Libyan Oil Sales forecasting using ARIMA models

Mustafa . M . Ali . AlfAki, Prof. Dr. Ajet paul & Dr. (Ms.): Shalini B. Masih

***Abstract*— This study aims to apply modern forecasting for the study and analysis of monthly time series data for the sales of Azzawiya oil Refining Company–Libya, and represented in Fuel Oil For the period (2008 - 2014) by using the ARIMA models known as Box-Jenkins Method in all its stages (Diagnostic, Estimation, Adequacy Test of model, Forecasting). In order to find the best models for forecasting sales of the Fuel oil for the next period.**

The statistical program Eviews.5 was used in the analysis. The results of the analysis of the data were shown after estimating ARIMA models. And through the trade-offs between them using Forecasting Accuracy Measurements, it was appeared that the most appropriate model to forecast sales of the company for the Fuel oil is ARIMA (2,1,2); Based on this model the forecasting of monthly sales has been done for the period of upcoming six years. The forecasting values are consistent with the values of the original time series which show the efficiency of the model.

***Index Terms*— Autoregressive, Moving Averages, Autocorrelation Function, Partial Autocorrelation Function, Box-Jenkins, Mixed Models, Mixed Autoregressive Moving Average Models, ARIMA Models, Forecasting.**

I. INTRODUCTION

Oil sector in Libya is one of the important and vital sectors especially at present time, because this sector is most prominent and most important sector for various productive and service sectors alike as one of the most important sectors that contributes to advance the progress towards the betterment. The Libyan government has paid the greatest attention to the oil sector by important role played by the sector in improving the living standard of its citizens, especially when we know that Libya is among the limited sources countries. That is why it was necessary to study the oil sector in order to forecast the sales of oil products produced by Azzawiya Oil Refining Company.Inc.Libya, by using time series models, and main purpose of time series method is to obtain a model that can be used to describe the problem and lead to forecast the future of the studied phenomenon.

II. LITERATURE REVIEW

There are many previous studies that have used the time-series model method in forecasting with economic nature, and most of the references listed at the end of the research contain practical examples on it, and we will review the most important studies.

Bopp and Lady.(1991) evaluate the hypothesis that futures and spot prices perform similarly as forecasting subsequent cash prices. Allen. (1994) The procedures and data used to forecast Midwest diesel prices are identified starting with a presentation of the preliminary procedures.LeBlanc and Coibion. (2001). Who postulated that the best predictor of future prices of oil is futures prices. While they found that futures prices are unbiased predictors of future spot prices, the prediction errors were large. Muhammad Jalal Studied. (2005) the forecasting of fuel sales of national company's marketing and distribution of petroleum products (NAFTAL). Rangsan Nochai , Titida Nochai. (2006)Studied model of forecasting oil palm price of Thailand in three types as farm price, wholesale price and pure oil price for the period of five years, (2000 – 2004).

Abeerhasan Jaburi. (2010) Studied the forecasting of Iraqi oil prices and forecasting by using the time-series and she used two methods of forecasting one of them is (Box-Jenkins) another is dual exponential boot models. Jon et al. (2011) Provide application of ACE in modeling parameters.

It has been observed that no previous research has focused on development of models to estimate natural gas price using time series analysis to find relationship between crude oil and natural gas prices.

From the above mentioned studies it is clear that ARIMA can be used to forecast. The authors tried to find out best ARIMA model, but in most of the articles the authors used ARIMA to forecast. And the present study is designed to select the best ARIMA model to forecast.

III. DATA AND METHODOLOGY

the data under study using the ARIMA models, have been getting this data from the Azzawiya Oil Refining Company–Libya, which is the company's sales of oil product that is produced and represented in Fuel oil for the period (January 2008 – December 2014), to analyze this data was used statistical software (Eviews.5).

Study of the series stationary:

At this stage the time series for original data is drown to know initially about some characteristics of this series, the following Graph represents that:

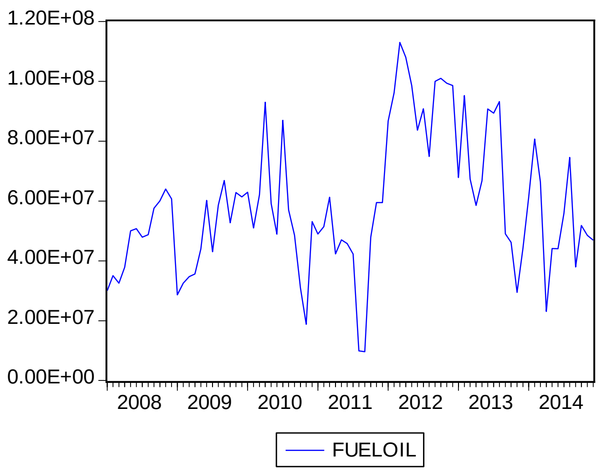
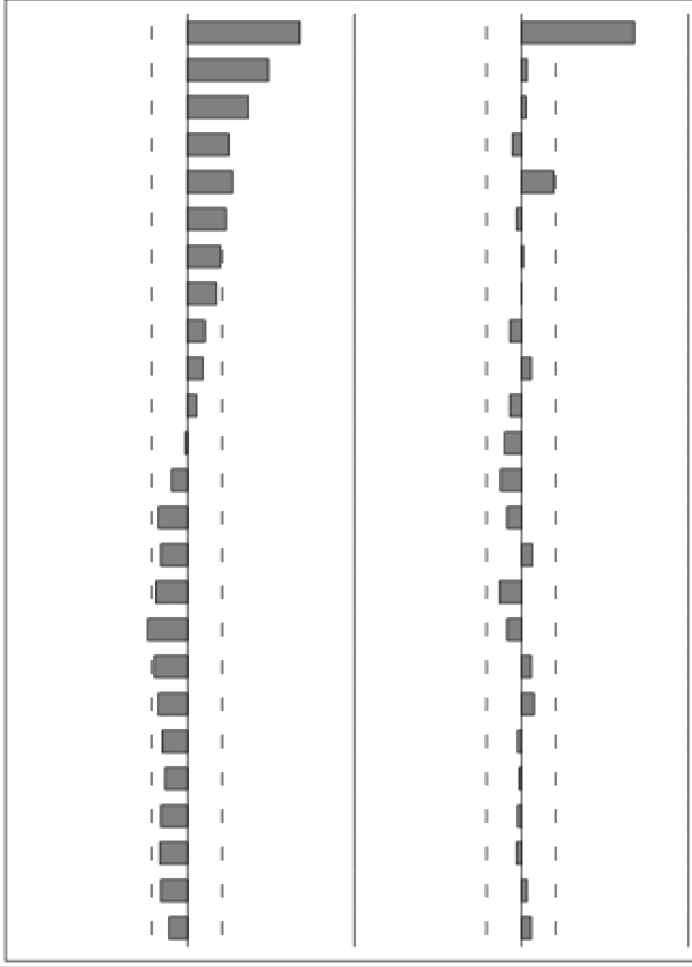


Figure (1): Graphical representation of the series Fuel Oil
Source: Eviews.5

Manuscript received March 24, 2015.

Through the Figure (1) of Jet Fuel series we see that the series spread randomly, and therefore, this graphical presentation gives us no answers whether the series is stationary or not? So we draw the autocorrelation function (ACF), and the partial autocorrelation function(PACF) of data and draw confidence interval of (ACF) and (PACF) to detect the stationary or non-
stationary of time series, as well as the use of the test Q - Box Ljung to ensure stationary of the
series, as in the following table:

Table (1): ACF and PACF of the series Fuel Oil



Date: 02/01/15 Time: 11:36
Sample: 2008M01 2014M12
Included observations: 84

	AC	PAC	Q-Stat	Prob
1	0.697	0.697	42.287	0.000
2	0.506	0.039	64.840	0.000
3	0.382	0.031	77.870	0.000
4	0.259	-0.053	83.934	0.000
5	0.279	0.198	91.038	0.000
6	0.245	-0.029	96.586	0.000
7	0.209	0.012	100.70	0.000
8	0.187	0.001	104.03	0.000
9	0.109	-0.066	105.16	0.000
10	0.105	0.059	106.25	0.000
11	0.063	-0.069	106.64	0.000
12	-0.016	-0.101	106.67	0.000
13	-0.091	-0.123	107.50	0.000
14	-0.177	-0.084	110.73	0.000
15	-0.159	0.067	113.37	0.000
16	-0.194	-0.133	117.39	0.000
17	-0.240	-0.084	123.57	0.000
18	-0.202	0.059	128.04	0.000
19	-0.174	0.080	131.39	0.000
20	-0.151	-0.021	133.96	0.000
21	-0.135	-0.016	136.03	0.000
22	-0.156	-0.023	138.88	0.000
23	-0.170	-0.025	142.30	0.000
24	-0.163	0.039	145.49	0.000
25	-0.112	0.059	147.03	0.000

Source: Eviews.5

Through the table (1) of the original series of Correlation Coefficients and figures of ACF and PACF, we note that there is non- stationary in the data of original series as there are some values outside the confidence interval, and that the significant value of the Coefficients autocorrelation function using the test (Q - stat) Ljung & Box is:

$\chi^2_{25,0.05} = 37.652$ Q – stat = 147.03 >

We then test the unit root of original series Fuel oil by using Advanced Dickey Fuller test
(ADF) as in the following table:

Table (2): ADF test results for the original series (Fuel Oil)

Prob.	Test Critical Values 5%	T -student	ADF – Test
0.0037	-2.8968	-3.847	With constant
0.018	-3.4649	-3.8634	With constant and trend
0.2079	-1.9448	-1.2035	Without constant and trend

Source: Eviews.5

Through the data of the table (2) we conclude:
The statistical value calculated for Dickey Fuller test in the case (without constant and trend) is

greater than the corresponding table value. i.e. we accept the hypothesis of a unit root.
The results of this test indicate non-stationary of the series, and to make it stationary we make differences from First level while taking the natural logarithm of the data and then we get the following figure:

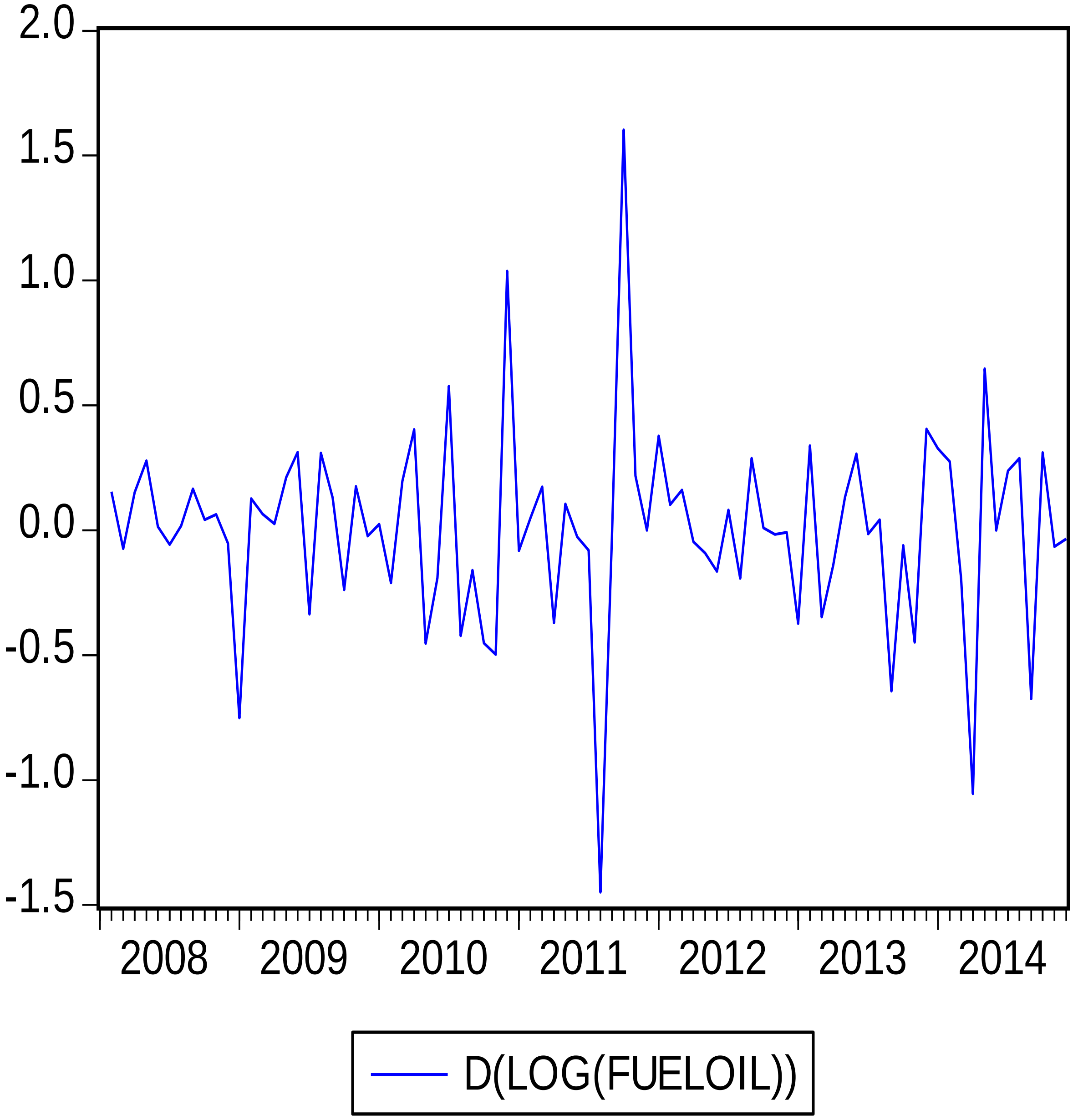
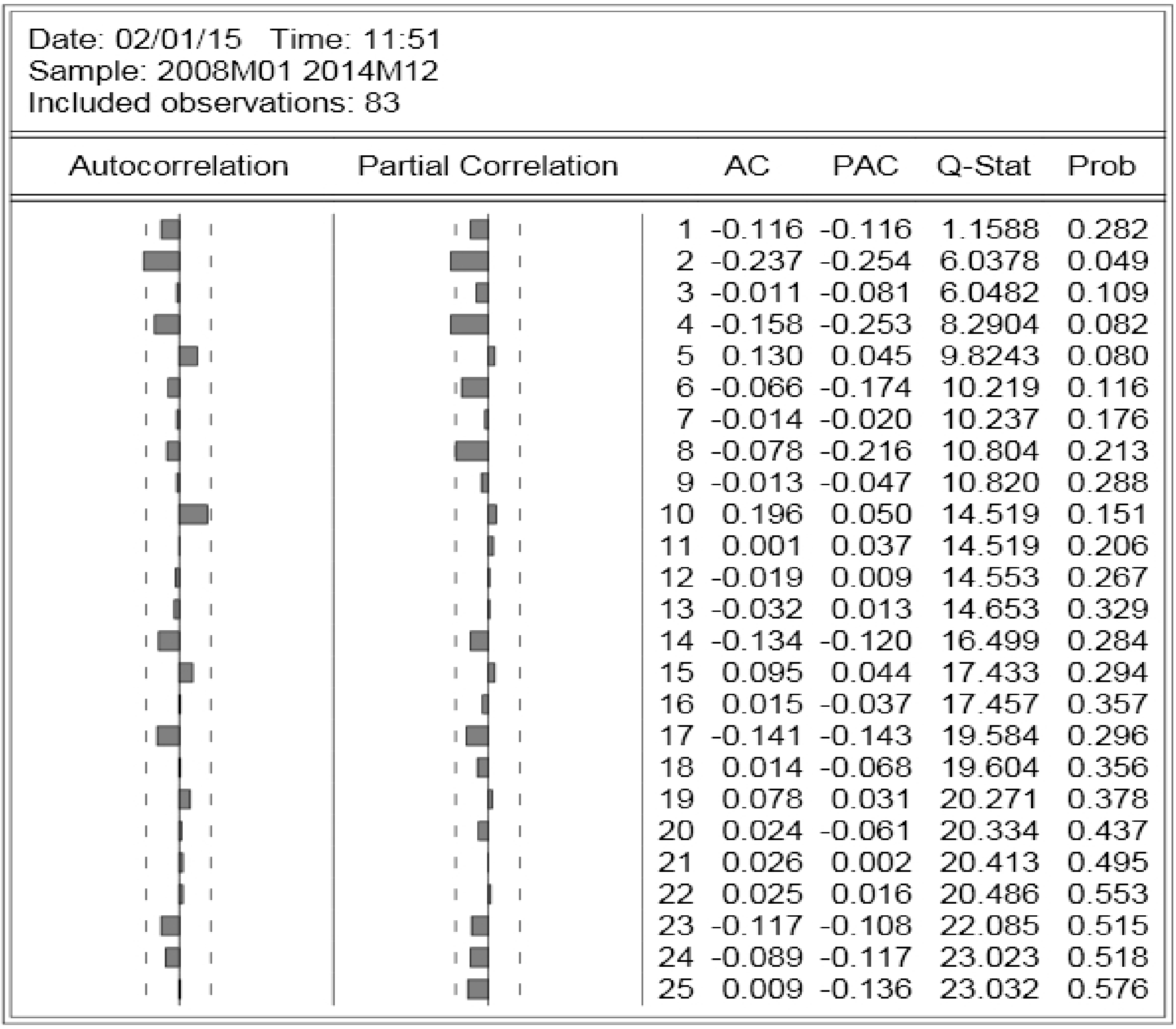


Figure (2): Graphical representation of the series D(Log(Fuel Oil)) Source: Eviews.5

Through the Figure (2), it is clear that the series oscillation Around the zero value (0), and does not increase with time (after taking the first difference) while taking the natural logarithm of data, and to ensure the stationary of the time series we draw ACF and PACF of data and draw confidence interval for ACF and PCF to detect the stationary or non-stationary of the time series,
as well as the test Q - Box Ljung is also used to make sure the stationary of the series, as in the following table:
Table (3.3.3.3): ACF and PACF for D(log(Fuel Oil))



Date: 02/01/15 Time: 11:51
Sample: 2008M01 2014M12
Included observations: 83

	AC	PAC	Q-Stat	Prob
1	-0.116	-0.116	1.1588	0.282
2	-0.237	-0.254	6.0378	0.049
3	-0.011	-0.081	6.0482	0.109
4	-0.158	-0.253	8.2904	0.082
5	0.130	0.045	9.8243	0.080
6	-0.066	-0.174	10.219	0.116
7	-0.014	-0.020	10.237	0.176
8	-0.078	-0.216	10.804	0.213
9	-0.013	-0.047	10.820	0.288
10	0.196	0.050	14.519	0.151
11	0.001	0.037	14.519	0.206
12	-0.019	0.009	14.553	0.267
13	-0.032	0.013	14.653	0.329
14	-0.134	-0.120	16.499	0.284
15	0.095	0.044	17.433	0.294
16	0.015	-0.037	17.457	0.357
17	-0.141	-0.143	19.584	0.296
18	0.014	-0.068	19.604	0.356
19	0.078	0.031	20.271	0.378
20	0.024	-0.061	20.334	0.437
21	0.026	0.002	20.413	0.495
22	0.025	0.016	20.486	0.553
23	-0.117	-0.108	22.085	0.515
24	-0.089	-0.117	23.023	0.518
25	0.009	-0.136	23.032	0.576

Source: Eviews.5

Through the table (3) of modified series of correlation coefficients and figures of ACF and
PACF we note that there is stationary in the data of series and most of the values within the
confidence interval, and that the Significant value of autocorrelation coefficients by using the test
(Q - stat) Ljung & Box was:

$\chi^2_{25,0.05} = 37.652$ Q – Stat = 23.032 <

We then test by Advanced Dickey Fuller and the estimation of models is as follows:

Table (4): ADF test results for series D(Log(Fuel Oil))

Prob.	Test Critical Values 5%	T -student	ADF – Test
0	-2.8977	-8.5494	With constant
0	-3.4662	-8.5176	With constant and trend
0	-1.9448	-8.6012	Without constant and trend

Source: (Eviews.5)

Through the data of the table (4) we conclude; All calculated statistics of Dickey Fuller in all models are less than the corresponding tabular values, i.e. Reject the hypothesis of a unit root in the series, and the series D(Log(Fuel oil)) is stationary

• Model Identification Stage:

After ensuring the stationary of the Series D(Log (Fuel oil)), one can identify Model, through the table (3), we observe the presence of moving averages models MA (1), MA (2), through the autocorrelation function. And the presence of autoregressive models AR (1), AR (2), through partial autocorrelation function, and for more accuracy in reconciling the best model among Box-Jenkins models, possible models of ARIMA (p, d, q) have been applied, and calculation of each of them SC, AIC and MSE are as shown in the following table:

Table (5): Compared to a set of values of AIC , SIC , MSE

Models	AR	MA	MSE	AIC	SIC
ARIMA (1,1,1)	1	1	0.1024	0.6320	0.7200
ARIMA (1,1,2)	1	2	0.1307	0.8765	0.9645
ARIMA (2,1,2)	2	2	0.0860	0.4583	0.5470

Source: Eviews.5

From the table (5), we find that the model ARIMA (2,1,2) is the one who gives the lower

values of the previous standards, these values were as follows:

MSE = 0.0860 , AIC= 0.4583 , SC= 0.5470
So this model was relied on to be an appropriate model for this series.

• Estimating the model parameters:
After identifying the best model which is ARIMA (2,1,2), we estimate the specific model
parameters by Least-squares.
Table (6): Estimates of model parameters ARIMA (2,1,2)

Dependent Variable: DFUELOIL
Method: Least Squares
Date: 02/04/15 Time: 08:00
Sample (adjusted): 2008M04 2014M12
Included observations: 81 after adjustments
Convergence achieved after 50 iterations
Backcast: OFF (Roots of MA process too large)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.046262	0.006900	6.704605	0.0000
AR(2)	0.412789	0.084109	4.907774	0.0000
MA(2)	-1.285462	0.049828	-25.79797	0.0000

R-squared	0.452891	Mean dependent var	0.004523
Adjusted R-squared	0.438863	S.D. dependent var	0.398903
S.E. of regression	0.298815	Akaike info criterion	0.458347
Sum squared resid	6.964633	Schwarz criterion	0.547030
Log likelihood	-15.56305	F-statistic	32.28385
Durbin-Watson stat	2.154717	Prob(F-statistic)	0.000000

Inverted AR Roots	.64	-.64
Inverted MA Roots	1.13	-1.13
	Estimated MA process is noninvertible	

Source: Eviews.5

From the table (6) We note that the model estimate of ARIMA (2,1,2) according to the
statistical T – student, That all the coefficients of the model are statistically significant
(less than 0.05), and that this model has gross significance because:
prob (F-Statistic) < 0.05
Thus, the estimated model of series Fuel oil can be written as follows:

$$\hat{Z}_t = \varepsilon_t + 0.4128\ Z_{t-2} + 1.2855\ \varepsilon_{t-2}$$

• Check the efficiency of the model:

After the parameters of diagnosed model have been estimated, it is necessary to check the
efficiency of this Model, and primarily it is identified through auto correlation coefficients and
partial auto correlation coefficients of the series residuals (errors) to ensure the stationary of the
residuals series we do the following tests:
Test of White Noise (Study of stationary of the residuals):
The following table represents Auto correlation function and partial Auto correlation function for
residuals of the Model for ARIMA (2,1,2)
Table (7): ACF and PACF for Residual ARIMA (2,1,2)

Date: 02/04/15 Time: 08:17
Sample: 2008M04 2014M12
Included observations: 81
Q-statistic probabilities adjusted for 2 ARMA term(s)

	AC	PAC	Q-Stat	Prob
1	-0.157	-0.157	2.0775	
2	-0.016	-0.042	2.0995	
3	-0.038	-0.049	2.2261	0.136
4	-0.088	-0.106	2.9037	0.234
5	0.100	0.068	3.7911	0.285
6	-0.043	-0.025	3.9549	0.412
7	-0.004	-0.019	3.9564	0.556
8	-0.036	-0.044	4.0749	0.667
9	-0.024	-0.027	4.1298	0.765
10	0.146	0.127	6.1504	0.630
11	-0.003	0.040	6.1514	0.725
12	-0.032	-0.026	6.2484	0.794
13	-0.023	-0.017	6.2994	0.853
14	-0.159	-0.152	8.8521	0.716
15	0.062	-0.012	9.2483	0.754
16	-0.062	-0.074	9.6471	0.788
17	-0.125	-0.171	11.276	0.733
18	-0.031	-0.108	11.381	0.785
19	0.044	0.032	11.594	0.824
20	-0.001	-0.058	11.594	0.867
21	0.002	-0.040	11.595	0.902
22	-0.012	-0.017	11.610	0.929
23	-0.107	-0.119	12.926	0.911
24	-0.096	-0.140	14.013	0.901
25	0.008	-0.075	14.020	0.926

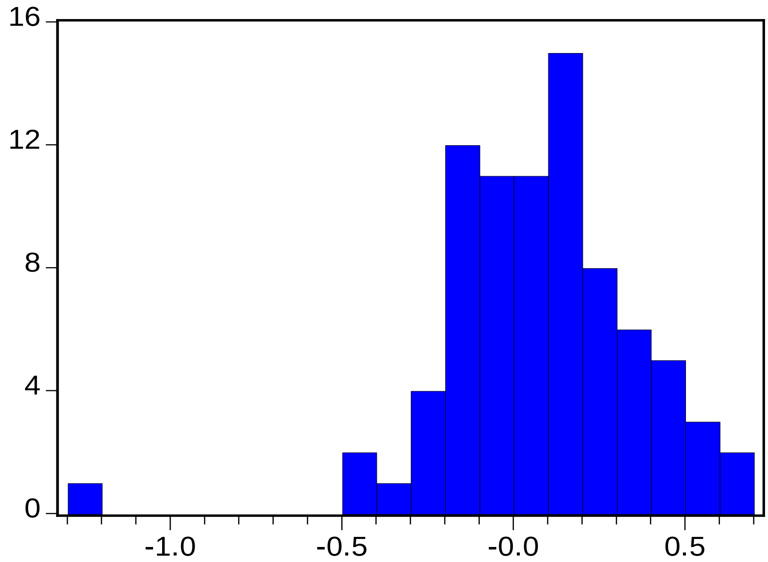
Source: Eviews.5

From the table (7) we note that the most of the coefficients fall within the confidence interval, as well as the statistical (Q):

$\chi^2_{25,0.05} = 37.652$ Q – stat 14.020 <

Therefore Residuals represent white noise.

Normal distribution test:
In order to ensure that the residuals follow the normal distribution, Jarque - Bera test has been used where the results of this test are shown in the following diagram:



Series: Residuals
Sample 2008M04 2014M12
Observations 81

Mean	0.076638
Median	0.076350
Maximum	0.616684
Minimum	-1.211594
Std. Dev.	0.284800
Skewness	-0.973057
Kurtosis	6.765853
Jarque-Bera	60.64541
Probability	0.000000

Figure (3): Testing the hypothesis of normal distribution of residuals series

Source: Eviews.5

Note that almost symmetric, and the residuals are distributed an average of close to zero, and the variance Autocorrelation test for errors:

To carry out this test, Durbin-Watson's statistical has been used, as in the following table:

Table (8): Autocorrelation test for residuals Series Fuel Oil

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	10.24263	Prob. F(2,76)	0.000115
Obs*R-squared	12.51966	Prob. Chi-Square(2)	0.001912

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 02/01/15 Time: 12:41
Sample: 2008M04 2014M12
Included observations: 81
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.047278	0.010657	4.436380	0.0000
AR(2)	-0.576272	0.129897	-4.436380	0.0000
MA(2)	-0.341408	0.076956	-4.436384	0.0000
RESID(-1)	0.000239	0.103495	0.002308	0.9982
RESID(-2)	0.681597	0.175800	3.877113	0.0002

R-squared	0.154564	Mean dependent var	0.076638
Adjusted R-squared	0.110067	S.D. dependent var	0.284800
S.E. of regression	0.268670	Akaike info criterion	0.269073
Sum squared resid	5.485937	Schwarz criterion	0.416878
Log likelihood	-5.897442	F-statistic	3.473603
Durbin-Watson stat	1.845175	Prob(F-statistic)	0.011648

Source: Eviews.5

From the table (8), we find that:
Durbin-Watson = 1.8452

We note that:
Value of D-W lies within the confidence interval, it has no Auto correlation of errors, we accept the hypothesis H_0. Then all the results of the residuals tests confirm the validity of the estimated model ARIMA (2,1,2) to represent the time series, and can be used in the forecasting process.

Forecasting Stage:
After deciding the diagnosed and best model, its level, estimating its parameters and testing
its efficiency, now we are in the last stage of Box jenks Models and this is forecasting stage
which is the target of the study, among the estimated model:

$$\hat{Z}_t = \varepsilon_t + 0.4128\, Z_{t-2} + 1.2855\, \varepsilon_{t-2}$$

The relation of forecasting for the period h will be as follows:

$$\hat{Z}_t = \varepsilon_{t+h} + 0.4128\, Z_{t+h-2} + 1.2855\, \varepsilon_{t+h-2}$$

Where: $\varepsilon_{t+h} = 0$

Test of forecasting accuracy:

It is necessary to insure the accuracy of the forecasting by available model while
experiencing it on the previous values of the series Fuel Oilby using the test Chow Forecast
test as in the following table:
Table (9): Chow Forecast Test

Chow Forecast Test: Forecast from 2014M03 to 2014M12

F-statistic	1.663017	Prob. F(10,68)	0.107802
Log likelihood ratio	17.72144	Prob. Chi-Square(10)	0.059848

Source: Eviews.5

From the table (9) we note that:

$F_C = 1.6630$

$F_{0.95}(6, 68) = 2.2312$

FC < Ft and so we accept HO ie. The model of forecasting is fixed for the last nine months of the series which means that the forecasting after the year 2014 will be accurate in greater level.
Forecasting of company's Fuel oil product for nine years (2015- 2020) will be as in the following:

Table (10): Fuel Oil sales forecast for (2015 – 2020)

Year / Month	2015	2016	2017	2018	2019	2020
Jan.	60270976	1.22E+08	2.13E+08	3.71E+08	6.46E+08	1.12E+09
Feb.	63394863	1.28E+08	2.23E+08	3.88E+08	6.76E+08	1.18E+09
March	72231613	1.34E+08	2.33E+08	4.06E+08	7.08E+08	1.23E+09
April	75785165	1.40E+08	2.44E+08	4.26E+08	7.42E+08	1.29E+09
May	82181903	1.47E+08	2.56E+08	4.46E+08	7.77E+08	1.35E+09
June	86135778	1.54E+08	2.68E+08	4.67E+08	8.14E+08	1.42E+09
July	91518383	1.61E+08	2.81E+08	4.89E+08	8.52E+08	1.48E+09
Aug.	95880475	1.69E+08	2.94E+08	5.12E+08	8.92E+08	1.55E+09
Sep.	1.01E+08	1.77E+08	3.08E+08	5.37E+08	9.35E+08	1.63E+09
Oct.	1.06E+08	1.85E+08	3.23E+08	5.62E+08	9.79E+08	1.71E+09
Nov.	1.11E+08	1.94E+08	3.38E+08	5.89E+08	1.03E+09	1.79E+09
Dec.	1.16E+08	2.03E+08	3.54E+08	6.16E+08	1.07E+09	1.87E+09

Source: Eviews.5

The figure (4) represents the Fuel oil Sales of the company for which the forecasting has been made according the Model ARIMA (2, 1, 2) :

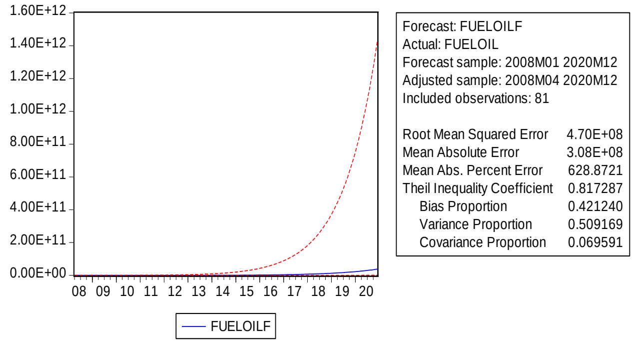


Figure (4): Graphical representation of sales forecasting Fuel Oil Source: Eviews.5

IV. CONCLUSIONS:

Through the applied study of time series analysis using ARIMA models to forecast sales of Azzawiya oil Refining Company-Libya for Fuel oil, the research scholar has reached the following conclusions:While comparing ARIMA models with other statistical methods for forecasting it was found that ARIMA models are the best forecasting models. But at the same time it is very difficult method which needs more observation when it used for economic forecasting.Time series of sales of company are an independent series which do not need any preliminary transformations.After the study and analysis of the series of company's sales for Fuel oil it was found that this is non-stationary series in compare to original data, which have been converted into stationary series after taking differences.After estimating ARIMA models and through the trade-offs between them using accurate forecasting measurements, it was found that the most appropriate model to forecast sales of the company for Fuel oil product is ARIMA (2,1,2)At the time of diagnosis of residuals series for company's product for Fuel oil , this series was independent and random which indicate the diagnostic accuracy of ARIMA models, proposed to use in the series of monthly sales. And the Autocorrelation Function and Partial Autocorrelation Function of the residuals come under the confidence interval which confirms the randomness of residuals (Errors) and relying on them for normal distribution with a mean of (zero) and variance ($\frac{1}{\sqrt{n}}$)

The estimated model is the best model to use in forecasting of the company's sales in future. When the forecasting values and values of original series were compared using approved models, the result confirmed that the forecasting values are consistent with the values of original series because of the very small differences between them. It shows the efficiency of the models.According to the forecasting extracted from the estimated model that the company's sales will rise in the coming years.

REFERENCES:

[1] Box, G. and Pierce, D, (1970)., "Distribution of Autocorrelations in Autoregressive Moving AverageTime Series Models", Journal of the American Statistical Association, 65, 1509-1526.

[2] Box , G.E.P and G.M Jenkins, (1976)."Time series Analysis : forecasting and control".San Francisco : Holden day , Inc. 2^{nd} Ed , 1976.

[3] Chatfied, C,.(1980). " The analysis of time series, An introduction" second edition chapman and hall,London-New york.

[4] Abraham, B. and Ledoter.J, (1983) ."Statistical Methods for Forecasting", John Wiley, New York.

[5] Mohammed, Aziz Menem, (1987). "Introduction to Time Series Analysis and index numbers ",Faculty of Management and Economics, Mustansiriya University, Baghdad.

[6] Makridakis, Spyros et al (1998). Forecasting: Methods And pplications. Jon Willy & New York.

[7] Allyrdi, Adnan Hashim, (1990). "Statistical forecasting techniques and methods of its application",Faculty of Administration and Economics, University of Basra, Iraq.

[8] Bopp, A. E., and G. M. Lady (1991), "A Comparison of Petroleum Futures versus Spot Prices as Predictors of Prices in the Future", Energy Economics 13, 274–282.

[9] Vande, Walter, (1992). "Time series of the destination and Applied models Box –Jenkins",Arabization of Abdul Hamid Mardi Azzam,Mareh Publishing House, Riyadh, Saudi Arabia.

[10] Brockwell, P. and R. Davis. (2002). Introduction to Time Series and Forecasting , New York.

[11] Ghannam, Hamad Abdullah, (2003). "Time Series Analysis of Stock Price Index in Saudi Arabia using the methodology Box - Jenkins ", Journal ofKing Abdulaziz University, Saudi Arabia.

[12] Shaarawi, Samir Mustafa,(2005)." Introduction to the analysis of modern time series."Center of Scientific Publications- King Abdulaziz University.

[13] Liu, L. M., (2006), "Time Series Analysis and Forecasting", second edition, Scientific Computing Associates Corp. USE

[14] Fabian TorbenBosler,(2010)."Models for Oil price prediction and forecasting", San Diego State University.

[15] AbeerhasanJaburi, (2010). " forecasting of Iraqi oil prices and forecasting by using the time-series.",Journal of Babil University , Humanities ,Volume 18 / Issue No.1.

[16] Ron Alquist, Lutz Kilian and Robert J. Vigfusson,(2011)."Forecasting. The Price of Oil.",Bank of Canada Working Paper 2011-15.