Prediction of Stock prices in Oil Sectors using ARIMA Model

A.V. Jadhav[#1] K.B.Kamble[#2]
[#1] Associate Professor, N.E.S. Science College, Nanded, (M.S.) India,
[#2]Siddharth College of Commerce and Economics, Fort Mumbai, India

Abstract- *The paper investigates the trend of four specific prices on the Bombay Stock Exchange. These are the shares of public sector Oil companies HPCL, IOCL, BPCL, and ONGC. An Autoregressive Moving Average model has been used for modelling purposes. BSE to a certain extent reflect the degree of inflection which has a great signification on national economic by time series model Autoregressive .This model is a simple and practical model in financial time series analysis which has relatively high for forecast accuracy. The paper utilizes monthly secondary data from web site www.bseindia.com through the statistical analysis of BSE from the year of, January 2, 2012 to October, 12, 2017. ADF unit root test of Autocorrelation function ACF diagram and partial autocorrelation function PACF diagram to the prediction of model. The prediction of the models showed that the ARIMA model is valid and forecast accuracy is relative high.*

Keywords - ***BSE***- *Bombay Stock India,* ***ARIMA***- *Autoregressive Integrated Moving Average,* ***ACF***- *Autocorrelation Function,****PACF***- *Partial Autocorrelation Function,* ***HPCL***- *Hindustan petroleum Corporation Limited ,****IOCL***- *Indian Oil Corporation Limited,* ***BPCL***- *Bharat Petroleum Corporation Limited,* ***ONGC***- *Oil and Natural Gas corporation.*

Introduction: This model is suited to the time series analysis data for forecast feature point in the series. This model can be applied in some cases where integrated part of model can be applied to remove non stationary Box- Jenkins Methodology: Autocorrelation and Partial Autocorrelation functions. The different Box Jenkins models are identified by the number of Auto regression parameters (p), the degree of differencing (d) and number of moving averages parameter (q). Any such model can be written using uniform notation ARIMA (p, d, q) first for prediction the investigation of appropriate model type by looking at autocorrelation and partial autocorrelation .The sample autocorrelation coefficient (ACF) of log k is computed for (n-k) pairs.

$$Y_K = \frac{\Sigma(y_{t-1} - \bar{y})(y_t - \bar{y})}{\Sigma\Sigma(y_{t-1} - \bar{y})^2}$$

This quantity measures the linear relationship between the time series operations separated by a log of k time units. The autocorrelation coefficient is analysed to determine the appropriate order p of the model. The partial autocorrelation coefficient (PACF) of log k, denoted by φ_{kk} is measures of the correlation between $\mathbf{y_t}$ and $\mathbf{y_{t-k}}$ after adjusting for the presences of $\mathbf{y_{t-1}}$, $\mathbf{y_{t-2}}$,……. ,$\mathbf{y_{t-k+1}}$ One method of computing the partial autocorrelation of log k is to perform a regression of $\mathbf{y_t}$ on y_{t-1} through $\mathbf{y_{t-k}}$, using the resulting the coefficient of the $\mathbf{y_{t-k}}$ formed as the estimate of φ_{kk}. When φ_{kk} is graphed for $\mathbf{log_1}$, $\mathbf{log_2}$……. $\mathbf{log_k}$, the result is partial autocorrelation graph (PACF) of the series. If the ACF trails off and the PACF shows picks them an autoregressive (AR) model with order q equal to the number of significant PACF. Spikes are considered the 'best' model. If Autocorrelation function (ACF) and Partial Autocorrelation function trail of then an Autoregressive moving Averages (ARIMA).model is used with p and q equal to one. If the data had to be different for it to become stationary then the ARIMA model is used.

Literature Review:, One of the methods that commonly used for forecasting time series data is Autoregressive Integrated Moving Average (ARIMA). The ARIMA model was developed by George Box Dan Gwilym Jenkins and it is called time series method of Box-Jenkins. In their book, Box and Jenkins strongly recommend that forecasts of ARIMA processes be made using the Difference Equation form because it is the simplest approach. Defu Zhanga, et.al. (2) proposed a hybrid model FTSGA on Fuzzy time series and genetic algorithm FTSGA improved the performance by applying the operations of genetic algorithm such as selection, mutation and crossover to interactively find a good discourse partion TALEX is chosen as the experimental outcomes shows that comparing with other model based on Fuzzy time series FTSGA can significantly reduce the root mean square error and improved accuracy. This model can achieve more suitable partion of the universe which can improve the prediction result considerable.Jose Manue Azevede Rui Almena, Pedro Almenda in (3) proposed a model in which literature review of the use of data mining with time series data forecasting on short time stock forecast. Research is associated with the combined use of fundamental and technical indicators'. Preethi, B .Santhi in "stock market forecasting techniques survey' journal of theoretical and applied information technology.Asiri (2008) applied the Dicky fuller unit test and ARIMA model in (5) as well as exponential smoothing techniques, to measures performance Bahrain stock exchange

(BSE). Their result show evidence that stock return followed a random walk process no drift and trend. In another research mobark et al (2008) .In (4) investigated the return series on Bangladesh's Dhaka Stock exchange (DSE) to see if they are sovereign and resemble the RWH. They used both parametric and nonparametric test with daily data from 1988 to 2008. The result shows that the return did not trail the RWH and the important Auto-Correlation Coefficient at dissimilar lags rejected the weak from efficiency.

2. METHODOLOGY:

a) Time series modelling:

The model used in the study is the ARMA. To test for the stationary of data we used time series plot, (ACF) unit root test. After the stationary of the time series was attained, ACF and PACF of the stationary series are employed to select the order of the AR process and the order of the MA process of the ARIMA model. An account for stabilizing or making the data stationary. In practice, one or two level of differencing are often enough to reduce a time series to apparent stationary. The data used in this study are from the Oil companies from January 2, 2012 to October, 12, 2017 It is the monthly data of BSE stock Price Index.

Let X_t be a stationary time series with mean μ and variance σ^2, and assume for case of notation that t takes on integer values t = ±0, ±1, the auto covariance function of X_t at lag k is defined as:

X_t = **E(** $X_{t-}\mu)(X_{t-k}-\mu)$ **(1)**

The autocorrelation function at lag k is defined as:

$$\rho(k) = \frac{\gamma_{(k)}}{\gamma(0)} = \frac{\gamma(k)}{\sigma_x{}^2} \quad (2)$$

The partial autocorrelation denoted as φ_{kk} was obtained by substituting γ_k for $\hat{\gamma}_k$ by a recursive method given by Durbin (1960) as follows

$$\varphi_{k+1} = \frac{\gamma_{k+1-\sum_{j=1}^{k}\hat{\varphi}_{kj}\gamma_{k+1-j}}}{1-\sum_{j=1}^{k}\varphi_{kj}\gamma_{jk}} \quad \text{and}$$

$$\hat{\varphi}_{kj}\gamma_{k+1j} = \hat{\varphi}_{kj} - \hat{\varphi}_{k+1,k+1}\hat{\varphi}_{k,k+1-j} \quad (3)$$

b) Autoregressive Filter (AR (p))

The mathematical representation of the autoregressive model of order **p** is defined below:

$$X_{t=}\varphi_1 X_{t-1} - \varphi_2 X_{t-2} \cdots \varphi_p X_{t-p} + a_t \quad (4)$$

Where a_t the white noise, p is denote the order of the AR and denote φ_P the AR parameter.

c) Moving Average Filter (Ma (q))

The mathematical representation of the moving average model of order q denoted as MA (q) is given by:

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \quad (5)$$

where a_t is the white noise, q denote the order of the MA and θ_q denote the MA parameter's autoregressive Moving Average Filter(ARMA (P,Q)) .An ARMA (p,q) where p and q denotes the order of the autoregressive and moving average models respectively of stochastic process X_t is given.

$$X_t = \varphi_1 X_{t-1} - \varphi_2 X_{t-2} \cdots \varphi_P X_{t-p} - a_t - \theta_1 a_{a-1} - \theta_{t-2} \cdots - \theta_q \quad (6)$$

Where φ_P and θ_q are the AR and MA parameters **and** a_t **is the white noise.**

Autoregressive Integrated Moving Average Filter ARIMA (p, d, q)

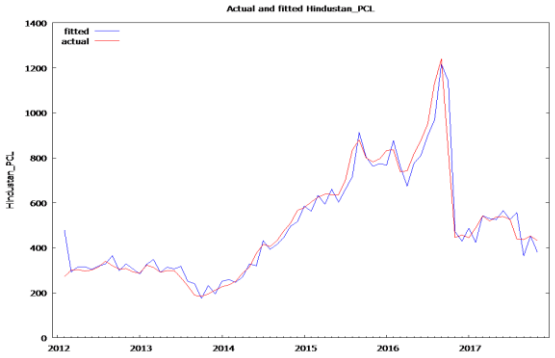
Box and Jenkins (1976) developed a methodology for fitting ARIMA models to different data. These are known as autoregressive integrated moving-average (ARIMA) Models. The ARIMA (P, d, q) where p denote the order of the AR, d denote the order of differencing and q denote the order of MA. The mathematical representation of ARIMA (p, d, q) model is given by

$$\nabla^d \varphi_{(B)} X_T = \theta_{(B)} a_t$$

Where ∇**= (1 -B), B** is the backward shift operator given as

$\mathbf{BX_t = X_{t-1}}$, φ *and* θ are the AR and MA parameters, respectively

Fig-1(a)HPCL



.Fig-1(b) OICL

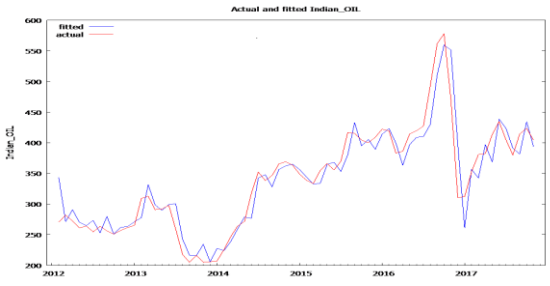


Fig.1(c) BPCL

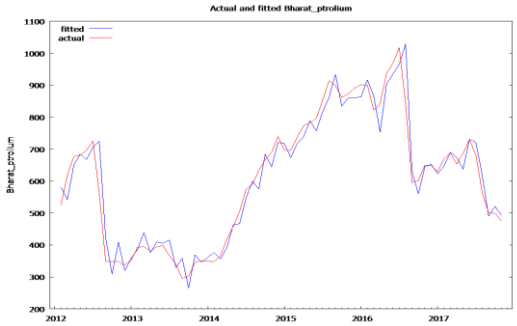


Fig.1(d) ONGC

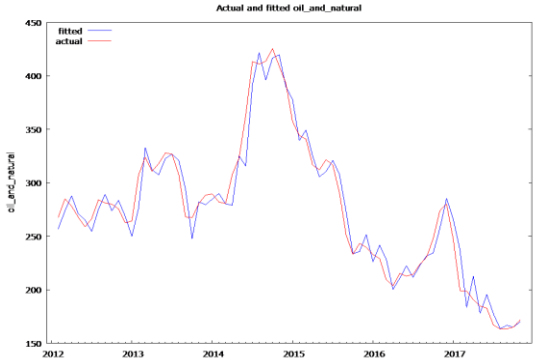


Fig.1-Graph showing the trend of the stock price actual and fitted.

Table.1: Summary Statistics Average stock price of four companies

Company	Mean	S.D.	C.V.	Skewness	Kurtosis
HPCL	495.80	240.67	0.48542	0.89931	0.21684
IOCL	339.2234	82.9594	0.24 4557	0.407921	-5.69968
BPCL	611.523	199.861	0.32 6825	0.0565948	-1.12888
ONGC	275.570	64.8643	0.23 8302	0.340758	-0.20264

Table.1: Summary Statistics Average stock price of four companies

As comparatively in four sector ONGC and HPCL, mean of ONGC 275.570 and SD = 64.8643 while HPCL MEAN 495.80 and SD = 240.67. In HPCL sector found to be more variation in time series. All series are symmetric (skewness coefficient $\neq 0$) and trend to have kurtosis normal However normality test suggest not presence of normality in the data.

Test for stationary: Stationary of the series is the pre-requisite of any time series to develop any forecasting model. In this study we have used gretl software to test stationary the result are displayed in table (2)

company		coefficient	Std.Error	t-test	P Value
HPCL	θ	475.960	100.928	4.7158	< 0.00001
	φ_1	0.860019	0.0574093	14.9805	< 0.00001
IOCL	θ	334.614	33.9821	9.8468	< 0.00001
	φ_1	0.851919	0.0621288	13.7121	< 0.00001
BPCL	θ	588.771	90.4331	6.4741	< 0.00001
	φ_1	0.888648	0.0511836	17.3620	< 0.00001
ONGC	θ	262.319	34.8925	7.5179	< 0.00001
	φ_1	0.932141	0.0405164	23.0065	< 0.00001

Table (2): Partial Autoregressive (φ) and Moving Average test.

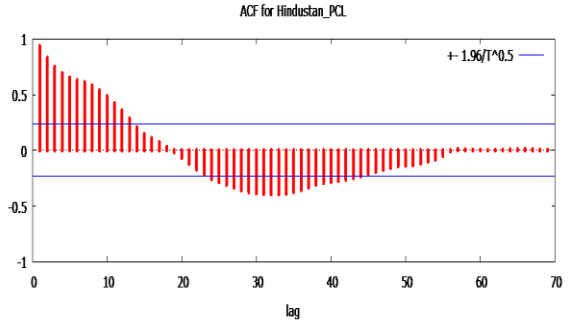
All the four companies' monthly average prices (Open, High, Low and low) are level. For all the series correlogram observed suggested AR (1) to be best with no MA. This is confirmed by observing the AIC (1,0,1) No other values of p and q were used as Box-Jenkins method recommends total number of parameters to be less than 2. Gretl software is used for data analysis in this paper.

Autoregressive Model Order Identification Process

The identification of the order of the fitted Autoregressive Model done by Plotting its Akaike Information Criterion (AIC) as shown in fig .2 in detecting the correct order for the fitted autoregressive model, it is necessary to examine the value at which the AIC gives a minimum value, bearing in mind that the first AIC value is for order zero. The minimum value at which the plot Akaike Information Criterion gives a minimum is 8 which make the order of the fitted autoregressive model to be 7.

Statistics of the four series is provided in Table 1.

Fig .2 **ACF Graph**



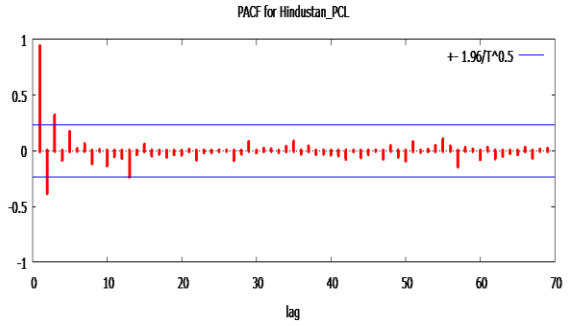
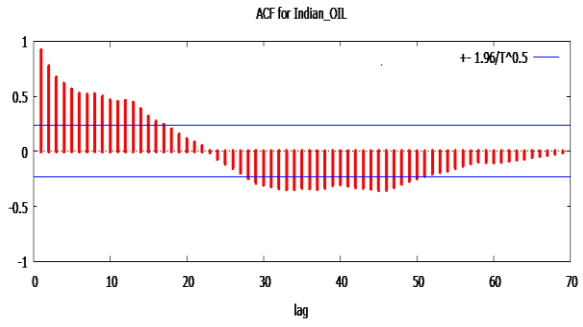


fig. 2(a) HPCL



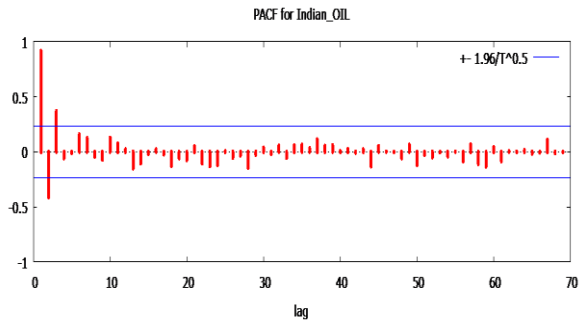
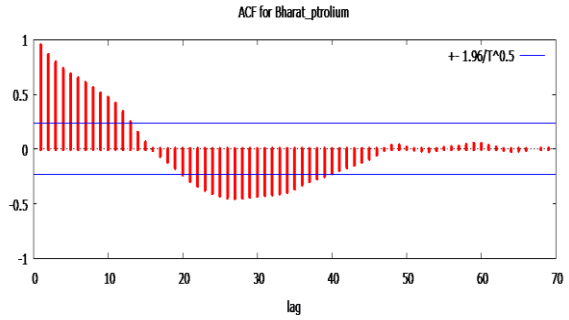


fig. 2(a) IOCL



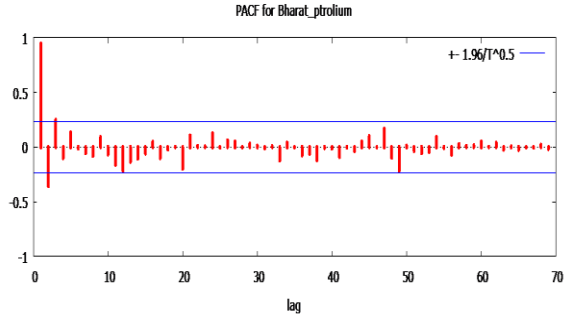
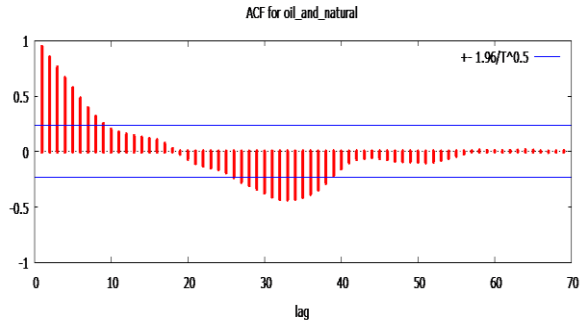


fig. 2(a) BPCL



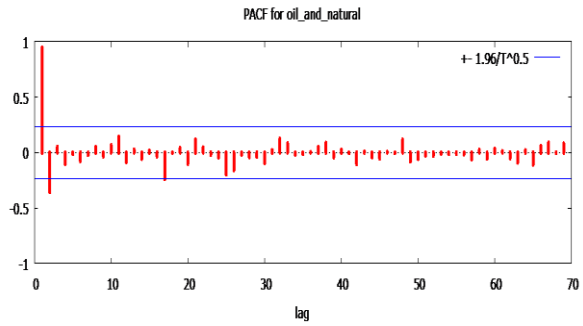


fig. 2(a) ONGC

The reasonable fit of the fitted autoregressive model was assessed by carrying out diagnostic checks. By observing the plots of Autocorrelation and Partial Autocorrelation of residual, as showed in fig.2, if they all possess the property of stationary. The plots of the Autocorrelation and Partial Autocorrelation of this residual in fig. 2 respectively also implies that the residual possess stationary property with most of the plotted points decaying to zero sharply which make the order of the fitted autoregressive model to

be 8.The reasonable fit of the fitted autoregressive model was assessed by carrying out diagnostic checks. By observing the plots of Autocorrelation and Partial Autocorrelation of residual, as showed in fig.2, if they all possess the property of stationary. The plots of the Autocorrelation and Partial Autocorrelation of this residual in fig. 2 respectively also implies that the residual possess stationary property with most of the plotted points decaying to zero sharply

Result and Discussion: The study deals with the closing price of 4 different companies pertaining to Oil sector is chosen for the study. The firms are Hindustan Petroleum Corporation Ltd., Bharat petroleum Corporation Ltd., Indian Oil Corporation Ltd.and Oil and Natural Gas Corporation Ltd. the series The descriptive Statistics of the four series is provided in Table 1 Summary Statistics Average stock price of four companies for all the series. The general trend of all the series is to increase, which can be observed in Fig 1- line graph of the averages prices. Among the four Indian Oil and Bharat petroleum performs comparatively better than the rest. In table 2. Partial AR(φ_1) and MA ($\boldsymbol{\theta}$) test the statistics standard error of sector HPCL constant =100.92, φ_1 = 0.0574093 , and ONGC

Constant = 34.895, φ_1 = 0.0405164 is less as compare to IOCL constant =33.9821, φ_1=0.0621288 while BPCL constant =90.4331, φ_1 =0.0511836.

References:

1. Box, G.E.P. and G.M. Jenkins, (1970). *Time series analysis: Forecasting and control*. Holden-Day, San Francisco
2. Qisen Caia, Defu Zhonga, Bo Wua, Stehpen C.H. Leungb,"Anovel stock forecasting model based on fuzzy time series and genetic algorithm" International conference on computational science,18(2013) pp1155-1162
3. Jose ManulAzevedo, Rui Almedia, Pedro Almeida" using data in short term stock prediction" A Literature review' International journal of intelligence science2 (2012) pp176-180
4. Mobarak,A.A., Mollahs and Bhuyan R (2008) Market efficiency in emerging stock market Evidence market finance 7(2010) pp17-41
5. Chicanos T.Okany (2014)" Effiect of Oil price movement on stock prices in the Nigerian Equity Market" International Tanning Institute, Central Bank of Nigeria Research Journal of finance and Accounting.
6. Diamond, D., & Dybvig, P. (1983). Bank Qisen Caia, Defu Zhonga, Bo Wua, Stehpen C.H. Leungb,"Anovel stock forecasting model based on fuzzy time series and genetic algorithm" International conference on computational science,18(2013) pp1155-1162.
7. Asche, F., Osmundsen, P. & Tveterås, R. (2001). Market Integration for Natural Gas in Europe. International Journal of Global Energy Issues, 16 (4), 300-312
8. Box, George Ep and George C. Tiao,"Intervation analysis with applications to economic and Environmental problems," Journal of the American Statistical Association 70.349 (1975)
9. Mondal, Shit, Goswami,(2014) study of effectiveness of time series modeling (ARIMA) in forecasting stock prices; International journal of computer science Engineering and Applications(IJCSEA) Vol.4. no.2
10. Young H.Kim, Edword L. et al. An ARIMA Model Approach to the behavior of Weekly Stock prices of fortune 500 firm and S& P small cap 600 firms.
11. Rahnao Jin, Sha wang, Fang Yan& Jie Zhu: The application of ARIMA model in 2014 shanghais. Composit stock price index, Journal of Applied Mathematics and Statistics3 (4) (2015) 199.203
12. Javier, c, Rosario E, Francisco and Antonio, j.c. ARIMA model to predict Next Electricity price IEEE Transactions on power systems 18(30 (2003)014-1020,
13. Anderson, B and Led older, J: Statistical Methods for FORCASTING, John Wiley and sons, Newyork 1983
14. Ardian Harri, Lanier Nalley, & Darren Hudson (2009). The Relationship between Oil, Exchange Rates, and Commodity Prices. Journal of Agricultural and Applied Economics, 41(2), 501-510.
15. Anurag Agnihotri, and Anand Sharma (2011) Study of convergence of spot & future prices in commodity market (With reference to Zeera,Zink and Natural gas for(2005-2010) International journal of Multidisciplinary Research I(2), 101-113