



Future estimation for Electricity interruption in Dohuk governorate (Kurdistan-Iraq) using SARIMA model

ID No. 2196

(PP 201 - 215)

<https://doi.org/10.21271/zjhs.23.1.14>

Nabeel George Nacy
Salahaddin University-
Dep. Of statistics

Mohammed AbdulMajeed Badal
Salahaddin University- Dep. Of
statistics
sinjary2013@gmail.com

Samaher Tareq Ibrahim
Dohuk University- Dep. Of
statistics

Received: 04/04/2018**Accepted: 23/10/2018****Published: 15/02/2019**

Abstract

Seasonality, in time series, refers to a regular pattern of changes that repeat for S time period, where S defines the numbers of timer periods until the pattern repeats again. Seasonality, of course, usually causes the time series to be non stationary [4], seasonal differencing is defined as a difference between a value and a value with lag that is a multiple of S .

In this paper, we will discuss the daily average of hours of electricity interruption per month which represents the gap between the amount of energy available and energy required for consumption. We took the data of Dohuk governorate (within Kurdistan region of Iraq) for the period from Jan. 2010 to Dec. 2016. Since this series have seasonal affects, we used SARIMA model with $S = 12$ and the seasonal and non-seasonal differences are 0 or 1, to choose the appropriate model, that gives least value of BIC , $RMSE$, MAE , $MAPE$ and largest value of R^2 to forecast the periods of daily interruption for the next months.

We concluded that the best model is $SARIMA(1,1,1)(0,1,0)12$, and the expected values of the daily average of hours of electricity interruption per month in Dohuk Governorate are increasing and it is expected that the entirely lack of electricity supply during the month of December 2018 if the time series continues this pattern.

Key words: Time series, SARIMA model, BIC , $RMSE$, MAE , $MAPE$, R^2 , ACF , $PACF$, Box-Ljung statistics, electricity interruption.

1. Introduction:

A *time series* is a series of data points indexed (or listed or graphed) in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time[3]. Thus, it is a sequence of discrete-time data. Time series are very frequently plotted via line charts. Time series are used in statistics, signal processing, econometrics, mathematical finance, weather forecasting, intelligent transport and trajectory forecasting, earthquake prediction, electroencephalography, control engineering, astronomy, communications engineering, and largely in any domain of applied science and engineering which involves temporal measurements[5].

In statistics, prediction is a part of statistical inference. One approach to such inference is known as predictive inference, but the prediction can be undertaken within any of the several approaches to statistical inference. Indeed, one description of statistics is that it provides a means of transferring knowledge about a sample of a population to the whole population, and to other related populations, which is not necessarily the same as prediction over time. When information is transferred across time, often to specific points in time, the process is known as forecasting. Forecasting on time series is usually done using automated statistical software packages and programming languages, such as *R*, *S*, *SAS*, *SPSS*, *Minitab*, *Pandas* (Python) and others.

Time series analysis comprises methods for analyzing time series data to extract meaningful statistics and other characteristics of the data. *Time series forecasting* is the use of a model to



predict future values based on previously observed values. While regression analysis is often employed in such a way as to test theories that the current values of one or more independent time series affect the current value of another time series, this type of analysis of time series is not called "time series analysis", which focuses on comparing values of a single time series or multiple dependent time series at different points in time. Time series data have a natural temporal ordering. This makes time series analysis distinct from cross-sectional studies, in which there is no natural ordering of the observations[2].

Time series analysis is also distinct from analysis where the observations typically relate to geographical locations.

A stochastic model for a time series will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values.

Methods for time series analysis may be divided into two classes: frequency-domain methods and time-domain methods. The former include spectral analysis and wavelet analysis; the latter include auto-correlation and cross-correlation analysis. In the time domain, correlation and analysis can be made in a filter-like manner using scaled correlation, thereby mitigating the need to operate in the frequency domain.

Additionally, time series analysis techniques may be divided into parametric and non-parametric methods. The parametric approaches assume that the underlying stationary stochastic process has a certain structure which can be described using a small number of parameters (for example, using an autoregressive or moving average model). In these approaches, the task is to estimate the parameters of the model that describes the stochastic process. By contrast, approaches explicitly estimate the covariance or the spectrum of the process without assuming that the process has any particular structure.

Methods of time series analysis may also be divided into linear and non-linear, and univariate and multivariate, the aim to project concluded that the best model is SARIMA and the expected values of the daily average of hours of electricity interruption per month

2. Theoretical aspect:

2-1- Autoregressive integrated Moving Average (ARIMA):

The process $\{Y_t; t \in Z\}$ is an autoregressive moving average (ARMA) process of order (p, q) , denoted with $Y_t \sim ARMA(p, q)$, if :

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + u_t - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q} \quad \dots (1)$$

where $u_t \sim WN(0, \sigma^2)$, and $\phi_0, \phi_1, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$ are $(p+q+1)$ constants and the polynomials $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ and

$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ have no common factors.

In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. Both of these models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). ARIMA models are applied in some cases where data show evidence of non-stationary, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied one or more times to eliminate the non-stationary.

The AR part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged (i.e., prior) values. The MA part indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past. The I (for "integrated") indicates that the data values have been replaced with the difference between their values and the previous values (and this differencing

process may have been performed more than once). The purpose of each of these features is to make the model fit the data as well as possible.

Non-seasonal ARIMA models are generally denoted $ARIMA(p,d,q)$ where parameters p , d , and q are non-negative integers, p is the order (number of time lags) of the autoregressive model, d is the degree of differencing (the number of times the data have had past values subtracted), and q is the order of the moving-average model.

When two out of the three terms are zeros, the model may be referred to base on the non-zero parameter, dropping "AR", "I" or "MA" from the acronym describing the model. For example, $ARIMA(1,0,0)$ is $AR(1)$, $ARIMA(0,1,0)$ is $I(1)$, and $ARIMA(0,0,1)$ is $MA(1)$.

2-2- Seasonal ARIMA Model:

The seasonal *ARIMA* model incorporates both non-seasonal and seasonal factors in a multiplicative model. One shorthand notation for the model is:

$$SARIMA(p, d, q) \times (P, D, Q)S,$$

with p = non-seasonal AR order, d = non-seasonal differencing, q = non-seasonal MA order, P = seasonal AR order, D = seasonal differencing, Q = seasonal MA order, and S = time span of repeating seasonal pattern[11].

Without differencing operations, the model could be written more formally as:

$$\Phi(B^s)\phi(B)(X_t - \mu) = \Theta(B^s)\theta(B)u_t \quad \dots(2)$$

The non-seasonal components are:

$$AR: \phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad \dots(3)$$

$$MA: \theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q \quad \dots(4)$$

The seasonal components are:

$$Seasonal AR: \Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps} \quad \dots(5)$$

$$Seasonal MA: \Theta(B^s) = 1 + \Theta_1 B^s + \dots + \Theta_Q B^{Qs} \quad \dots(6)$$

Where B is operating on Y_t , has the effect of shifting the data back one period.

$$BY_t = Y_{t-1} \quad \dots(7)$$

Two applications of B to Y_t shifts the data back two periods:

$$B(BY_t) = B^2 Y_t = Y_{t-2} \quad \dots \quad \text{and so on}$$

Identifying a Seasonal Model:

Step 1: Do a time series plot of the data. Examine it for features such as trend and seasonality. You'll know that you've gathered seasonal data (months, quarters, etc.) so look at the pattern across those time units (months, etc.) to see if there is indeed a seasonal pattern[1].

Step 2: Do any necessary differencing. The general guidelines are:

- If there is seasonality and no trend take a difference of lag S . For instance, take a 12^{th} difference for monthly data with seasonality.
- If there is linear trend and no obvious seasonality, take a first difference. If there is a curved trend, consider a transformation of the data before differencing.

- If there is both trend and seasonality, apply both a non-seasonal and seasonal difference to the data, as two successive operations (in either order).
- If there is neither obvious trend nor seasonality, don't take any differences.

Step 3: Examine the *ACF* and *PACF* of the differenced data (if differencing is necessary).

We're using this information to determine possible models. This can be tricky going involving some (educated) guessing. Some basic guidance:

Non-seasonal terms: Examine the early lags (1, 2, 3, ...) to judge non-seasonal terms. Spikes in the *ACF* (at low lags) indicate non-seasonal *MA* terms. Spikes in the *PACF* (at low lags) indicated possible non-seasonal *AR* terms.

Seasonal terms: Examine the patterns across lags that are multiples of *S*. For example, for monthly data, look at lags 12, 24, 36, and so on (probably won't need to look at much more than the first two or three seasonal multiples). Judge the *ACF* and *PACF* at the seasonal lags in the same way you do for the earlier lags.

Step 4: Estimate the model(s) that might be reasonable based on Step 3. Don't forget to include any differencing that you did before looking at the *ACF* and *PACF*.

Step 5: Examine the residuals (with *ACF*, Box-Pierce, and any other means) to see if the model seems good. Compare *AIC* or *BIC* values if you tried several models.

If things don't look good here, it's back to Step 3 (or maybe even Step 2).

2-3- Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF):

An important guide to the persistence in a time series is given by the series of quantities called the sample autocorrelation coefficients, which measure the correlation between observations at different times. The set of autocorrelation coefficients arranged as a function of separation in time is the sample autocorrelation function (r_k), or the *ACF*.

$$r_k = \frac{c_k}{c_0} = \frac{\sum_{t=1}^{N-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^N (Y_t - \bar{Y})^2} \quad \dots (8)$$

Where $c_k = \frac{1}{N} \sum_{t=1}^{N-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})$; $k = 0, 1, 2, \dots, K \leq \frac{N}{4}$ is the auto covariance;

$\bar{Y}$ is the mean of the time series and *N* is the number of the observations.

The partial autocorrelation coefficients $\hat{\varphi}_k$ are calculated as follows:

$$\hat{\varphi}_1 = r_1 \quad ; \quad \hat{\varphi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2} \quad ; \quad \dots \dots etc.$$

$$\hat{\varphi}_k = \frac{\begin{vmatrix} 1 & r_1 & r_2 & \dots & r_{k-2} & r_{k-1} \\ r_1 & 1 & r_1 & \dots & r_{k-3} & r_2 \\ r_2 & r_1 & 1 & \dots & r_{k-4} & r_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{k-1} & r_{k-2} & r_{k-3} & \dots & r_1 & r_k \end{vmatrix}}{\begin{vmatrix} 1 & r_1 & r_2 & \dots & r_{k-2} & r_{k-1} \\ r_1 & 1 & r_1 & \dots & r_{k-3} & r_2 \\ r_2 & r_1 & 1 & \dots & r_{k-4} & r_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{k-1} & r_{k-2} & r_{k-3} & \dots & r_1 & 1 \end{vmatrix}} \quad \dots (9)$$

The collection $\{\hat{\varphi}_k\}$ is called the sample partial autocorrelation function (*SPACF*).

2-4- Stationary:

A stationary time series is one whose properties do not depend on the time at which the series is observed[9].

- $E[y_t] = \mu$, for all $t \Rightarrow$ Stationary around mean.
- $Cov[y_t, y_{t-k}] = \gamma_k$, for all $t \Rightarrow$ Stationary around variance.
- $Var[y_t] = \gamma_0$, for all t

Any series that are not stationary are said to be *non stationary*.

So, time series with trends, or with seasonality, are not stationary — the trend and seasonality will affect the value of the time series at different times. On the other hand, a



white noise series is stationary — it does not matter when you observe it, it should look much the same at any period.

Some cases can be confusing a time series with cyclic behavior (but not trend or seasonality) is stationary. That is because the cycles are not of fixed length, so before we observe the series we cannot be sure where the peaks and troughs of the cycles will be.

A stationary time series will have no predictable patterns in the long-term. Time plots will show the series to be roughly horizontal (although some cyclic behavior is possible) with constant variance.

Stationarity implies that the covariance and correlation of observations taken k periods apart are only a function of the lag k , but not of the time point t .

A time series $\{Y_t\}$ is called strictly stationary if its *p.d.f.* does not depend on time t that means the random vectors:

$$(Y_{t_1}, Y_{t_2}, \dots, Y_{t_n})^T \text{ and } (Y_{t_1+\tau}, Y_{t_2+\tau}, \dots, Y_{t_n+\tau})^T$$

have the same joint distribution for all sets of indices $\{t_1, \dots, t_n\}$ and for all integers τ and $n > 0$.

A time series $\{Y_t\}$ is called weakly stationary if:

$E(Y_t)$ does not depend on t .

$\text{Cov}(Y_t, Y_{t+s})$ depends on s and not t .

Weak stationary does not imply strict stationary.

To make a time series stationary — compute the differences between consecutive observations. This is known as differencing.

Transformations such as logarithms can help to stabilize the variance of a time series. Differencing can help stabilize the mean of a time series by removing changes in the level of a time series, and so eliminating trend and seasonality.

As well as looking at the time plot of the data, the *ACF* plot is also useful for identifying non-stationary time series. For a stationary time series, the *ACF* will drop to zero relatively quickly, while the *ACF* of non-stationary data decreases slowly. Also, for non-stationary data, the value of r_1 is often large and positive.

2-5- Fitting the model:

The selection of the model is important, as under-fitting a model may not capture the true nature of the variability in the outcome variable, while an over-fitted model loses generality. Akaike Information Criteria (*AIC*) is then a way to select the model that best balances these drawbacks. Once a best model is selected, traditional null-hypothesis testing can then be used on the best model to determine the relationship between specific variables and the outcome of interest:

$$AIC = 2K - 2\log(L(\hat{\theta}/y)) \quad \dots(10)$$

where K is the number of estimable parameters (degrees of freedom) and $\log(L(\hat{\theta}/y))$ is the log-likelihood at its maximum point of the model estimated. Further refined this estimate to correct for small data samples:

$$AICc = AIC + \frac{2K(K+1)}{n-K-1} \quad \dots(11)$$

where n is the sample size and K and *AIC* are defined above. If n is large with respect to K , this correction is negligible and *AIC* is sufficient. *AICc* is more general, however, and is generally used in place of *AIC*. The best model is then the model with the lowest *AICc* (or *AIC*) score. It is important to note that the *AIC* and *AICc* scores are ordinal and mean nothing on their own.

Bayesian Information Criteria (*BIC*) is an estimate of a function of the posterior probability of a model being true, under a certain Bayesian setup, so that a lower *BIC* means that a model is more likely to be the true model:

$$BIC = 2 \log n - 2\log(L(\hat{\theta}/y)) \quad \dots(12)$$



The Box-Ljung test is a diagnostic tool used to test the lack of fit of a time series model. The test is applied to the residuals of a time series after fitting an $ARMA(p,q)$ model to the data. The test examines m autocorrelations of the residuals. If the autocorrelations are very small, we conclude that the model does not exhibit significant lack of fit[10].

the Box-Ljung test is defined as:

H_0 : The model does not exhibit lack of fit (autocorrelations up to lag k equal zero, i.e. the data values are random and independent up to a certain number of lags (m)).

H_1 : The model exhibits lack of fit.

Given a time series Y of length n , the test statistic is defined as:

$$Q = n(n+2) \sum_{k=1}^m \frac{\hat{r}_k^2}{n-k} \quad \dots(13)$$

where $\hat{r}_k$ is the estimated autocorrelation of the series at lag k ,

m is the number of lags being tested.

The Box-Ljung test rejects the null hypothesis (indicating that the model has significant lack of fit) if: $Q > \chi^2_{1-\alpha, h}$ where $\chi^2_{1-\alpha, h}$ is the chi-square distribution table value with h degrees of freedom and significance level α .

Because the test is applied to residuals, the degrees of freedom must account for the estimated model parameters so that $h = m - p - q$.

Furthermore, for forecasting values, (where n is the number of forecasted errors):

Mean Square Error $MSE = \frac{\sum e_i^2}{n} \quad \dots(14)$, $e_i = y_i - \hat{y}_i$.

Root Mean Square Error $RMSE = \sqrt{MSE} \quad \dots(15)$

Mean Absolute Percentage Error $MAPE = \frac{1}{n} \sum \frac{|e_i|}{y} * 100\% \quad \dots(16)$

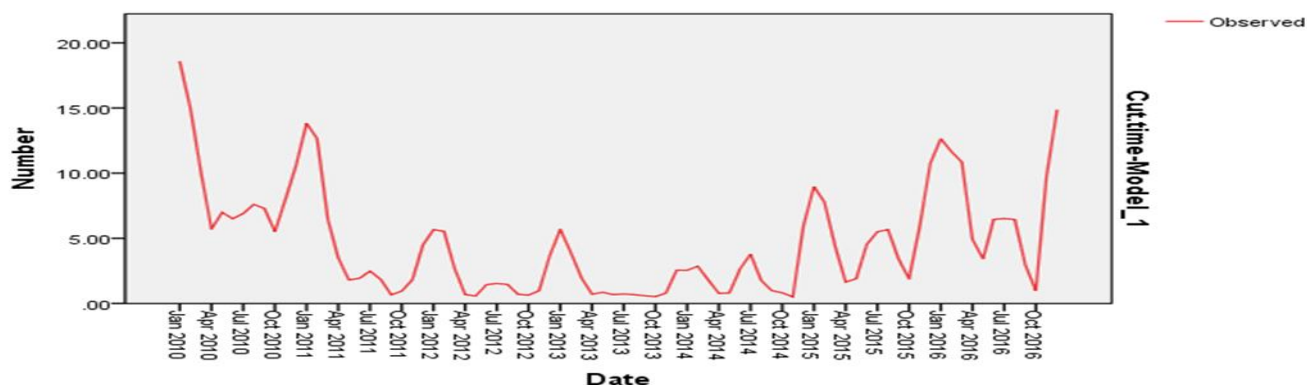
Mean Absolute Deviation $MAD = \sum \frac{|e_i|}{n} \quad \dots(17)$

3. Practical aspect:

In the last forty years ago, Iraq faces a range of economic problems due to the deterioration of infrastructure because of misguided policies. One of these problems is the electric power. To study this topic, we relied on the daily average of hours of electricity Interruption per month, which represents the gap between the amount of energy available and energy required for consumption. We take the data of Dohuk governorate (within Kurdistan region of Iraq) for the period from Jan. 2010 to Dec. 2016 (the table below), to forecasting the daily rate of electricity interruption periods for the future months, using the seasonal time series model (SARIMA):

Table(1):Daily average of hours of electricity Interruption per month for years (2010-2016) in Duhuk

Month	2010	2011	2012	2013	2014	2015	2016
Jan.	18.6	13.83	5.66	5.69	2.54	8.96	12.65
Feb.	15.0	12.63	5.54	3.90	2.86	7.76	11.66
Mar.	10.1	6.48	2.67	1.98	1.80	4.41	10.83
Apr.	5.7	3.47	0.68	0.71	0.77	1.64	4.90
May	7.0	1.81	0.57	0.86	0.81	1.90	3.42
June	6.5	1.93	1.44	0.68	2.65	4.53	6.45
July	6.9	2.48	1.53	0.73	3.79	5.50	6.52
Aug.	7.6	1.85	1.46	0.67	1.77	5.67	6.44
Sep.	7.3	0.66	0.70	0.59	0.99	3.42	2.95
Oct.	5.5	0.95	0.64	0.52	0.82	1.87	0.97
Nov.	8.0	1.81	0.98	0.80	0.50	5.90	9.73
Dec.	10.6	4.48	3.66	2.55	5.95	10.74	14.88



Figure(1): seasonal time series of Daily average of hours of electricity Interruption(2010-2016)

Before forecasting we must check the stationary for series has satisfied all assumption or not by using Phillips-perron test for unit root ,we say not stationary (Diagnostic checking for residually)

Table(2): Phillips-perron test for unit root for Stationary

	Test Statistic	1% critical value	5% critical value	10% critical value
Z(rho)	-5.907	-12.300	-7.460	-5.380
Z(t)	-2.866	-2.644	-1.950	-1.604

Then, the seasonal factors are:

Table(3): Seasonal factors % in each months from years (2010-2016)

Seasonal Decomposition

Seasonal Factors

Series Name: cut.time

Period Seasonal Factor

(%)

1 225.1

2 202.6

3 115.0

4 47.4

5 44.5

6 75.3

7 82.8

8 71.5

9 44.9

10 35.9

11 71.4

12 183.6

Seasonal Decomposition

Seasonal Factors

Series Name: cut.time

Period Seasonal Factor

(%)

1 225.1

2 202.6

3 115.0

4 47.4

5 44.5

6 75.3



7 82.8
8 71.5
9 44.9
10 35.9
11 71.4
12 183.6

Month	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
Seasonal factor %	225.1	202.6	115.0	47.4	44.5	75.3	82.8	71.5	44.9	35.9	71.4	183.6

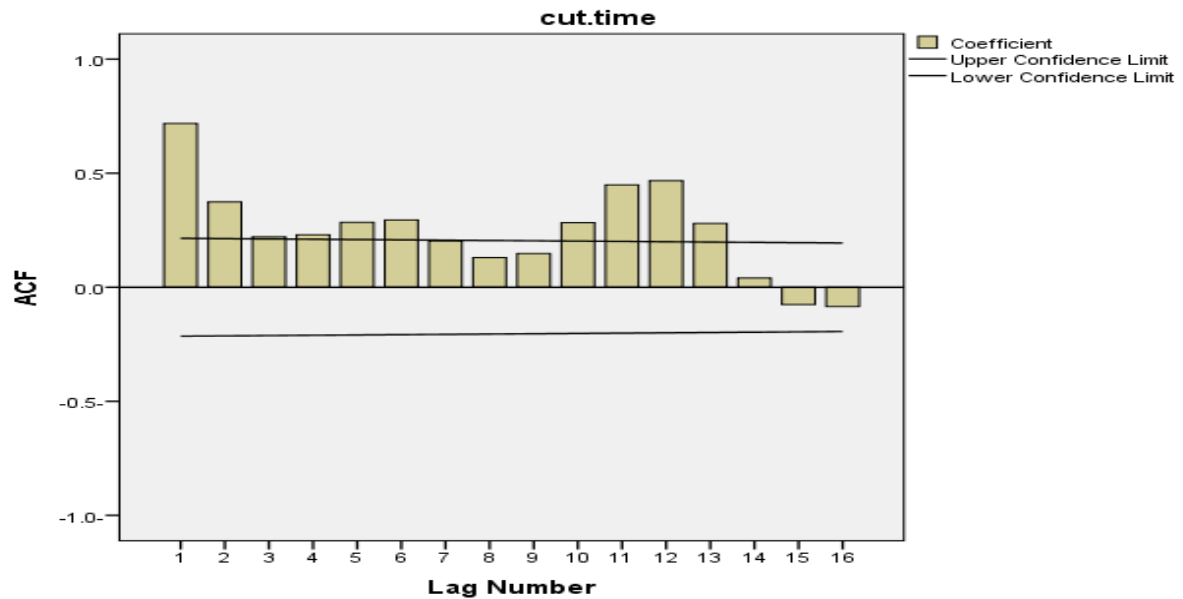
From the above table, we notice the seasonal fluctuations, which are high during the first three months of the year (at winter), decreasing in the following two months (at spring), then increasing in the summer months, fall in autumn (September and October) and rise again with winter (last two months of the year) and so on.

Table(4): model Parameters for original data of electricity Interruption in Duhuk

Lag	PAC	AC	Std. Error ^a	Box-Ljung Statistic		
				Value	df	Sig. ^b
1	.718	.718	.107	44.860	1	.000
2	-.292-	.374	.107	57.178	2	.000
3	.189	.221	.106	61.534	3	.000
4	.120	.230	.105	66.297	4	.000
5	.099	.284	.105	73.692	5	.000
6	.038	.295	.104	81.738	6	.000
7	-.097-	.203	.103	85.613	7	.000
8	.077	.130	.103	87.218	8	.000
9	.080	.148	.102	89.315	9	.000
10	.239	.283	.101	97.150	10	.000
11	.214	.448	.101	117.048	11	.000
12	-.010-	.467	.100	138.964	12	.000
13	-.187-	.280	.099	146.928	13	.000
14	-.199-	.040	.098	147.096	14	.000
15	-.106-	-.075-	.098	147.692	15	.000
16	-.165-	-.085-	.097	148.458	16	.000

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.



Figure(3):Autocorrelation Function(ACF)

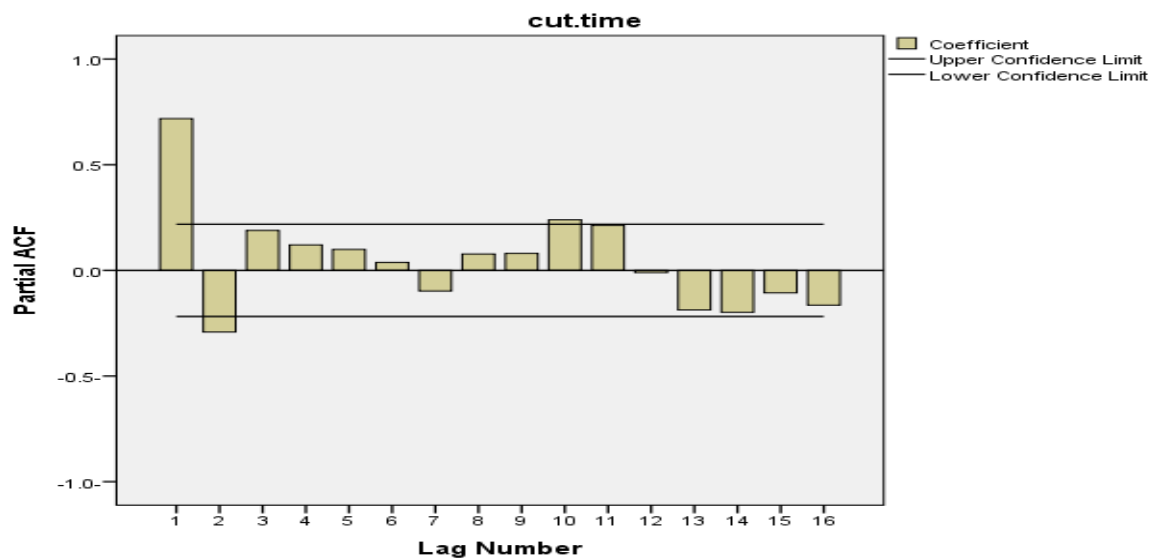


Figure (4): Partial Autocorrelation Function (PACF)

The “suspension bridge” pattern in the *ACF* is typical of a series that is both non stationary and seasonal. Clearly, we need at least one order of differencing. If we take a non seasonal difference ($SARIMA(0,1,0)(0,0,0)$ or $ARIMA(0,1,0)$), the corresponding plots are as follows:

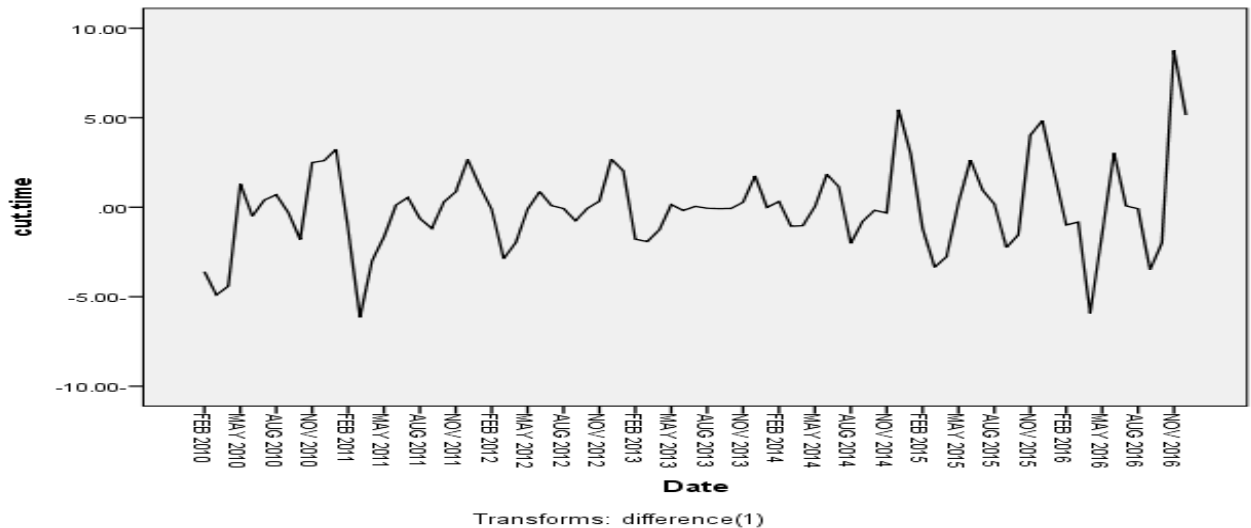


Figure (5): Forecasting SARIMA Model (0,1,0)(0,0,0)

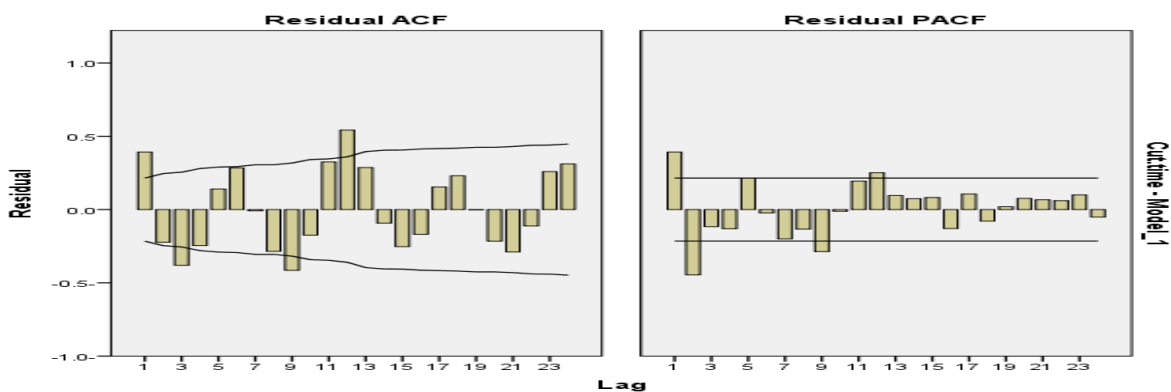


Figure (6): Residual (ACF) & Residual (PACF)

The differenced series (the residuals of a random-walk-with-growth model) looks more or less stationary, but there is still strong autocorrelation at the seasonal period ($lag\ 12$). Because the seasonal pattern is strong and stable, we will want to use an order of seasonal differencing in the model ($SARIMA(0,0,0)(0,1,0)12$), the corresponding plots are as follows:

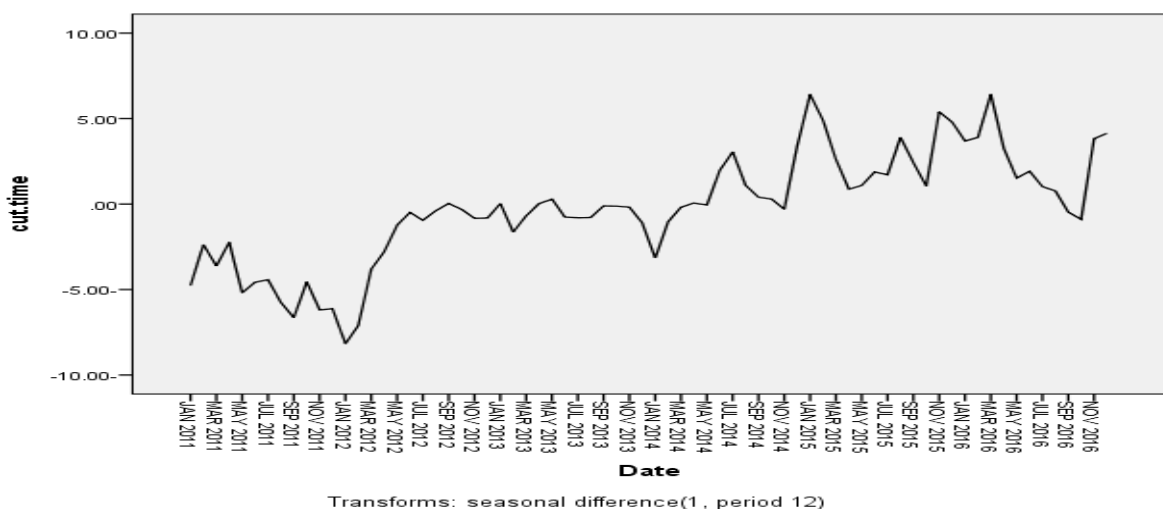


Figure (7): Forecasting SARIMA Model (0,0,0)(0,1,0)12

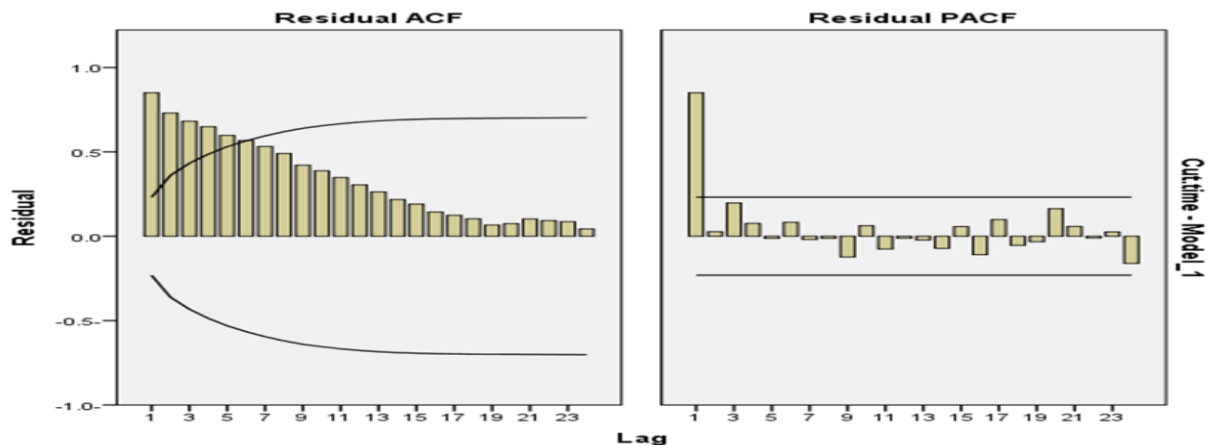


Figure (8): Residual (ACF) & Residual (PACF) for SARIMA Model(0,0,0)(0,1,0)12

The seasonally differenced series shows a very strong pattern of positive autocorrelation. This could be signal the need for another difference. If we take both a seasonal and non-seasonal difference ($SARIMA(0,1,0)(0,1,0)12$), the corresponding plots are as follows:

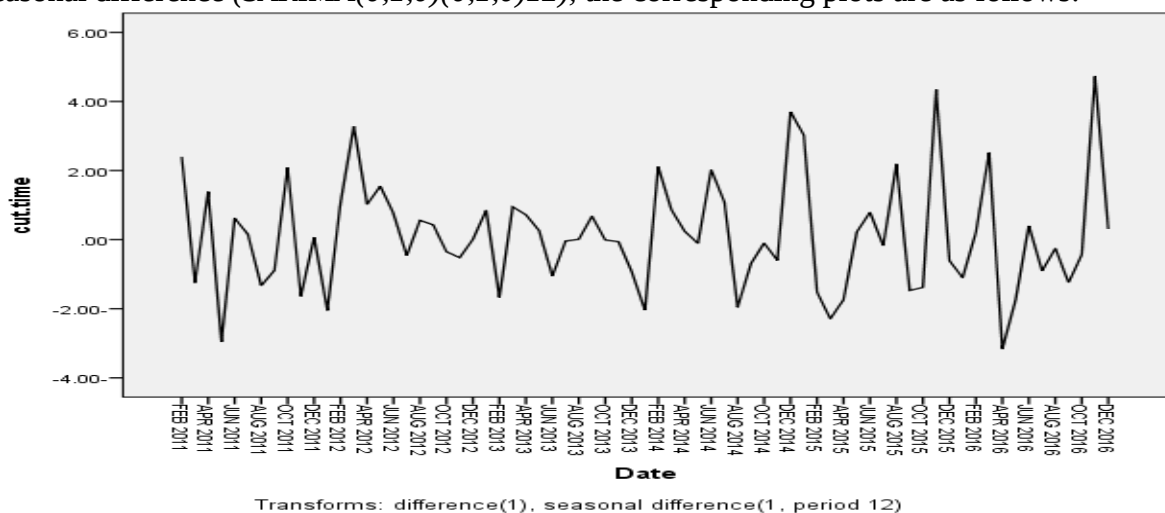


Figure (9): Forecasting SARIMA Model SARIMA (0,1,0)(0,1,0)12

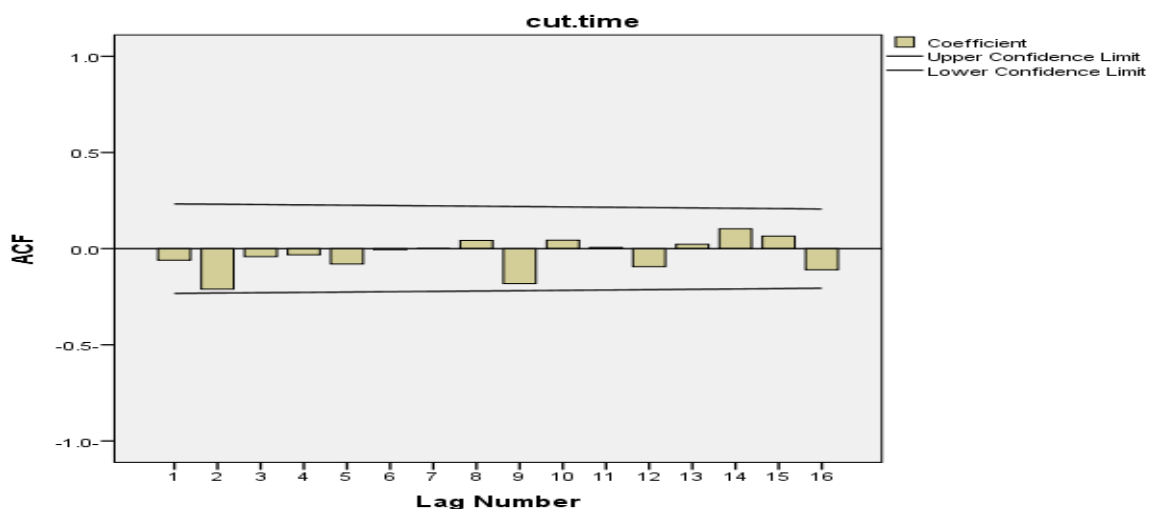


Figure (10): Autocorrelation Function (ACF) for SARIMA Model(0,1,0)(0,1,0)12

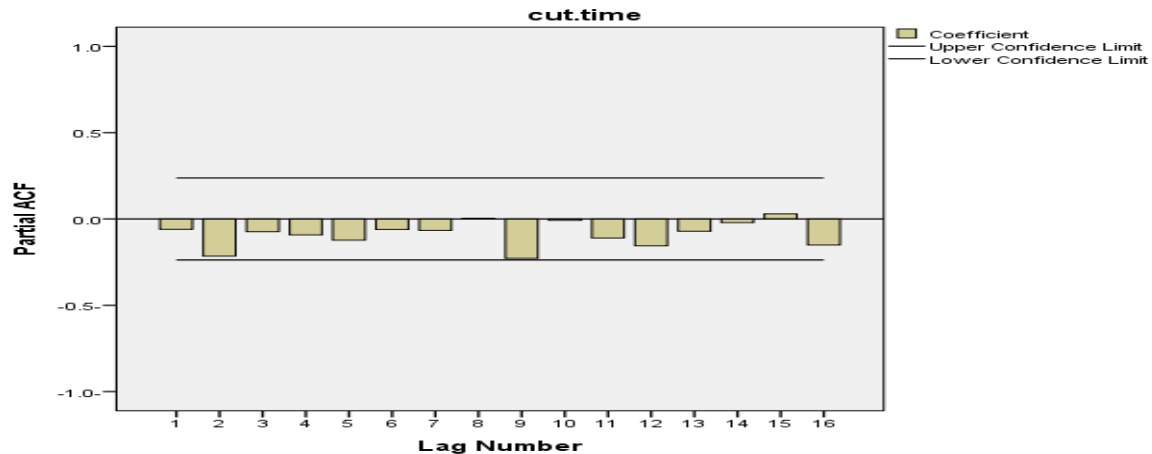


Figure (11): Partial Autocorrelation Function (PACF) for SARIMA Model(0,1,0)(0,1,0)12

The correct order of differencing is a calculation of the error statistics of the series at each level of differencing. We can compute these by fitting the corresponding *ARIMA* models in which only differencing is used:

Table(5) : Statistics of SARIMA Models

<i>SARIMA model</i>	R^2	<i>RMSE</i>	<i>MAE</i>	<i>MAPE</i>	<i>BIC</i>
<i>(0,1,0)(0,0,0)12</i>	.596	2.424	1.687	50.675	1.824
<i>(0,0,0)(0,1,0)12</i>	.201	3.243	2.410	91.004	2.412
<i>(0,1,0)(0,1,0)12</i>	.780	1.617	1.222	54.266	1.021

The smallest errors, in the models with the estimation period and validation period, are obtained by *SARIMA(0,1,0)(0,1,0)12*, which uses one difference of each type. This, together with the appearance of the plots above, strongly suggests that we should use both a seasonal and non-seasonal difference.

In the analysis that follows, we will try to improve these models through the addition of seasonal *SARIMA* terms:

Table (6) : Statistics of Seasonal SARIMA Models

No.	<i>SARIMA model</i>	R^2	<i>RMSE</i>	<i>MAE</i>	<i>MAPE</i>	<i>BIC</i>
1	<i>(0,1,0)(0,1,1)12</i>	0.783	1.620	1.206	54.276	1.084
2	<i>(0,1,0)(1,1,0)12</i>	0.783	1.619	1.204	54.235	1.084
3	<i>(0,1,0)(1,1,1)12</i>	0.783	1.630	1.204	54.502	1.157
4	<i>(0,1,1)(0,1,0)12</i>	0.782	1.624	1.227	56.510	1.089
5	<i>(0,1,1)(0,1,1)12</i>	0.784	1.628	1.216	56.239	1.154
6	<i>(0,1,1)(1,1,0)12</i>	0.784	1.627	1.215	56.258	1.154
7	<i>(0,1,1)(1,1,1)12</i>	0.785	1.635	1.214	57.088	1.224
8	<i>(1,1,0)(0,1,1)12</i>	0.783	1.629	1.212	55.354	1.156
9	<i>(1,1,0)(1,1,0)12</i>	0.783	1.629	1.210	55.362	1.156
10	<i>(1,1,0)(1,1,1)12</i>	0.784	1.638	1.207	56.038	1.227
11	<i>(1,1,1)(0,1,0)12</i>	0.810	1.526	1.162	55.625	1.025
12	<i>(1,1,1)(0,1,1)12</i>	0.812	1.529	1.160	56.587	1.089
13	<i>(1,1,1)(1,1,0)12</i>	0.812	1.531	1.159	56.132	1.092

From the two tables above, we conclude that the model (*SARIMA(1,1,1)(0,1,0)*) is the best, which gives us the lowest values for each of *RMSE*, and *BIC*, and approximately lowest



value for MAE and largest value for R^2 . So, we will rely on this model to estimate the predictions of the next months of the years 2017 and 2018:

The test of the parameters of the model is:

Table (7): Test of the parameters of predictions electricity Interruption for years 2017-2018

By the model $SARIMA(1,1,1)(0,1,0)$

Parameters	Estimate	S.E.	t	Sig.
Constant	0.126	0.035	3.588	0.001
AR – Lag 1	0.692	0.140	4.947	0.000
Difference	1			
MA – Lag 1	1.000	7.338	0.136	0.892
Seasonal Difference	1			

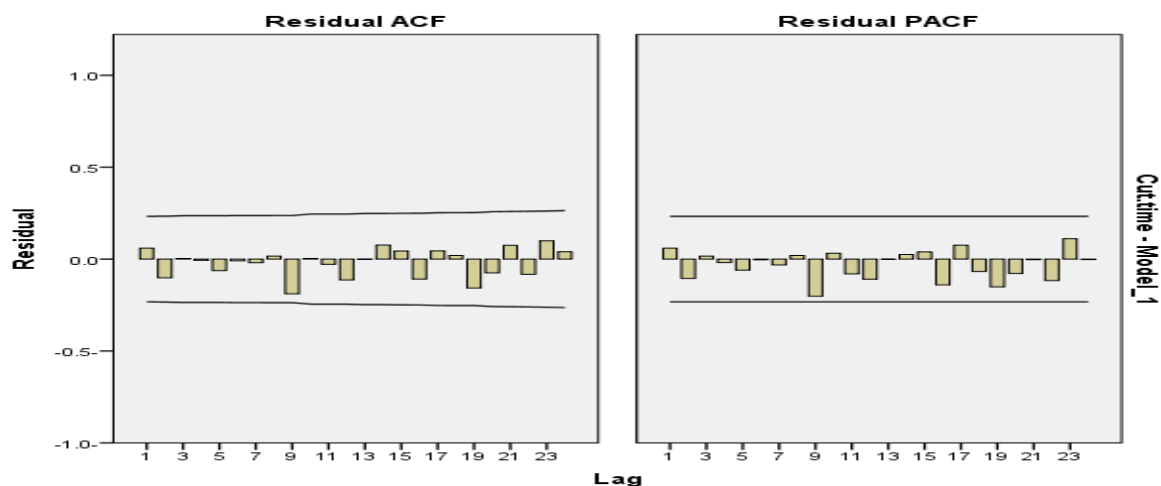


Figure (12): Residual of (ACF)& (PACF) for $SARIMA(1,1,1)(0,1,0)12$

Therefore, the forecasting values of the daily average of hours of electricity Interruption per month in Dohuk Governorate during the years 2017 and 2018, using the above model $SARIMA(1,1,1)(0,1,0)12$, will be as follows:

Table (8): Forecasting of electricity Interruption per month in Dohuk Governorate of $SARIMA(1,1,1)(0,1,0)12$ since 2017-2018

Year	months	forecast	LCL	UCL
2017	Jan.	16.94	13.94	19.95
	Feb.	16.10	12.42	19.78
	Mar.	15.41	11.43	19.39
	Apr.	9.61	5.50	13.73
	May	8.27	4.08	12.46
	Jun.	11.43	7.20	15.56
	Jul.	11.63	7.38	15.88
	Aug.	11.68	7.41	15.94
	Sep.	8.31	4.04	12.58
	Oct.	6.46	2.19	10.73
	Nov.	15.35	11.07	19.62
	Dec.	20.62	16.35	24.00
2018	Jan.	22.81	17.50	24.00
	Feb.	22.10	16.33	24.00
	Mar.	21.53	15.54	24.00
	Apr.	15.86	9.76	21.96
	May	14.64	8.48	20.80



June	17.93	11.73	24.00
July	18.25	12.04	24.00
Aug.	18.43	12.20	24.00
Sep.	15.19	8.96	21.42
Oct.	13.47	7.23	19.70
Nov.	22.48	16.24	24.00
Dec.	24.00	21.64	24.00

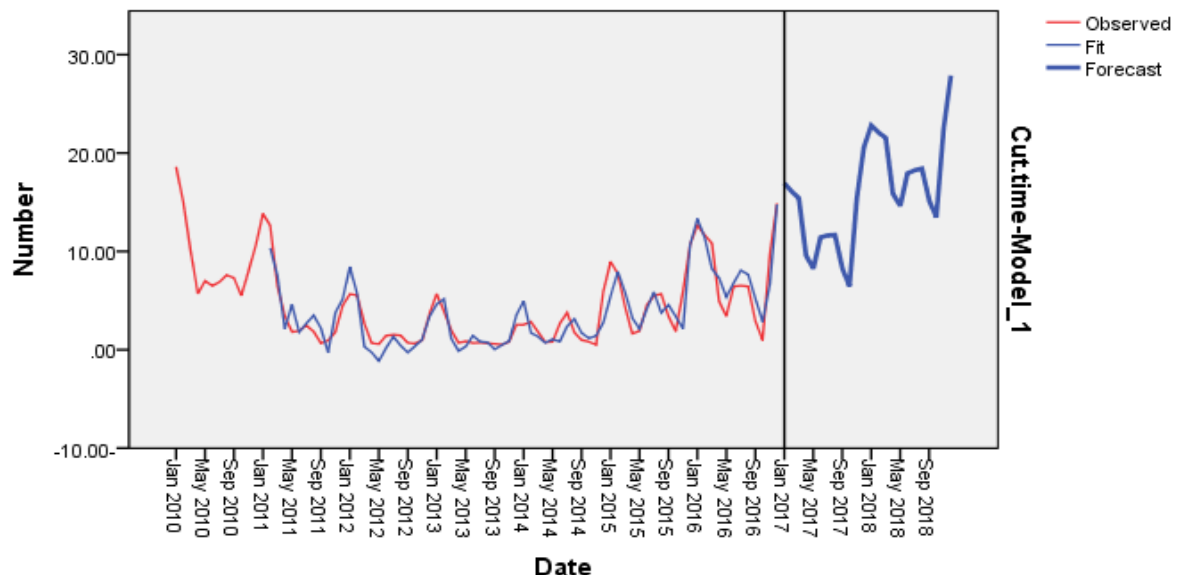


Figure (13): Fitting Model for predictions of electricity Interruption of SARIMA (1,1,1)(0,1,0)12 since 2017-2018

4. **Conclusions and Recommendation:**

- 1-We conclude that the expected values of the daily average of hours of electricity interruption per month in Dohuk Governorate are increasing and it is expected that the entirely lack of electricity supply during the month of December 2018 if the time series continues this pattern.
- 2- It should pay attention to increase the production capacity of electricity and educate citizens to rationalize the use of electricity to avoid this crisis.
- 3-Other alternatives can be used for electrical power, such as general or private generators that supply electricity to citizens at fairly high prices, which in turn affect environmental pollution and create high noise because they are used within residential complexes. Solar energy, which is environmentally friendly, is also available, but its use is very limited despite the availability of all the necessary supplies. It can also be used to provide electricity to some public sites such as street lighting.

5. **References:**

1. Adhistya, E.P., Indriana, H., Isna, A. (2013), "SARIMA (Seasonal ARIMA) Implementation on time series to forecast the number of Malaria incidence", Information Technology and Electrical Engineering, conference on Yogyakarta, Indonesia.
2. Akapanta, A.C., Okorie, I.E., Okoye, N.N. (2015) "SARIMA Modeling of frequency of Monthly Rainfall in Umuhia, Abia State of Nigeria, American Journal of Mathematics and Statistics, 5(2):82-87.
3. Box, G. E. P., Jenkins, G. M. (1970), "Time Series Analysis Forecasting and Control, Holden-Day, San Francisco, CA.
4. Box, G.E. P., Jenkins, G.M., and Reinsel, G. C. (1994). "Time Series Analysis: Forecasting and Control", 3rd ed. Prentice Hall, Upper Saddle River, N.J.
5. Brockwell, P., Davis, R. (2002), "Introduction to time series and forecasting", New York: Springer.

6. Chan, N. H. (2002), "Time Series Applications to Finance", John Wiley & sons, INC., publication, USA; ISBN 0-471-41117-5.
7. Chu, F. (2009), "Forecasting tourism demand with ARIMA-based methods", Tourism Management, Vol 32, No. 5, pp 740-751.
8. Franses, Ph. H. and Dijk, D. V. (2000), "Nonlinear Time Series Models in Empirical Finance" Cambridge University Press, ISBN 0 511 01100 8 virtual.
9. Gikungu, S. W., Waititu, A. G. (2015), "Forecasting Inflation Rate in Kenya Using SARIMA Model American Journal of Theoretical and Applied Statistics 4, 15-18.
10. Luo, C. S., Zhou, L., Qingfeng, W. (2013), "Application of SARIMA Model in Cucumber Price Forecast", Applied Mechanics and Materials, Vols. 373-375, pp. 1686-1690.
11. Mira, S. K., Ahmad M. R. (2015), "Time Series Models for Average monthly Solar radiation in Malaysia", Research and Education in Mathematics, International Conference Kuala Lumpur, Malaysia.

إنقطاعات الكهرباء والتقدير التنبؤي لها في محافظة دهوك - كردستان العراق باستخدام نموذج SARIMA

ملخص

تشير السلسلة الزمنية الموسمية إلى نمط منتظم من التغيرات التي تتكرر لفترة S من الزمن. وعادة السلسلة الموسمية تسبب في عدم استقرار السلسلة الزمنية. ويعرف الفرق الموسمي بأنه الفرق بين قيمة وقيمة أخرى التي يكون الفارق الزمني فيها من مضاعفات S من الفترات. في هذا البحث تطرقنا إلى المعدل اليومي لساعات انقطاع الكهرباء حسب الاشهر والتي تمثل الفجوة بين الحمل الكهربائي الشهري الذي يتم توريده لمحافظة دهوك (ضمن كردستان العراق) والطلب على الكهرباء واستهلاكها خلال الفترة من كانون الثاني 2010 إلى كانون الأول 2016. ولكون هذه السلسلة لها تأثيرات موسمية، فإننا استخدمنا نموذج $SARIMA$ لعدد من الأشهر وهي $S = 12$. ونعتمد الفروقات الموسمية وغير الموسمية 0 أو 1 ، لاختيار النموذج المناسب، الذي يعطي أدنى قيمة من $BIC, RMSE, MAE, MAPE$ وأكبر قيمة إلى R^2 ، لغرض التنبؤ بمعدل ساعات القطع اليومية خلال الأشهر المقبلة. إستنتجنا إن أفضل نموذج هو $SARIMA(1,1,1)(0,1,0)_{12}$ ، وإستناداً لهذا النموذج لاحظنا تزايد المعدل اليومي لساعات الإنقطاع، ومن المتوقع أن يكون عجز كلي لتجهيز الكهرباء خلال شهر كانون الاول من عام 2018 إذا استمرت السلسلة الزمنية بهذا النمط.

Appendix:

CUT TIME EQUATIONS:

Seasonal Factors (Seasonal components): $F_t = \{ \text{medial average } (S_{lt+p}, S_{lt+2p}, \dots, S_{lt+qp}), 1 \leq t \leq L - [Lp]p \text{ medial average } (S_{lt+p}, S_{lt+2p}, \dots, S_{lt+(q-1)p}), L - [Lp]p < t \leq [p2] \text{ medial average } (S_{lt}, S_{lt+p}, \dots, S_{lt+(q-1)p}), \dots$

Model Seasonal Algorithm: $X_t = TC_t S_{lt}$, $t = 1, \dots, n$ Additive Model $X_t = TC_t + S_t + I_t$, $t = 1, \dots, n$ where TC_t is the "trend-cycle" component, S_t is the "seasonal" component, and I_t is the "irregular" or "random".

R-squared: $R^2 = 1 - \frac{\sum (Y(t) - \hat{Y}(t))^2}{\sum (Y(t) - \bar{Y})^2}$ Related Topics Goodness-of-Fit Statistics (TSMODEL algorithms) Mean Squared Error (TSMODEL algorithms) Mean Absolute Percent Error (TSMODEL...)

Goodness-of-fit statistics are based on the original series $Y(t)$.

Let k = number of parameters in the model, n = number of non-missing residuals. Related Topics Mean Squared...

Normalized $BIC = \ln(MSE) + k \ln(n)$ Related Topics Goodness-of-Fit Statistics (TSMODEL algorithms) Mean Squared Error (TSMODEL algorithms) Mean Absolute Percent Error (TSMODEL)

$MAPE = 100n \sum |Y(t) - \hat{Y}(t)| / Y(t)$ Related Topics Goodness-of-Fit Statistics (TSMODEL algorithms) Mean Squared Error (TSMODEL algorithms) Maximum Absolute Percent Error (TSMO)

$MSE = \sum (Y(t) - \hat{Y}(t))^2 / (n - k)$ Related Topics Goodness-of-Fit Statistics (TSMODEL algorithms) Mean Absolute Percent Error (TSMODEL algorithms) Maximum Absolute Percent Error (TSM)

$MAE = \ln \sum |Y(t) - \hat{Y}(t)|$ Related Topics Goodness-of-Fit Statistics (TSMODEL algorithms) Mean Squared Error (TSMODEL algorithms) Mean Absolute Percent Error (TSMODEL algorithms)