International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 04, 2020
ISSN: 1475-7192

Using time series in the prediction of Iraqi GDP for the period (2019-2028)

[1]Sawsan Qassim Hadi

Abstract
Gross domestic product (GDP) is defined as the value of output of goods and services achieved within a year and is an important measure of the size of the economy's output. This study aims to predict the Iraqi GDP for the following period, where the annual time series data for the period from 1970 to 2018 were obtained from the National Accounts Directorate / Central Statistical Organization / Ministry of Planning. Using the GRETL, the appropriate statistical model for the Iraqi GDP is an ARIMA (0.2,1). According to the data of this study, the Iraqi GDP is predicted for the next ten years (2019-2028) based on the ARIMR model (0.2). 1) The results showed an increase in the Iraqi GDP.

Key words*: Box - Jenkins methodology, GDP, ARIMA models, GRETL, prediction.*

Introduction

Time series analysis is an important statistical topic in the analysis of many phenomena, time series is a set of observations taken at intervals of time as a result of tracking this phenomenon for a relatively long period of time and often this period of time is regular.

The time series is one of the main topics that have become very widely used in various sciences, since mathematical statistical procedures in time series analysis. Most of these analyzes give important functions of estimation in addition to other points that are very important in making decisions in many subjects and it helps to suit some Mathematical and statistical models of the problem to be studied, where it clearly identifies the equation and inferred from estimating its parameters to predict towards the future and make the right decisions, if the series is stable and there are no problems.

Prediction is generally defined as an unknown estimation, especially with respect to future events, to identify the course of the phenomenon under study in the future, and thus can be defined as a rational attempt to estimate possible future variables by knowing the behavioral variables of that phenomenon, as prediction is an important practical methods used in Planning processes and decision-making areas.

Prediction is defined as a process of estimating what will happen in the future depending on the trend of the phenomenon in the past using one of the known prediction models. The issue of predicting the future value of the phenomenon depends on the movement of the value of this phenomenon or changes in the past, and depending on these variables can be expected this future value of the phenomenon in the next certain period assuming that the changes will take a certain pattern in the future in light of what happened in the past . The main objectives of time series analysis are to obtain an accurate description of the time series, build an appropriate model for interpreting time series behavior, and use the results to predict future time series behavior.

Reference review

Time series analysis is a scientific and statistical methods that are important in the natural sciences, industry, trade, economics, health care, natural resources, population growth, etc., to represent the relationship to the data of any time series and explain its behavior, there are many studies and research, including:

[1] College of Administration and Economics, University of Baghdad, Iraq; sawsan.q@yahoo.com

The researchers (McNelis and Siddiqui 1994) conducted a study on the time series (1975-1994) to determine the impact of financial liberalization on the phenomenon of double deficit in New Zealand by analyzing the behavior of total government debt, external debt, government deficit and current account deficit. In order to determine the stability of the time series used, and to determine the degree of integration, they used the Johansen Juselius test, the multivariate regression vector, and the error correction vector model (ECM) to analyze the causal relationship between the variables.

Dagum (1980) developed a new method that differs from the old one by using Box-Jenkins models for predicting beyond and before the original time series, which was a method of quasi-parameter analysis of the time series.

The researchers (Wheelwright and Makridakis 1973) used two types of random and non-random time series and found that increasing the parameters in the equation of the random chain reduces MSE while the frequency of the process to obtain the optimal feature values is constant, but in the case of non-random series, which assumed their values (1, 2, 1, 2,…) The number of parameters increased by increasing the frequency of the process accordingly in order to obtain the lowest value of (MSE). The research also included comparing (MSE) between them and used three methods of prediction: regression method and Box-Jenkins method. The AF method is seasonal.

The researchers (Maravall and Kaiser 2000) also made observations to analyze the time series of ARIMA models and their practical application in some studies in Spain. They concluded that ARIMA models can be considered as stable ARMA models with different ranks. For example, ARIMA (1,1,1) Mind the ARMA model (2,1).

The researchers (Dobre and Alexandru 2008) in their research using the Box - Jenkins method in the analysis of time series to predict unemployment rates in January, February and March (2008) based on monthly data of unemployment rates in Romania has been found that the most appropriate model is ARIMA (2,1, 2).

The researchers (Sham et al. 2014) found the best model for predicting the number of people with hand, foot and mouth disease (HFMD) in Sawak, Malaysia using ARMA models and based on HFMID data and found that the most suitable model is ARMA (1,4).

Define time series:

There are multiple definitions of the time series concept, all of which focus on three main aspects: components, arrangement, and uses. The time series can be defined as a set of values observed for a phenomenon over a specific time period and following a specific pattern. The process of determining the mathematical model of the time series is based on the basic assumption that the pattern followed in the past will continue in the future, and is also defined as a set of observations (Xt), each of which is recorded at a specific time (t).) (Brockwell, Davis, and Calder 2002)

Stability of Time Series: Stationary Time Series (Wheelwright, Makridakis, and Hyndman 1998) (Box 2013)

1- Stationary Time Series: -

Stability of data is an important process in analyzing time series as well as in finding the appropriate mathematical model, and that the time series in the period (t, th) may sometimes be identical to the drawing of the series in another period (s, sh) and this indicates that there is a time homogeneity Which is called stability. The stability of the time series is achieved when the general trend is lacking, depending on the graph of the observations, as well as if it has an arithmetic mean and constant variations free of effects, which are said to be stable when the following conditions are met: the theoretical side

Define a time series:

There are multiple definitions of the time series concept, all of which focus on three main aspects: components, arrangement, and uses. The time series can be defined as a set of values observed for a phenomenon over a specific time period and following a specific pattern. The process of determining the mathematical model of the time series is based on the basic assumption that the pattern followed in the past will continue in the future, and is also defined as a set of observations (Xt), each of which is recorded at a specific time (t).) (Brockwell, Davis, and Calder 2002)

Stability of Time Series: Stationary Time Series (Wheelwright, Makridakis, and Hyndman 1998) (Box 2013)

1- Stationary Time Series: -

Stability of data is an important process in analyzing time series as well as in finding the appropriate mathematical model, and that the time series in the period (t, th) may sometimes be identical to the drawing of the series in another period (s, sh) and this indicates that there is a time homogeneity Which is called stability. The stability of the time series is achieved when the general trend is lacking, depending on the graph of the observations, as well as if it has an arithmetic mean and constant variations free of effects, which are said to be stable when the following conditions are met:

1. Proof of arithmetic mean $E(xt)=\mu$

2 . Prove the value of the contrast $Var(x_t) = E(x_t-\mu\ {}_X{}^2\sigma =)^2$

3. Auto Covariance, a common time difference function (Wei 1989)

$\gamma_x(k) = cov\ (x_t\ , x_{t-k}\ ) = E[(x_t-\mu)(x_{t-k}-\mu)] = E[(x_{t+k}-\mu)(x_t-\mu)] = \gamma_x(-k\ )$

The common distribution function shall be as follows:

$F(x_{t1,\ \dots,}\ x_{tn}) = (x_{t1},\ x_{t2},\ \dots,\ x_{tn}) = F(x_{t1+k},\ \dots, x_{tn+K}) = (x_{t1},\dots,\ x_{tn})$

K is a displacement variable, and K γ is the subjective covariance of the time series and is a symmetrical even function around zero.

2. Non Stationary Time Series: (Box 2013)

We can say that most models of time series are in an unstable state and can be detected through the function of self-correlation and partial self-correlation, noting that their value does not go to zero after the second or third displacement, but retains its large value for a number of displacements. The two time series are:

Stationary Mean: (Brockwell et al. 2002) (Wei 1989)

Stability is achieved in the medium if the series does not show a general trend. The instability around the mean means that the time series does not fluctuate around a constant medium. This instability can be eliminated by taking the appropriate differences, and we can get a stable time series after taking d differences.

$Y_t=\Delta^d\ X_t$

Where (Δ= 1-B) is called the backward difference operator and is calculated as follows:

$X_t = (1-B)X_t = X_t - X_{t-1}\Delta$

B: Rear difference indicator.

In general and to a number of differences d calculated in the following general formula:

$X_t = (1-B)^d\ X_t$ $d = 1,2,3,4\Delta^d$

B. Stationary in Variance: (Box 2013) (Wei 1989)

This type of stability is achieved in time series when different fluctuations do not appear in the time series form. On the time series it will shift from the instability state to the state of stability around the variance. Power conversions can generally be used to stabilize the variations.

$$X_t^\lambda = \begin{cases} X_t^\lambda & \text{if } \lambda \neq 0 \\ LnX_t & \text{if } \lambda = 0 \end{cases}$$

Where:

Xt: represents the value of the unconverted series.

λ: represents the value of the conversion parameter and ranges from (, 0 1 ±, 2/1 ±, 3/1 ±, 4/1 ±).

To find the appropriate type of conversion, we use the Rang-Mean plot method. The motives behind the conversion process are to obtain residues with a constant variation.

Box - Jenkins Models: (Box and Pierce 1970) (Box 2013) (Wei 1989)

The study of time series, the application of their models and their use in future forecasting have received considerable attention since 1927. One of the first to use time series models was the world, Yule, using self-regression models (AR). This development was followed by another progress by researchers Box and Jenkins in 1970 to include evolution. The study of moving averages models (MA) followed by the study of mixed models ((ARMA) and then the model of self-regression integrated for the moving media (ARIMA) These models are suitable to represent the time series that are already stable or were conversions, and therefore we have to identify the types of models It is as follows:-

1- Auto Regressive Model (Chatfield 2000) (Box 2013)

This model is self-regressing and of the grade (P) and can be represented by the following mathematical formula:

$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + ... + \varphi_p X_{t-p} + a_t \quad ...(1)$

Since:

Xt: represents the time series in time t.

حقيقيةi: real numbers whose absolute values are smaller than one and represent the model parameters and (i = 1, 2,…, p).

P: represents the model's rank (self-regression rank).

at: independent random variable that follows the normal distribution with an arithmetic mean of zero and a constant variation of 2at and that its common variance is zero. This variable represents a series of random errors and is called white noise.

The above equation can be written in terms of random error and using the Bounce Factor (B) as follows:

$(1 - \varphi_1 B^1 - \varphi_2 B^2 - ... - \varphi_p B^p) X_t = a_t \quad ...(2)$

$\varphi_p(B) X_t = a_t$

2- Moving Average Model: (Chatfield 2000) (Brockwell et al. 2002)

There are several ways to present the MA model, such as a simple extension in the residue series, or AR model, but using random variations that have occurred in the past, and to find out if you can get a better representation of the time series, and the mathematical form of the model of q and symbolized by MA (q) are:

$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - ... - \theta_q a_{t-q} \quad ...(3)$

As:

θi: real numbers whose absolute values are smaller than one and represent the model parameters and (i = 1, 2,…, q).

at: represents a series of random errors.

The equation above can be rewritten using the Bounce Factor (B) as follows:

$X_t = (1 - \theta_1 B^1 - \theta_2 B^2 - ... - \theta_q B^q) a_t$

3- ARMA (Auto Regressive- Moving Average Model): (Box and Pierce 1970) (Chatfield 2000)

This model combines the characteristics of the two previous models and has a rank (p, q) to obtain a model with greater flexibility in the representation of time series data. It takes the following mathematical formula:

$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + ... + \varphi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - ... - \theta_q a_{t-q}$

The equation above can be rewritten by using the back bounce (B) as follows:

$(1 - \varphi_1 B^1 - \varphi_2 B^2 - ... - \varphi_p B^p) X_t = (1 - \theta_1 B^1 - \theta_2 B^2 - ... - \theta_q B^q) a_t \quad ...(4)$

4 - Auto Regressive-Integrated Moving Average Model (ARIMA): (Box 2013) (Chatfield 2000)

This model is used in case the time series is unstable. , q Rank Moving Average, p Rank Self Regression. D is used to make the time series stable, and the mathematical form of this model is as follows:

$\varphi_p(B)(1-B)^d X_t = \theta_q(B) a_t$ …(5)

Since

$\varphi_p(B) = 1- \varphi_1 B^1 - \varphi_2 B^2 -... - \varphi_p B^p$

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

Building a Time Series Model (Box and Jenkins 1976)

The construction of the time series model is carried out through several phases, some of which are repeated until a suitable time series model is adopted to be adopted in the last phase, which is the forecasting phase.

Diagnosis

It is through this stage, which is the most important step in the construction of time series models, and the first stage of the algorithm established by researchers Box and Jenkins in 1976, and the basis on which the model of the time series. The graph is one of the main steps in the analysis of time series. Characteristics of the phenomenon studied and geographically called the graph of the phenomenon. If the values of (Xt) take the direction of increasing or decreasing with time, then the time series is unstable, but if it is distributed around its average over time so that it can be confined between two upper and lower values then the series is stable. There are several ways to detect whether the time series is stable or unstable, including:

The Autocorrelation Function: (Box and Pierce 1970)

The ACF is defined as a measure of the degree of relationship between the values of the same variable at different time intervals. This function is one of the important methods used to know the behavior of the time series and its stability through a successive K time, and its value is between (1). , -1) that is -1 k 1 and that the generated function of the self-correlation is as follows: -

…(6) $$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}(X_t-\bar{X})(X_{t+k}-\bar{X})}{\sum_{t=1}^{n}(X_t-\bar{X})^2}$$

Since:

Rate Series Views: Xt

X : represents the mean of the chain.

The statistical distribution of self-correlation coefficients is a normal distribution with a mean of zero and variance (1 \ n) where n represents the sample size.

Partial Autocorrelation Function (Box 2013) (Wei 1989)

Partial correlation function (PACF) is also an important function and is used to know the behavior of the series and diagnosis as well as the function of self-correlation mentioned above. This function is related to the self-correlations between Xt + k, Xt)) and note the rest of the values of the time series. To estimate the partial self-correlation coefficients of the sample, we follow the following mathematical formula:

…(7) $$\hat{\varphi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^{k}\hat{\varphi}_{kj}\,\hat{\rho}_{k+1-j}}{1-\sum_{j=1}^{k}\hat{\varphi}_{kj}\,\hat{\rho}_j}$$

Since:

j = 1, 2, 3,…, k

PACF is another important tool as well as an ACF that helps us to study the stability of time series and determine the degree of model (AR), (MR) or mixed model (ARIMA) to represent time series data.

Unit Root Test (Pankratz 1983)

Unit root tests are used to determine the degree of integration of the time series of the variables under study and to determine whether the variables are stable or not. These tests detect the general direction of the vehicle. We will talk about the Dicky-Fuller test as one of the best tests in standard studies.

Dickey - fuller test

This test was developed by David Dickey and William Fuller and is one of the important tests used to determine the stability of irrigation time series and be of two types:

Dicky-Fuller test (simple)

This test was developed by Dickey and Fuller in 1979 to test the hypothesis that the chain contains the root of the unit (ie, it is non-static) against the alternative hypothesis of the chain's silence (Harris 1995). Dickey Fuller is tested by the following three equations (Gujarati 2004):

1. No fixed limit and time direction. $\Delta X_t = \alpha_1 X_{t-1} + e_t$

2. Fixed limit and no time direction. $\Delta X_t = \alpha_0 + \alpha_1 X_{t-1} + e_t$

3. With fixed limit and time trend $\Delta X_t = \alpha_0 + \alpha_1 X_{t-1} + \delta_t + e_t$

Test Augmented Dickey - fuller: -

There was much criticism of the simple Dickey-Fuller test. To avoid these criticisms, in 1981, Dickie and Fuller developed this test to a new test called the Dickey-Fuller Extended Test (Dickey and Fuller 1981). Recent applied economic studies used in time series analysis show that it is more efficient than the methods used to process data that suffers from the root of ADF because it does not maintain the correlation error between residues and the ADF model is described by the following equation (Sinoha-Lopete 2006):

$$\Delta X_t = \alpha + \beta_t + \gamma X_{t-1} + \; \ldots\ldots + \delta P \, \Delta X_{t-p} + \varepsilon_t \qquad \ldots(8)$$

Since:

X_t: Time series to test.

Δ: the first difference of the time series.

α, β, γ, δ: parameters to be estimated.

P: number of time bounces.

White noise has a mean of zero and constant variability and its elements are not self-related

After estimating the regression equation parameters, the following two hypotheses are tested:

The series (Xt) is unstable, which contains the root of the entity $H_0: \gamma = 0$

Series (Xt) is stable $H_1: \gamma < 0$

The value of the extracted ADF is calculated according to the following formula:

$$\ldots(9) \quad ADF = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

The calculated ADF value is compared to the tabular t-value assumed by Mackinnon (1991). If the absolute value of ADF is greater than the absolute table value, then we reject the null hypothesis H0 and accept the alternative hypothesis H1. Then the first difference of the time series is required, then the second difference is taken, tested and so on until a stable series is obtained.

Determine the rank of the model and estimate the parameters of the diagnostic model

Model rank (Box 2013) (Wei 1989)

After identifying the types of models and the characteristics and characteristics of each type must be know the rank of the model and this is done through several statistical criteria, the correct choice of the rank of the model leads to the accuracy of the model reached. The researcher (Wei, 1989) indicates that the time series by drawing a function ACF if it suffers from decay and that the series drawing of the PACF function suffers from the presence of segments, this is proof that the model AR and the model rank is in the cutting area. Dependence on the behavior of the functions of the correlation functions, but in such cases can be used by many statistical criteria in the diagnosis of the appropriate model and determine its rank, including mention Hannan - Quinn Criterion (Hannan - Quinn Criterion) as well as Akaike (Akaike) and Bayes standard (Bayesian n Criterion) and other criteria, It is statistical criteria we recall.

Akaike Information Criterion

This criterion was proposed by Akaike in 1973.This criterion shows that minimum values are preferred when testing a model that measures competing models of unstable alternatives and is formulated as follows (Roberts and Nord 1985):

…(10) $$AIC = n\,Ln\,(\sigma_\epsilon^2) + 2h$$

As:

n: number of views (sample size).

2_ϵ ^ 2: Estimates of variance of the remainder of the model.

h: number of parameters (model rank).

Schwarz Criterion (SC):

This criterion is used to determine the number of hysteresis intervals h when regression is performed, after which the hysteresis period that achieves the lowest SC value is tested and Schwarz assumes the following function (Schwarz 1978):

…(11) $$SC = \ln(\sigma_\epsilon^2) + h\,l\,n(n)$$

3. Hanna - Quinn Criterion (H-Q): Al-Nasser and Jumma 2013

The researchers Hanna & Quinn (1979) proposed a new criterion for determining the rank of the studied model called Hanna - Quinn Criterion, which is abbreviated (H-Q) and formulated as follows:

…(12) $$H - Q = Ln\,(\sigma_\epsilon^2) + 2h\,C\,Ln\,(Ln\,n)/n$$

Since:

C: represents a constant amount such that C> 2.

Estimation of model parameters (Wei 1989)

The stages of constructing the Box - Jenkins model are sequenced. There are several methods that can be used to estimate model parameters of the time series, including Moment, Exact Maximum Likelihood, Conditional Maximum Likelihood, Approximate Maximum Likelihood, and Least. And many other common methods. The method used to estimate the parameters of the model in the statistical program used is one that is the greatest possible function.

Exact Maximum Likelihood Function

The Likelihood function of the Xt time series depends on maximizing the function by making the sum of the least error squares possible and the formula according to the following equation:

…(13)

$$L(\emptyset, \theta, \sigma_a^2 [W) = (2\pi\sigma_a^2)^{\frac{-n}{2}} \left| M_n^{(p,q)} \right|^{\frac{1}{2}} \exp\left(\frac{-S(\emptyset,\theta}{2\sigma_a^2}\right)$$

When the sample is large and after omitting the amount | $M_n^{(p,q)}$ | In the above formula and then take the logarithm of the function possible, the equation becomes as follows:

…(14) $$Ln\,L(\emptyset, \theta, \sigma_a^2 [W) = (\frac{-n}{2} Ln \frac{S(\phi,\theta)}{2\sigma_a^2})$$

For large samples, the information matrix is equal to (∅,)), which is available after the matrix of derivatives of the order n x (p + q) is calculated.

$$a = \theta^{-1}(B)\phi(B)$$

Approximate Maximum Like hood Estimation is obtained from the above equation after taking its partial derivatives.

$$\left.\begin{aligned}\frac{\partial a_t}{\partial \theta_i} &= \theta^{-1}(B)B^i a_t\\ \frac{\partial a_t}{\partial \emptyset_j} &= \emptyset^{-1}(B)B^j a_t\end{aligned}\right] \qquad i = 1, 2, \ldots \qquad j = 1, 2, \ldots$$

The common variations of the estimates are calculated by using the inverse of the information matrix $[(\emptyset, \theta)]^{-1}$.

Testing the suitability of the personalized model This is the next stage after determining the rank of the model and estimating the parameters of the diagnostic model. During this stage, the time series is recalculated using the previously reached model and then the residue is calculated and the residue is tested (equal to the difference between the actual and estimated values). Or reached this stage works under the following hypothesis: $H_0: \rho_1 = \rho_2 = \cdots = \rho_k = \rho_m = 0 \qquad k = 1, 2, \ldots, m$

$H_1: \rho_k \neq 0$ for some values of k

There are many tests used to check the suitability of the diagnostic model and here we will mention some of these tests.

1- Box and Pierce: (Box and Pierce 1970)

This test was developed in 1970 by Box and Pierce researchers. $Q_{B\&P} = n \sum_{k=1}^{m} \hat{\rho}_k^2 \sim \chi^2_{(m-j),\alpha}$ …(15)

If the value extracted from the test is smaller than the tabular value at the specified significance level and the degree of freedom, this indicates the suitability of the diagnostic model of time series data and accept the null hypothesis. Similar to this test, Ljung and Box (Edward 2011) (Ramon 2008) in 1978 created a test for random string error testing by calculating the residual correlation coefficients for a range of displacements. The test formula is as follows: $Q_m = n(n+2)\sum_{k=1}^{m} \frac{\widehat{\rho_k^2}}{n-k} \sim \chi^2_{(m-p)}$ …(16)

Since:

n: Represents the number of views of the time series.

m: represents the number of offsets of self-correlation.

P: number of estimated parameters in the model.

2 k_k ^ 2: The squares of the self-correlation coefficients for the remainder of the model series represent rt = Xt - M and then for the series rt2 the value of the test statistic Q (m) is compared with the tabular value of the kai square test with a degree of freedom (mp) ie _ ((mp)) ^ 2 At the level of significance if Q _ ((m)) < _ ((mp)) ^ 2 or P-Value ≥0.05 means not to reject the null hypothesis H0, that is, the errors are random. This means that the errors are not related to each other and this means that the diagnostic model is appropriate.

Test Confidence Interval: (Box 2013) (Wei 1989)

In this test the sequence of residues {at} of the pre-diagnosed model is taken. This is assumed to be normally distributed and eat at + k = 0 and the chain test after estimating its self-correlations (at) r k. Zero and variance of 1 / n, that is (r ^ k (a) ~ N (0, 1 / n).

The estimation of the correlation coefficient for the remainder is equal.

…(17) $\hat{r_k}(a) = \frac{\sum_{t=1}^{n-k} at\ at{+}1}{\sum_{t=1}^{n} a^2 t}$

If the correlation coefficients for the remainder in the diagnostic model fall within the confidence limits and 95% confidence level in.

$pr\{\hat{r}_k (a) | < 1.96 \frac{1}{\sqrt{n}} \} = 1-\alpha$

This indicates that the diagnostic model is appropriate for the time series data under study.

Forecasting: (Makridakis, Wheelwright, and Hyndman 1998)

The final stage of building the Box - Jenkins model is the prediction stage. After determining the type and order of the model, then estimating the model parameters and examining the suitability of the diagnostic model through several tests. We can predict the future values of the time series and what the data will be in the future by using the diagnostic model, according to the following formula (Montgomery and Contreras 1977):

$$\hat{X}_{t+L} = E[X_{t+L} | X_t, Z_{t-1}, Z_{t-2}, \dots] \quad \text{for } L \geq 1 \quad (18)$$

If the model AR (1), the best prediction of the number of steps (L) is:

$$\hat{X}_{t+L} = \varphi^L X_{t-1+L} \quad L \geq 1$$

If the model AR (2), the best prediction of the number of steps (L) is:

$$\hat{X}_{t+L} = \varphi_1^L X_{t-1+L} + \varphi_2^L X_{t-2+L} \quad L \geq 2$$

In the case of MA (q), the best prediction of the number of steps (L) is: $\hat{X}_{t+L} = a_{t+L} - \theta_1^L a_{t-1+L} - \theta_2^L a_{t-2+L} - \dots - -\theta_q^L a_{t-q+L}$

In the case of the mixed model ARMA (p, q), the best prediction of the number of steps (L) is:

$$\hat{X}_{t+L} = \varphi_1^L X_{t-1+L} + \varphi_2^L X_{t-2+L} + \dots + \varphi_p^L X_{t-P+L} + a_{t+L} - \theta_1^L a_{t-1+L} - \theta_2^L a_{t-2+L} - \dots - -\theta_q^L a_{t-q+L}$$

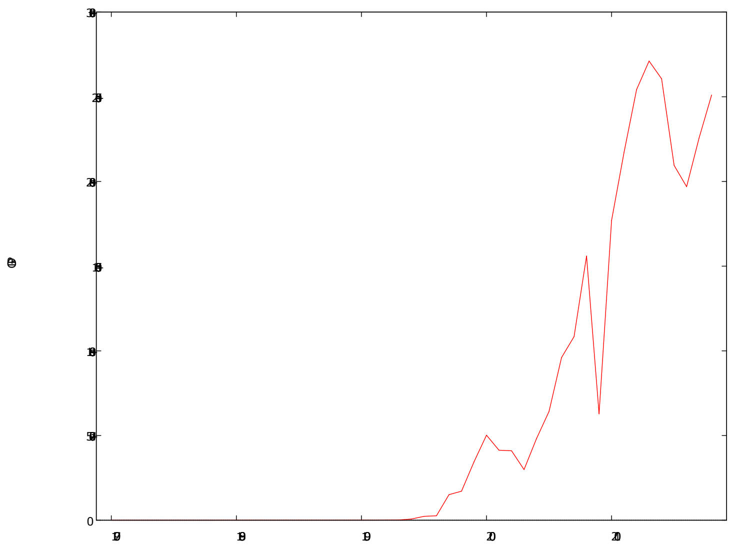
Application side:

This includes the practical aspect where the data of the study are described for the Iraqi GDP for the period (1970-2018) m obtained from the Ministry of Planning (Central Statistical Organization / Directorate of National Accounts) where the application of modern time series analysis method of applying models (Box - Jenkins), based on the Stability Stability Test, Model Diagnostics, Estimation and Proficiency Test based on GRETL, and then the best GDP forecast model for the next 10 years from 2019 to 2028.

Data analysis

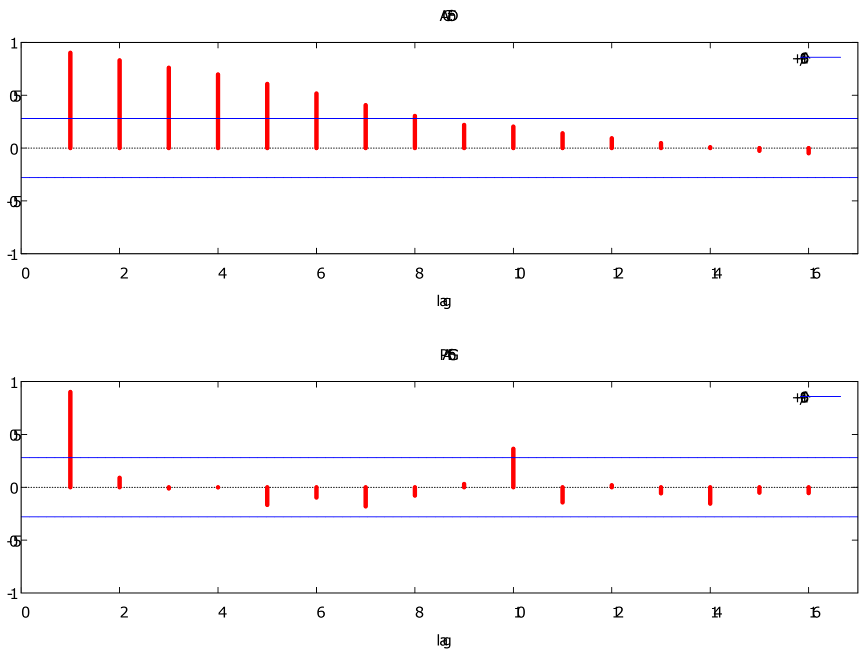
After data collection, which is the first phase of the Box - Jenkins methodology, we draw the data series represented by GDP using GRETL to identify the behavior of the series and its initial characteristics.

Figure (1) Draws the original GDP time series from 1970 to 2018.



From Figure (1) above, it is clear that the time series is unstable due to the large variation between GDP values by years. For more accuracy, we draw both the ACF function and the PACF partial correlation function, respectively:

Figure 2 Draws the ACF self-correlation function and the PACF partial correlation function for the time series



From Figure (2), we notice that the ACF function is outside the confines of confidence. This indicates the lack of stability of the series and the same form of PACF that the first displacement is outside the confines of the confidence of the partial self-correlation coefficients.

To get rid of the instability of the series, the natural logarithm and the second difference were taken to reach the stability of the time series. The graph of the resulting series becomes as shown in Figure (3).

Figure 3 Draw the time series after taking the natural logarithm and the second difference

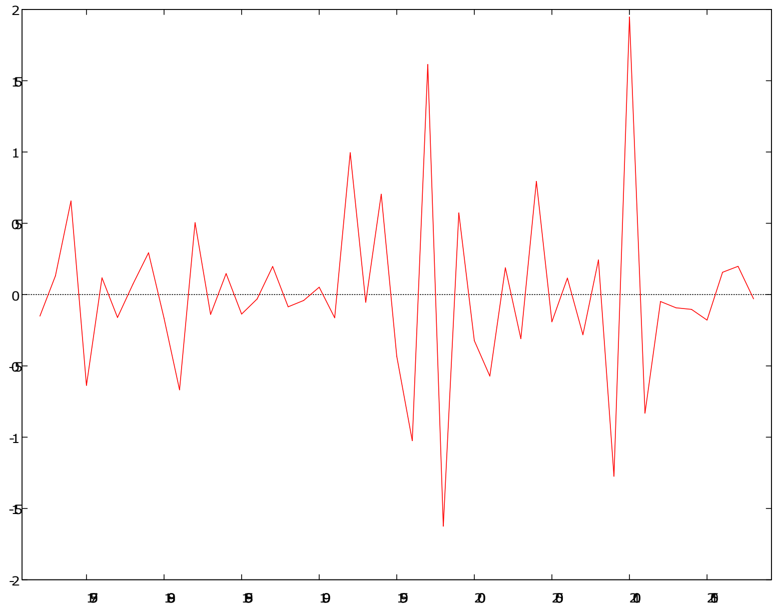
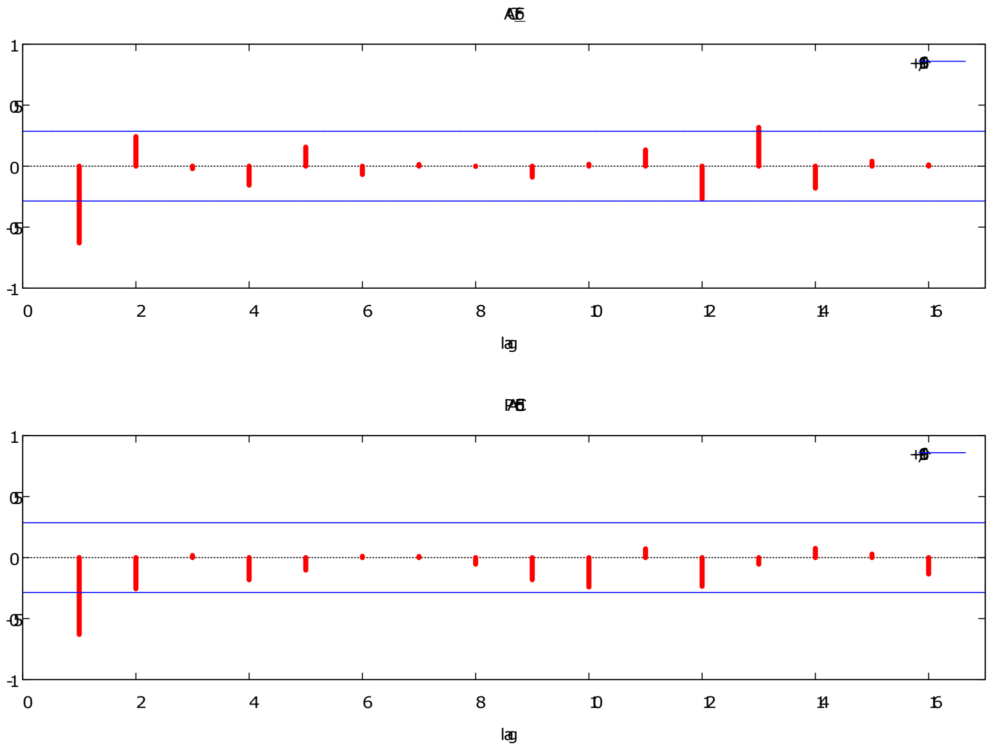


Figure 4 Plot the functions of self-correlation and partial self-correlation after taking the natural logarithm and the second difference



In Figure (3) and (4) above, we note that the series has achieved stability and morale and to make sure that we are .finally doing the Dickey Fuller Augmented Test results as shown in Table (1) below

Table 1 shows the results of the expanded Dicky Fuller test

	Without constant	With Constant	With constant and trend
Estimated value	-2.34875	-2.35393	-2.3625
Test statistic	-4.50428	-4.45355	-4.38148

p- value	7.439e-006	0.0002333	0.002284

It is noticed from Table (1) that the values of (p- value) for the three models are less than 0.05, ie that the time series is stable and thus we have achieved stability.

Testing and estimating the model:

At this stage, an appropriate model is identified for the time series under study.

Table (2) represents the proposed models (ARIMA (p, d, q)) of the series and differentiation criteria.

	ARIMA(2,2,2)	ARIMA(0,2,2)	ARIMA(2,2,1)	ARIMA(1,2,2)	ARIMA(0,2,1)	ARIMA(0,2,0)
AIC	72.54733	70.16260	71.16580	71.23100	69.36460	91.69158
H-Q	76.72466	72.94749	74.64692	74.71212	71.45327	93.08403
SC	83.64821	77.56319	80.41654	80.48174	74.91504	95.39188

It is clear from Table (2) that the best model can be selected is ARIMA (0,2,1). The parameters of the model were estimated according to the maximum approximation method and the results are as follows:

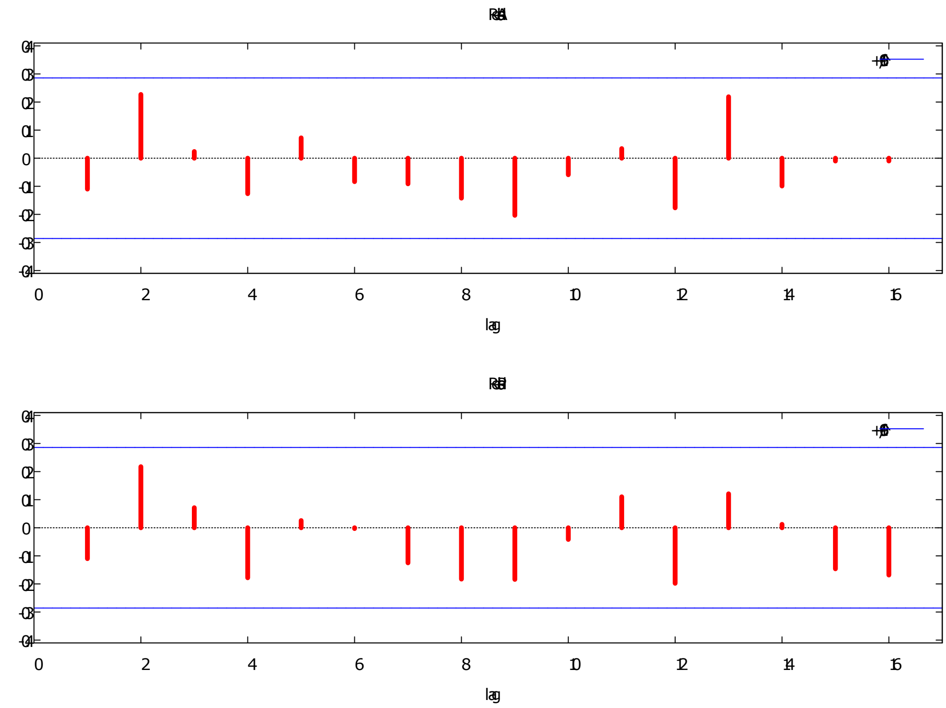
Table (3) represents the estimated value and significance of the parameter.

P-Value	Values Z	Std. error	value	
0.0001	−7.237	0.101775	−0.736507	θ_1

Diagnosis

This stage is one of the most important phases of the analysis where it is to verify the suitability and efficiency of the model by analyzing the residues through the drawing of the functions of self-correlation and partial self-correlation of the remainder of the model selected as shown below Figure (5):

Figure 5: Drawing of ACF and PACF for the residues.



Note from Figure 5 that the values of the correlation coefficients for all residues are within confidence limits, which means that the sequence of residues is random and the model used is good and appropriate.

Through the application of the test (Ljung-Box), which depends on the residues to check the suitability of the model the results emerged as follows:

Ljung-Box Q' = 15.2628

with p-value = P(Chi-square(15) > 15.2628) = 0.4327

Note that (Ljung-Box Q '= 15.2628) is less than the tabular value Chi-square and the degree of freedom (15) and a significant level (0.05) is equal to (25.00), which means that the errors are not related to each other.

Forecasting

After determining the model ARIAM (0,2,1), which was chosen statistically acceptable and after the necessary tests to ensure its validity. We will predict Iraqi GDP for the period 2019-2018 as shown in Figure (6) below, and Table (4) shows the predictive values and the maximum and minimum values of forecasting based on the gretl program.

Figure 6 shows the time series graph of the Iraqi GDP and the predicted values.

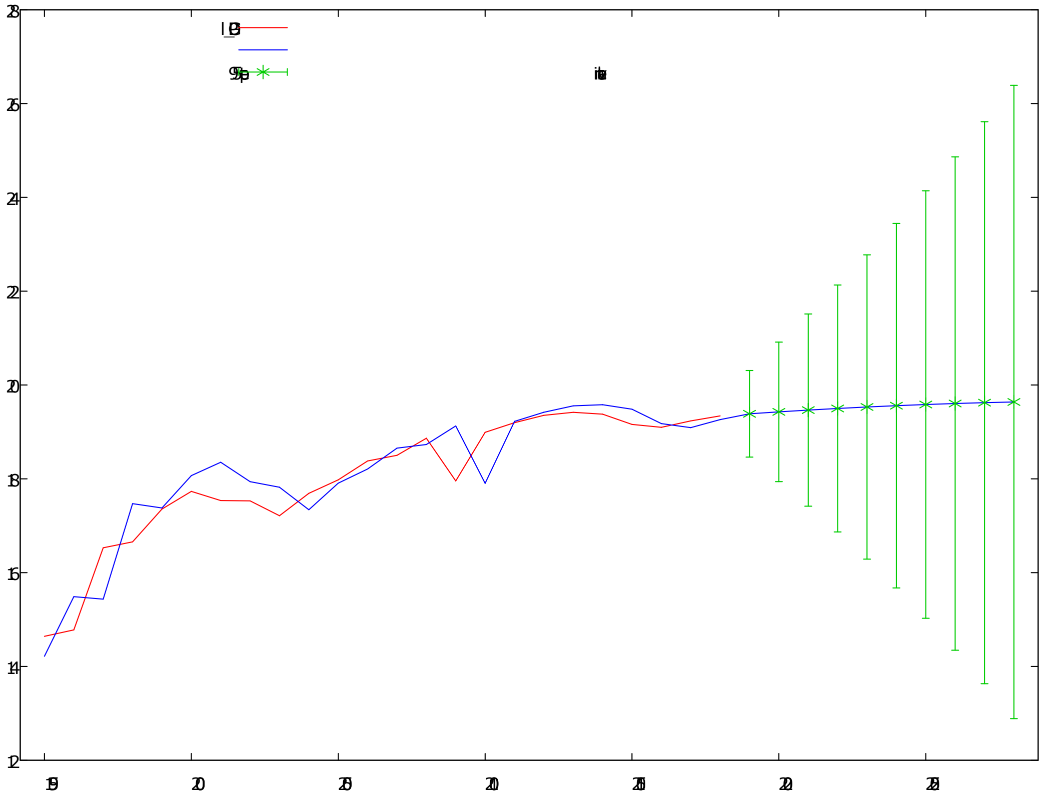


Table (4) shows predicted values and confidence limits.

Year	Std. error	Predicted values	minimum	maximum
2019	0.470874	262562577	104333181	660759177
2020	0.758738	273660816	61854784	1210742930
2021	1.04531	284288464	36641792	2205455719
2022	1.34293	294326425	21170092	4092001289
2023	1.65495	303684533	11850730	7782940033
2024	1.98234	312307863	6415324	15203629208
2025	2.32522	320085914	3357812	30515482051
2026	2.68339	326976872	1699934	62899245163
2027	3.05651	332882456	832926	133051219493
2028	3.44419	337778192	395300	288626701559

From the results of table (4) we note that there is an increase in the values of GDP for the coming years.

Conclusions

The time series of the Iraqi GDP for the period (1970-2018) when the analysis was found to be unstable in the average and variance, so the conversions were made on the time series by taking the natural logarithm and the first difference and then the second difference of data, has been achieved stability through the use of correlation functions

Self-correlation and partial self-correlation, as well as an extensive Dicky Fuller test were conducted and the results confirmed the stability of the time series, and then obtained the appropriate model to represent the time series that has the lowest value of the three criteria (AIC - SC - HQ) and ARIMA (0.2,1) For the purpose of use in a Prediction has been confirmed the diagnosis of the specimen chosen through the use of statistical tests and the health of the two test (Ljung - Box) and test residuum draw self-correlation functions and self-correlation partial residuals suggesting an appropriate form prescribed. Finally, using the selected model ARIMA (0.2,1), the Iraqi GDP was predicted for the period from 2019 to 2028, and according to the values obtained we note that there is an increase in the values of Iraqi GDP for the coming years.

References

Al-Nasser, A. H. and A. A. Jumma. 2013. "Introduction to Applied Time Series Analysis." *Al-Jazeera Printing and Publishing*.

Box, George. 2013. "Box and Jenkins: Time Series Analysis, Forecasting and Control." Pp. 161–215 in *A Very British Affair*. Springer.

Box, George E. P. and Gwilym M. Jenkins. 1976. "Time Series Analysis: Forecasting and Control San Francisco." *Calif: Holden-Day*.

Box, George E. P. and David A. Pierce. 1970. "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models." *Journal of the American Statistical Association* 65(332):1509–26.

Brockwell, Peter J., Richard A. Davis, and Matthew V Calder. 2002. *Introduction to Time Series and Forecasting*. Vol. 2. Springer.

Chatfield, Chris. 2000. *Time-Series Forecasting*. Chapman and Hall/CRC.

Dagum, Estela Bee. 1980. *The X-II-ARIMA Seasonal Adjustment Method*. Statistics Canada, Seasonal Adjustment and Time Series Staff.

Dickey, David A. and Wayne A. Fuller. 1981. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica: Journal of the Econometric Society* 1057–72.

Dobre, Ion and Adriana AnaMaria Alexandru. 2008. "Modelling Unemployment Rate Using Box-Jenkins Procedure."

Edward, Ngailo. 2011. "Modelling and Forecasting Using Time Series Garch Models: An Application of Tanzania Inflation Rate Data." *University Of Dar Es Salaam*.

Gujarati, D. N. 2004. "Basic Econometrics.(4 Th Edtn) The McGraw− Hill Companies."

Harris, Richard I. D. 1995. "Using Cointegration Analysis in Econometric Modelling."

Makridakis, Spyros, Steven C. Wheelwright, and Rob J. Hyndman. 1998. "Forecasting Methods and Applications, John Wiley& Sons." *Inc, New York*.

Maravall, Agustín and Regina Kaiser. 2000. *Notes on Time Serie Analysis, ARIMA Models and Signal Extraction*. Universidad Carlos III de Madrid. Departamento de Estadística.

McNelis, Paul D. and Anjum Siddiqui. 1994. "Debt and Deficit Dynamics in New Zealand: Did Financial Liberalization Matter?" *International Economic Journal* 8(3):71–87.

Montgomery, Douglas C. and L. E. Contreras. 1977. "A Note on Forecasting with Adaptive Filtering." *Journal of the Operational Research Society* 28(1):87–91.

Ngo, Theresa Hoang Diem and Warner Bros. 2013. "The Box-Jenkins Methodology for Time Series Models." Pp. 1–11 in *Proceedings of the SAS Global Forum 2013 conference*. Vol. 6.

Pankratz, A. 1983. "Forecasting with Univariate Box-Jenkins Models: Concepts and Cases. John Wily & Sons." *Inc. USA*.

Ramon, Haydee L. 2008. *Forecasting the Volatility of Philippine Inflation Using GARCH Models*. BSP Working

Paper Series, Series.

Roberts, David L. and Stephen Nord. 1985. "Causality Tests and Functional Form Sensitivity." *Applied Economics* 17(1):135–41.

Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." *The Annals of Statistics* 6(2):461–64.

Sham, Noraishah Mohammad, Isthrinayagy Krishnarajah, Mahendran Shitan, and Munn-Sann Lye. 2014. "Time Series Modelon Hand, Foot and Mouth Disease in Sarawak, Malaysia." *Asian Pacific Journal of Tropical Disease* 4(6):469–72.

Sinoha-Lopete, Ramona. 2006. "Export-Led Growth in Southern Africa."

Wei, William. 1989. *Time Series Analysis: Univariate and Multivariate Methods*. Vol. 33.

Wheelwright, Steven C. and Spyros Makridakis. 1973. "An Examination of the Use of Adaptive Filtering in Forecasting." *Journal of the Operational Research Society* 24(1):55–64.

Wheelwright, Steven, Spyros Makridakis, and Rob J. Hyndman. 1998. *Forecasting: Methods and Applications*. John Wiley & Sons.