



التنبؤ بتعويضات صندوق الضمان الاجتماعي في العراق باستخدام نموذج ARMAX □

أ.م.د. فراس احمد محمد
جامعة بغداد /كلية الادارة والاقتصاد /قسم الاحصاء
zaid.khalil.ismail@gmail.com

الباحث /زيد خليل إسماعيل
جامعة بغداد / كلية الإدارة والاقتصاد / قسم الإحصاء
zaid.khalil.ismail@gmail.com

Received: 21/6/2020

Accepted :26/7/2020

Published : January / 2021

هذا العمل مرخص تحت اتفاقية المشاع الابداعي نسب المُصنّف - غير تجاري - الترخيص العمومي الدولي 4.0
[Attribution-NonCommercial 4.0 International \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)



مستخلص البحث

اكتسبت السلاسل الزمنية اهمية كبيرة وطبقت بشكل في المجالات الاقتصادية والمالية والصحية والاجتماعية و تستخدم في التحليل عن طريق دراسة التغيرات والتنبؤ بمستقبل الظاهرة ومن اهم نماذج الصندوق الأسود نموذج "ARMAX" وهو نموذج مختلط مكون من الانحدار الذاتي مع متوسطات متحركة مع مدخلات خارجية. وان بناء النموذج يتألف من عدة مراحل وهي تحديد رتبة النموذج وعملية تقدير معالم النموذج ثم عملية التنبؤ لمعرفة حجم التعويضات الممنوحة للعمال مستقبلا من اجل ايفاء الصندوق بالتزاماته المستقبلية. وتم تحديد رتبة النموذج وفق المعايير FPE, MDL, AIC, وتقدير النموذج بطريقة المربعات الصغرى الاعتيادية وطريقة المربعات الصغرى التكرارية بواسطة مرشح كالمن وعامل التغاضي للحصول على افضل طريقة يمكن التنبؤ من خلالها للمستقبل باستخدام بيانات اخذت من دائرة الضمان الاجتماعي للعمال ومن الفترة 2013/1 الى 2019/6 حيث تم مقارنة طريقة المربعات الصغرى OLS مع طريقة المربعات الصغرى التكرارية RLS بواسطة مرشح كالمن وعامل التغاضي ووجد ان طريقة OLS هي الافضل وفق مقاييس المقارنة MAPE, RMSE, وانها اعطت نتائج دقيقة وقريبة من القيم الحقيقية.

المصطلحات الرئيسية للبحث / معيار اكاكي AIC - معيار خطأ التنبؤ FPE - معيار اكاكي البيزي BIC - طريقة المربعات الصغرى الاعتيادية OLS - طريقة المربعات الصغرى التكرارية RLS - مرشح كالمن - عامل التغاضي - التنبؤ

بحث مستل من رسالة ماجستير

1. المقدمة Introduction

تستند البيانات المعتمدة على الزمن على فلسفة رصد القيم التي تعبر عن الظاهرة خلال فترات زمنية متتالية لاكتشاف اسلوب تطور الظاهرة واستخدامه فيما بعد للتنبؤ بتلك الظاهرة في المستقبل ويسمى هذا النوع من البيانات بالسلاسل الزمنية ، وتعد السلسلة الزمنية من اهم اساليب التنبؤ المستقبلي (الاستشراف ، التكهون) وان اهداف استخدام السلسلة الزمنية هو التوصل الى وصف دقيق للعملية التي تتولد منها السلسلة والتحقق من البيانات الشاذة ان وجدت وبناء انموذج ملائم والتحقق من ملائمة هذا الانموذج لغرض التنبؤ للمستقبل .ومن اهم نماذج السلاسل الزمنية نماذج الصندوق الاسود نموذج ARMAX اذ يعد هذا الانموذج المختلط والمكون من الانحدار الذاتي والمتوسطات المتحركة مع المدخلات الاضافية من النماذج الكفوءة في التنبؤ بقيمة الظاهرة لذلك تم توظيف هذه النماذج للتنبؤ بسلسلة المخرجات والتي تمثل التعويضات بتأثير سلسلة المدخلات التي تمثل الاشتراكات لمعرفة القيم المستقبلية حسب النموذج ARMAX

يهدف هذا البحث الى بناء نموذج سلسلة زمنية للتنبؤ بقيمة تعويضات صندوق الضمان الاجتماعي الممنوحة للعمال جراء تحقق الاخطار المؤمنة على العمال وهي الوفاة والعجز والمرض ولغرض التقاعد من خلال تأثير الاشتراكات المستحصلة من العمال شهريا على واردات الصندوق لنظام التأمينات الاجتماعية وذلك باستخدام نموذج ARMAX والاستفادة من التنبؤ من قبل الجهات المالية والرقابية لوضع استراتيجيات لتلافي خطر العجز في الاموال ولتعظيم رأس المال من خلال استغلال الفائض .

تأتي اهمية البحث لأنه يمثل اداة تنبؤ لصندوق الضمان الاجتماعي ولكونه اداة فعالة لضمان الوفاء بالتزامات الصندوق المالية ولاعداد خطة لمواجهة الخطر واتباع استراتيجيات يمكن الاعتماد عليها واداة لتقويم الصندوق واستخدام البحث من خلال الجهات الرقابية لتلافي الاخطار المستقبلية قبل حدوثها ولتأمين حياة العامل من اجل ان يعيش حياة كريمة .

ومن الدراسات السابقة في سنة 1986 قدم الباحث David بحثا ناقش فيه التنبؤ بالتباينات لمتعدد المتغيرات بالاعتماد على تحديد معالم نموذج ARMAX بحيث ان المقدر الجديد وعملية التشخيص تعملان مع جميع انواع الفقدان في البيانات غير التامة [3].

وايضا من الدراسات تطرق الباحثان في سنة 1999 Noah & John الى توقعات صندوق الضمان الاجتماعي باقتراح نموذج اكتوبري تضمن طريقة بناء النموذج ومعالجة المخرجات بتأثير المدخلات للتنبؤ بتمويل صندوق الضمان الاجتماعي [14].

وفي سنة 2015 قام الباحثان Gokhan & Bahattin بنشر بحث ناقشا فيه كيفية التنبؤ في عدد الحاصدات خلال الفترة 2009-2020 في تركيا معتمدين على استخدام نماذج ARMAX للتنبؤ للفترة المذكورة معتمدين على بيانات حقيقية للنتائج المحلي الاجمالي الزراعي في تركيا [5].

2. الطرائق والاساليب

نموذج الانحدار الذاتي والمتوسطات المتحركة مع مدخلات خارجية (ARMAX Model)

Autoregressive Moving Average With Exogenous Input Model

يقوم هذا النموذج بوصف سلسلة المخرجات للمتغير التابع $y(t)$ للظاهرة المراد التنبؤ بها من

خلال دمج نموذج ARIMA مع سلسلة المدخلات الخارجية $x(t)$ وتم تطوير هذا النموذج

وهو مطبق في مجالات الرقابة والاقتصاد القياسي والنموذج يكون بالصيغة الاتية [2] [4] [11] [13]:

$$y(t) + a_1y(t-1) + \dots + a_nay(t-n_a) = b_1u(t-n_k) + \dots + b_nbu(t-n_k-n_b+1) + c_1e(t-1) + \dots + c_nce(t-n_c) + e(t)$$

حيث ان

na = رتبة AR للانحدار الذاتي

nk = زمن التأخير

nb = عدد حدود المدخلات

nc = رتبة MA للمتوسطات المتحركة

وبهذا يمكن كتابة النموذج بالشكل الاتي :

$$A(L) y(t) = B(L) u(t - n_k) + C(L) e(t)$$

حيث ان

$$A(L) = 1 + a_1L^{-1} + a_2L^{-2} + \dots + a_nL^{-na}$$

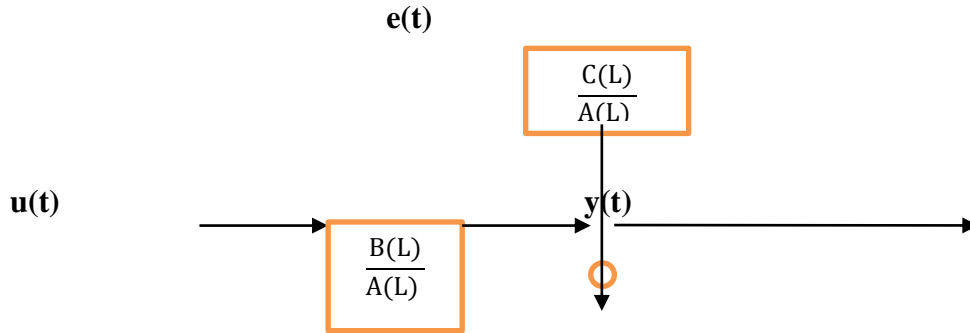
$$B(L) = b_1L^{-1} + b_2L^{-2} + \dots + b_nL^{-nb+1}$$

$$C(L) = 1 + c_1L^{-1} + c_2L^{-2} + \dots + c_{nc}L^{-nc}$$

ومعلمات النموذج

$$\theta = [a_1 \ a_2 \dots a_{na} \ b_1 \ b_2 \dots b_{nb} \ c_1 \ c_2 \dots c_{nc}]^T$$

= L الازاحة الى الخلف



شكل يوضح هيكلية انموذج ARMAX

الارتباط الذاتي (Autocorrelation function ACF)

تقوم دالة الارتباط الذاتي بتوضيح الارتباطات بين مشاهدات السلسلة الزمنية اي الارتباطات الداخلية بين السلسلة الزمنية وهو عبارة عن مقياس يقوم بقياس مدى قوة الارتباط لمشاهدات المتغير عند ازاحة زمنية معينة .

$$\rho_k = \frac{\text{cov}(X_t, X_{t+k})}{\sigma_{X_t} \cdot \sigma_{X_{t+k}}} \quad \rho_k = \gamma_k / \gamma_0$$

ويقدر معامل الارتباط الذاتي عند الازاحة k

$$r_k = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2}$$

وتتراوح معاملات الارتباط بين (-1,1) وهي اداة للتعرف على السلسلة .

دالة الارتباط الذاتي الجزئي (Partial Autocorrelation Function PAC)

وتستخدم هذه الدالة مع دالة الارتباط الذاتي في تحليل نماذج بوكس جينكنز كشيء اساسي لغرض تمييز النموذج وتتقوم بقياس الارتباط الجزئي بين القيم المتتالية لفترتين مع ثبوت الفترات الاخرى اي بين Y_t و Y_{t-k} وفي تحديد رتبة النموذج .

$$P(kk) = \frac{rk - \sum_{j=1}^{k-1} Pk-1,j rk-j}{1 - \sum_{j=1}^{k-1} Pk-1,j rj} \quad k = 2, 3, \dots$$

$$P(kj) = P(k-1, j) - P(kk) P(k-1, (k-j)) \quad j = 1, 2, 3, \dots, k-1$$

معايير تحديد رتبة النموذج

عملية اختيار النموذج الملائم ذلك عن طريق عدة معايير يقوم باستخدامها لكي يصل الى النموذج المطلوب وجميعها تكون مرشحة وغير مرفوضة ولهذا السبب وضعت هذه المعايير للمساعدة في تحديد النموذج الافضل [15]

معايير أكايكي (Akaike Information Criteria AIC)

اقترح هذا المعيار من قبل العالم Akaike عام 1973-1974 وعن طريق هذا المعيار اختيار رتبة النموذج الملائم من بين النماذج المرشحة بحيث يكون النموذج الافضل يقابل اقل قيمة من معيار AIC ويمثل الرتبة الاكثر ملائمة ويتم التعبير عنه كالآتي [13]:

$$AIC(B) = \ln(\sigma^2_{(B)}) + \frac{2B}{n}$$

حيث ان :

$$B = \text{عدد معلمات النموذج} \quad n = \text{عدد المشاهدات} \quad \sigma^2_{(B)} = \text{مقدار تباين الخطأ}$$

معايير (FPE)

هو معيار خطأ التنبؤ ووصف من قبل العالم اكاكي في عام 1969 وهو خطأ التباين ويحسب بالصيغة الآتية [11]:

$$FPE(B) = (\sigma^2_{(B)}) + 2\lambda \frac{B}{n}$$

وان صيغة λ كما يلي :

$$\lambda = \sigma^2_{(B)} \frac{1}{1 - \frac{B}{n}}$$

وعند تعويض قيمة λ في المعادلة السابقة

$$FPE(B) = (\sigma^2_{(B)}) \frac{n+B}{n-B}$$

حيث ان :

$$B = \text{عدد معلمات النموذج} \quad n = \text{عدد المشاهدات} \quad \sigma^2_{(B)} = \text{مقدار تباين الخطأ}$$

معايير اكاكي البيزي (BIC)

يسمى هذا المعيار بمعيار الوصف الاصغر ويكتب (MDL) اختصارا الذي اقترحه العالم Akaike عام 1979 وايضا العالم Rissanen عام 1978 ويكتب بالصيغة الآتية [11]:

$$MDL = \ln(\sigma^2_{(B)}) + \frac{B}{n} \ln(n)$$

حيث ان :

$$B = \text{عدد معلمات النموذج} \quad n = \text{عدد المشاهدات} \quad \sigma^2_{(B)} = \text{مقدار تباين الخطأ}$$

تقدير معلمات النموذج**طريقة المربعات الصغرى الاعتيادية (Ordinary least square method (OLS)**

تم تطوير طريقة المربعات الصغرى بواسطة Gauss في عام 1795 وهي من الحلول المثلى والاكثر تطبيقا في التحسين الخطي والهدف هو ايجاد ناتج النموذج \hat{Y} الذي يقدر مخرجات النموذج y ويمكن استخدام هذه الطريقة لتقدير معلمات نموذج ARMAX لجعل مجموع مربعات الخطأ اقل ما يمكن .

وعلى افتراض ان النموذج يكون بالصورة التالية [16] [13] [9]:

$$y(t) + a_1y(t-1) + \dots + a_nay(t-n_a) = b_1u(t-n_k) + \dots + b_nbu(t-n_k-n_b+1) + \dots + c_1e(t-1) + \dots + c_nce(t-n_c) + e_t$$

$$y(t) + a_1y(t-1) + \dots + a_nay(t-n_a) \quad Y = \text{المخرجات}$$

$$b_1u(t-n_k) + \dots + b_nbu(t-n_k-n_b+1) \quad U = \text{المدخلات}$$

$$c_1e(t-1) + \dots + c_nce(t-n_c) = \quad e = \text{الاضاء}$$

وان

$$\phi^T(t) = [-y(t-1) - \dots - y(t-n_a) \quad u(t-n_k) + \dots + u(t-n_k-n_b+1) \quad e(t-1) + \dots + e(t-n_c)]$$

والمعاملات الخاصة بالنموذج هي :

$$\theta = [a_1 \ a_2 \ \dots \ a_{n_a} \ b_1 \ b_2 \ \dots \ b_{n_b} \ c_1 \ c_2 \ \dots \ c_{n_c}]^T$$

فان

$$y(t) = -a_1 y(t-1) - \dots - a_{n_a} y(t-n_a) + b_1 u(t-n_k) + \dots + b_{n_b} u(t-n_k-n_b+1) + c_1 e(t-1) + \dots + c_{n_c} e(t-n_c)$$

$$y(t) = \varphi^T(t)\theta + e(t)$$

وبما ان الاخطاء غير متوفرة كمشاهدات لذلك نقوم بتقدير المعلمات كالآتي :

$$\theta = [a_1 \ a_2 \ \dots \ a_{n_a} \ b_1 \ b_2 \ \dots \ b_{n_b}]$$

$$\varphi^T(t) = [-y(t-1) - \dots - y(t-n_a) \ u(t-1) + \dots + u(t-n_b)]$$

ولكي نحصل على مقدر خطأ للتنبؤ :

$$\hat{\theta}_{PE} = \arg \min_{\theta} V_n$$

ولكي نجعل مجموع المربعات للخطأ اقل مقدار :

$$V_n = \frac{1}{n} \sum_{t=1}^n (y(t) - \varphi^T(t)\hat{\theta})^2$$

$$\hat{y}(t) = \varphi^T(t)\hat{\theta}$$

ومن خلال اشتقاق المعادلة ومساواتها للصفر

$$\frac{d}{d\theta} V_n = -\frac{2}{n} \sum_{t=1}^n \varphi(t) (y(t) - \varphi^T(t)\hat{\theta}) = 0$$

$$\sum_{t=1}^n \varphi(t)y(t) = \sum_{t=1}^n \varphi(t)\varphi^T(t)\hat{\theta}$$

$$\hat{\theta} = [\sum_{t=1}^n \varphi(t)\varphi^T(t)]^{-1} [\sum_{t=1}^n \varphi(t)y(t)]$$

ويمكننا اعادة الصياغة بالشكل التالي

$$y(t) = \varphi^T(t)\theta + (c_1 e(t-1) + \dots + c_{n_c} e(t-n_c)) + e(t)$$

ونحصل على المقدرات بالصيغة الآتية :

$$\hat{\theta} = [\sum_{t=1}^n \varphi(t)\varphi^T(t)]^{-1} [\sum_{t=1}^n \varphi(t)y(t)]$$

طريقة المربعات الصغرى التكرارية (RLS) recursive least square

تقوم هذه الطريقة بحساب المعلمة $\hat{\theta}(t)$ في الوقت t وحساب المعلمة $\hat{\theta}(t-1)$ عند الفترة $t-1$ عن طريق بعض التعديل البسيط وهذه الصيغة هي تعاقب من طريقة المربعات الصغرى وخطوات هذه الطريقة تبدأ من تعريف تقدير المعلمة حسب طريقة المربعات الصغرى الاعتيادية وكالآتي :

ان تقدير المعلمة في الزمن t هو [12] [16] [17]:

$$\hat{\theta}(t) = [\varphi(t)\varphi^T(t)]^{-1} [\varphi(t)y(t)]$$

اما في الزمن $t+1$ هو :

$$\hat{\theta}(t+1) = [\varphi(t+1)\varphi^T(t+1)]^{-1} [\varphi(t+1)y(t+1)]$$

وان حدود $y(t+1)$ و $\varphi(t+1)$ تكتب بالشكل الآتي :

$$\varphi^T(t+1) = \begin{bmatrix} \varphi^T(t) \\ \varphi(t+1) \end{bmatrix}$$

$$y(t+1) = \begin{bmatrix} y(t) \\ y(t+1) \end{bmatrix}$$

$$\varphi(t+1)\varphi^T(t+1) = [\varphi(t) \ \varphi^T(t+1)] \begin{bmatrix} \varphi^T(t) \\ \varphi(t+1) \end{bmatrix}$$

$$= \varphi(t)\varphi^T(t) + \varphi^T(t+1)\varphi(t+1)$$

$$\varphi(t+1)y(t+1) = [\varphi(t) \ \varphi^T(t+1)] \begin{bmatrix} y(t) \\ y(t+1) \end{bmatrix}$$

$$= \varphi(t)y(t) + \varphi^T(t+1)y(t+1) \dots \dots \dots (1)$$

ويتم تعريف مصفوفة التغيرات في الزمن t :

$$p(t) = [\varphi(t) \quad \varphi^T(t)]^{-1} = [\sum_{t=1}^n (\varphi(t) \varphi^T(t))]^{-1}$$

اما في حالة الزمن $t+1$ فان مصفوفة التغيرات تصبح :

$$p(t+1)^{-1} = p(t)^{-1} + \varphi^T(t+1)\varphi(t+1) \dots \dots \dots (2)$$

ويمكن اعادة كتابة مصفوفة التغيرات

$$\hat{\theta}(t) = p(t)\varphi(t)y(t)$$

$$\hat{\theta}(t+1) = p(t+1)\varphi(t+1)y(t+1) \dots \dots \dots (3)$$

ومن خلال توظيف المعادلة (1) و (3) وبعد الترتيب نجد ان :

$$\hat{\theta}(t+1) = p(t+1)[\varphi(t)y(t) + \varphi^T(t+1)y(t+1)]$$

$$= p(t+1)\varphi(t)y(t) + p(t+1)\varphi^T(t+1)y(t+1) \dots \dots \dots (4)$$

وبضرب طرفي المعادلة بمصفوفة التغيرات $p(t)^{-1}$ نجد :

$$\varphi(t)y(t) = p(t)^{-1}\hat{\theta}(t)$$

وعند دمج معادلة (2) و (3) نحصل على :

$$\varphi(t)y(t) = p(t+1)^{-1}\hat{\theta}(t) - \varphi^T(t+1)\varphi(t+1)\hat{\theta}(t)$$

ومن خلال تعويض الحد $\varphi(t)y(t)$ في العلاقة رقم (4)

$$\hat{\theta}(t+1) = p(t+1)[p(t+1)^{-1}\hat{\theta}(t) - \varphi^T(t+1)\varphi(t+1)\hat{\theta}(t)]$$

$$+ p(t+1)\varphi^T(t+1)y(t+1)$$

وبعد تبسيط الصيغة :

$$\hat{\theta}(t+1) = \hat{\theta}(t) + p(t+1)\varphi^T(t+1)[y(t+1) - \varphi(t+1)\hat{\theta}(t)] \dots \dots (5)$$

الحد $[y(t+1) - \varphi(t+1)\hat{\theta}(t)]$ يمثل الفرق بين سلسلة المخرجات وقيمة التنبؤ بخطوة واحدة

والذي يمثل تقدير خطأ التنبؤ في الزمن $t+1$

اي ان

$$e(t+1) = y(t+1) - \varphi(t+1)\hat{\theta}(t)$$

وان متجه المعلمات $\hat{\theta}(t+1)$ في المعادلة (5) سيكون :

$$\hat{\theta}(t+1) = \hat{\theta}(t) + p(t+1)\varphi^T(t+1)e(t+1)$$

وان مصفوفة $p(t+1)$ يتم ايجادها عن طريق قانون معكوس المصفوفة

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

وحسب المعادلة رقم (2) فان الحدود تكون

$$A = p(t)^{-1} \quad C = 1 \quad B = \varphi^T(t+1) \quad D = \varphi(t+1)$$

والتي تمثل مصفوفة مربعة يكون المحدد لها غير مساوي للصفر اي ان المصفوفة تمتلك معكوس وبالتالي فان

القيمة تكون :

$$p(t+1) = p(t)[Im - \varphi^T(t+1)(1 + \varphi^T(t+1)p(t)\varphi(t+1))^{-1}\varphi(t+1)p(t)]$$

وبعد التبسيط :

$$p(t+1) = p(t) - \frac{p(t)\varphi^T(t+1)\varphi(t+1)p(t)}{1 + \varphi^T(t+1)p(t)\varphi(t+1)}$$

وتتمثل المربعات الصغرى التكرارية بنهجين :

النهج الاول : مرشح كالمن Kalman Filter

يعتبر مرشح كالمن نهج مباشر عن طريقه يتم تقدير ومعالجة المعلمات خلال الزمن للحصول على

تقديرات دقيقة وذات جودة للوصول للحالة المثلى لخوارزمية الطريقة عن طريق فرض قيم صحيحة لمصفوفة

التغيرات $R_1(t)$ وبالصيغة الاتية [13] [16]:

نقوم بتقدير خطأ التنبؤ

$$e(t+1, \theta) = y(t+1) - \varphi(t+1)\hat{\theta}(t)$$

ويمثل الحد $\hat{\theta}(t+1)\varphi(t+1)$ من المعادلة مقدار الخطأ للمخرجات
ومن ثم نقدر مصفوفة التغيرات $R1(T)$

$$p(t+1) = R1(t) + [I - K(t+1)\varphi(t+1)]p(t)$$

ثم نقوم بتقدير حد الكسب $K(t+1)$ وكالاتي :

$$K(t+1) = p(t+1)\varphi^T(t+1) = \frac{p(t)\varphi^T(t+1)}{[1 + \varphi(t+1)p(t)\varphi^T(t+1)]}$$

ونقوم بتقدير متجه المعلمات وفق الصيغة :

$$\hat{\theta}(t+1) = \hat{\theta}(t) + K(t+1)e(t+1)$$

النهج الثاني : عامل التفاضل Forgetting factor

يتلخص هذا النهج بتقليل دالة الخسارة الى اقل حد و يعرف بعامل التفاضل من خلال اضافة (λ) الى دالة الخسارة وبالشكل الاتي [16][13][11][8]:

$$V_i(\theta) = \sum_{i=1}^t \lambda^{t-i} e^2(i)$$

$$= \sum_{i=1}^t \lambda^{t-i} (y(i) - \varphi^T(i)\theta)^2$$

سابقا كانت دالة الخسارة تستخدم $(\lambda = 1)$ لكن مع نهج عامل التفاضل تكون فيه قيمة λ عدد موجب اقل من واحد $(\lambda < 1)$ ويقع بين $(0.99 - 0.95)$ الذي يعطي اهمية نسبية للبيانات عن البيانات السابقة من خلال زيادة (t) وتسقيط البيانات التي حصلنا عليها مسبقا

$$\hat{\theta}(t+1) = \hat{\theta}(t) + K(t+1)[y(t+1) - \varphi(t+1)\hat{\theta}(t)]$$

او نكتب الصيغة بدلالة خطأ التنبؤ

$$\hat{\theta}(t+1) = \hat{\theta}(t) + K(t+1)e(t+1)$$

حيث ان $K(t+1)$ يمثل حد الكسب او معامل المنفعة او الربحية الذي يفسر مقدار قيمة خطأ التنبؤ وتكون صيغته كالاتي :

$$K(t+1) = p(t+1)\varphi^T(t+1) = \frac{p(t)\varphi^T(t+1)}{[\lambda + \varphi(t+1)p(t)\varphi^T(t+1)]}$$

لذا فان مصفوفة التغيرات لعامل التفاضل تكون بالشكل الاتي :

$$p(t+1) = \frac{1}{\lambda} [1 - K(t+1)\varphi(t+1)]p(t)$$

مقارنة الطرق للتقدير :

متوسط مربعات الخطأ MSE

تتم مقارنة الطرق عن طريق متوسط مربعات الخطأ وكالاتي :

$$MSE = \frac{1}{n} \sum_{t=1}^n (\theta - \hat{\theta})^2$$

حيث ان

$n =$ حجم العينة $\hat{\theta} =$ القيمة التقديرية $\theta =$ القيمة الحقيقية
المقدر الأفضل يكون يقابل اقل قيمة من MSE
وايضا تتم المقارنة ب :

متوسط مطلق الخطأ Mean Absolute Error

افضل المقدرات الذي يقابل اقل قيمة ويكتب بالصيغة الاتية [17]:

$$MAE = \frac{1}{n} \sum_{t=1}^n |\theta - \hat{\theta}_t|$$

بحيث ان

$n =$ حجم العينة $\hat{\theta} =$ القيمة التقديرية $\theta =$ القيمة الحقيقية

مقياس متوسط مطلق الخطأ النسبي (MAPE)

ان افضل المقدرات الذي يقابل اقل قيمة ويكتب بالصيغة [1],[6]:

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{\theta - \hat{\theta}_i}{\theta} \right|}{n} * 100$$

بحيث ان

$$n = \text{حجم العينة} \quad \hat{\theta} = \text{القيمة التقديرية} \quad \theta = \text{القيمة الحقيقية}$$

فحص ملائمة النموذج

يتم فحص ملائمة النموذج للتنبؤ عن طريق تحسينه او تطويره او الابقاء عليه وتعد هذه المرحلة من اهم المراحل لأنها تحدد القبول بالنموذج او القيام باستبداله , ويتم اختبار النموذج عن طريق الاختبار الاتي :

اختبار (Ljung – Box)

في هذا الاختبار قام الباحثان Ljung&Box في عام 1978 بتعديل اختبار Box&Prece ويستخدم الاختبار للتأكد من عشوائية البواقي بحساب معاملات الارتباط للبواقي لعدد من الازاحات حسب الفرضية [10]:

$$H_0 : P_1 = P_2 = \dots = P_k = P_h \quad \text{for } k = 1, 2, \dots, h$$

$$H_1 : P_h \neq 0$$

بان الاختبار يأخذ بنظر الاعتبار اقتراب او ابتعاد الارتباط الذاتي عن اعلى قيمة h التي دخلت في احتساب احصاءة الاختبار كما يلي :

$$Q^* = n(n+2) \sum_{k=1}^h (n-k)^{-1} r^2 k$$

ويتم المقارنة مع توزيع مربع كاي من الدرجة (h-m) اي $X^2(h-m)$

حيث ان :

$r^2 k$ = معامل الارتباط الذاتي = n = حجم العينة = h = عدد الازاحات = m = عدد المعالم
نقبل فرضية العدم اي ان البواقي عشوائية غير مرتبطة $Q^* < X^2(h-m)$ فاذا كانت القيمة المحسوبة اما اذا كانت اكبر نرفض فرضية العدم اي ان البواقي غير عشوائية

التنبؤ

بعد الحصول على القيم (na,nb,nc) لمعاملات النموذج من خلال طرق التقدير يصبح النموذج كالآتي [2]:
 $y(t) + a_1 y(t-1) + \dots + a_n y(t-n) = b_1 u(t-1) + \dots + b_n u(t-n) + c_1 e(t-1) + \dots + c_n e(t-n) + e_t$

اما التنبؤ لخطوة واحدة فقط يكون بالصيغة الاتية :

$$\hat{y}_t(1) = \hat{y}_t(t+1) = a y(t) - \dots - a_n y(t-n+1) + b_1 u(t) + \dots + b_n u(t-n+1) + c_1 e(t) + \dots + c_n e(t-n+1)$$

وهذه تعتبر الخطوة الاخيرة والمراد تحقيقها من اجل معرفة القيم المستقبلية لحالة صندوق الضمان ودراسة الحالة لتجاوز وتلافي الاخطار التي من الممكن مواجهتها وتعيق التزامات الصندوق المالية .

3. تحليل البيانات**1- متغيرات الدراسة والعلاقة بينهما**

ان الصندوق يعتمد على متغيرين اساسيين :

1- الاشتراكات : الية تسديد شهرية ومقدارها 17% من راتب العامل يسدد العامل منها 5% ويسدد صاحب العمل 12% عن كل عامل يعمل لديه , وهذا مقابل ضمان العامل ضد الاخطار .

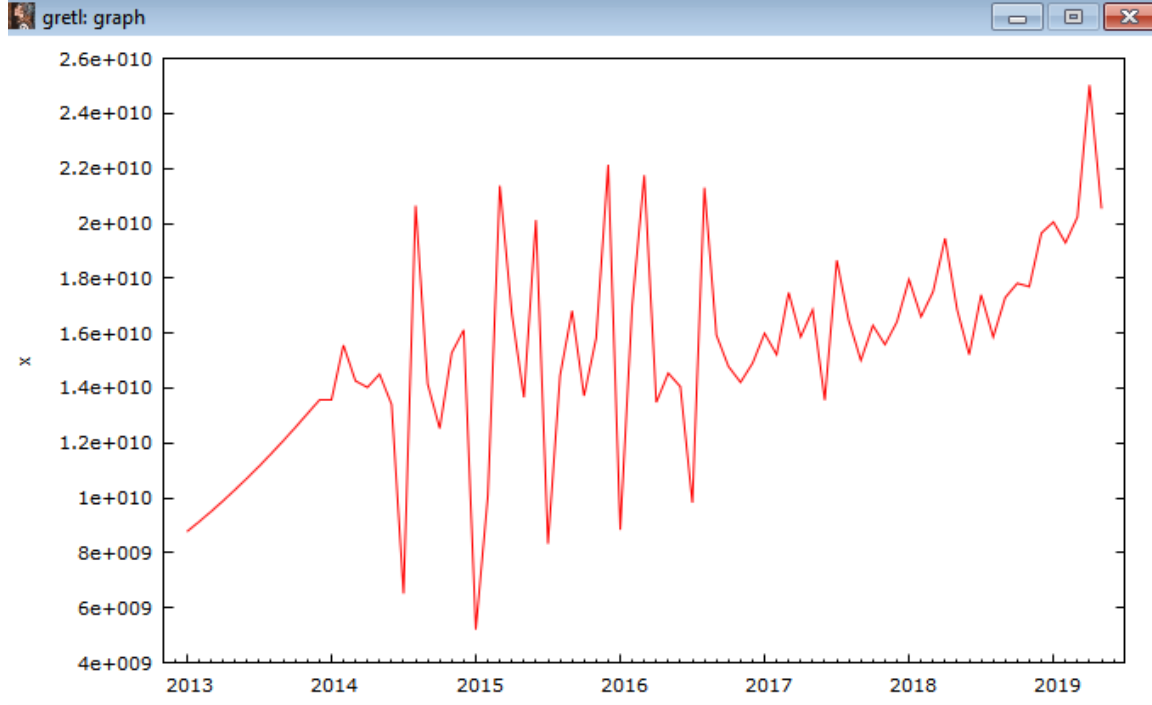
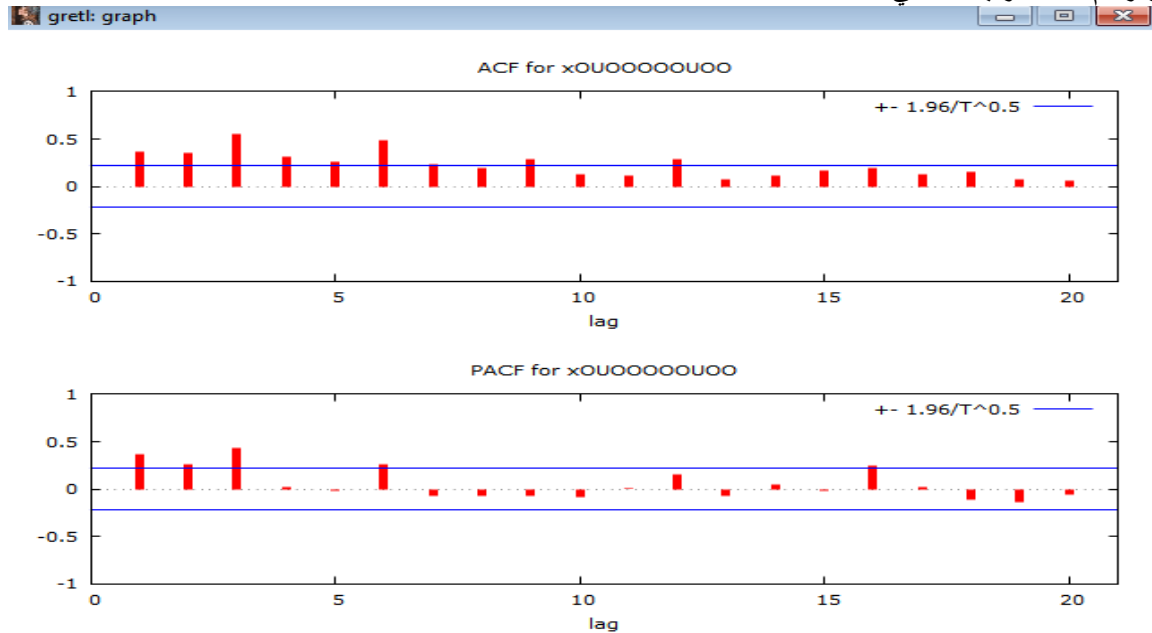
2- التعويضات : عند تحقق احد الاخطار مثل العجز او المرض او الاصابة او الوفاة اثناء العمل وتتمثل التعويضات برواتب تقاعدية ومكافآت نهاية الخدمة .

ان العلاقة بين الاشتراكات والتعويضات علاقة تأثير قوية وذلك لان الاشتراكات تؤثر في التعويضات فعند ازدياد العمال المشتركين تحت مظلة الضمان الاجتماعي فذلك يعني ازدياد منح التعويضات ولكن زيادة التعويضات لا تؤثر في زيادة الاشتراكات لان منح التعويضات هو من جراء تحقق احد الاخطار .

2- مراحل بناء نموذج ARMAX (na,nc,nb,nk)

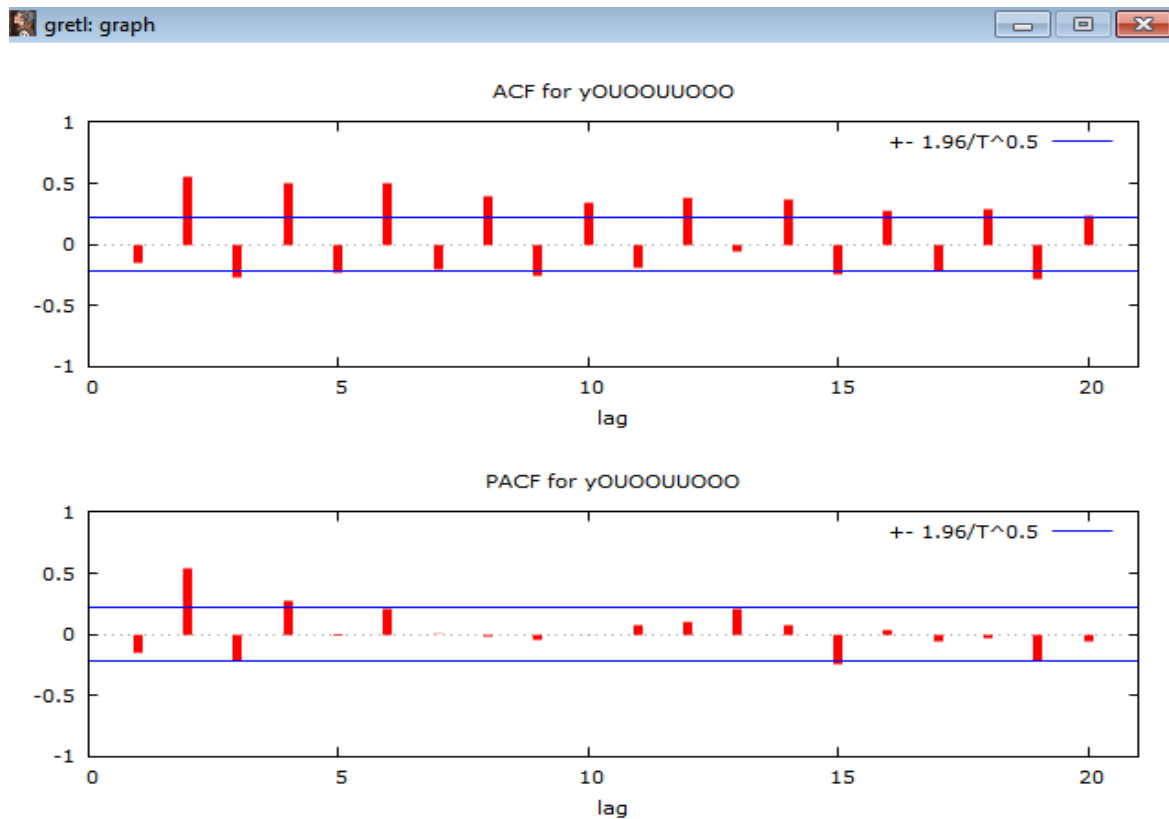
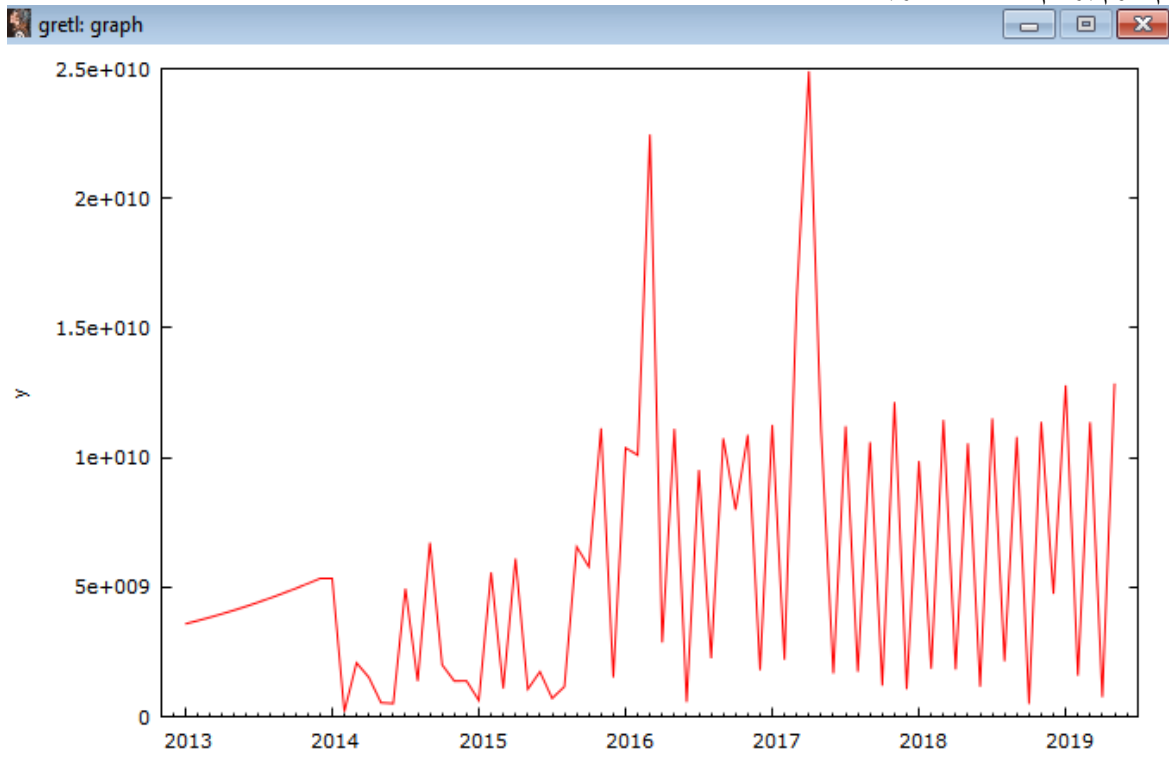
مرحلة التعرف على السلسلة :

تمثلت بيانات سلسلة المدخلات X_t بسلسلة الاشتراكات وسلسلة المخرجات Y_t بسلسلة التعويضات وبيانات السلسلتين للفترة من شهر كانون الثاني 2013 الى شهر حزيران 2019 وتم الحصول على البيانات من وزارة العمل والشؤون الاجتماعية نبدأ بالتحرف على السلسلة من خلال الرسم لسلسلة المدخلات X_t والمخرجات Y_t التي سيتم التنبؤ بها بتأثير سلسلة المدخلات عليها بنموذج ARMAX .

ورسم دالة الارتباط الذاتي لسلسلة المدخلات X_t 

نلاحظ وجود تزايد في قيم الاشتراكات من خلال الرسم ووجود علاقة ارتباط بين قيم السلسلة الزمنية من خلال رسم دالة الارتباط الذاتي والارتباط الجزئي وان الارتباطات بعضها يقع خارج حدود الثقة .

ثم نقوم برسم سلسلة المخرجات Y_t



نلاحظ وجود تذبذب في سلسلة التعويضات ووجود علاقة ارتباط بين قيم السلسلة الزمنية من خلال رسم السلسلة ورسم دالة الارتباط الذاتي والارتباط الجزئي وان الارتباطات تقع خارج حدود الثقة .

تحديد رتب النموذج ARMAX(na,nc,nb,nk)

يتم التعرف على النموذج من خلال تحديد رتب na, nc, nb وتم تحديد الرتب عن طريق ايجاد المعايير AIC,FPE,MDL للنماذج وفق برنامج بلغة ماتلاب MATLAB من قبل الباحث لاختيار الافضل عن طريق تحديد النموذج الذي يقابل اقل القيم من المعايير من بين النماذج المرشحة المبينة في الجدول الاتي:

جدول مقارنة رتب النماذج المرشحة

مقارنة نماذج (ARIMAX(na , nc , nb , nk) المرشحة			
MODELS	AIC	FPE	MDL
ARMAX(1,1,1,0)	44.72131121	2.643e+19	44.81195386
ARMAX(1,1,1,1)	44.7475171	2.714e+19	44.83815975
ARMAX(1,1,1,2)	44.73308799	2.675e+19	44.82373064
ARMAX(1,2,1,0)	44.59457014	2.329e+19	44.715427
ARMAX(1,2,1,1)	44.62502459	2.402e+19	44.74588145
ARMAX(1,2,1,2)	44.6893612	2.56e+19	44.81021806
ARMAX(2,1,2,0)	44.24057515	1.635e+19	44.39164623
ARMAX(2,1,2,1)	44.2598595	1.667e+19	44.41093058
ARMAX(2,1,2,2)	44.36098185	1.844e+19	44.51205293
ARMAX(2,2,2,0)	44.24014919	1.634e+19	44.42143448
ARMAX(2,2,2,1)	44.24867802	1.649e+19	44.42996332
ARMAX(2,2,2,2)	44.27107224	1.686e+19	44.45235754
ARMAX(3,1,3,0)	44.29947761	1.735e+19	44.51097712
ARMAX(3,1,3,1)	44.28487881	1.71e+19	44.49637832
ARMAX(3,1,3,2)	44.31590508	1.763e+19	44.52740459
ARMAX(3,2,3,0)	44.27994511	1.702e+19	44.52165884
ARMAX(3,2,3,1)	44.29000267	1.718e+19	44.53171639
ARMAX(3,2,3,2)	44.31121987	1.756e+19	44.5529336

نلاحظ من الجدول اعلاه ان رتبة النموذج الافضل الذي يمثل العلاقة بين المدخلات والمخرجات هو ARMAX(2,2,2,0) وذلك لان اقل المعايير هي في هذا النموذج وهي :
 $na=2, nc=2, nb=2, nk=0$ اي ان $AIC=44.24014919, FPE=1.634e+19$
 من بين عدة نماذج تم تشخيصها للوصول للنموذج الافضل .

مرحلة تقدير النموذج ARMAX(2,2,2,0)

بعد الانتهاء من تحديد رتبة النموذج وتحديد المعايير الاحصائية المذكورة مسبقا يتم تقدير المعلمات باستخدام برنامج بلغة ماتلاب MATLAB من قبل الباحث وفق الطرائق المذكورة سابقا :

المربعات الصغرى الاعتيادية OLS

وكانت النتائج وفق متعددات الحدود

$$A(L) = 1 + 0.3236 L^{-1} - 0.713 L^{-2}$$

$$B(L) = 0.06225 + 0.1755 L^{-1}$$

$$C(L) = 1 + 0.5463 L^{-1} - 0.4537 L^{-2}$$

وبهذا يمكننا كتابة النموذج بالشكل الاتي :

$$\hat{Y}(t) = -0.3236 Y_{(t-1)} + 0.713 Y_{(t-2)} + 0.06225 u(t) + 0.1755 u_{(t-1)} + e_{(t)} + 0.5463 e_{(t-1)} - 0.4537 e_{(t-2)}$$

طريقة المربعات الصغرى التكرارية باستخدام مرشح كالمن

كانت نتائج التقدير وفق متعددات الحدود :

$$A(L) = 1 + 0.764 L^{-1} - 0.3117 L^{-2}$$

$$B(L) = 0.3014 + 0.2359 L^{-1}$$

$$C(L) = 1 + 0.1349 L^{-1} + 0.2885 L^{-2}$$

وبهذا يمكننا كتابة النموذج بالشكل الآتي :

$$\hat{Y}(t) = -0.764 Y_{(t-1)} + 0.3117 Y_{(t-2)} + 0.3014 u(t) + 0.2359 u_{(t-1)} + e_{(t)} + 0.1349 e_{(t-1)} + 0.2885 e_{(t-2)}$$

طريقة المربعات الصغرى التكرارية بواسطة عامل التغاضي

كانت نتائج التقدير وفق متعددات الحدود :

$$A(L) = 1 + 0.1294 L^{-1} - 0.2187 L^{-2}$$

$$B(L) = 0.07029 + 0.335 L^{-1}$$

$$C(L) = 1 - 0.1992 L^{-1} + 0.1673 L^{-2}$$

وبهذا يمكننا كتابة النموذج بالشكل الآتي :

$$\hat{Y}(t) = -0.1294 Y_{(t-1)} + 0.2187 Y_{(t-2)} + 0.07029 u(t) + 0.335 u_{(t-1)} + e_{(t)} - 0.1992 e_{(t-1)} + 0.1673 e_{(t-2)}$$

جدول نتائج تقديرات المعلمات للنموذج

نتائج التقديرات لمعاملات نموذج ARMAX						
طريقة التقدير	ϕ_1	ϕ_2	θ_1	θ_2	ϕ_1	ϕ_2
OLS	0.3236	-0.713	0.5463	-0.4537	0.06225	0.1755
RLS-KF	0.764	-0.3117	0.1349	0.2885	0.3014	0.2359
RLS-EF	0.1294	-0.2187	-0.1992	0.1673	0.07029	0.335

مقارنة طرق التقدير

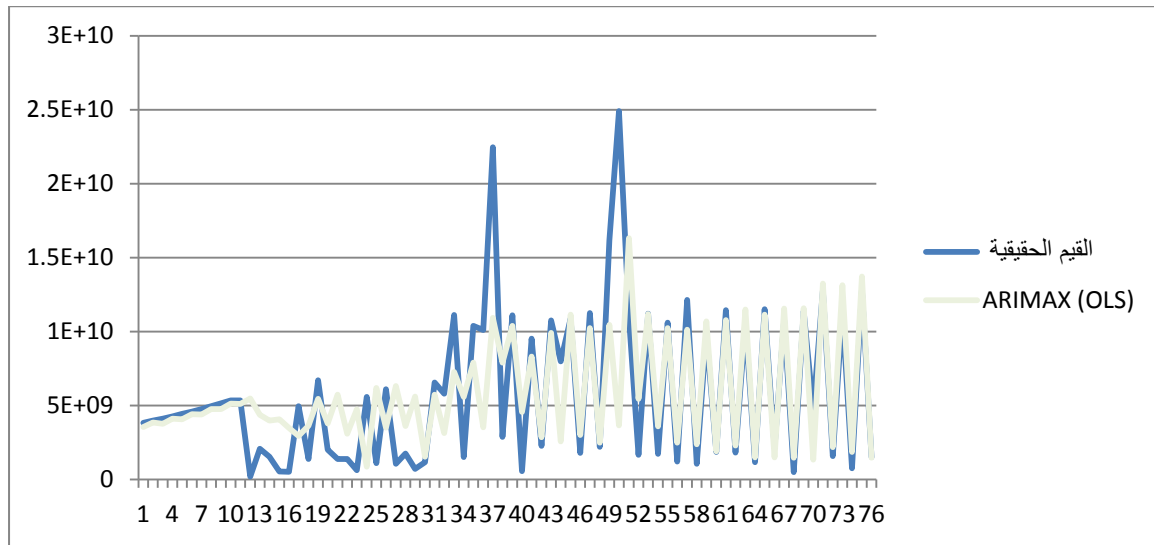
تمت مقارنة طرق التقدير لنموذج $ARMAX(2,2,2,0)$ باستعمال مقياس الجذر التربيعي لمتوسط مربعات الخطأ $RMSE$ ومقياس الخطأ المطلق المنوي $MAPE$ بعد إيجاد الأخطاء عبر برنامج ماتلاب من اعداد الباحث للطرق الثلاثة وكانت النتائج كما يأتي :

جدول يوضح مقارنة طرق التقدير		
طريقة التقدير	RMSE	MAPE
OLS	3766961640	1.376553632
RLS-KF	7480554792	3.314804278
RLS-EF	4830579675	2.522835243

نلاحظ ان افضل نموذج التي يتم تقدير معلمته وفق طريقة OLS حسب المقاييس :

$$MAPE = 1.376553632 \quad RMSE = 3766961640$$

بالمقارنة مع المقاييس الأخرى للطريقة التكرارية بعامل التغاضي ومرشح كالمن ونقوم بالتنبؤ داخل السلسلة بطريقة المربعات الصغرى OLS نرسم القيم الأصلية والقيم التنبؤية



اختبار Ljung -Box

جدول يوضح اختبار Ljung -Box للبواقي			
TEST	Q* (Lag)	Q* - Statistic	p-value
Ljung -Box	Q*(5)	1.437184856	0.230595436
Ljung -Box	Q*(10)	5.619058725	0.467185216

لغرض اختبار عشوائية النموذج المقترح والجدول الآتي يوضح نتائج الاختبار القيم التي تم تقديرها :

ويتضح من الجدول ان القيمة المحسوبة اصغر من القيمة الجدولية $X^2(h - m)$ وهذا يعني قبول فرضية العدم بان البواقي للنموذج عشوائية وهذا يدل على ملائمة النموذج للتنبؤ .
3-التنبؤ بخطوة الى الامام

القيمة الحقيقية	ARMAX(OLS)
10,784,790,130	14,732,238,590

e الخطأ المطلق	ARMAX(OLS)
	3,947,448,460

مناقشة النتائج

تم تحديد رتب نموذج ARMAX وفق المعايير (AIC , FPE,MDL) وتقدير معاملات النموذج وفق طريقة المربعات الصغرى الاعتيادية وطريقة المربعات الصغرى التكرارية بواسطة مرشح كالمن وعامل التغاضي وتمت مقارنة الطرق وفق مقاييس المقارنة RMSE,MAPE فوجد ان طريقة OLS هي الافضل وبعد اختبار ملائمتها للتنبؤ تمت مقارنة القيم التنبؤية مع القيم الاصلية فوجد انها قريبة منها وهذا يدل على الدقة في التنبؤ واعتماد هذا النموذج للتنبؤ بالتعويضات لصندوق الضمان الاجتماعي لدراسة الحالة المالية للصندوق .

الاستنتاجات Conclusions

بعد بناء نموذج ARMAX و إجراء عملية التنبؤ توصل الباحث من خلال النتائج التي تم الحصول عليها في الجانب التطبيقي الى عدة استنتاجات كالآتي :

تم استخدام ثلاث معايير في تحديد رتبة النموذج وهما (AIC , FPE,MDL) والتي اتفقت اثنان منها على نفس الرتبة تفوقت طريقة المربعات الصغرى الاعتيادية على طريقة المربعات الصغرى التكرارية بواسطة عامل التغاضي ومرشح كالمن وفق المقاييس المقارنة RMSE,MAPE و جاءت الطريقة التكرارية لمرشح كالمن في المرتبة الاخيرة و تم اختبار جودة نموذج ARIMAX (2,2,2,0) بواسطة اختبار Ljung-Box وكانت النتائج ملائمة واعطى نموذج ARMAX (2,2,2,0) نتائج قريبة من القيم الحقيقية وهذا يدل على دقة التنبؤ واعتماد النموذج في دراسة تعويضات صندوق الضمان الاجتماعي التي اعطت مؤشر مرتفع فيجب دراسة تزايد التعويضات بصورة دورية لتجنب الاضرار في المستقبل وتلافي العجز .

الاعمال المستقبلية

اعتماد نموذج ARMAX للتنبؤ بتعويضات صندوق الضمان الاجتماعي ودراسة حالة الصندوق وفق نموذج ARMAX بعدة مدخلات ومخرج واحد ومتابعة التطورات والطرانق الجديدة .

المصادر

- 1- Caihong Li & Wenheng Sun (2012), " The Study on Electricity Price Forecasting Method based on Time Series ARMAX Model and Chaotic Particle Swarm Optimization " , International Journal of Advancements in Computing Technology(IJACT)
- 2- Chi-man & Jacob (2008) , " statistical modelling and Forecasting Schemes for Air-Conditioning System " , Pao Yue-Kong library , the Hong kong Polytechnic Univeristy , Hong Hom , Kowloon , Heng Kong .
- 3-- David s. Stoffer (1986) " Estimation and Identification of Space-Time ARMAX Models in the Presence of Missing Data " , Journal of the American Statistical , September 1986 .
- 4-Gautam,Bijaya (2010), "Spectral Estimation of Electroencephalogram Signal Using ARMAX Model and Particle Swarm Optimization " , The Faculty of College of Graduate Studies Lamar University.
- 5- Gökhan Unakitan & Bahattin Akdemir (2015) " Prediction of Combine's Number By Using ARMAX Model in Turkey " , Balkan and Near Eastern Journal of Social Sciences .
- 6- J. Hyndman Rob & B. Koehler Anne (2006) , " Another look at measures of forecast accuracy " , elsevier International Journal of Forecasting 22 , PP. [679 – 688]
- 7- J. Willmott Cort & Matsuura Kenji (2005) , " Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance " , CLIMATE RESEARCH Clim Res
- 8- Kanjilal,P.P,(1995)"Adaptive Prediction and Predictive Control"
,Peter Peregrines Ltd . London
- 9- Kumar P.R. & Varaiya Pravin (1986), " Stochastic Systems : Estimation, Identification, and Adaptive Control " , Prentice-Hall, Inc , Englewood Gifts, New Jersey .
- 10- Ljung Greta , Box E.P(1978)," On a measure of lack of fit in time series models " .Biometrika,Vol (65),No(2),pp(297-303)

- 11- Ljung Lennart (1999) " System Identification : Theory for the User 2 edition " , Linkoping University sweden , Prentice Hall International
- 12- Morata,A & Loira,L (2017)" Yeast-Industrial Applications " .Rijeka, Croatia
- 13- Nelles Oliver (2001) , " Nonlinear System Identification " , springer – verlag 1st edition .
- 14-- Noah Meyerson &John Sabelhaus (1999) " Uncertainty in Social Security Trust Fund Projections " , Congressional Budget Office, Washington, D.C. 20515 , National Tax Journal Vol. LIII, No. 3, Part
- 15-- Saeed Matroushi(2011) " Hybrid Computational Intelligence Systems based on Statistical and Neural Networks Methods for Time Series Forecasting: The Case of Gold Price" , Lincoln University.
- 16- Soderstrom Torsten & Stoica Peter (2001) , " System Identification" , Prentice Hall International .
- 17- Wellstead,P.E& Zarrop,M.B (1991) "Self – Tuning Systems : Control and Signal Processing " . Chichester , UK:John Wiley & Sons

Predicting Social Security Fund compensation in Iraq using ARMAX Model

Firas Ahmmed Mohammed

Ass.Prof. Department of statistics,
College of Management and Economics,
Baghdad University, Baghdad, Iraq

drfirasmohana@gmail.com ﻻﻏﻤﺎﻥ

Zaid Khalil Ismail

Master's student student.
Department of Statistics, College of
Management and Economics,
Baghdad University, Baghdad, Iraq
zaid.khalil.ismail@gmail.com

Received: 21/6/2020

Accepted :26/7/2020

Published : January / 2021



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

Abstract

Time series have gained great importance and have been applied in a manner in the economic, financial, health and social fields and used in the analysis through studying the changes and forecasting the future of the phenomenon. One of the most important models of the black box is the "ARMAX" model, which is a mixed model consisting of self-regression with moving averages with external inputs. It consists of several stages, namely determining the rank of the model and the process of estimating the parameters of the model and then the prediction process to know the amount of compensation granted to workers in the future in order to fulfill the future obligations of the Fund. , And using the regular least squares method and the frequency squares method by "kalmen filter" and "forgeting factor" to obtain the best predictable way for the future using data from the Social Security Department for workers and from the period 1/2013 to 6/2019 where The method of least squares "OLS" was compared with the method of iterative least squares "RLS" by "kalman Filter" and found that "OLS" method is the best according to the comparison measures "RMSE, MAPE" and it gave accurate results close to the real values.

Keywords of research / Akaike Standard (AIC) –Prediction Error Standard(FPE) - Bayesian Akaike Standard (BIC) - Ordinary least square Method (OLS) - recursive least square Method(_RLS) - Kalman Filter – Forgetting factor - Prediction