ISSN (1681- 6870)

Time Series Modeling to Forecast on Consuming Electricity: A case study Analysis of electrical consumption in Erbil City from 2014 to 2018

Dler Hussein Kadir
Department of Statistics
College of Administration and Economics
Salahaddin University
dler.kadir@su.edu.krd

Abstract:

Time series analysis and forecasting have become a major tool in different applications in hydrology and environmental management fields. Among the most effective approaches for analyzing time series data is the model introduced by Box and Jenkins, ARIMA (Autoregressive Integrated Moving Average). Approach: In this study, we used Box-Jenkins methodology to build ARIMA model for electricity consumption data taken for Erbil region station for the period from 2014-2018. Results: In this research, ARIMA (1, 1, 1) (0, 1, 1)12 model was developed. This model is used to forecasting the monthly consumption for the upcoming 2019 year in each month to help decision makers establish priorities in terms of electricity demand management. Conclusion/Recommendations: An intervention time series analysis could be used to forecast the peak values of producing electricity in megawatt for Erbil city.

Keywords: Time Series Modeling, Forecasting, ARIMA, Moving Average, Consuming Electricity and Autoregressive.

نمذجة السلاسل الزمنية للتنبؤ باستهلاك الكهرباء: دراسة حالة تحليل استهلاك الكهرباء في مدينة أربيل من 2014 إلى 2018

دلير حسين قادر
جامعة صلاح الدين – كلية الادارة والاقتصاد – قسم الاحصاء

المستخلص: أصبح تحليل السلاسل الزمنية والتنبؤ بها أداة رئيسية في التطبيقات المختلفة في مجالات الهيدرولوجيا وإدارة البيئة. من بين أكثر الأساليب فعالية لتحليل بيانات السلاسل الزمنية النموذج الذي قدمه Box and Jenkins، ARIMA (المتوسط المتحرك الانحدار التلقائي). المنهج: في هذه الدراسة استخدمنا منهجية Box-Jenkins لبناء نموذج ARIMA لبيانات استهلاك الكهرباء المأخوذة لمحطة منطقة أربيل للفترة من 2014-2018. النتائج: في هذا البحث تم تطوير نموذج 12(0,1,1)(1,1,1) ARIMA. يستخدم هذا النموذج للتنبؤ بالاستهلاك الشهري لعام 2019 القادم في كل شهر لمساعدة صانعي القرار على تحديد الأولويات فيما يتعلق بإدارة الطلب على الكهرباء. الاستنتاج / التوصيات: يمكن استخدام تحليل السلاسل الزمنية للتدخل للتنبؤ بقيم الذروة لإنتاج الكهرباء بالميجاواات لمدينة أربيل.

الكلمات المفتاحية: نماذج السلاسل الزمنية، التنبؤ، ARIMA ، المتوسطات المتحركة

 ISSN (1681- 6870)

1 Introduction

Many methods and approaches for formulating forecasting models are available in the literature. This research exclusively deals with time series forecasting model, in particular, the Auto-Regressive Integrated Moving Average (ARIMA). These models were described by Box and Jenkins [1].

The Box-Jenkins approach possesses many appealing features. It allows the manager who has only data on past years' quantities, rainfall as an example, to forecast future ones without having to search for other related time series data, for example temperature. Box-Jenkins approach also allows for the use of several time series, for example temperature, to explain the behavior of another series.

Box-Jenkins (ARIMA) modeling has been successfully applied in various water and environmental management applications. The followings are examples where time series analysis and forecasting are effective.

2 Electricity in Kurdistan

As a result of the First Gulf War in 1991 and the ensuing internal conflicts, the electricity supply system in the three northern governorates suffered severe damage (e.g. several distribution and transmission lines were put out of commission, many substations were destroyed, power station were ruined by explosives). In 1994, the governorates of Duhok, Erbil and Sulaymaniyah had been cut off from the national grid and Erbil and Sulaymaniyah relied on the hydropower stations of Dokan and Derbandikhan in Sulaymaniyah governorate for their power supply while Duhok had no power supply for almost one year.

By early 1998, the electricity generation, substations and transmission and distribution systems became very weak and power cuts of up to about 5 hours were a regular practice. In certain areas, electricity supply was limited to 3 to 5 hours daily, further reduced to about one hour per day or no supply in some areas.

The electricity sector of the Kurdistan Region of Iraq has significantly improved. In 2013, it had reached more than 20 hours supply and brought in private sector investments.

In this project, we are going to study the amount consumed MW in Erbil 2014 to 2018, our main objective is to find the best model to predict for future demand in Erbil city as people are suffering from lack of electricity supply in the area.

3 Materials and Methods

The main stages in setting up a forecasting ARIMA model includes model identification, model parameters estimation and diagnostic checking for the identified model appropriateness for modeling and forecasting. Model Identification is the first step of this process. The data was examined to check for the most appropriate class of ARIMA processes through selecting the order of the consecutive and seasonal differencing required to make series stationary, as well as specifying the order of the regular and seasonal autoregressive and moving average polynomials necessary to adequately represent the time series model. The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are the most important elements of time series analysis and forecasting. The ACF measures the amount of linear dependence between observations in a time series that are separated by a lag k. The PACF plot helps to determine how many auto regressive terms are necessary to reveal one or more of the following characteristics: time lags where high correlations appear, seasonality of the series, trend either in the mean level or in the variance of the series.

4 Time Series

In statistics, signal processing and financial mathematics, a **time series** is a sequence of data points, measured typically at successive times spaced at uniform time intervals. **Time series analysis** comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. **Time series forecasting** is the use of a model to forecast future events based on known past events: to predict data points before they are measured. There are two types of data indeed. The first one is that when the observation is recorded as continuing such as Temperature. The second is when the data is recording discretely as rainfall data for example [3].

4.1 Autoregressive Model

Let $e_1, e_2, \dots$ be a purely random process with mean zero and variance σ^2. Then we can define an autoregression process z_t of order p, written AR(p), if

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \cdots + \varphi_p z_{t-p} + e_t \qquad (1)$$

This looks just like a multiple regression model, except that the regressors are just the past values of the series. Autoregressive series are stationary processes provided the variance of the terms are finite and this will depend on the value of the φ's. Autoregressive processes have been used to model time series where the present value depends linearly on the immediate past values as well as a random error.

4.2 ARMA Models

We can combine the moving average (MA) and the autoregressive models (AR) processes to form a mixed autoregressive/moving average process as follows:

$$Z_t = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \cdots + \emptyset_p Z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \qquad (2)$$

which is formed by a $MA(q)$ and an $AR(p)$ process. The term used for such processes is an ARMA process of order (p, q). An advantage of using an $ARMA$ process to model time series data is that an $ARMA$ may adequately model a time series with fewer parameters than using only an MA process or an AR process. One of the fundamental goals of statistical modeling is to use the simplest model possible that still explains the data – this is known as the principle of parsimony [3].

4.3 Multiplicative Seasonal Model

In rare case all of the seasonal models are combined with a non-seasonal model in order to reach an appropriate form and it is symbolized by $ARMA(p, q)\ (P, Q)_s$ and its backward shift form is provided below;

$$(B)\Phi(B^s)Z_t = \theta(B)\Theta(B^s)\varepsilon_t \qquad (3)$$

The variance is;

 ISSN (1681- 6870)

$$\gamma_0 = (1 + \theta^2)(1 + \Theta^2)\sigma_\varepsilon^2 \quad (4)$$

4.4 Autoregressive Integrated Moving Average ARIMA

Most time series in their raw form are not stationary. If the time series exhibits a trend, then, as we have seen, we can eliminate the trend using differencing $\nabla^k Z_t$. If the differenced model is stationary, then we can fit an ARMA model to it instead of the original non-stationary model. We simply replace Z_t in (21) by $\nabla^k Z_t$. Such a model is then called an *autoregressive integrated moving average* ($ARIMA$) model. The word "integrated" comes from the fact that the stationary model that was fitted based on the differenced data has to be summed (or "integrated") to provide a model for the data in its original form. Often, a single difference $k = 1$ will suffice to yield a stationary series. The notation for an $ARIMA$ process of order p for the AR part, order q for the MA part and differences d is denoted $ARIMA(p, d, q)$. Its Backward Shift Operator form can be written as below [2].

$$(B)W_t = \theta(B)\varepsilon_t \quad (5)$$

where;

$$W_t = \nabla^d Z_t$$

$$\emptyset(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p$$

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

4.5 Integrated Multiplicative Seasonal Model

It is a merged model from seasonal part ARIMA and non-seasonal part, and the symbol of this model is $ARIMA(p, d, q)\ (P, D, Q)_S$. The general formula derived by *Backward Shift Operator* is;

Where ∇_S^D: is a degree of seasonal difference

$$\nabla_S^D Z_t = (1 - B^S)^D Z_t \qquad D = 0,1,2, \ldots$$

$$= Z_t - Z_{t-S}$$

5 Stages of building model

The most important purpose of building time series model is to be used for future and can be reliable to make serious decisions in some circumstances. Thus, there are several stages to find the best fitted time series model [6].

5.1 Identification

Identifying model is the most vital part in time series modeling. At this stage, the first step in this phase is the formulation of the original time series to get to know their characteristics (direction, changes, rotating, seasonal changes…) and then test series in terms of stationary properties around the mean and variance. Finally, we calculate the autocorrelation function and partial autocorrelation function.

- If the autocorrelation function is exponential decay and the partial autocorrelation function interrupted after a period (p), then the pattern is the AR(p)
- If the partial autocorrelation function is exponential decay and the autocorrelation function interrupted after a period (q), then the pattern is the MA(q)
- If both autocorrelation function and partial autocorrelation function are exponential decay, then the patterns are most likely to be ARMA(p, q)

5.2 Estimation

After identifying the character of the model and decide which model is to set on the pattern, then the estimation of the parameters comes forward and needs to be estimated. In addition, there are several methods to do so;

a. Least-Square Method
b. Maximum Likelihood Method
c. Yule-Walker Method

5.3 Model Diagnostic Checking

Model diagnose is the end of times series modeling which tells the researchers how good the model fits the series. The procedure is easy, once estimation is done; we put the original series into the model and compute the residuals. The behavior of the residual does provide a good feeling about the selected model.

5.3.1 Autocorrelation of Residuals Test

When there are no significant differences between the autocorrelation function of the fitted series and autocorrelation function of the original series, this indicates that the model is appropriate. However, if the autocorrelation function does not have a significant difference, we calculate the autocorrelation of residuals r_ε. If the autocorrelation function of the residuals is outside the below range, it gives us enough evidence that the model is appropriate for the data;

$$-1.96\, S(r_\varepsilon) \le \rho\,(\varepsilon) \le 1.96\, S\,(r_\varepsilon) \qquad (6)$$

$S(r_\varepsilon)$: The standard deviation of residual

$$r_\varepsilon = \frac{\sum_{t=1}^{n-k} \hat{\varepsilon}_t \hat{\varepsilon}_{t+k}}{\sum_{t=1}^{n} \hat{\varepsilon}_t^2}$$

$$S(r_\varepsilon) = N^{\frac{-1}{2}} \left[1 + 2\sum_{k=1}^{k-1} r_\varepsilon\right]$$

5.3.2 Goodness of Fit Test

Each of the [1,3] studied mixed model (p, d, q) ARIMA and is supposed to be:

$$Q = n \sum_{k=1}^{m} r_k^2\,(\hat{\varepsilon}) \sim \chi^2_{(m-p-q)} \tag{7}$$

One can determine the appropriateness of the model through this test, which is in accordance with the following hypothesis:

H_0: Model appropriate
H_1: Model not appropriate

Thus, if $p - value < 0.05$, the null hypothesis is rejected and it means the model is not fitted the data well and vice versa.

5.4 Forecasting

It is known to predict that the future value of certain behavioral knowledge of the phenomenon with having the least possible error compared to reality. After the completion of the first three phases, identifying the model, determine the appropriate model, and checking the appropriateness of the model, we use the fitted model to predict. A good predictor is the one which gets mean squares error (MSE) as small as possible;

$$E[\varepsilon_t^2(l)] = E\big[(Z_{t+l} - \hat{Z}_t(l))\big]^2 \tag{8}$$

$\hat{Z}_t(l)$: Predicted value
t: The original time period.
By making equation (7) to zero, we have;

$$\hat{Z}_t(l) = E(Z_{t+l})$$

Substituting equations (8) into (7), the MSE is of $\hat{Z}_t(I)$:

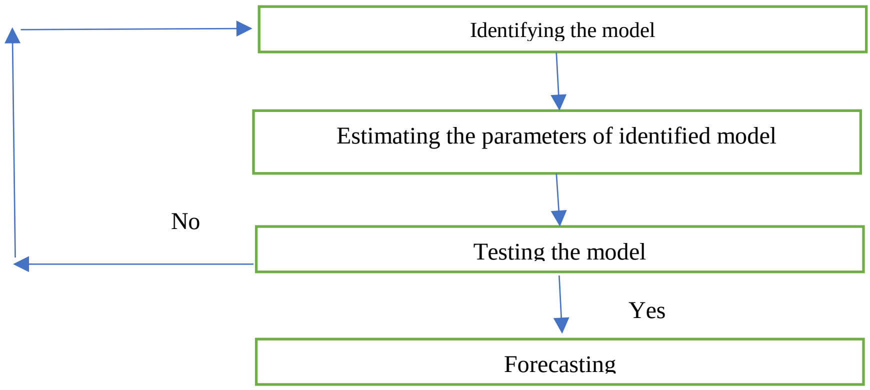
$$MSE = E[(Z_{t+l} - E(Z_{t+l})]^2 = Var(Z_{t+l}) \tag{9}$$

Thus we get the optimal prediction of the conditional expectation for (Z_{t+l}) equals to $\widehat{Z}_t(I)$, so;

$$\hat{Z}_t(l) = E(Z_{t+l}|Z_t, Z_{t-1}, \dots, Z_1)$$

ISSN (1681- 6870)

It can illustrate the stages of building the time series model in the following chart:



6 Case Study

In this section, we are going to display several models based on our data in order to find a suitable model to predict future values. The data was taken from a government organization for weather record sets from 2014 to 2018. The analysis is done by using STATGRAPHIC version 16.

6.1 Time series analysis

The purpose of time-series analysis to find an appropriate model for electricity in the city of Erbil, and then take advantage of these models for the purposes of prediction and control usage electricity source for the future.

First of all, we need to plot the data and see its behavior. From the plot, it can be seen that whether it has trend or seasonality shape. This is the most important step because all of the checking models are based on this step. If you go wrongly, it would take a long time till you gain the best model. Therefore, one has to be very careful at this stage. Since the data is a daily, Fig. 1, shows that there is a seasonal cycle of the series and the series is not stationary. The ACF and PACF of the original data, as shown in Fig. 2, show that the electricity data is not stationary.

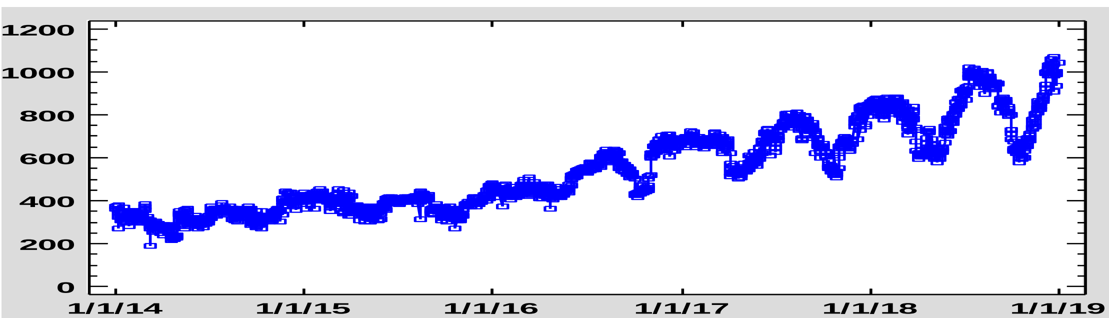
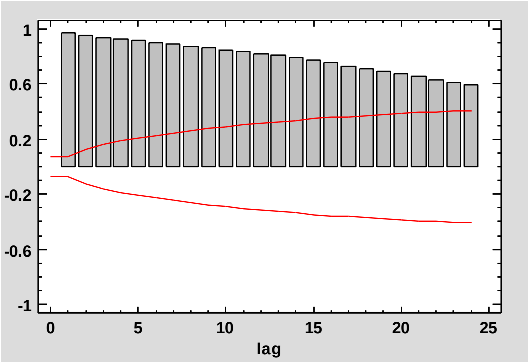


Figure 1: Time Series plot of the original electricity data



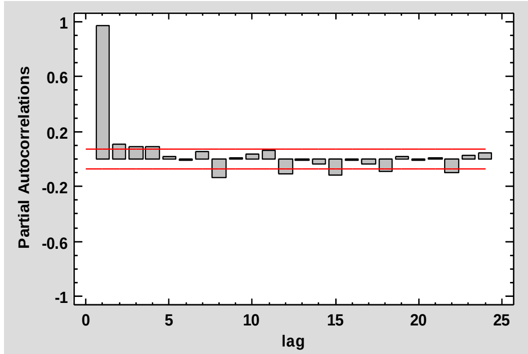


Figure 2: ACF&PACF for the original electricity data

In order to start modeling to our data, the data must be randomness. Although, it is obvious that the data is randomness regarding to the plots, we still check it by using Box & Pierce test. If p-value is smaller than 0.05 (p-value <0.05) then we reject null hypothesis which is shown below;

Table 1: Test of randomness for the data

Data	Hypotheses	P-Value
Electricity Data in Erbil	H_0: *The series is randomness* H_1: *The series is not randomness*	0.0000

It is clear that the above result indicates non stationary process and also all the values of autocorrelation functions do not lie in the below confidence interval as given below;

$$-1.96\, S(r_k) \leq \quad \rho_k \quad \leq 1.96\, S(r_k)$$

Where k is the lag and $S(r_k)$ is the standard deviation of the autocorrelation.

In order to fit an ARIMA model stationary data in both variance and mean are needed. We could attain stationary in the variance could be attained by having log transformation and differencing of the original data to attain stationary in the mean. For our data, we need to have seasonal first difference, $d = 1$ of the original data in order to have stationary series. After that, we need to test the ACF and PACF for the differenced series to check stationary.

As a result of the above transformation now we test our data to see the differences between those results and test as well. The below result shows enough information that our data is stationary now;

Issue No. 46/ 2020
ISSN (1681- 6870)

Table 2: Randomness Test for the transformed data

Data	Hypotheses	P-Value
Electricity Data in Erbil	H_0: *The series is randomness* H_1: *The series is not randomness*	0.163

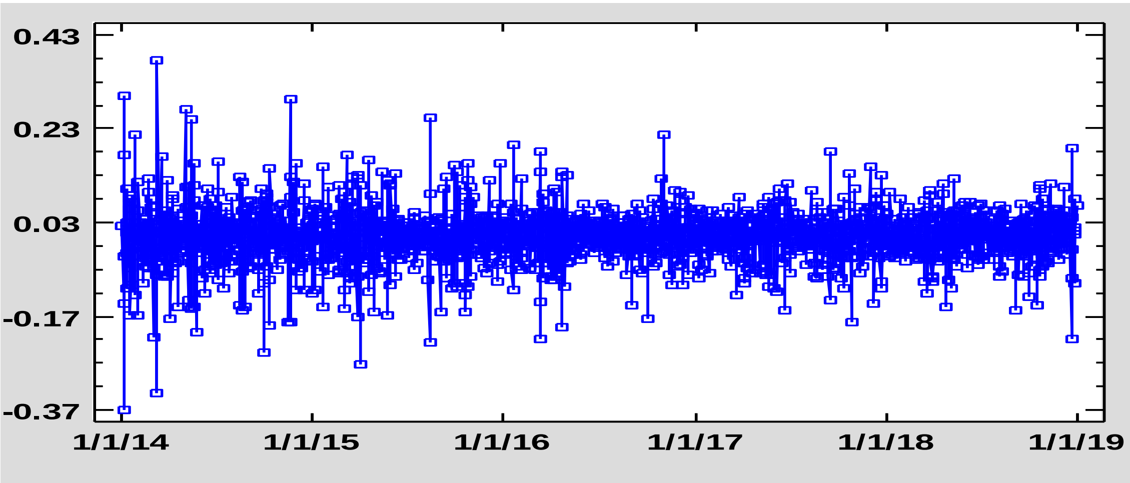
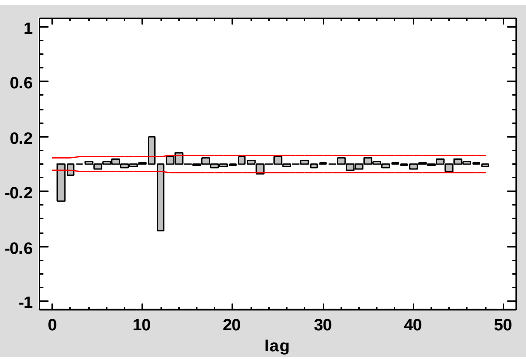
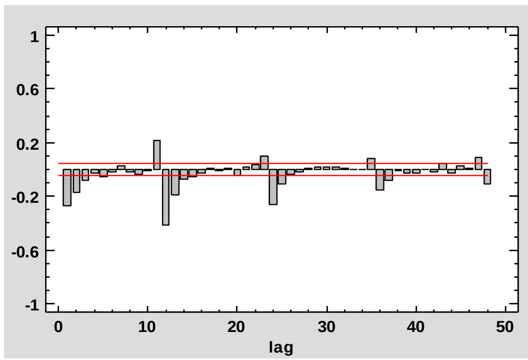


Figure 3: Time Series Plot of Transformed Data



(a)



(b)

Figure 4: (a) Autocorrelation (ACF); (b): Partial Autocorrelation (PACF) for first order seasonal differencing and de-seasonalized original electricity data

As shown in Fig. 4 the ACF and PACF for the differenced and de-seasonalized the data are almost stable which support the assumption that the series is stationary in both the mean and the variance after having 1st order seasonal difference. Therefore, an ARIMA $(p, 1, q)\ (P, 1, Q)^{12}$ model could be identified for the differenced and de-seasonalized electricity data. After ARIMA model was identified above, the p, q, P *and* Q parameters need to be identified for our model.

 ISSN (1681- 6870)

6.2 Choosing the appropriate model

After stabilizing the time series around mean and variance, next we need to define models by setting several values into the p, q, P *and* Q. The following is MSE values for 5 different models provided in Table 5.

Table 3 : Fitted Models andtheir MSE

Fitted Models	**MSE**
ARIMA(1,1,1)x(1,1,1)12 with constant	22.7566
ARIMA(1,1,1)x(0,1,1)12	22.7532
ARIMA(0,1,0)x(1,1,1)12	23.5066 NS
ARIMA(1,1,0)x(1,1,0)12	27.4866
ARIMA(0,1,1)x(0,1,1)12	22.8582
ARIMA(1,1,0)x(0,1,1)12	23.0072
ARIMA(0,1,0)x(1,1,2)12	23.5095
ARIMA(0,1,1)x(0,1,2)12	22.8588

As shown from Table 5 there are several models fitted on the data and less value of MSE indicates the more suitable model. However, there are other assumptions which the model must meet such as, significant parameters, the randomness of residuals is the most important. Thus, here we select ARIMA (1,1,1)x(0,1,1)12 as it does not have the smallest MSE but its parameter is highly significant and we always look forward with the smallest

Table 4: Result of ARIMA (1,1,1) (0,1,1)12

Parameter	Estimate	Stnd. Error	T		P-value
AR(1)	0.241263	0.0581368	4.14992		0.000033
MA(1)	0.579839	0.0494229	11.7322		0.000000
SMA(1)	0.979969	0.00149198	656.824		0.000000

And the form of the above model can be written as followings;

$$(1 - \varphi_1 \beta)\nabla_{12}\, z_t = (1 - \theta_1 \beta^{12}) e_t$$

6.3 Testing the Fitted Model

After identifying the model, it has to be tested whether it is appropriate for the data or not. The most important part of doing this is to check the capability of the model how far it goes with the original series. Thus, it can be reliable to predict future cases. Now, we test the residual autocorrelation. Since all of the values inside the confidence interval, it means the variable of residual autocorrelation function is randomness and thus the model is appropriate for the data.

$$-1.96\,S(r_{\varepsilon}) \leq \rho_{\varepsilon} \leq 1.96\,S(r_{\varepsilon})$$

6.4 Goodness of Fit Test

This is the final step when one chooses the best fitted model is to check it mathematically rather than looking at the autocorrelation function of the residual. We can test it again by Box & Pierce. As $p - value > 0.05$ it means the model is good to be used for forecasting which is explained in the next section. The Table illustrates the above discussion;

Table 5: Goodness of Fit Test for Fitted Model

Series Data	Fitted Model	Hypotheses	P-Value
Electricity Data in Erbil	*ARIMA(1,1,1)x(0,1,1)12*	$H_0: Fitted\ model\ approperiate$ $H_1: Fitted\ model\ not\ apprperiate$	0.348

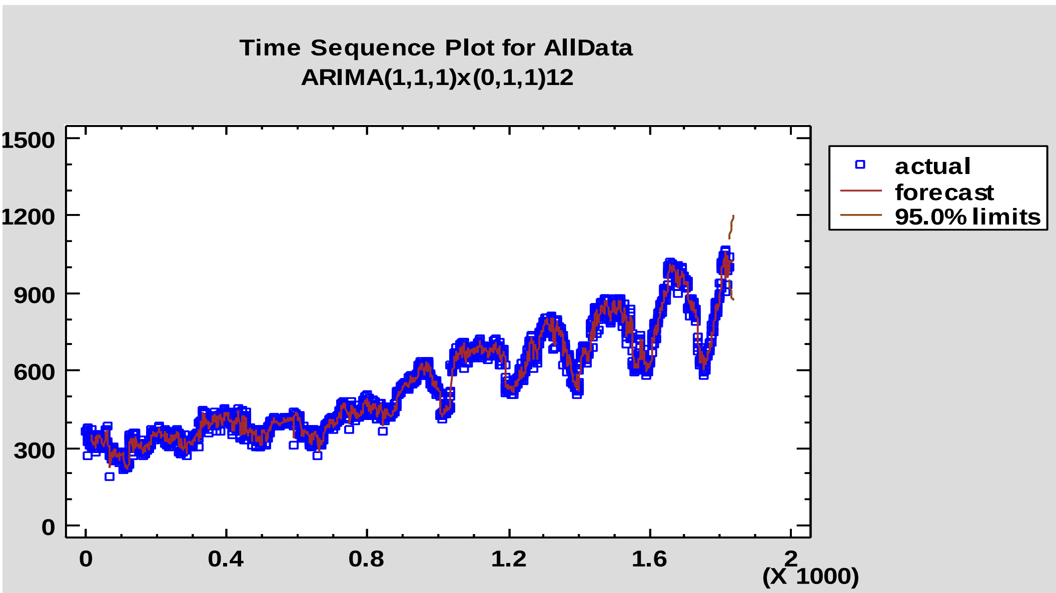
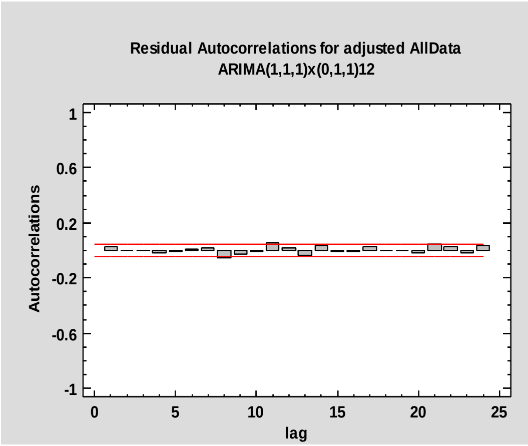


Figure (5): Forcasting plot of the estimated best fit model

 ISSN (1681- 6870)



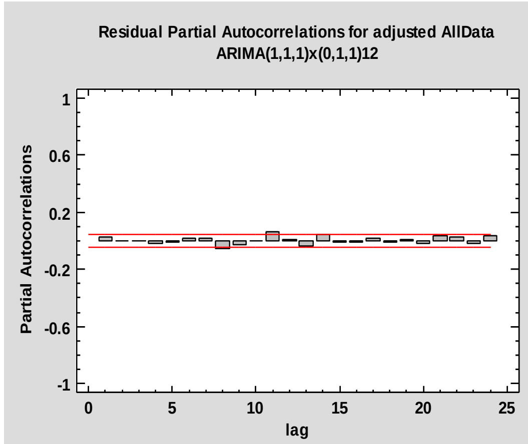


Figure (6): ACF and PACF of residual model

Since all lags are within the two red lines from both graphs, it means that the lags are not significant and that is what we are anticipating.

6.5 Forecasting

After diagnosing the fitted model and selected as the best one, final step comes forward forecasting. This is the last part of modeling in time series analysis.

Table 6 : Predicted electricit usage based on the orginal data as well as estimated model

2019	Forecast	Lower 95.0% Limit	Upper 95.0% Limit
Jan	1030.946	950.786	1107.61
Feb	1053.372	932.268	1119.75
Mar	1076.656	921.968	1130.13
Apr	1098.708	914.135	1140.04
May	1125.77	903.592	1145.02
Jun	1151.763	895.704	1151.95
Jul	1136.83	897.785	1170.07
Aug	1178.617	891.557	1177.39
Sep	1204.565	880.556	1178.07
Oct	1232.018	877.594	1188.13
Nov	1259.604	875.416	1198.7
Dec	1287.395	870.229	1205.02

7 Conclusion

Time series analysis is an important tool in modeling and forecasting. ARIMA(1,1,1) (0,1,1)12 model give us information that can help the decision makers establish strategies, priorities and proper use of electricity resources in Erbil. This piece of information is quite meaningful and appropriate to predict the exact monthly needed electricity data. Therefore, it is worth mentioning that individual data should not be used in decision making by depending on our model. However, an intervention time series analysis can be tested to see if we can improve our model performance in forecasting the peak values of electricity data.

References

[1] Box, G.E.P. and G.M. Jenkins, 1976. Time Series Analysis: Forecasting and Control. Revised Edn., Hoden-Day, San Francisco.

[2] Brillinger, D.R. (1975). Time Series: Data Analysis and Theory. New York: Holt, Rinehart & Winston Inc.

[3] Chatfield, C. (2004), The Analysis of Time Series, Chapman & Hall

[4] Langu, E.M., 1993. Detection of changes in rainfall and runoff patterns. J. Hydrol., 147: 153-167.

[5] Margherita Gerolimetto, (2010). ARIMA and SARIMA models.

[6] McQuarrie, A.D.R. and C-L. Tsai (1998). Regression and Time Series Model Selection, Singapore: World Scientific.

[7] Robert H. Shumway and David S. Stoffer (2015). Time Series Analysis and It's Applications.

[8] Walter Vandaele, 1983. Applied time series and Box-Jenkins Models. Academic Press Inc., Orlando, Florida.