

using some statistical methods to predict lost electricity in holy Karbala Governorate

استعمال بعض الاساليب الاحصائية لأغراض التنبؤ بالطاقة الكهربائية الضائعة في
محافظة كربلاء المقدسة *

أ. د عدنان كريم نجم الدين
جامعة كربلاء / كلية الإدارة والاقتصاد / قسم الإحصاء
حسنين علي هاشم

البحث مستل

المستخلص

تم في هذا البحث استخدام نماذج السلاسل الزمنية الموسمية لدراسة وتحليل البيانات الشهرية عن الطاقة الكهربائية الضائعة في محافظة كربلاء المقدسة للمدة (2012-2016) لما تمتاز به هذه النماذج من دقة ومرونة عاليتين في تحليل السلاسل الزمنية . واطهرت نتائج التطبيق ان النموذج الكفؤ والملائم لتمثيل بيانات السلسلة الزمنية هو النموذج الموسمي المضاعف من الدرجة $SARIMA(0,1,1) (0,1,1)_{12}$ ووفقا لنتائج تقدير هذا النموذج تم التنبؤ بالطاقة الكهربائية الضائعة للمدة من كانون الثاني 2017 ولغاية كانون الاول 2018 ، حيث اظهرت هذه القيم تناسقا مع مثيلاتها في السلسلة الزمنية الاصلية .

Abstract

The seasonal time series models were used to study and analyze the monthly data on lost electricity in the holy governorate of Karbala for the period 2012-2016 because of the high accuracy and flexibility of these models in time series analysis.

The results of the application showed that the efficient and appropriate model to represent the time series data is the seasonal multiplier model of the class $SARIMA(0,1,1) (0,1,1)_{12}$.

According to the results of this model estimate, the lost electricity was predicted for the period from January 2017 to December 2018. The predicted values were consistent with those in the original time series.

المقدمة :

مما لا شك فيه ان اهمية الطاقة الكهربائية التي اصبحت في عالمنا المعاصر تدخل في جميع مفاصل الحياة اليومية ولها اهمية بالغة لجميع القطاعات الاقتصادية ، ومن البديهي ان دراسة اي ظاهرة من الظواهر يعتمد على تحليل البيانات الخاصة بتلك الظاهرة في مده زمنية مناسبة (سواء أكانت سنوات ام اشهر ام مدد زمنية اخرى)، وعلى ضوء البيانات المتوفرة يتم تقدير والتنبؤ للمدد اللاحقة .

وتجدر الإشارة الى ان المنظومة الكهربائية تتكون من ثلاث مراحل وهي (مرحلة الانتاج ، مرحلة النقل ، مرحلة التوزيع) وصولا الى المستهلكين في مختلف اصنافهم الحكومي والتجاري والعائلي .

كما وان الضائعات في الطاقة الكهربائية هي ظاهرة طبيعية في كل مرحلة من مراحل المنظومة الكهربائية وضمن نسب مقبولة عالميا وتحدث نتيجة الى المرور في التيار الكهربائي في المراحل المختلفة والاسباب عديدة .

مشكلة البحث

تكمن مشكلة البحث في كون الضائعات من الطاقة الكهربائية تمثل ركيزة مهمة تؤثر في كميات الطاقة الكهربائية المجهزة ومن ثم فان مستوى التجهيز يتأثر سلبيا مع انتشار هذه الظاهرة ، مما يستوجب دراستها وبيان اسبابها واقتراح الخطوات اللازمة للحد منها ومعالجتها .

فرضية البحث

- 1- الفرضية الصفرية (فرضية العدم): يفترض الباحث عدم وجود علاقة ذات دلالة احصائية بين نسبة الضائعات في الطاقة الكهربائية وطرائق الجباية المعتمدة حالياً
- 2- الفرضية البديلة: توجد علاقة ذات دلالة احصائية بين نسبة الضائعات في الطاقة الكهربائية وطرائق الجباية المعتمدة .

هدف البحث

يهدف البحث الى التنبؤ بالطاقة الكهربائية الضائعة عن طريق اختيار افضل نموذج للتنبؤ من خلال نماذج بوكس جنكنز، وباستعمال البرنامج الاحصائي Gretl .

1- الجانب النظري : مفاهيم رئيسية Fundamental Concepts

1-1 السلسلة الزمنية : Time Series

بالإمكان تعريف السلسلة الزمنية بانها مجموعة من القياسات او المشاهدات التي يتم جمعها عن ظاهرة ما في مدد زمنية معينة وغالبا ما تكون هذه المدد الزمنية متساوية كأن تكون (سنة، شهر ، يوم ... الخ) بحيث يتشكل لدينا توزيع ذو بعدين أولهما الزمن الذي يمثل المتغير المستقل، والبعد الاخر للتوزيع يمثل قيم الظاهرة المدروسة [8].
وتقسم السلاسل الزمنية حسب سلوكها الى : السلاسل الزمنية الموسمية Seasonal time series والسلاسل الزمنية غير الموسمية Non- seasonal time series كما تصنف السلاسل الزمنية الى السلاسل الزمنية المستقرة Stationary time series والسلاسل الزمنية غير المستقرة Non-Stationary time series
ويتم الكشف [7] عن مركبات السلسلة الزمنية عن طريق الاختبارات الاحصائية ، منها اختبار دانيال لكشف مركبة الاتجاه العام إذ يعد من اهم المركبات التي تكون منحني السلسلة الزمنية ، ويستعين هذا الاختبار بمعامل ارتباط لسبيرمان Spearmanis Correlation Coefficient والذي يأخذ الصيغة الاتية :.

$$R_S = 1 - \frac{6 \sum d_i^2}{n(n^2-1)} \dots\dots\dots(1-1)$$

إذ ان : d_i تمثل الفرق بين رتب قيم السلسلة الزمنية وبين رتب قيم الزمن ، n تمثل عدد المشاهدات .
ثم نقارن بين القيمة المحسوبة لمعامل الارتباط الرتبي والقيمة الجدولية للمعامل نفسه فاذا كانت القيمة المحسوبة اكبر من القيمة الجدولية فان السلسلة تحتوي على الاتجاه العام وبالعكس
وللكشف [5] عن المركبة الموسمية (الفصلية) تستخدم احد الاختبارات الاحصائية الاكثر تداولاً وهو اختبار كروسكال - ويليز
(Kruskall-Wallis) ويرمز له بالرمز KW وتعطي علاقته كما يأتي :

$$KW = \frac{12}{n(n+1)} \sum_{i=1}^K \frac{R_i^2}{n_i} - 3(n+1) \dots\dots\dots(1-2)$$

إذ ان :

n : مشاهدات السلسلة الزمنية بعد ازالة الاتجاه العام

R_i : تمثل مجموع رتب المشاهدات

n_i : مشاهدات السلسلة الزمنية عند k و k طول الموسم وتقرن قيمة (kw) مع قيمة (χ^2) الجدولة بدرجة حرية $(k-1)$ وبمستوى معنوية (α) معين فاذا كانت قيمة kw اصغر من قيمة (χ^2) الجدولة فان السلسلة الزمنية لا تحتوي على تأثيرات موسمية .

2-1 استقرارية السلسلة الزمنية Tim Series Stationary [9][17][19]

تعد استقرارية السلسلة خاصية مهمة في تحليل السلاسل الزمنية ، وكذلك في ايجاد النموذج الرياضي المناسب لها، فالسلسلة الزمنية المستقرة هي السلسلة التي تكون متجانسة زمنياً أي ان رسم السلسلة في المدة $\{t, t+h\}$ يكون بعض الاحيان مطابقاً لرسم السلسلة في مدة أخرى $\{s, s+h\}$ ، ويمكننا القول إن السلسلة مستقرة اذا حققت الشروط الاتية :
(1) ثبوت متوسط قيمها عبر الزمن :

$$E(X_t) = M \dots\dots\dots(1-3)$$

$$va(X_t) = E(X_t - \mu)^2 = \sigma_X^2 \dots\dots\dots(1-4) \quad (2) \text{ ثبوت قيمة التباين :}$$

3-1 عدم الاستقرار في السلاسل الزمنية Non-Stationary Time series [11][20]

ان أغلب نماذج السلاسل الزمنية هي من النوع غير المستقر ، وللتعرف على حالة عدم الاستقرار في السلاسل الزمنية يمكن فحص دالة الارتباط الذاتي إذ ان قيم معاملات الارتباط الذاتي الجزئي في هذه الحالة لا تؤول الى الصفر بعد الإزاحة الثانية أو الثالثة وإنما تبقى قيمها كبيرة لعدد من الإزاحات ، وهناك نوعان من السلاسل الزمنية غير المستقرة .
الأول: عدم الاستقرار على المتوسط (Non stationary about the mean) فتعني عدم تذبذب السلسلة الزمنية حول متوسط ثابت فالسلسلة الزمنية تكون غير مستقرة على المتوسط وفي هذه الحالة يتم اخذ الفروق المناسبة (differences) لتحقيق الاستقرار في السلسلة الزمنية .
الثاني: عدم الاستقرار به على التباين (Non stationary about the Variance) وفي هذه الحالة عندما تتذبذب السلسلة الزمنية حول تباين غير ثابت. ويمكن إزالته عن طريق التحويل اللوغاريتمي او الاسي (تحويلات القوى) (power transformation) لغرض تحويل السلسلة غير المستقرة حول التباين الى مستقرة .

4-1 دراسة استقرار السلسلة الزمنية :

يتم الكشف عن عدم استقرار السلسلة الزمنية باستعمال مجموعة من المعايير والاختبارات وتتمثل في :

1-4-1 دالة الارتباط الذاتي (ACF) Auto-correlation function [12][4]

تعرف دالة الارتباط الذاتي بأنه مقياس لدرجة العلاقة بين قيم السلسلة مع نفسها لمدة زمنية مزاحة ، تعد الارتباطات الذاتية صفة مميزة للعملية العشوائية وهي ذات أهمية كبيرة ؛ لأنها احدى اساليب التحديد فيما اذا كانت العملية العشوائية مستقرة ام لا، فاذا كانت كذلك فيتم اختيار احد النماذج المناسبة من مجموع نماذج العمليات العشوائية المستقرة .
إن الصيغة الرياضية لاجاد دالة الارتباط الذاتي للعملية العشوائية عند الإزاحة k هي:

$$\rho_k = \frac{cov(X_t, X_{t+k})}{\sqrt{var(X_t)var(X_{t+k})}} = \frac{E(X_t - \mu)(X_{t+k} - \mu)}{\sqrt{E(X_t - \mu)^2 E(X_{t+k} - \mu)^2}} \dots\dots\dots(1-5)$$

وعندما تكون العملية العشوائية مستقرة فإن المتوسط والتباين يكونان ثابتين ومتساويين لكل المدد وان $X_t = \sigma_{X_{t+k}}$ وعليه فان $\rho_k = \frac{cov(X_t, X_{t+k})}{\sigma_{X_t} \sigma_{X_{t+k}}}$ ولذلك تكون الحصيلة كالآتي:

$$\rho_k = \frac{E(X_t - \mu)(X_{t+k} - \mu)}{\sigma_X^2} = \frac{\gamma_k}{\gamma_0} \dots\dots\dots(1-6)$$

4-1-2 دالة الارتباط الذاتي الجزئي [12] Partial Auto-Correlation Function (PACF)

هناك دالة اخرى تسهم في تشخيص الانموذج المناسب، تعرف هذه الدالة بدالة الارتباط الذاتي الجزئي والتي يرمز لها اختصاراً (PACF) وتعد هذه الدالة مؤشراً بقياس العلاقة بين X_t و X_{t+k} للسلسلة نفسها مع افتراض ثبوت بقية قيم السلسلة الزمنية ، لذلك فهي أداة مهمة في تحليل السلسلة الزمنية و تساعدنا في تحديد درجة الانموذج ونوعه، ويمكن تقديره بواسطة المربعات الصغرى او مجموعة المعادلات التقريبية لتقدير الارتباط الذاتي الجزئي .
لنفرض لدينا انموذج الانحدار وبعد الإزاحة k نحصل على الانموذج:

$$X_t = \phi_{k_1} X_{t-1} + \phi_{k_2} X_{t-2} + \dots + \phi_{k_k} X_{t-k} + u_t \dots\dots\dots(1-7)$$

عندما k=1 فإن الانموذج (1-7) يمكن كتابته بصيغة انموذج انحدار جزئي

$$X_t = \phi_{11} X_{t-1} + u_t \dots\dots\dots(1-8)$$

5-1 نماذج السلاسل الزمنية الموسمية: Seasonal Time Series Models [1][12][18]

1-5-1 نماذج الانحدار الذاتي الموسمية Seasonal Autoregressive Models SAR(P)

بافتراض ان (X_t) تمثل انحرافات قيم السلسلة الزمنية المستقرة عن وسطها الحسابي ، فاذا كانت القيمة الحالية (X_t) تعتمد بصورة خطية على قيم الموسم نفسه من الاعوام السابقة $(X_{t-s}, X_{t-2s}, \dots, X_{t-ps})$ ، فضلا عن التغيير العشوائي في المدة الحالية (a_t) ، فيقال ان السلسلة تتبع عملية انحدار ذاتي موسمي من الرتبة (P) ، والذي يرمز لها بالرمز $ASR(P)$. ويمكن صياغة نموذج الانحدار الذاتي الموسمي بالصيغة الاتية:

$$X_t = \Phi_1 X_{t-s} + \Phi_2 X_{t-2s} + \dots + \Phi_p X_{t-ps} + a_t \quad \dots\dots\dots(1-9)$$

إذ ان :

X_{t-is} : قيم مشاهدات السلسلة الزمنية الموسمية ، $i=0,1, 2, \dots, p$

Φ_{is} : معالم الانحدار الذاتي الموسمي ، $i=0,1, 2, \dots, p$

S: طول المدة الموسمية .

a_t : الخطأ العشوائي .

وباستعمال عامل الازاحة الخلفي (back shift operator) يمكن التعبير عن هذا النموذج كالآتي :

$$\Phi(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_p B^{pS} \quad \dots\dots\dots(1-10)$$

فاذا كان النموذج من الرتبة الاولى اي ان قيمة $(P=1)$ ، وان طول الموسم (12) اي ان $(S=12)$ فان النموذج يكون $AR(1)_{12}$ كالآتي :

$$X_t = \Phi_1 X_{t-12} + a_t \quad \dots\dots\dots(1-11)$$

وباستعمال عامل الازاحة الخلفي يمكن التعبير عن هذا النموذج كالآتي :

$$(1 - \Phi_1 B^S) X_t = a_t \quad \dots\dots\dots(1-12)$$

2-5-1 نماذج المتوسطات المتحركة الموسمية: Seasonal Moving Average Models(SMA)

ان الصيغة العامة لنموذج المتوسطات المتحركة الموسمية من الرتبة (Q) ، أي $SMA(Q)$ ستأخذ الشكل الآتي :

$$X_t = a_t + \Theta_1 a_{t-s} - \Theta_2 a_{t-2s} - \dots - \Theta_Q a_{t-Qs} \quad \dots\dots\dots(1-13)$$

إذ ان :

Θ_{is} : معالم نموذج الاوساط المتحركة الموسمية ، $i=1,2,\dots,Q$

وباستعمال عامل الازاحة الخلفي يمكن التعبير عن هذا النموذج كالآتي :

$$X_t = \Theta_S(B^S) a_t = (1 - \Theta_S B^S - \Theta_{2S} B^{2S} - \dots - \Theta_{QS} B^{QS}) a_t \quad \dots\dots\dots(1-14)$$

فاذا كان النموذج من الرتبة الاولى اي ان قيمة $(Q=1)$ ، وطول الموسم (12) اي ان $(S=12)$ فان النموذج يكون $MA(1)_{12}$ كالآتي :

$$X_t = a_t - \Theta_1 a_{t-12} \quad \dots\dots\dots(1-15)$$

ان دالة الارتباط الذاتي للنموذج المتوسطات المتحركة الموسمية تنقطع بعد الازاحة Q_S (تؤول الى الصفر بعد الرتبة (Q) ، في حين ان دالة الارتباط الذاتي الجزئي تتناقص اسيا .

3-5-1 النموذج المختلط (الانحدار الذاتي والايوساط المتحركة) الموسمي

Seasonal Mixed(Autoregressive and Moving Average) Model (SARMA)

الصيغة العامة للنموذج الموسمي المختلط من الرتبة (P,Q) ستأخذ الشكل كالآتي :

$$X_t = \Phi_S X_{t-S} + \Phi_{2S} X_{t-2S} + \dots + \Phi_{PS} X_{t-PS} + a_t - \theta_S a_{t-S} - \theta_{2S} a_{t-2S} \dots - \theta_{QS} a_{t-QS} \dots \dots \dots (1-16)$$

وباستعمال مشغل الازاحة للخلف يمكن التعبير عن هذا النموذج كالآتي :

$$\Phi_S(B^S)X_t = \theta_S(B^S)a_t$$

$$(1 - \Phi_S B^S - \Phi_{2S} B^{2S} - \dots - \Phi_{PS} B^{PS})X_t = (1 - \theta_S B^S - \theta_{2S} B^{2S} - \dots - \theta_{QS} B^{QS}) \dots (1-17)$$

وعليه فان الصيغة النموذج ARMA(1,1)₁₂ تكون بالشكل الآتي :

$$X_t = \Phi_1 X_{t-12} + a_t - \theta_1 a_{t-12} \dots \dots \dots (1-18)$$

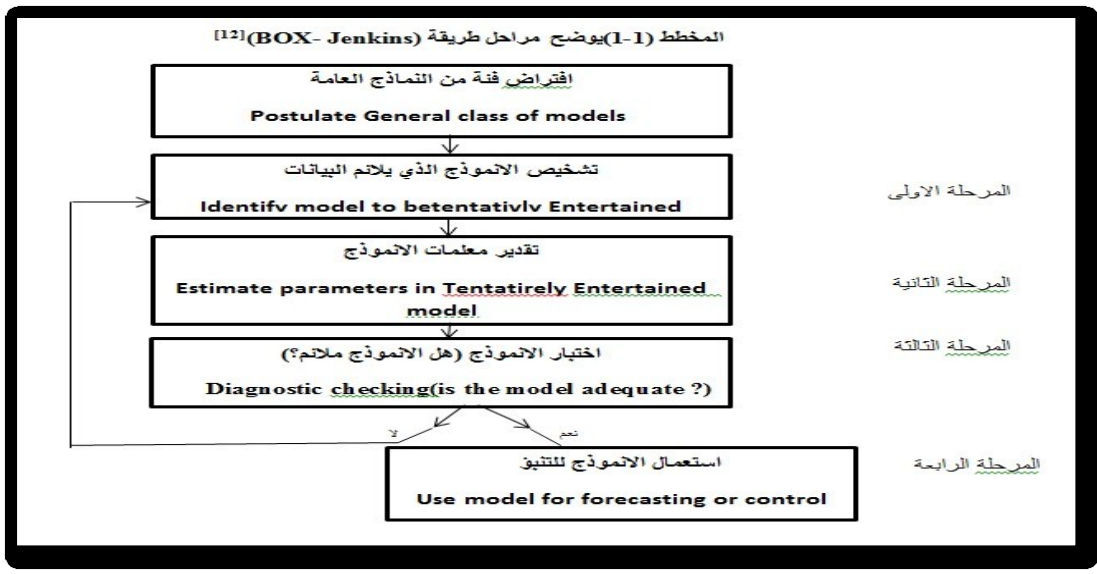
Multiplicative Seasonal الموسمي المضاعف 4-5-1 النموذج Model(SARIMA)

على الاغلب يتضمن بناء النماذج الموسمية استعمال معالم متتالية (غير الموسمية) ، لذا يمكن دمج جميع النماذج السابقة الموسمية وغير الموسمية في مجموعة واحدة من النماذج تسمى النماذج الموسمية العامة المضاعفة . وسنرمز الى تلك الفئة من النماذج بالرمز $ARIMA(p, d, q) \times (P, D, Q)_S$ وبافتراض ان (X_t) سلسلة مستقرة يمكن التعبير عن نماذج (ARIMA) المضاعفة كالآتي :

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - \Phi_1 B^S - \dots - \Phi_P B^{PS})X_t = (1 - \theta_1 B - \dots - \theta_q B^q)(1 - \Theta_1 B^S - \dots - \Theta_Q B^{QS})a_t \dots \dots (1-19)$$

6-1 طريقة بوكس – جنكنز Box – Jenkins method [3]

يقصد بطريقة بوكس – جنكنز (Box and Jenkins) هي تلك المنهجية التي طبقها كل من (Gwilyn & George Box) و (Jenkins) على السلاسل الزمنية عام 1970 وهذه المنهجية تعتمد في صيغتها على نموذج الانحدار الذاتي (Autoregressive Model) و أنموذج المتوسطات المتحركة (Moving Average Model) و الأنموذج المختلط – (Auto Regressive – Moving Average Model) ويمكن توضيح مراحل بناء النموذج حسب منهجية بوكس- جنكنز في التنبؤ كما في المخطط الآتي



1-6-1 المرحلة الاولى :التشخيص Identification [16][12]

تعد مرحلة التشخيص من اهم مراحل السلاسل الزمنية المهمة للوصول الى النموذج الملائم ، واولى خطوات مرحلة التشخيص هي رسم بيانات السلسلة الزمنية وعن طريق الرسم تكون لدينا فكرة جيدة عن مدى استقراره السلسلة من عدمها ، اي مدى احتواء السلسلة على الموسمية او الاتجاه العام او القيم شاذة ، وعدم الاستقرارية الذي يقود الى التحويلات الممكنة على البيانات ، لذلك فان رسم السلسلة يبين حاجتها الى التحويل المناسب لتستقر في متوسطها او تباينها قبل اي تحليل ، والخطوة الثانية هي حساب (ACF) ، (PACF) للسلسلة التي تم تحويلها في الخطوة الاولى وفحص (ACF),(PACF) لتحديد فيما اذا كانت السلسلة تحتاج الى اخذ فرق ثاني ، ان المهمة الرئيسية لهذه المرحلة هي تشخيص واختبار الانموذج الافضل ، الذي يمثل السلسلة الزمنية من النماذج العامة ARMA او ARIMA باستعمال معاملات الارتباط الذاتي والارتباط الذاتي الجزئي . وهناك عدة معايير للتحقق من دقة النموذج مقارنة بنماذج مقترحة .

1-1-6-1 معيار معلومات اكاكي Akaike Information Criterion(AIC) [10]

وضع الباحث اكاكي عام 1973 معياراً للتعرف على نوعية ومطابقة الانموذج في تحديد الرتبة ويرمز له (AIC) وتحديد الرتبة باختيار الانموذج الذي يقابل اقل قيمة للمعيار AIC وصيغته هي :

$$AIC = n \ln(\sigma_E^2) + 2k \dots\dots\dots(1-20)$$

إذ ان : k عدد المعلمات في النموذج

$$\sigma_E^2 : \text{مقدار تباين الخطأ}$$

n : عدد المشاهدات

2-1-6-1 معيار شوارتز Schwartz Bayesian Criterion (SBC) [3]

اقترح الباحث Schwartz عام 1978 معياراً على غرار معيار BIC ويتضمن اقتراح المعيار البيزي في اختبار الانموذج ويدعى (schwartz bayesian eriterion) ويرمز له (SBC) وصيغته هي :

$$SBC(P) = n \ln(\sigma_E^2) + M \ln(n) \dots\dots\dots(1 - 21)$$

إذ ان :

M و n هي عدد مشاهدات السلسلة والعدد الكلي لمعلمات الانموذج على الترتيب .

3-1-6-1 معيار معلومة بيز Bayesian information Criterion (BIC) [6]

في عام (1978,1979) قام الباحث اكاكي بتطوير معيار(AIC) الى معيار معلومة بيز(BIC) وصيغته هي :

$$BIC = 2\ln(\sigma_u^2) + M \ln(n) \dots\dots\dots(1-22)$$

إذ ان :

M, n : هي عدد مشاهدات السلسلة والعدد الكلي لمعلمات الانموذج على الترتيب .

$$(\sigma_E^2) \text{ يمثل متوسط مربعات الخطاء (MSE).}$$

ويتم اختيار الانموذج المناسب لبيانات السلسلة الذي يقابل القيمة الاقل لمعيار (BIC) .

1-6-1-4 معيار حنان - كوين Hannan-Quinn (H-Q) [3]

في عام 1979 اقترح الباحثان (Hannan و Quinn) معياراً لتحديد رتبة الانموذج ويرمز له (H-Q) وصيغته الرياضية هي :

$$H - Q = \ln \hat{\sigma}_a^2 + 2 M \ln(\ln n) / n ; C > 2 \dots\dots\dots(1-23)$$

M:دالة رتبة الانموذج (p+q).
C: ثابت .

[2] فضلا عن ان هناك اختبارات تعتمد في تقييم النماذج المتنبأ بها ومنها الاتي

1-متوسط مربعات الخطأ (MSE) Mean Squar Errorr

ويحسب وفق الصيغة الاتية :

$$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2 \dots\dots\dots(1-24)$$

$$e_t^2 = \sum (Y_t - \hat{Y}_t)^2 \dots\dots\dots(1-25)$$

إذ ان:

e_t : تمثل أخطا التنبؤ للسلسلة الزمنية في الزمن t .

n: تمثل عدد المشاهدات في السلسلة الزمنية .

2- المتوسط المطلق للخطأ (MAE) Mean Absolute Error

وصيغته الرياضية هي :

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \dots\dots\dots(1-26)$$

3- المتوسط المطلق للخطأ النسبي (MAPE) Mean Absolute percentage Error

وصيغته الرياضية هي :

$$MAEP = \frac{1}{n} \sum_{t=1}^n |Pe_t| \dots\dots\dots(1-27)$$

إذ ان :

pe_t : يمثل نسبة الخطأ (percentage error) ، ويحسب كالآتي :

$$pe_t = \left(\frac{X_t - e_t}{X_t} \right) * 100$$

X_t : مشاهدات السلسلة الزمنية

p : رتبة النموذج

1-6-2 المرحلة الثانية (مرحلة تقدير المعلمات) Estimation [12][15]:

ان المرحلة الثانية من مراحل دراسة السلاسل الزمنية وتحليلها هي عملية تقدير المعلمات وتأتي بعد عملية تشخيص الأنموذج

الملائم للسلسلة الزمنية وتوجد هناك عدة طرائق لتقدير معلمات الأنموذج ومنها :

1- طريقة الامكان الاعظم Maximum likelihood Method

2- طريقة المربعات الصغرى الاعتيادية Ordinary Least Squares Method

3- طريقة العزوم Moments Method

4- طريقة المربعات الصغرى الشرطية Conditional Least Square Method

ومن اهم هذه الطرائق هي طريقة الإمكان الأعظم (Maximum Likelihood Method) لتقدير معلمات الإنموذج وذلك بجعل مربعات الاخطأ اقل ما يمكن وفق هذه الطريقة وصيغة دالة الإمكان الأعظم تكون على النحو الآتي :

$$L(\theta, \phi, \sigma_e^2 | X_t) = (2\pi)^{-\frac{n}{2}} (\sigma_e^2)^{-\frac{n}{2}} \text{Exp} \left[-\frac{1}{2\sigma_e^2} S(\theta, \phi) \right] \dots (1-28)$$

إذ ان :

$S(\theta, \phi)$ تمثل مجموع مربعات الاخطاء اي :

$$S(\theta, \phi) = \sum_{t=1}^n e_t^2(\theta, \phi)$$

وعند اخذ اللوغاريتم تكون الدالة كما يأتي:

$$\text{Ln}L(\theta, \phi, \sigma_e^2) = -\frac{n}{2} \text{Ln}(2\pi \sigma_e^2) - \frac{S(\theta, \phi)}{2\sigma_e^2} \dots (1-28)$$

3-6-1 المرحلة الثالثة (اختبار دقة الانموذج) : Model Diagnostic Checking

بعد الانتهاء من مرحلتي التشخيص والتقدير ينبغي التحقق من دقة الانموذج المستعمل في هذه المرحلة اذ يتم اختبار النموذج لمعرفة مدى ملائمته لتمثيل بيانات الظاهرة المدروسة واستعمالها للحصول على تنبؤات مستقبلية إذ يوجد العديد من الاختبارات التي يمكن استعمالها لهذا الغرض ومنها .

1-3-6-1 اختبار حدي الثقة confidence interval test [9]

ويتم هنا اختبار قيم معاملات الارتباط الذاتي للبقاوي ويرمز له بـ $r_k(a_t)$ ، وبما ان معاملات الارتباط الذاتي للبقاوي تتوزع توزيعاً طبيعياً حسب ما اثبته كل من (Box and Pierce) عام 1970 بوسط حسابي مساويا للصفر وتباين مقداره $\left(\frac{1}{n}\right)$. وان معامل الارتباط الذاتي يأخذ الصيغة الآتية :

$$r_k(a_t) = \frac{\sum_{t=1}^{n-k} (a_t a_{t+k})}{\sum_{t=1}^n a_t^2} \dots (1-29)$$

إذ ان :

a_t : تمثل البقاوي .

وللتحقق من صحة كفاءة الانموذج نختبر قيمة $r_k(a_t)$ فاذا كانت قيمته ضمن حدود الثقة

$$\left(-\frac{1.96}{\sqrt{n}} \leq r_k(a_t) \leq \frac{1.96}{\sqrt{n}} \right) ، يعد الانموذج جيداً وكفوفاً بمعنى ان السلسلة الزمنية تمتلك اقل خطأ [5]$$

2-3-6-1 اختبار بوكس - بيرس Box-Pierce test [17][13]

وضع الباحثان Box وPierce عام 1970 اختباراً لاخطاء الانموذج المشخص وبيان مدى كفاءة الانموذج المحدد باستعمال معاملات الارتباط الذاتي للبقاوي والتي تعطى بالعلاقة الآتية :

$$Q_{B-P} = n \sum_{k=1}^m \hat{\rho}_k^2 \dots (1-30)$$

اذ ان :

$\hat{\rho}_k$: تمثل مقدار الارتباط الذاتي المقدر من بيانات العينة عند الازاحة K .

n : تمثل عدد مشاهدات الانموذج المشخص .

M : اكبر ازاحة مده ازاحة للارتباطات الذاتية.

K : تمثل عدد الازاحات

ومن ثم يتم مقارنة Q المحسوبة ب χ^2 الجدولية بدرجة حرية $(m - p - q)$ ومستوى دلالة (α) معين .

فاذا كانت Q المحسوبة اصغر من القيمة الجدولية هذا يدل على عدم رفض فرضية العدم ، ويشير الى ان الانموذج ملائم وان اخطاء الانموذج عشوائية . اما اذا كانت Q المحسوبة اكبر من القيمة الجدولية فهذا يعني الانموذج غير ملائم لبيانات السلسلة الزمنية وفي هذه الحالة لا بد من اعادة المرحلة الاولى اي تشخيص انموذج اخر لتمثيل السلسلة وتقدير معالمها .

1-3-6-3 اختبار لجنج- بوكس Ljung and Box test [14][17]

قام الباحثان (ljung ,G.M and Box ,G.E.P) في عام 1978 بوضع اختبار على غرار اختبار Box and Pierce عام 1970 وهو اختبار احصائي يستعمل لاختبار اي مجموعة من الارتباطات الذاتية في السلسلة الزمنية كونها تختلف معنويًا عن الصفر ام لا. ويمكن استخدام هذا الاختبار ايضا للتحقق من عشوائية البيانات باستعمال مجموعة من الاراحات وصيغة الاختبار هي :

$$Q_{(L-B)} = n(n + 2) \sum_{k=1}^m \frac{\hat{p}_k^2}{(n-k)} \dots\dots\dots(1-31)$$

ويتم اجراء المقارنة نفسها للاختبار السابق في اختبار $Q_{(B-p)}$ وان هذه الصيغة تعطي نتائج جيدة وكفوة.

1-4-6-4 المرحلة الرابعة (التنبؤ) Forecasting [20]

ان المرحلة الاخيرة من مراحل منهجية (Box-Jenkins) هي التنبؤ وفي هذه المرحلة يتم ايجاد القيم المستقبلية للسلسلة الزمنية عن طريق استعمال النموذج الحالي والمقدر في فترة زمنية معطاه وذلك تبعا لأصغر خطأ ممكن، والتنبؤ بالقيم المستقبلية لمشاهدات السلسلة الزمنية هو عبارة عن التوقع الشرطي (Conditional Expectation) في المدة $(t+1)$ عن الزمن (t) . ان هذا التنبؤ يتم بعد تقدير معالم النموذج $ARIMA(P,d,q)$ ، والذي يكون قد تجاوز مختلف مراحل الاختبارات السابقة ، فأنا نستطيع التنبؤ بالقيم المستقبلية $(L=1,2,\dots)$ لمدة قادمة وباستعمال صيغة معادلة الفروق لهذا النموذج وكلاتي :

$$X_t = \phi_1 X_{t-1} + \dots + \phi_{p+d} X_{t-p-d} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} + e_t \dots(1-32)$$

وبكتابة المعادلة عند المدة $(t+1)$ واخذ التوقع الشرطي عند المدة (t) نحصل على معادلة التنبؤ عند زمن الاصل (t) والى $(L=1,2,\dots)$ مدة تنبؤيه .

اي ان :

$$E(X_{t+l}) = \phi_1 E(X_{t+l-1}) + \phi_2 E(X_{t+l-2}) + \dots + \phi_{p+d} E(X_{t+l-p-d}) - \theta_1 E(e_{t+l-1}) - \theta_2 E(e_{t+l-2}) - \dots - \theta_q E(e_{t+l-q}) + E(e_{t+l}) \dots\dots(1-33)$$

2- الجانب التطبيقي :

1-2 بيانات البحث

بيانات البحث تمثل كمية الطاقة الكهربائية الشهرية الضائعة مقاسة ب كيلو واط /ساعة (kwh) لمحافظة كربلاء المقدسة وللمدة الزمنية من كانون الثاني 2012 الى كانون الأول 2016 . وقد تم الحصول على البيانات الخاصة بالبحث من وزارة الكهرباء / مديرية توزيع كهرباء كربلاء المقدسة والجدول رقم (1-2) يبين كمية الطاقة الكهربائية الضائعة للمدة الزمنية المذكورة انفاً .

الجدول (1-2) كمية الطاقة الكهربائية الشهرية الضائعة (كانون الثاني 2012 ولغاية كانون الاول 2016) المقاسة بكيلو واط / ساعه

السنة / الشهر	2012	2013	2014	2015	2016
كانون ثاني	30391327	62910756	105461499	118937036	183844255
شباط	34236531	45472252	45895269	58944177	104393106
اذار	15023204	29577121	34358757	91382834	87039040
نيسان	43264509	38268764	62396609	58214617	74476066
ايار	26909936	51743043	105770326	140573401	109194884
حزيران	34567324	63484626	142907824	170566887	195242706
تموز	46947592	78345065	123367202	197424161	232344370
اب	40957748	72941125	114586374	218302229	231694630
ايلول	59715107	60522705	121736132	181651258	203692295
تشرين أول	31838536	48832600	74549025	129357804	95137846
تشرين ثاني	47160734	29631641	71297390	105032021	150526140
كانون أول	52106952	59113988	157037184	183606859	173149708

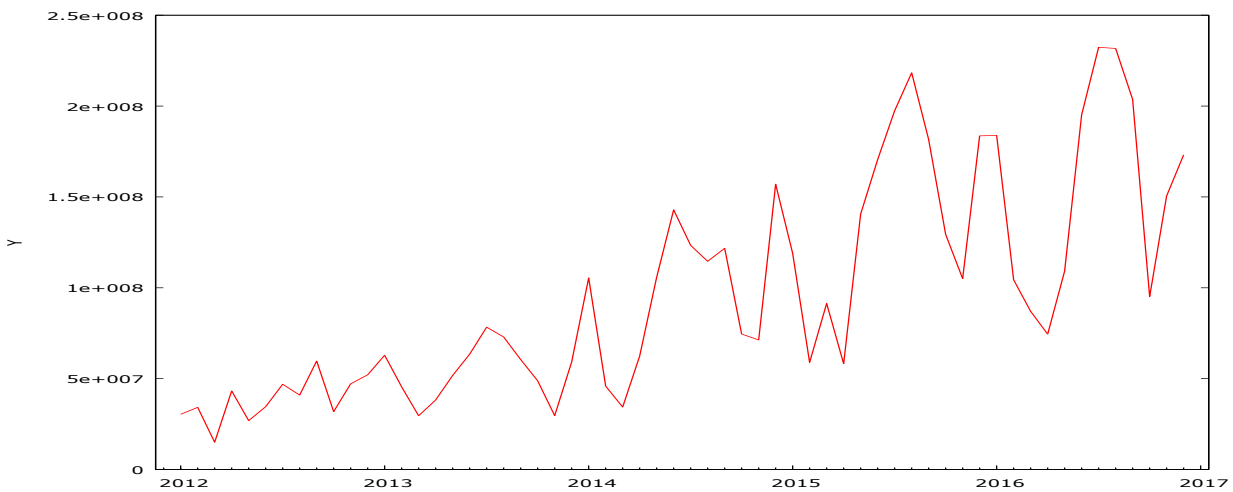
المصدر: وزارة الكهرباء ، مديرية توزيع كهرباء كربلاء المقدسة .

لوحظ ان الضائعات لعام 2016 تستحوذ على نسبة 31% من الطاقة الكهربائية المنتجة وقد ترتفع هذه النسبة الى 58% في بعض الاشهر والتي تسمى بأشهر الذروة مثل (كانون ثاني، تموز، اب)

2-2 استقراريه السلسلة:

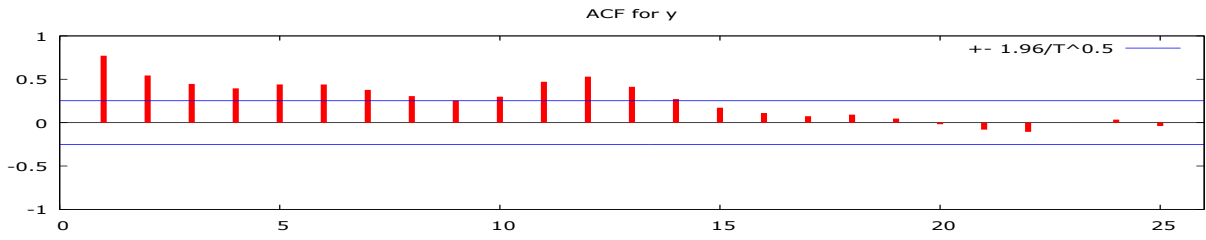
بعد جمع البيانات والتي تعد المرحلة الاولى من منهجية (Box- Jenkins) نرسم سلسلة البيانات (time Plot) باستخدام البرنامج الاحصائي Gretl كما في جدول (1-2) الممثلة لكمية الطاقة الكهربائية الضائعة للتعرف على سلوك السلسلة وخصائصها الأولية والشكل (1-2) يمثل رسم السلسلة :

الشكل (1-2) سلسلة كمية الطاقة الكهربائية الشهرية الضائعة (كانون الثاني 2012 ولغاية كانون الاول 2016)

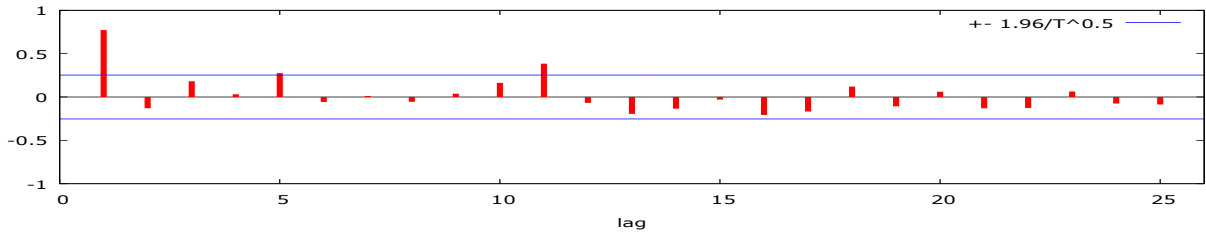


عن طريق الشكل (1-2) نلاحظ عدم استقرار السلسلة الزمنية ولمزيد من الدقة نرسم كل من دالة الارتباط الذاتي ACF والارتباط الذاتي الجزئي PACF وكما في الشكل (2-2) :

الشكل (2-2) أ- دالة الارتباط الذاتي ACF

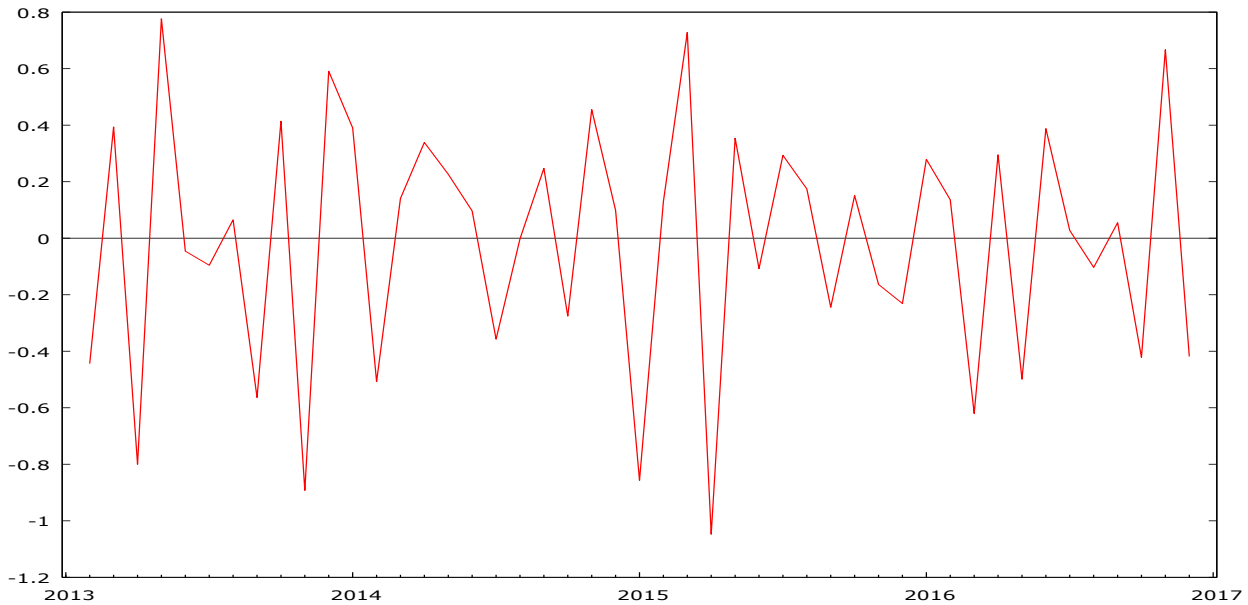


الشكل (2-2) ب-دالة الارتباط الذاتي الجزئي PACF



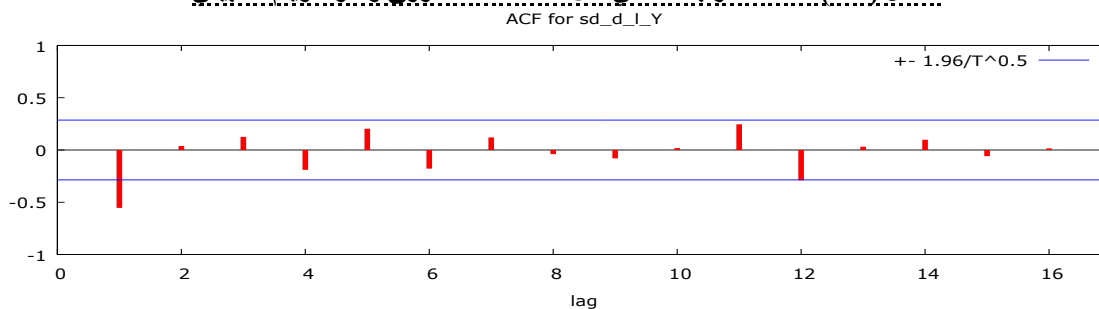
اذ نلاحظ من الشكل (2-2) بان معاملات الارتباط الذاتي ACF خارج حدي الثقة ($\pm \frac{1.96}{\sqrt{T}} = 0.253$) بمستوى معنوية (95%) وكذلك معاملات الارتباط الذاتي الجزئي PACF خارج حدود الثقة لمعاملات الارتباط الذاتي الجزئي وهذا مؤشر على عدم استقرار السلسلة ، لذا ترفض فرضية العدم التي تشير إلى تساوي معاملات الارتباط الذاتي مع بعضها ومساواتها للصفر اي استقرار السلسلة وتقبل الفرضية البديلة بعدم استقرار السلسلة. ويمكن ملاحظة ظهور قمة عند الازاحة (12 و 24) وهذا مؤشر على ضرورة اخذ التغيرات الموسمية في الاعتبار عند اختيار الانموذج الملائم .

الشكل (3-2) سلسلة بيانات الطاقة الكهربائية الضائعة بعد الفرق الاعتيادي الاول و الفرق الموسمي الاول واللوغاريتم الطبيعي

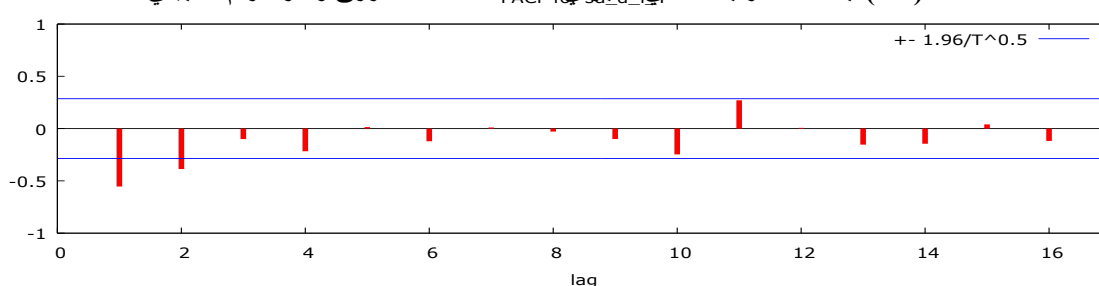


وعن طريق الشكل (3-2) المذكور انفاً يمكن ملاحظة الاستقرار في السلسلة بعد أخذ الفرق الاعتيادي الاول والفرق الموسمي واللوغاريتم الطبيعي ويمكن رسم كل من دالة الارتباط الذاتي ACF والارتباط الذاتي الجزئي PACF لسلسلة الفروق واللوغاريتم الطبيعي للتأكد من الاستقرار والرسم (4-2) يوضح ذلك :

الشكل (4-2) أ- دالة الارتباط الذاتي ACF لسلسلة الفروق واللوغاريتم الطبيعي



الشكل (4-2) ب- دالة الارتباط الذاتي الجزئي PACF لسلسلة الفروق واللوغاريتم الطبيعي



2-3 تشخيص وتقدير الأنموذج واختيار الأنموذج الأفضل

لتحديد رتبة الأنموذج (ARIMA) الملائم عن طريق دراسة ومقارنة السلوك النظري لدالتي (ACF), (PACF) لعدد من نماذج ARIMA الشائعة الاستعمال وهذا الموضوع يتطلب معرفة وخبرة طويلة في مجال السلاسل الزمنية وبشكل أدق يتم توفيق عدد من النماذج واختيار الأنموذج الأفضل على وفق معايير المفاضلة للنماذج المعنوية المعلمات والنماذج المقترحة هي كما موضحة في الجدول رقم (2-2)

الجدول (2-2) يمثل نماذج (Box-Jenkins) المقترحة لسلسلة كمية الطاقة الكهربائية الضائعة

Model	AIC	SBC	H-Q	MSE	MAE	MAPE
SARIMA(0,1,1)×(0,1,1) ₁₂	30.597	36.148	32.68	0.101	0.233	1.286
SARIMA(1,1,0)×(0,1,1) ₁₂	36.481	42.032	38.57	0.114	0.266	1.461
SARIMA(2,1,3)×(0,1,0) ₁₂	34.083	45.184	38.26	0.095	0.247	1.354

أن أفضل أنموذج من نماذج بوكس جنكنز هو SARIMA(0,1,1)×(0,1,1)₁₂ إذ نلاحظ انه يمتلك اقل قيمة لثلاثة معايير وهي معيار حنان كوين (H-Q) ومعيار معلومة اكيكي (AIC) ومعيار معلومة شوارتز (SBC) والتي تستعمل للمفاضلة بين النماذج المقترحة، والجدول (3-2) يمثل تقدير معلمات الانموذج الافضل ومعنويتها واخطائها المعيارية .

الجدول (3-2) معلمات الانموذج SARIMA(0,1,1)×(0,1,1)₁₂ ومعنويتها

	Coefficient	std. error	z	p-value
theta_1	-0.792993	0.104009	-7.624	2.45 e -014***
theta_1	- 0.407735	0.2064468	- 1.975	0.0483**

الأنموذج التقديري هو :

$$(1-B)Y_t = \epsilon_t - 0.792993\epsilon_{t-1} - 0.407735\epsilon_{t-2}$$

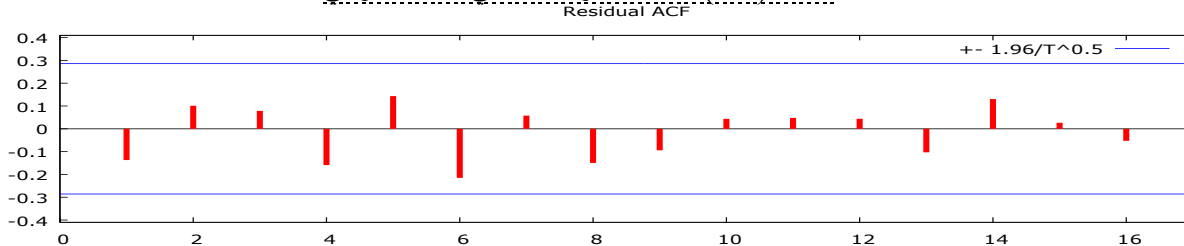
4-2 اختبار دقة الإنموذج

بعد تشخيص الإنموذج وتحديد درجته وتقديره لابد من التأكد من صحة ملائمة الأنموذج وكفاءته وتم ذلك عن طريق :

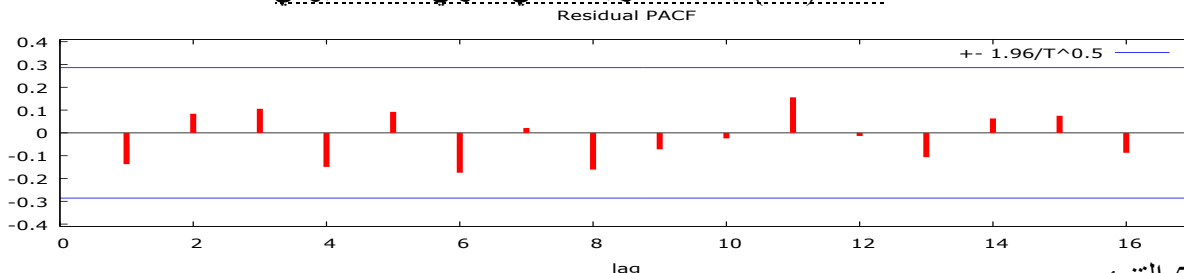
1-4--2 اختبار البواقي

تم استخراج معاملات الارتباط الذاتي والارتباط الذاتي الجزئي للبواقي (الأخطاء) للإنموذج المقدر ورسمها ونلاحظ من الشكل (5-2) ان قيم معاملات الارتباط الذاتي للبواقي (Residual ACF) جميعها تقع ضمن حدود الثقة مما يعني ان سلسلة البواقي عشوائية وان الإنموذج المستعمل جيد وملائم . :-

الشكل (5-2) أ- دالة الارتباط الذاتي ACF للبواقي



الشكل (5-2) ب- دالة الارتباط الذاتي الجزئي PACF للبواقي



5-2 التنبؤ

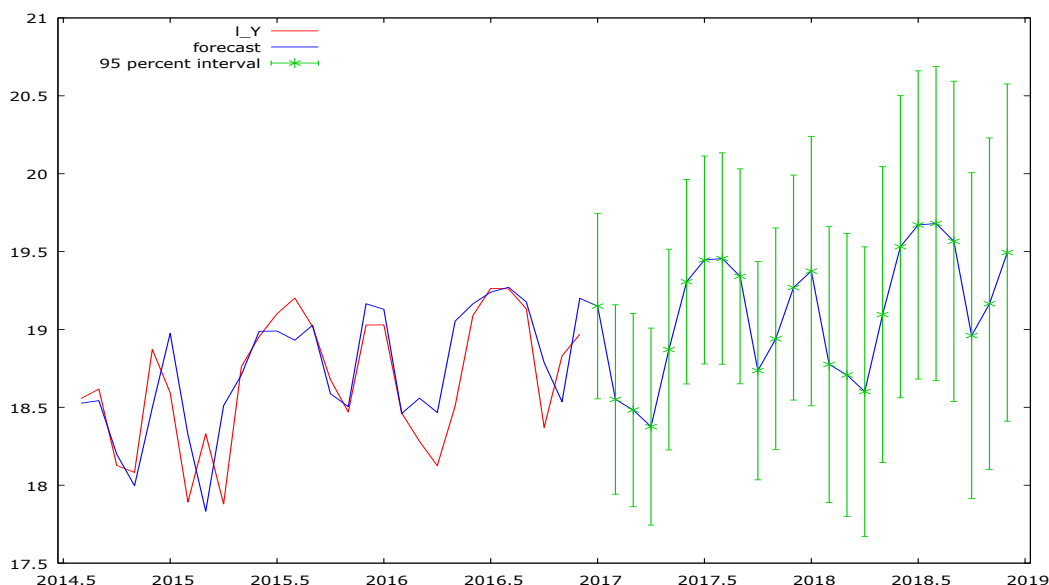
بعد اجتياز الأنموذج لاختبارات التشخيص يصبح بالإمكان استعمال الأنموذج للتنبؤ والجدول رقم (4-2) يمثل القيم التنبؤية للمدة الزمنية (2017-2018) لسلسلة كمية الطاقة الكهربائية الضائعة وان الشكل (4-2) يمثل رسم القيم التنبؤية للسلسلة وقد تم الحصول على هذه القيم التنبؤية باستعمال الأنموذج SARIMA(0,1,1)×(0,1,1)₁₂. من النتائج القيم المتنبأ بها نلاحظ ان الضائعات من الطاقة الكهربائية لمدينة كربلاء المقدسة تزداد للأشهر القادمة (كانون الثاني 2017-كانون الاول 2018)

الجدول (4-2) المعدلات الشهرية المتنبأ بها وحدود الثقة 95% لعامي (2017 و2018) المقاسة ب كيلو واط/ ساعة

الشهر	القيم المتنبأ بها لعام 2017	القيم المتنبأ بها لعام 2018
كانون الثاني	207,370,374	259,631,013
شباط	113,888,334	142,590,009
اذار	106,421,693	133,241,656
نيسان	95,655,332	119,762,000
ايار	156,863,569	196,395,688
حزيران	242,466,472	303,571,887
تموز	278,867,412	349,146,443
اب	281,419,782	352,341,700
ايلول	251,081,212	314,357,679
تشرين اول	137,018,156	171,548,915
تشرين ثاني	168,105,783	210,470,909
كانون اول	233,529,742	292,400,794

الشكل (2-6) يمثل رسم القيم التنبؤية لسلسلة كمية الطاقة الكهربائية الصانعة وفق إنموذج

$$\text{SARIMA}(0,1,1) \times (0,1,1)_{12}$$



4-1 الاستنتاجات :

- 1- تعد الصانعات في الطاقة الكهربائية من الظواهر غير الحضرارية والسلبية والتي لها تأثير مباشر على مستويات تجهيز الكهرباء في الشبكة الوطنية على مستوى المحافظات كافة وبشكل خاص محافظة كربلاء المقدسة إذ لوحظ أنها تستحوذ على ما نسبته 31% من الطاقة الكهربائية المنتجة وقد ترتفع هذه النسبة إلى 58% في بعض الأشهر والتي تسمى بأشهر الذروة مثل (كانون ثاني، تموز، اب).
- 2- نلاحظ من خلال دراسة السلسلة الزمنية و الاختبارات الاحصائية ان السلسلة غير مستقرة وان هناك اتجاه عام واضح في السلسلة فضلا عن احتوائها على المركبة الموسمية حيث انها تعيد نفسها كل 12 شهراً .
- 3- تم اختيار افضل نموذج من بين النماذج الممكنة اعتمادا على معايير المفاضلة (AIC, BIC, H-Q). وقد تم فحص ملائمة النموذج المقترح احصائيا من خلال الاختبارات التالية : معنوية المعالم المقدرة وتحليل دالة الارتباط الذاتي للبقايا والتوزيع الطبيعي للبقايا .
- 4- وجد ان النموذج الكفؤ والملائم لتمثيل بيانات السلسلة الزمنية هو النموذج الموسمي المضاعف $(0,1,1)_{12}$ SARIMA(0,1,1) .

4-2 التوصيات :

- من خلال النتائج التي تم التوصل اليها نوصي بما يأتي
- 1- اصدار قوانين صارمة ضد المتلاعبين بمقاييس الطاقة والربط غير القانوني وحالات التجاوز على الشبكة والزام المتجاوزين بدفع غرامات مالية كبيرة للحد من هذه الظاهرة السلبية.
 - 2- نظرا لأهمية هذا القطاع الحيوي لا بد من استعمال نماذج بوكس جنكيز للتنبؤ بالقيم المستقبلية ، إذ امكن الحصول على تقديرات جيدة للمؤشرات الشهرية للمدة المراد التنبؤ بها ويمكن ان تعطي المسؤولين المعنيين مؤشرات ذات أهمية بالغة.
 - 3- اعتماد النموذج الذي تم التوصل اليه النموذج الموسمي المضاعف $(0,1,1) \times (0,1,1)_{12}$ SARIMA عن طريق نماذج بوكس جنكيز للاستفادة منه في التخطيط للفترات القادمة كونه يعطي نتائج افضل .
 - 4- تعميم هذه الدراسة الى دراسات مناظرة على مستوى المحافظات الأخرى واجراء مقارنة بينهما

المصادر

1. الركابي ، مخلص عبد الله ، "استخدام منهجية بوكس و جنكنز في التنبؤ بأعداد المعتمرين من الخارج شهريا " ، معهد خادم الحرمين الشريفين لأبحاث الحج والعمرة ، جامعة ام القرى، 2014
2. الشاروط ، محمد حبيب ، و صاحب ، دعاء عبد الكريم، " التنبؤ باستهلاك الطاقة الكهربائية في محافظة القادسية باستخدام نماذج السلاسل الزمنية ونماذج الشبكات العصبية الاصطناعية" ، مجلة القادسية لعلوم الحاسوب والرياضيات ، المجلد السادس ، العدد الاول، 2014.
3. شومان ، عبد اللطيف حسن ، والصراف ، نزار مصطفى ، " السلاسل الزمنية والأرقام القياسية " ، الناشر دار الدكتور للعلوم الإدارية والاقتصادية ، بغداد، 2013 .
4. كنيهر ، عباس لفته ، "بناء نموذج سلاسل زمنية للتنبؤ باعداد المرضى المراجعين لمستشفى الزهراء في محافظة واسط " ، بحث منشور في مجلة كلية الادارة والاقتصاد ، جامعة واسط ، 2010 .
5. محمد، شيخي ، " طرق الاقتصاد القياسي محاضرات وتطبيقات " ، دار الحامد للنشر والتوزيع ، عمان، 2012
6. المشهداني ، ايمان محمد ، وكاظم ، جاسم محمد ، " استعمال نماذج بوكس و جنكنز للتنبؤ بأعداد المرضى المصابين بمرض التهاب الكبد الفيروسي في العراق " ، مجلة العلوم الاقتصادية والإدارية ، المجلد 22 ، العدد 89 ، 2016 .
7. مولود حشمان ، " السلاسل الزمنية وتقنيات التنبؤ قصيرة المدى" ، ديوان المطبوعات الجامعية الجزائر، 2010 .
8. الناصر ، عبد المجيد حمزه ، ورشيد ، ظافر حسين، " الارقام القياسية وتحليل السلاسل الزمنية " ، 2013
9. والتر فاندل ، "السلاسل الزمنية من الوجة التطبيقية ونماذج بوكس جنكنز " ، تعريب ومراجعة د. عبد المرضي ، حامد عزام ، د. احمد حسين هارون ، دار المريخ – السعودية ، 1990 .

المصادر الاجنبية

10. Akaike, Hirotugu , " A Bayesian Extension of the Minimum AIC Procedure of Autoregressive Model Fitting", Biometrika, Vol. 66, No. 2, PP. 237-242,1979
11. Anderson, t.w, "The statistical Analysis of Time Series", John Wily, New york. 1971.
12. Box G . E . P , and Jenkins , G . M , "Time Series Analysis and control" , San Francisco , Holden Day San Francisco , 1976
13. Box, G. E. P. and Pierce, David A. , "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models", JASA, Vol. 65, No. 332,1970
14. Brockwell. P. J, & Davis . A. R , "Introduction to Time Series and Forecasting ", Second Edition", springer –verlag, New York, USA, , 2002.
15. Cryer, D. & Chan ,K , "Time Series Analysis With Applications in R", Second Edition, Springer Science +Business Media, LLC, New York, USA, 2008
16. Enders, W, " Applied Econometric Time Series", John Wiley and Sons Inc New york, 1995
17. Gujarati .D.N & Porter . D. C , "Basic Econometrics 5 th" , Published by McGraw-Hill/Irwin, a business unit of The McGraw-Hill Companies, New York, 2009
18. TSAY. R.S, "Analysis of Financial Time Series", Second Edition, John Wiley & Sons, Inc, Printed in the United States of America., (2005).
19. Voind ,H.D , "Time Series analysis", Economic Fordham University , Bronx ,New York, . 1999 .
20. Wei, William ,W.S, " Time Series Analysis Univariate and Multivariate Methods" ,2nd , Addison ,Wesley publishing Company, 2006 .