DOI: https://doi.org/10.24297/jam.v23i.9576

Using The Box-Jenkins Method In Time Series To Predict The Monthly Electrical Loads In (Babylon Governorate - Shomali District)

Hayder Kadim Mohammed[1] , Ali Kazim Jari[2] , Wissam Sadiq Khudair[3]

[1,2]Directorate of Education Al-Qadisiyah, Ministry of Education, Iraq

[3] Directorate of Education Babylon, Ministry of Education, Iraq.

[1] Header.Kadim@gimal.com ,[2] alshabanijj@gmail.com [3] wissamhsse12@gmail.com

Abstract:

The topic of time series analysis is considered one of the important statistical topics to explain the phenomena that occur during a specific period of time. Time sequence examination objects to find an accurate account of the sequence, build a suitable perfect to interpret its behavior, and use the effects to predict the future time series . We using the Box-Jenkins method in the period sequence to predict the monthly electrical loads in (Babylon Governorate - Shomali district), and we have found that the studied time series is unstable in the mean and variance, we note that the time series is stable in the nasty and alteration. Autocorrelation and incomplete autocorrelation coefficients are used for the original data. Through these coefficients, we conclude that the appropriate model for the data is (3-1-2) ARMA. This model was chosen as it obtained the least (ARAM), and thus the model is appropriate for the data and the use of predictive values until the year (2022).

Keywords: Box-Jenkins Method, Time Series Analysis, Appropriate Model For The Data Is (3-1-2) ARMA, Period Sequence.

2010 AMS Mathematics Subject Classification: 37M10, 62M10**.**

Introduction

The losses in electrical energy are a natural phenomenon in every stage of the electrical system and within internationally acceptable rates and occur as a result of the passage of electric current in the different stages. As for the scientific side, this term (electricity) is ambiguous with other concepts related to it, which we are not going to discuss. The science of electricity It did not witness any progress until the 17th and 18th centuries, as applications related to electricity remained scarce and engineers could not use electricity in an applied manner in the industrial field. Residential housing was not used until the end of the 19th century. and multiple and surprising uses of electricity as a source of energy were not witnessed. showed the possibility of using it in an infinite number of applications in life [1].

Static According to one of the controversial theories, the people of the Parthians knew electroplating based on the discovery of the Baghdad battery in 1936 [2] Benjamin Franklin also conducted comprehensive research on electricity in the eighteenth century, although the early nineteenth century witnessed progress in the science of electricity because in the late. The same century witnessed the greatest progress in the field of electrical engineering from a mere baffling curiosity to a major and indispensable tool in modern life and became the defining force of the second industrial revolution [3]. There are many studies and research related to forecasting using time series models, the most prominent of which are:

In 1993, the researcher dealt with hadiths [5] in his research, where he used the Box-Jenkins method to predict the production of the yellow corn crop in Iraq, based on the annual data of the yellow corn crop for the period (1941-1990), which he obtained from the Central Statistical Organization, in the Ministry of Planning, as He reached to build a forecasting model by which the yellow corn crop production was predicted in Iraq for the year 2000. In 2004, the researcher Ibrahim [4] predicted the temperatures in Khartoum state using one of the Box-Jenkins models for the time series, and in this research, one of the Box-Jenkins models was applied to the monthly averages of temperatures in the state of Khartoum for the period (1991-2001), and the appropriate model was found. ARIMA(1,0,0)x(0,0,1) 12 Monthly temperature forecasts were made for the year 2002.

In 2012, the researchers Omran and Zaalan [8] conducted a study, and it was concluded that the variables that affect the consumption of electric energy in the Kingdom of Saudi Arabia were destroyed, namely the variables of gross domestic product, the number of residents, and the variable of electric energy consumption in the previous periods, and it was concluded that the best model For prediction is the Box-Jenkins model. In 2013, researchers Al-Safawi and Ghanem [6] conducted a comparative study between the chain prediction methods for the daily peak load of electric power in the city of Mosul, and after addressing the instability, the ARIMA (0,1,2)x(0,1,1) model was chosen to represent

Data of the daily peak load series of electric power in the city of Mosul, because this model has the lowest value for the criteria of Akaki information and Bayes information for Schwartz compared to the Box-Jenkins models, and therefore it is the most suitable model for the series that can be used for prediction purposes.

Also, in 2013, the researcher Ghafel [10] he Box-Jenkins methodology was adopted in building the time series model, and then choosing the best model to predict the future values of electric power production. A practical application was made on the time series For the production of electric power for six years including (69 months) using the statistical program Statgraphics, and the results indicate that the ARIMA (1,0,2) model is better than the ARIMA (1,0,1) model, according to the performance measures of the prediction methods. In 2016, the researcher Omar [9] studied the prediction of the population of Iraq using Boxes-Jenkins models until 2020. The study concluded that the appropriate model is (3, 1, 3) ARIMA because it has the least average square error for this category of data.

In this research, the use of the Box-Jenkins method to predict the monthly electrical loads in Babil Governorate for the period from (2018-2022) for a period of five years, represented by the duration and by months as a time series, was studied and an analytical study was conducted using Box-Jenkins models.

Literature Review :

Box-Jenkis style

The researchers developed a number of statistical predictive methods and models that have proven their efficiency and the accuracy of their results in many cases. Among these methods is the (Box-Jenkins) method, which is based on a set of probabilistic models called Box-Jenkins models. It is used to represent the time series data of a particular phenomenon, as it is considered one of the General methods for finding future expectations of values appearing in the future within certain limits, which is one of the important, more advanced and complex methods than other methods, and is considered the most powerful and effective in many cases (3) Box Jenkins Models:

The first type: It is represented by non-seasonal models. It means models that deal with time series that do not contain the season element.

The second type: represented by seasonal models, which are those models that deal with the regular time series of the season element, and they repeat themselves, i.e. repeat their highs and lows after each fixed period of time.

A: AUTOREGRESSIVE (AR) MODELS:

It is denoted by (P) AR, anywhere P signifies the degree of the model, which is a positive integer:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t , \exists t, i = 1......p$$

(φ_1 ... .φ_p are modes , c constant , ε_t random variable).

B: MOVING MEDIA MODELS (AM):

It is usually denoted (q) AM, where q is a positive integer indicating the degree of the model:

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} , \exists t, i = 1......q$$

(θ_1........ θ_q are modes,μ The expected value of the variable for X_t , ε_{t-1} , ε_t independent variables for X_t).

C: Mixed models (autoregressive and moving media) ARMA: These models consist of merging the two previous models, the autoregressive model and the moving media model. This model is the most commonly used Box Jenkins model due to its flexibility and suitability for various types of data. This model is symbolized by the symbol (q, p) ARMA, where (q, p) represents the two degrees of the model. The general form of this model can be written in the following form:

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Stages of building a Jenkins box

The process of building the model for the time series needs great experience and effort, especially for (BOX-GENKIS) models, and this building process is an iterative method. It takes the following stages (3).

1- Diagnosis

2- Estimation

3- Diagnostic checking

4- Forecasting

1- Diagnosis: The diagnosis stage of the model is one of the most important stages in time series analysis. The diagnosis of the model is based on the studied data set and on understanding the basic characteristics of the time series, especially the autocorrelation and partial autocorrelation functions. Before starting the diagnosis and choosing the appropriate model to represent the time series data, it must be tested first to identify what If those chains are stable or not, to treat them, then if they are stable, we move on to studying and defining the appropriate model, but if it is the other way around, we must make them stable either by using differences if the instability is in the mean, or by using transformations.

The diagnostic stage includes the following steps:

A- We draw series data, and drawing data is the first step in analyzing any time series. Through drawing, we have a good idea of whether the series contains seasonality, a general trend, abnormal values, or instability that leads to possible transformations in the data.

The drawing of the series shows its need for the appropriate transformation to settle in its averages or variances before any analysis.

b- We calculate and check the IACF, PACE, ACF of the sample drawn from the original series to determine the degree of differences (in the case of instability). Where $d<0$ (and it is often ($d=0,1,2$) and that the consequences of using unnecessary differences are less serious than the consequences of underestimating the importance of differences.

c- We calculate and check the IACF, PACE, ACF of the sample to diagnose the model. There is a binary between ARMA(1,0) or AR(1) models and ARMA(0,1) or MA(1) models according to the three functions. The problem becomes more complicated in the case of mixed models (p, q) ARMA, because relying on IACF, PACF, ACF to diagnose the model and determine its rank is not effective, since the above functions in this case behave similarly, which is the behavior of gradual decrease.

2 - Estimating the parameters: After defining the proposed model to represent the studied phenomenon, then comes the estimation of the parameters, as there are several methods.

1- The method of determination.

2- OILS method of least squares.

3- The M.L.E. Most Likelihood Method.

4- Linear estimation method.

5- Conditional least squares method.

3- Choosing the accuracy of the model: At this stage, the model is tested to see its suitability for representing the studied apparent data and using it to obtain future predictions, as there are many tests that can be used for this purpose, which depend in their calculation on the residuals to detect whether there is any factor other than randomness. Within these residuals, as we assume that these residuals are random and devoid of any effect, and from these tests:

A- The box-pierce test: This test is symbolized by the symbol (q), as it was based on the equality of autocorrelation to zero. This means testing the null hypothesis.

In this test, the computed value of (q) is compared with the tabular value of X^2 at a degree of freedom (h-m) and a certain significant level.

B- The BOX-Liung test: This test is called the (Portmantean) test, and this test is symbolized by Q. In general, the ACF autocorrelation function for stable time series has a special form, as it decreases with increasing degrees of errors. The autocorrelation function of the stable series also It decreases quickly and is close to zero, as the calculated value of Q is compared with the tabular value of X^2 at a temperature (H-M) and a certain level, and then the judgment is given as in the previous test.

4- Forecasting: Forecasting is the ultimate goal of time series analysis, and the transition to this stage is not done until after making sure that the initial model that was tested in the identification stage and a transaction was estimated had passed all the tests that were conducted with high efficiency. This method has made a distinct qualitative leap in The field of time series analysis so that it has become the most widespread and used method by researchers because of its many advantages that ultimately lead to obtaining an excellent system and reliable prediction in most time series. In other words, the use of this method means getting rid of the defects of other traditional methods, i.e. used in Time series analysis.

Results And Discussion :

- **Data collection method**

We must, while addressing the practical side of our research, talk about the method through which the data was collected, as the Electricity Distribution Directorate was used in the form of paper data, then it was entered into the Excel program and extracted in the form of the following tables:

20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	month
162	163	149	114	99	100	106	93	87	92	97	111	112	122	96	69	74	82	73	70	R

41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21
117	100	111	121	117	129	131	123	137	138	163	153	138	125	134	135	128	122	119	134	157

62	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	24
173	169	153	171	185	189	192	171	159	140	138	132	128	127	129	131	132	150	159	140	121

83	82	81	80	79	78	77	76	75	74	73	72	71	70	69	68	67	66	65	64	63
250	274	323	316	323	341	284	239	241	300	297	217	265	253	277	290	292	227	169	177	182

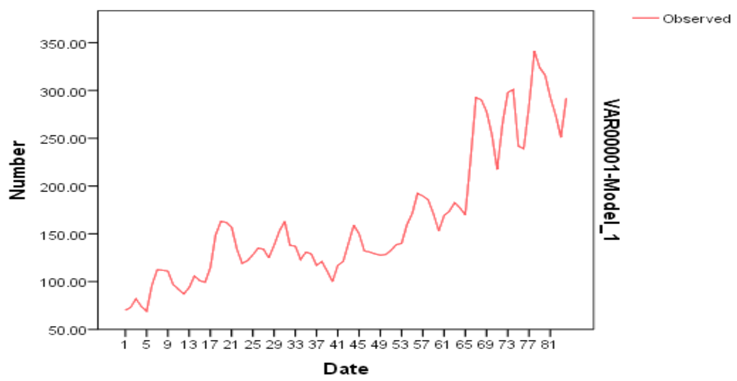

Figure (1) Time series drawing

We notice that the time series is unstable in the arithmetic mean and variance, and that when we take the first difference, we notice the following drawing .

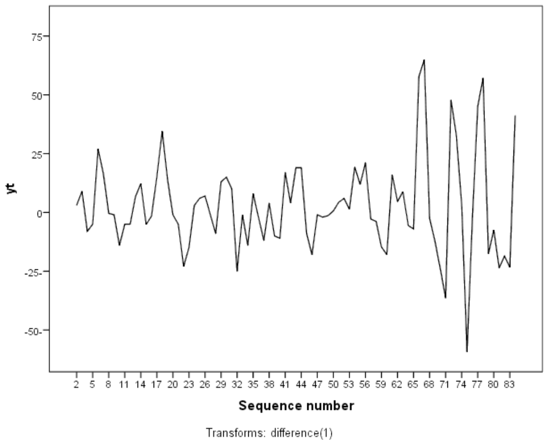

Figure (2) Drawing the time series (a)

We notice that the time series is stable in variance and unstable in the arithmetic mean. When taking the first differences and logarithms, we notice the following drawing.

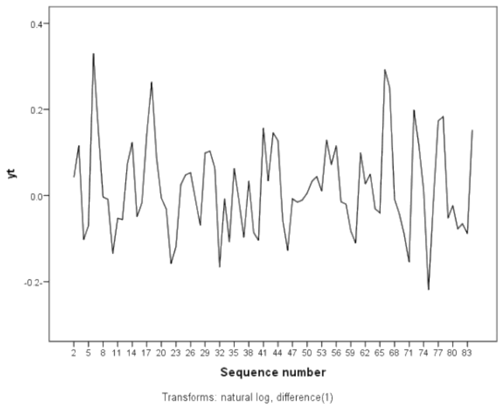

Figure (3) Drawing the time series (b)

We notice that the time series is stable in the arithmetic mean and variance, noting that the data are normally distributed.

When analyzing the data using the(SPSS) statistical program, the following results were obtained: autocorrelation and partial autocorrelation function drawing.

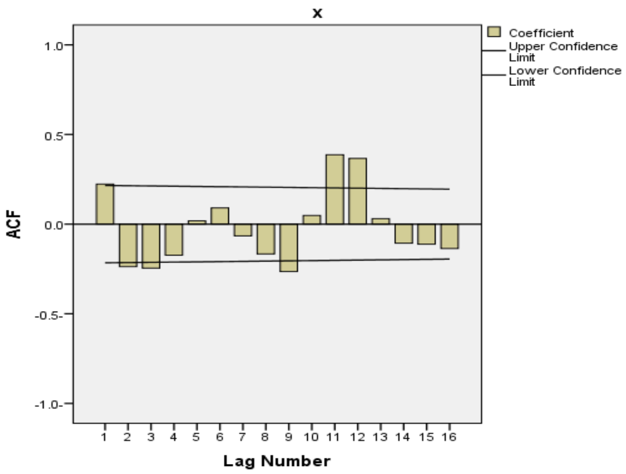

Figure 4: Autocorrelation function for time series

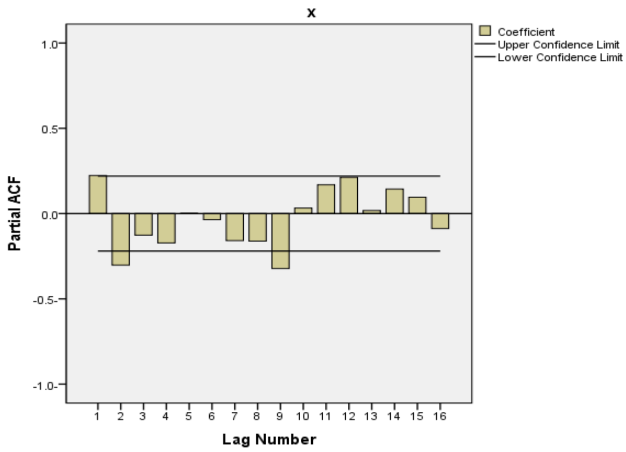

Figure (5) Partial correlation function for the time series

Autocorrelations

Series: x

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	.223	.108	4.291	1	.038
2	-.237-	.107	9.191	2	.010
3	-.245-	.106	14.505	3	.002
4	-.173-	.106	17.192	4	.002
5	.018	.105	17.223	5	.004
6	.090	.104	17.971	6	.006
7	-.065-	.104	18.369	7	.010
8	-.166-	.103	20.964	8	.007
9	-.264-	.102	27.618	9	.001
10	.048	.102	27.839	10	.002
11	.387	.101	42.503	11	.000
12	.366	.100	55.849	12	.000
13	.030	.100	55.941	13	.000
14	-.106-	.099	57.097	14	.000
15	-.112-	.098	58.395	15	.000
16	-.136-	.097	60.352	16	.000

a. The underlying process assumed is independence (white noise).
b. Based on the asymptotic chi-square approximation.

Partial Autocorrelations
Series: x

Lag	Partial Autocorrelation	Std. Error
1	.223	.110
2	-.302-	.110
3	-.126-	.110
4	-.172-	.110
5	.001	.110
6	-.035-	.110
7	-.158-	.110
8	-.161-	.110
9	-.322-	.110
10	.032	.110
11	.169	.110
12	.212	.110
13	.018	.110
14	.144	.110
15	.096	.110
16	-.087-	.110

It is clear from the drawing of the autocorrelation and partial autocorrelation functions that the time series can follow one of the autoregressive models with the moving average (ARMA), which is the (3-1-2)ARMA model, and we will use the kies Box-jun model.

No	ARIMA	RMSE	MAPE
1	1-1-1	20.601	8.456
2	1-1-0	19.908	8.260
3	1-1-2	18.413	7.823
4	1-1-3	17.843	7.748
5	2-1-0	18.273	7.941
6	2-1-1	17.551	7.776
7	2-1-2	17.664	7.776
8	2-1-3	17.689	7.794
9	3-1-0	17.604	7.921
10	3-1-1	17.676	7.885
11	**3-1-2**	**17.497**	**7.854**
12	3-1-3	17.519	7.703
13	0-1-1	10.393	7.997
14	0-1-2	19.326	8.111
15	0-1-3	17.751	7.831

- **Model (3-1-2) ARMA :**

a) ARIMA Model Parameters

					Estimate	SE	T	Sig.
VAR00001-Model_1	VAR00001	No Transformation	Constant		2.747	1.224	2.245	.028
			AR	Lag 1	-.153-	.238	-.643-	.522
				Lag 2	.176	.232	.760	.450
				Lag 3	-.562-	.128	-4.390-	.000
			Difference		1			
			MA	Lag 1	-.388-	.267	-1.452-	.150
				Lag 2	.424	.239	1.774	.080

b) Residual analysis: The residuals or prediction errors are the true values minus the values

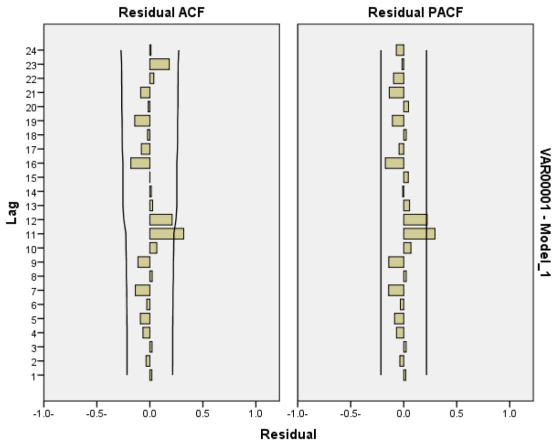

Figure (6) The autocorrelation and partial autocorrelation functions of the residuals

We note from the drawing of the autocorrelation and partial autocorrelation functions. It is clear from drawing the autocorrelation and partial autocorrelation of the user model that most of the values fall within the limits of confidence, and this indicates the randomness of the residuals. Therefore, the used model is an acceptable model that can be relied upon in analyzing and calculating future predictive values.

c) Forecasting

Forecasts from period 84
95% Limits

Period	Forecast	Lower	Upper	Actual
85	313.658	279.374	347.941	
86	328.402	274.191	382.613	
87	309.884	246.766	373.002	
88	307.517	242.615	372.418	
89	299.097	233.245	364.950	
90	315.284	248.019	382.549	
91	316.338	245.339	387.336	
92	329.094	254.091	404.097	
93	321.729	242.844	400.613	
94	329.422	248.205	410.639	
95	323.050	239.559	406.541	
96	334.483	249.066	419.900	
97	330.753	242.713	418.794	
98	341.999	251.610	432.387	
99	336.654	243.592	429.716	
100	346.540	251.416	441.663	

101	341.168	243.753	438.582
102	351.719	252.426	451.012
103	347.046	245.498	448.594
104	357.649	254.197	461.101
105	352.726	247.055	458.397
106	362.962	255.477	470.448
107	358.013	248.438	467.588
108	368.321	257.019	479.623
109	363.573	250.248	476.898
110	373.880	258.865	488.895
111	369.135	252.149	486.121
112	379.320	260.696	497.944
113	374.597	254.070	495.123
114	384.749	262.636	506.861
115	380.109	256.149	504.069
116	390.227	264.720	515.733
117	385.632	258.325	512.938
118	395.684	266.869	524.498
119	391.130	260.563	521.697
120	401.134	269.097	533.171
121	396.637	262.891	530.383
122	406.596	271.413	541.778
123	402.148	265.296	539.000
124	412.054	273.797	550.312
125	407.653	267.764	547.542
126	417.511	276.246	558.776
127	413.160	270.298	556.021
128	422.970	278.760	567.179
129	418.667	272.894	564.440
130	428.428	281.333	575.524
131	424.174	275.546	572.801
132	433.887	283.961	583.812

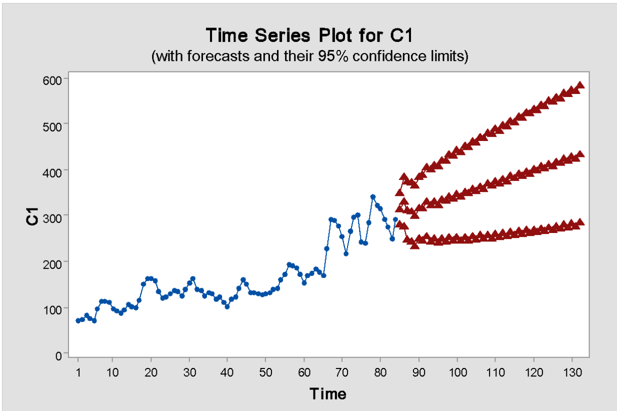

Figure (6) Plotting predictive values

Conclusion And Recommendations

- **conclusions**

1) We conclude from the time series that the variance and the mean are unstable
2) Through the coefficients of autocorrelation and partial autocorrelation, we found that the model is not stable for the time series data
3) We notice that there is seasonality in the time series, where there is a rise in the summer season and a decrease in the winter season
4) The stability of the time series after taking the natural logarithm and taking the first difference. The ARIMA model (3-1-2) was chosen as the best model for the RMSE and MHPE standards as the lowest standards.

- **Recommendations**

1) We recommend taking into account the results of the research, especially in the Electricity Department Directorate in Al-Qadisiyah Governorate and related departments.
2) We recommend the researcher to use the ARIMA (3-1-2) model to predict future electrical loads.

References

1. Al-Naqash, Iftikhar Abdul-Hamid: Time series analysis to predict the consumption of electrical energy for the city of Baghdad. Master's thesis ' College of Administration and Economics, University of Baghdad (1982)
2. Al-Jader, Buthaina Abdel-Aziz: Application of Box-Jenkins models to time series for rain forecasting. Master's thesis, to the Department of Statistics 'College of Administration and Economics, University of Baghdad (1982)
3. Ameen, Mohamed Abdel-Aal, Analytical Statistical Study of Electric Energy Consumption Models in Iraq 'Master's Thesis, Department of Statistics, College of Administration and Economics, University of Baghdad (2006)
4. Zaki, Izzat Hazem: The use of neural networks in predicting time series by applying them to the consumption of electric energy in Nineveh Governorate, master's thesis, College of Computer Science and Mathematics, University of Mosul (2004)
5. Ibrahim, Bassam Younes, "Temperature Prediction in Khartoum State Using Box-Genghis Time Series Models," University of Sudan, research published in the Sudan Journal of Science and Culture, Volume V, Issue 2, (2004)
6. Al-Hadithi, Issam Mouloud, "A study on forecasting the production of the yellow corn crop in Iraq for the period (1941-1991), master's thesis in statistics, University of Baghdad, College of Administration and Economics, (1993)
7. Al-Safawi, Safaa Younis and Ghanem, Iman Ibrahim, "A Comparison of Prediction Methods for the Daily Peak Load Chain of Electric Energy in the City of Mosul," Iraqi Journal of Statistical Sciences, special issue of the proceedings of the Sixth Scientific Conference of the College of Computer Science and Mathematics, (2013)
8. Tohma, Saadia Abdel-Karim, "Using Moses time series models to predict the consumption of electric energy in the city of Fallujah," Anbar University Journal of Economic and Administrative Sciences, Volume Four, Number Seven, (2011).
9. Omran, Kholoud Musa and Zaalan, Risan Abdel-Imam, "The Use of Some Statistical Methods to Predict Electric Energy Consumption in the Kingdom of Saudi Arabia," Journal of Economic Sciences, Volume VIII, Issue 29, (2012).
10. Omar, Fawzia Ghaleb, "Forecasting the Population of Iraq Using Box-Jenkins Models to 2020," Economic Sciences, Volume Eleven, Issue 4, (2016).
11. Ghafel, Amal Ali, "Using Box-Jenkins ARIMA Models in Predicting Electric Energy Production," Karbala University Journal, Volume Eleven, Issue Two, (2013).
12. Al-Ghannam, Muhammad bin Abdullah, "Analysis of the time series of the stock price index in the Kingdom of Saudi Arabia using the Box Genghis methodology," a research published in the Journal of King Abdulaziz University, Economics and Administration, Volume 12, Issue 2, (2003).