استعمال نماذج بوكس وجنكنز للتنبؤ باعداد المرضى المصابين بمرض التهاب الكبد الفايروسي في العراق

م. د. ايمان محمد المشهداني / كلية الادارة والاقتصاد / جامعة بغداد
الباحث/ جاسم محمد كاظم

المستخلص

يعد أسلوب تحليل السلاسل الزمنية من الطرائق الإحصائية المهمة المتبعة في هذا المجال ، فدراسة ظاهرة معينة على فترة زمنية معينة وتحليلها يمكننا التعرف على نمطها والعوامل المؤثرة فيها واستعمالها للتنبؤ بالقيم المستقبلية للظاهرة والذي يساعد في وضع إسلوباً للتنبؤ في وضع الخطط السليمة للتنمية الاقتصادية.

يهدف البحث الى اختيار أفضل إنموذج للتنبؤ بعدد الإصابات بالتهاب الكبد الفايروسي بإستعمال نماذج بوكس – جينكنز غير الموسمية في التنبؤ المستقبلي .

تم جمع البيانات من وزارة الصحة / قسم الاحصاء الصحي والحياتي للمدة (من كانون الثاني ٢٠٠٩ ولغاية كانون الاول ٢٠١٣) وتم استعمال البرنامج الاحصائي MINITAB16 لتحليل السلسلة الزمنية ومن خلال معايير المفاضلة (معيار معلومات بيز BIC ، معيار اكيكي AIC) تبين ان افضل إنموذج لعدد المصابين بالتهاب الكبد الفايروسي في العراق هو الإنموذج ARIMA(1,1,0) .

المصطلحات الرئيسية للبحث/ نماذج بوكس وجنكنز - التهاب الكبد الفايروسي.

*البحث مستل من بحث الحصول على الدبلوم العالي في الاحصاء التطبيقي .

مجلة العلوم الاقتصادية والإدارية
المجلد ٢٢ العدد ٨٩
الصفحات ٤٠٧-٤٢٤

المقدمة

يعد مرض التهاب الكبد الفيروسي احد أهم الأمراض الانتقالية المسببة للمراضة والوفيات في العالم بشكل عام والدول النامية ومنها العراق ، وتعد أنواعه الخمسة (A,B,C,D,E) متوطنة في العراق وبدرجات مختلفة لذلك ادخل برنامج السيطرة على هذا المرض ضمن برامج الصحة العامة ومنذ بداية السبعينيات من القرن الماضي .

ان السلاسل الزمنية باساليبها المختلفة التقليدية منها والحديثة تعد ادوات علمية مهمة تلعب دوراً مهماً في التنبؤ بقيم اتجاه وسلوك الظاهرة خلال فترة من الزمن وبناء إنموذج احصائي تنبؤي لذلك يمكن من خلاله السيطرة على سلوك الظاهرة مستقبلاً

هدف البحث

يهدف البحث الى التنبؤ بعدد الإصابات بمرض التهاب الكبد الفايروسي من خلال الحصول على أفضل إنموذج من بين نماذج بوكس وجنكيز ومن ثم التمكن من وضع الخطط المستقبلية في علاج وتوفير الأدوية والمستلزمات الأخرى الخاصة بالمرض .

عينة البحث

تمثلت عينة البحث بالبيانات الشهرية لعدد المصابين بالتهاب الكبد الفايروسي في العراق عدا اقليم كردستان للمدة من (كانون الثاني ٢٠٠٩) ولغاية (كانون الاول ٢٠١٣) وقد تم اخذ عينة البحث من وزارة الصحة – قسم الاحصاء الصحي والحياتي ، تمثل مجتمع البحث بالمصابين بالتهاب الكبد الفايروسي في العراق عدا اقليم كردستان.

الاستعراض المرجعي

هناك الكثير من الدراسات والبحوث المتعلقة بمسألة استعمال السلاسل الزمنية والحصول على نماذج يتم من خلالها التنبؤ بجميع الظواهر وسوف نستعرض أهم وابرز هذه الدراسات والبحوث :

في عام (١٩٧٠) ظهرت اسهامات (Box, Jenkins) [18] الواسعة في مجال السلاسل الزمنية حيث استنبطا أسلوباً ذا مزايا عديدة في تحليل السلاسل الزمنية بصورة عامة والنماذج المختلطة (ARIMA) بصورة خاصة من حيث التكهن والسيطرة ، وكذلك اقترن اسماهما بصورة متصلة مع نماذج (ARIMA) في كتابهم الشهير "Time series analysis: forecasting and control".

في عام ١٩٨١ قام الباحث عبد الرسول، محمود جواد [12] بدراسة احصائية تطبيقية للمقارنة بين النماذج الاسية ونماذج بوكس وجنكنز في التوقعات المستقبلية مع تطبيق عملي على السلاسل الزمنية لكميات المبيعات الشهرية لعدة مجاميع من الادوية وقد توصل الى ان نماذج بوكس وجنكنز الموسمية افضل من نماذج التمهيد الاسي بالاعتماد على معياري (MSE) و (MPE) .

في عام ٢٠٠٥ استعمل العاني، احمد حسين بتال [8] نماذج (ARIMA) في التنبؤ الاقتصادي وتم إجراء تطبيق عملي على سلسلة زمنية لمنتج طبي (٧٠ مشاهدة) باستعمال البرنامج الإحصائي STATGRAPHICS وتبين من النتائج ان الإنموذج ARIMA(1,0,1) حقق قدرة تنبؤية أعلى من الإنموذج ARIMA(1,0,0) حسب اختبارات القدرة التنبؤية .

في عام (٢٠٠٧) قام الباحثان البياتي والمخلافي ، د.عصام حسين وفؤاد عبده اسماعيل [1] بنشر بحث استعمال اسلوب بوكس – جنكينز للتنبؤ بإنتاجية العمل في مصنع اسمنت عمران في القطاع الصناعي اليمني وتبين ان افضل إنموذج للتنبؤ هو الإنموذج المختلط ARMA(1,2) بحسب المعايير (AIC , MSE , SBS).

في عام ٢٠١٠ قام الباحثان مطر والياس، ظافر رمضان وانتصار ابراهيم [16] بنشر بحث تحليل ونمذجة السلسلة الزمنية لتدفق المياه الداخلة الى مدينة الموصل، دراسة مقارنة، وتبين بان نماذج بوكس جنكنز متفوقة على نموذج الشبكات العصبية الاصطناعية بحسب معياري (MSE, RMSE) .

وفي العام نفسه قامت الباحثة الجبوري، عبير حسن [3] بنشر بحث التنبؤ باسعار النفط العراقي لعام ٢٠١٠ باستعمال السلاسل الزمنية وتبين ان الإنموذج ARIMA(0,2,1) متفوقة على إنموذج هولت Holt من خلال المعايير (MAPE , MAE , RMSE) .

وفي العام نفسه قام الطائي ، فاضل عباس [7] بنشر بحث "التنبؤ والتمهيد للسلاسل الزمنية باستعمال التحويلات مع التطبيق " وتبين ان إنموذج ARIMA(5,0,2) افضل إنموذج للتنبؤ بمعدلات الامطار لامتلاكه اقل قيمة للمعيارين (MSE,AIC) وبعد اجراء التحويل بطريقة الجذر التربيعي كان الإنموذج الافضل هو ARIMA(1,0,0) لامتلاكه اقل قيمة للمعيارين (MSE,AIC) ، وعند اجراء تمهيد الاسي كان الإنموذج الافضل هو ARIMA(1,0,0) لامتلاكه اقل قيمة للمعيارين (MSE,AIC)

في عام ٢٠١٢ استخدمت الباحثة طعمة ، سعدية عبد الكريم [11] السلاسل الزمنية للتنبؤ بأعداد المصابين بالأورام الخبيثة في محافظة الانبار وتوصلت الى أن نموذج الانحدار الذاتي من الدرجة الثانية ARIMA(2,1,0) أفضل نموذج للتنبؤ .

وفي عام ٢٠١٤ قامت الباحث السلطاني ، حلى مثنى [4] بالتنبؤ بإنتاج محصولي الشلب والقطن في العراق باستعمال السلاسل الزمنية وتبين ان أفضل النماذج هو إنموذج ARIMA(1,1,1) والذي له اقل قيمة من المعايير الاتية (MSE , RMSE MAPE) .

وفي العام نفسه قامت الباحث عمران ، خالدة كريم [14] باستعمال نماذج بوكس جنكنز للتنبؤ بانتاج التمور في العراق وتوصلت الى ان الإنموذج ARIMA(2,1,1) هو الافضل حسب معياري MSE , AIC

قائمة الرموز والمصطلحات

الرمز	المصطلح باللغة الانكليزية	المصطلح باللغة العربية
ACF	Auto correlation function	دالة الارتباط الذاتي
PACF	Partial Auto correlation function	دالة الارتباط الذاتي الجزئي
AML	Approximate Maximum Likelihood.	الإمكان الأعظم التقريبية
EML	Exact Maximum Likelihood.	الإمكان الأعظم المضبوط
LS	Least Square.	المربعات الصُغرى
RMSE	Root Mean Square Error	الجذر التربيعي لمتوسط مربعات الخطأ
MAE	Mean Absolute Error	متوسط القيم المطلقة للخطأ
MAPE	Mean Absolute Percentage Error	متوسط نسب القيم المطلقة للخطأ
BIC	Bayesian information criterion	معيار معلومات بيز
AR	Auto Regressive	الانحدار الذاتي
MA	Moving Average	المتوسط المتحرك
ARIMA	Autoregressive-Integrated-Moving Average Models	نماذج الانحدار الذاتي المتكاملة مع المتوسطات المتحركة
AIC	Akaike Criterion	معيار أكيكي

السلاسل الزمنية (Time Series)[9]

أن الكثير من الظواهر ذات العلاقة بالزمن تسجل على فترات زمنية محددة ، غالباً ما تكون متساوية فقد تكون فترات زمنية سنوية أو ربع سنوية أو شهرية كما في هذا البحث أو أسبوعية أو يومية وهكذا، ومن الأمثلة على ذلك كثيرة مثل عدد المراجعين الى مستشفى سنوياً ، استهلاك الكهرباء لعدة شهور، درجات حرارة الجو في مدينة معينة يومياً كل ما سبق يعتبر أمثلة على متسلسلات زمنية .

وردت عدة تعاريف للسلسلة الزمنية منها :

- هي مجموعة من المشاهدات التي تتولد على التوالي خلال الزمن .
- هي مجموعة من المشاهدات المرتبطة مع بعضها يتم تسجيلها في فترات زمنية متعاقبة لظاهرة ما.

إنموذج بوكس جنكنيز (Box-Jenkins) [10],[12],[15]

من بين ابرز طرق التنبؤ الاقتصادي هي نماذج (ARIMA) Autoregressive-Integrated-Moving Average Models) أي نماذج الانحدار الذاتي المتكاملة مع المتوسطات المتحركة ، وتم صياغة هذه المنهجية من قبل (Box و Jenkins) عام ١٩٧٠م ولذلك تسمى بصيغة (Box-Jenkins) وتعتمد هذه المنهجية على الدمج بين نماذج الانحدار الذاتي (AR) والمتوسطات المتحركة (MA) .

١-إنموذج الانحدار الذاتي العام (General Autoregressive model: AR(p))

يعرف على انه القيمة الحالية للسلسلة الزمنية Y_t يعبر عنها بدلالة المجموع الموزون للقيم السابقة للسلسلة الزمنية (Y_{t-1} , Y_{t-2}, …) مضافاً إليها قيمة الخطأ الحالي (u_t) .
والصيغة العامة لإنموذج الانحدار الذاتي من الدرجة p الذي يرمز له [AR(p)] هي :

$$Y_t = m + a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots + a_p Y_{t-p} + u_t \quad \ldots\ldots\ldots(1)$$

إذ أن :
m : يمثل ثابت الإنموذج
$a_1, \ldots, a_p$ **: هي معلمات الإنموذج**
t : يمثل الزمن
$(u_1, u_2, u_3 \ldots, u_t)$ **تمثل سلسلة الاخطاء العشوائية بمتوسط مقداره صفر وتباين** σ_u^2 **وتغاير ذاتي مقداره صفر .**

٢- الإنموذج العام للأوساط المتحرك (General moving average model: MA(q))

في هذا الإنموذج يعبر عن القيمة الحالية للسلسلة الزمنية (Y_t) بدلالة المجموع الموزون للقيم السابقة للأخطاء (u_t, u_{t-1}, ..)، والصيغة العامة لهذا الإنموذج من الدرجة q الذي يرمز له (MA(q)) هي :

$$Y_t = m + u_t - b_1 u_{t-1} - b_2 u_{t-2} - \ldots - b_q u_{t-q} \quad \ldots\ldots\ldots(2)$$

اذ ان $b_1, \ldots, b_q$ **هي معلمات الإنموذج**
وان m يمثل ثابت الإنموذج
t: يمثل الزمن

٣- الإنموذج المختلط (Mixed Autoregressive – Moving average model(ARMA) **في هذا الإنموذج يتم جمع الإنموذجين السابقين ((MA(q)،AR(p)) للحصول على إنموذج له مرونة اكبر في تمثيل بيانات السلسلة الزمنية والصيغة العامة للإنموذج المختلط ((ARMA(p,q))هي :**

$$Y_t = m + a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots + a_p Y_{t-p} + u_t - b_1 u_{t-1} - \ldots - b_q u_{t-q} \quad \ldots\ldots(3)$$

m يمثل ثابت الإنموذج
$a_1, \ldots, a_p$ **: هي معلمات الإنموذج**
$b_1, \ldots, b_q$ **هي معلمات الإنموذج**
t : يمثل الزمن

منهجية بوكس جنكنز (Box – Jenkins approach(ARIMA) [13]

تتضمن منهجية بوكس – جنكنز مايأتي

أولاً - جمع البيانات [5]

افترض (بوكس وجنكنز) توفر على الأقل (٥٠) مشاهدة هي الحد الأدنى في السلسلة وهناك من يرى استعمال حجم عينة اقل يصل الى (٣٠) مشاهدة كحد أدنى لتحليل وبناء إنموذج للسلسلة الزمنية والاختبارات التي تساعد على تشخيص السلسلة هي رسم السلسلة والسلوك العام للسلسلة واختباري دالة الارتباط الذاتي والارتباط الذاتي الجزئي .

ثانياً - فحص استقرارية السلسلة [13],[6]

يقصد بالاستقرارية ان مشاهدات السلسلة الزمنية تتذبذب بشكل عشوائي حول متوسط ثابت، ويمكن القول ان السلسلة الزمنية المستقرة هي السلسلة التي تكون متجانسة زمنيا أي ان رسم السلسة في الفترة {t,t+h} يكون مطابقا لرسم السلسلة في فترة أخرى{s,s+h} وفي الحقيقة ان حالة الاستقرارية هي نادرة جداً وتوضع لغرض تسهيل التعامل الرياضي مع السلاسل الزمنية وهي على نوعين :

١- الاستقرارية التامة (strictly stationary): تكون السلسلة الزمنية $(Y_t ; t=1,2,\ldots,n)$ مستقرة بشكل تام اذا كان التوزيع المشترك لأية مجموعة من المشاهدات لايتأثر بازاحة كل الفترات الزمنية للمشاهدات الى الأمام أوالى الخلف بأية كمية صحيحة

ويمكن التعبير عن ذلك بطريقة الاحتمالات بالشكل :

$$\Pr(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m}) = \Pr(Y_{t_1+k}, Y_{t_2+k}, \ldots, Y_{t_m+k}) \quad \ldots\ldots\ldots(4)$$

إذ أن t تمثل أي فترة زمنية وان (m, k) تمثل أي زوج من القيم الصحيحة

٢- الاستقرارية من الدرجة الثاني (Second –order Stationary)

يقال للسلسلة الزمنية $(Y_t ; t=1,2,\ldots,n)$ انها ذات استقرارية من الدرجة الثانية اذا تحققت الشروط الاتية :

1- $$E(Y_t) = \mu \quad \ldots\ldots\ldots(5)$$

إذ أن μ يمثل ثابت الإنموذج ويكون ثابت لجميع قيم t

2- $$\operatorname{var}(Y_t) = E(Y_t - \mu)^2 = \sigma_y^2 = \gamma_o \ldots\ldots(7)$$

إذ أن $\hat{\gamma}_o$: هو تباين العملية العشوائية ويكون ثابتا لجميع قيم t

3- $$\operatorname{cov}(Y_t, Y_{t+k}) = E(Y_t - \mu)(Y_{t+k} - \mu) = \gamma_k \quad \ldots\ldots\ldots(9)$$

إذ أن γ_K هو التغاير الذاتي (Auto covariance) للعملية العشوائية عند الإزاحة k (lag k) ويكون ثابتا لجميع القيم الصحيحة الى k .

عدم الاستقرارية في السلاسل الزمنية(Non-Stationary Time Series) [4]

إن حالة الاستقرارية هي حالة نادرة جدا فكثير من السلاسل الزمنية التي تواجهنا في التطبيق العملي تكون غير مستقرة ، وللتعرف على حالة عدم الاستقرارية في السلاسل الزمنية يمكن فحص دالة الارتباط الذاتي اذ ان قيمها في هذه الحالة لا تؤول الى الصفر بعد الإزاحة الثانية والثالثة وإنما تبقى قيمها كبيرة لعدد من الإزاحات، وهناك نوعين من عدم الاستقرارية في السلسلة الزمنية :

الأول: هو عدم الاستقرارية حول المتوسط (Non stationary about the mean)

عندما لاتتذبذب السلسلة الزمنية حول متوسط ثابت فالسلسلة الزمنية تكون غير مستقرة حول المتوسط وفي هذه الحالة يتم اخذ الفروق المناسبة لتحقيق الاستقرارية في السلسلة الزمنية .

الثاني:هو عدم الاستقرارية حول التباين (Non Stationary around the variance)

وفي هذه الحالة نجد ان السلسلة الزمنية لا تتذبذب حول تباين ثابت، ومن اجل تحقيق الاستقرارية في السلسلة الزمنية يتم إجراء التحويل اللوغاريتمي أو الآسي (تحويلات القوى) (Power transformation) للسلسة الزمنية.

وتستعمل دالة الارتباط الذاتي (Auto Correlation Function) ودالة الارتباط الذاتي الجزئي (Partial Auto Correlation Function) للكشف عن استقرارية او عدم استقرارية السلسلة الزمنية :

أ ـ دالة الارتباط الذاتي (Auto Correlation Function):[15]

تعرف دالة الارتباط الذاتي بأنه مقياس لدرجة العلاقة بين قيم السلسلة مع نفسها لفترة زمنية بفرق ازاحات مختلفة، وهي ذات اهمية كبيرة لأنها احدى اساليب تحديد فيما اذا كانت العملية العشوائية مستقرة ام لا، إن الصيغة الرياضية لإيجاد دالة الارتباط الذاتي للعملية العشوائية بفرق ازاحة k هي :

$$\rho_k = \frac{\operatorname{cov}(Y_t , Y_{t+k})}{\sqrt{\operatorname{var}(Y_t)\ \operatorname{var}(Y_{t+k})}} \qquad \dots (11)$$

والسلسلة المستقرة تحتوي المشاهدات $(Y_1,Y_2,\dots,Y_n)$ لـ n من الأزمنة المتعاقبة .

يمكن الاستفادة من تحليل الارتباط الذاتي في هذه المجالات :

1-وجود الاستقرارية أو عدم الاستقرارية في السلسلة والمقصود بالاستقرارية عدم وجود نمو أو هبوط في المشاهدات بعبارة أخرى إنّ المشاهدات تكون أفقية حول محور السينات .

2 ـ تحديد الموسمية في السلسلة الزمنية ويقصد بالموسمية وجود نمط معين في السلسلة يعيد نفسه خلال فترات زمنية محددة، مثل ازدياد مبيعات سلعة معينة في الشتاء وانخفاضها في الصيف ، إذ تشير هذه المبيعات الى وجود نمط الموسمية طوله (١٢) شهراً.

ب ـ دالة الارتباط الذاتي الجزئي Partial Auto correlation function [15]

هناك دالة أخرى تسهم في تشخيص الإنموذج المناسب، تعرف هذه الدالة بدالة الارتباط الذاتي الجزئي ويشار إليها اختصاراً (PACF) وتعد هذه الدالة مؤشراً يقيس العلاقة بين Y_t و Y_{t+k} للسلسلة نفسها مع افتراض ثبوت بقية قيم السلسلة الزمنية ويساعد في عملية تحديد رتبة الإنموذج ونوعه، لذلك فهو احد الأدوات المستعملة في تحليل السلاسل الزمنية ويمكن تقديره بواسطة المربعات الصغرى أو مجموعة المعادلات التقريبية لتقدير الارتباط الذاتي الجزئي .

وأن الحالة العامة لمعادلات الارتباط الذاتي الجزئي ϕ_{kk} هي:

$$\varphi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \rho_{k-1} & \rho_{k-2} & \cdot & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \rho_{k-1} & \rho_{k-2} & \cdot & \cdots & \rho_1 & 1 \end{vmatrix}} \qquad \dots\dots\dots\dots\dots\dots(14)$$

إنّ المعلمات $(\phi_{KK}; k = 1,2,\dots)$ تسمى بدوال الارتباط الذاتي الجزئي (PACF)، إنّ (PACF) أداة مهمة أخرى فضلا عن دالة (ACF) في تحليل السلاسل الزمنية إذ تساعدنا في تحديد درجة الإنموذج (AR) أو (MA) وكذلك الإنموذج المختلط غير المستقر (ARIMA) الملائم لتمثيل بيانات السلسلة الزمنية.

ثالثاً - تحديد رتبة الإنموذج Model order selection criteria [2],[6],[13]

هناك نماذج مختلفة الدقة يمكن ان توفق في تحليل السلاسل الزمنية وان اختيار الإنموذج الاحسن لايكون سهلا في كثير من الاحيان، لذا فقد وضعت معايير عدة لمقارنة النماذج واختيار رتبة الإنموذج ، وتأتى اهمية اختيار رتبة الإنموذج من كون اختيار رتبة ادنى من الرتبة الفعلية للإنموذج يؤدي الى عدم اتساق معلمات الإنموذج بينما يؤدي اختيار رتبة اعلى من الرتبة الفعلية للإنموذج الى زيادة تباين الإنموذج وهذا يؤدي الى فقدان الدقة بسبب الزيادة في عدد المعلمات للإنموذج المختار.

وتستند معايير اختيار رتبة الإنموذج الى احصاءات البواقي الناتجة من مطابقة الإنموذج غير المتحيز .

ومن معايير اختيار رتبة الإنموذج :

أ- تحديد الرتبة من خلال ادوات التشخيص

وضع (B-J) اسلوبا لتحديد رتبة الإنموذج فضلا عن انه كونه اداة من ادوات التشخيص (معرفة نوع الإنموذج) اذ يتم تحديد إنموذج (ARMA) الملائم لها عن طريق دراسة ومقارنة السلوك النظري لدالتي (ACF),(PACF) لعدد من نماذج ARMA الشائعة الاستعمال، إنّ دالة الارتباطات الذاتية للإنموذج (AR(p)) تكون متناقصة بشكل أسي أو بشكل موجات جيبية متضائلة، أمّا دالة الارتباطات الجزئية فإنّها تنقطع بعد الإزاحة p أما دالة الارتباط الذاتي لإنموذج MA(q) تنقطع بعد الإزاحة q، أما دالة الارتباط الذاتي الجزئي فإنّها تتناقص بشكل أسي أو بشكل موجات جيبية متضائلة بينما ان دالة الارتباطات الذاتية للإنموذج المختلط بعد الإزاحة (q-p) تسلك سلوكا مشابها إلى دالة الارتباط الذاتي لإنموذج (AR(p) (تتناقص بشكل أسي أو بشكل موجات جيبية)، اما دالة الارتباط الذاتي الجزئي له بعد الإزاحة (p-q) فانها تتناقص بشكل أسي أو بشكل موجات جيبية (كما في دالة الارتباط الذاتي الجزئي لإنموذج ((MA(q)).

ب- معيار أكاي(Akaike) (AIC)

في عام ١٩٧٤م وضع اكاي معيار اعتمد فيه على ان يكون الفارق بين كثافة الإنموذج وبين الكثافة الحقيقية للمشاهدات ضئيلاً أو بعبارة أخرى تدنئة تباين الإنموذج مقارنة بزيادة عدد المعالم المقدرة وصيغته :

$$AIC = Log\left(\sigma_u^2\right) + 2V \qquad \ldots\ldots\ldots(15)$$

ج- معيار معلومات بيز Bayesian Information Criterion (BIC)

وهنالك طرائقُ أخرى لتقييم الإنموذج ومنها طريقة معيار معلومات بيز(Bayesian (Information Criterion : وصيغته هي :

$$BIC = 2Ln\left(\sigma_u^2\right) + V\ Ln\left(M\right) \qquad \ldots\ldots\ldots(16)$$

إذ (M و V) هي عدد مشاهدات السلسلة والعدد الكلي لمعلمات الإنموذج على التوالي ، و $\left(\sigma_u^2\right)$ يمثل متوسط مربعات الخطأ (MSE) .

على وفق هذا المعيار فالإنموذج الأفضل هو الإنموذج الذي يعطي اقل قيمة من (BIC)

رابعاً: تقدير معاملات النموذج

ان تقدير معلمات النموذج إذا كان نموذج انحدار ذاتياً لا تطرح اية مشكلة، حيث يمكن استعمال طريقة المربعات الصغرى الاعتيادية، وفي هذه الحالة فان اي برنامج إحصائي يعطي معلمات الانحدار الخطي المتعدد يفي بالغرض .

أما في حالة نموذج ARMA فإن تقدير المعلمات يصبح معقداً وتوجد عدة خوارزميات مقترحة لتقدير النموذج فعلى سبيل المثال يمكن استعمال طريقة المربعات الصغرى الاعتيادية او طريقة الامكان الاعظم .

خامساً :اختبار معنوية المعالم(Significant test parameters) [13]

يقصد باختبار معنوية المعالم اختبار صحة الإنموذج واختبار حد الخطأ وان البواقي (Residuals) الناتجة من تطبيق الإنموذج يجب أن تتوزع بشكل عشوائي ولمعرفة كون الارتباطات الذاتية تتوزع بشكل عشوائي أم لا فانه يمكن اللجوء إلى اختبار (Ljung-Box) واختبار البواقي:

١- اختبار (Ljung-Box)

ان اختبار (*Ljung-Box*) هو اختبار إحصائي يستعمل لاختبار أية مجموعة من الارتباطات الذاتية في السلسلة الزمنية كونها تختلف معنويا عن الصفر ام لا، ويمكن استعمال هذا الاختبار أيضاً للتحقق من عشوائية البيانات (إجمالاً) باستعمال مجموعة من الإزاحات .

وفي الحقيقة إن صيغة اختبار (Ljung-Box) هي صيغة معدلة لصيغة (Q_{BP}) اذ ان كل من (Ljung,G.M&Box,G.E.P) قام بتعديل صيغة اختبار (Q_{BP}) الأصلية التي اقترحها كل من (*Box & Pierce*) بالشكل الآتي:

$$Q = m(m+2)\sum_{i=1}^{k}\frac{r_i^2(u)}{(m-i)} \dots\dots (17)$$

اذ ان : $m=(n-d-SD)$

٢- اختبار البواقي

يتم استخراج معاملات الارتباط الذاتي والجزئي للبواقي (الأخطاء) للإنموذج المقدر ورسمها فإذا كانت جميع قيم معاملات الارتباط الذاتي للبواقي تقع ضمن حدود الثقة يعني أن سلسلة البواقي عشوائية وان الإنموذج المستعمل جيد وملائم .

سابعاً : التنبؤ (Forcasting)

بعد اجتياز الإنموذج لاختبارات التشخيص تأتي المرحلة الأخيرة في طريقة بوكس جينكنز (Box – Jenkins method) وبالإمكان استعمال الإنموذج الذي تم اختياره للتنبؤ بالقيم المستقبلية للسلسلة الزمنية (لبيانات خارج السلسلة الزمنية) .

إنموذج ARIMA لعدد المصابين بالتهاب الكبد الفايروسي

تم تطبيق نماذج (Box- Jenkins) من جمع البيانات وتحليلها باستعمال الحاسب الآلي وبالاعتماد على البرنامج الجاهز Minitab16 إذ تم تحليل السلسلة الزمنية لعدد المصابين بالتهاب الكبد الفايروسي وتشخيص الإنموذج الملائم لها ومن ثم اختيار أفضل إنموذج واستعمال هذا الإنموذج للتنبؤ.

جمع البيانات

عينة الدراسة تمثلت بعدد المصابين بالتهاب الكبد الفايروسي لجميع الأنواع في العراق عدا اقليم كردستان للمدة الزمنية من كانون الثاني ٢٠٠٩ الى كانون الأول ٢٠١٣ .

وقد تم الحصول على البيانات الخاصة بالبحث من وزارة الصحة / قسم الإحصاء الصحي والحياتي والجدول رقم (١) يبين بعدد المصابين بالتهاب الكبد الفايروسي لجميع الأنواع في العراق عدا اقليم كردستان للمدة من كانون الثاني ٢٠٠٩ الى كانون الأول ٢٠١٣

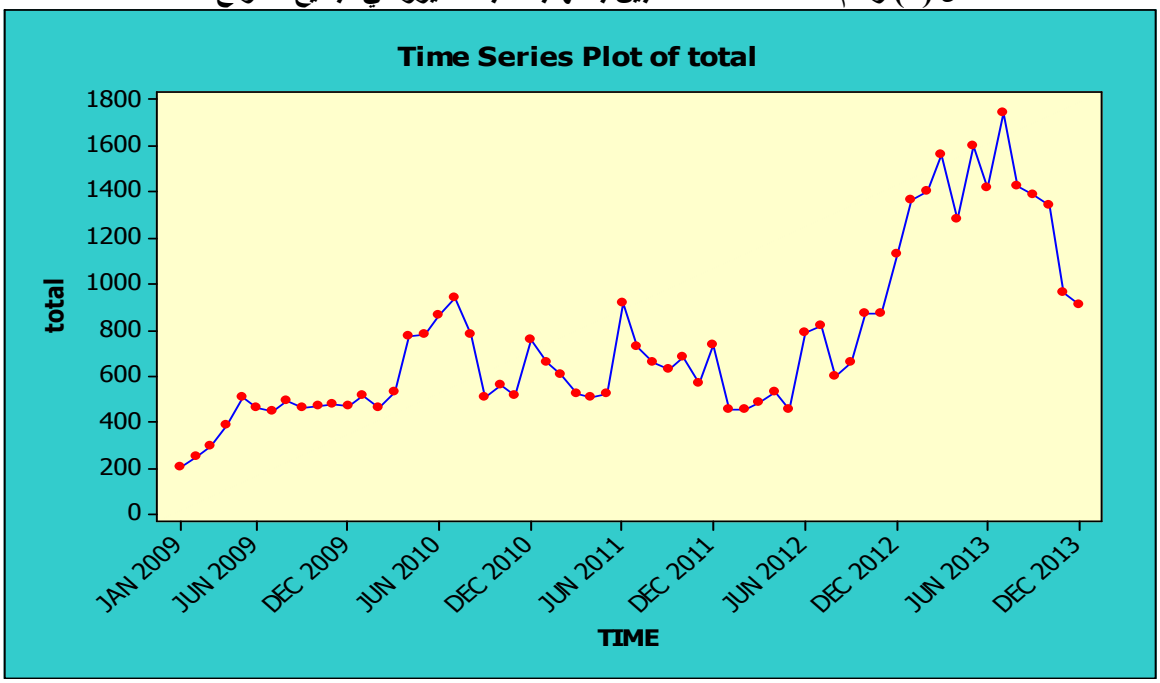
الجدول رقم (١) يمثل عدد المصابين بالتهاب الكبد الفايروسي في العراق عدا اقليم كردستان للمدة (كانون الثاني ٢٠٠٩ – كانون الأول ٢٠١٣)

	2009	2010	2011	2012	2013
كانون الثاني	205	512	657	458	1365
شباط	253	464	605	452	1400
اذار	292	533	520	488	1564
نيسان	389	770	210	527	1279
ايار	504	779	525	455	1596
حزيران	460	862	920	791	1416
تموز	447	943	729	817	1742
اب	491	779	656	599	1428
ايلول	464	511	626	658	1390
تشرين الاول	469	564	679	868	1339
تشرين الثاني	474	513	566	869	963
كانون الاول	466	757	734	1128	910

استقرارية السلسلة

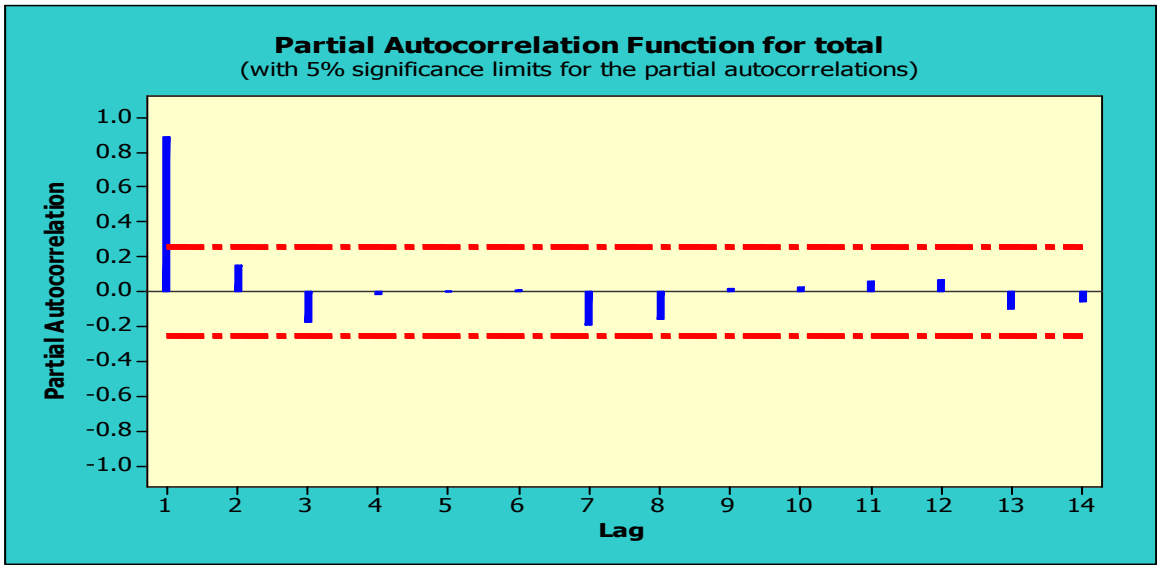
بعد جمع البيانات والتي تعد اول مرحلة من منهجية بوكس جنكيز نرسم سلسلة البيانات عدد المصابين بالتهاب الكبد الفايروسي لجميع الأنواع للتعرف على سلوك السلسلة والشكل (١) يمثل رسم السلسلة :

شكل (١) رسم سلسلة عدد المصابين بالتهاب الكبد الفايروسي لجميع الأنواع

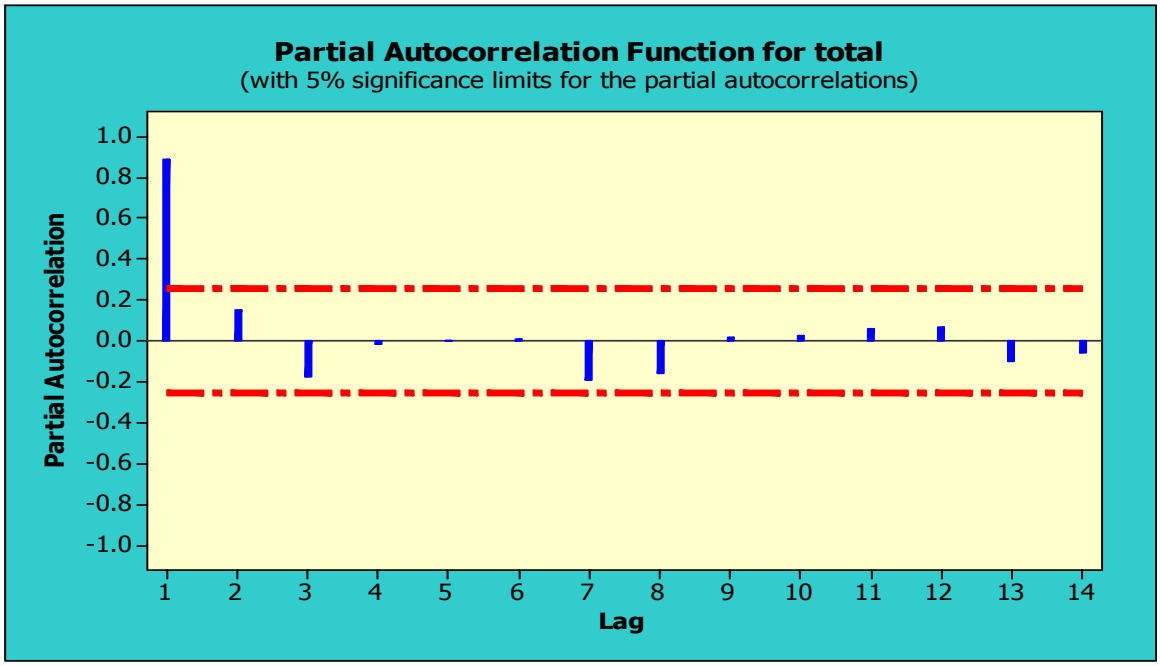


من خلال شكل رقم (١) نلاحظ عدم استقرار السلسلة الزمنية ولمزيد من الدقة نرسم كل من دالة الارتباط الذاتيACF والارتباط الذاتي الجزئي PACF على التوالي :

شكل رقم (٢) رسم دالة الارتباط الذاتي للسلسلة ACF

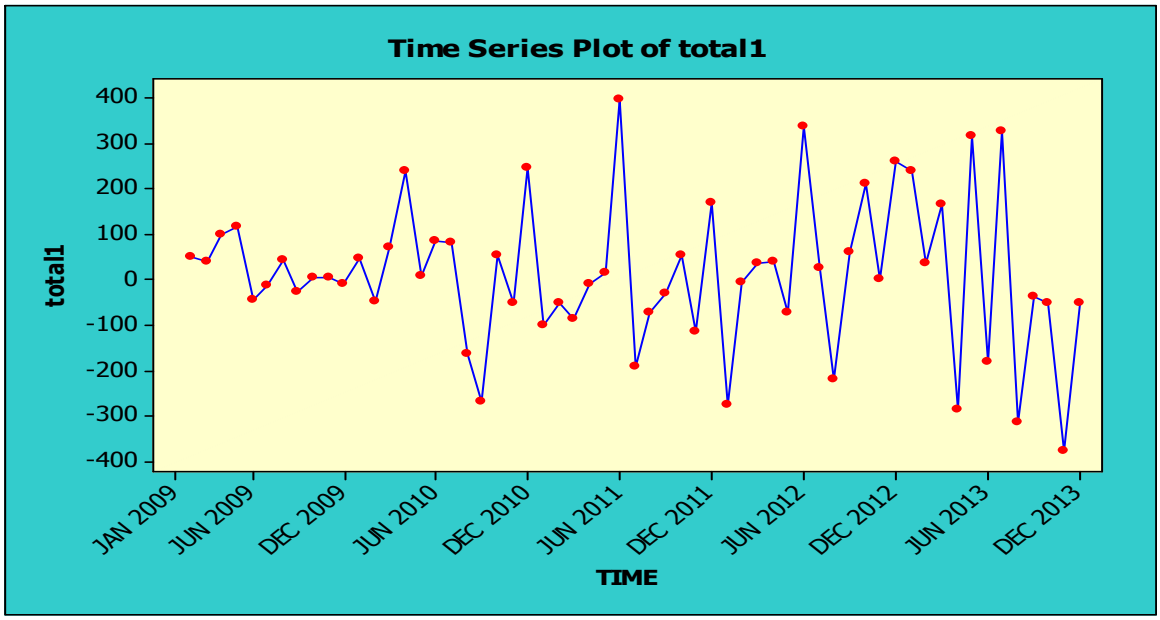


شكل رقم (٣) رسم دالة الارتباط الذاتي الجزئي للسلسلة PACF

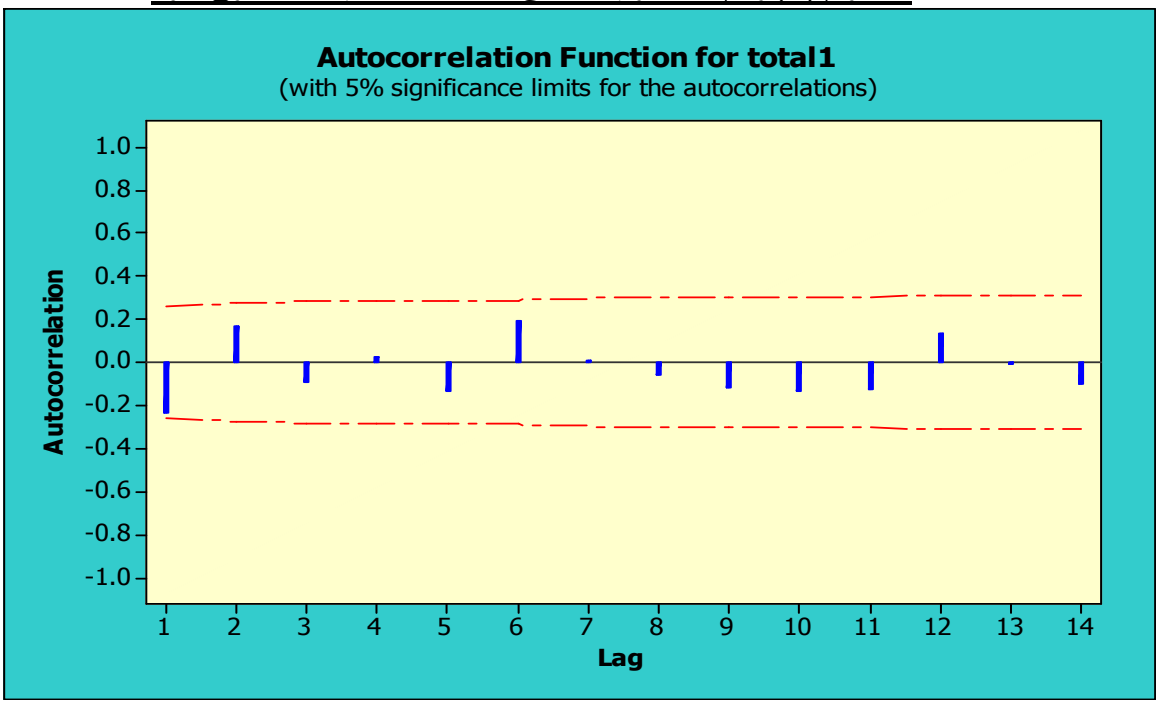


حيث نلاحظ من الشكل (٢) بان معاملات الارتباط الذاتي ACF خارج حدود الثقة بمستوى معنوية (95%) وكذلك من الشكل (٣) والخاص بسلوك معاملات الارتباط الذاتي الجزئي PACF بان الإزاحة الأولى خارج حدود الثقة لمعاملات الارتباط الذاتي الجزئي وهذا مؤشر على عدم وجود استقرارية في السلسلة حسب سلوك ولذلك نأخذ الفروق حيث تتحقق الاستقرارية بعد أخذ الفرق الأول فيصبح الشكل البياني للسلسلة الناتجة كما يظهر في الشكل رقم (٤) إذ يبدو أن السلسلة اصبحت مستقرة ورسم كل من دالتي ACF و PACF بعد أخذ الفرق الأول ومعاملاتها كما يظهر في الشكلين (٥) و (٦) يؤكد ذلك :

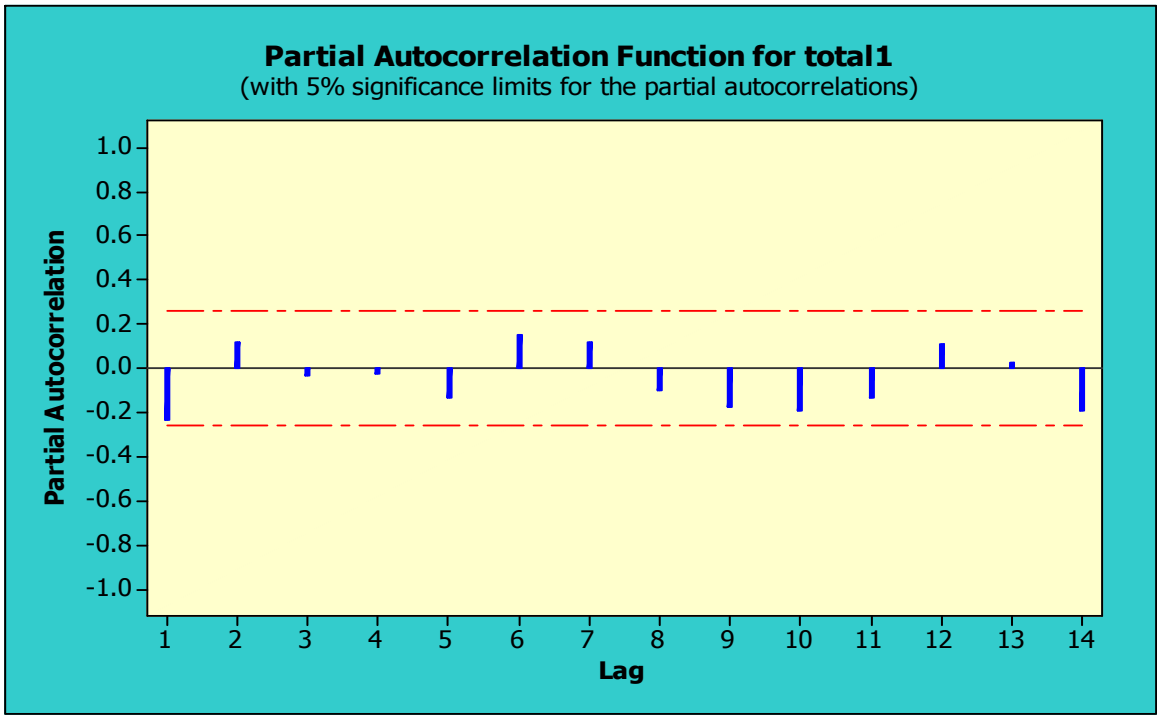
شكل (٤) رسم السلسلة عدد المصابين بالتهاب الكبد الفايروسي لجميع الأنواع بعد الفرق الاول



شكل رقم (٥) رسم دالة الارتباط الذاتي للسلسلة ACF بعد أخذ الفرق الأول



شكل رقم (٦) رسم دالة الارتباط الذاتي الجزئي للسلسلة PACF بعد أخذ الفرق الأول



تشخيص وتقدير الإنموذج واختيار النموذج الأفضل

لتحديد رتبة النموذج بشكل ادق تم توفيق عدد من النماذج واختيار النموذج الأفضل بحسب معايير المفاضلة والنماذج المقترحة هي كما موضحة في الجدول رقم (٢)

الجدول رقم (٢) يمثل نماذج بوكس جنكينز المقترحة لسلسلة عدد المصابين بالتهاب الكبد الفايروسي لجميع الأنواع

الإنموذج	MSE	BIC	AIC	الملاحظات
ARIMA(1,1,0)	25487	606.7646	602.6095	الإنموذج الافضل
ARIMA(0,1,1)	25776	607.4298	603.2747	
ARIMA(1,1,1)	25524	610.9277	604.6951	
ARIMA(2,1,0)	25594	611.0893	604.8567	
ARIMA(2,1,1)	25963	616.0114	607.7012	
ARIMA(2,1,2)	26441	621.1653	610.7776	
ARIMA(0,1,2)	25691	611.3125	605.0799	
ARIMA(1,1,2)	25671	615.3441	607.0339	
ARIMA(3,1,0)	26033	616.1702	607.8601	
ARIMA(3,1,1)	24687	617.1156	606.7279	
ARIMA(3,1,2)	24671	621.1549	608.6896	
ARIMA(3,1,3)	25012	626.0423	611.4995	
ARIMA(0,1,3)	26010	616.1181	607.8079	
ARIMA1,1,3)	24671	617.0773	606.6896	
ARIMA(2,1,3)	25536	623.1881	610.7228	

أن أفضل إنموذج من نماذج بوكس جنكيز هو ARIMA(1,1,0) كون قيم معياري (معيار معلومات بيزBIC ومعيار اكيكي AIC) الذين يستخدما للمفاضلة بين النماذج المختلفة اقل قيمة من بين جميع قيم المعيارين (AIC وBIC) للنماذج المقترحة والإنموذج الرياضي هو

$$Y_t = 14.91 - 0.2380 Y_{t-1} + u_t$$

اختبار دقة الإنموذج

بعد تشخيص النموذج وتحديد درجته وتقديره لابد من التأكد من صحة ملائمة النموذج وكفاءته وتم ذلك من خلال

١- اختبار (Ljung-Box)

من خلال تطبيق احصاءة (Ljung-Box) لفحص ملائمة الإنموذج وظهر بان (Q=12.7) عند الازاحة K=12 ومن خلال قيمة P-Value وهذا يدل على ان الاخطاء غير مرتبطة ببعضها البعض وتزداد قيمة P-Value كلما تزداد عدد الازاحات وهذا دليل على إن الإنموذج جيد وملائم وكفوء

وكما موضح ادناه

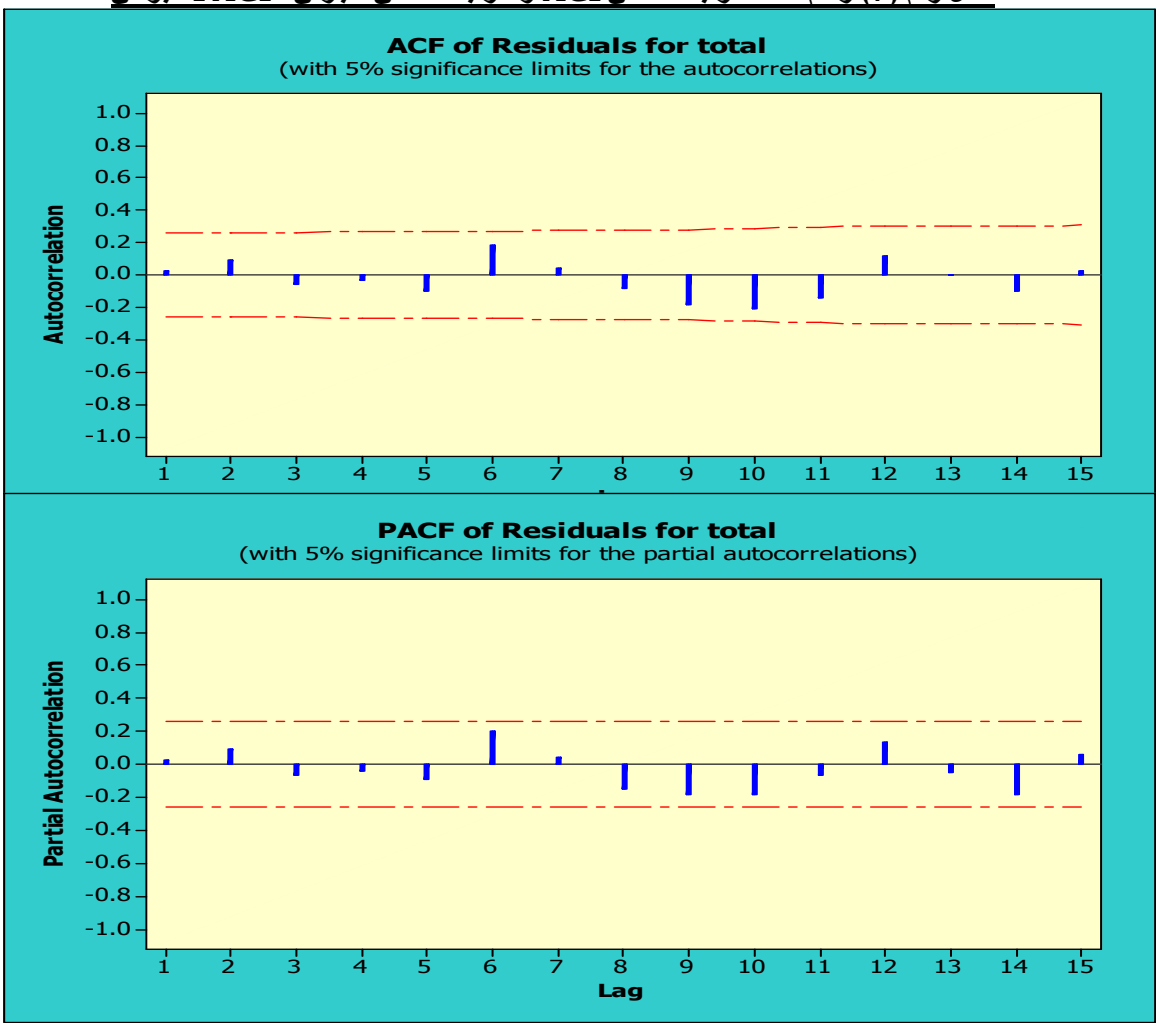
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	12.7	17.9	26.9	33.9
DF	10	22	34	46
P-Value	0.242	0.711	0.802	0.906

٢- اختبار البواقي

تم استخراج معاملات الارتباط الذاتي والجزئي للبواقي (الأخطاء) للنموذج المقدر ورسمها ونلاحظ من الشكل (٣-٣٩) ان جميع قيم معاملات الارتباط الذاتي للبواقي تقع ضمن حدود الثقة مما يعني ان سلسلة البواقي عشوائية وان النموذج المستخدم جيد وملائم . :-

شكل رقم (7) رسم دالة الارتباط الذاتيACF والارتباط الذاتي الجزئي PACF للبواقي



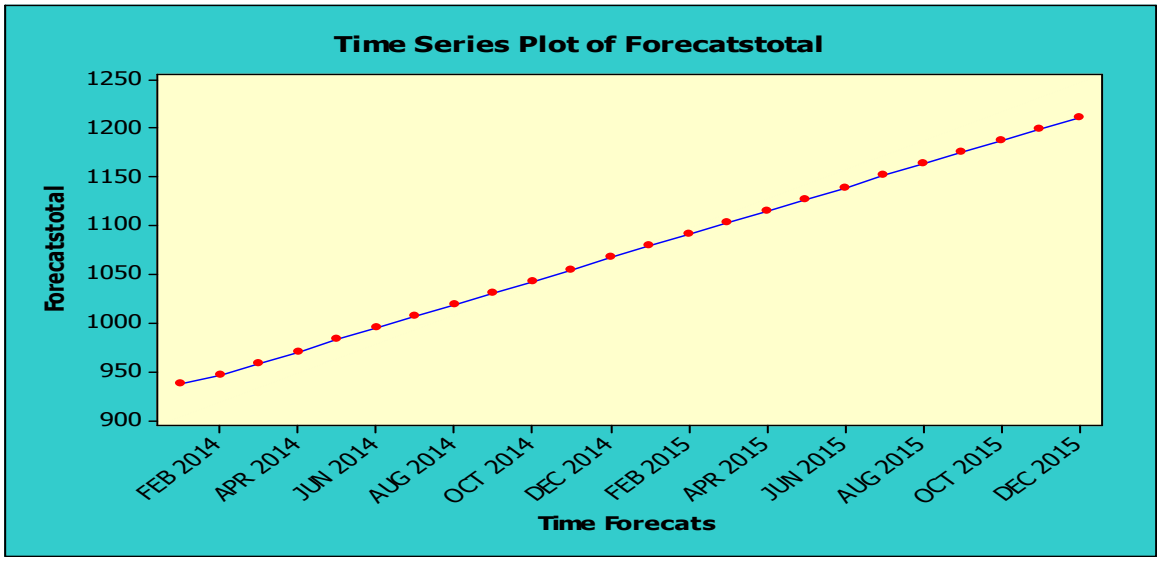
التنبؤ

بعد اجتياز الإنموذج لاختبارات التشخيص يصبح بالإمكان استعمال الإنموذج للتنبؤ والجدول رقم (٣) يمثل القيم التنبؤية للمدة ٢٠١٤ - ٢٠١٥ عدد المصابين بالتهاب الكبد الفايروسي لجميع الأنواع كما ان الشكل رقم (٨) يمثل رسم القيم التنبؤية لسلسلة عدد المصابين بالتهاب الكبد الفايروسي لجميع الأنواع مع حدود الثقة وقد تم الحصول على هذه القيم التنبؤية باستعمال الإنموذج ARIMA(1,1,0) كما يأتي:

الجدول رقم (٣) يمثل القيم التنبؤية عدد المصابين بالتهاب الكبد الفايروسي لجميع الأنواع

التاريخ	القيمة التنبؤية	التاريخ	القيمة التنبؤية
JAN 2014	938	JAN 2015	1079
FEB 2014	946	FEB 2015	1091
MAR 2014	959	MAR 2015	1103
APR 2014	971	APR 2015	1115
MAY 2014	983	MAY 2015	1127
JUN 2014	995	JUN 2015	1139
JUL 2014	1007	JUL 2015	1151
AUG 2014	1019	AUG 2015	1163
SEP 2014	1031	SEP 2015	1175
OCT 2014	1043	OCT 2015	1187
NOV 2014	1055	NOV 2015	1200
DEC 2014	1067	DEC 2015	1212

الشكل رقم (8) يمثل رسم القيم التنبؤية لسلسلة عدد المصابين بالتهاب الكبد الفايروسي لجميع الأنواع بحسب إنموذج ARIMA(1,1,0)



الاستنتاجات والتوصيات

١- الاستنتاجات

تم التوصل إلى عدد من الاستنتاجات بناءً على النتائج التي تم التوصل اليها وأهمها هي:

١- إن السلسلة الزمنية لعدد المصابين بالتهاب الكبد الفايروسي غير مستقرة في المتوسط لذا تم اخذ الفرق الأول لتحقيق الاستقرارية فيها ، وبعد اخذ الفرق الأول اصبحت السلسلة مستقرة .

٢- بالاعتماد على معياري (معيار معلومات بيز BIC ، معيار اكيكي AIC) تم اختيار الإنموذج ARIMA(1,1,0) لعدد المصابين بالتهاب الكبد الفايروسي عن طريق المفاضلة بين مجموعة من النماذج .

٣- كان توفيق الإنموذج ARIMA(1,1,0) المعتمد لعدد المصابين بالتهاب الكبد الفايروسي توفيقاً جيداً (النموذج ملائم) بعد تطبيق اختبار Ljung-Box واختبار البواقي.

٢- التوصيات

١- التأكيد على ضرورة تدوين البيانات بصورة دقيقة ومستمرة في وزارة الصحة – قسم الإحصاء الصحي والحياتي لجميع المحافظات وبضمنها إقليم كردستان لعدم وجود بيانات خاصة بالإقليم .

٢- الاعتماد على القيم المتنبأ بها في هذا البحث حسب النماذج المختارة لوضع الخطط والسياسات الملائمة لتقليل الإصابة بالتهاب الكبد الفايروسي ولجميع الأنواع .

٣- إجراء دراسات مستقبلية تتضمن استعمال نماذج السلاسل الزمنية الغير خطية والنماذج الثنائية لعدد المصابين بالتهاب الكبد الفايروسي في العراق .

المصادر

١- البياتي و المخلافي ، د.عصام حسين و فؤاد عبده اسماعيل ، استعمال اسلوب بوكس – جنكينز للتنبؤ بإنتاجية العمل في مصنع اسمنت عمران في القطاع الصناعي اليمني ، مجلة الادارة والاقتصاد ، العدد الثالث والستون /٢٠٠٧ .

٢- التميمي ، د. رعد فاضل، الانحدار والسلاسل الزمنية أساليب إحصائية تطبيقية متقدمة باستعمال نظام Minitab ، كتاب ، بغداد ٢٠١٣ .

٣- الجبوري ، عبير حسن علي ، التنبؤ بأسعار النفط العراقي لعام ٢٠١٠ ، مجلة جامعة بابل ، العلوم الإنسانية، المجلد ١٨ العدد ١ لعام ٢٠١٠ .

٤- السلطاني ، حلا مثنى محمد ، التنبؤ بمحصولي الشلب والقطن في العراق باستعمال السلاسل الزمنية، بحث دبلوم عالي ، كلية الإدارة والاقتصاد، جامعة بغداد ٢٠١٤

٥- الشمراني ، د.محمد موسى ، مقارنة بين بعض الاساليب الاحصائية التقليدية ونماذج بوكس جنكنز في تحليل بيانات السلاسل الزمنية ، مجلة ام القرى للعلوم التربوية والنفسية ، المجلد الخامس ، العدد الاول يناير ٢٠١٣ .

٦- الصراف وشومان د.نزار مصطفى ، د.عبد اللطيف حسن، (السلاسل الزمنية والأرقام القياسية)، كتاب، دار الدكتور للعلوم الإدارية والاقتصادية ، بغداد ٢٠١٣.

٧- الطائي ، فاضل عباس ، "التنبؤ والتمهيد للسلاسل الزمنية باستعمال التحويلات مع التطبيق"، بحث منشور، المجلة العراقية للعلوم الإحصائية العدد ١٧ لعام ٢٠١٠

٨- العاني، احمد حسين بتال "استعمال نماذج ARIMA في التنبؤ الاقتصادي /٢٠٠٥ .

٩- الكحلوت ، د.جمال رشيد ، مبادئ في الاحصاء والاحتمالات / الطبعة الثالثة ، المملكة العربية السعودية / ٢٠٠٤ .

١٠- بري، د.عدنان ماجد بري ، طرق التنبؤ الإحصائي ، كتاب ، جامعة الملك سعود/ ٢٠٠٢ .

١١- طعمة ، سعدية عبد الكريم ، استعمال السلاسل الزمنية للتنبؤ بأعداد المصابين بالأورام الخبيثة في محافظة الانبار، مجلة جامعة الانبار للعلوم الاقتصادية والادارية ، المجلد ٤ / العدد ٨ /٢٠١٢ .

١٢- عبد الرسول ، محمود جواد "دراسة احصائية تطبيقية للمقارنة بين النماذج الاسية ونماذج بوكس وجينكنز في التوقعات المستقبلية مع تطبيق عملي"، رسالة ماجستير إحصاء، كلية الإدارة والاقتصاد، جامعة بغداد ١٩٨١ .

١٣- عبود، نزار جابر " بناء نموذج لتقدير إنتاج الثروة السمكية في العراق " بحث دبلوم عالي ، كلية الإدارة والاقتصاد، جامعة بغداد ٢٠١٥ .

١٤- عمران ، خالدة كريم " التَنبـؤ بإنتـاجُ التمورفي العِـراق بإستعمال نماذج السلاسل الزمنية" ، بحث دبلوم عالي ، كلية الإدارة والاقتصاد، جامعة بغداد ٢٠١٤ .

١٥- ماجد ، هيثم حسون ، استعمال أساليب السلاسل الزمنية لمعالجة الاختلافات الموسمية في الرقم القياسي لسعر المستهلك ، بحث دبلوم عالي ، كلية الإدارة والاقتصاد، جامعة بغداد ٢٠١٢ .
١٦- مطر والياس ، ظافر رمضان و انتصار ابراهيم ،" تحليل ونمذجة السلسلة الزمنية لتدفق المياه الداخلة الى مدينة الموصل " بحث منشور، المجلة العراقية للعلوم الاحصائية (١٨) لعام ٢٠١٠ .

▪ المصادر الأجنبيـــة

17- Baldwin, Andrew, 1990, "Seasonal Baskets in Consumer Price Indexes," Journal of Official Statistics, Vol. 6 (No. 3.
18- Box, G.E.P., & Jenkins, G.M. (1970), "Time series analysis: forecasting and control", Holden-Day, San Francisco.
19- Brown, R.G. & Meyer, R.E.," The fundamental theorem of exponential smoothing ",O.R . vol. (9),NO.(6), p.(673),(1960).
20- William R.Bell and Steven C. Hillmer,(1984) (Seasonal Adjustment of Economic Time Series) Statistical Research Series Census/SRD/RR- March 30, 1984.

Box and Jenkins use models to predict the numbers of patients with hepatitis Alvairose in Iraq

Abstract:

The time series of statistical methods mission followed in this area analysis method, Figuring certain displayed on a certain period of time and analysis we can identify the pattern and the factors affecting them and use them to predict the future of the phenomenon of values, which helps to develop a way of predicting the development of the economic development of sound

The research aims to select the best model to predict the number of infections with hepatitis Alvairose models using Box - Jenkins non-seasonal forecasting in the future.

Data were collected from the Ministry of Health / Department of Health Statistics for the period (from January 2009 until December 2013) was used statistical program MINITAB16 for the analysis of time series and through differentiation standards (standard biz BIC information, standard Waikiki AIC) shows that the best model for the number of injured Alvairose liver inflammation in Iraq is a model ARIMA (1,1,0).

Key words : Box and Jenkins use models , hepatitis Alvairose .