

The Prediction of the Rate of the Dropout of the Primary Schools Students by Using the Genetic Algorithm

Sabah Manfi Redha

Department of Statistics, College of Administration and Economics, Baghdad University, Iraq
Baghdad, Iraq, Tel. +964 1 778 7086
drsabah@coadec.uobaghdad.edu.iq

Abstract

In this research, the ARIMA model of the time series has been applied for the prediction of the rate of the dropout of the primary schools for the male and female students during the period (2007-2015) by estimating the autocorrelation and partial coefficients. It shows that the time series is unstable. After estimating autocorrelation and partial coefficients, it manifests that the appropriate ARIMA models (1,1,0) for males and ARIMA (1,1,0) for females and ARIMA (1,1,0) for males and females together. Also, it has been assured that these models are good and give accurate predictions and close to the reality through statistical calculation Q. It turns out that the selected models are appropriate and good. Finally, the prediction of the rate of the dropout of the primary schools for the male and female students during the period (2007-2015) and required results were obtained by using the Genetic Algorithm through the applications (MATLAB R2013a Version: 8.1) and (GRETl 2016).

Keywords: prediction, ARIMA model, genetic algorithm

1. Introduction

Nations progressed through education and learning. When educated generations learn since a young age, a civilized and educated society arises. Hence, this Society will progress in terms of industry, agriculture, medicine and in all fields, and that the first stage of a person's life is the basis of education. So, the primary school is one of the most important stages in a student's life, and if properly prepared, a student will be interested in studying and will achieve his/her education which will benefit his/her own society. However, if it is improperly performed, the education will be a burden on the students. This will lead the student to think to drop out of school. In addition to other reasons, that may affect the student. So, from this point, this subject is selected, which is the prediction of the rate of the students' dropout of the primary schools. The prediction model (Box-Jenkins) will be examined according to gender and to predict this model, a Genetic Algorithm will be used.

2. Theoretical aspect

2.1. First: Time series

Is a set views on specific phenomenon is taken for a period of time and can be expressed in the form as follows (1):

$$[X_t]_{t=-\infty}^{\infty} = \dots, X_{t-1}, X_t, X_{t+1}, \dots \quad (1)$$

Also, the future values of the time series can be predicted by using the fixed form and which does not contain a random variable e_t , and it is called determinism (Deterministic). While most of the times series are characterized in incidental (Stochastic). It means that the future values are subjected to probabilistic distribution by using a random model that contains a random variable e_t . In some types of time series, it can be watched the changing values in every moment of the time periods and these types are called Continuous time series, and that most of the time, series include the views of the variable that can be viewed at equal points in time where it is called Discrete time series. This type can be obtained on either by recording phenomenon values in fixed times or by

collecting phenomenon values in a fixed time (2). The components of the time series are General Trend, Seasonal Variations, Cyclical Variations, Random Variations (1,4). The time series is considered static if the static is constant that collects the data around and which is free from the influence of the general trend and seasonal variations. The time series has a constant statics and variable and contrasting is shown as follows (2,4):

$$\mu = E(X_t) \quad (2)$$

$$\sigma_X^2 = \text{var}(X_t) - E(X_t - \mu)^2 \quad (3)$$

$$\gamma_k = \text{cov}(X_t, X_{t+k}) = E(X_t - \mu)(X_{t+k} - \mu) \quad , k = 0, 1, 2, \quad (4)$$

If $X_1, X_2, \dots, X_n$ is values observation the time series $\{X_t\}$ and if $C_k, \hat{\sigma}_X^2, \bar{X}$ estimates are for σ_X^2, Y_k , respectively the

$$\bar{X} = \frac{1}{N} \sum_{t=1}^N X_t \quad (5)$$

$$\hat{\sigma}_X^2 = \frac{1}{N-1} \sum_{t=1}^N (X_t - \bar{X})^2 \quad (6)$$

$$C_k = \frac{1}{N-1} \sum_{t=1}^N (X_t - \mu)(X_{t+k} - \mu) \quad (7)$$

The static time-series can be distinguished from non-static through self-correlation coefficients values where approaching its value to zero after the second and the third period with regard to the static time-series, while non-static time-series have significant differences close to zero after the seventh or eighth period, and there is a kind of time-series called seasonal time series if the series is repeating itself in each fixed period of time that is (8):

$$X_t = X_t + S \dots \quad (8)$$

where S represents the length of the season and can be known and distinguished through self-correlation coefficients, which is positive and so large that can differ significantly from zero at the time periods values.

2.2. Second: Autocorrelation Function (ACF):

The autocorrelation function measures the correlation strength between the time series values of the phenomenon $\{X_t\}$ within different time periods and it is used to determine the degree and the appropriate form by random errors residual e_t for the purpose of knowing the extent of stability of time series or the series can be unstable. The mathematical formula is illustrated as follows (9):

$$\rho_k = \frac{\text{cov}(X_t, X_{t+k})}{\sqrt{\text{var}(X_t) \text{var}(X_{t+k})}} = \frac{Y_k}{Y_0} \quad , \quad k = 0, 1, 2, \dots \quad (9)$$

Since the variation of static time series is fixed and equal for each different time periods, it is estimated as follows:

$$\sigma_{X_t} = \sigma_{X_{t+k}} \quad (10)$$

2.3. Third: Partial Autocorrelation Function (PACF):

The partial autocorrelation coefficient function contributes also in determining the appropriate model, and reveals the link between time series X_t with fixing of other variables and a mathematical formula are as follows (2):

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{kk}\rho_{j-k}, \quad j = 1, 2, \dots, k \quad (11)$$

Since ϕ_{kj} represents the gradient regression.

2.4. Fourth: Formulation of model ARIMA

2.4.1. Autoregressive Model (AR):

This model is considered of great benefit for the analysis of time series and it is the most frequently used and this model can be represented by the format of the degree of P this model will be as follows ⁽¹⁾:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + e_t \quad (12)$$

Since:

e_t represents the random error, and ϕ represents the model parameters, and P represents autoregressive rank.

And this model can be written X using the background displacement factor B is described as follows:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) X_t = e_t$$

Since:

$$\begin{aligned} \phi_p(B) X_t &= e_t \\ (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) X_t &= \phi_p(B) X_t \end{aligned} \quad (13)$$

$$X_t = (\phi_p(B))^{-1} e_t \quad (14)$$

Therefore, the autoregressive process called class P is irreversible.

2.4.2. Moving Average Model (MA):

This model represents the current correlation views of the time series with the time series of the same error for previous periods and this model can be performed as mathematical degree q are as follows ⁽⁴⁾

$$X_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (15)$$

As the: θ represents the model parameters, q represents the rank of moving circles.

By using posterior displacement factor B we get on the following formula:

$$X_t = \theta_q(B) e_t \quad (16)$$

The formula is as follows:

$$\theta_q B = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \quad (17)$$

2.4.3. Mixed Autoregressive Moving Average Model (ARMA (p,q))

This model is considered as a cross between a self-regression model and moving circles, and represents the current correlation time series values with the previous values of the time series of the same series and with the same series error for previous periods of time and a model formula will be as follows ⁽²⁾

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad (18)$$

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} + e_t \quad (19)$$

By using posterior displacement factor B the formula would be as follows:

$$\phi(B)X_t = \theta(B)e_t \quad (20)$$

If that:

$$\begin{aligned} \phi(B) &= (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_q B^q) \\ \theta(B) &= (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \end{aligned}$$

And that the model parameters are $\phi_1, \phi_2, \dots, \phi_p, \mu, \sigma^2$, and $\theta_1, \theta_2, \dots, \theta_q$. Which is estimated from observations of the time series, and e_t is considered independent random errors and a series of normal distribution an arithmetic mean of zero and constant variance σ_e^2 .

2.4.4. Autoregressive Integrated Moving Average (ARIMA (p,d,q):

Most of the time, series are not stable and are recognized by functions autocorrelation where the value does not reach zero after the second and third displacement remain significant for a number of displacements, so indiscriminate forms are used to convert the series to a stable series and then using backward differences operator and its form is illustrated as follows:

$$\nabla^d X_t = (1 - B)^d X_t \quad (21)$$

If we substitute a X_t with a $\nabla^d X_t$ in equation (20) we get a new model that deals with a particular type of unstable time series and it is called unstable mixed model and a mathematical formula for this model is as follows:

$$\phi(B)\nabla^d X_t = \phi(B)(1 - B)^d X_t = \theta(B)e_t \quad (30)$$

2.5. Fifth: Building time series model (Box-Jenkins):

The style of Box-Jenkins to build a time series of works in accordance with the self-regression model and moving circles for a range of views of certain phenomenon, and going through several important stages for the construction of the model are as follows^(1,2):

2.5.1. Model Identification

In this stage, the stability of the time series must be checked, and the required p,d,q is selected to select the ARIMA model. The functions of autocorrelation ACF and partial autocorrelation PACF must be designed to know the time series. However, if autocorrelation is located within the limits of the confidence interval 95%, the autocorrelation coefficient P_k is close to zero and a series is considered static of zero rank and we are working on the values of the original views without transferring them. If the form of autocorrelation function does not fall within the limits of the confidence interval of 95%, the autocorrelation coefficient P_k will not come close to zero and the series is considered static and must take its differences in order to convert the series into a static one; when taking the first difference, it will become the d = 1 value. However, if the second difference becomes the d = 2 value of ARIMA (p, d, q) model. So, this can be summarized the diagnostic process through Table 1, as follows (1,2):

Table 1. Diagnosis of the set of ARIMA models

Model	ACF	PACF
AR(1)	Cede engineering starting from P_1	Zero after P_{kk1}
AR(2)	Cede engineering starting from P_2	Zero after P_{kk2}
AR(p)	Cede engineering starting from P_p	Zero after P_{kkp}
MA(1)	Zero after P_1	Ceding after P_{kk1}
MA(2)	Zero after P_2	Ceding after P_{kk2}
MA(q)	Zero after P_q	Ceding after P_{kkq}
ARMA(1,1)	Cede engineering starting from P_1	Ceding after P_{kk1}
ARIMA(p,d,q)	Cede engineering starting from P_p	Zero after P_{kkp}

Since: P is the coefficient of autocorrelation function, P_{kk} represents coefficient partial autocorrelation function.

2.5.2. Estimation and Testing

At this stage, to build the ARIMA model, estimating appropriate model is based on personalized parameters p , d , q , and using of appropriate equations for estimating; after the diagnosis and assessment stages are accomplished, it must be tested the efficient and estimated model. It is confirmed that Box-Jenkins is to check the residual e_t to find out the stability and random residuals. The confidence limit must be tested which uses the value of the autocorrelation coefficients of the residuals and should be located between the extent set at probability 95% according to the illustrated formula :

$$P_r \left\{ -1.96 \left[\frac{1}{\sqrt{n}} \right] \leq P_k(a_t) \leq +1.96 \left[\frac{1}{\sqrt{n}} \right] \right\} = 0.95 \quad (31)$$

Since a_t is a white noise factor (when the residual is distributed according to a normal distribution with zero average and constant variation), and when the above formula is achieved, it refers to random transactions of residuals autocorrelation. The diagnosed form is considered good, convenient, highly qualified and predictable. For the purpose of examining residual series, the Box-Jenkins test is used to make sure that random residual have been used to estimate parameter model and formula will be as follows ^(1,4)

$$Q = T(T+2) \sum_{k=1}^K \frac{1}{T+k} P_k^2 \quad (32)$$

2.5.3. Prediction stage

In this stage, the future values are predicted to observe the time series and it will $T+L$ if we assume that $X_t(L)$ represents the value at time (T) . The prediction can be got by taking the conditional expectation at the original time t of the model after writing period $T+L$ so $E_t(X_{t+1} / X_{t-1}, \dots)$. Therefore, we get the future prediction of the time series. The prediction formula model can be written as follows:

where writing the equation (19) will be as follows:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_{p+d} X_{t-p-d} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} + e_t \quad (33)$$

This equation can be written at the period $T+L$ and by taking the conditional expectation at period (T) , we will get equation prediction at the original time are as follows:

$$E(X_{t+1}) = X_t(L) = \phi_1 E(X_{t+L-1}) + \dots + \phi_{p+d} E(X_{t+L-p-d}) - \theta_1 E(e_{t-1}) - \dots - \theta_q E(e_{t+L-q}) + E(e_{t+L}) \quad (34)$$

2.6. Sixth: Genetic Algorithm (GA)

As one of the Optimization and search methods, it can be categorized as one of the developmental Algorithms Methods, as it's called (Genetic and Evolutionary Algorithm) (GEA). It depends on the reproduction of (Darwin's perspective) and that is to be, the living creatures try to adequate and adapt with their surroundings, without any alteration to its conditions through generations, and if they are not capable of adjusting or evolving, they will eventually go extinct. The high qualification creatures will prevail, while the low qualification creatures will fade and die. As for the genetic mutation, it's very rare to occur; it's one of the factors that assist to develop the genetic characteristics which are impacted by the genes, the Jainism Algorithm quotes the expressions and terms of the genetics, such as generations, transit, mutations etc.

And in this Algorithm, we get to use solutions inside a random society chosen according to the available data, every solution will be shaped as codes, these codes formed as a chain of chromosomes and its size is either bilateral or eightfold etc. and every Gene inside the chromosome is symbolized as Bit .

Numbers should be allocated according to the size of the function as a primary solution for this Algorithm, then we start by making fixed steps (fixed quantifications); a new solution is formed by applying the genetic commands and there are other solutions within this procedure by choosing random samples until we get The trade-off function ^(3,5)

2.7. Seventh: Genetic Algorithm Steps

There are several steps to apply the Genetic Algorithm ^(3,5)

1. *Initialization: as a start, generating many individual solutions randomly for the primary form of chromosome, and its size depends on the nature of the trouble, but normally there are hundreds or thousands of possible solutions.*
2. *Selection: choosing a proportion of the current chromosome to produce new generation throughout the successive generation, these are chosen by depending on the Fitness Function and there is another way by choosing a random set of chromosomes, but this procedure can take a very long time.*
3. *Reproduction: it's the process of reproducing a new generation of chromosomes that can be chosen through testing process then make a Hybridization and mutation process to produce children.*
4. *The Crossover Process: through choosing parents that are chosen due to the choosing test, by couple mating to produce two new children, this continues until a new group of chromosomes is found plus a new group of parents.*

There are so many methods in the Process of Crossover:

- A. One point Crossover: this process leads to producing the next generation of inhabitants that are inconsistent to the first generation, where all the data will be assembled around this point to change the data with the condition of not repeating it.
- B. Two points Crossover: this process leads to producing the next generation of inhabitants that are inconsistent to the first generation, where all the data will be assembled around these two points to change the data with the condition of not repeating it.
- C. The Crossover Interfaces: this process leads to disconnecting the chromosome from a part that is different from the second chromosome which leads to change the length of the chromosome.
- D. Mutation: it is a descent sudden change which results from The Crossover Process, these forms change in the chromosome's shape by changing the chromosome components (Bit), the result is not due to parents. Generally, the Reproduction results in producing new chromosomes which can be implemented on by the Fitness Function to produce new children.
- E. Termination: it is the process of finding a new generation and it continues until the causes of completion takes place, reaching to the best solution or to the exact number of generation or achieving the desired Budget like the count of time or money or reaching the Local Minimum, with no lapsing ability.

3. The Practical Side

3.1. First: Data Description

The privet data of dropout during the period of (2007_2015) has been taken to serve the purpose of applying one of (Box-Jenkins) models to predict the proportion of student's dropout in the primary stage for males and females, and to analyze the time data that has been used in the GRETL 2016 program.

3.2. Second: Time Series Stability

After collecting all the students' dropout proportion for both males and females in the primary stage, the first step of the Box-Jenkins is to draw the time chain data to understand the chain's attitude.

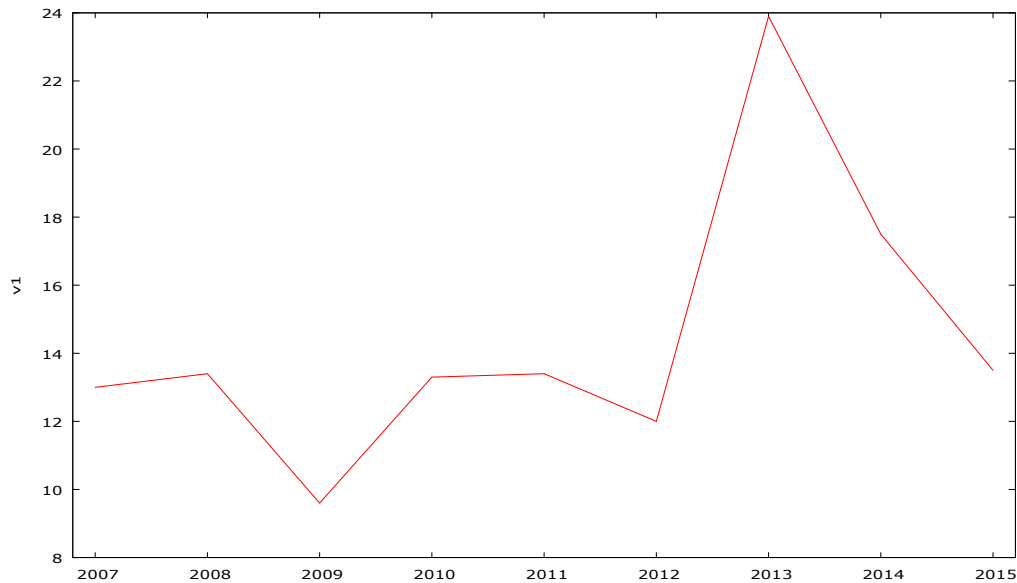


Figure 1. Drawing the time series for males primary stage

We get to notice the instability of the time series, and to be more precise we draw each (Autocorrelation Function) ACF, and (Partial Autocorrelation Function) PACF in a row to ensure the stability according to the figure 2.

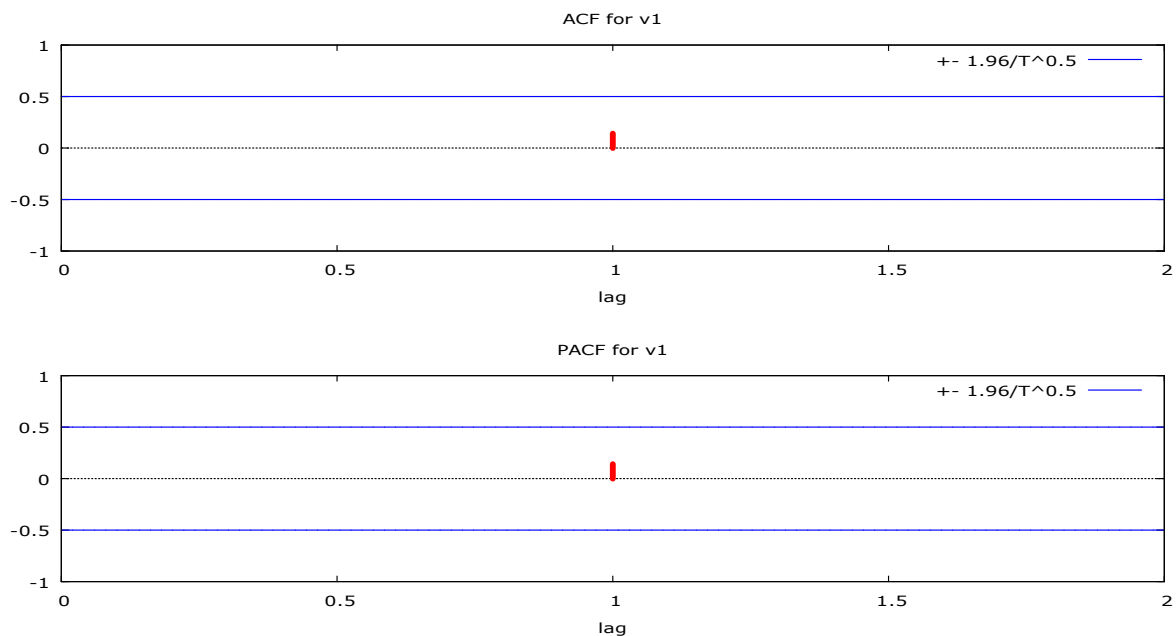


Figure 2. Drawing of autocorrelation function and partial correlation for males primary stage

The Unit Radix Dickey-Fuller Test is used to assure the series' stability. The results are:

Dickey-Fuller Test

Estimated Value = 0.276151, Statistic Test = 0.87796, P-Value = 0.8984

We get to notice from the values above P-Value = 0.8984 on the abstract level of 0.05 which leads to accepting the Null Hypothesis and refusing the Alternative Hypothesis (Existence of a

Radix Unit) implies that the time series is instable. By taking the first difference, it is observed that the stability of the Time Series has been accomplished . See figure 3:

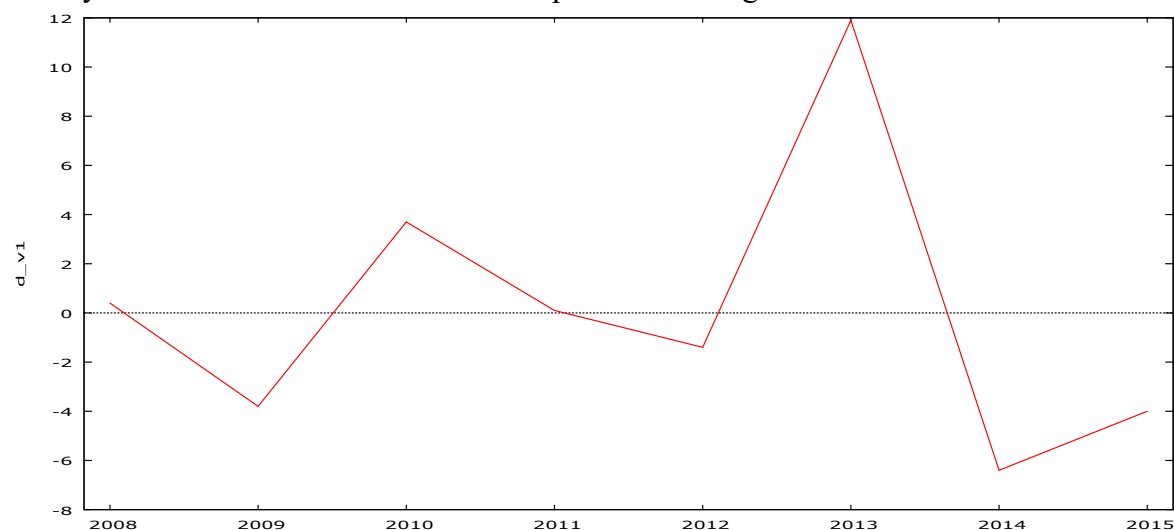


Figure 3. Drawing the time series for Males primary stage students after getting the first difference

We get to notice the stability of the time series, and to be more accurate we draw each (Autocorrelation Function) ACF, and (Partial Autocorrelation Function) PACF in a row to ensure the stability according to the figure (4).

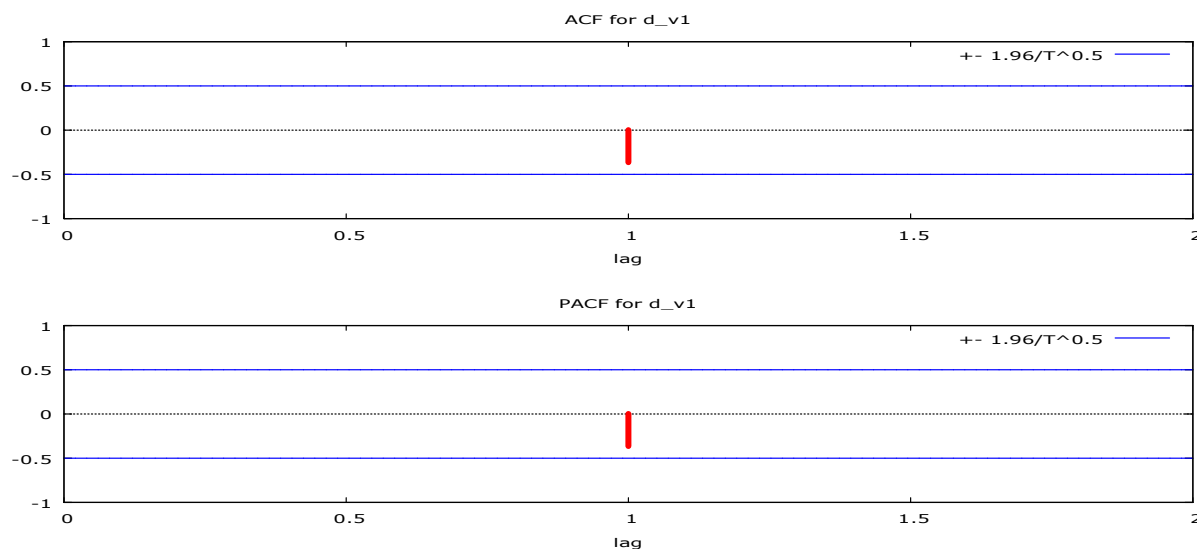


Figure 4. Drawing of autocorrelation function and partial correlation for males primary stage students

We use the Unit Radix Dickey-Fuller Test to assure the series' stability. The results are:

Dickey-Fuller Test

Estimated Value = 0.233403, Statistic Test = 0.125769, P-Value = 0.6405

The values above P-Value = 0.6405 is noted on the abstract level of 0.05 which leads to refusing the Null Hypothesis and accepting the Alternative Hypothesis (The Nonexistence of a Radix Unit) implies that the time series is stable. Figure 5 represents the Time series of Female Primary Stage Students:

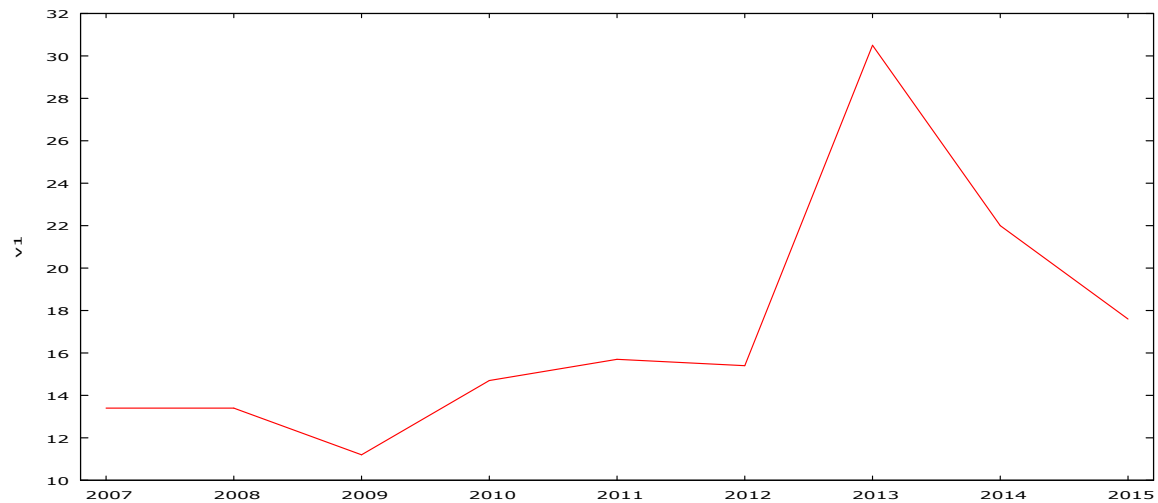


Figure 5. Drawing the time series for primary stage female students

The instability of the time series is noticed and to be more precise we draw each (Autocorrelation Function) ACF, and (Partial Autocorrelation Function) PACF in a row to affirm the stability according to the figure 6.

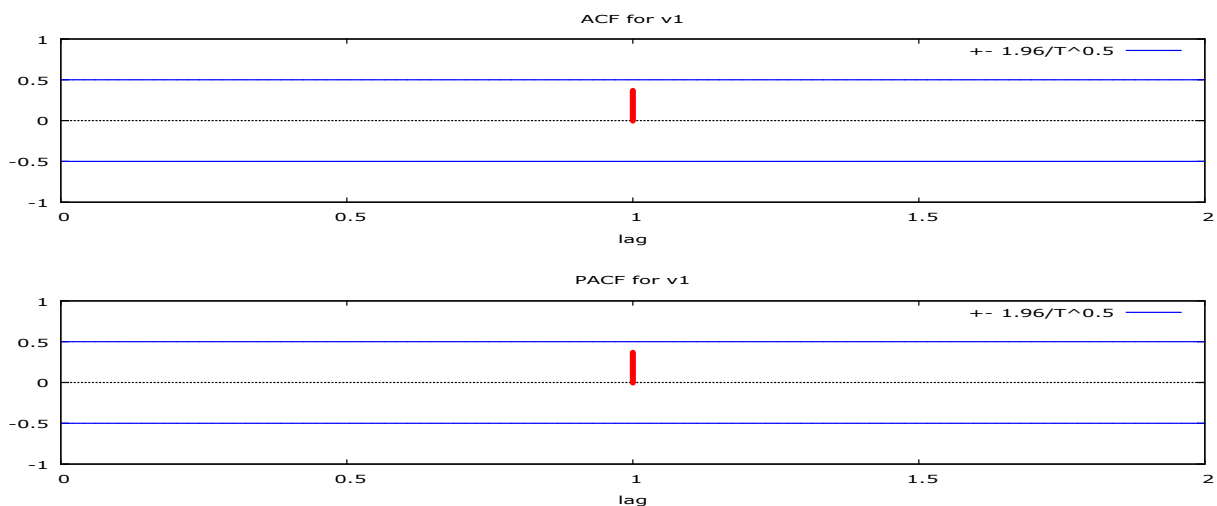


Figure 6. Drawing of auto correlation function and partial correlation for females primary stage students

We use the Unit Radix Dickey-Fuller Test to assert the series' stability. The results are:

Dickey-Fuller Test

Estimated Value = 0.47242, Statistic Test = 1.22797, P-Value = 0.9445

We notice from the values above P-Value = 0.9445 on the abstract level of 0.05 which leads to accepting the Null Hypothesis and refusing the Alternative Hypothesis (Existence of a Radix Unit) implies that the time series is instable. By taking the first difference, it is noticed that the stability of the time series has been accomplished. See Figure 7:

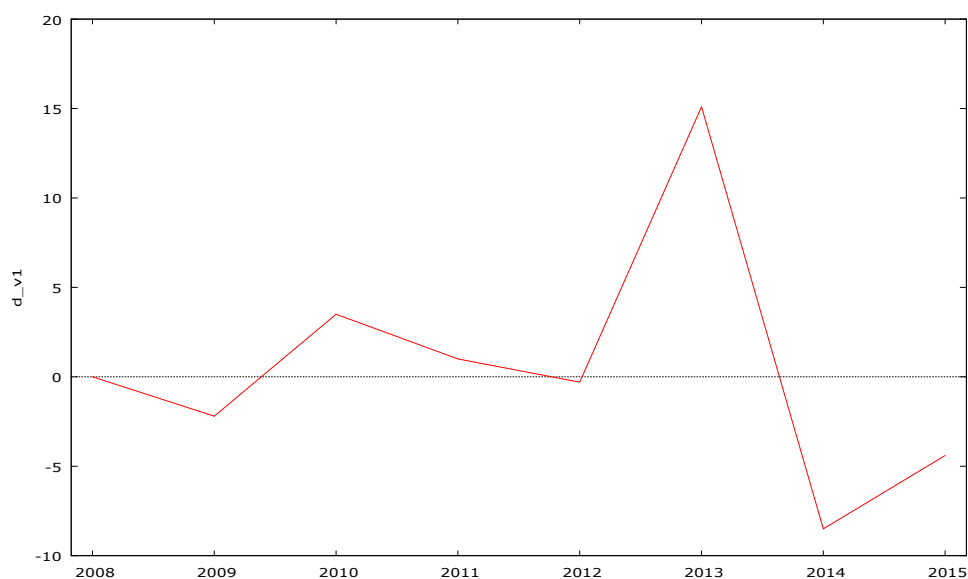


Figure 7. Drawing the time chain for Females primary stage students after the first difference

The stability of the time series is observed and to be more accurate we draw each (Autocorrelation Function) ACF, and (Partial Autocorrelation Function) PACF in a row to assure the stability according to the figure (8).

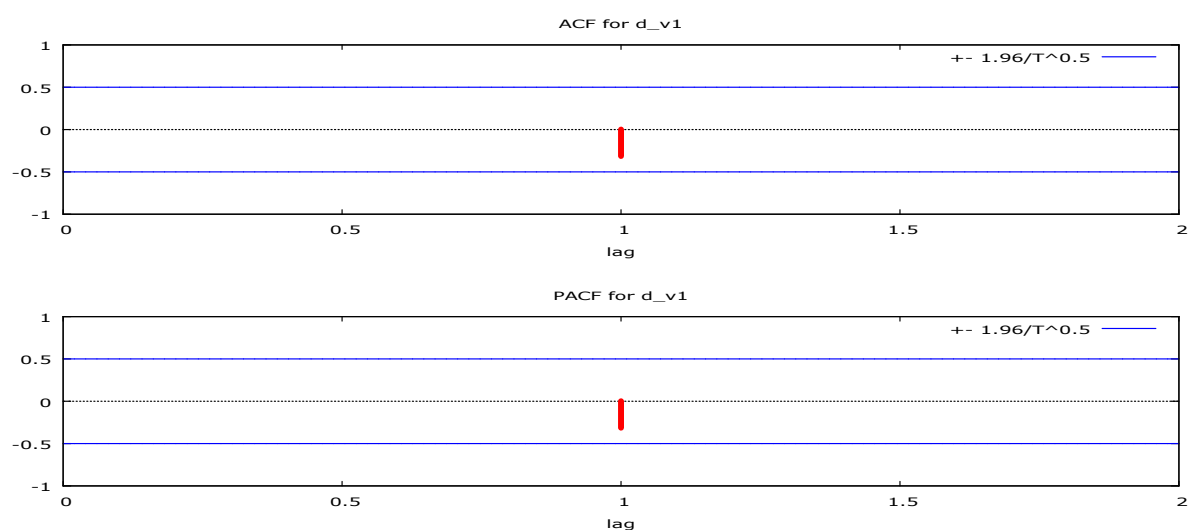


Figure 8. Drawing of autocorrelation function and partial correlation for Females primary stage students

The Unit Radix Dickey-Fuller Test is used to ensure the series' stability. The results are:
Dickey-Fuller Test

Estimated Value = 0.736458 , Statistic Test = 0.380545 , P-Value = 0.794

We notice from the values above P-Value = 0.794 on the abstract level of 0.05 which leads to refusing the Null Hypothesis and accepting the Alternative Hypothesis (The Nonexistence of a Radix Unit) implies that the time series is stable. Figure (9) represents the time series of females and males in the primary stage students:

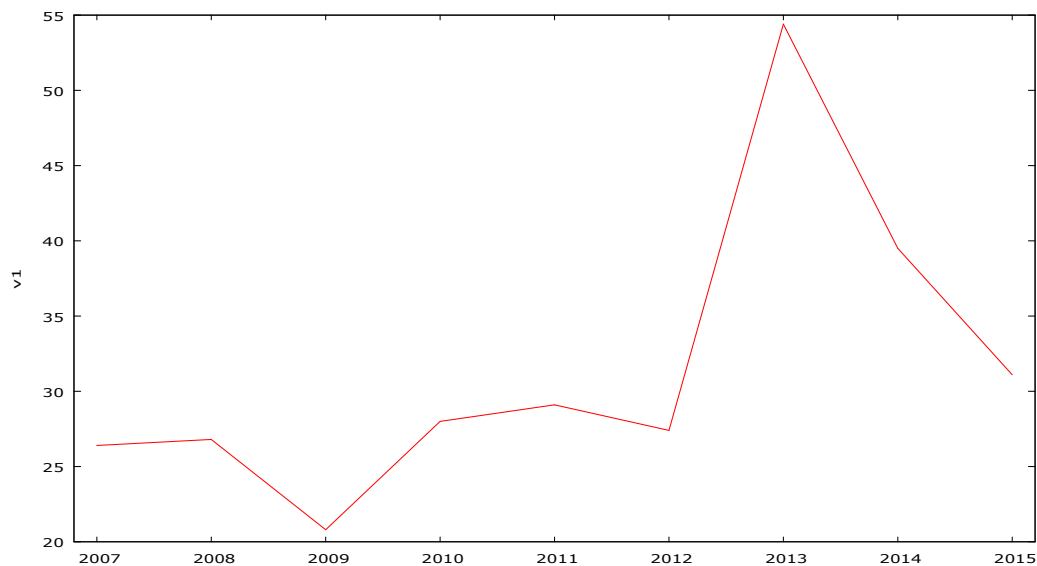


Figure 9. Drawing the time series for males and females primary stage students

The instability of the time series is recognized, and to be more accurate, we draw each (Autocorrelation Function) ACF, and (Partial Autocorrelation Function) PACF in a row to assure the stability according to the figure (10).

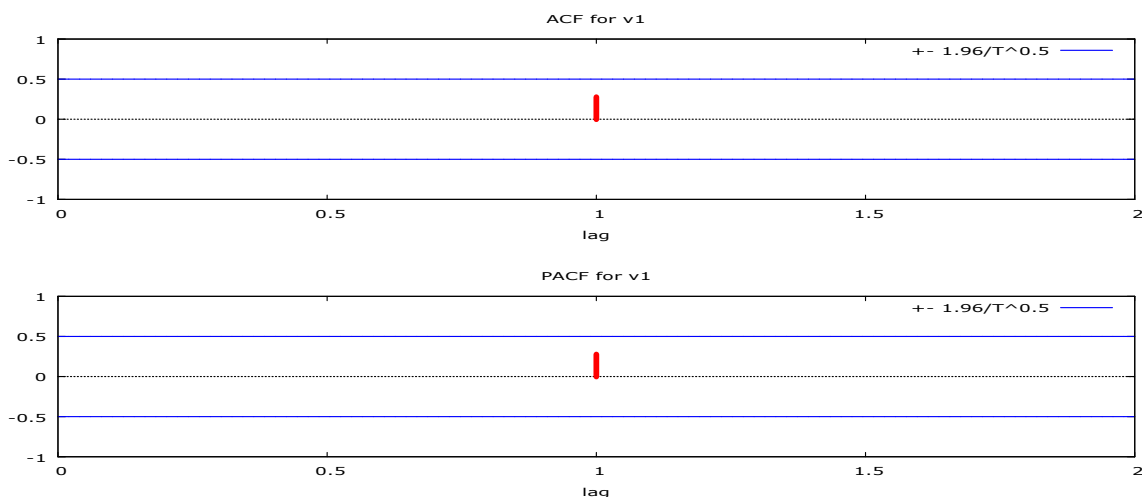


Figure 10. Drawing of autocorrelation function and partial correlation for Males and Females primary stage students

We use the Unit Radix Dickey-Fuller Test to ensure the series' stability. The results are:

Dickey-Fuller Test

Estimated Value = 0.369693, Statistic Test = 1.01829, P-Value = 0.9194

We notice from the values above P-Value = 0.9194 on the abstract level of 0.05 which leads to accepting the Null Hypothesis and refusing the Alternative Hypothesis (Existence of a Radix Unit) implies that the time series is instable. By taking the first difference, we notice that the stability of the time series has been achieved. See Figure 11:

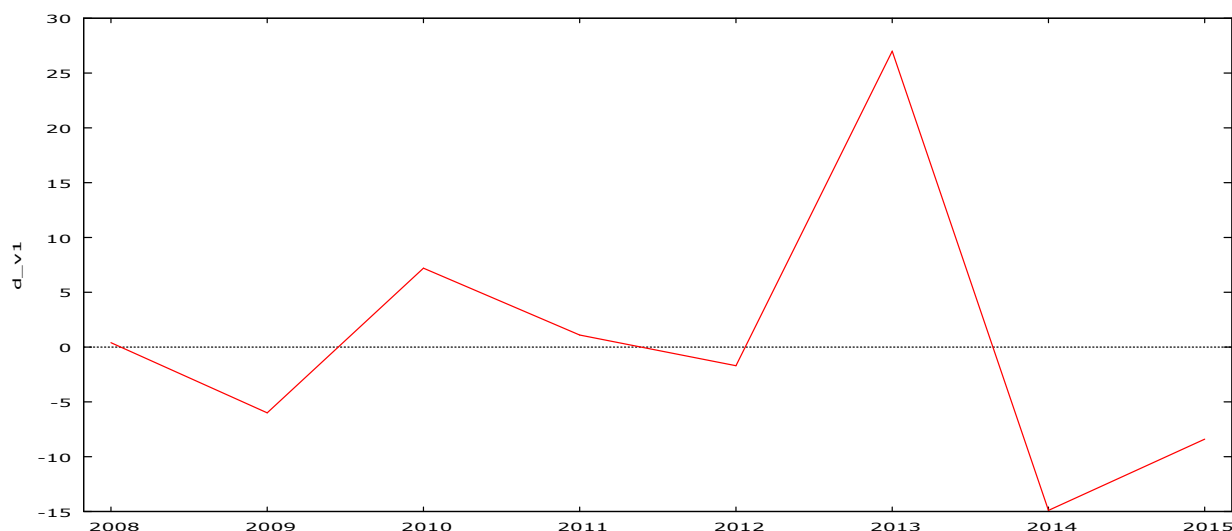


Figure 11. Drawing the time series for males and females primary stage after the First difference

The stability of the time series is noticed, so to be more precise we draw each (Autocorrelation Function) ACF, and (Partial Autocorrelation Function) PACF in a row to ensure the stability according to the Figure 12.

We use the Unit Radix Dickey-Fuller Test to assure the series' stability. The results are:

Dickey-Fuller Test

Estimated Value = 0.215889, Statistic Test = -0.114984, P-Value = 0.719

The values above P-Value = 0.719 is noticed on the abstract level of 0.05 which leads to refusing the Null Hypothesis and accepting the Alternative Hypothesis (The Nonexistence of a Radix Unit) implies that the time series is stable.

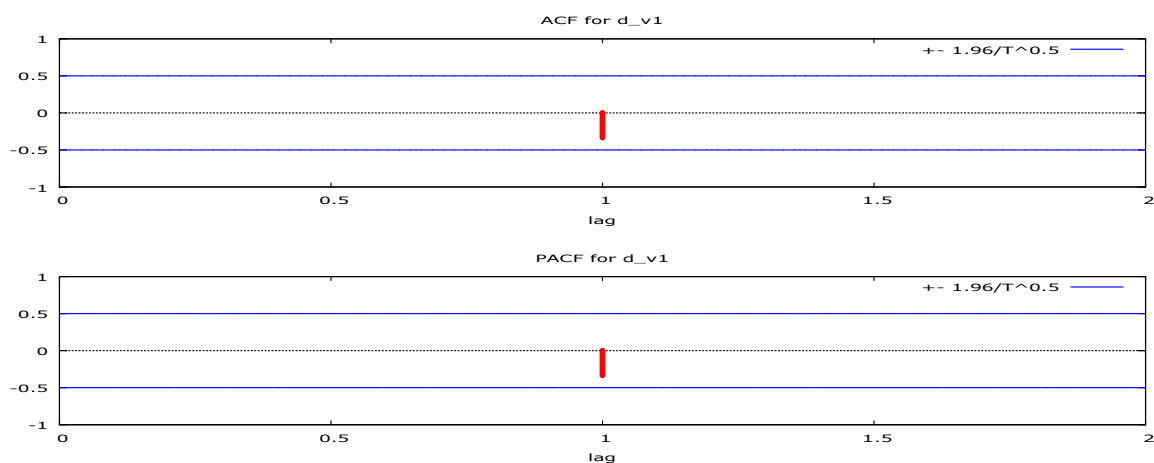


Figure 12. Drawing of autocorrelation function and partial correlation for females primary stage students

3.3. Third: Diagnosing the Time Chain Model ,estimate it and choosing the optimum

Accomplishing the time series Stability for the dropout proportion for both male and female students, let us move to the next step of diagnosing the appropriate model representing the Male and Female Time Chain. The (Akaike criterion), (Hannan-Quinn criterion) and (Schwarz criterion) all have been adopted to differentiate between set of models. The model with teachers estimates and has been adopted to represent the model tested results , with the grades and the comparison standers of the primary stage:

Table 2. The results of the models tested with the ranking and benchmarks primary stage

Akaike	Hannan-Quinn	Schwarz	Used grades	The primary stage
52.30670	50.96962	52.19852	ARIMA(1,1,0)	males
54.82685	53.48977	54.71867	ARIMA(1,1,0)	females
63.28090	61.94382	63.17272	ARIMA(1,1,0)	females+males

Table 2 shows the best model and according to comparison models which is established by its intangibility. It has been given the least value standers for the three used models, the model teachers have been estimated according to (the Maximum Faculty). Table (3) presents the Model Milestones and the Intangible Milestones for the males of primary stage students, as it's shown:

Table 3. The estimated and spirits parameters for the primary stage of males

	p-value	Z	std. error	coefficient	Parameters
**	0.0497	1.963	0.277415	-0.544476	phi_1

The reckoned model for males' primary stage is formed as:

$$\hat{Z}_t = -0.544476z_{t-1}$$

Table 4 presents the Model and Intangible Milestones for females of primary stage students as it is shown:

Table 4. The (estimated and spirits features of the primary stage for females

	p-value	Z	std. error	coefficient	Parameters
**	0.0562	1.910	0.279002	-0.532785	phi_1

The reckoned model for females primary stage is formed as:

$$\hat{Z}_t = 0.532785 z_{t-1}$$

Table 5 shows the Model and Intangible Milestones for Males Females primary stage students as it is presented:

Table 5. The (estimated and spirits parameters for primary school male and female)

	p-value	Z	std. error	coefficient	Parameters
**	0.0536	1.930	0.278258	-0.536956	phi_1

The reckoned model for males and females primary stage is formed as:

$$\hat{Z}_t = 0.532785 z_{t-1}$$

3.4. Forth: Testing the Accuracy Model:

After diagnosing and evaluating the models, the accommodating and the sufficiency of the models must be checked for Males primary stage, through applying the compute (Ljung-Box Q) to check the model accommodation on the function level 0.05 so the Q value occurs of males primary stage students:

$$\text{Ljung-Box } Q' = 1.28784,$$

$$\text{With p-value} = P(\text{Chi-square}(1) > 1.28784) = 0.2564$$

Knowing that the Tabulated value χ^2 equals 3.841, whereas the Q value is less than χ^2 Tabulated value, so it accepts the Null Hypothesis which indicates the emptiness of the evaluated model out of the contrast accordance trouble.

It's possible to notice the two parameters functions (Autocorrelation and Partial correlation Functions) of the residues for males primary stage, in which the residues value is located within the confidence interval limits which means the residues series is random and the Evaluated Model is good and convenient as it is presented:

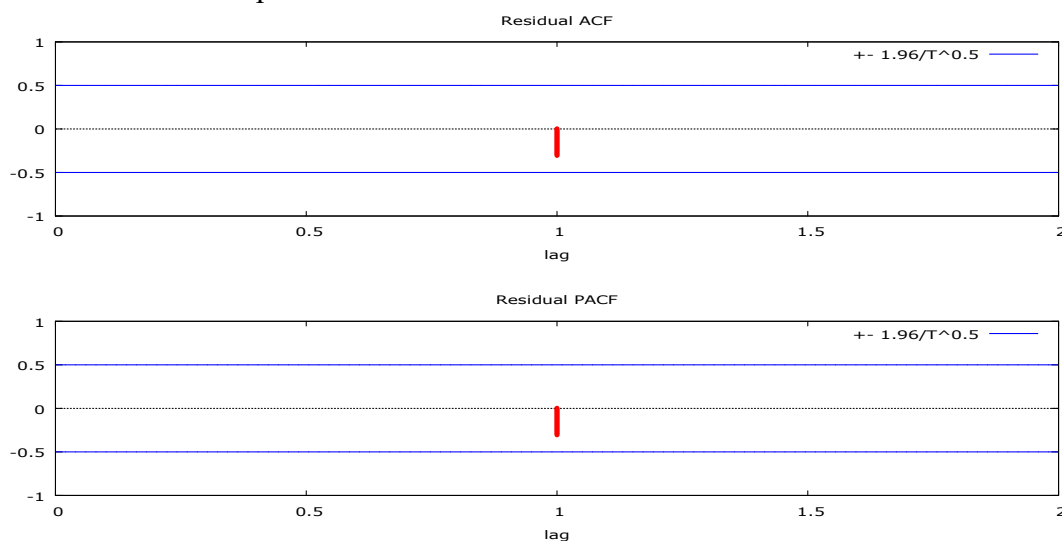


Figure 13. Drawing of autocorrelation Function and partial correlation of the residues for primary stage males students

After diagnosing and evaluating the models, the accommodating and the sufficiency of the models must be checked for primary stage female students, through applying the compute (Ljung-Box Q) to check the model accommodation on the function level 0.05 so the Q value occurs of primary stage female students:

Ljung-Box $Q' = 0.966626$,

With $p\text{-value} = P(\text{Chi-square}(1) > 0.966626) = 0.3255$

However, the Tabulated value χ^2 equals 3.841 whilst the Q value is less than χ^2 Tabulated value, so it accepts the Null Hypothesis which indicates the emptiness of the evaluated model out of the contrast accordance trouble.

It's possible to notice the two parameters functions (Autocorrelation and Partial Correlation Functions) of the residues for females primary stage students, in which the residues value is located within the confidence interval limits, which means the residues series is random and the evaluated model is good and convenient as it is shown:

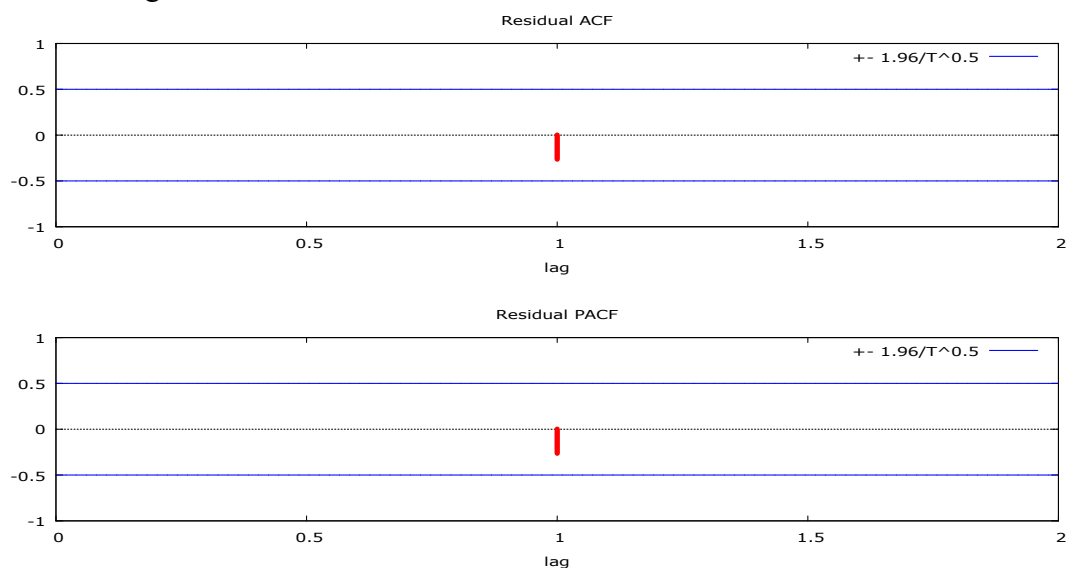


Figure 14. Drawing of autocorrelation function and partial correlation of the residues for primary stage males students

After diagnosing and evaluating the models, the accommodating and the sufficiency of the models must be checked for males and females of primary stage students, through applying the compute (Ljung-Box Q) to check the model accommodation on the Function level 0.05 so the Q value occurs of males and females of primary stage students:

Ljung-Box $Q' = 1.10306$,

With $p\text{-value} = P(\text{Chi-square}(1) > 1.10306) = 0.2936$

Note that the Tabulated value χ^2 equals 3.841 while the Q value is less than χ^2 Tabulated value, so it takes the Null Hypothesis which manifests that the emptiness of the evaluated model out of the contrast in accordance trouble.

It's possible to notice that the two parameters functions (Autocorrelation and Partial correlation Functions) of the residues for male females primary stage, in which the residues value is located within confidence interval limits, which means the residues series is random and the evaluated model is good and convenient as it is presented.

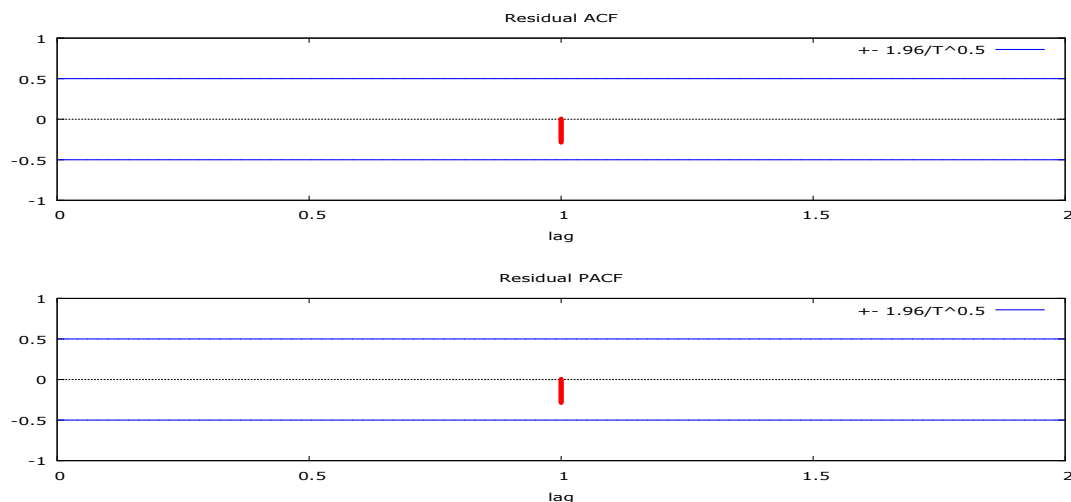


Figure 15. Drawing of autocorrelation function and partial correlation of the residues for males and females primary stage students

3.5. Fifth: Prediction

The time series values have been predicted using the Genetic Algorithm for the next five years, and it manifests the described results as shown in table (6):

Table 6. Predict the results of the standard deviation and confidence limits for each prediction primary stage

2020	2019	2018	2017	2016	Stage
11.2	11.2	10.8	11.0	11.1	Males
17.8	17.7	17.7	17.1	17.3	Females
31.6	31.1	29.9	30.1	29.8	Males and Females

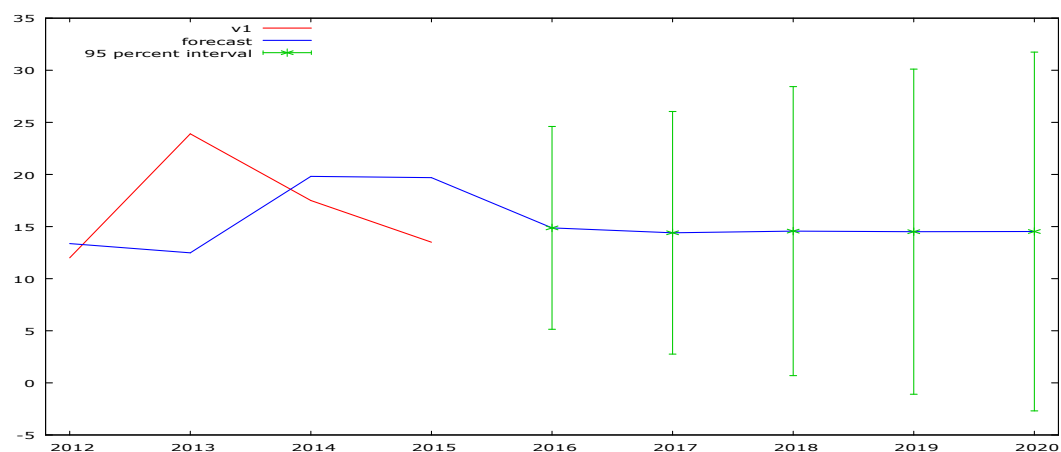


Figure 16. Drawing of the time series for males of primary stage students and its prediction

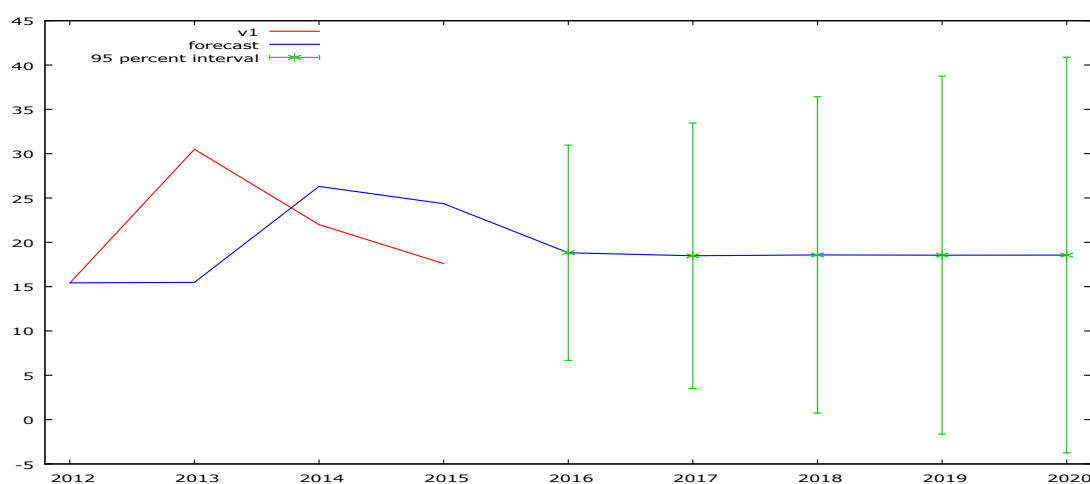


Figure 17. Drawing of the time series for females of primary stage students and its prediction

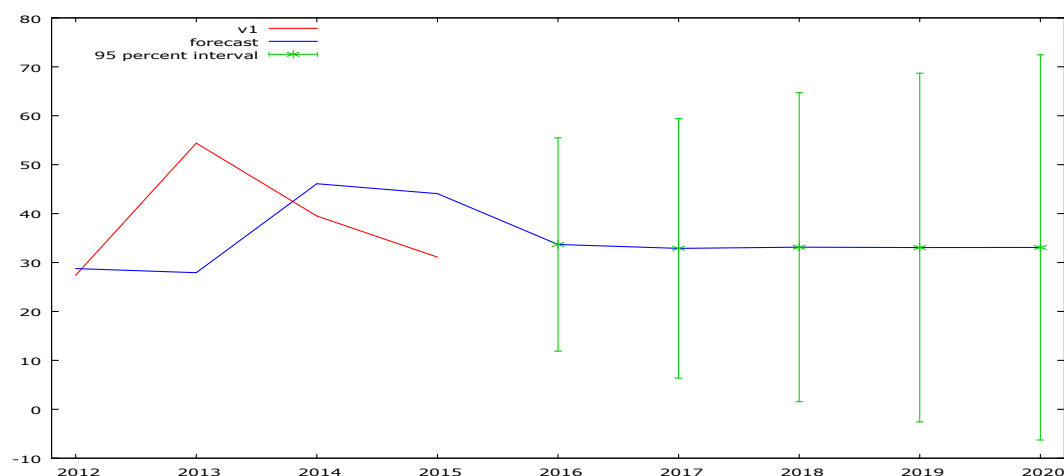


Figure 18. Drawing of the time series for males and females of primary stage students and its prediction)

4. Conclusions and Recommendations

Through the results, the study has reached to:

1. There is a noticeable rise in the dropout rate of both males and female students in the predicted years compared with the original data, so the dropout rate would reach to (32.6%)

in 2020; that is a very high percentage and it would put Iraq in a great embarrassment in the global education ranking.

2. There is a noticeable rise in the dropout rate for the predicted years compared with the original data, so the dropout rate would reach (11.2%) in 2020.
3. There is a noticeable rise in the dropout rate for the predicted years compared with the original data, so the dropout rate would reach to (17.8%) in 2020.
4. Female students' dropout rate is much bigger than male students dropout rate since several families are convinced with the idea that girls should stay at home after primary education.
5. Also, some families have financial difficulties which lead to the increase the dropout rate, and fathers tend to make their children work after the primary education to make a profit of their financial source.

Therefore, we recommend the following:

1. Increase the awareness and importance of girl's education in the society, through mass media and other facilities.
2. There must be an intervention by the government on child labor.

References

- Kasyok, Al. 2013. Simple Steps for Fitting Arima Model to Time Series Data for Forecasting Using R. *International Journal of Science and Research (IJSR)*, ISSN (Online): 2319-7064.
- Ong, C.-S., Huang, J. J., & Tzeng, G. T. (2005). Model identification of ARIMA family using genetic algorithms. *Applied Mathematics and Computation* 164 (2005) 885–912.
- Alabsi, F., Naoum, R. (2012). Comparison of Selection Methods and Crossover Operations using Steady State Genetic Based Intrusion Detection System. *Journal of Emerging Trends in Computing and Information Sciences*, Vol. 3, No.7, July 2012, ISSN 2079-8407.
- Asadul, I. (2007). Explaining And Forecasting Investment Expenditure In Canada: Combined Structural And Time Series Approaches (1961-2000). *Applied Econometrics and International Development*, Vol.7-1 (2007).
- Razali, N. M., Geraghty, J. (2011). Genetic Algorithm performance with different selection strategies in solving TSP, in *Proceedings of the World Congress on Engineering*, ISBN 978-988-19251-4-5, Vol II, July 6 - 8, London, U.K.