Use Box - Jenkins strategy for making a time sample to forecasting to Oil prices

Jassim Nassir Hussain[1]
University in Karbala
department of statistics
Jasim.nasir@uokerbala.edu.iq

Saif Hosam Raheem[2]
University of Al-Qadisiyah
department of statistics
Saif.hosam@qu.edu.iq

Mariam Mahdi Enad[3]
University in Karbala
department of statistics
mariam.m@uokerbala.edu.iq

Abstract

The strategy of Box-Jenkins in the field of time models building is one of the most important models which take interest to predicting in time series by using past and current values for predicting in the short term only, which also known as the Box Jenkins model.

This study aims to analyze the time series of oil prices for the period from 01/03/2017 to 08/01/2021 by using the Box-Jenkins strategy in time analysis that includes (diagnosis, estimation, model suitability testing, prediction) to find the appropriate model for predicting oil prices in Iraq by depending on the daily data of recorded prices.

The results of the application showed that the appropriate and efficient model is the ARIMA (2,1,2) integrated self-regression model.

Keywords : time series, forecasting , Box-Jenkins method.

Article Received: 18 October 2020, Revised: 3 November 2020, Accepted: 24 December 2020

Introduction

The financial crises had very important effect on retrogradation & fluctuation of oil prices and the deterioration of world economy for most counties and companies. With all these fluctuations but the oil prices raised slightly. The countries and companies care about the Fluctuations of oil prices and what resulted from benefits and loses for these companies and countries. From this the importance of the study of oil prices and possibility of making mathematical model framing for representation the prices series and use this model to predict the oil prices in the future.

The Research's aim

This study aims to use the temporal series for analyzing the temporal series that connected with oil prices and predict on these prices in the future and recognize on the Box-Jenkins way in analysis.

Research's method

This study includes three axes, the first one assigned for the theoretical part which include the concept of time series and box-Jenkins samples, the second axes dealt with the practical part, while the third we introduced the conclusions and suggestions .

Theoretical part :

1- Time series:

Time series defines as Consecutive time order for observation group for variable , these observations arrange according to seeing time . (AlMotawaly, 1989:15)

It can also define as a group of observations happened in specific time periods usually these periods can be equal. The time series divided into:

1-1 Stationary time series

For estimating any time series, the data stability will be checked which make the prediction for the future easy. (Cryer, 1986: 200). The time series will be stable if has arithmetic mean with stable disparity with time continuity. Also the time series will be stable without presence any local direction and different fluctuation in series shape. (Box-Jenkins,, 1976: 22)

2-1 Seasonal Time Series

In a case of existence of regular pattern which repeats after (s) from the time series called (Seasonal time series) and the period called the Seasonal period (s) which will be month or a year for each season as :

$$f(t+s)=f(t)$$

2. Box-Jenkins method :

This method can be used for complicit time series, also the other prediction cases which more than one pattern can be found a the same time.

This way as others, we can see advantages and disadvantages for this method:

- The disadvantage of these method is the high cost in contrast with other methods and the understanding and applying difficulty, since it treats many patterns that be complicated.
- The advantages of these method are high accuracy because it treats complicated and mixed time series at the same time, also it assumes temporary pattern which stops the data under study, we achieve less value for error. (Alzubaidy , 1980: 5)

3- Unit Root test :

The Box-Jenkins time series is checking for random context, and for applying their ways, the creator context should be stable.

The unsability which can face in the time series which represents the real observer comes out from that this series either be trend stationary or difference stationary. (Authman Naqar, 2011: 128)

3.1 **Dickey-Fuller test**

This test depends on three simple equation assumed existence of random context from self pattern in 1st level as:

1- $\Delta X_t = \alpha_1 X_{t-1} + e_t$
2- $\Delta X_t = \alpha_0 + \alpha_1 X_{t-1} + e_t$
3- $\Delta X_t = \alpha_0 + \alpha_1 X_{t-1} + B_t + e_t$

The testing hypothesis $H_0: a_1 = 0$, existence of not stability , comparing test $t = \frac{\alpha_1}{SE(\alpha_1)}$ with theoretical values by Dickey and fuller. (Othman Nakar, 2011: 130)

2.3. Augmented Dickey–Fuller test

If the model changes were time series, most times , insert the time series in regression model leads to unknown results like coefficient of determination (R2) even there is no real relation among these changes, this called spurious regression , so it must sure the stability of these series for each variable in stable shape, and for stationary of time series for the study sample, this request test of unit root test , although the multiple tests of unit root test but Dickey fuller is the important and popular one in modern studies. (Albasheer, 2009: 13)

3. ARMA model

3.1 Autoregressive Model

General formula of autoregressive model from (P) degree will be: (AlJebouri, 2010: 13)

$$Z_t=\Phi_0+\Phi_1 z_{t-1}+\Phi_2 z_{t-2}+\ldots\ldots+\Phi_p z_{t-p}+a_t$$

This equation represents the autoregressive model from (P) degree , and (Φj) represents parameter group AR(p) which can be like:

$$\Phi(B)z_t=\Phi_0+a_t$$

whereas

$$\Phi(B)=1-\Phi_1 B-\Phi_2 B^2-\ldots\ldots\ldots-\Phi_p B^p$$

whereas

Z_t Observation

(Φ_i) parameters of the model (i=1,2,…….p)

Φ_0 constant

P order of the model

a_t Error term $N\sim(0,\sigma_a^2)$

2.4 Moving Average Model

The general formula for autoregressive model from (q) degree will be :

$Z_t = \theta_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$

This formula called moving average (MA) from (q) degree and its simple MA(q) which can be written as :

$Z_t = \theta_0 + \theta(B) a_t$

Where

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$

The autoregressive correlation function (ACF) for the movable average close or Interrupted to zero after displacement (q) whereas the autoregressive correlation function will dwindle in Exponentiation shape. (Alsaraf, 1981: 24)

3.4 Auto Regressive-Moving Average Models (ARMA)

This model can be written by formula from (p.q) degree as :

$Z_t = \Phi_0 + \Phi_1 z_{t-1} + \Phi_2 z_{t-2} + \dots + \Phi_p z_{t-q} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$

$\Phi(B) z_t = \Phi_0 + \theta(B) a_t$

It can be called ARMA(p,q) and the modul stability if the root of equation $\Phi(B)=0$ outside The boundaries of a single circle (AlMutawaly, 1989: 24)

5. Sample's estimating

5.1 Identification

In this step, test the best sample to represent the time series from the mixed samples, specify an idea about the cost (p,d,q) which the sample need it which required two steps:

5.1.1. series stability achievement

knowing the series stability only through examing the graph for the values of time series or through deviding the time series to aparts which length is proportional to the length of the seasonal cycle , whether the series not seasonal so the Lengths of subtotals are equall for almost 8 views . then calculate the arithmetic mean & standard deviation for all groups, then draw the pairs of these means and deviation. (Alwardi, 1990 : 16)

when the graph is Scattered randomly, so the series will have steady contrast, whereas the graph is Scattered randomly around a straight line with a slope up or down, this eplains the unstability of this series in contrast, for achieving this stability in contrast, the Appropriate conversion for the time series, from these conversion is logarithm conversion according to :

$Z^*_t = \log z_t$

$z^*_t = \sqrt{z_t}$

Or Radical transformation in average will be according to draw Autoregressive Average model for the sample, if the diagram inclines slowly to zero when the displacement periods will prove the unstability of the series in the average. (AlMotawalli. 1989: 29)

$\nabla z_t = z_t - z_{t-1}$

5.1.2 Chose one of the ARMA for representing the series

After achieving the series stability in average and contrast, one of the ARMA samples will choose through studying the ACF and PACF with the conduct of Autoregressive Average and partial Autoregressive Average for the stable linear samples. (Alwardi, 1990: 20) in a case of being more than one sample, the contrasting process according to the less value for MSE.

5.2. Estimation :

for achieving the Preliminary for Autoregressive Average sample we use Yule-walker :

$\gamma(h) = \emptyset_1 \gamma(h-1) + \dots + \emptyset_p \gamma(h-p)$, h=1,………p

$\sigma_w^2 = \gamma(0) - \emptyset_1 \gamma(1) - \dots - \emptyset_p \gamma(p)$

5.3 verification of suggestion sample :

After finding the estimates of the suggestion sample statics, checking from sample accuracy through testing the knowledge either different from zero and product test that the static different from zero if it will be outside the ($\pm$1.96)

The most famous test to relevance the sample is the Box-Pierce which used to test the Statistical significance test for For self-links for others according to: (Box-Pierce,1970:20)

$Q = n\sum_{K=1}^{L} r_k^2(a) \sim x^2_{((L-m),\alpha)}$

Which is :

L: The number of seasonal displacements

M: The number of estimated features

5.4 Forecasting :

Define as the scientific method in research to achieve the unknown data through the attached data in this research:

Application section

1- data description :

Data which is used in this part is daily time series about (1034) watch except the Days of non-trading which represent the prize of daily recorded oil barrel in OPEC which we got it from the official website of the organization https://www.opec.org for the period 03/01/2017 to 08/01/2021.

2- draw the graphical series :

Before analysis the time series of this study as mentioned in shape (1), we find repeated fluctuate with different of increasing between a year and other , this explains general orientation and seasonal factors.

graph (1) The time series

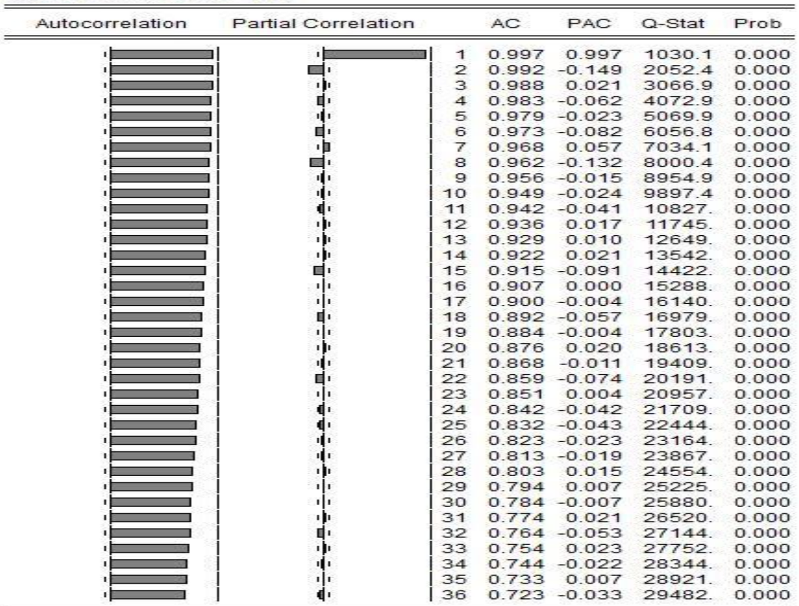
3. Testing the time series stability :

By noticing the graph (1) see for time series, we find that the series tends to fluctuate in time change which explains that the series is unstable, this certified by the graph for autocorrelation function and partial autocorrelation in graph (2), also the data of chart (1) which shows Unit Root test by using ADF test, the hypothesis of this test as:

H_0: the series not stationary

H_1: the series is stationary

The (P-value) (0.7589) in chart (1) we find that it is maximal from (0.05), which means that it is un significance, so here we refuse null hypothesis, the study series is not stable.



Date: 01/18/21 Time: 14:29
Sample: 1/03/2017 1/08/2021
Included observations: 1034

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.997	0.997	1030.1	0.000
		2	0.992	-0.149	2052.4	0.000
		3	0.988	0.021	3066.9	0.000
		4	0.983	-0.062	4072.9	0.000
		5	0.979	-0.023	5069.9	0.000
		6	0.973	-0.082	6056.8	0.000
		7	0.968	0.057	7034.1	0.000
		8	0.962	-0.132	8000.4	0.000
		9	0.956	-0.015	8954.9	0.000
		10	0.949	-0.024	9897.4	0.000
		11	0.942	-0.041	10827.	0.000
		12	0.936	0.017	11745.	0.000
		13	0.929	0.010	12649.	0.000
		14	0.922	0.021	13542.	0.000
		15	0.915	-0.091	14422.	0.000
		16	0.907	0.000	15288.	0.000
		17	0.900	-0.004	16140.	0.000
		18	0.892	-0.057	16979.	0.000
		19	0.884	-0.004	17803.	0.000
		20	0.876	0.020	18613.	0.000
		21	0.868	-0.011	19409.	0.000
		22	0.859	-0.074	20191.	0.000
		23	0.851	0.004	20957.	0.000
		24	0.842	-0.042	21709.	0.000
		25	0.832	-0.043	22444.	0.000
		26	0.823	-0.023	23164.	0.000
		27	0.813	-0.019	23867.	0.000
		28	0.803	0.015	24554.	0.000
		29	0.794	0.007	25225.	0.000
		30	0.784	-0.007	25880.	0.000
		31	0.774	0.021	26520.	0.000
		32	0.764	-0.053	27144.	0.000
		33	0.754	0.023	27752.	0.000
		34	0.744	-0.022	28344.	0.000
		35	0.733	0.007	28921.	0.000
		36	0.723	-0.033	29482.	0.000

Graph (2) autocorrelation function and partial autocorrelation for the original time series

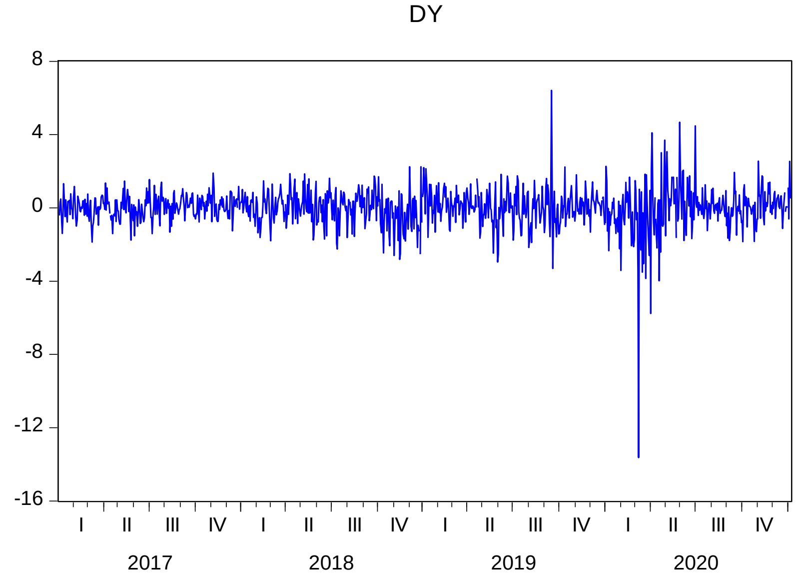
Null Hypothesis: Y has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic - based on SIC, maxlag=21)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.681700	0.7589
Test critical values:	1% level	-3.966965	
	5% level	-3.414173	
	10% level	-3.129195	

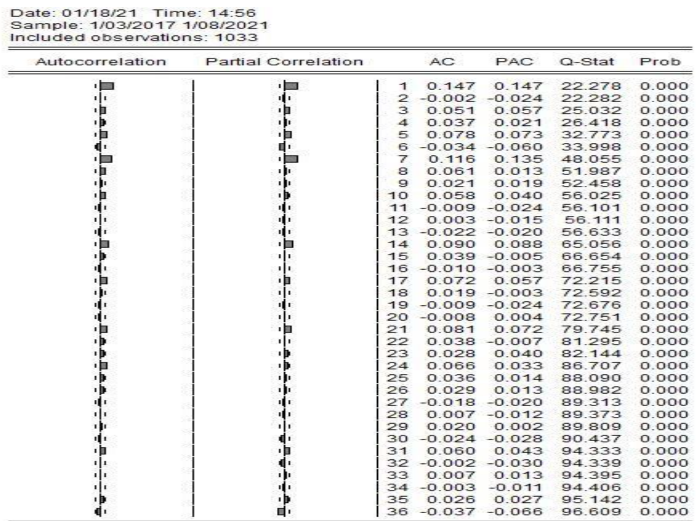
*MacKinnon (1996) one-sided p-values.

Chart (1) Results of Unit Root test by using ADF test for the original time series

For the series being unstable, it required to mention the first different for the watch ∇y_t, we found the the losing the general direction in the attitude, this can be ssen in graph(3), it explain the time series stability, the graph (4) which represents the autocorrelation function and partial autocorrelation functions for the time series (after chooseing the first different), we also can note that the Unit Root test by using ADF method , we find P-value is (0.0000) which value is less than (0.05), here it is significance which we didn't agree with the null hypothesis, so the series was achieved to stability.



Graph (3) the time series curve after choosing the first different



Date: 01/18/21 Time: 14:56
Sample: 1/03/2017 1/08/2021
Included observations: 1033

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.147	0.147	22.278	0.000
		2	-0.002	-0.024	22.282	0.000
		3	0.051	0.057	25.032	0.000
		4	0.037	0.021	26.418	0.000
		5	0.078	0.073	32.773	0.000
		6	-0.034	-0.060	33.998	0.000
		7	0.116	0.135	48.055	0.000
		8	0.061	0.013	51.987	0.000
		9	0.021	0.019	52.458	0.000
		10	0.058	0.040	56.025	0.000
		11	-0.009	-0.024	56.101	0.000
		12	0.003	-0.015	56.111	0.000
		13	-0.022	-0.020	56.633	0.000
		14	0.090	0.088	65.056	0.000
		15	0.039	-0.005	66.654	0.000
		16	-0.010	-0.003	66.755	0.000
		17	0.072	0.057	72.215	0.000
		18	0.019	-0.003	72.592	0.000
		19	-0.009	-0.024	72.676	0.000
		20	-0.008	0.004	72.751	0.000
		21	0.081	0.072	79.745	0.000
		22	0.038	-0.007	81.295	0.000
		23	0.028	0.040	82.144	0.000
		24	0.066	0.033	86.707	0.000
		25	0.036	0.014	88.090	0.000
		26	0.029	0.013	88.982	0.000
		27	-0.018	-0.020	89.313	0.000
		28	0.007	-0.012	89.373	0.000
		29	0.020	0.002	89.809	0.000
		30	-0.024	-0.028	90.437	0.000
		31	0.060	0.043	94.333	0.000
		32	-0.002	-0.030	94.339	0.000
		33	0.007	0.013	94.395	0.000
		34	-0.003	-0.011	94.406	0.000
		35	0.026	0.027	95.142	0.000
		36	-0.037	-0.066	96.609	0.000

Graph (4) autocorrelation function and partial autocorrelation function for time series after choosing the first different

Null Hypothesis: D(Y) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=21)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-27.68414	0.0000
Test critical values:	1% level	-3.436468	
	5% level	-2.864130	

10% level	-2.568201

*MacKinnon (1996) one-sided p-values.

Chart (2) Results of Unit Root test by using ADF test for the original time series After taking the first difference

4.

specify the suggestion module :

Chart (3) explains five suggestion modules ARMA and the standers of choosing good module, these standers are (AIC, SIC, H-QIC), ARMA (1,1,1) has been chosen as the best one to represent the data.

	ARIMA (1,1,0)	ARIMA (0,1,1)	ARIMA (1,1,1)	ARIMA (2,1,1)	ARIMA (1,1,2)	ARIMA (2,1,2)**
Akaike info criterion	2.976055	2.975118	2.973432	2.971508	2.971961	2.964085
Schwarz criterion	2.995185	2.994248	2.997344	3.000203	3.000655	2.997562
Hannan-Quin criter.	2.983314	2.982377	2.982506	2.982397	2.98285	2.976788

Chart (3) Contrasting between the suggestions modules

5. Estimation of the Model Parameters

we used Eviwes program to estimate the module statics as :

Dependent Variable: DY
Method: ARMA Generalized Least Squares (Gauss-Newton)
Date: 02/02/21 Time: 21:15
Sample: 1/04/2017 1/08/2021
Included observations: 1033
Convergence achieved after 53 iterations
Coefficient covariance computed using outer product of gradients
d.f. adjustment for standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.665751	101.6309	0.065588	0.9477
X	-9.04E-06	0.000138	-0.065550	0.9477
AR(1)	0.242458	0.112225	2.160465	0.0310
AR(2)	0.659906	0.093735	7.040141	0.0000
MA(1)	-0.100437	0.103729	-0.968263	0.3331
MA(2)	-0.723554	0.079553	-9.095186	0.0000

Chart (4) the results of estimating of suggestion modul (2,1,2) ARIMA

So the estimating formula for the sample will be :
DY = C(1) + C(2)*X + [AR(1)=C(3),AR(2)=C(4),MA(1)=C(5),MA(2)=C(6)

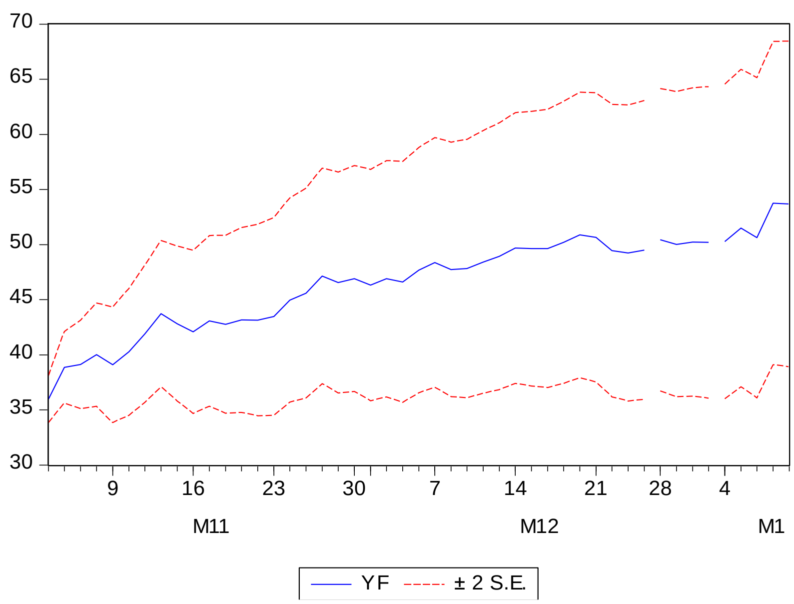
6

. Forecasting

The oil price was predicted by using the sample Estimator for a month and the results that explained in chart (5)m the time series was represented as graph (5) which can be seen that it has the same behaviou of the original series.

35.97	43.07	46.32	49.63	50.23
38.86	42.76	46.90	49.64	50.19
39.12	43.16	46.61	50.19	50.28
40.01	43.14	47.69	50.87	51.50
39.09	43.48	48.38	50.64	50.61
40.26	44.96	47.74	49.44	53.76
41.90	45.60	47.81	49.23	53.69
43.73	47.15	48.41	49.50	
42.81	46.55	48.93	50.44	
42.07	46.92	49.69	50.02	

Chart (5) estimating oil prices (3/11/2020 - 18/01/2021)



Forecast: YF	
Actual: Y	
Forecast sample: 11/03/2020 1/08/2021	
Included observations: 47	
Root Mean Squared Error	0.842643
Mean Absolute Error	0.626978
Mean Abs. Percent Error	1.371933
Theil Inequality Coefficient	0.008980
Bias Proportion	0.149741
Variance Proportion	0.020636
Covariance Proportion	0.829623

Graph (5) value curve that estimating for oil prices for (3/11/2020 - 18/01/2021)

Conclusion :

1- According to estimates that have been calculated, the oil prices even with the world finicial crises will rise in the next months

2- Box- Jenkins is the best method for estimating the time series for this study.

3. depend on ARIMA (2,1,2) for represent the time series.

Recommendations

According to the results of our study we recommend :

1- Benefit from high oil prices can be used to develop and establish Hydrocarbon projects, also increase the investments by the Foreign companies.

2. Benefit from oil oprices to develop the industrial projects.

3. Benefits from the oil incomes for Improving the living siruation of the Iraqi citizen and improving projects.

References

1- R. H. Saif "Use Box-Jenkins models for predicting traffic accidents in AL-Qadisiya province", Al-Muthanna Journal of Administrative and Economic Sciences, vol.7, pp. 190-200, 2017.
2- Box- G. E. P & Jenkins, G.M.T.(1976), "Time series Analysis Forecasting and Control, San Francisco", Holden-day ,USA.
3- Cryer ,Johnathan .D(1986) "Time series Analysis", R .R .Donnelley & Sons Com .USA.
4- O‘Donovan ,Thomas M,1983"Short Term Forecasting ,An Introduction to the Box-Jenkins "Johan Wiley and Sona ,New York.
5- Box ,G.E.P.& Pierce , D.A,1970,"Distrubution of the Residual Autocorrelation in Autoregressive-integrated moving Average Time Series Models" , JASA,VOL.65,P.
6- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.
7- Kendall, M. G. (1953). The Analysis of Economic Time-Series—Part I: Prices. Journal of the Royal Statistical Society: Series A (General), 116(1), 11-25.
8- Mills, T. C. (2019). Applied time series analysis: A practical guide to modeling and forecasting. Academic press.
9- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). Introduction to time series analysis and forecasting. John Wiley & Sons.
10- Agung, I. G. N. (2018). Advanced Time Series Data Analysis: Forecasting Using EViews. John Wiley & Sons.