Prediction of Stock prices in Oil Sectors using ARIMA Model

A.V. Jadhav¹  K.B.Kamble²

¹Associate Professor, N.E.S. College of Engineering, Nanded, (M.S.) India,
²Siddharth 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 the 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 parameters (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.

\[ Y_k = \frac{\sum(y_{t-1} - \bar{y})(y_{t-k} - \bar{y})}{\sum(y_{t-1} - \bar{y})^2} \]

This quantity measures the linear relationship between the time series operations separated by a log k of 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 \( \Phi_{kk} \) is measures of the correlation between \( y_t \) and \( y_{t+k} \) after adjusting for the presences of \( y_{t+1}, y_{t+2}, \ldots \ldots \ldots, y_{t+k} \). One method of computing the partial autocorrelation of log k is to perform a regression of \( y_t \) on \( y_{t+1} \) through \( y_{t+k} \) using the resulting the coefficient of the \( y_{t+k} \) formed as the estimate of \( \Phi_{kk} \). When \( \Phi_{kk} \) is graphed for \( \log_1, \log_2, \ldots \ldots, \log_k \) 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 Zhang, 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 partition 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 Manuel Azevede Rui Almenda, 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 \( \mu \) and variance \( \sigma^2 \), and assume for case of notation that \( t \) takes on integer values \( t = \pm 0, \pm 1, \) the auto covariance function of \( X_t \) at lag \( k \) is defined as:

\[
X_t = \text{E}(X_{t-k}) (X_{t-k} - \mu)
\]

The autocorrelation function at lag \( k \) is defined as:

\[
\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\sigma_x^2}{\sigma_x^2}
\]

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

\[
\hat{\phi}_k = \frac{\hat{\gamma}_{k-1}}{1 - \hat{\phi}_1} - \hat{\phi}_1 \hat{\phi}_{k-1}
\]

b) Autoregressive Filter (AR (p))

The mathematical representation of the autoregressive model of order \( p \) is defined below:

\[
X_t = \varphi_1 X_{t-1} - \varphi_2 X_{t-2} \ldots \varphi_p X_{t-p} + \alpha_t
\]

Where \( \varphi_p \) and \( \alpha_t \) are the AR and MA parameters and \( \alpha_t \) 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

\[
\nabla_d \varphi(B) X_t = \theta(B) \alpha_t
\]

Where \( \nabla = (1 - B) \), B is the backward shift operator given as

\[
BX_t = X_{t-1}
\]

\( \varphi \) and \( \theta \) are the AR and MA parameters, respectively

Fig-1(a) HPCL

Fig-1(b) OICL

Fig.1(c) BPCL

order of the autoregressive and moving average models respectively of stochastic process \( X_t \) is given:

\[
X_t = \varphi_1 X_{t-1} - \varphi_2 X_{t-2} \ldots \varphi_p X_{t-p} - \alpha_t - \theta_1 \alpha_{t-1} - \theta_2 \alpha_{t-2} - \ldots - \theta_q \alpha_{t-q}
\]
Table 1: Summary Statistics Average stock price of four companies

<table>
<thead>
<tr>
<th>Company</th>
<th>Mean</th>
<th>S.D.</th>
<th>C.V.</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPCL</td>
<td>495.80</td>
<td>240.67</td>
<td>0.48</td>
<td>0.899</td>
<td>0.216</td>
</tr>
<tr>
<td>IOCL</td>
<td>339.22</td>
<td>594.85</td>
<td>0.24</td>
<td>0.407</td>
<td>-5.699</td>
</tr>
<tr>
<td>BPCL</td>
<td>611.52</td>
<td>594.85</td>
<td>0.32</td>
<td>0.056</td>
<td>-1.128</td>
</tr>
<tr>
<td>ONGC</td>
<td>275.57</td>
<td>64.85</td>
<td>0.23</td>
<td>0.340</td>
<td>-0.202</td>
</tr>
</tbody>
</table>

Table (2): Partial Autoregressive (p) 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.
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. Among the four Indian Oil and Bharat petroleum performs comparatively better than the rest. In table 2. Partial AR(\(\varphi\)) and MA(\(\theta\)) test the statistics standard error of sector HPCL constant =100.92, \(\varphi_1=0.0574093\), and ONGC
\[\text{Constant } = 34.895, \quad \varphi_1 = 0.0405164 \text{ is less as compare to } IOCL \quad \text{constant } =33.9821, \quad \varphi_1 =0.0621288 \text{ while BPCL constant } =90.4331, \quad \varphi_1 =0.0511836.\]

**References:**