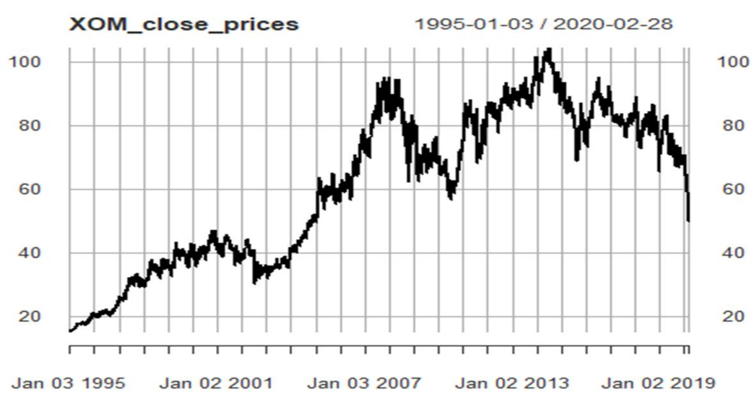
V. REVIEW OF RESULTS

1) By seeing the graph of the Closing prices of Exxon Mobil Corporation we observe that the series is non stationary.
2) By obserbing the graph of ACF and PACF there is an identification of stationarity in the data.
3) By augmented Dicky Fuller test the null and alternative hypothesis is defined as

Here H0: The data is stationary
H1: The data is not stationary
Here the value of P is greater than 0.05 so there is no evidence to reject the null hypothesis .
So the data is non stationary

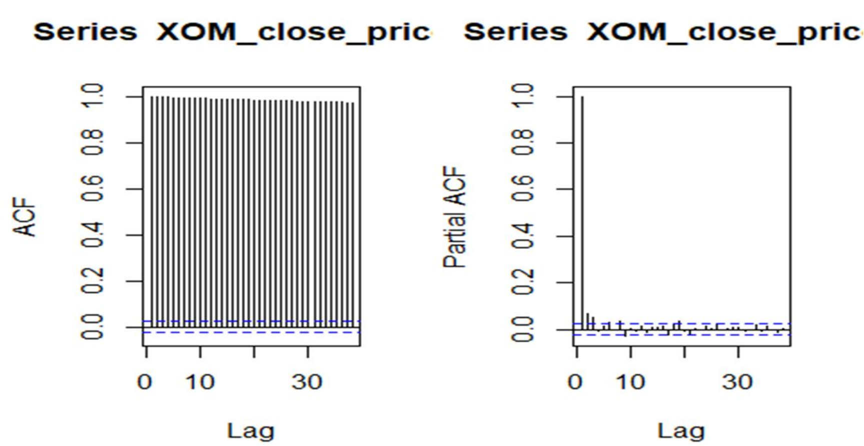
4) Fit ARIMA model by using auto.arima function . auto.arima fuction calculates the best arima model for this data ,for which model having the lowest AIC that model is the best model.For this data, here the order of AR is 0 & MA is 2,the difference required for converting non stationary data to stationary is 1.The AIC value for this model is 17082.15
5) Fit ARIMA(1,1,1) model to the data ,in this model the order of p,d,q is one.The AIC value for this model is 17086.75
6) Forecast the values of 10 time periods by using these two models ARIMA(0,1,2),ARIMA(1,1,1) .
7) Identify the best model by using the accuracy measures.The RMSE for two models is 0.9316196 and 0.9319578.so the smallest RMSE occurs at ARIMA(0,1,2).So ARIMA(0,1,2) is the best model to compare ARIMA(1,1,1).

```
plot(XOM_close_prices)
```



3.ACF&PACF plots

```
Acf(XOM_close_prices)
pacf(XOM_close_prices)
```



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

.ADF test for testing Stationarity

```
print(adf.test(XOM_close_prices))

##
##  Augmented Dickey-Fuller Test
##
## data:  XOM_close_prices
## Dickey-Fuller = -0.84025, Lag order = 18, p-value = 0.9581
## alternative hypothesis: stationary
```

5. arima(0,1,2) model

```
Fit1<-auto.arima(XOM_close_prices, seasonal=FALSE)

## Series: XOM_close_prices
## ARIMA(0,1,2)
##
## Coefficients:
##           ma1      ma2
##       -0.0995  -0.0612
## s.e.   0.0126   0.0129
##
## sigma^2 estimated as 0.8683:  log likelihood=-8538.08
## AIC=17082.15   AICc=17082.16   BIC=17102.41

Fit2= arima(XOM_close_prices, order = c(1,1,1))
fit4
##
## Call:
## arima(x = XOM_close_prices, order = c(1, 1, 1))
##
## Coefficients:
##          ar1      ma1
##       0.4338  -0.5371
## s.e.  0.0958   0.0899
##
## sigma^2 estimated as 0.8687:  log likelihood = -8540.37,  aic = 17086.75
```

```
To forecast the future Stock prices by using the function
term=10
fcast1 <-forecast(fit, h=term)
fcast1
##      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 6335       51.53763 50.34343 52.73184 49.71126 53.36401
## 6336       51.46589 49.85884 53.07294 49.00812 53.92366
## 6337       51.46589 49.57187 53.35990 48.56923 54.36254
## 6338       51.46589 49.32299 53.60878 48.18861 54.74316
## 6339       51.46589 49.10015 53.83162 47.84781 55.08397
## 6340       51.46589 48.89657 54.03521 47.53645 55.39532
## 6341       51.46589 48.70797 54.22380 47.24802 55.68375
## 6342       51.46589 48.53147 54.40030 46.97808 55.95369
```

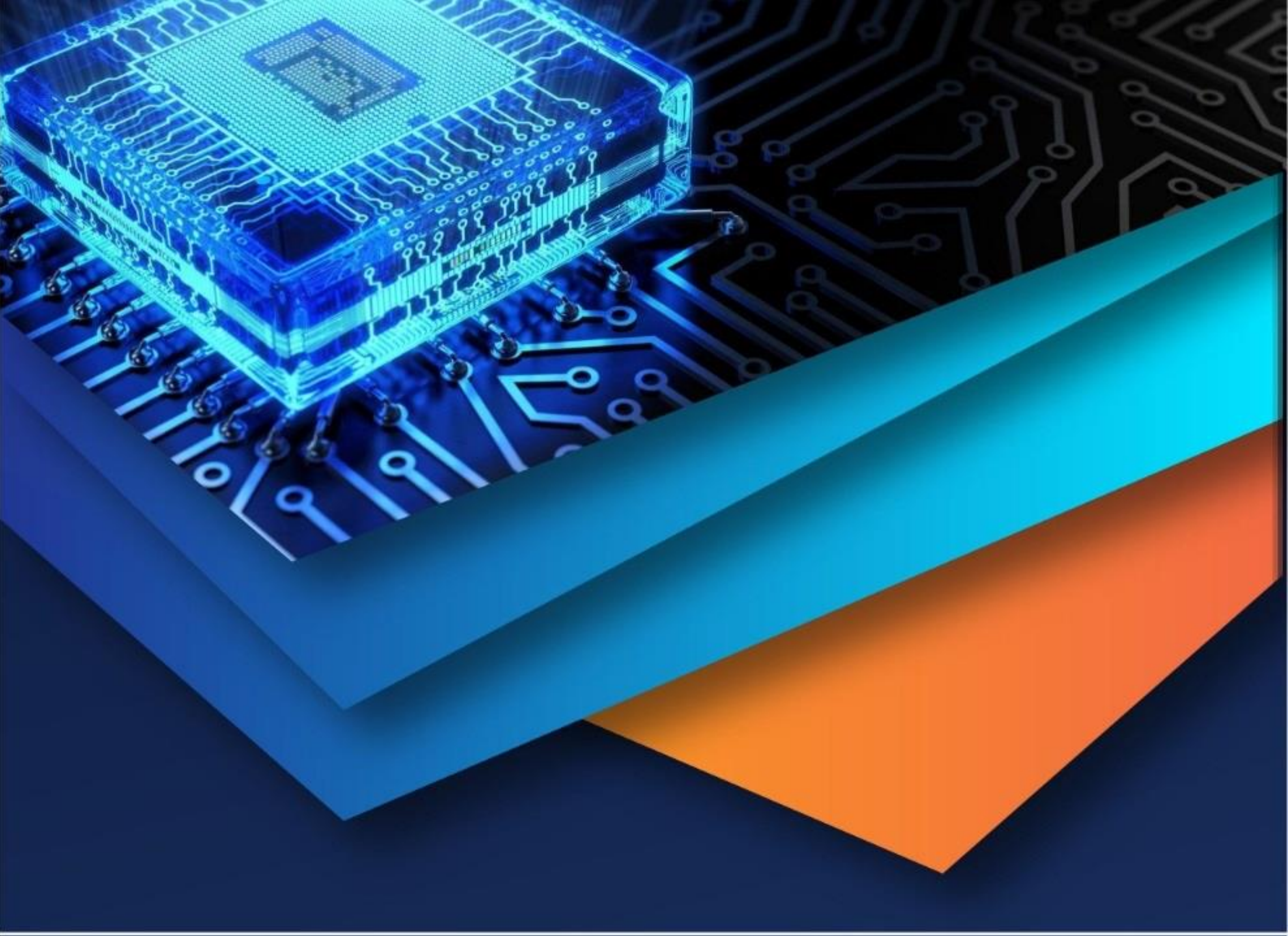
```
## 6343      51.46589 48.36500 54.56677 46.72349 56.20828
## 6344      51.46589 48.20702 54.72475 46.48188 56.44989

Fcast2 <- forecast(fit2, h=term)
fcast2
##     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 6335      51.54808 50.35363 52.74252 49.72133 53.37482
## 6336      51.59496 49.99063 53.19928 49.14135 54.04856
## 6337      51.61529 49.71549 53.51509 48.70980 54.52078
## 6338      51.62411 49.47984 53.76839 48.34473 54.90349
## 6339      51.62794 49.26856 53.98731 48.01959 55.23629
## 6340      51.62960 49.07483 54.18436 47.72242 55.53677
## 6341      51.63032 48.89476 54.36587 47.44664 55.81399
## 6342      51.63063 48.72579 54.53547 47.18805 56.07321
## 6343      51.63077 48.56608 54.69545 46.94374 56.31779
## 6344      51.63082 48.41429 54.84736 46.71155 56.55010

accuracy(fcast1)
##                   ME      RMSE       MAE        MPE     MAPE      MASE
## Training set 0.00683194 0.9316196 0.6301223 0.01002681 1.068281 0.9996295
##                      ACF1
## Training set -0.001030687

accuracy(fcast2)
##                    ME      RMSE       MAE        MPE     MAPE      MASE
## Training set 0.007045708 0.9319578 0.6300612 0.0104624 1.068072 0.9995326
##                    ACF1
## Training set 0.001707472
```

VI. CONCLUSION AND FUTURE RESEARCH WORK

By seeing the graph of Exxon Mobil Corporation closed prices we notice that the series is non stationary.And also by Augmented Dicky Fuller test the series must be non stationary. By using auto.arima function it shows ARIMA(0,1,2) which has the least AIC.This model is compared with the ARIMA(1,1,1) . Comparing these two models by using the accuracy measures. The model ARIMA(0,1,2) contains least RMSE.So ARIMA(0,1,2) is the best model. Future reasearch work

REFERENCES

[1] SHISKIN, J., YOUNG, A.H., and MUSGRAVE, J.C. (1967), "The X-l 1 Variant of Census Method II Seasonal Adjustment Program," Technical Paper 15, Bureau of the Census, U.S. Dept. of Commerce.
[2] YOUNG, A.H. (1968), "Linear Approximations to the Census and BLS Seasonal Adjustment Methods," Journal of the American Statistical Association, 63,1445.
[3] WALLIS, K.F. (1974), "Seasonal Adjustment and the Relations Be- tween Variables," Journal of the American Statistical Association, 69, 18. WHITTLE, P. (1963), Prediction and Regulation, New York: D. Van Nostrand.
[4] TIAO, G.C., BOX, G.E.P., and HAMMING, W. (1975), "A Statistical Analysis of the Los Angeles Ambient Carbon Monoxide Data 1955-1972," Journal of the Air Pollution Control Association, 25, 1130.
[5] TIAO, G.C., and HILLMER, S.C. (1978), "Some Consideration of Decomposition of a Time Series," Biometrika, 65, 497.
[6] PIERCE, D.A. (1978), "Seasonal Adjustment When Both Deterministic and Stochastic Seasonality Are Present," in Seasonal Analysis of Economic Time Series, ed. Arnold Zellner, U.S. Department of Com- merce, 242. (1980), "Data Revisions With Moving Average Seasonal Ad- justment Procedures," Journal of Econometrics, 14, 95.
[7] Arumugam, P., & Anithakumari, V. (2013). “Fuzzy Time Series Method for Forecasting Taiwan Export Data.” International Journal of Engineering Trends and Technology, 4(8), 3342-3347.
[8] Adebayo, F. A., Sivasamy, R., & Shangodoyin, D. K. (2014). “Forecasting Stock Market Series with ARIMA Model.” Journal of Statistical and Econometric Methods, 3(3), 65-77.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY