

## **Comparison Between ARIMA And Fourier ARIMA Model To Forecast The Demand Of Electricity In Sulaimani Governorate**

### **Botan Karim Ahmed**

Department of Information technology, College of Commerce, University of Sulaimani, Sulaimani, Iraq

**Botan.ahmed@univsul.edu.iq**

### **Sham Azad Rahim**

Department of Statistics and Computer, College of Commerce, University of Sulaimani, Sulaimani, Iraq

**Sham.rahim@univsul.edu.iq**

### **Bestan Bahaalddin Maarroof**

Department of Information technology, College of commerce, University of Sulaimani, Sulaimani, Iraq

**Bestan.Maarroof@univsul.edu.iq**

### **Hindreen Abdullah Taher**

Department of Statistics and Computer, College of Commerce, University of Sulaimani, Sulaimani, Iraq

**Hindreen.taher@univsul.edu.iq**

---

### **ARTICLE INFO**

#### **Article History:**

Received: 6/8/2020

Accepted: 21/9/2020

Published: Summer 2020

#### **Keywords:**

*Trend, ARIMA, Stationary, Box-Jenkins Models.*

---

### **ABSTRACT**

Electric energy is accounted as one of the major goods in human life, and also have a great role in progressing and developing the several sectors as economics, manufactures and any other sector related to daily use. In this study the monthly demand of electricity in Sulaimani governorate have been used, the main goal of the study is to choose appropriate model to forecast the monthly demand of electric in Sulaimani governorate for 12 months in 2020, the analyzing, results and comparison shows that FSARIMA(0,0,0)x(2,1,0)<sup>4</sup> is appropriate model



Doi:

10.25212/lfu.qzj.5.3.36

for this mission of forecasting which has minimum AIC among the other candidate models that equal to 0.28

## **1. Introduction<sup>[7]</sup>**

Decision making is so sensitive when it is based on the forecasting methods, therefore the researcher should be careful during selection of forecasting methods, these methods are depending on the number of observation and its run-time.

The methodology of this paper is dependent on Seasonal Autoregressive Integrated Moving Average (SARIMA) model because there is no doubt that the using of electric contain seasonal effect, and the four season in our country are distinguished. The form of SARIMA model is  $(p,d,q) \times (P,D,Q,s)$  where the first part  $(p,d,q)$  is non-seasonal ARIMA model and the second part  $(P,D,Q,s)$  is the seasonal. The seasonal part of the model consists of terms that are similar to the non-seasonal components of the model, but involve backshifts of the seasonal period. The application section contain the data that was used in this paper that comes from directory of electricity in Sulaimani city for monthly demand, while fitted six different models which are presented in table 3.3, and the maximum likelihood estimation method have been used to estimate the parameter of the best model that is adequate the data under consideration.

## **2. Materials and Methods**

### **2.1 Time series**

A range of data variables is said to be time series data, which is consisting of sequential observations on a quantifiable variable(s), which is making more than one interval time. Usually, the observations are arranged according to time and taken at orderly intervals (days, months, years). There are many use of time series data in daily life including; (Economics, Finance, Environmental, and Medicine), while a set of

variables  $x_t$  where each of them represent to certain time  $t$ , and written as:  $\{X_1, X_2, \dots, X_t\}$  or  $\{X_T\}$ , where  $T = 1, 2, \dots, t$

and  $X_t$  is the value of the variable at period of time  $t$ , then the goal is to create a model of the form:

$$X_t = f(X_{t-1}, X_{t-2}, \dots, X_{t-n}) + e_t \quad \dots\dots\dots (1)$$

Where  $X_{t-1}$  is  $X_t$  variable for attitude of lag 1 that is the past observations value,  $X_{t-2}$  is the  $X_t$  variable for attitude of lag 2 means two past observations value, etc., and it represents noise value which doesn't keep a pattern of forecasting. The  $X_t$  with  $x_{t-1}$  value is highly correlated if a time series keep a pattern repeating, where the cycle is an observations number in a steady cycle [8].

## **2.2 Time series Analysis** [1] [4]

The processes and expansion of modern organizations are completely need time series data also public and personal foundation are using time series data to manage the networks, more use this type of data to understanding of thousands of time series data that consist of economic and financial information so in any field of life time series are necessary.

The reality about data information dots that accepted through the time, use time series analysis and also it must be acceptable inside template like (AR, trend or seasonal variation). as previously defined time series analysis acclimated expectation in statistics and other quality of data so that is mean use time series to drawing suitable inference, so the main purpose of time series analysis is forecasting. Forecasting is used to predict model for future values based on formerly observed time series values [8].

## **2.3 Component of Time Series** [2] [10]

Data type and patterns display from the time series plots is a fundamental step in choosing a good model and forecasting procedure to a timeseries. Classical approaches of time series analysis are mainly affected with decaying the variation in series in terms of patterns in the time series data. The sources of variation are mostly classified into four main components. The components are:

### **2.3.1 Trend**<sup>[13]</sup>

A trend pattern exists when there is a long-term secular increase or decrease in the data the sales of many companies, the gross national production, and many other business or economic indicators follow a trend pattern in their movement over time.

### **2.3.2 Seasonal Variation**<sup>[18]</sup>

Is a timeseries the variations of seasonal are a short-term fluctuation that occurs in a year periodically, then it repeats year after year. The responsible for a pattern is customs of people or conditions of repetitive of seasonal variations as a major factor. In general, when a series is affected by seasonal factors the seasonality is occur

### **2.3.3 Cyclical Variations**<sup>[22]</sup>

A cyclical pattern exists when the data are influenced by longer term fluctuations in economic which is related with the business cycle. The products sell such as automobiles and main devicedisplay this type of pattern.the major distinction between a seasonal and a cyclical pattern is firstconstant length and repeat on a regular periodic basis while the latter varies in length and magnitude <sup>[4]</sup>.

### **2.3.4 Irregular Variation**<sup>[19]</sup>

The short duration fluctuations in a time series that is named by Irregular variation, which means it happened unsystematically variation at occurrence; these variations are also referred to as residual variations, cyclical and seasonal variations. Irregular

fluctuations result due to the occurrence of unforeseen events like floods, earthquakes, wars, famines, etc.

### **2.3.5 Stationary and Non-Stationary Series <sup>[11]</sup>:**

constant mean level is called stationary series decreasing or increasing regularly according to time with fixed variance. Non-stationary series have regular trends, such as linear, quadratic, and so on. Differencing can be used for change non-stationary series to stationary and is called “non-stationary in the homogeneous sense.” Stationary is used as a tool in time series analysis, where the raw data are often changed to become stationary. For example, economic data dependent on a non-stationary price level and they are often seasonal.

unreliable and dummy results product from non-stationary time series where cause understanding and less forecasting. while the problem solution occurs when the data connected by time become stationary, so for the random walk the process is non-stationary with or without a drift, so the process convert to stationary by differencing. Differencing the grade is the easiest way to produce a non-stationary mean stationary (flat). The times number you have to difference the grade to create the stationary process chooses the value of  $d$ . If  $d$  equal to 0, the model is stationary previously and does not have any trend. When the series is differenced once,  $d=1$  and linear trend is removed. When  $d=2$  and both linear and quadratic trend are removed. For non-stationary series,  $d$  values of 1 or 2 are usually adequate to produce the mean stationary. If the time series data analyzed exhibits a deterministic trend, the spurious results can be avoided by detrending. Sometimes merge the stochastic and deterministic trend at the same time cause non-stationary series and to avoid obtaining misleading results both differencing and detrending should be applied, as differencing will remove the trend in the variance and detrending will remove the deterministic trend <sup>[2][3]</sup>.

A process stationary has the property that the mean, variance and autocorrelation don't change over time. The mathematical term of Stationary can be defined in:

1. The mean  $\mu(t) = E(y(t))$
2. The variance  $\sigma^2(t) = \text{Var}(y(t)) = \gamma(0)$

There are two kinds of stationary:

### **2.4 Non-Stationary around Variance**

In the case fluctuation of time series about the contrast and this discrepancy is not fixed, it means that the series is stationary about the contrast, and there are transformations to convert the string non stationary to a series of stationary, including the conversion logarithmic and transfers of power and the square root of the absence of stationary, about the contrast non-fixed and turn it into a series of fixed and stationary contrast by applying the following formula<sup>[10]</sup>:

$$X_{\tau} = \begin{cases} X_{\tau}^{\lambda} & \lambda \neq 0 \\ X_{\tau} & \lambda = 0 \end{cases}$$

Where:  $X_t$ : the original series

### **2.5 Non-Stationary around the Mean** <sup>[6][9]</sup>

The basic conditional in being a stationary time series about mean and middle hard as the changes that occur in the qualities and characteristics of chains with time makes it unstable so you must remove the property not stability, these chains are used difference method (Difference) to convert the string unstable to a series stable in terms of time difference and take her first and be in the following format:

$$\Delta X = X_{t_i} - X_{t_{i-1}}$$



$$W_t = \Delta X_t = X_t - X_{t-1}$$

Where:

$\Delta$ : the difference factor

$W_t$ : the new series

$X_t$ : the authentic series

### 2.6 Box-Jenkins Models:<sup>[20][23]</sup>

In 1970 the George-Box and Gwilyn Jenkins applied his model in time series data. It's model show that if the data stationary or not and solve the non-stationary data by differencing one or more than to become stationary data with the "I" standing for "Integrated" of an ARIMA model, also the box-Jenkins is used to solve many time series problem. This methodology depends on parts of procedure which is [autoregressive (AR), moving average (MA) and autoregressive moving average (ARMA)] that can be explained as follow:

#### 2.6.1 Autoregressive (AR) model <sup>[9][12]</sup>

The order of AR model can be determined based on the statistically significant partial autocorrelation function (PACF), the (P) stands for the order of AR model, the AR model can be written as follow:

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

By using backshift operator equation (3) can be rewrite as follow:

$$\phi_p(B_p)X_t = a_t \dots \dots \dots (3)$$

Where:  $\phi_p(B_p) = (1 - \phi_1 B_1 - \dots - \phi_p B^p)$

$X_t$  : is the origin series.

$a_t$  : is white noise,  $a_t \sim N(0, \sigma_a^2)$

$\phi_p$  : is the predicted PACF.

To calculate Variance-Covariance the equation (3) should be multiplied by  $(X_{t-k})$  and taking expectation so we get:

$$E(X_t X_{t-k}) = E(\phi_1 X_{t-1} X_{t-k} + \phi_2 X_{t-2} X_{t-k} + \dots + \phi_p X_{t-p} X_{t-k} + a_t X_{t-k}) \dots (4)$$

Note:

$$E(X_t X_{t-k}) = \lambda_k .$$

$$E(a_t X_{t-k}) = 0 .$$

Then:

$$\lambda_k = \phi_1 \lambda_{k-1} + \phi_2 \lambda_{k-2} + \dots + \phi_p \lambda_{k-p} ; K > 0 \dots (5)$$

To get the ACF the equation (5) should be divided by the variance of the series ( $\gamma_0$ ).

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \dots (6)$$

Note:

$$\frac{\lambda_k}{\lambda_0} = \rho_k, \gamma_0 = \sigma_X^2$$

Then the PACF for the AR(P) model can be estimated by using Yule-Walker equations

$$\rho_j = \phi_k \rho_{j-1} + \phi_{k(k-1)} \rho_{j-2} + \dots + \phi_{kk} \rho_{j-p} \dots (7)$$

**2.6.2 Moving Average (MA) Model <sup>[10][16]</sup>:**



the number of significant ACF is defined order of moving average model and  $-\Theta_1$  is the coefficient of dependency of observations ( $X_t$ ) on the error term  $e_t$  and the previous error term  $a_{t-1}$ , the MA(q) model can be written as follow:

$$X_t = a_t - \Theta_1 a_{t-1} - \Theta_2 a_{t-2} - \dots - \Theta_q a_{t-q} \dots \dots \dots (8)$$

Equation (8) can be reformulate with back shift operator as follow:

$$X_t = \Theta(B)a_t \dots \dots \dots (9)$$

Where:

$$\Theta(B) = 1 - \Theta_1 B - \dots - \Theta_q B^q$$

The Var-Cov of MA(q) model is:

$$\gamma_k = \begin{pmatrix} \sigma_a^2 (-\Theta + \Theta_1 \Theta_{k+1} + \dots + \Theta_q \Theta_{q-k}) & ; K = 1, 2, \dots, q \\ 0 & ; K > q \end{pmatrix} \dots \dots (10)$$

$$\gamma_0 = \sigma_a^2 \sum_{i=0}^q \Theta_i^2 \dots \dots \dots (11)$$

Note:

$$\Theta_0 = 1.$$

And the ACF is:

$$P_k = \begin{pmatrix} \frac{-\Theta_k + \Theta_1 \Theta_{k+1} + \dots + \Theta_q \Theta_{q-k}}{1 + \Theta_1^2 + \Theta_2^2 + \dots + \Theta_q^2} & ; k = 1, 2, \dots, q \\ 0 & ; k > q \end{pmatrix} \dots \dots (12)$$

**2.6.3 Autoregressive Moving Average Model (ARMA)<sup>[14][17]</sup>**

There is a large family of models which is named "Autoregressive-Moving Average Models" and abbreviated by ARMA. Many of researchers in different

application fields prove that ARMA models fits more than other traditional methods for forecasting [10]. The more general model is ARMA model because it is mix model between AR(p) and MA(q) models and it is called an ARMA model of order (p,q). The ARMA(p,q) is given by [19]:

$$\phi_p(B)X_t = \Theta_q(B)a_t \quad \dots\dots\dots 13$$

Where:

$$\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

$$\Theta_q(B) = 1 - \Theta_1 B - \dots - \Theta_q B^q$$

We write equation (13) as:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + a_t - \Theta_1 a_{t-1} - \dots - \Theta_q a_{t-q} \dots (14)$$

**2.6.4 The Autoregressive Integrated Moving Average Models (ARIMA)<sup>[19][15]</sup>**

When you have not stationary time series, by difference operator can be change non-stationary to stationary, the time series data after differencing is called adjusted data and the fitted model is called integrated model which combines both autoregressive and moving average models.

Generally, all AR(p) and MA(q) models can be figured as ARIMA(1,0,0) this tells there is no differencing and no MA part. The general form of ARIMA model is written as (p,d,q) while a degree of AR part is P and the differencing degree is D also MA for q<sup>[11]</sup>.

$$W_t = \nabla^d X_t = (1 - B)dX_t$$

The general ARIMA process is of the form

$$X_t = \sum_{t=1}^p \phi_i X_{t-1} + \sum_{t=1}^q \Theta_i a_{t-1} + \mu + a_t \dots \dots \dots (15)$$



**2.7 Seasonal Time Series Models<sup>[13]</sup>**

Seasonal model perhaps used in a manner similar to that described thus far for no seasonal series. For mix models, the exact order of the AEMA process –p and q – and the length of seasonality must be specified. The sequential ARMA models are much simpler to apply. one starts by specifying as many parameters as the length of seasonality and then inspecting the autocorrelation of residual. In the seasonal model the form of seasonality shown by the significant ACF and PACF patterns. The number of times over year where seasonal revision happen is further important for yearly seasonality the ACF spikes are raise patterns at seasonal lags according the regular non-seasonal variation once a year. If the seasonality is quarterly, there will be eminent ACF spikes four times a year <sup>[5]</sup>.

**2.8 Seasonal Autoregressive Integrated Moving Average (Seasonal ARIMA) Models<sup>[21][22]</sup>**

The seasonal difference is an important tool in non-stationary seasonal modeling. The difference of seasonality of period s for the series {X<sub>t</sub>} is symbol by ∇<sub>s</sub>X<sub>t</sub> and is expressed as:

$$\nabla_s X_t = X_t - X_{t-s} \quad \dots\dots\dots (16)$$

The difference of seasonality will be n-s if the time series of size n while s is data values and lose because of the difference of seasonality so a time series X<sub>t</sub> is called a seasonal multiple ARIMA by non-seasonal order p,d and q and seasonal periods . if the differenced series expressed as below:

$$W_t = \nabla^d \nabla_s^D X_t \quad \dots\dots\dots (17)$$

the ARMA model (p×q)(P×Q)s with seasonal period s. leads to X<sub>t</sub> be an ARIMA (p,d,q)(P,D,Q)s model with seasonality in a period s. the non-seasonal part is equal with the seasonal part of any seasonal ARIMA model because they have an ARIMA

factor and differencing order. To generalize ARIMA model to agree with seasonality use the test box and Jenkins, also the popular multiply SARIMA model in the form:

$$\phi(B)(1 - B^4)X_t = \Theta(B)(B^4)a_t \dots\dots\dots (18)$$

Where the backward shift operator denoted by B,  $\Phi$ ,  $\phi$ ,  $\Theta$ , and  $\Theta$  are polynomials with p, l, q, degrees and Q so  $a_t$  is random process with mean (0) and constant variance  $\sigma_a^2$ . Consider a Seasonal ARIMA (0, 1, 1) × (0, 1, 1)<sub>4</sub> model for instance. The model specification is:

$$(1 - B^4)(1 - B)X_t = (1 - \Theta_1 B)(1 - \Theta_1 B^4)a_t \dots\dots\dots (19)$$

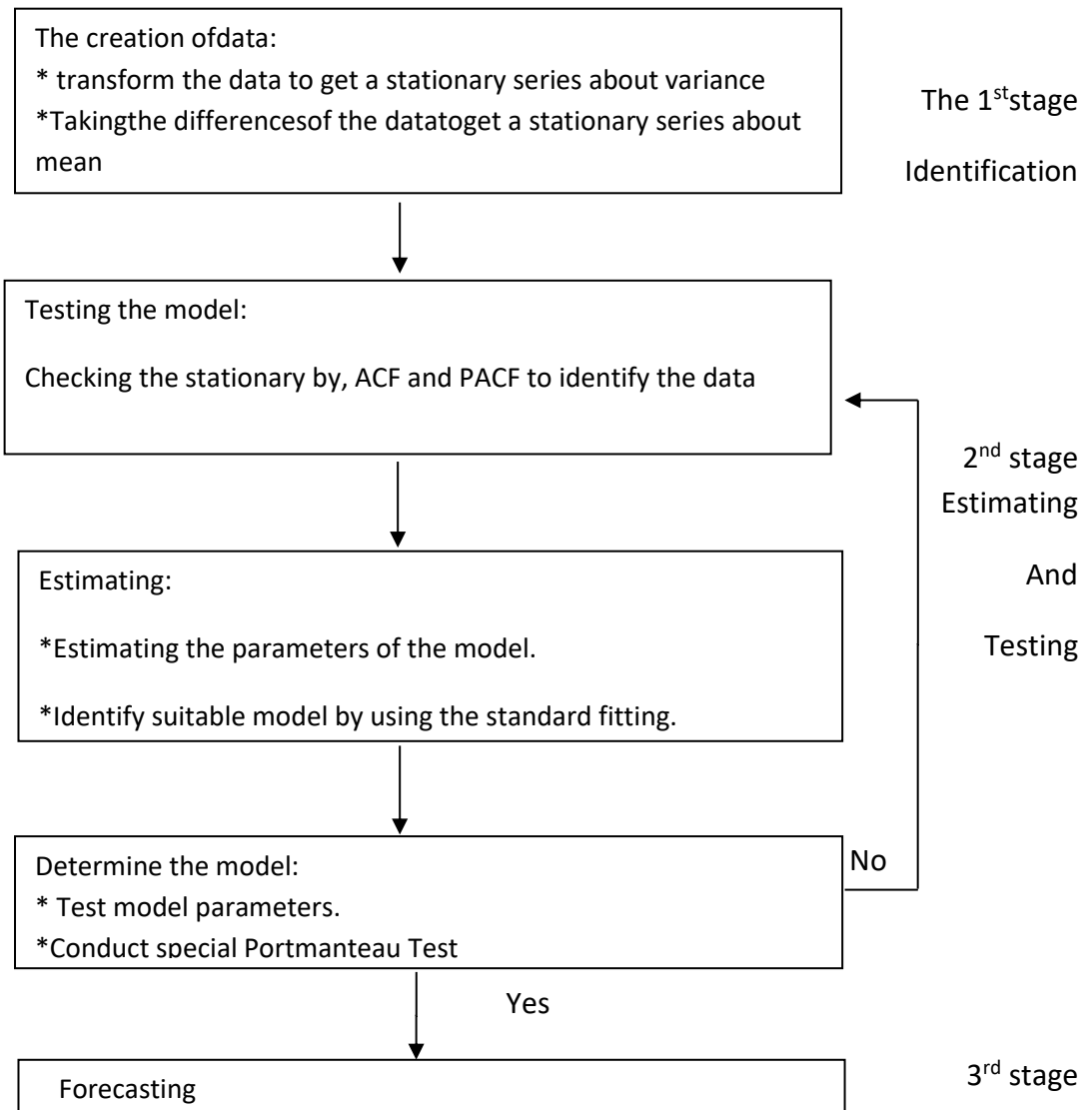
By expansion, we have:

$$X_t = X_{t-1} + X_{t-4} - X_{t-5} + a_t - \Theta_1 a_{t-1} - \Theta_1 a_{t-4} + \Theta_1 \Theta_1 a_{t-4}$$

The predicts of the parameters of the model in (19) may be obtained using the method of maximum likelihood [5], [15].

**2.9 Time series analysis steps:<sup>[18]</sup>**

build any model of Box – Jenkins step by step to forecast and can represent the stages of the scheme follows (1):



Applying

**Figure 2.1:** Represents the Box-Jenkins step by step scheme for forecasting

**2.10 Augmented Dickey-Fuller test**

Dickey-Fuller is used to test the hypothesis when exists a non-stationary series, and can be tested in regression equation [9].

$$\Delta X_t = \beta_0 + a_t + \beta_1 X_{t-1} + \sum_{i=1}^p \lambda_i \Delta X_{t-i} + \varepsilon_t \dots \dots \dots (20)$$

Where a random walk,  $a_t = a_{t-1} + a\varepsilon_t$  is allowed [9].

**2.11 Autocorrelation Function (ACF)[4]**

In a time series the order of correlation between neighboring observations measures by ACF. The autocorrelation coefficient is estimated from sample observation using the formula [10]:

$$\rho_k = \frac{\sum_{r=1}^n (X_r - \mu_x)(X_{r+k} - \mu_x)}{\sum_{r=1}^n (X_r - \mu_x)^2} \dots \dots \dots (21)$$

Thus, the autocorrelation function at lag k is characterized as:

$$\rho_k = \frac{\lambda_k}{\lambda_0} , \quad k = 0, \pm 1, \pm 2, \dots$$

**2.12 Partial Autocorrelation Function (PACF)[12]**

correlation between  $X_t$  variable and  $X_{t-m}$  is known by partial autocorrelation function after take of the impact of the intervening variables  $X_{t-1}, X_{t-2}, \dots, X_{t-m+1}$  which locate

within (t, t-m) period, partial autocorrelation function will be denoted by  $\phi_{mm}$ , PACF is calculated by iteration <sup>[17]</sup>.

$$\phi_{00} = 1$$

$$\phi_{11} = \rho_1$$

$$\phi_{mm} = \frac{\rho_{kk} - \sum_{j=1}^{m-1} \phi_{m-1,j} \rho_{m-j}}{1 - \sum_{j=1}^{m-1} \phi_{m-1,j} \rho_j}, m = 2,3, \dots \dots \dots (22)$$

Therefore  $\phi_{mj} = \phi_{m-1,j} - \phi_{mm} \phi_{m-1,m-2}, j = 1,2, \dots, m - 1$

**2.13 Model Selection Criteria<sup>[7]</sup>**

**2.13.1 Akaike Information Criterion AIC<sup>[2]</sup>**

Akaike Information Criterion AIC is defined as

$$AIC = -2 \log L + 2p \dots \dots \dots (23)$$

where L: maximized likelihood function

p: number of emotional parameters.

smallest AIC is the best. The likelihood function part reverses the appropriate of fit of the model to the data, while 2p is described as a penalty. Since L usually increases with p, AIC reaches the minimum at a certain p. AIC is depend on the information theory.

**2.14 Estimating the Parameters of an ARMA Model<sup>[9]</sup>**

Estimating the parameters of the ARMA model used Iterative method. At every point sum square residual should be calculated of suitable grid of the parameter values, and the sufficient values are given minimum sum of squared residuals. For an ARMA (1,1) the model is given by

$$X_t - \mu = \phi_1(X_{t-1} - \mu)a_t + \theta_1 a_{t-1} \dots\dots\dots (24)$$

Given N observation  $X_1, X_2, \dots, X_N$ , we guess values for  $\mu, \phi_1, \theta_1$ , set  $a_0 = 0$  and  $Y_0 = 0$  and then calculate the residuals recursively by

$$a_1 = X_1 - \mu$$

$$a_2 = X_2 - \mu - \phi_1(X_1 - \mu) - \theta_1$$

$$a_N = X_N - \mu - \phi_1(X_1 - \mu) - \theta_1 a_{N-1}$$

The residual sum of squares  $\sum_{t=1}^N a_t^2$  is calculated. Then other values of  $\mu, \phi_1, \theta_1$ , are tried until the minimum residual sum of squares is found [7], [12].

Note: It antiquated found that the most of the stationary time series occurring in practices can be fitted by AR(1), AR(2), MA(1), MA(2), ARMA(1,1) or white noise models that are customarily needed in practice<sup>[10]</sup>.

**2.15 Models Forecasts<sup>[22]</sup>**

After building model the main goal in this type of data is create forecast for the series. it also show an important role in assessing the forecasts accuracy. The power or ability to forecast is conclusive test of an ARIMA model. In order to obtain a forecast with a least error, there are seven features of a good ARIMA model taken into account [13], [15]. First, it has the minimum coefficients number which show in the data set. Secondly, a sufficient AR model must not be nonstationary. Thirdly, the MA must be invertible. Fourth, insufficient model the residuals must be independent. Fifth, the distribution of residuals of a good model must be distributed normal. From the exist series according time t, namely,  $X_1, X_2, X_3, \dots, X_{t-1}, X_t$ , we can forecast  $X_{t+h}$ , that will use h time units ahead. In this case, time t is the forecast source and the lead time forecast. This forecast is denoted and estimated as

$$\hat{X}_t(L) = E(X_{t+h} | X_1, X_2, \dots, X_t) \dots\dots\dots (25)$$



Once an enough and satisfactory model is fitted to the series of benefit, forecasts can be produced using the model [13].

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + a_t - \theta_1 a_{t-1} \dots - \theta_q a_{t-q} \dots \dots \dots (26)$$

The one-step-before forecast for  $t + 1$  is shown as follow:

$$x_{t+1} = \phi_1 x_t + \dots + \phi_p x_{t-p+1} + a_{t+1} - \theta_1 a_t \dots - \theta_q a_{t-q+1} \dots \dots (27)$$

**2.16 Fourier Residual Modifications (FR)[20][22]**

The first conception of the Fourier model is derived from Tien. Meanwhile, the forecasting of primary series trends originally uses ARIMA, Then the ARIMA model is adaptive using the Fourier series. The main goal behind using Fourier series is to gain a better forecasting with a high precision. To achieve the accuracy of forecasting models, the Fourier series has been widely and profitably applied in adapting the residuals in time series forecasting models Fourier Residual Modification ARIMA is:

- 1- According to the original sequence to build the ARIMA model, the residual time series as the difference between the real value and model fitted value is obtained by:

$$e_t = Z_t - \hat{Z}_t \quad t = 1, 2, \dots, N \quad \dots \dots (28)$$

- 2- Transfer  $e_{it}$  into Fourier series from

$$e_t = 0.5a + \sum_{i=1}^F \left[ \alpha_i \cos\left(\frac{2\pi i}{T}(t)\right) + b_i \sin\left(\frac{2\pi i}{T}(t)\right) \right] \dots \dots (29)$$

where  $t = 2, 3, \dots, m$  ,  $T = m - 1$  and  $F = \left(\frac{m - 1}{2}\right) - 1$

3- from  $et = PC$  to get the value of

$$C = (P^T P)^{-1} P^T [et]^T \dots\dots (30)$$

Which

$$p =$$

$$\left[ \begin{array}{c} \frac{1}{2} \cos\left(2 \frac{2\pi}{T}\right) \sin\left(2 \frac{2\pi}{T}\right) \cos\left(2 \frac{2\pi 2}{T}\right) \sin\left(2 \frac{2\pi 2}{T}\right) \dots \cos\left(2 \frac{2\pi F}{T}\right) \sin\left(2 \frac{2\pi F}{T}\right) \\ \frac{1}{2} \cos\left(3 \frac{2\pi}{T}\right) \sin\left(3 \frac{2\pi}{T}\right) \cos\left(3 \frac{2\pi 2}{T}\right) \sin\left(3 \frac{2\pi 2}{T}\right) \dots \cos\left(3 \frac{2\pi F}{T}\right) \sin\left(3 \frac{2\pi F}{T}\right) \\ \vdots \\ \frac{1}{2} \cos\left(m \frac{2\pi}{T}\right) \sin\left(m \frac{2\pi}{T}\right) \cos\left(m \frac{2\pi 2}{T}\right) \sin\left(m \frac{2\pi 2}{T}\right) \dots \cos\left(m \frac{2\pi F}{T}\right) \sin\left(m \frac{2\pi F}{T}\right) \end{array} \right]$$

se least square method where:

$$C = [\alpha_{(0)}, \alpha_{(1)}, b_{(1)}, \alpha_{(2)}, b_{(2)}, \dots, \alpha_{(F)}, b_{(F)}]$$

4- substitute all data into the Fourier series to get the value of  $\hat{e}_t$

$$\hat{e}_t = 0.5\alpha^{(0)} + \sum_{i=1}^F \left[ \alpha_i \cos\left(\frac{2\pi i}{T}(t)\right) + b_i \sin\left(\frac{2\pi i}{T}(t)\right) \right] \dots\dots\dots (31)$$

5- the final prediction value is show in equation (32)

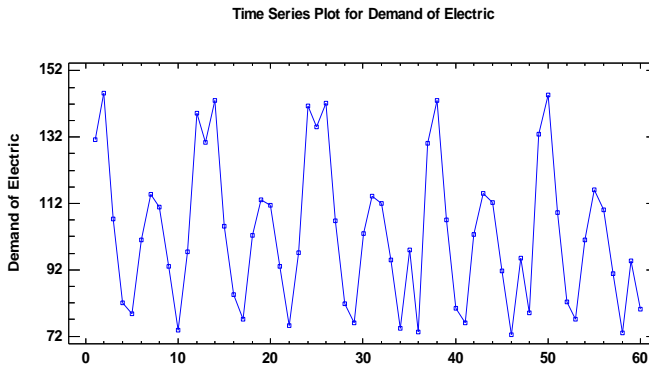
$$\hat{Z}_{i(t)} = \hat{Z}_i + \hat{e}_t, t = 1, 2, \dots, N \dots\dots\dots (32)$$

### 3.1 Data Description

The dataset used for the analysis in this study came from the direcorey of electricity of Sulaimani city which is contained one variable and deals with monthly demand of electricity since January 2015 up to December 2019. These data measurement is Gigahertz.

### 3.2 Applications

The time series plots are display observations on the y-axis against equally spaced time intervals on the x-axis. They are used to evaluate patterns, knowledge of the general trend and behaviors in data over time. The time series plot of monthly Demand of Electriciyin Sulaimani city displayed in Figure 3.1 below:



**Figure 3.1:** Monthly plot of Demand of Electricity in Sulaimanicity

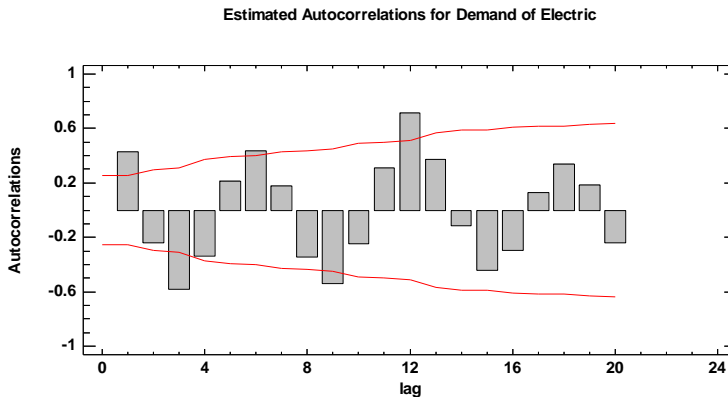
Figure 3.1 illustrate that the data of time series is not random. The plot shows consistent style of short-term changes for data which indicates the existence of seasonal fluctuations. This series varies randomly over time and there are seasonal fluctuations. For further testing of the stationery of the time series, we applied Augmented Dickey-Fuller test for monthly Demand of Electriciyin Sulaimani city. The augmented Dickey-Fuller tests the hypotheses which is stated that the series is non-stationary series. Table 3.1 shows the results of ADF of the data of the time series of monthly Demand of Electriciyin Sulaimani city.

**Table 3.1:** ADF test for monthly Demand of Electric in Sulaimani city

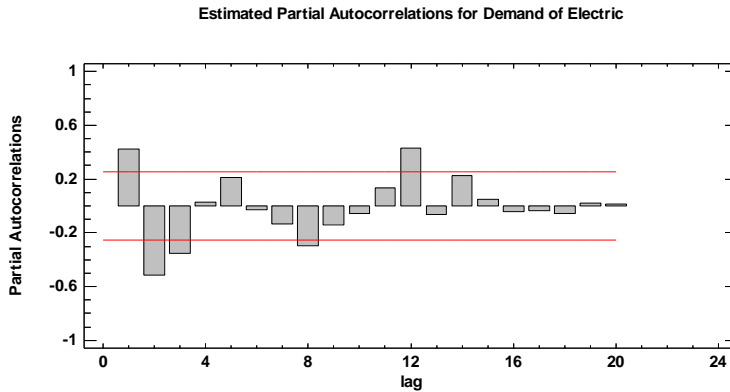
Test	t-Statistic	P-value
------	-------------	---------

ADF	0.22471	0.9992
-----	---------	--------

Table (3.1) explain that the p-value of the ADF test equals 0.9992 and it is greater than 0.05. This result indicates that the data of time series of monthly Demand of Electricity in Sulaimani city is not random, and demonstrates this results by the examination of the autocorrelation and partial autocorrelation functions as shown below.

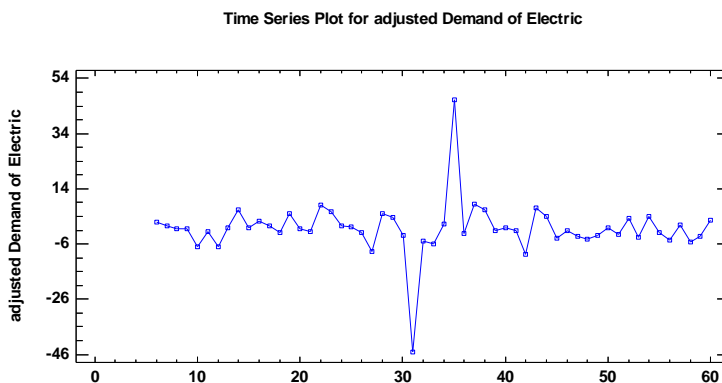


**Figure 3.2:** Autocorrelation Function for the monthly Demand of Electricity in Sulaimani city



**Figure 3.3:** Partial Autocorrelation Function for the monthly Demand of Electricity in Sulaimani city

All the above results and plots support that the series is not random at the level, which needs some treatments to be transformed to a random series. Also, the differencing of a series by suitable transformation. We see that the time series for the first-Seasonal differenced series in Figure 3.4 show that the data series is stationary.

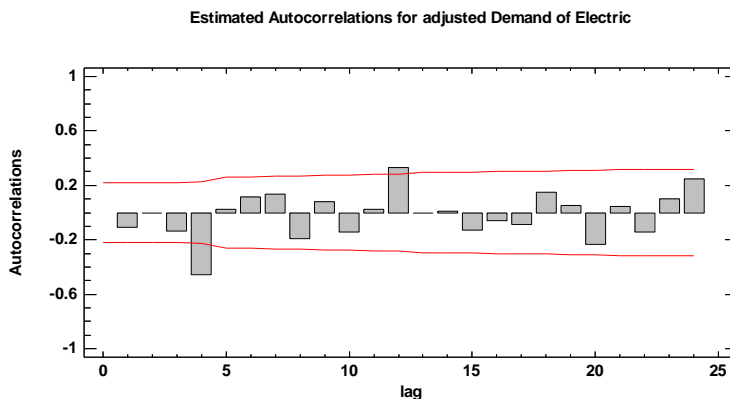


**Figure 3.4:** Time series plot of the first difference of monthly Demand of Electricity in Sulaimani city

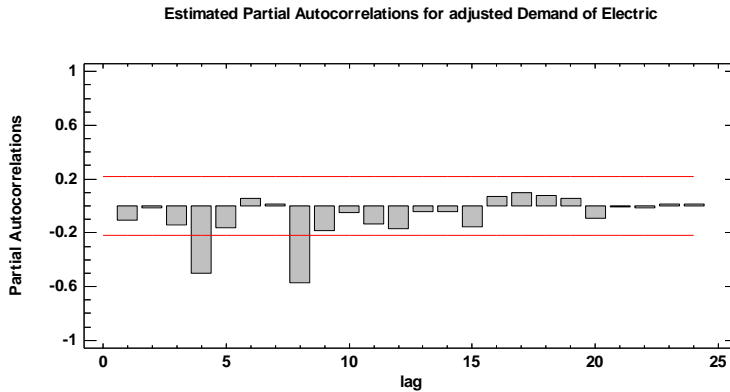
**Table 3.2:** ADF test of the first seasonal difference for monthly Demand of Electricity in Sulaimani city

Test	t-Statistic	P-value
ADF	-4.84370	0.000

Table (3.2) explain that the p-value of the ADF test equals 0.000 and it is less than 0.05. This result indicates that the non-stationarity hypotheses of the seasonal differenced monthly Demand of Electricity in Sulaimani city is rejected and this demonstrates by estimating the autocorrelation and partial autocorrelation function (ACF and PACF) for the first-differenced series in Figure 3.5 and 3.6



**Figure 3.5:** Autocorrelation Function for the first-seasonal differenced series of the monthly Demand of Electricity in Sulaimani city



**Figure 3.6:** Partial Autocorrelation Function for the first-differenced series of the monthly Demand of Electricity in Sulaimani city

The results above demonstrate the success of differencing of the time series data of the monthly Demand of Electricity in Sulaimani city. Thus, the data series became stationary.

### 3.3 Model Identification

This section shows how we determine the order of the seasonal ARIMA model. We computed all relevant criteria to determine appropriate SARIMA model of Demand of Electricity in Sulaimani city. Those are the ACF and PACF in addition to RMSE, MAE, MAPE, and AIC criteria. To take a decision must be scanning all the plots of their coefficients of the series as shown in the figure (3.5 and 3.6), it is obvious from the AC, PAC in a time series data that it is requiring to measure the changes in seasonal during identifying, estimating. The following models have been examined and estimated as shown in table (3.3) below. The adequate seasonal model is chosen based on AIC criteria if it shows the minimum value as it is shown in table (3.3).

**Table 3.3:** SARIMA Models Criteria for the monthly demand of electric in Sulaimani governorate

Model	AIC	RMSE	MAPE
<b>ARIMA(0,0,0)x(2,1,0)<sup>4</sup></b>	<b>4.53022</b>	<b>9.3176</b>	<b>3.5580</b>
ARIMA(1,0,0)x(2,1,0) <sup>4</sup>	4.57962	9.4005	3.5730
ARIMA(0,0,1)x(2,1,0) <sup>4</sup>	4.57974	9.4007	3.5801
ARIMA(0,0,0)x(2,1,1) <sup>4</sup>	4.58149	9.4061	3.5935
ARIMA(1,0,0)x(2,1,1) <sup>4</sup>	4.63101	9.4915	3.5947

It is shown in table (3.3) that the SARIMA(0,0,0)x(2,1,0)<sup>4</sup> model produced the value of AIC with the smallest values. This means that the SARIMA (0,0,0)x(2,1,0)<sup>4</sup> model is the best among all the other models, which is the most suitable model that can be obtained for the monthly demand of electricity in Sulaimani city.

**3.4 Parameters Estimation:**

The former section indicate that the SARIMA (0,0,0)x(2,1,0)<sup>4</sup> model is the sufficient model with the minimum amount of RMSE, MAPE and AIC criteria ,also used maximum likelihood estimation for estimate the parameter because it is the best and more suitable method for estimation The results of the parameters prediction of the model are shown in table (3.4).

**Table3.4:** Parameter Estimates of SARIMA (0,0,0)x(2,1,0)<sup>4</sup> Model with their coefficients

Parameter	Estimate	Std. Error	t	P-value



SAR (1)	-1.01421	0.0393474	-25.7758	0.000000
SAR(2)	-0.973588	0.0281807	-34.548	0.000000

It is shown in table (3.4) that the p-value for the parameters SAR(1) and SAR(2) coefficients are less than  $\alpha = 0.05$ . This indicates that these coefficients are significantly different from zero, as it is shown for this model, the AIC criteria are the smallest values among the other models. Thus, the final model is SARIMA (0,0,0)x(2,1,0)<sup>4</sup>.

### 3.5 Comparison

Depending on the residuals of the SARIMA (0,0,0)x(2,1,0)<sup>4</sup> model we computed the Fourier SARIMA (0,0,0)x(2,1,0)<sup>4</sup> model and then comparing them to select the best model between them which has minimum AIC to forecast the demand of electric in Sulaimani city for the comes period of time, the comparative it is shown in table (3.5)

**Table 3.5:** shows the comparison between SARIMA and FSARIMA models

Model	AIC	RMSE	MAPE
ARIMA(0,0,0)x(2,1,0) <sup>4</sup>	4.53022	9.3176	3.5580
<b>FARIMA(0,0,0)x(2,1,0)<sup>4</sup></b>	<b>0.28314</b>	<b>2.5823</b>	<b>1.256</b>

Then the appropriate model to represent the monthly demand of electricity in Sulaimani city is FARIMA(0,0,0)x(2,1,0)<sup>4</sup> with minimum AIC.

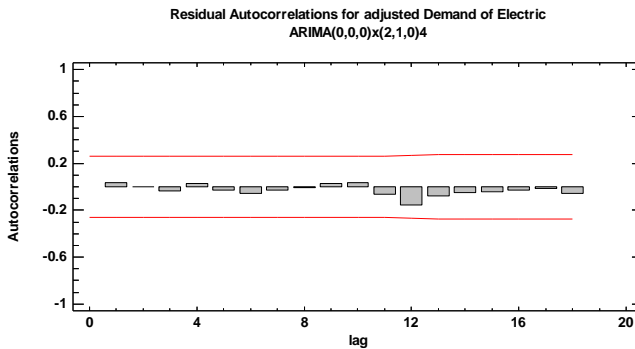
therefore, the mentioned model is used to forecast the monthly demand of electricity in Sulaimani city as it is shown below

the residuals of selected model have been tested to figure out if it is statistically significant or not, the below table represents the result of the test.

**Table 3.6:** ADF test of the residuals of FSARIMA(0,0,0)x(2,1,0)<sup>4</sup> model

Test	t-Statistic	P-value
ADF	-9.0070	0.000

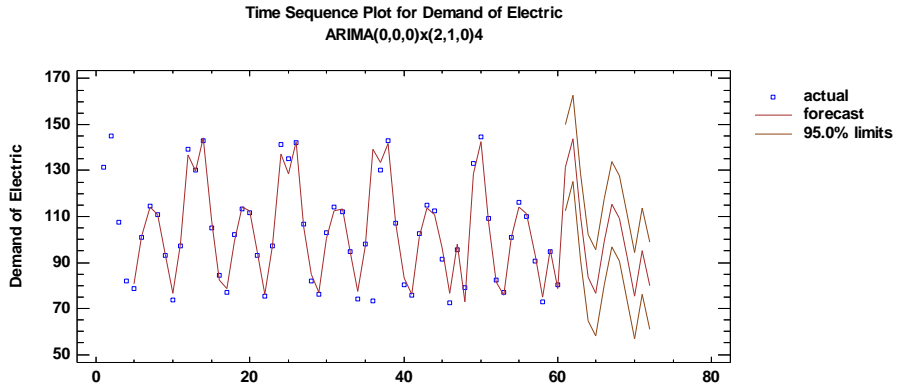
From the above table it is clear that the p-value of ADF test is less than 0.05 that implies the residuals is random in another word the FSARIMA(0,0,0)x(2,1,0)<sup>4</sup> model is statistically significant.



**Figure 3.6:** Autocorrelation Function for residuals of FSARIMA (0,0,0)x(2,1,0)<sup>4</sup> model

### 3.6 Forecasting for weight of imported equipment:

After getting the final model FSARIMA (0,0,0)x(2,1,0)<sup>4</sup> of the data of the monthly demand of electricity in Sulaimani city that has been expressed above. We forecasted the monthly demand of electricity in Sulaimani city in 2020 for 12 months. The forecasting of time series for monthly demand of electricity in Sulaimani city have been plotted as in figure (3.8).



**Figure 3.7:** Plot of the data and the forecasts with 95% confidence interval are represented.

Figure 3.7 shows the result that the behavior of forecasted values is the same as original series of monthly demand of electricity in Sulaimani city.

**Table 3.7:** Represents the monthly demand of electricity and the forecasted values depending on FSARIMA (0,0,0)x(2,1,0)<sup>4</sup>Model.

Year	Month	Demand	Forecasted
2015	1	131.3	
	2	145.1	
	3	107.4	
	4	82.1	
	5	78.8	79.269
	6	100.9	101.000

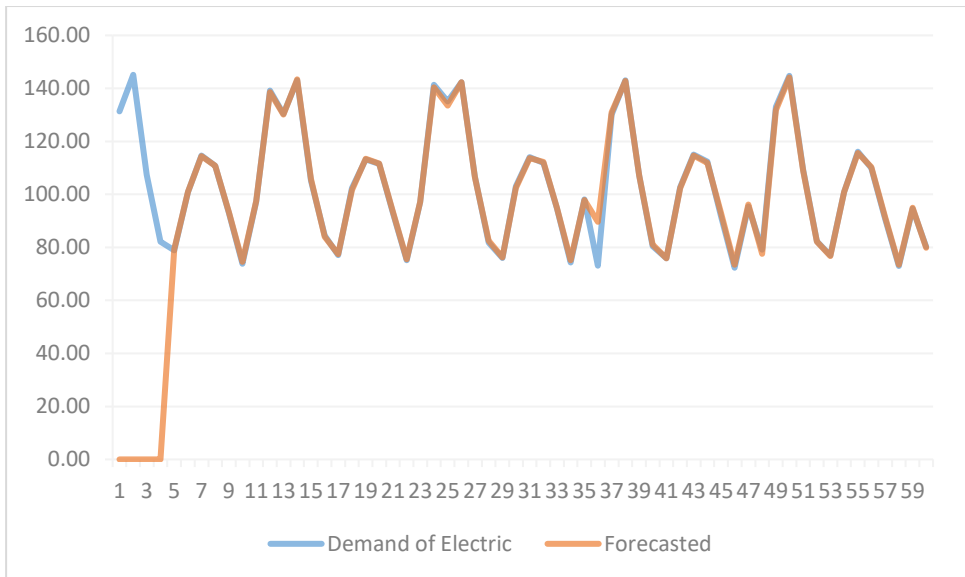
Year	Month	Demand	Forecasted
2018	1	130.1	130.928
	2	143.1	142.748
	3	107.1	107.128
	4	80.5	81.116
	5	75.9	75.932
	6	102.6	102.311

	7	114.6	114.462
	8	110.8	110.836
	9	93.0	93.137
	10	73.9	74.528
	11	97.3	97.497
	12	139.2	138.619
2016	1	130.3	130.165
	2	143.1	143.394
	3	105.2	105.851
	4	84.4	83.924
	5	77.1	77.461
	6	102.4	101.571
	7	113.2	113.428
	8	111.5	111.684
	9	93.0	93.428
	10	75.2	75.472
	11	97.1	97.149
	12	141.4	140.373
2017	1	135.1	133.510
	2	142.4	142.360

	7	115.0	114.652
	8	112.3	111.921
	9	91.6	92.813
	10	72.3	73.420
	11	95.4	96.137
	12	79.1	77.555
2019	1	133.0	131.871
	2	144.7	144.124
	3	109.4	108.940
	4	82.3	82.166
	5	77.0	76.666
	6	100.9	100.863
	7	116.1	115.681
	8	110.0	110.347
	9	90.8	91.460
	10	73.0	73.463
	11	94.7	94.923
	12	80.2	79.843
2020	1		130.930
	2		144.090

	3	106.7	106.435
	4	81.8	82.500
	5	76.0	76.241
	6	103.1	102.467
	7	114.1	113.739
	8	111.9	112.184
	9	94.9	94.913
	10	74.3	75.130
	11	98.1	97.877
	12	73.2	89.707

	3		109.865
	4		83.122
	5		76.753
	6		99.300
	7		115.541
	8		109.361
	9		92.014
	10		74.959
	11		94.635
	12		80.688



**Figure 3.8:** Scatter plot of the data and the forecasts with 95% confidence interval are represented

The results of the forecasted values in table (3.8) for the year 2020 are all fallen into the boundaries of the 95% confidence intervals. This confirms that the forecasting is very efficient.

**Table 3.8:** Forecast future value with 95% confidence interval

Period	Forecast	Lower Limit 95.0%	Upper Limit 95.0%
61.0	130.930	112.583	150.065
62.0	144.090	125.179	162.66
63.0	109.865	91.123	128.605
64.0	83.122	64.7145	102.196
65.0	76.753	58.0265	95.5118
66.0	99.300	80.3931	117.878
67.0	115.541	96.614	134.099
68.0	109.361	90.4236	127.909
69.0	92.014	73.8874	111.404
70.0	74.959	56.7513	94.2678
71.0	94.635	76.2252	113.742
72.0	80.688	61.1526	98.6691

Table 3.8 shows that the quantities of monthly demand of electricity in Sulaimani city for the 12 months of 2020 have been forecasted. It is also shown from these results that the forecasted values are all between the upper and lower boundaries of the 95% confidence intervals. This supports that the forecasting is efficient.

#### **4. Conclusion and Recommendations**

##### **4.1 Conclusion**

From the results, the following conclusions can be summarized:

1. The origin series of monthly demand of electricity is not random, the randomness achieved after taking first seasonal differencing.
2. A seasonal effect with repeating itself every four periods is exist in the data series.
3. The best postulated model that represents the data is FSARIMA  $(0,0,0) \times (2,1,0)^4$  model with minimum AIC between the other candidate models which is equal to 0.28.
4. Based on FSARIMA  $(0,0,0) \times (2,1,0)^4$ , the monthly demand of electricity in Sulaimani for the year 2020 for 12 months have been forecasted. The forecasted values are presented that the heist demand of electricity will be in January, February and July which are (130.930, 144.090 and 115.541) respectively and the forecasted values are fallen in boundaries of the 95% confidence intervals.

##### **REFERENCES:**

1. A.D, j., d. F., & tucker, m. (2007). new ARIMA models for seasonal time series and their application to second adjustment and forecasting.
2. agyemang, b. (2012). *autoregrissive moving average (ARIMA) intervention analysis model for the major crimes in chana*. math department kwame nkrumah university.
3. akinmutmi, c. (1996-2011). a sesonal arima modelling of mathematics.
4. alptekin, k. m. (2007). *building cost index forecasting with time series analysis*. turkey.
5. chatfield, C. (1996). the analysis of time series.



6. cochrane, J. (2005). *time series for macroeconomics and finance*. chicago: university of chicago.
7. cryer, J. D. (2008). *time series analysis* (2 ed.). new york.
8. dtreg. (2015, 3). *time series analysis*.
9. eni, d., & adeyeye, f. J. (2015). seasonal ARIMA modeling and forecasting of rainfall in warri town nigeria. *geoscience and environment protection* .
10. etuk, e. H. (2014, 4 18). an additive seasonal box-jenks model for nigerian monthly savings deposit rates.
11. Reinsel, George. C. (1994). *time series analysis forecasting and control*. (3, Ed.)
12. gyasi-agyei, k. (2012). *analysis and modeling of prevalence of measles in the ashti region of chana*. west africa : university of science and technology .
13. hamilton, J. (1994). *time series analysis* . amirica united state : princeton university.
14. hudu, m. (2009). *interrupted time series analysis of the rate of inflation in chana*. west africa.
15. hurvich, C., & C.L, T. (1989). *regression and time series model selection in small sample* .
16. mynsbrugge, j. V. (2010). *bidding strategies using orice based unit commitment in a deregulated power market* .
17. permanasari, A. e., hidayat, i., & alfi bustoni, i. (2013). SARIMA seasonal ARIMA implementation on time series to forcast the number of malaria incidence. *information technology and electrical engineering*.
18. petrevska, b., & delcev, g. (2016). predicting tourism demand by A.R.I.M.A models.
19. sariaslan, n. (2010). *the effect of temporal aggregation on univariate time series analysis* . turkey.
20. shumway, R. S. (2006). *time series analysis and its applection*. (2, Ed.) new york.
21. singh, e. h. (2013). forecasting tourist inflow in bhutan using seasonal ARIMA. *international journal of science and resarch (IJSR)*, 2 (9).
22. spyros, M. w., & R.J, h. (1998). *forecasting model and application*. (3, Ed.)
23. Wei. William. (1989). *time series analysis univariate and multivariate method* . university of tempie.



پوخته.

بهره‌می کارها یه‌کنکه له هه‌ره گرن‌گترین پیداو‌یستیه‌کانی ژیان وه‌رولنیکی گرن‌گی هه‌یه له پین‌ش‌خستنی زوربه‌ی بواره‌کانی ژیاندا وه‌ک لایه‌نی نابوری پین‌شه‌سازی و زور لایه‌نی که‌ی په‌پوهندی دار به ژیان‌ی رورژانه‌ی هاو‌لاتیه‌وه . له‌م لیکولینه‌وه‌یه‌دا ریژه‌ی خواست له‌سه‌ر به‌کار هینانی بهره‌می کارها به‌ش‌یوه‌ی مان‌گانه له‌پاریزگای سلیمانی وه‌رگیراوه . وه نامان‌جی سه‌ره‌کی له‌م لیکولینه‌وه‌یه بریتیه له دیاریکردنی مودیلکی گون‌جاو بو پین‌ش‌بینی ریژه‌ی خواست له‌سه‌ر کارها له پاریزگای سلیمانی بو ماوه‌ی 12 مان‌گ له سالی 2020 وه ده‌رنه‌ن‌جام و لیکولینه‌وه‌کان ده‌ریان خست که  $FSARIMA(0,0,0) \times (2,1,0)^4$  بریتیه له مودیله گون‌جاوه‌که گرن‌گی نه‌م پین‌ش بینی کردنه‌ش نه‌وه‌یه که ریژه‌ی نه‌کایکی زور نزمه به‌راورد به مودیله‌کانی که وه به‌سه‌ساته به 0.28.

## مقارنه بین نموج ARIMA و نموج FARIMA للتنبؤ بالطلب على الكهرباء في

### محافظة السليمانية

#### المخلص:

تعتبر الطاقة الكهربائية أحد أهم الطاقات المستهلكة في حياة الانسان، كما أن لها دورًا كبيرًا في تقدّم وتطور العديد من القطاعات منها القطاع الزراعي و الصناعي وقطاعات اخرى ذات العلاق، تم استخدام الطلب الشهري على الطاقة الكهربائية من قبل المستخدم في محافظة السليمانية حيث ان الهدف الرئيسي من هذا الدراسة هو اختيار النموذج المناسب للتنبؤ به لعام 2020، بعد اجراء التحليل أظهرت الدراسة ان نموذج هو  $FSARIMA(0,0,0) \times (2,1,0)^4$  هو النموذج المناسب والافضل حيث ان النموذج المختار قد تخطى كل الاختبارات الاحصائية اللازمة لاختيار النموذج المثالي للتنبؤ بالبيانات و ان AIC لهذا النموذج تساوي 0.28 الأقل بين نماذج المقارنة.