AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

استعمال الأنموذج التكاملي [ARIMA (P، d، q)] للتنبؤ في بيانات طلبة المدارس الثانوية لمحافظة كربلاء

[ARIMA (p، d، q)] integrative model for To predict the data of secondary school students in the Karbala Governorate

م .د. فائز حامد سلمان

Faez Hamed Salman

كلية الإدارة والاقتصاد – جامعة بغداد

Faiz.H@coadec.uobaghdad.edu.iq

المستخلص

يُعد التعليم أساس تقدم المجتمعات المعاصرة؛ لأنه مؤشراً لرقي المجتمعات وتطورها والسمو بثقافتها إلى الأعلى عن طريق التركيز على التعليم ، وإن معدل النمو السكاني المرتفع نسبياً (3.4 %) حسب تقديرات وزارة التخطيط /الجهاز المركزي للإحصاء أدى إلى ظهور مشكلة كثافة الطلبة داخل الصف المدرسي إذ بلغ معدل الكثافة (43) تلميذ لكل صف دراسي وما يرتبط بهذه الأعداد من التزامات يجب التخطيط لها ومواجهتها مثل توفير الأعداد الكافية من المدارس والملاكات التعليمية والكتب المجانية وغيرها من خلال التنبؤ بأعداد هؤلاء الطلبة بأسلوب احصائي يعتمد عليه وعدم استعمال أسلوب التخمين المتبع في المؤسسة المعنية حسب علم الباحث الأمر الذي إلى تنبؤات عشوائية لا تعكس التغيرات التي تطرا على اعداد الطلبة مع الزمن ،وهدف البحث إلى التنبؤ بعدد طلبة المرحلة الثانوية في محافظة كربلاء المقدسة للمدة الزمنية (2025- 2030) عن طريق الأنموذج التكاملي ($ARIMA_{(P،d،q)}$) وذلك من خلال جمع بيانات السلسلة الزمنية الممتدة من (1970- 2023) م لعدد الطلبة الثانوي وتوصل الباحث إلى افضل انموذج للتنبؤ هو $ARIMA_{(2،1،3)}$ وظهرت القيم التنبؤية لطلبة الثانوي في تصاعد مستمر لتصبح عام 2030 (**279567**) طالب.

الكلمات المفتاحية: السلسلة الزمنية ، $ARIMA_{(P،d،q)}$، دالة الارتباط الذاتي ، دالة الارتباط الذاتي الجزئي، التنبؤ .

Abstract

Education is considered the pillar and foundation of the progress of contemporary societies، as it is an indicator of the progress and development of societies and the advancement of their culture to the highest level by focusing on education. The relatively high population growth rate (3.4%) according to estimates by the Ministry of Planning/Central Bureau of Statistics led to the emergence of the problem of student density within the school class، as it reached The density rate is (43) students per class، and the obligations associated with these numbers must be planned and faced، such as providing sufficient numbers of schools، educational staff، free books، etc.، by predicting the number of these students in a reliable statistical manner and not using the estimation method used in the institution in question، according to knowledge. The researcher made random predictions that do not reflect the changes that occur in the number of students over time، The research aimed to predict the number of secondary school students in the Holy Governorate of Karbala for the period of time (2025-2030) through the integrative model ARIMA $_{(P،d،q)}$ by collecting time series data extending from (1970-2023) AD. For the number of secondary students، the researcher arrived at the best prediction model، which is ARIMA$_{(2،1،3)}$. and the predictive values for secondary school students appeared to be constantly increasing، In 2030، there will be 279،567 students.

المقدمة : Introduction

يُعد التعليم عِماد وأساس تقدم المجتمعات المعاصرة كونه مؤشراً لرقي المجتمعات وتطورها والرقي بثقافتها إلى الاعلى عن طريق التركيز على التعليم ، ويعدُّ التعليم الثانوي مرحلة حاسمة ومهمة لطلبة التعليم العام، إذ يفترض في هذه المرحلة أن يعد الطلبة إعدادا متكاملا وشاملاً ومزوداً بالمعلومات الأساسية والمهارات التي من شانها تنمي شخصيتهم المعرفية والنفسية والاجتماعية والعقلية والبدنية ...الخ ،

تم تناول موضوع السلاسل الزمنية في البحث والتي تعد وأحدة من أهم الادوات الإحصائية الحديثة التي اصبح تدأولها في مختلف مجالات الحياة وبكثرة ، إذ نستطيع من خلالها الكشف عن سلوك السلسلة الزمنية من حيث طبيعة التغيرات التي تحدث على قيم الظاهرة مع الزمن وتشخيص الأسباب وتفسير العلاقات بينهما والتوصل إلى النتائج ومن ثم التنبؤ بما سيجري على قيم الظاهرة من تغير في المستقبل استنادا على ما حدث لها في الماضي.

استخدم الباحث في هذا البحث الأنموذج التكامل$ARIMA_{(P،d،q)}$ وهو أحد نماذج بوكس وجنكز للتنبؤ بأعداد طلبة الثانوي في محافظة كربلاء المقدسة للمدة (2025- 2030) م.

مشكلة البحث : Research Methodology

إن معدل النمو السكاني المرتفع نسبياً (3.4 %) حسب تقديرات وزارة التخطيط /الجهاز المركزي للإحصاء أدى إلى ظهور مشكلة كثافة الطلبة داخل الصف المدرسي إذ بلغ معدل الكثافة (43) تلميذ لكل صف دراسي وما يرتبط بهذه الأعداد من التزامات يجب التخطيط لها ومواجهتها مثل توفير الأعداد الكافية من المدارس والملاكات التعليمية والكتب المجانية وغيرها من خلال التنبؤ بأعداد هؤلاء الطلبة بأسلوب احصائي يعتمد عليه وعدم استعمال أسلوب التخمين المتبع في المؤسسة المعنية حسب علم الباحث الأمر الذي إلى تنبؤات عشوائية لا تعكس التغيرات التي تطرأ على أعداد الطلبة مع الزمن .

هدف البحث : Objective of Research

يهدف البحث إلى التنبؤ بعدد طلبة المرحلة الثانوية في محافظة كربلاء المقدسة للمدة الزمنية (2025- 2030) عن طريق الأنموذج التكاملي ($ARIMA_{(P،d،q)}$) ؛ وذلك من خلال جمع بيانات السلسلة الزمنية الممتدة من (1970-2023) م لعدد الطلبة الثانوي.

حدود البحث : Research Window

إطار البحث زمنياً تمثل في المدة (1970- 2023)م ، مكانياً تمثل في محافظة كربلاء المقدسة.

مصدر البيانات : Data Source

حصل على البيانات من وزارة التخطيط / الجهاز المركزي للإحصاء ، والمديرية العامة للتربية في محافظة كربلاء المقدسة .

1 - الجانب النظري

1-1: السلسلة الزمنية :Time Series [6]،[1]

كل مجموعة من المشاهدات لظاهرة معينة، دونت على وفق مدة زمنية متساوية ومتتابعة (ساعات ، أيام ، أسبوعي ، فصلي ، سنوي وغيرها) تسمى سلسلة زمنية مثل صادرات النفط اليومية لدولة العراق. ومن فوائدها التنبؤ عن القيم التي ستأخذها الظاهرة في المستقبل.

ورياضيا هي العلاقة الدالية بين متغيرين وتكتب بالشكل الآتي :

$$Y= f(t) \quad \dots(1)$$

حيث أن : y يمثل قيم الظاهرة، t يمثل الزمن ، f(t) تمثل دالة بدلالة الزمن

وعند وجود متغيرات أخرى مع الزمن تؤثر على الظاهرة y تستعمل العلاقة الرياضية الآتية :

$$Y= f(t، x_1، x_2، \dots ، x_3) \quad \dots (2)$$

وغالبا ما تكون السلسلة الزمنية لأعداد الطلبة ذات اتجاه عام، الذي يعرف على أنه معدل النمو السلبي أو الإيجابي لقيم ظاهرة معينة المحددة خلال مدة زمنية طويلة، بمعنى أن ميل الخط يكون صاعداً أو هابطا، ويكون مستقيما في حالة عدم وجود معدل النمو لقيم الظاهرة خلال المدة الزمنية، يرمز لاتجاه العام بالرمز (T)، مثل نمو عدد الطلبة لبلد معين هو النمو الإيجابي، أما النمو السلبي مثل تناقص عدد الاميين .

2-1: الأنموذج التكاملي: Auto Regressive Integrated Moving Average Model [2][3]

هو أحد نماذج بوكس وجنكز ويعد من ابرز طرائق التنبؤ والذي يدعى (انموذج الانحدار الذاتي والمتوسطات المتحركة التكاملية $\mathrm{ARIMA}_{(P،d،q)}$، وتمثل (p) رتبة الانحدار الذاتي أما (q) رتبة الوسط المتحرك في حين (d) تمثل رتبة الفروق وأن ($d \geq 1$). ، قدمه العالمان بوكس وجنكز عام 1970. وأثبتت هذه الأنموذج كفاءة عالية في النمذجة والتنبؤ لشتى مجالات الحياة لما يمتاز به عن بقية النماذج لكونه لا يفترض الاستقلال بين المشاهدات، وإنما يعتمد نمط الارتباط الموجود بين المشاهدات في عملية النمذجه والتنبؤ ، فتكون قيم التنبؤات أكثر مصداقية ودقة كون السلسلة الزمنية تمر بعدة مراحل (التعريف – التشخيص – التقدير – التنبؤ) ، مع ملاحظة استعمال الأنموذج في حال كون السلسلة الزمنية غير مستقرة بعد استعمال الفروقات وتحويلات ما لجعلها مستقرة.

وانموذج ($\mathrm{ARIMA}_{(P،d،q)}$) للسلسلة الزمنية y_t هو :

$$\varphi_P(\beta)(1-B)^d Y_t = \theta_q(\beta)a_t \qquad \dots(3)$$

إذ ان :

إذ $\emptyset(B)$:عامل ذاتي الانحدار المستقر، وجذور المعادلة $\emptyset(B) = 0$ تقع خارج دائرة الوحدة.

أما $\theta(B)$: معامل المتوسطات المتحركة وجذور المعادلة $\theta(B) = 0$ تقع خارج دائرة الوحدة.

وبصورة عامة الفرق(d) يمثل بالمعادلة:

$$\nabla^d \mathrm{Y_t} = (1-B)^{\mathrm{d}} \mathrm{Y_t} \qquad \dots(4)$$

∇. : معامل الفروق ، B : عامل الارتداد الزمني .

3-1: مراحل تنفيذ انموذج (ARIMA) Stages of Application

يمكن تنفيذ الأنموذج بأربع مراحل وكما يأتي :

المرحلة الأولى / جمع البيانات : Data Collection [4]

يكون حجم العينة (50) مشاهدة فأكثر لسلسلة الزمنية ولكن بعضهم ارتأى بأن يكون حجم العينة (30) مشاهدة فأكثر لتحليل سلسلة الزمنية .

المرحلة الثانية / فحص استقرارية السلسلة : Check Stationary of Time Series [3]

السلسلة الزمنية المستقرة هي التي تكون متجانسة زمنياً أي أن رسم السلسة في المدة t} ،{t+w يكون مطابقا لرسم السلسلة في مدة أخرىs}،{s+w ، يوجد الكثير من الاختبارات التي يمكن بواسطتها الكشف عن الاستقرارية ومنها رسم السلسلة بيانياً، استعمال دالة الارتباط الذاتي(ACF) والارتباط الذاتي الجزئي (PACF)، واختبار ديكي فولر المطور Augmented Dickey Fuller، وعند ثبوت عدم استقرارية السلسلة تحويل إلى سلسلة زمنية مستقرة باستعمال بعض التحويلات المناسبة ، فقد تؤخذ الفروق للسلسلة الزمنية إذا كانت السلسلة غير مستقرة في المتوسط ، أما إذا كانت السلسلة الزمنية غير مستقرة في التباين، فإنه يتم تحويلها إلى الصيغة اللوغارتمية أو يؤخذ لها الجذر التربيعي ...الخ

- دالة الارتباط الذاتي Auto Correlation Function (ACF) [9][5]

إن الصيغة الرياضية لهذه الدالة هي :

$$P_K = \frac{\operatorname{cov}(Y_t, Y_{t+k})}{\sqrt{\operatorname{var}(Y_t)\ \operatorname{var}(Y_{t+k})}} = \frac{\mathrm{E}(Y_t-\mu)(Y_{t+k}-\mu)}{\sqrt{\mathrm{E}(Y_t-\mu)^2\ \ \mathrm{E}(Y_{t+k}-\mu)^2}} \ldots\ldots(5)$$

إذ إن:

Y_t: قيم مشاهدات السلسلة الزمنية ، Y_{t+k} : قيم مشاهدات السلسلة الزمنية القادمة.

$Var(Y_t)$: تمثل تباين السلسلة الزمنية.

تعدُّ السلسلة الزمنية مستقرة عندما تقع معاملات الارتباط الذاتي بعد الازاحتين الثانية أو الثالثة ضمن حدود الثقة الآتية وبمستوى معنوية 95% :

$$-1.96 * 1/\sqrt{n} \le r_k \le 1.96 * 1/\sqrt{n} \quad k=1,2,3 \ldots\ldots(6)$$

n: عدد مشاهدات السلسلة الزمنية ، r_k: معاملات الارتباط الذاتي المقدرة.

- **دالة الارتباط الذاتي الجزئي Partial Auto Correlation Function (PACF) [8]، [14]**

ان معاملات الارتباط الذاتي الجزئي تكتب رياضياً كالآتي :

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}} \ldots(7)$$

إذ تسمى المعلمات $\left(\emptyset_{kk}; k = 1،2، \dots\right)$ بدوال الارتباط الذاتي الجزئي$(PACF)$

- **اختبار ديكي فولر المطور Augmented Dickey Fuller (ADF) [10]،[7]**

يعدُّ اختبار دكي فولر الافضل والاقوى في تشخيص استقرارية السلسلة الزمنية ويعتمد على ثلاث معادلات وهي :-

المعادلة الأولى بدون حد ثابت وصيغتها هي :

$$\Delta Y_t = \emptyset_1 Y_{t-1} + \sum_{j=1}^{p} B_j \, \Delta Y_{t-j} + a_t \qquad \dots (8)$$

Δ-- Y_t : الفرق الأول للسلسلة ، B_j : معلمة دكي فولر ، Δ : فروقات الفجوة الزمنية .

المعادلة الثانية تحتوي على حد ثابت وصيغتها هي :

$$\Delta Y_t = \emptyset_0 + \emptyset_1 Y_{t-1} + \sum_{j=1}^{p} B_j \, \Delta Y_{t-j} + a_t \qquad \dots (9)$$

المعادلة الثالثة تحتوي على حد ثابت واتجاه عام وصيغتها هي :

$$\Delta Y_t = \emptyset_0 + \emptyset_1 Y_{t-1} + \sum_{j=1}^{p} B_j \, \Delta Y_{t-j} + \delta t + a_t \qquad \dots (10)$$

وان اختبار دكي فولر المطور يستعمل لاختبار الفرضيتين الآتيتين :

$H_0: B - 1 = 0$ وجود جذر الوحدة (عدم استقرارية السلسلة الزمنية)

$H_1: B - 1 < 0$ عدم وجود جذر الوحدة (استقرارية السلسلة)

وتقارن احصاء الاختبار $t = \frac{\hat{B}-1}{SE\hat{B}_1}$ مع القيم الجدولية التي وضعها (ديكي فولر)، فإذا كانت قيمة(t) المحسوبة أكبر من الجدولية (جداول ديكي فولر) ترفض فرضية العدم وتقبل الفرضية البديلة، ومن ثم تكون السلسلة مستقرة .

المرحلة الثالثة / تشخيص الأنموذج Identification [11]،[13][10]

يتم الاعتماد في تشخيص الأنموذج الملائم للسلسلة الزمنية على بعض المعايير الإحصائية، إذ اختير الأنموذج الذي يملك أقل قيم للمعايير الاحصائية والمعايير هي :

1- **معيار أكأيكي Akaike Information Criterion (AIC)**

ان الصيغة الرياضية لهذا المعيار تكتب كما يأتي:

$$AIC(M) = n\,Ln(\hat{\sigma}_a^2) + 2M \qquad \dots(11)$$

$$\hat{\sigma}_a^2 = \frac{\sum_{t=0}^{n}(y_t - \hat{y}_t)^2}{n} \qquad \dots(12)$$

n: تمثل عدد المشاهدات ، $\hat{\sigma}_a^2$: يمثل مقدار تبأين الخطأ.

$\hat{\sigma}_a^2$: يمثل مقدار تبأين الخطأ ، M : رتبة الأنموذج المختار أو عدد معلمات الأنموذج.

2- **حنان– كوين $(H-Q)$ Hannan-Quinn Criterion**

ان الصيغة الرياضية لهذا المعيار تكتب كما يأتي

$$H - Q(M) = Ln(\hat{\sigma}_a^2) + \frac{2MCLn\big(Ln(n)\big)}{n} \quad ; \quad C > 2 \qquad \dots(13)$$

إذ إن:

C: ثابت ، M: عدد معلمات الأنموذج أو رتبة الأنموذج المختار.

n: عدد مشاهدات السلسلة الزمنية ، $\hat{\sigma}_a^2$: تقدير تبأين الخطأ

3- **معيار شوارتز البيزي Schwartz Bayesian Criterion (SBC)**

ان الصيغة الرياضية لهذا المعيار تكتب كما يأتي:

$$SBC(M) = n \times Ln(\hat{\sigma}_a^2) + M \times Ln(n) \qquad \dots(14)$$

M: عدد معلمات الأنموذج أو رتبة الأنموذج المختار ، n: عدد مشاهدات السلسلة الزمنية.

4-1 :تقدير معلمات الأنموذج Estimation Model Parameters [12][14]

تم استعمال طريقة الامكان الأعظم المضبوطة Exact Maximum Likelihood Function لتقدير معلمات الأنموذج وكما يأتي:

إن دالة الإمكان Likelihood function للسلسلة الزمنية Y_t تعتمد على تعظيم دالة الامكان وذلك بجعل مجموع مربعات الاخطاء أقل ما يمكن والصيغة على وفق المعادلة الآتية :

$$L\big(\emptyset,\theta,\sigma_a^2 \lfloor W\big) = (2\pi\sigma_a^2)^{\frac{-n}{2}} \left|M_n^{(p,q)}\right|^{\frac{1}{2}} \exp\left(-\frac{S(\emptyset,\theta}{2\pi\sigma_a^2}\right) \qquad \dots(15)$$

وكون العينة كبيرة وعند اهمال المقدار $\left|M_n^{(p،q)}\right|$ في الصيغة اعلاه وأخذ اللوغارتم لدالة الإمكان فإن المعادلة تصبح بالشكل الآتي :

$$Ln\ L(\emptyset،\theta،\ \sigma_a^2|W) = \left(-\frac{n}{2}\right|n(2\pi\sigma_a^2) - \frac{S(\phi،\theta)}{2\pi\sigma_a^2} \qquad \dots(16)$$

وعند العينات الكبيرة الحجم فان مصفوفة المعلومات التي تساوي$^{I}(\emptyset،\theta)$ والمتوفرة بعد ان يتم حساب مصفوفة المشتقات ذات الرتبة (p+q) x n ويكون

$$a = \theta^{-1}(B)\phi(B) \qquad \dots(17)$$

ويتم الحصول على مقدرات الإمكان الأعظم المضبوط من المعادلة رقم (17) بعد أخذ المشتقات الجزئية لها أي أن:

$$\left.\begin{aligned}\frac{\partial a_t}{\partial \theta_i} &= \theta^{-1}(B)B^i a_t \\ \frac{\partial a_t}{\partial \emptyset_j} &= \emptyset^{-1}(B)B^j a_t\end{aligned}\right\} \quad \dots\dots(18)$$

5-1:فحص ملائمة الأنموذج : Diagnostic Checking Model [6] [10]

لتأكد من صحة مدى ملائمة الأنموذج المقدر المحدد تستعمل الفرضية الآتية :

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_3 = 0 \qquad \dots(19)$$

$$H_1: \rho_i \neq \rho_l \quad ; \ i \neq l \qquad \dots(20)$$

توجد عدة اختبارات للتحقق من مدى ملائمة الأنموذج المشخص نذكر البعض منها :

- **اختبار حدي الثقة: Confidence Interval Test**

يتم أخذ سلسلة البواقي $\{a_t\}$ والخاصة بالأنموذج المشخص التي تفترض بأنها تتوزع توزيعاً طبيعياً، وان $(Ea_t a_{t+k} = 0)$ وأختبارها بعد تقدير الارتباطات الذاتية لها $r_k(a)$، وان تلك المعاملات تتوزع توزيعاً تقاربياً طبيعياً بوسط حسابي (صفر) وتبأين $(\frac{1}{\sqrt{n}})$ عندما يزداد حجم العينة n ، أي أن:

$$r_k(a) \sim N\left(0، \frac{1}{\sqrt{n}}\right) \qquad \dots(21)$$

وأن تقدير معامل الارتباط الذاتي للبواقي هي:

$$r_k(a) = \frac{\sum_{t=1}^{n-k} a_t a_{t+1}}{\sum_{t=1}^{n} a_t^2} \qquad \dots (22)$$

فإذا كانت معاملات الارتباطات الذاتية لبواقي الأنموذج المشخص واقعة ضمن حدي الثقة وبمستوى ثقة 95%، أي ان:

$$pr\left\{\hat{r_k}(a)\right| < 1.96\frac{1}{\sqrt{n}}\} \quad =1- \alpha \qquad \dots(23)$$

فهذا يشير إلى إن الأخطاء عشوائية ومن ثم فإن الأنموذج المشخص يكون ملائماً لبيانات السلسلة الزمنية.

- **اختبار Ljung and Box**

إن مفهوم اختبار ليونغ- بوكس يعتمد على مدى معنوية عدد الازاحات الزمنية دفعة وأحدة. والصيغة الرياضية للاختبار هي :

$$Q_{L-B} = n(n+2)\sum_{k=0}^{n}(\frac{{r_k}^2(a)}{n-k}) \qquad \dots (24)$$

إذ أن:

k: عدد معاملات الارتباط الذاتي في المجموعة ، r_k: معاملات الارتباط الذاتي المقدرة.

وتقارن قيمة Q_{L-B} المحسوبة مع $\chi^2_{(m-p-q)}$ بمستوى معنوية معين، فإذا كانت القيمة المحسوبة أقل من قيمة $\chi^2_{(m-p-q)}$ نقبل فرضية العدم القائلة بأن معاملات الارتباط الذاتي للأخطاء المقدرة تتوزع توزيعاً عشوائياً، أي أن الأنموذج المشخص ملائم .

6-1: التنبؤ Forcasting [8][14]

اخيرا بعد الحصول على الأنموذج الملائم تأتي مرحلة التنبؤ، ويعدُّ التنبؤ جيد في حال كون الاخطاء الناتجة عن التقدير صغيرة جداً والتباين أقل ما يمكن.

نفرض أن الأنموذج الملائم الذي تُوصلَ إليه من خلال المراحل السابقة هوARIMA(q،d،p) ، يمكننا التنبؤ بالقيم المستقبلية إلى (L=1،2،…) مدة قادمة. فيكون أنموذج ARIMA(q،d،p) كما في المعادلة ادناه :

$$Y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \qquad \dots (25)$$

إذ إن:

Y_t: القيم المسجلة في الزمن t ، Y_{t-1}: القيم المقاسة في الزمن t-1.

ϕ_i: معلمات الأنموذج الانحدار الذاتي ، حيث (i = 1،…،p) وان $(-1 < \emptyset_i > 1)$.

a_t: الخطأ العشوائي عند الزمن t ، a_{t-1}: الخطأ العشوائي عند الزمن t-1 ، وهكذا.

θi: معلمات الأنموذج الأوساط المتحركة ، حيث (i = 1،...،q).

ولمدة L المستقبلية حسب المعادلة التالية:

$$Y_t = \mu + \phi_1 Y_{t+L-1} + \phi_2 Y_{t+L-2} + ... + \phi_p Y_{t+L-p} + a_{t+l_} - \theta_1 a_{t+L-1} - .\theta_2 a_{t+L-_2} ... - \theta_q a_{t+L-q} \quad ...(26)$$

إذ ان :

μ : مقدار ثابت .

ومن ثم التنبؤ بلانموذج يكون كمايأتي :

$$Y_t = \mu + \phi_1 EY_{t+L-1} + \phi_2 EY_{t+L-2} + ... + \phi_p EY_{t+L-p} + a_{t+L_} - \theta_1 Ea_{t+L-1} - .E\theta_2 a_{t+L-_2} - ... - \theta_q Ea_{t+L-q} \quad ..(27)$$

2 - الجانب التطبيقي :

تم تحليل البيانات السلسلة الزمنية بالاعتماد على البرنامج الجاهز(Gretl9،) . لغرض تحديد الأنموذج الافضل في تقدير أعداد طلبة الثانوي المبينه في الجدول (1) وقد اتبعت خطوات التحليل الآتية :

جدول (1) اعداد طلبة الثانوي

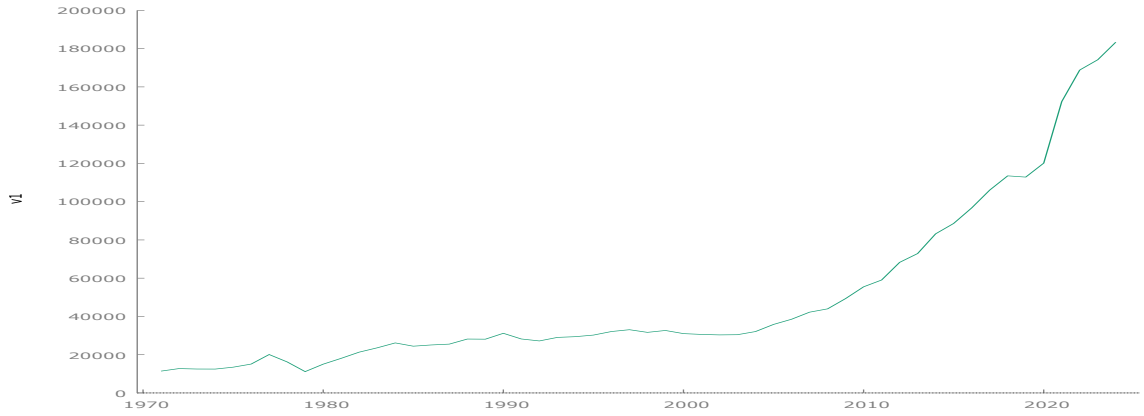
السنة	الأعداد	السنة	الأعداد	السنة	الأعداد	السنة	الأعداد	السنة	الأعداد
1970	11423	**1981**	21328	**1992**	29027	**2003**	32097	**2014**	88600
1971	12742	**1982**	23594	**1993**	29416	**2004**	35820	**2015**	96727
1972	12500	**1983**	26154	**1994**	30247	**2005**	38509	**2016**	106032
1973	12474	**1984**	24443	**1995**	32104	**2006**	42223	**2017**	113489
1974	13458	**1985**	25084	**1996**	33065	**2007**	43953	**2018**	112839
1975	15085	**1986**	25516	**1997**	31663	**2008**	49333	**2019**	120134
1976	20075	**1987**	28167	**1998**	32650	**2009**	55481	**2020**	152323
1977	16210	**1988**	28093	**1999**	31018	**2010**	59031	**2021**	168820
1978	11148	**1989**	31179	**2000**	30586	**2011**	68264	**2022**	174192

183381	**2023**	72870	**2012**	30339	**2001**	28213	**1990**	15054	**1979**
		83190	**2013**	30454	**2002**	27192	**1991**	18109	**1980**

1-2: فحص استقرارية السلسلة الزمنية

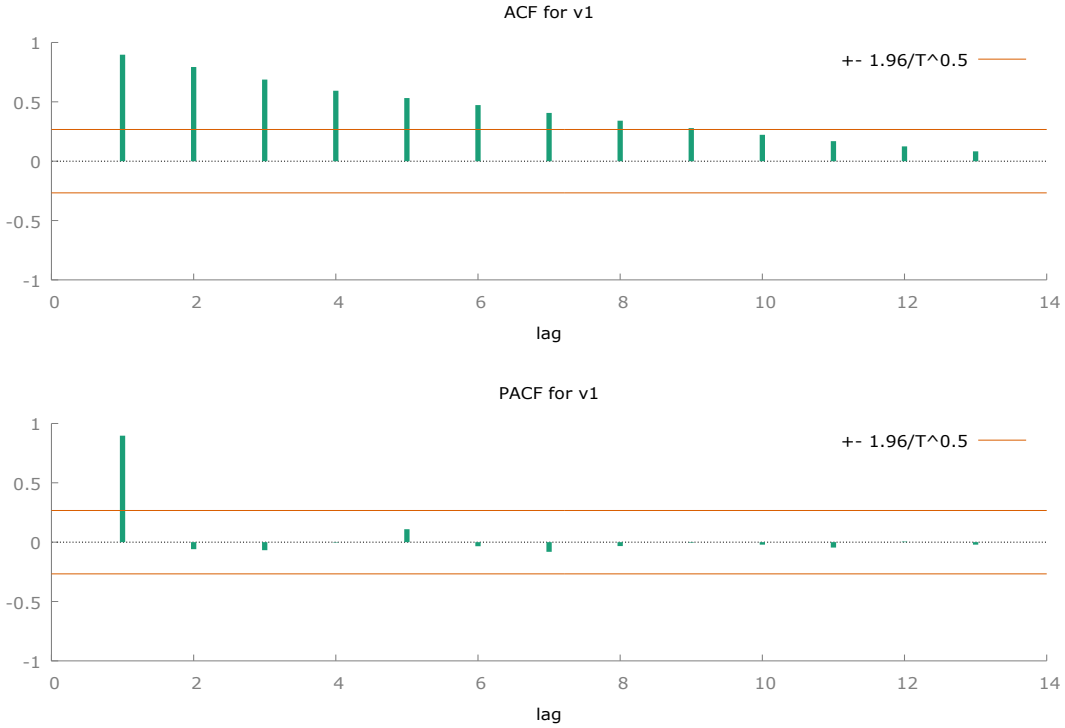
لغرض فحص الاستقرارية تم إجراء الآتي:

1- رسم سلسلة بيانات أعداد الطلبة بيانياً والشكل (1) يوضح عدم استقرارية السلسلة من ناحية المتوسط والتباين، إذ إن السلسلة لها اتجاه عام متزايد.



الشكل (1) سلسلة الثانوي

2- معنوية معاملات الارتباط الذاتي (ACF) والارتباط الذاتي الجزئي(pACF) نلحظ في الشكل (2) أغلب معاملات الارتباط الذاتي تقع خارج حدود الثقة خصوصا بعد الازاحة الثانية أو الثالثة، مما يدل على عدم استقرارية البيانات أي نرفض فرضية العدم $(H_O: \rho = 0)$.



الشكل (2) يبن فترات الثقة لمعاملات دالتي(ACF)،PACF لسلسلة الثانوي

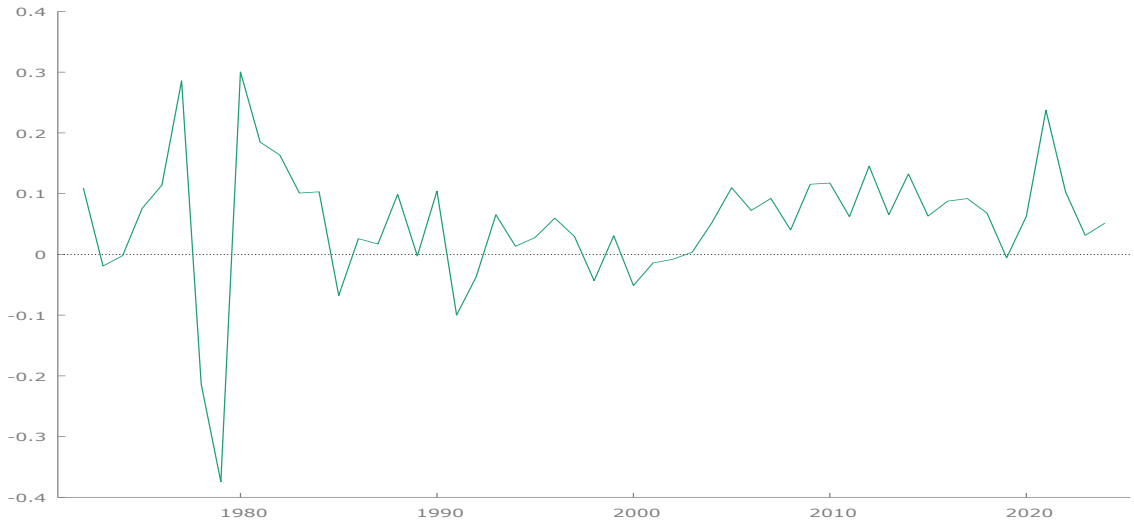
3- استعمال اختبار دكي فولر المطور اثبت الاختبار ان السلسلة غير مستقرة لكون ظهور قيمة (P-Value) أكبر من (0.05) لمعادلات (بدون القاطع، مع القاطع، مع القاطع والاتجاه) والجدول (2) يبين نتائج الاختبار كالآتي :

جدول (2) يبين اختبار دكي فولر لسلسلة بيانات اعداد الطلبة

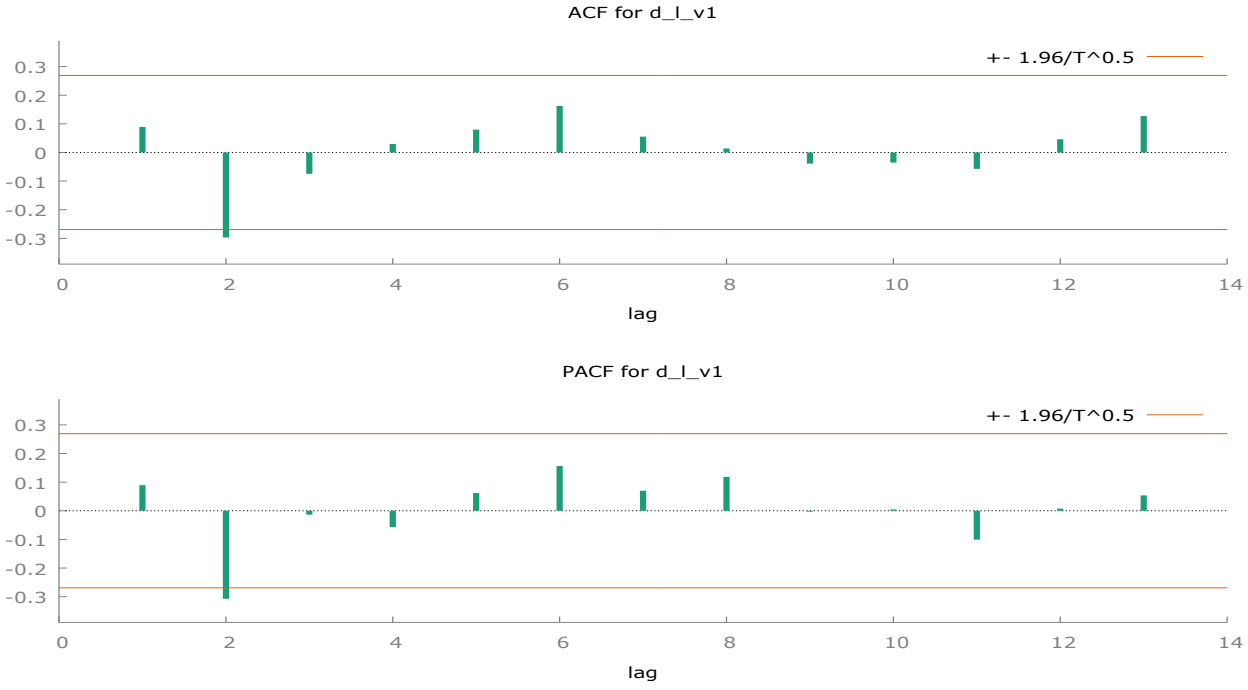
Model	estimated value	test statistic	p-value
test without constant	0.0273394	3.25219	0.101
test with constant	0.025344	1.74262	0.231
with constant and trend	0.0194255-	0.577378-	0.243

2-2 : استقرارية السلسلة : لكي نستعمل الأنموذج التكاملي ARIMA يجب أن نحقق الاستقرارية في التباين والمتوسط، وعليه تم أخذ اللوغارتم الطبيعي لقيم السلسلة ثم أخذ الفرق الأول، وبالإمكان مشاهدة استقرارية السلسلة من خلال الآتي :

1- الرسم البياني لقيم السلسلة يوضح استقرار السلسلة حول المتوسط والتباين وكذلك معاملات الارتباط الذاتي والارتباط الذاتي الجزئي غير معنوية؛ بسبب وقوع المعاملات بعد الازاحتين الثانية أو الثالثة ضمن حدود الثقة، وكما موضح أدناه :



الشكل (3) يمثل سلسلة الثانوي بعد أخذ اللوغارتم الطبيعي والفرق الأول



الشكل (4) يبن فترات الثقة لمعاملات دالتي ACF)،(PACF لسلسلة الثانوي بعد أخذ اللوغارتم والفرق الأول

2- اكد اختبار دكي فلور المطور استقرارية السلسلة حيث ظهرت قيمة p-value أقل من 0.05 ولكل الاختبارات مما يؤدي إلى قبول الفرضية البديلة التي تنص على استقرارية السلسلة، وكما يأتي:

جدول (4) يبين اختبار دكي فولر لسلسلة بيانات اعداد الطلبة بعد أخذ اللوغارتم والفرق الأول

Model	estimated value	test statistic	p-value
test without constant	**-0.56632**	**-2.37938**	**0.01677**
test with constant	**-0.910645**	**-6.48366**	**7.547e-09**
with constant and trend	**-0.933199**	**-6.58514**	**3.616e-08**

3-2:تحديد (اختيار) الأنموذج الملائم :

بعد الحصول على سلسلة زمنية مستقرة تأتي مرحلة اختيار الأنموذج الملائم إذ إن ظهور أكثر من نموذج قد يلائم بيانات السلسلة الزمنية يؤدي إلى استعمال معايير احصائية عديدة منها معيار اكايكي (AIC)، معيار شوارتز البيزي (SBC) ومعيار حنان-كوين (H-Q) لمعرفة الأنموذج الأفضل ملائمة ،ومن خلال الجدول (5) يمكن ذلك .

جدول (5) يبين قيم معايير المفاضلة بين نماذج $\mathrm{ARIMA}_{(P,d,q)}$ **المعنوية**

Model	AIC	SBC	H-Q	
0)،1،(0	−84.76360	−80.82302	−83.24824	**معنوي**
0)،1،(1	-123.7316	-117.9955	-121.5473	**غير معنوي**
1)،1،(0	-132.1096	-126.3736	-129.9253	**غير معنوي**
1)،1،(1	-130.9427	-123.2947	-128.0303	غير معنوي
0)،1،(2	-137.5171	-129.8690	-134.6047	**غير معنوي**
(1،1،2)	-135.7937	-126.2336	-132.153	غير معنوي
(2،1،2)	-136.0472	-124.5751	-131.6786	غير معنوي
(2،1،0)	-134.1869	-126.5388	-131.2745	غير معنوي
(2،1،1)	-135.8345	-126.2744	-132.1939	غير معنوي
(0،1،3)	-135.5509	-125.99089	-131.9104	غير معنوي

(1،1،3)	-135.4006	-123.9285	-131.0319	**غير معنوي**
(2،1،3)	-134.0831	-125.9908	-131.9104	غير معنوي
(3،1،3)	-134.1030	-118.8069	128.2782-	غير معنوي
(3،1،0)	-138.5735	-129.0134	-134.9329	**غير معنوي**
(3،1،1)	-138.6943	-127.2221	-134.3256	غير معنوي
(3،1،2)	−83.90399	−70.11195	−78.60024	معنوي

4-2 : تقدير معلمات الأنموذج :

بعد توفيق عدد من النماذج، كما في الجدول أعلاه تم التوصل إلى 2 نماذج معنوية وباعتماد على معايير المقارنة وباختيار أقل قيم لهما فإن الأنموذج الأكثر ملائمة لتمثيل السلسلة الزمنية هو $ARIMA_{(3،1،2)}$.

وعلى هذا الاساس تتم تقدير معلمات الأنموذج بطريقة الامكان الأعظم المضبوطة والمستعملة في البرنامج GRETL، وكما في الجدول (6).

جدول (6) يبين معلمات ومعنوية الأنموذج $ARIMA_{(3،1،2)}$

Paramters	**Coefficient**	**Std.error**	**Z**	**P.value**
Const	0.0551461	0.0139572	3.951	7.78e-05
phi_1	1.20621	0.105995	11.38	. 5.27e-030
phi_2	−0.795751	0.118457	−6.718	1.85e-011
theta_1	−1.17976	0.181173	−6.512	7.43e-011
theta_2	0.495830	0.281114	1.764	0.0778
theta_3	0.333228	0.179810	1.853	0.0639
Adjusted R-squared	0.983227			

5-2: اختبار ملائمة النموذج :

بعد اختبار النموذج تبين أنه ملائم وذلك من خلال تطبيق اختبار Ljung-Box؛ إذ تبين أن قيمة الاختبار هي (2.589) ومن خلال قيمة P-Value يودي إلى قبول الفرضية البديلة التي تنص على وجود ارتباط بين الاخطاء العشوائية.

$H_0: no\ autocorrelation$

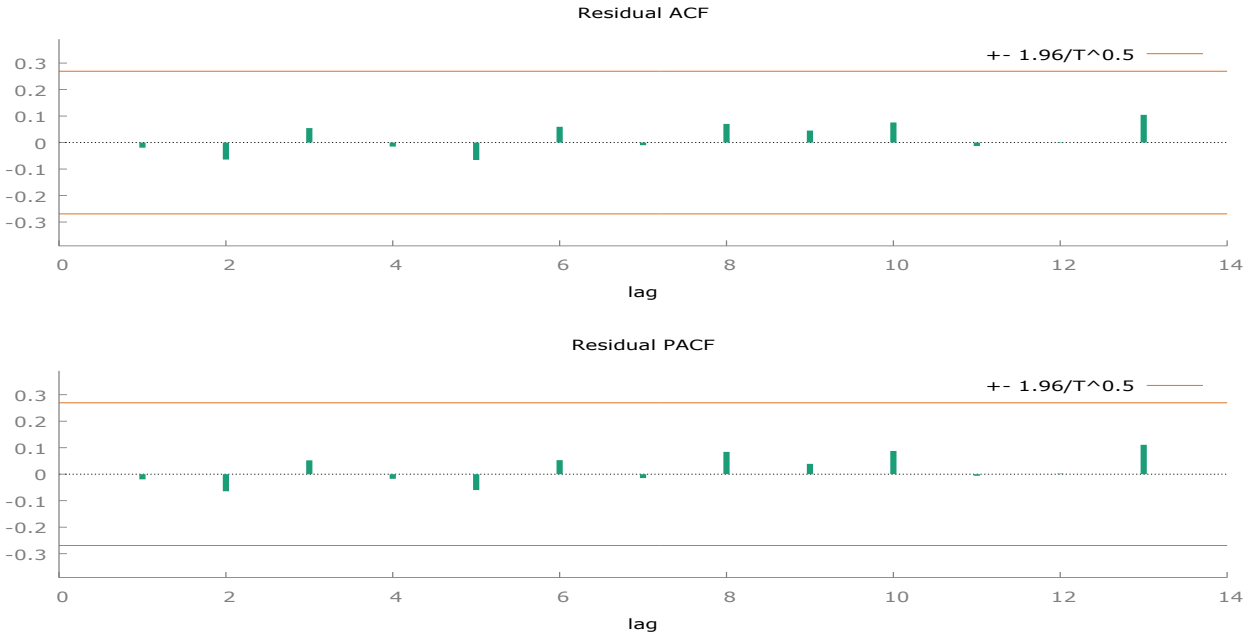
$H_1: autocorrelation$

Test for autocorrelation up to order 13

Ljung-Box Q' = 2.58733،

with p-value = P(Chi-square(8) > 2.58733) = 0.9575

وعند إجراء اختبار البواقي للنموذج المقدر ورسمها، كما في الشكل رقم (5) نلحظ أن معاملات الارتباط الذاتي للبواقي تقع جميعها ضمن حدود الثقة مما يعني أن الأنموذج المستعمل ملائم.



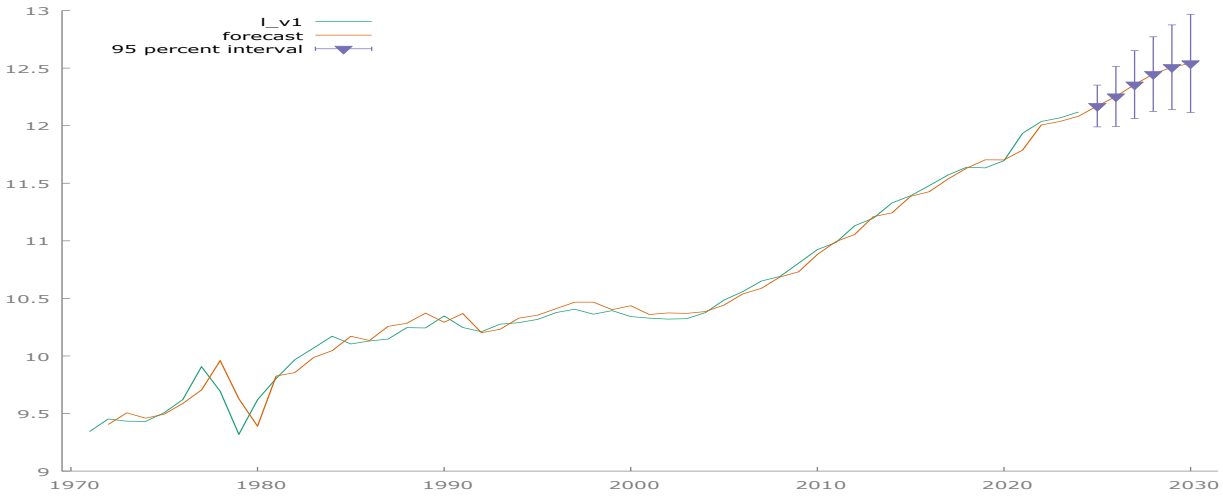
شكل (5) يبين فترات الثقة لمعاملات (ACF، (PACFلسلسلة البواقي لأنموذج $ARIMA_{(3)،1،2}$

6-2:التنبؤ : Forecasting

بعد اختيار النموذج الملائم والكفوء تأتي المرحلة الأخيرة، وهي التنبؤ بأعداد الطلبة الثانوي، وكما في الجدول (8) والشكل (7).

جدول (8) القيم المتنبأ بها لعدد طلبة المرحلة الثانوية للمدة الزمنية (2025- 2030)

القيم التنبؤية	**193150**	**209818**	**232582**	**254740**	**270763**	**279567**
السنة	**2025**	**2026**	**2027**	**2028**	**2029**	**2030**



الشكل (7) رسم القيم التنبؤية لسلسلة اعداد طلبة الثانوي

4- الاستنتاجات

- ظهرت السلسلة الزمنية لأعداد الطلبة الثانوي غير مستقرة .
- ظهر الأنموذج $ARIMA_{(2،1،3)}$ ملائم للتنبؤ .
- أعداد الطلبة الثانوي التنبؤية للمدة الزمنية $(2025 - 2030)$ في تزايد مستمر .

5- التوصيات

- الاستفادة من نتائج البحث من الجهات ذات العلاقة.
- التنبؤ بأعداد الطلبة الثانوي لمحافظة كربلاء المقدسة حسب قواطع المحافظة لمعرفة احتياجات كل قاطع من المستلزمات التربوية، إذ ان القاطع الواحد يظم مجموعة مناطق.
- استعمال طرائق تقدير اخرى ومقارنتها مع أنموذج $ARIMA_{(2،1،3)}$

المصادر

1- Ahmed، Abu Dharr Yusuf Ali، (2023)، "Using ARIMA models to Forecasting price index series In Sudan for the period from 1970 - 2012" Journal of the College of Economics and Political Science، No. 23

2- Albahli، S. & Yar، G.N.A.H. (2022). Defect Forecastingion Using Akaike and Bayesian Information Criterion. Comput. Syst. Sci. Eng.، 41(3)، 1117-1127

3- Al-Borai، Sarah Saad، and Amal Al-Sayed Abdel-Ghani، (2019) "Forecasting using the autoregressive model and fractional integral moving averages (ARFIMA) (p، d، q)، an applied study،" Scientific Journal of Financial and Commercial Studies and Research، Faculty of Commerce - Damietta University، Volume 4، No 1

4- Al-Dub، Tariq Abdul Aziz ، (2021)،" Using time series analysis models to Forecasting total crimes inKuwait"، Journal of Gulf and Arabian Peninsula Studies، Volume 47، No. 81

5- Alsharif، M.H.، Younes، M.K. & Kim، J. (2019). Time series ARIMA model for Forecastingion of daily and monthly average global solar radiation: The case study of Seoul، South Korea. Symmetry، 11(2)، 240

6- Dahrouj، Samir Farag Rashid ، (2023)، " using SARiMA Models to forecat seasonal Time series "Journal of Business Research، Zagazig College of Commerce، Volume 45; No.4

7- Hamza، Lama Karim، and Iqbal Mahmoud Alwan،(2020)،"Using Time Series Analysis to Forecastinging of the Amount of Solid Waste for Baghdad Governorate From 2008 - 2018" Journal of Al-Rafidain University College of Science، Proceedings of the Fifteenth and Second International Scientific Conference on Statistical Applications - Iraqi Society for Statistical Sciences ، Issue 46

8- Kumar، A.، Deepankar، J.، PK، M. & Kumar، A. (2019). Wheat yield forecasting in Haryana: A time series approach. Bull. Env. Pharmacol. Life Sci، 8(3)، 63-69

9- Kumar، N. & Susan، S. (2020). COVID-19 pandemic Forecastingion using time series forecasting models. In 2020 11th international conference on computing، communication and networking technologies (ICCCNT) (pp. 1-7). IEEE .وISO 690.

10-kadhim Al-Quraishi، O. A. (2021، May). Choosing the best eestimated regression equation for data subject to geometric distribution (Student data as a case study). In *Journal of Physics: Conference Series* (Vol. 1879، No. 3، p. 032045). IOP Publishing.

11-Mdloul، N.S.، Ahmed، J.S. & Battal، A. H. (2022). Using time series methods to Forecasting the value of agricultural output and some financial indicators affecting it in Iraq for the period (2021q1-2025q4). Tikrit Journal for Agricultural Sciences. 22(1)، 78-95

12- *Percival، D. B.، & Walden، A. T. (2000)." Wavelet methods for time series analysis and its statistical applications". (pp. 1 613). doi.org/10.1017/cbo9780511841040.*

13-*Percival، D. B.، & Walden، A. T. (2000)." Wavelet methods for time series analysis and its statistical applications". (pp. 1 613). doi.org/10.1017/cbo9780511841040.*

14-Saleem، N، Akbar، A.، Imon، A.R. & Al Mamun، A.S. M.(2022) Time series analysis modeling and forecasting of gross domestic product of pakistan.و journal of mechanics of continua and mathematical sciences

15-Zidan، Imad Abdel Sattar Taha، (2021)، "Application of time series in Forecastinging the number of visitors to the Arts Library at Kafrelsheikh University،" International Journal of Library and Information Sciences، Volume 8، No.4