P: ISSN NO.: 2321-290X RNI : UPBIL/2013/55327 VOL-6* ISSUE-8* (Part-2) April- 2019
E: ISSN NO.: 2349-980X **Shrinkhla Ek Shodhparak Vaicharik Patrika**

Stock Market Analysis Using Box-Jenkin Methodology

Abstract

Stock price prediction has always seek attraction of investors because of the financial benefit and complex nature of market. In this paper, we have studied the trend of Shirpur gold refinery daily stock data from 2010 to 2017. We propose a study on the effectiveness of Box-Jenkins methodology with Akaike Information Criterion (AIC) and BIC on Shirpur Gold Refinery. This was carried out by studying stationarity of the trend with the help of augmented dickey fuller test. Using Box-Jenkins methodology, prediction of stock index was studied based on data taken.

Keywords: AIC (Akaike Information Criterion, BIC (Bayesian Information Criterion), WAP, Difference-WAP,ACF, PACF, Standard Error.

Mayuri Pandya
Head & Professor
Dept. of Statistics,
Maharaja Krishna Kumar Sinhji
Bhavnagar University,
Bhavnagar, Gujarat, India

Monica Sharma
Research Scholar
Dept. of Statistics,
Maharaja Krishna Kumar Sinhji
Bhavnagar University,
Bhavnagar, Gujarat, India

Introduction

A time series can be define as a data collected at uniform time interval of consecutive point of time. It is an important part in statistics which analyze the nature of the data and helps in predicting or forecasting values of the series based on those characteristics. Prediction will always continue to attract researchers as well as investors in order to improve existing predictive models and profit gaining respectively. One can allow making investment decisions and developing effective strategy for profit gaining investment which everyone seeks for. Due to complex nature and factors affecting to stock market, sometimes it become difficult task to predict stock prices. Stock market can be termed as organized trading of securities through the various physical and electronic exchanges. It is the most essential areas of a market economy because investors can buy ownership of company by owning shares as well as company can access capital from market. By buying shares of ownership, investors stand to possibly gain money by profiting from companies future prosperity. Although there are millions to be gained by buying shares and them for a profit, not all investors are successful in gaining a return on their investment and even fewer are successful in making a lot of money. Due to stock market fluctuation, there are chances for losing capital when price of stock at time when it was bought was high and selling price of that particular stock decline. The only solution for not losing money would be for investors to sell their shares before they begin to decline. Thus, Stock price prediction is required and is regarded as one of the most difficult task. Future values may be predicted from past observations when consecutive observations are dependent. The times series can be said as deterministic if the future values of the time series can be predicted. Not all the time, predicted values can be obtained from past data.

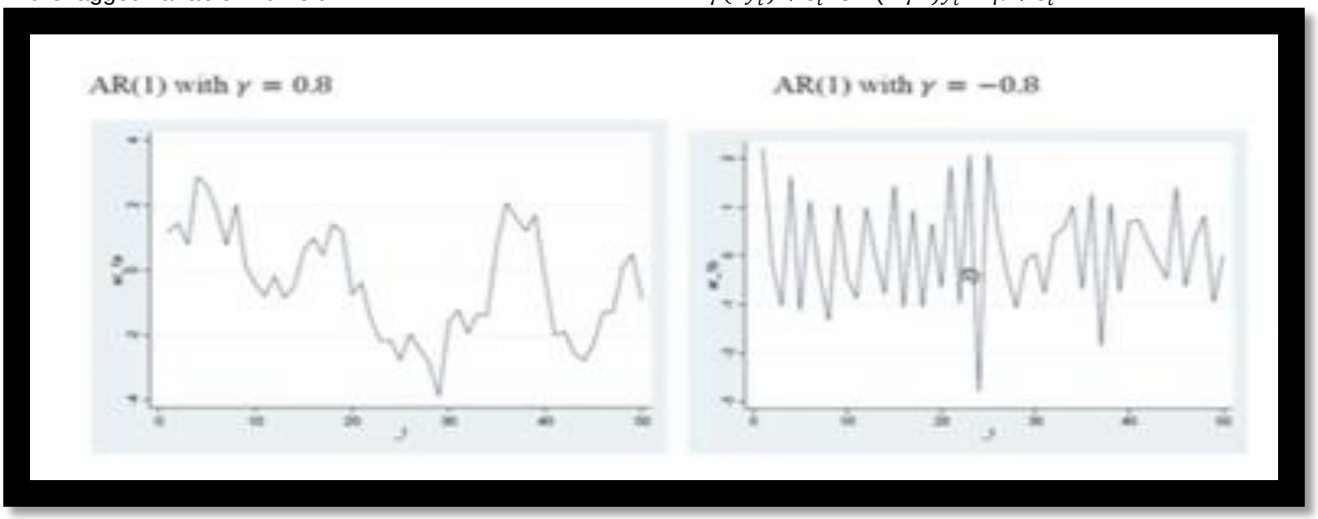
Methodology

In 1970, Box and Jenkins introduced the combination Autoregressive, Integrated and Moving Average, popularly known as ARIMA model. It is known as Box-Jenkins methodology too, compiled as a group of activities to identify, estimate and predict ARIMA models. ARIMA as well as individual models are most well-known methods in financial and economical forecasting. Understanding ARIMA, one need to understand white noise. White noise is defined as the assumption that each element in a series is a random draw from a population with zero mean and constant variance. For correction of violations of white noise assumptions, Autoregressive (AR) and Moving average (MA) models are considered best. ARIMA models may display efficient approximate forecasts. Below we can see the individual effect of the model on particular set of data as well as combination of 2 models, ARMA and lastly ARIMA model.

Autoregressive (AR) models

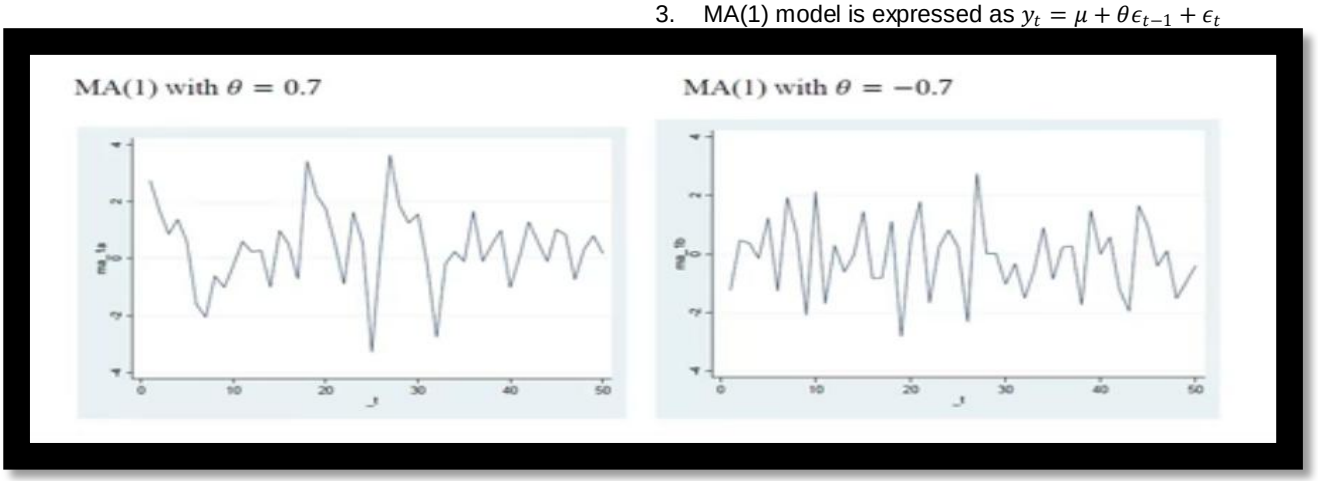
1. Autoregressive (AR) models are models in which the value of a variable in one period is related to its values in previous periods.
2. AR(m) is an autoregressive model with m lags:

3. $y_t = \mu + \sum_{i=1}^{m} \gamma_i \, y_{t-i} + \epsilon_t$
4. where μ is a constant and γ_m is the coefficient for the lagged variable in time t-m.
5. AR(1) is expressed as: $y_t = \mu + \gamma y_{t-1} + \epsilon_t = \mu + \gamma(Ly_t) + \epsilon_t$ Or $(1-\gamma L)y_t = \mu + \epsilon_t$.



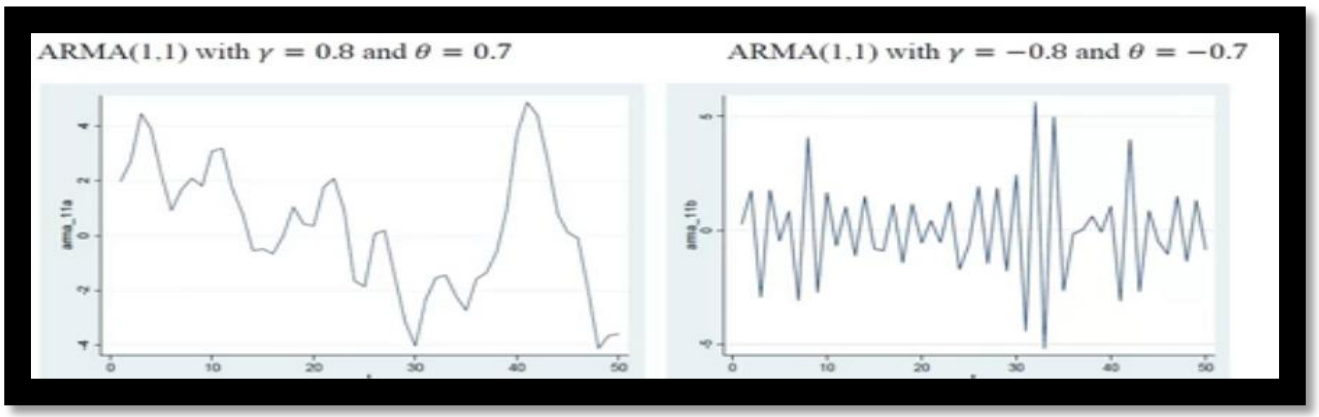
Moving average (MA) models:

1. Moving average (MA) models describe the possibility of a relationship between a variable and the residuals from previous periods.
2. MA(n) is a moving average model with n lags : $y_t = \mu + \sum_{i=1}^{n} \theta_i \, \epsilon_{t-i} + \epsilon_t$
   Where θ_n is the coefficient for the lagged error term in time t-n.
3. MA(1) model is expressed as $y_t = \mu + \theta \epsilon_{t-1} + \epsilon_t$



Autoregressive moving average (ARMA) models

1. Autoregressive moving average (ARMA) models combine both m autoregressive terms and n moving average terms, also ARMA (m,n).
2. $y_t = \mu + \sum_{i=1}^{m} \gamma_i \, y_{t-i} + \epsilon_t + \sum_{i=1}^{n} \theta_i \, \epsilon_{t-i}$



3. Modeling an ARMA (m,n) process requires stationarity.
4. A stationary process has a mean and variance that do not change over time and the process does not have trends.
5. An AR(1) disturbance process: $u_t = \rho u_{t-1} + \epsilon_t$
6. Is stationary if $|\rho|<1$ and ϵ_t is white noise.

The best part of this methodology is that one can make changes in models according to their data. Every model has its own importance in studying time series. Yule, Slutsky and Yaglom introduced idea ARMA models. Then after, some modification were made by introducing Integrated model by Box and Tiao in (1970) which is well-known as ARIMA model or Box-Jenkins methodology. Using this model in stock market for forecasting price can be done as WAP are not randomly generated values and can be used as discrete time series model, hence trend can be studied and prediction can be carried out. Further if data includes input variables, it is considered as ARIMAX model derived from ARIMA model.

Since, for accuracy of result or forecasting one need to transform the time series data rather than simple prediction. This paper focus not only on the prediction based on previous data, but with the help of Box-Jenkins methodology, it is figure out the efficiency of the predicted result for which investor can get an idea for investing.

The ARIMA model for Weighted Average Price of stock market is developed using subsection described below. Stock WAP data used for this paper are historical daily stock prices obtained from BSE, Indian Stock Exchange. The data composed of many elements such as open price, low price, high price, close price and WAP etc respectively. This paper deals with WAP (Weighted Average Price) as it represents the average price of the index. It also reflects average of all the activities of the index in a trading day. To determine the best ARIMA model among several experiments are performed, different ARIMA models can generated with the following criteria.

1. Comparatively small of BIC (Bayesian or Schwarz Information Criterion)
2. Relatively small Standard error.
3. Q-statistics and Correlogram for observing that the selected model are white noise and have stationarity into it. Stationarity test is carried out using Augmented Dickey Fuller test (ADF).

ARIMA (m, d, n) Model for Shirpur Gold Refinery Pvt. Ltd.

Shirpur Gold Refinery (SGR) stock data used in this study covers the period from 4th Jan, 2010 to 17th Oct, 2017 having a total number of 2028 observations.

Figure 1: Graphical representation of stock WAP index

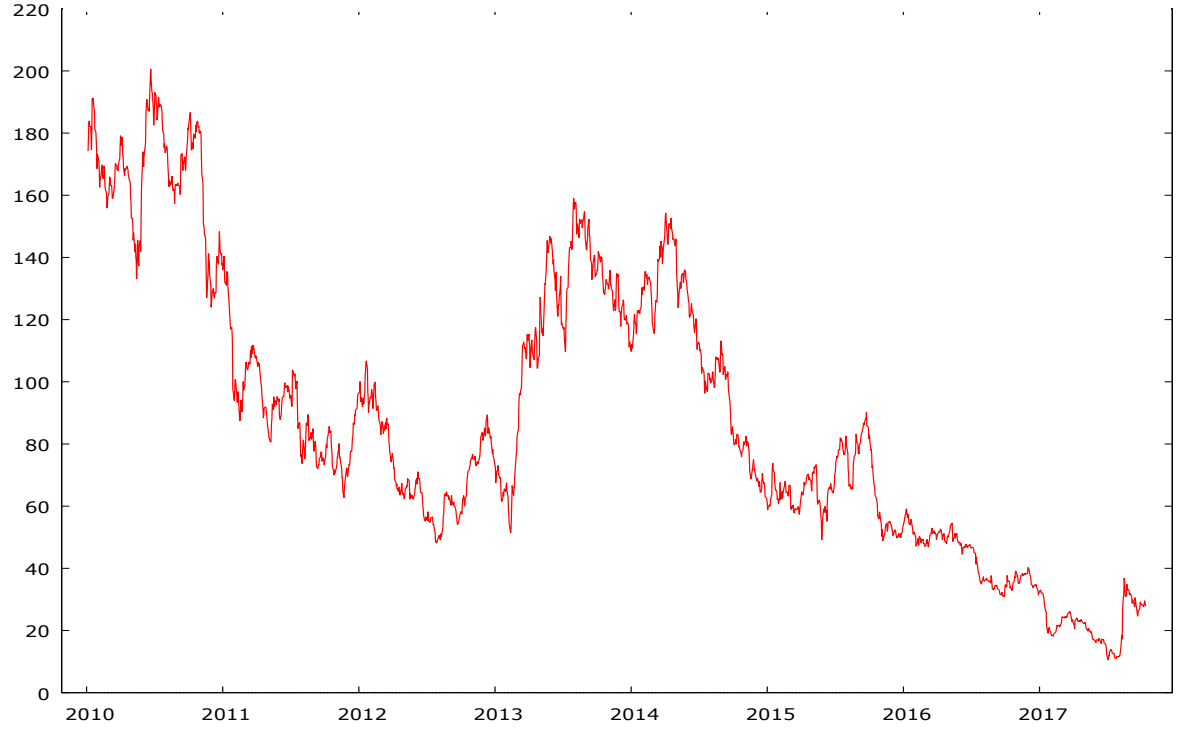
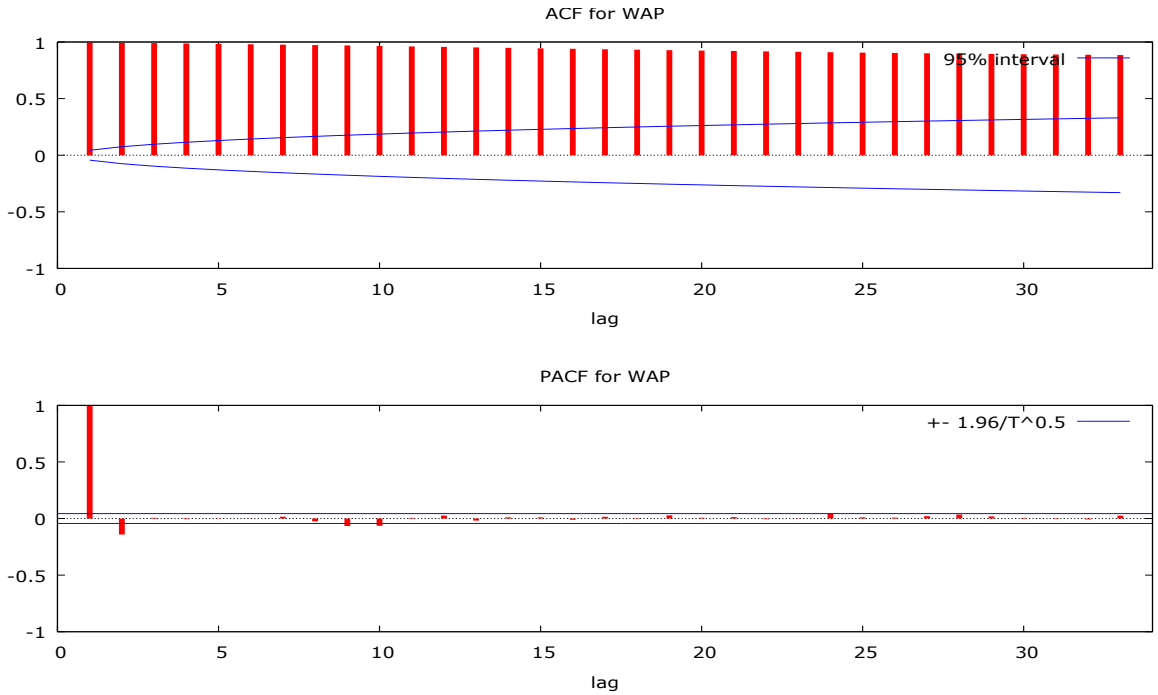


Figure 1 shows the original pattern of the time series of WAP to study whether the time series is stationary or not. From the figure shown above the time series have random walk pattern.

P: ISSN NO.: 2321-290X RNI : UPBIL/2013/55327 VOL-6* ISSUE-8* (Part-2) April- 2019
E: ISSN NO.: 2349-980X **Shrinkhla Ek Shodhparak Vaicharik Patrika**

Figure 2a: Graphical representation of ACF and PACF



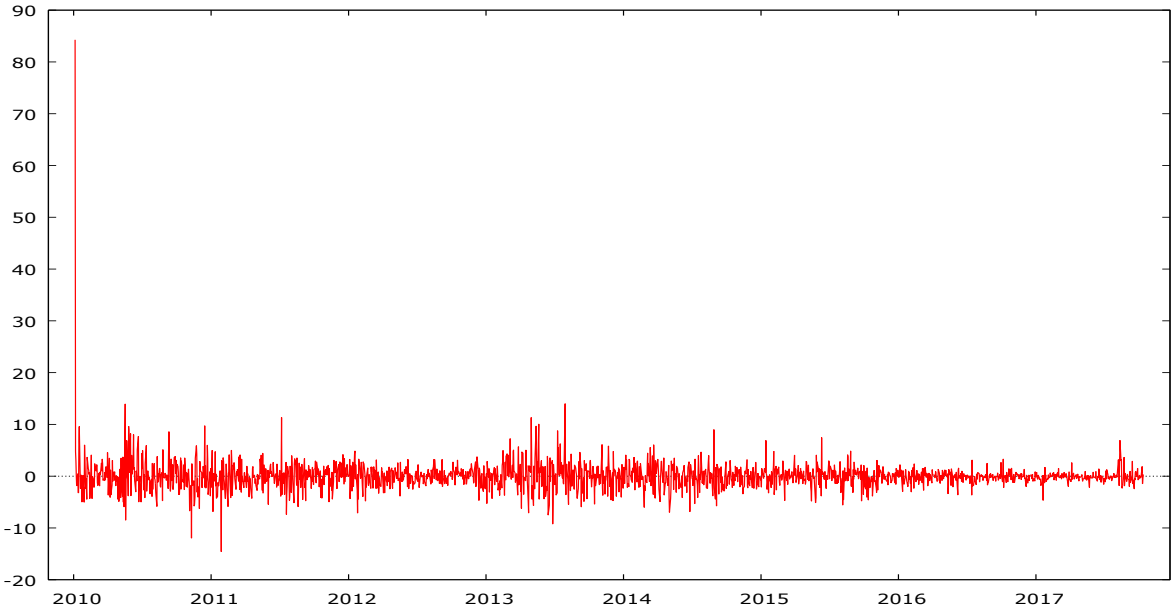
From the figure 2a, the ACF dies down extremely slowly which simply means that the time series is non-stationary. If the series is not stationary, it is converted to a stationary series by differencing. Figure 2b is the Correlogram of Shirpur Gold refinery time series. After the first difference, the series "D-WAP" (Differenced-WAP) of stock index becomes stationary as shown in figure 3 and figure 4 of line graph and the ADF test after differencing respectively.

Figure 2b: Correlogram of Shirpur Gold Refinery

*, **, * indicate significance at the 1%, 5%, 10% levels
using Bartlett standard errors for ACF

LAG	ACF		PACF		Q-stat.	[p-value]
1	0.9973	***	0.9973	***	2021.1864	[0.000]
2	0.9939	***	-0.1411	***	4029.5701	[0.000]
3	0.9905	***	0.0063		6024.9373	[0.000]
4	0.9870	***	-0.0053		8007.2381	[0.000]
5	0.9835	***	0.0021		9976.5688	[0.000]
6	0.9800	***	-0.0006		11933.0094	[0.000]
7	0.9767	***	0.0164		13876.9816	[0.000]
8	0.9732	***	-0.0245		15808.1554	[0.000]
9	0.9694	***	-0.0665	***	17725.1153	[0.000]
10	0.9651	***	-0.0642	***	19626.2460	[0.000]
11	0.9608	***	0.0064		21511.3680	[0.000]
12	0.9566	***	0.0262		23381.0967	[0.000]
13	0.9524	***	-0.0180		25235.3120	[0.000]
14	0.9483	***	0.0105		27074.2833	[0.000]
15	0.9442	***	0.0103		28898.4290	[0.000]
16	0.9401	***	-0.0106		30707.7286	[0.000]
17	0.9361	***	0.0155		32502.4365	[0.000]
18	0.9321	***	0.0043		34282.6036	[0.000]
19	0.9282	***	0.0281		36048.8115	[0.000]
20	0.9244	***	0.0072		37801.4941	[0.000]
21	0.9207	***	0.0131		39541.1133	[0.000]
22	0.9170	***	-0.0060		41267.7418	[0.000]
23	0.9133	***	-0.0023		42981.4427	[0.000]
24	0.9099	***	0.0413	*	44683.1417	[0.000]
25	0.9066	***	0.0109		46373.3272	[0.000]
26	0.9034	***	0.0078		48052.3475	[0.000]
27	0.9003	***	0.0209		49720.8237	[0.000]
28	0.8975	***	0.0352		51379.7159	[0.000]
29	0.8949	***	0.0187		53029.7375	[0.000]
30	0.8923	***	0.0059		54671.1777	[0.000]
31	0.8898	***	-0.0043		56304.0488	[0.000]
32	0.8872	***	-0.0084		57928.3012	[0.000]
33	0.8848	***	0.0256		59544.5593	[0.000]

Figure 3: Graphical representation of the Sgr stock price index after differencing




```
Augmented Dickey-Fuller test for WAP
testing down from 33 lags, criterion BIC
sample size 2027
unit-root null hypothesis: a = 1

   test with constant
   including one lag of (1-L)WAP
   model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
   estimated value of (a - 1): -0.00212794
```

Figure 4: Augmented Dickey-fuller test for WAP

```
   asymptotic p-value 0.3901
   1st-order autocorrelation coeff. for e: 0.014

   with constant and trend
   including one lag of (1-L)WAP
   model: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
   estimated value of (a - 1): -0.00398998
   test statistic: tau_ct(1) = -2.37995
   asymptotic p-value 0.3901
   1st-order autocorrelation coeff. for e: 0.014
```

In above figure the model checking was done with Augmented Dickey Fuller (ADF) unit root test on "D-WAP" of Shirpur gold refinery stock index. The result confirms that the series becomes stationary after the first-difference of the series.

Table 1 shows the different values of the several ARIMA model experimented. ARIMA (2, 1, 2) turn out to be the best for Shirpur Gold Refinery Stock Index. The model returned the smallest Bayesian or Schwarz information criterion of 9073.069 and standard error of 2.143 as shown in Table 1.

In statistically forecasting, the best model selected can be expressed as follows:

$$Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \varepsilon_t$$

Where $\varepsilon_t = Y_t - \widehat{Y}_t$ (i.e. the difference between the actual value of the series and the forecast value.)

Table 1: Statistical Results Of Different Arima Parameters

Models	BIC	S.E.
(1,0,0)	9219.16	2.345
(1,0,1)	9102.933	2.272
(2,0,0)	9112.595	2.277
(0,0,1)	18533.84	23.24
(0,0,2)	16310.92	13.42

(1,1,0)	9099.838	2.277
(0,1,0)	9217.538	2.346
(0,1,1)	9089.997	2.272
(1,1,2)	9085.498	2.2698
(2,1,0)	9078.208	2.272
(2,1,2)	**9073.069**	**2.143**

Results and Discussion

Table 1 showed the best ARIMA model among the list of different models tested with BIC and S.E. values. Table 2 is the result of the predicted values of ARIMA (2, 1, 2) considered the best model for SGR stock index.

Table 2: Sample of Empirical Results of Arima (2,1,2) Of SGR Stock Index

For 95% confidence intervals, z(0.025) = 1.96

	WAP	prediction	std. error	95% interval
2010-01-04	174.230849			
2010-01-05	178.880623	174.159834		
2010-01-06	182.846891	179.942903		
2010-01-07	183.897396	183.517528		
2010-01-08	182.186822	184.010027		
2010-01-11	182.112542	181.775968		
2010-01-12	178.778888	182.250340		
2010-01-13	174.662457	177.878562		
2010-01-14	180.279472	173.847868		
2010-01-15	191.103320	181.865281		
2010-01-18	191.216322	193.311780		
2010-01-19	188.712768	190.636608		
2010-01-20	188.213752	188.378882		
2010-01-21	186.852194	188.241416		
2010-01-22	181.429334	186.510050		
2010-01-25	180.181822	180.135024		
2010-01-26	177.326142	180.185732		
2010-01-27	174.407442	176.462557		
2010-01-28	168.589494	173.785030		
2010-01-29	173.106078	167.108199		
2010-02-01	171.731896	174.493293		
2010-02-02	170.828723	170.738160		
2010-02-03	167.844077	170.758492		
2010-02-04	162.588228	166.927357		
2010-02-05	164.882090	161.343807		
2010-02-08	166.755339	165.621321		
2010-02-09	168.303096	166.781202		
2010-02-10	168.946306	168.539635		
2010-02-11	169.753571	168.917314		
2010-02-12	165.537004	169.874414		
2010-02-15	165.395840	164.335181		
2010-02-16	169.364934	165.621695		
2010-02-17	168.016786	170.169578		
2010-02-18	163.282836	167.335936		
2010-02-19	162.012047	162.214958		
2010-02-22	160.870834	161.888935		
2010-02-23	158.361066	160.452806		
2010-02-24	155.895904	157.688512		
2010-02-25	157.621554	155.295758		
2010-02-26	159.443643	158.051253		
2010-03-01	160.783047	159.599119		
2010-03-02	163.239199	160.952861		
2010-03-03	165.891473	163.720729		
2010-03-04	165.215891	166.361067		
2010-03-05	163.185202	164.875809		
2010-03-08	163.194473	162.751993		
2010-03-09	161.389384	163.276541		
2010-03-10	158.987938	160.830512		
2010-03-11	159.131751	158.464073		
2010-03-12	159.399521	159.213976		
2010-03-15	161.941095	159.319611		
2010-03-16	163.437368	162.512140		
2010-03-17	166.996001	163.566635		
2010-03-18	170.197415	167.827965		
2010-03-19	170.072950	170.752295		
2010-03-22	169.436357	169.906902		
2010-03-23	168.320182	169.363699		
2010-03-24	168.682403	168.066255		
2010-03-25	167.948293	168.839171		
2010-03-26	169.486694	167.682761		
2010-03-29	171.541829	169.928789		
2010-03-30	171.664790	171.897071		
2010-03-31	176.151675	171.580204		
2010-04-01	179.116882	177.327028		
2010-04-02	176.075956	179.537069		
2010-04-05	178.719413	175.240681		
2010-04-06	177.194338	179.695359		
2010-04-07	172.640341	176.515263		
2010-04-08	172.642769	171.704233		
2010-04-09	168.932876	172.877362		

2017-09-12	27.760477	28.448475
2017-09-13	30.543308	27.493811
2017-09-14	29.825410	31.243229
2017-09-15	29.000562	29.349568
2017-09-18	26.428420	28.882899
2017-09-19	25.502495	25.726762
2017-09-20	24.640510	25.381459
2017-09-21	25.267581	24.331332
2017-09-22	26.431545	25.399560
2017-09-25	26.680411	26.571588
2017-09-26	27.413460	26.608913
2017-09-27	29.173038	27.545086
2017-09-28	28.966237	29.512817
2017-09-29	28.290491	28.759217
2017-10-02	28.392493	28.140903
2017-10-03	28.092945	28.408296
2017-10-04	27.990765	27.949874
2017-10-05	27.787457	27.946117
2017-10-06	27.638492	27.681346
2017-10-09	29.532568	27.563842
2017-10-10	28.619204	29.969501
2017-10-11	27.995345	28.193697
2017-10-12	28.254489	27.918259
2017-10-13		28.274029 2.271056 23.822840 - 32.725218

Forecast evaluation statistics

Mean Error	-0.00066386
Root Mean Squared Error	2.2712
Mean Absolute Error	1.5724
Mean Percentage Error	-0.0037879
Mean Absolute Percentage Error	1.955
Theil's U	0.96418
Bias proportion, UM	8.5436e-008
Regression proportion, UR	0.0017082
Disturbance proportion, UD	0.99829

Figure 5: Graph of Actual Stock Price v/s Predicted values of Sgr Stock Index

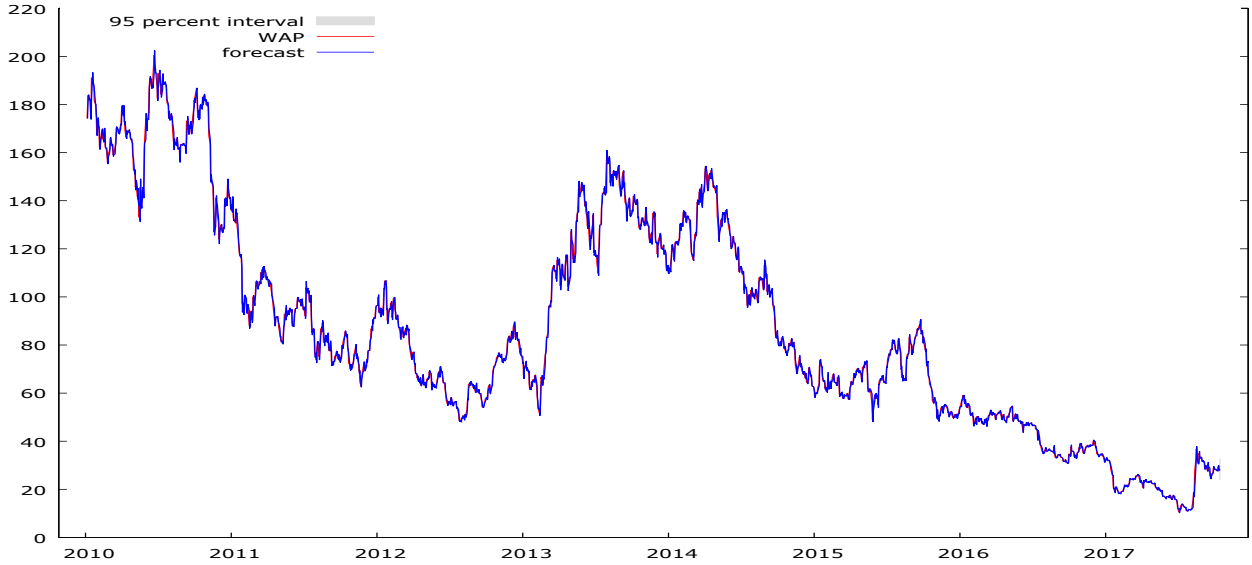


Figure 5: graphical representation of the accuracy of the predicted WAP level against actual WAP of the ARIMA model selected. From figure, it can be observed there are very minor difference between predicted value and actual value which indeed is satisfactory result.

Conclusion

This paper has presented the process of Box-Jenkin's methodology and building different ARIMA model for WAP stock Prediction. Though it also depends on other factor that affects the stock market. Therefore data was selected on long-term basis to include some of the factors. The results obtained shows satisfactory prediction of stock price with best ARIMA model among different models, which can guide the investor to earn profit through their decision.

References

B.G. Tabachnick and L.S. Fidell, "Using multivariate statistics", 4th ed., Person Education Company, USA 2001.

Box, G.E, Jenkins, G.M., and Reinsel, G.C.(1994). Time Series Analysis: Forecasting and Control. 3rd Edition. New Jersey: Prentice – Hall Inc.

C. Wang, "A comparison study of between fuzzy time series model and ARIMA model for forecasting Taiwan Export", Expert System with Applications, vol.38, no.8, pp.9296-9304, 2011.

Devi, B. Uma, D. Sundar, and P. Alli. "An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50."

Meyler, G. Kenny and T. Quinn, "Forecasting Irish Inflation using ARIMA Models", Central Bank of Ireland Research Department, Technical Paper, 3/RT/1998.

Pai, Ping-Feng, and Wei-Chiang Hong. "An improved neural network model in forecasting arrivals." Annals of Tourism Research 32.4 (2005): 1138-1141