استخدام طريقة تقاطع الشرعية اللامعلمية لتقدير رتب نماذج السلاسل الزمنية

أ.م.أميرة جابر محيسن
كلية علوم الحاسبات والرياضيات
جامعة القادسية

الخلاصة

من المسائل المهمة التي تواجه الباحث عند ملائمة نموذج رياضي لسلسلة زمنية معينة هي مسألة تقدير رتبة النموذج ، وتهدف هذه الدراسة الى استخدام طريقة تقاطع الشرعية اللامعلمية لتقدير رتب نماذج الانحدار الذاتي للسلاسل الزمنية وتطبيق هذه الطريقة باستخدام أسلوب المحاكاة ، كما تم تطبيقها باستخدام بيانات واقعية.

1- المقدمة

لايخفى على المهتم بموضوع السلاسل الزمنية الأهمية الكبيرة لنماذجها في الوصف والتحليل والتنبؤ المستقبلي Forecasting بقيم الظاهرة قيد البحث ، وقد يذهب الأمر إلى ابعد من هذا فيصل إلى اختبار الفرضيات المختلفة المتعلقة بالظاهرة ، فيؤدي الأمر بالنتيجة إلى القرار الذي يتوقف مدى صوابه على مدى صواب اختيار النموذج وطريقة التقدير الملائمة والمعايير المستخدمة في الاختبار وما إلى ذلك.

وتعد عملية بناء النماذج الرياضية من الأمور المهمة في تحليل السلاسل الزمنية ، ومن المسائل المهمة التي تواجه الباحث عند ملائمة نموذج رياضي لسلسلة زمنية معينة هي مسألة تقدير رتبة النموذج ، حيث قد نجد عدم إمكانية استخدام الأساليب الإحصائية الكلاسيكية لتقدير رتب نماذج السلاسل الزمنية وذلك لطبيعة المشاهدات ، وهنالك عدة محاولات لتقدير رتب نماذج الانحدار الذاتي ، ففي عام (1953) قام الباحث Moran باستخدام دالة الترجيح الشرطية لاختيار الرتبة المناسبة لنموذج انحدار ذاتي لبيانات بيئية ،[6] إلا ان هذه الطريقة احتوت على بعض المعوقات التي لم تشجع على انتشارها ، وفي عام (1970) اقترح الباحثان Box and Jenkins طريقة تقريبية لتقدير رتبة النموذج وذلك بالاعتماد على كل من دالة الارتباط الذاتي ودالة الارتباط الذاتي الجزئي ،[3] ومن مزايا هذه الطريقة انها تعتمد على الخبرة الشخصية وتعطي نتائج دقيقة في حالة النماذج الخطية فقط ، في حين ان الدراسات المعاصرة أظهرت بان اغلب السلاسل الزمنية الواقعية هي ذات طبيعة غير خطية ، وعليه تظهر الحاجة الملحة الى طريقة لاتتقيد بافتراض ان النموذج الرياضي يكون خطياً بل من الممكن ان يكون خطياً أو غير خطي ، وأفضل طريقة هي الطريقة اللامعلمية حيث اقترح الباحثان Cheng and Tong في عام (1992) استخدام طريقة تقاطع الشرعية Cross-Validatory Approach لتقدير رتب نموذج الانحدار الذاتي،[4] وفي العراق درس الباحث باسل يونس ذنون هذه الطريقة في عام (2002) حيث اثبت كفاءتها باستخدام أسلوب المحاكاة وباستخدام بيانات واقعية بيئية ، [1] .

وفي البحث الحالي تم استخدام هذه الطريقة اللامعلمية لتقدير رتبة نماذج الانحدار الذاتي وتم تطبيقها باستخدام أسلوب المحاكاة وطبقت ايضاً على بيانات واقعية.

2- هدف الدراسة

ان الهدف الرئيسي لهذا البحث هو استخدام طريقة تقاطع الشرعية Cross-Validatory Approach اللامعلمية لتقدير رتب نماذج الانحدار الذاتي للسلاسل الزمنية وتطبيق هذه الطريقة باستخدام أسلوب المحاكاة ، كما تم تطبيقها باستخدام بيانات واقعية عبارة عن سلسلة زمنية تمثل معدل وفيات الأطفال الرضع في العراق للسنوات 1950-1998.

3- الجانب النظري

3-1 الطريقة اللامعلمية لتقدير الرتبة

لنفرض ان لدينا المشاهدات $\{X_t , t = 1,2,...,n\}$ من السلسلة الزمنية المستقرة بقوة Strictly Stationary ذات التباين المحدود وبفرض ان هذه السلسلة ممثلة بنموذج الانحدار الذاتي العام المحدود الآتي:-

$$X_t = F(X_{t-1}, e_t) + ... + X_{t-d} \qquad (1)$$

حيث ان:-

$F(.)$:- هي دالة غير معروفة ، ممكن ان تكون خطية وممكن ان تكون غير خطية.

e_t :- هو خطأ عشوائي يمثل تشويش النموذج.

ان الهدف الرئيس الذي نريد تحقيقه هو تقدير رتبة النموذج (d) .

ان الخطوة الاولى في التقدير تتطلب تقدير $F(.)$ وبالاستفادة من أسلوب التقدير اللبي يمكننا تقدير $F(.)$ لامعلمياً ، [4],[2],[1]

لنفرض أننا نتعامل مع فضاء ذو بعد (1) وان لدينا الدالة اللبية $K(\cdot)$ ، بحيث أن $K : R' \to R'$ وان $K(\cdot)$ تحقق الشروط الآتية ، [2],[1] :-

$$\left.\begin{array}{l}(i) K(u) \geq 0 \forall u \\ (ii) \int_{R'} K(u)du = 1 \\ (iii) \int |u| K(u)du < \infty\end{array}\right\} \qquad (2)$$

ولو كنا في فضاء ذو بعد (1) وحيث $u = (u_1, u_2, ..., u_l) \in R'$ فان الدالة اللبية عندئذ تعرف كالآتي :-

$$K_l(u) = \prod_{i=1}^{l} K(u_i) \qquad (3)$$

أما إذا رمزنا إلى $(X_{t-1},...,X_{t-d})$ بالرمز (Y_t) وأن (f) ترمز لدالة كثافة احتمال (Y_t) فان المقدار اللبي للدالة (f) بالاعتماد على المشاهدات $Y_r, Y_{r+1},...,Y_l$ حيث أن (r) عدد موجب يتم اختياره لجعل التأثيرات الطريفة أقل ما يمكن يرمز له بالرمز $\hat{f}_t$ ، كما أن المقدار اللبي للدالة (f) بالاعتماد على المشاهدات السابقة نفسها ما عدا المشاهدة (Y_i) يرمز له بالرمز $\hat{f}_{t,\backslash i}$ ، وقد اقترح الباحثان Cheng and Tong في عام (1992) المقدرين اللبين الآتيين عند $Y \in R^d$ ، [4],[1] :-

$$\hat{f}_{t,\backslash i}(y) = \left[(t-r)B(t)^d\right]^{-1} \sum_{s=r, s\neq 1}^{t} K_d((y - Y_s)/B(t)) \qquad (4)$$

$$\hat{f}_t(y) = \left[(t-r+1)B(t)^d\right]^{-1} \sum_{s=r}^{t} K_d((y - Y_s)/B(t)) \qquad (5)$$

حيث أن :-

t, r :- هما عددين صحيحين موجبين وان $t \leq i \leq r$.

وبالعودة إلى المعادلة (1) نلاحظ أن $F(X_{t-l},...,X_{t-d})$ تكافئ التوقع الشرطي $E(X_t \mid X_{t-l},...,X_{t-d})$ وبناءاً على هذا يكون مقدار التوقع الشرطي هو نفسه مقدر الدالة F ، فإذاً $\hat{F}_t$ يمثل المقدار اللبي للدالة F بالاعتماد على المشاهدات $Y_r, Y_{r+1},...,Y_l$ في حين ان $\hat{F}_{t,\backslash i}$ هو نفس $\hat{F}_t$ ما عدا المشاهدة Y_i لم تستخدم التقدير ، لذا عندما $Y \in R^d$ قيمة (i) تحقق $t \leq i \leq r$ ، فأن [4],[1]:-

$$\hat{F}_{t,i}(y)=\left[(t-r)B(t)^{d}\right]^{-1}\sum_{s=r,s\neq 1}^{t}X_sK_d((y-Y_s)/B(t))\hat{f}_{t,\backslash i}^{-1}(y)\ldots\ldots(6)$$

$$\hat{F}_{t}(y)=\left[(t-r+1)B(t)^{d}\right]^{-1}\sum_{s=r}^{t}X_sK_d((y-Y_s)/B(t))\hat{f}_{t,\backslash i}^{-1}(y)\ldots\ldots(7)$$

ان تقاطع الشرعية كما عرفه الباحثان Cheng and Tong عام (1992) يكون بالشكل الآتي:-

$$CV(d)=(n-r+1)^{-1}\sum_{t=r}^{n}(X_t-\hat{F}_{n,\backslash d})^2W(y_t)\ldots\ldots(8)$$

حيث أن :-

$W(.)$:- هي دالة وزن

أن المقدر اللامعلمي لرتبة النموذج(1) يمكن الحصول علية وذلك بالسماح لقيم (d) لكي تأخذ عدداً من القيم مثلاً $(1,2,\ldots,l)$ وحساب CV(d) في كل حالة ثم اخذ المقدر لكي يكون تلك القيمة من (d) التي تقابل اقل قيمة من قيم CV(d) فإذا كان $\hat{d}$ يمثل مقدر تقاطع الشرعية لرتبة النموذج (d) فأن :-

$$\hat{d}=\underset{d=1,2,\ldots,l}{Min}\,CV(d)\ldots\ldots(9)$$

3-2 الخوارزمية

لغرض ترجمة الطريقة اللامعلمية السابقة الى السابقة الى واقع علمي ممكن التطبيق فقد تمت مراعاة بعض الاعتبارات العملية :-

(1) فبالنسبة لسعة القيد B(t) فليس هنالك طريقة رصينة لتحديد الصيغة المناسبة لهذه الدالة وهنالك وجهات نظر متباينة بهذا الخصوص وفي هذه الدراسة اختيرت B(t) = K(SD) وكما مستخدم في الدراسات السابقة حيث أن ، [1] :-

$0<K<1$ و SD:- يمثل الانحراف المعياري للمشاهدات.

(2) كما تم مراعاة تحويل البيانات الى معيارية وذلك بقسمتها على SD .

(3) وبالنسبة لدالة الوزن $W(.)$ في المعادلة(8) فقد اختيرت لكي تساوي واحد .

(4) تم اختيار التوزيع الطبيعي القياسي لكي يكون هو الدالة اللبية أي أن :-

$$K(u)=(2\pi)^{-0.5}\exp(-u^2/2),\quad -\infty<u<\infty$$

وخطوات الخوارزمية المستخدمة هي :-

الخطوة الاولى :- يتم إدخال المشاهدات $X_1,\ldots,X_n$ و n وكذلك إدخال قيمة r وقيمةL.

الخطوة الثانية :- حساب الانحراف المعياري للمشاهدات ثم تحديد قيمة B(t).

الخطوة الثالثة :- تحويل المشاهدات الخام الى مشاهدات معيارية $Y_1,\ldots,Y_n$.

الخطوة الرابعة :- للقيم $d=1,2,\ldots,l$ ننفذ كل مما يأتي :-

(1)للقيم $tt=r,r+1,..n$.

(2) للقيم $S=r,r+1,\ldots,tt$.

(3) من العلاقة (3) نحسب :-

$$K_d(u)=\prod_{i=0}^{d}(2\pi)^{-0.5}\exp\left[-0.5\{y((tt-i)-y(S-i))/B(n)\}^2\right]$$

(4)نحسب المقدرات المبينة في العلاقات (4),(5),(6),(7).

(5)نحسب CV(d) من العلاقة (8) .

الخطوة الخامسة :- نؤشر اقل قيمة من قيم CV(d) وعندها تكون قيمة (d) المقابلة لهذه القيمة هي عبارة عن مقدر (d).

4- الجانب التطبيقي

لغرض اختبار كفاءة الطريقة اللامعلمية والتي تم توضيحها في الجانب النظري تم استخدام أسلوب المحاكاة ، اضافة الى استخدام بيانات واقعية عبارة عن سلسلة زمنية تمثل سلسلة وفيات الاطفال الرضع في العراق للسنوات (1950-1998) ، وتم استخدام برنامج Matlab للحصول على النتائج وكالاتي :-

4-1 دراسة المحاكاة

لقد تم توليد خمسة عشر مجموعة من المشاهدات المولدة بأستخدام نموذجين رياضيين هما:-

$$X_t = 0.3X_{t-1} - 0.9X_{t-7} + e_t$$

$$X_t = 0.4X_{t-1} + (-0.6 + 0.5\exp(-0.03X^2_{t-7}))X_{t-7} + e_t$$

حيث أن حد التشويش e_t قد احل محله مولد من التوزيع الطبيعي القياسي N(0,1) وان سعة القيد:-

$$B(t) = \frac{SD}{2}$$

ومن خلال ملاحظة النموذجين اعلاه نجد ان رتبة كل من النموذجين السابقين هي d=7 ، وقد تم توليد سبعة تكرارات حجم كل تكرار n=250 من كل من النموذجين السابقين ، وتم تطبيق الطريقة الموضحة في الجانب النظري لتقدير الرتب وتم تقدير قيم CV(d) لكلا النموذجين والنتائج التي تم الحصول عليها موضحة في الجدولين (1) و(2) ، حيث نلاحظ من الجدول رقم(1) ان اقل قيمة من قيمCV(d) كانت عند قيمة (d=7) ولجميع التكرارات وهذا يؤكد ان رتبة النموذج الاول هي (7) ، ومن الجدول رقم (2) يمكن ان نلاحظ ان اقل قيمة من قيم CV(d) هي عند قيمة (d=7) ولكافة التكرارات المولدة وهذا يعني ان رتبة النموذج رقم (2) هي(7) ، ومن خلال هذه النتائج نلاحظ كفاءة طريقة تقاطع الشرعية اللامعلمية في تقدير رتبة النموذج.

4-2 البيانات الواقعية

تم استخدام بيانات واقعية تمثل سلسلة حوادث معدل وفيات الأطفال الرضع في العراق للسنوات 1950-1998 بواقع (49) مشاهدة ولغرض التعرف على خصائص هذه السلسلة فقد تم رسمها مقابل الزمن وكما في الشكل رقم(1) حيث يمكن أن نلاحظ من هذا الشكل أن السلسلة مستقرة الى حد ما في المتوسط والتأثيرات الموسمية فيها ليست واضحة بشكل كبير ، ولغرض تشخيص وتحديد درجة النموذج استخدمنا في البداية الأسلوب الكلاسيكي وذلك باستخراج ورسم دالة الارتباط الذاتي كما في الشكل رقم(2) واستخراج ورسم دالة الارتباط الذاتي الجزئي كما في الشكل رقم (3) ، ويتضح من خلال دالتي الارتباط الذاتي للسلسلة والارتباط الذاتي الجزئي لها بأنه ليس هنالك نموذج أوساط متحركة تخضع له البيانات لكون دالة الارتباط الذاتي تتناقص تدريجياً بينما تخضع البيانات لنموذج الانحدار الذاتي من الدرجة الاولى لكون دالة الارتباط الذاتي الجزئي تتلاشى بعد أول إزاحة.

ولغرض الحصول على تحديد دقيق لدرجة النموذج تم استخدام الطريقة اللامعلمية التي تم عرضها في الجانب النظري ، حيث تم الحصول على قيم CV(d) ، وكما مبين في الجدول رقم(3) فمن خلال هذا الجدول يمكن أن نلاحظ أن اقل قيمة لـ CV(d) كانت عند d=1 ومن هنا يتضح أن رتبة النموذج هي d=1.

جدول رقم(1) يوضح نتائج تقدير رتبة النموذج الأول

d \ التكرار	الأول	الثاني	الثالث	الرابع	الخامس	السادس	السابع
1	0.99873	0.98125	1.13412	1.020061	0.93567	1.00892	0.98996
2	1.12011	0.96667	0.99888	0.97002	1.010203	1.18461	1.00408

3	1.71231	0.99099	1.31311	1.61011	1.001161	0.99989	1.00066
4	1.31121	1.102101	1.03261	1.00899	1.05342	1.12887	1.033233
5	1.11342	1.00123	1.08081	1.07887	1.08071	1.109887	1.008071
6	1.02322	1.06672	1.00651	0.97812	0.98976	0.98689	0.89999
7	*0.80001	*0.81011	*0.79899	*0.78986	*0.80023	*0.80115	*0.79865
8	0.98901	0.95907	0.89807	0.87321	0.96067	0.98544	0.85671
9	0.95011	0.89111	0.94344	0.85101	0.94221	0.96601	0.86012
10	0.93947	0.91811	0.97165	0.91231	0.93451	0.92931	0.90123
11	0.90023	0.87121	0.83156	0.90002	0.85124	0.82138	0.83102
12	1.00121	0.98911	0.90125	0.89981	0.91007	0.86081	0.91818
13	0.98002	1.12008	0.91607	0.98008	0.97201	0.89899	0.88077
14	1.01231	1.00181	0.93081	0.89971	0.87162	0.85075	0.81211
15	0.94031	0.82211	0.86014	0.80245	1.00019	0.90002	0.88001

جدول رقم(2) يوضح نتائج تقدير رتبة النموذج الثاني

التكرار d	الأول	الثاني	الثالث	الرابع	الخامس	السادس	السابع
1	1.08765	1.00241	0.99341	0.97651	0.99861	1.00386	0.99989
2	0.98765	0.99462	0.97801	0.96271	1.02361	1.05678	0.97801
3	0.96061	0.93033	0.95656	0.92232	0.95565	1.01161	1.00211
4	0.94999	0.96787	0.92099	0.96121	0.94981	0.92001	0.96002
5	0.90120	0.89901	0.91542	0.91114	0.89003	0.91212	0.89665
6	0.89231	0.91451	0.90234	0.89871	0.91891	0.89121	0.90002
7	*0.78001	*0.77781	*0.77999	*0.76661	*0.750608	*0.75661	*0.74198
8	0.96051	0.88145	0.93341	0.89806	0.88451	0.91678	0.89612
9	0.88967	0.870112	0.88565	0.90006	0.86051	0.87786	0.85011
10	0.90101	0.86651	0.90011	0.85061	0.85611	0.87111	0.84013
11	0.86501	0.88241	0.86511	0.90003	0.84044	0.82101	0.85011
12	0.82263	0.90004	0.84664	0.83232	0.87577	0.85002	0.82113
13	0.90234	0.84564	0.82381	0.82141	0.79781	0.80971	0.79899
14	0.79998	0.801211	0.78199	0.79165	0.80113	0.78103	0.79008
15	0.80021	0.78078	0.80199	0.78889	0.810121	0.82121	0.78102

جدول(3) يوضح نتائج تقدير رتبة نموذج سلسلة معدل وفيات الرضع

d	1	2	3	4	5	6	7	8	9	10
CV(d)	*18.013	18.758	18.909	18.989	18.568	18.795	18.689	18.453	18.667	18.962

5- الاستنتاجات

من خلال ما تقدم يمكن أن نستنتج الآتي :-

(1)أن الطريقة اللامعلمية تعطي تقديراً دقيقاً لرتبة نماذج السلاسل الزمنية سواء كان النموذج خطي أم غير خطي ، حيث نلاحظ من نتائج المحاكاة انه كان تقدير الرتبة للنموذج الأول 7 والتي تقابل اقل قيمة لـ CV(d) ، كما أن رتبة النموذج الثاني كانت 7 ايضاً وحسب قيمةCV(d) .

(2)أن الطريقة اللامعلمية أعطت نتائج دقيقة لتقدير رتبة نماذج السلاسل الزمنية باستخدام البيانات الواقعية ، حيث كان النموذج الممثل لسلسلة معدل الوفيات هو من الرتبة الاولى وهذا ما أكدته الطريقة اللامعلمية حيث كانت اقل قيمة من قيم CD(d) هي 18.013والمقابلة لقيمة d=1.

(3)من خلال دراسة البيانات الواقعية يتضح أن سلسلة الوفيات الشهرية مستقرة الى حد ما في المتوسط وأفضل نموذج يمثل هذه السلسلة هو نموذج الانحدار الذاتي من الرتبة الاولى و عليه بالامكان استخدام هذا النموذج للتنبؤ بحالات الوفيات ووضع الخطط المستقبلية لتقليلها.

6- التوصيات

(1)استخدام طريقة تقاطع الشريعة اللامعلمية لتقدير رتب نماذج السلاسل الزمنية لانها لسهولة تطبيقها ودقة نتائجها.

(2)أن الطريقة اللامعلمية غير محددة بخطية النموذج الرياضي للسلسلة الزمنية ، وبالتالي يمكن استخدامها للوصول الى نتائج دقيقة في حالة النماذج غير الخطية.

(3)أجراء دراسات أخرى لمقارنة هذه الطريقة مع طرق أخرى لتقدير رتب نماذج السلاسل الزمنية للتوصل الى أفضل طريقة لتقدير الرتب.

7- المصـــــــــــــــــادر

(1)ذنون ، د. باسل يونس ، 2000، "تقدير رتب نماذج الانحدار الذاتي للسلاسل الزمنية" بحث منشور في مجلة وقائع المؤتمر العلمي الثالث عشر للجمعية العراقية للعلوم الاحصائية للفترة من 2-2002/11/3

(2)ذنون ، د. باسل يونس، 1996"التقدير اللبي أسلوب بياني في التقدير الإحصائي " علوم الرافدين العدد 1 الجزء 7.

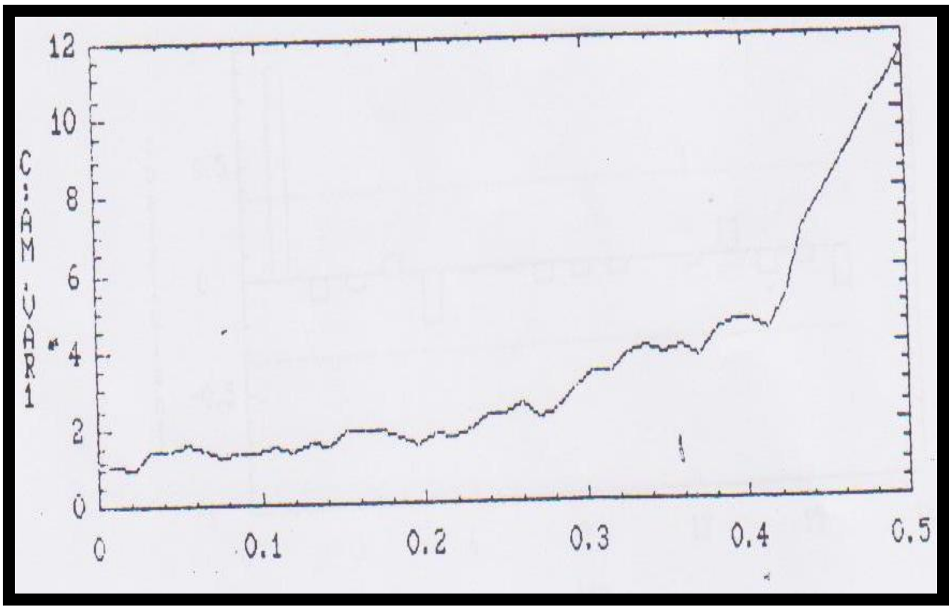
(3) Box , G.E.P.and Jenkins, E.M. (1970) "Time series Analysis Forecasting and Control " Holden- Day, San Francisco.

(4) Cheny , B. and Tong , H.(1992) "On consistent nonparametric order determination and chaos "

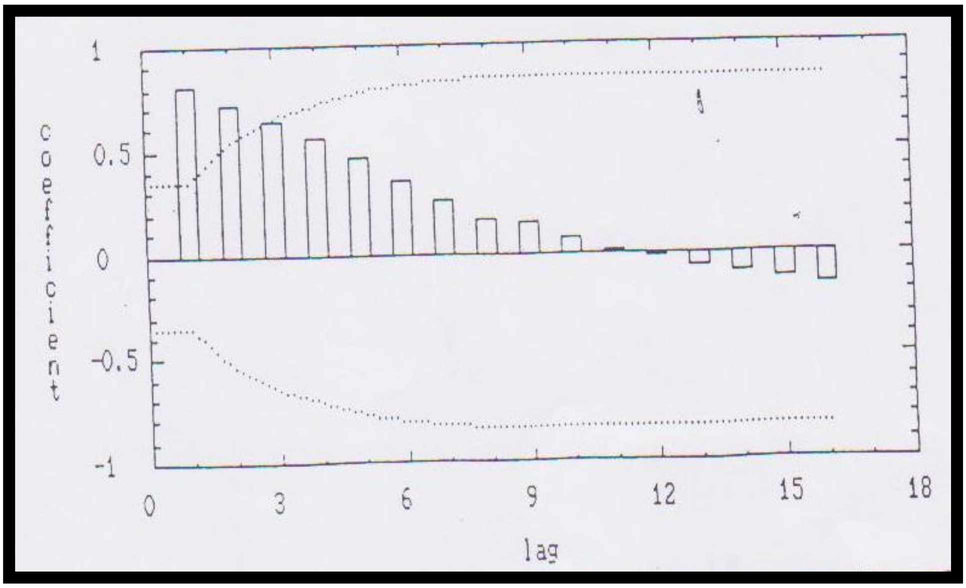
(5)Tong , H.(1983)"Threshold Models in non-linear Time series Analysis " Lecture Notes in stat . No.21, New York.

(6) Moran ,P.A.P.(1953)" The Statistical analysis of the Canadian lynx cycle "Int Aust .J.Zool, 1,,163-173.

شكل رقم (1) السلسلة الزمنية لمعدل وفيات الأطفال



شكل رقم (2) دالة الارتباط الذاتي لسلسلة معدل وفيات الأطفال



شكل رقم (3) دالة الارتباط الذاتي الجزئي لسلسلة معدل وفيات الأطفال