International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 13, Issue 5, 2020

Predicting the Numbers of Primary Schools in the City of Qadisiyah

Assistant Lecturer Ali Hamdullah Ahmed[a], Assistant Lecturer Hassanien Adel Salih[b], [a]Department of Educational Planning, Directorate-General for Education of Qadisiyah, [b]Department of Educational Planning, Directorate-General for Education of Najaf, Email: [a]Alihamdullah8220@gmail.com, [b]Hassanienadel19822020@gmail.com

The purpose of this study is to analyze the time series of the number of primary schools in Qadisiyah City. Also the development of a standard model to predict the future values of the number of schools that will be required in the future, and to contribute to the preparation of future plans in a scientific method, because the primary schools represent the base of the educational pyramid.

Key words: *Primary Schools, Model for predicting number of schools,*

Introduction

Iraq is characterized by free primary education and education at this stage is compulsory and this privilege, which is accompanied by high population growth rates, especially for the age group (1-12 years), as well as other factors has led to a problem in providing sufficient number of primary schools to accommodate the number of pupils. Hence the importance of future planning and forecasting according to studied and sober scientific contexts to provide an adequate number of schools and thus to develop a clear picture of decision makers and central and local governments for the importance of the primary stage, where this stage is the basis on which the rest of the educational stages are built. Therefore, primary education in any country is of great importance because it has a fundamental role in its advancement and progress, where primary schools represent the field where the children of society meet to acquire the basics of knowledge and science.

Research objective

This research aims to analyze time series and find a suitable formula for predicting the number of schools expected for each academic year through the use of the Box Jenkins method.

Research Methodology

The research includes two aspects, the theoretical and the applied aspect where on the theoretical side are reviewed some basic concepts of time series and the proposed method of forecasting and in the applied aspect is shown to analyze the time series of variables as well as the presentation of the predicted results as well as the most important results and recommendations reached.

The Theoretical aspect

Time series

On this side, some of the general and used concepts are presented in this study, which is the analysis of the time series where we can define the time series as (A group of observations that are linked to each other and these observations are recorded in successive periods of time and for a particular phenomenon). These series are of two types. They are intermittent time series and continuous time series. It must be noted that most of the time series used in the practical field are intermittent time series in which the time period is equal between any two observations.

Time series stability test

What is meant by Stability is that for both the mean and the variance of the time series to be constant, and most studies using time series data, the series are stable or unstable, but most of the time series are unstable, instability may be caused by the instability of the conditions surrounding the variable. By drawing the time series as well as the Auto-correlation function (ACF) and the partial auto-correlation function (PACF), we can judge the stability of the series or its lack of stability and the reason for the instability of the time series can be attributed to one of the following reasons:

1- There is a secular trend
2- The presence of seasonal fluctuations
3- The instability of the variance and the mean

There are several methods used to stabilize the time series, the most important of which are:

A- In the case of a secular trend for the series:
The most important methods used to remove the secular trend are:
1- Linear regression method is to estimate the secular trend, and then isolate and treat residuals as a stable time series.
2- Method for taking the differences: In this method, the values of observations are subtracted from each other for certain periods, as the difference from the first degree takes the following form

$$C_t = \nabla D_t = D_t - D_{t-1}$$

The difference from the second degree takes the following form

$$C_t = \nabla^2 D_t = \nabla D_t - \nabla D_{t-1}$$

B- In case of seasonal fluctuations:

To get rid of the seasonal element in the time series, we use the seasonal difference method by subtracting the values from each other according to the periods, for example

Quarterly differences $Ct = Dt - Dt\text{-}4$
Monthly differences $Ct = Dt - Dt\text{-}12$

C - In case of instability of variance and mean

In this case, the natural logarithm of the string data is obtained, the square root of it, or the reciprocal of the data, is obtained.

It should be noted that the most important statistical criteria used to describe the quality of the time series and its model are:

Auto-correlation function (ACF)

The self-correlation function can be defined as a measure of the degree of relationship between string values with itself for a different time period and is of great importance because it is one of the methods of determining whether the random process is stable or not, the mathematical formula for finding the function of self-correlation of the random process at displacement k is as follows:

$$\rho_k = \frac{\operatorname{cov}(X_t, X_{t+k})}{\sqrt{\operatorname{var}(X_t)\ \operatorname{var}(X_{t+k})}} = \frac{E(X_t - \mu)(X_{t+k} - \mu)}{\sqrt{E(X_t - \mu)^2\ E(X_{t+k} - \mu)^2}}$$

When the random process is stable, the average and the variance are constant and equal for all periods

$$\sigma_{Y_t} = \sigma_{Y_{t+k}}$$

So the

$$\rho_k = \frac{\mathrm{cov}(X_t, X_{t+k})}{\sigma_{x_t} . \sigma_{x_{t+k}}}$$

The result is therefore as follows:

$$\rho_k = \frac{E(X_t - \mu)(X_{t+k} - \mu)}{\sigma_x^2} = \frac{\gamma_k}{\gamma_0}$$

Because the contrast is fixed to the stable series

$$\gamma_0 = E(X_t - \mu)^2$$

Its estimate is: $\gamma_0 = \frac{1}{N}\sum_{t=1}^{N}(X_t - \bar{X})^2$

And $\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$

The stable series contains observations (X1, X2, ..., Xn) for n from successive times

4-1 partial Auto correlation function

The (PACF) is defined as a measure of the degree of correlation between the two observations $X_t \quad X_{t+k}$,and the validation of the rest of the string values, meaning $corr(X_t, X_{t+k} / X_{t+1}, ..., X_{t+k-1})$

The partial Auto-correlation function is no less important than the Auto-correlation function, as it is also important in diagnosing the model and determining its degree.

Autoregressive Model

The mathematical model of Autoregressive (P) will be represented in the following form:

Zt=0+ 1zt-1+ 2zt-2+……..+ pzt-p+at

This equation represents the Auto-regressive model (P and Φj) representing the set of features and symbolizes this model with the symbol AR (p), which can be reformulated as follows:

(B) zt= 0+at

Where

(B) =1− 1B- 2B2- …….. –pBp

And where

Zt Values for series observations
i Model feature
0 Fixed limit
P Model Degree
at Random mistakes that are distributed N~(0,a2)

6-1 Moving Average Model

The general version of the Autoregressive model with (q) degree will be as follows
Zt = 0 + at – 1 at-1 – 2 at-2 -…….- q at-q

Where this model is called the moving average model of degree (q) and symbolized by the symbol MA (q)

Which can be rewritten as:

Zt = 0 + (B) at

Where

(B) = 1- 1B – 2B2 -……..- qBq

The ACF of the motion average model is approaching zero after displacement (q), while the PACF is steadily diminishing.

Auto Regressive-Moving Average Models (ARMA)

The form can be written in the general form of the degree (p,q) as follows

Zt =0+ 1zt-1+ 2zt-2+……..+pzt-q+at-1 at-1 – 2 at-2 -…….-q at-q

Which can be written as follows?

(B)zt = 0 + (B) at

Which is symbolized by the symbol ARMA (p,q) and the stability of the model can be known if the roots of the equation Φ (B)=0 fall outside the boundaries of the single circle.

8Box Jenkins Methodology

There are steps that must be taken before starting to use the Box-Jenkins models in prediction, and they are as follows

1- The stage of diagnosis and identification of the proposed model:
Includes data preparation by representing data to stabilize contrast, and taking nuances to obtain a stable series. Choosing the appropriate model through examining the data and using the PACF and ACF
2- The stage of estimating the parameters of the model and selection:
It includes estimating the parameters of appropriate models and selecting the best model using a suitable standard, diagnostic examinations to examine the ACF and the PACF.
3- Checking the estimated model:
After finding the estimates of the parameters of the proposed model, the accuracy of the model is checked by testing whether it is morally different from zero and the product test is that the estimated parameter differs significantly from zero if it falls outside the limits of (± 1.96) and (among the most common tests to examine the suitability of the model are the Q-box statistics (Box-Pierce), which are used to test the statistical significance of the self-correlations of the residuals according to the following formula.

4- Prediction:
Prediction is defined as a scientific method in research to reach knowledge of anonymous data through known data related to the content of the research.

Estimates of Missing Values

Often not all values are available or present, meaning that some values are missing (It may not be possible to collect these values in normal circumstances). The (SPSS) system gives the possibility of compensating these missing values by statistical methods, but these values are approximate, and there are several methods used to compensate the lost values, of the most important:

1- Series mean:

As the sample mean is used to make up for the lost value

2- Mean of nearby points:
Here, the missing value is compensated by taking the mean of the values surrounding the missing value

3- Median of nearby points:
Here the value is offset by taking the median of the values surrounding the missing value

4- Linear interpolation:
As the last value before the lost value and the first value after the lost value are rounded up, and compensation is not made if any of these values is lost

5- Linear trend at point:
Here the line equation is calculated for the series and one of the calculated values is chosen on the line.

In the subject of our study, the second method (Mean of nearby points) was used to compensate the values that are not present for the year 2003.

The Practical aspect

In this research, the data of an annual series of 35 observations was used representing the number of schools in Qadisiyah province, which was obtained from the Ministry of Planning/ Central Bureau of Statistics as well as from the Directorate General of Al-Qadisiyah Education/ Department of Educational Planning/ Statistics Division, as shown in the table below, the data of this series extends from the school year 1982 / 1983 to the 2016/ 17 school year, where this data is collected annually according to a form dedicated to this purpose by school administrations and collected and prepared the summary in the general directorates of education in the provinces and send the summary to the Ministry of Education, in view of the events of 2003 and the acts of sabotage and looting of Iraq, statistical data from the Central Bureau of Statistics and the Directorate General of Education in Al-Qadisiyah province have been lost, so the special data for this year will be estimated by using one of the methods of estimating lost values, which have been indicated in the theoretical aspect.

#	Year	Schools	#	Year	Schools	#	Year	Schools	#	Year	Schools
1	1982	370	11	1992	387	21	2002	409	31	2012	671
2	1983	366	12	1993	397	22	2003	462	32	2013	684
3	1984	362	13	1994	402	23	2004	467	33	2014	697
4	1985	369	14	1995	403	24	2005	512	34	2015	707

5	1986	363	15	1996	396	25	2006	536	35	2016	721
6	1987	367	16	1997	397	26	2007	557			
7	1988	364	17	1998	380	27	2008	588			
8	1989	366	18	1999	381	28	2009	608			
9	1990	369	19	2000	382	29	2010	631			
10	1991	378	20	2001	383	30	2011	649			

Drawing of the graph series

In order to analyze the time series, its observations are drawn to know the general direction of it, and Figure No. (1) Shows the general direction of the series

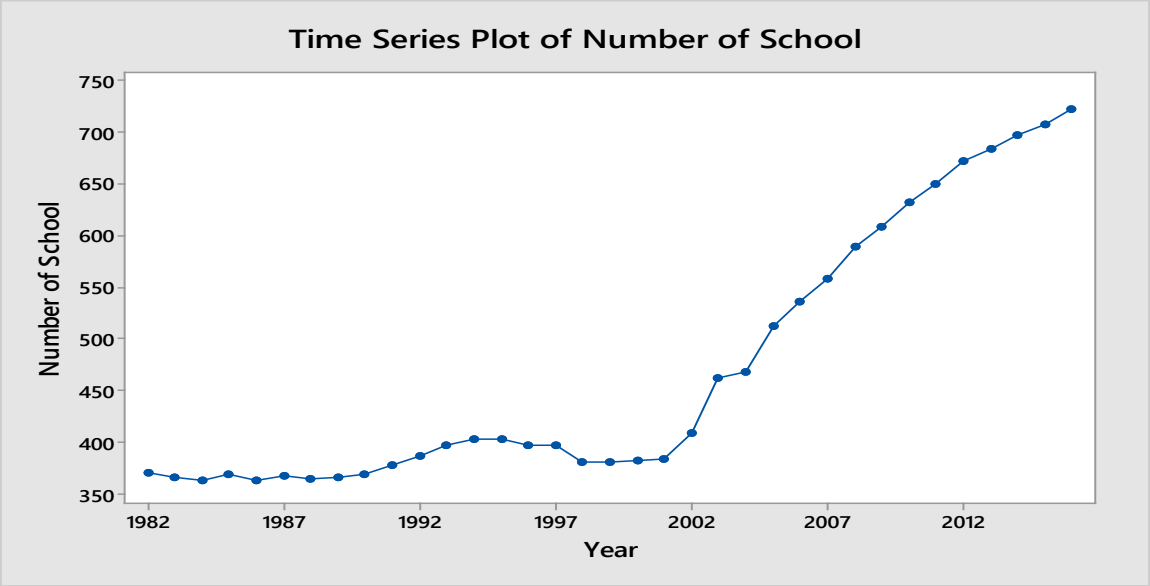


Figure (1) graph series

We observe by the drawing of the time series used and shown in Figure (1) an upward trend as we move forward, indicating instability in the time series (secular trend), as confirmed by the chart of the Auto-correlation function figure (2) and the partial auto-correlation as in Figure .(3)

Figure 2. Auto correction Function for No. school

Lag	ACF	T	LBQ	Lag	ACF	T	LBQ	Lag	ACF	T	LBQ
1	0.921937	5.45	32.37	6	0.433372	1.00	120.03	11	-0.012505	-0.03	128.95
2	0.836829	3.01	59.85	7	0.331027	0.74	125.10	12	-0.077390	-0.17	129.29
3	0.741544	2.17	82.11	8	0.234594	0.52	127.74	13	-0.123501	-0.27	130.19
4	0.640789	1.66	99.26	9	0.142415	0.31	128.75	14	-0.170009	-0.37	131.97
5	0.535603	1.29	111.64	10	0.061427	0.13	128.94	15	-0.194899	-0.42	134.43

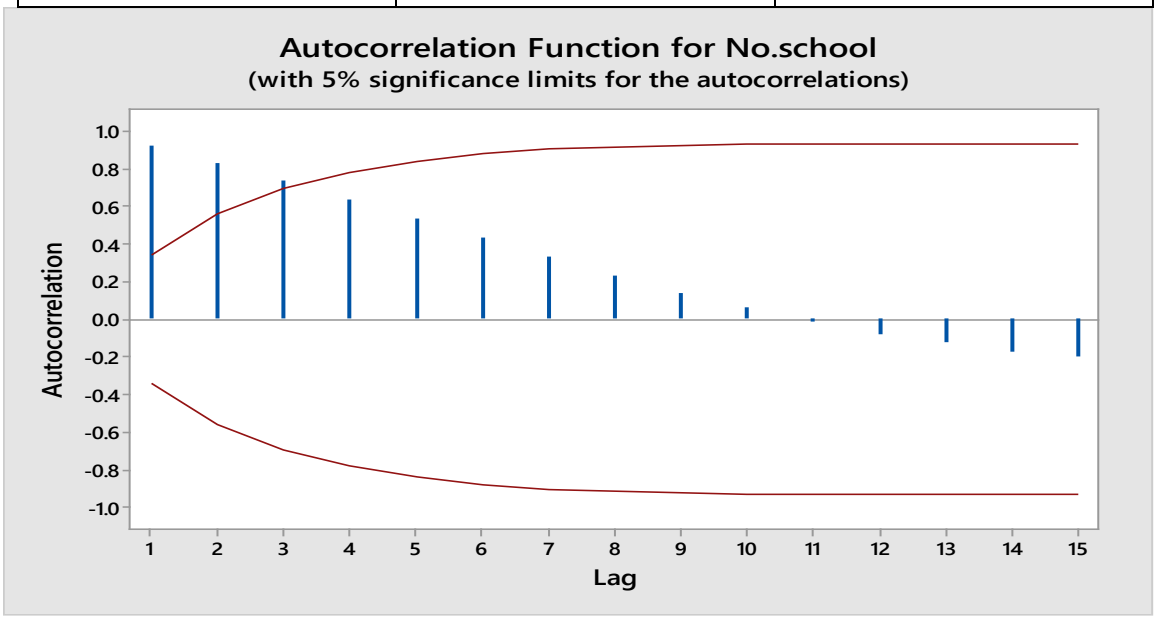
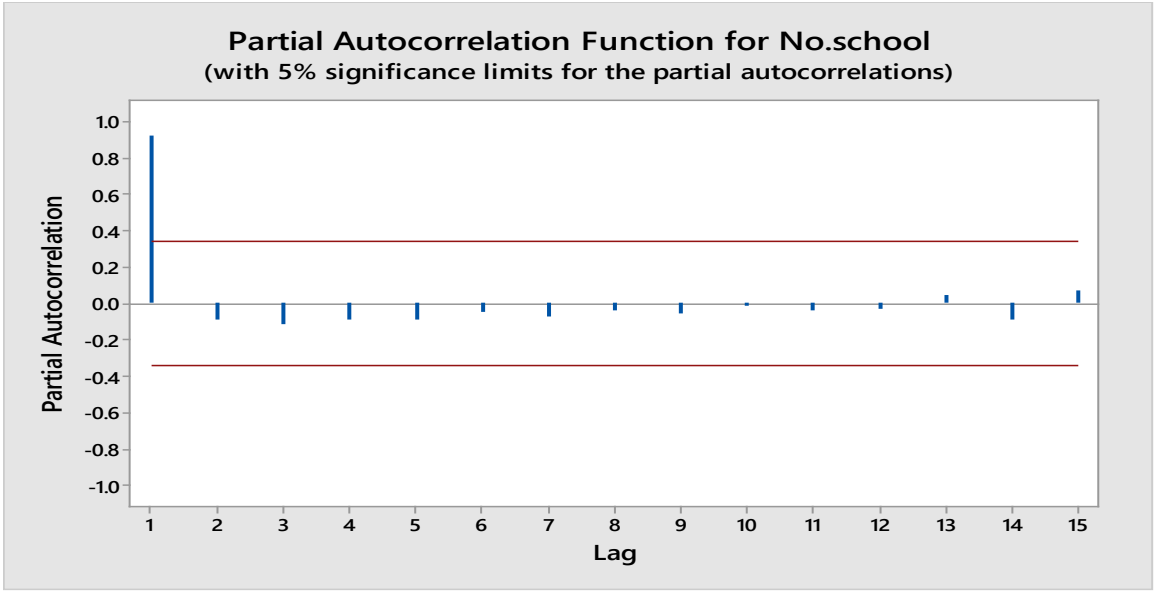


Figure 3. Partial Auto correction Function for No. school

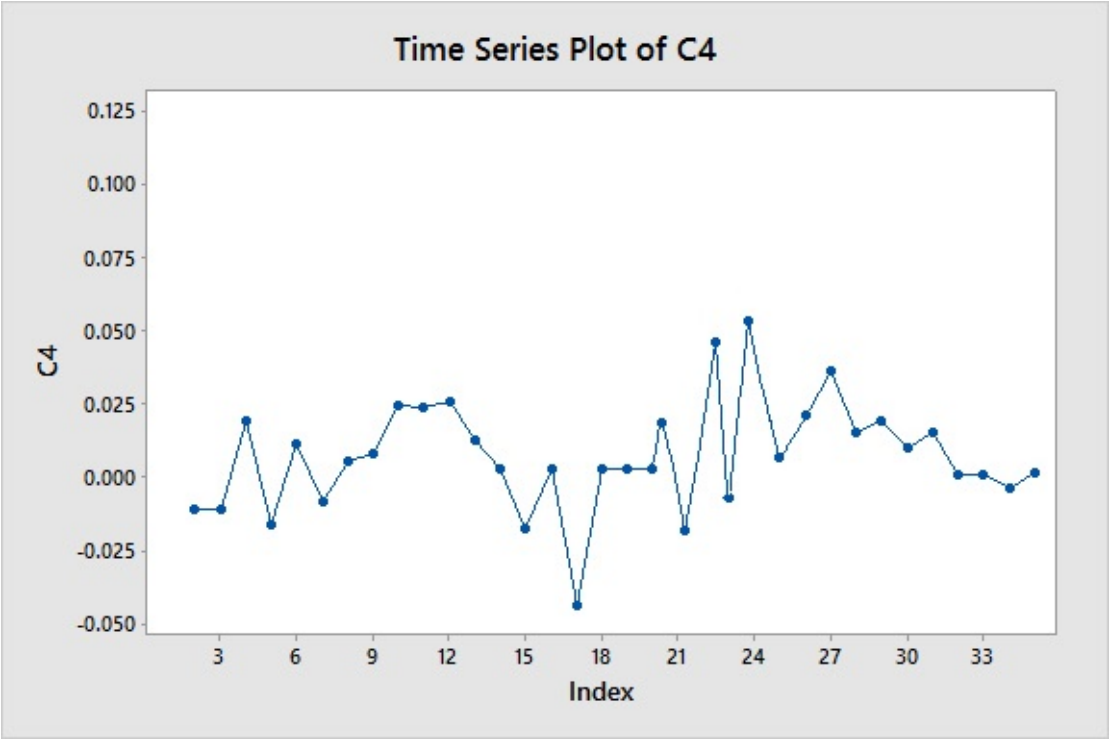


Partial Autocorrelation Function: No.school

Lag	PACF	T	Lag	PACF	T	Lag	PACF	T
1	0.921937	5.45	6	-0.043286	-0.26	11	-0.036766	-0.22
2	-0.087567	-0.52	7	-0.069821	-0.41	12	-0.027704	-0.16
3	-0.112755	-0.67	8	-0.037054	-0.22	13	0.044073	0.26
4	-0.086998	-0.51	9	-0.051446	-0.30	14	-0.085527	-0.51
5	-0.085987	-0.51	10	-0.009499	-0.06	15	0.068272	0.40

In order for the series to be stable, all the values of the auto-correlation coefficients of the series must be entered within the limits of confidence only the first or second displacement, it may be outside the limits of confidence and for the purpose of stabilizing the time series we will take the first difference of observations note the loss of the secular trend in his behavior and this is what we note in figure (4).

Figure 4 time series plot of C4

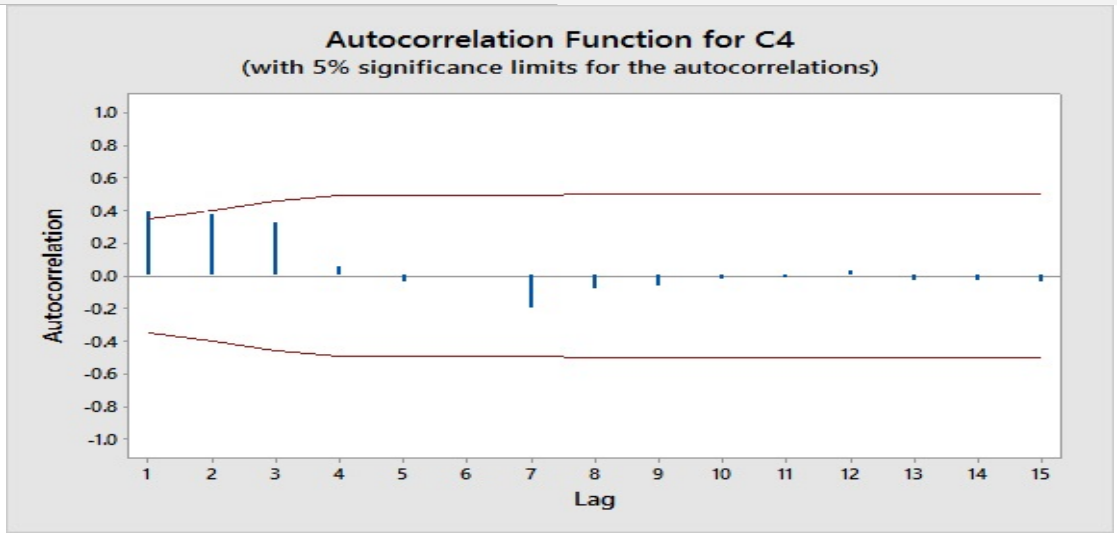


Time series after taking the first difference

Defining the proposed model

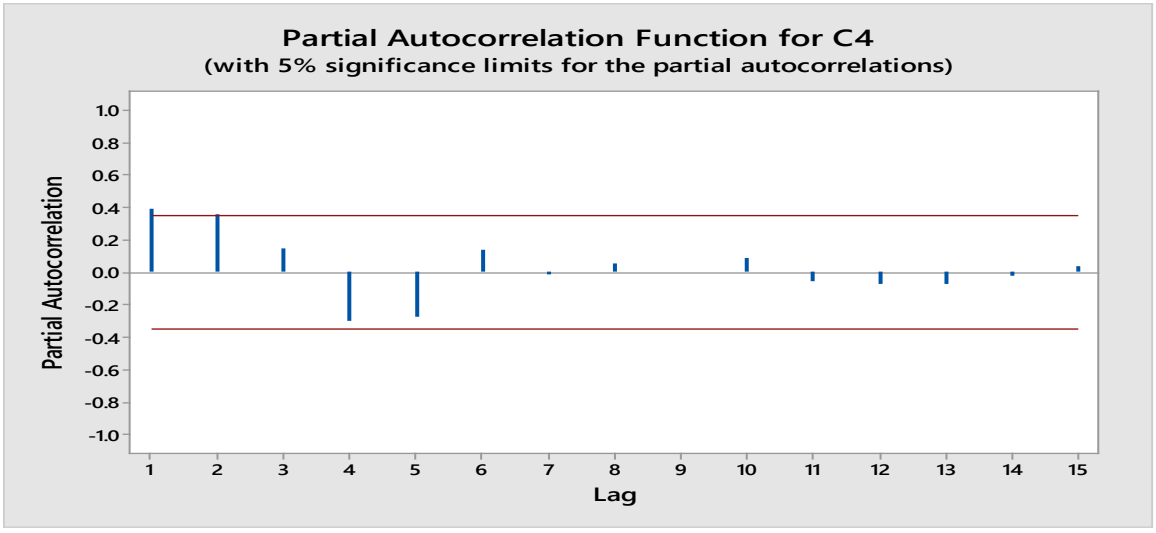
By studying the Auto-correlation and partial Auto-correlation functions of time series data after taking the first difference for them as shown in Figure No. (5 and 6) and comparing their behavior with the behavior of the Auto-correlation and partial auto-correlation functions of the set of models

Figure -5 Auto-correlation functions of C4



Lag	ACF	T	LBQ	Lag	ACF	T	LBQ	Lag	ACF	T	LBQ
1	0.389556	2.27	5.63	6	0.008162	0.03	18.91	11	-0.014683	-0.06	21.17
2	0.454181	2.32	13.52	7	-0.195942	-0.81	20.65	12	0.032233	0.13	21.23
3	0.362711	1.61	18.71	8	-0.077093	-0.31	20.93	13	-0.032451	-0.13	21.29
4	0.057110	0.24	18.85	9	-0.064705	-0.26	21.14	14	-0.027789	-0.11	21.34
5	-0.038109	-0.16	18.91	10	-0.021608	-0.09	21.16	15	-0.040235	-0.16	21.44

Figure -6 Partial Auto-correlation functions for C4



Lag	ACF	T	Lag	ACF	T	Lag	ACF	T
1	0.389556	2.27	6	0.137026	0.80	11	-0.054874	-0.32
2	0.356532	2.08	7	-0.008557	-0.05	12	-0.072912	-0.43
3	0.148528	0.87	8	0.051838	0.30	13	-0.074154	-0.43
4	-0.300354	-1.75	9	-0.001062	-0.01	14	-0.022653	-0.13
5	-0.273196	-1.59	10	0.090273	0.53	15	0.040850	0.24

Estimating the parameters of the model

The statistical program - 17 MINITAB - and the SPSS 24 to estimate the parameters of the model, the results have appeared as follows:

1- Verifying the estimated model

After completing the estimation process, we note the following:

1- Value achieves the reflexive condition $1 < \theta < 1$
2- T value is equal to 2.59 and when compared with the tabular value where (1,717) this indicates the morality of the parameter.
3- Good-matching model test (PRICE BOX) where its value is equal to (69.6) as it is greater than the tabular value at the degree of freedom of 22 and the moral level of 0.05 where the tabular value is 45.55 and therefore it can be said that it indicates the acceptance of the estimated model

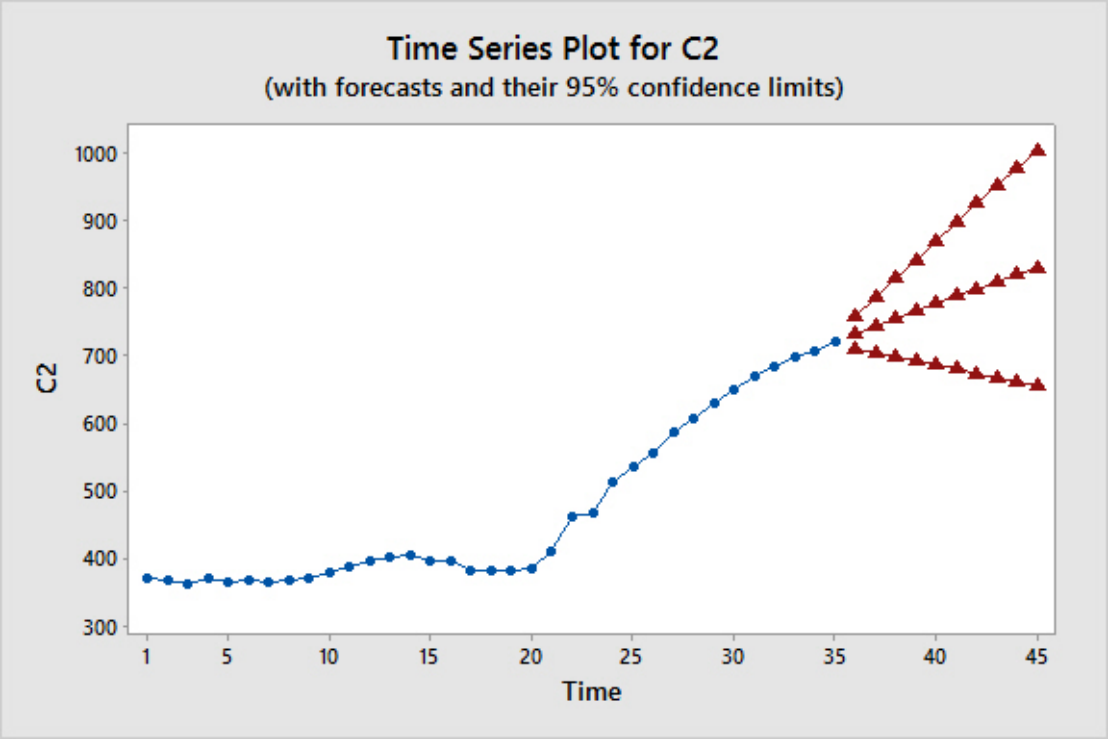
Prediction

The numbers of schools in Qadisiyah province were predicted using the estimated model for years (2026-2017) and the results were as in the table below -2- and the time series was also drawn as in Figure 7-

Table No.2 - Number of primary schools predicted for Al-Qadisiyah province for the period of 2026-2017

No	Year	Predicted No of Schools
1	2017	735
2	2018	751
3	2019	767
4	2020	783
5	2021	800
6	2022	818
7	2023	837
8	2024	857
9	2025	877
10	2026	898

Figure 7 graph series Number of primary schools predicted for Al-Qadisiyah province for the period of 2026-2017



Conclusions:

1- The number of primary schools in Qadisiyah province has increased exponentially over the years.
2- When this model is used to predict primary school numbers for the period 2017-2026, consistency has emerged with the original values of the last years.
3- The circumstances in which Iraq has experienced wars, economic blockades and political unrest have had an impact on the preparation of primary schools, and this is evident in the time series, where we note that the number of schools in some years has decreased over previous years.
4- Lack of attention to the reality of primary education, which we observe through the small annual increase in school numbers
5- Through the time series of primary schools, we note that there is no design for planning or future plans on the basis of which school numbers are predicted.

Recommendations:

1- We recommend using the model that has been reached in predicting primary school numbers, and to develop it to predict the expected numbers of primary schools each year.
2- Adoption of the predictions made in the study and thus the development of future plans for the primary education sector.
3- Use the Box Jenkins method in any of the future predictions of the time series of education indicators and at different stages.
4- The researcher recommends that the Ministry of Education and the local government in Qadisiyah province to pay attention to the planning aspect and develop scientific plans based on statistical studies and not to be satisfied with the technical aspects. Because of the educational aspect is of great importance in the progress and the development of the country.
5- To generalize this study to all educational levels to include middle school, middle school, high school and kindergartens.

REFERENCES

Almutwali, Ahmed Shaker Mohammed Tahir, 1989, (Used analysis of intervention in time series and its applications in environmental data (Master's thesis in Statistics, Salaheddine University, Faculty of Management and Economics).

Al-Sarraf, Nizar Mustafa, 1981, (Time Series Analysis, Using Statistical Technique for Economic Forecasts in Iraq), Master Thesis in Statistics, University of Baghdad, College of Administration and Economics.

Birri, Adnan Majed Abdul Rahman, 2002 (Statistical Forecasting Methods (Part 1). King Saud University.

Box- G. E. P &, Jenkins, G.M.T. (1976), "Time series Analysis Forecasting and Control, San Francisco", Holden-day, USA.

Crier, Johnathan .D (1986) "Time series Analysis", R .R .Donnelley & Sons Com .USA.

O'Donovan ,Thomas M,1983"Short Term Forecasting ,An. Introduction to the Box-Jenkins "Johan Wiley and Sona, New York

Othman Naqqar, Munther Al-Awad, 2011, (Box-Jenkins methodology for time series analysis and forecasting) Damascus University Journal for Economic and Legal Sciences - Volume 27 - No. 3. 2011

Shaarawi, Samir Mustafa, 2005, Introduction to the Modern Analysis of Time Series. Scientific Publishing Center, King Abdulaziz University.

Vandel, Walter, time series from an applied point of view, Boxes and Jenkins models, Arabization of Abd al-Mardi Azzam and Ahmed Harun. Mars Publishing House 1992.