أساليب الكشف عن الشوارد في السلاسل الزمنية ذات الدرجات الدنيا
((دراسة تجريبية))[+]

THE OUTLIERS DETECTION WAYS IN TIME SERIES FROM LOW ORDER ((EXPERIMENTAL STUDY))

سالم بدر محمد[*]

المستخلص :

أساليب الكشف عن الشوارد في السلاسل الزمنية ذات الدرجات الدنيا ((دراسة تجريبية)).إن تحليل السلاسل الزمنية ودقة المعلومات فيها يعتمد بشكل عالي على دقة البيانات وعدم احتواءها على شـــوارد ، والتي تؤثر على دقة النتائج ، وقد هدف البحث إلى دراسة أنواع الشوارد وكيفية التخلص مـــن تأثيرهـــا ، نظريا" وتجريبيا" لنماذج AR(1) و MA(1) و AR(2) و ARMA(1,1) .

وقد توصلنا إلى إن أساليب الكشف عن الشوارد فعالا" في الدرجات الدنيا ، وان الشوارد من نـــوع (Ao) أكثر تأثيرا" من (Io) كما انها تحتاج إلى اعداد اكبر من المكررات لتصحيح مواقع الشوارد .

Abstract :

The Outliers detection ways in time series from Low order(experimental study). That time series analysis and the accuracy of the data depened on the accuracy of data and the absence of Outliers of results .

The research amied to outliers detection ways and how to get the influence of theoretical and experimental models AR(1) , MA(1) , AR(2) and ARMA(1,1) .

We have the importe results in disclosure of outliers in the low order model , then the type (Ao) are high affect to type (Io) , and need to more frequency number to correct the place of outliers .

المقدمة :

ان توفر بيانات خالية من أي ملوثات قد يكون غير متوقع عبر الزمن ، وتؤثر على عملية التحليـــل الإحصائي الدقيق مالم تدرس وتحدد . الملوثات هي تلك القيم التي تكون خارجة عن النسق العام للبيانات مما يجعلها لا تتوافق مع بعض الفرضيات الأساسية الواجب توافرها للنموذج الإحصائي. وتحليل السلاسل الزمنية حقل يهتم بدراسة الظواهر التي تعتمد على البيانات المتعاقبة زمنيا" ومن اجل ذلك نقوم بتقدير معالم النمـــاذج ثم التنبؤ.

أكثر الطرائق التقليدية المستخدمة في تقدير معالم السلاسل الزمنية هي طريقة المربعات الـــصغرى (Least squares) وطريقة الامكان الأعظم (Maximum likelihood) وكذلك طريقة العـــزوم (Moment method)[١] ، [٢] .

[+] تاريخ استلام البحث ٢٠٠٧/٦/٦ تاريخ قبول النشر ٢٠٠٧/١١/٦

[*] استاذ مساعد/المعهد التقني/الانبار .

ان مقدرات هذه الطرائق قد تكون كفؤ ه ومناسبة في حالة كون توافر جميع شروط نمـــاذج الـــسلاسل الزمنية، إذ إن التوزيع يجب إن يكون طبيعيا (Normal Distribution) مع توافر صفتي الاســتقرارية (Stationary) والانعكاسية (Invertability) ولكن عند اختلاف الشروط نتيجة عامل معــين يــؤدي إلـــى اختلال في صفات المقدرات الجيدة، حيث ان الطرائق التقليدية تكون حساسة لأي تغير في توزيع الأخطاء ، وهو الناتج من الاختلاف في نسق البيانات حتى وان كان صغيرا" ، وهذا التغير يكون بسبب وجود الشوارد (Outliers) في البيانات والتي تظهر في الأخطاء مما يؤثر على التوزيع المفترض لهذه الأخطاء .

إن وجود القيم الشاردة في البيانات هو شيء وارد أي كانت تلك البيانات ومهما كانت دقـــة الحـــصول عليها ، وتنشأ عادة الشوارد من توزيعـــات ثقيلـــة الاطـــراف (Heavy Tailed Distribution) أو مـــن توزيعات خليطة (Mixture Distribution) أو من أخطاء المشاهدة ، وهي عموما" ما يكون لهـــا تـــأثير كبير على مقدرات الطرائق الاعتيادية للتقدير ، إذ تزداد حساسية تلك المقدرات بزيادة عدد الشوارد أو زيادة شرودها (ابتعادها عن نسق البيانات) عن مجموعة البيانات الممثلة للظاهرة (إن وجود قيمة شاردة واحـــدة فقط ضمن البيانات يؤدي إلى فقدان المقدرات خواصها الجيدة)[٣] .

ان عملية الكشف عن الشوارد وتحديد وطبيعتها هي المرحلة والاكثر اهمية في عملية معالجة تأثيرتها السلبية على مقدرات معالم النماذج المختلفة . ويعد مقياس نسبة الامكان من المقاييس المهمة والفعالة في هذا المجال اذا يعتمد على كون معالم النموذج والتباين معلومان او مجهولان وكذلك على الاسلوب المكرر لكشف الشوارد والتخلص منها .

١-٢ هدف البحث:

إن وجود القيم الشاردة في البيانات يجعل المقدرات حساسة وتقلل من صفاتها الجيدة ، وتعتمـــد دقـــة وكفاءة المقدرات على دقة اختيار طريقة التقدير الملائمة ، إذ إن هناك أسلوبان لدراسة النماذج في حالة وجود الشوارد ، الأول هو الكشف عن وجود الشوارد ومواقعها ومن ثم إزالة تأثيرها ثم القيام بعملية التقدير عـــن طريق الطرائق الاعتيادية والثاني هو الاستخدام المباشر لطرائق التقدير الحصينة والتي تعتمد أصـــلا" علـــى بعض الدوال الوزنية التي تعمل على جعل المقدرات التي نحصل عليه ذات خصائص جيدة .

قد جاءت فكرة هذا البحث لنحاول تحقيق الهدف التالي دراسة تأثير الشوارد في الـــسلاسل الزمنيـــة وكيفية الكشف عنها وتعريف أنواعها نظريا" وباستخدام المحاكاة .

٢- الجانب النظري

٢-١ السلاسل الزمنية :

إن معظم البيانات في الأعمال والاقتصاد والهندسة والعلوم التطبيقية وغيرها تأخذ شـــكل الـــسلاسل الزمنية ، وهي تلك النماذج التي تكون المشاهدات فيها معتمدة على بعضها بـــشكل طبيعـــي ، وان التحليـــل الإحصائي الذي يستخدم في هذه الحالات يسمى عادة بتحليل السلاسل الزمنية[٤] . كما إن التحليلات من هذا النوع يجب إن تحقق شروطا" معينة تجعل التحليل أكثر دقة ، فإذا اعتبرنا إن (X_t) هو متغيـــر مـــشاهدات الظاهرة معينة ، فان (X_t) قد تحمل بعض الشروط المهمة وقد لا تحملها ، ولذلك نفتـــرض (Z_t) هـــو التحويل المناسب لـــ (X_t) بواسطة (D) إذ أن :

$$Z_t = DX_t$$

وان Z_t ستحقق الشروط الاتية :

- إن العملية العشوائية (Z_t) تكون مستقرة .
- إن التوزيع المشترك لـ (Z_t , ... , Z_2 , Z_1) هو التوزيع الطبيعي المتعدد (Multivariate Normal Distribution) .

ومن أهم خصائص هذا التوزيع إن العزم الأول والعزم الثاني سوف يميزان التوزيع بشكل كامل[٥] .

ويتم تحليل السلاسل الزمنية اما بالاعتماد على الدالتين ، الدالة المولدة للتباينات المشتركة [AGF] (Auto covariance Generating function) والدالة المولدة للارتباطات المشتركة [ACF] (Auto covariance Correlation function) وهذا ما يسمى (Time Domain Analysis) . أو بالاعتماد على دالة الطيف (Spectral function) ويسمى (Frequency Domain Analysis) .

٢-٢ نموذج الانحدار الذاتي Autoregressive Model

هو نموذج عشوائي يتحقق عندما يوجد عدد محدد من الأوزان غير المساوية إلى الصفر ولتكن $\phi_k = 0 \quad k > \rho \qquad \phi_i \neq 0 \quad i = 1,2,.....\rho$ ، عندها تسمى العملية الناتجة بعملية الانحدار الذاتي الرتبة (ρ) ويرمز لها بالرمز $AR(\rho)$ ، ويعبر عنها بالصيغة التالية[١] :

$$Z_t = \phi_1 Z_{t-1} + \cdots\cdots + \phi_\rho Z_{t-\rho} + a_t \qquad .....2-1$$

أو بالصيغة التالية

$$\phi_\rho(B) Z_t = a_t$$

إذ أن

$$\phi_\rho(B) = (1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 \cdots\cdots \phi_\rho B^\rho)$$

وان (a_t) هي سلسلة الضوضاء البيضاء (White noise) بوسط صفر وتباين (σ_a^2) .

نماذج الانحدار الذاتي دائما قابلة للعكس أي تحمل صيغة الانعكاسية (Invertable) ولكنها ليست دائما مستقرة أي تحمل صيغة الاستقرارية (Stationary) ، وتكون مستقرة إذا كانت جذور المعادلة $\phi(B) = 0$ خارج دائرة الوحدة. ولإيجاد دالة التباين المشترك لهذا النموذج فانه يتم ضرب طرفي المعادلة بالحد Z_{t-k} واخذ التوقع فنحصل[٢] :

$$E(Z_t Z_{t-k}) = \phi_1 E(Z_{t-1} Z_{t-k}) + \cdots\cdots + \phi_\rho E(Z_{t-\rho} Z_{t-k}) + E(a_t Z_{t-k}) \qquad .....2-2$$

ومنها نحصل على صيغة التباين المشترك الاتية :

$$\gamma_k = \phi_1 \gamma_{k-1} + \cdots\cdots + \phi_\rho \gamma_{k-\rho} \qquad ,k > 0 \qquad .....2-3$$

إذ أن (Z_{t-k}) تعتمد على (a_{t-k}) و(a_{t-k-1}) و..... فقط ولذلك فهي غير مرتبطة مع (a_t) ، ولذلك فإنه $E(a_t Z_{t-k}) = 0$. ويمكن إيجاد دالة الارتباط الذاتي بقسمة الصيغة (٢-٣) على γ_0 فنحصل على الصيغة التالية :

$$\rho_k = \phi_1 \rho_{k-1} + \cdots\cdots + \phi_p \rho_{k-p} \qquad ,k > 0 \qquad .....2-4$$

وبالنسبة AR(1) فإن :

$$\gamma k = \begin{cases} \phi_1^k \dfrac{\sigma_a^2}{1-\phi_1^2} & ,k \neq 0 \\ \dfrac{\sigma_a^2}{1-\phi_1^2} & ,k = 0 \end{cases} \qquad \text{.....2-5}$$

$$\rho k = \phi_1^k \qquad \text{.....2-6}$$

٢-٣ نموذج الأوساط المتحركة Moving Average Model

النموذج يتحقق عندما يوجد عدد محدد من الأوزان غير المساوية للصفر ولتكن $-\theta_1, -\theta_2, \cdots\cdots, -\theta_q$ وان $-\theta_k = 0$ عندما k>q عند ذلك تكون العملية الناتجة هي نموذج الأوساط المتحركة من الرتبة (q) ويرمز لها بالرمز (MA(q)) ويعبر عنها بالصيغة التالية[١] :

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots\cdots - \theta_q a_{t-q} \qquad \text{.....2-7}$$

أو

$$Z_t = \theta(B) a_t$$

إذ أن

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots\cdots - \theta_q B^q)$$

ولان $1 + \theta_1^2 + \cdots\cdots + \theta_q^2 < \infty$ فإن عملية الأوساط المتحركة المحددة تكون دائما" مستقرة ، وتكون قابلة للانعكاس إذا كانت جذور المعادلة $\theta(B) = 0$ وهي دائما خارج دائرة الوحدة . ولإيجاد دالة التباين المشترك للنموذج نضرب الصيغة ٢-٧ بالحد (Z_{t-k}) ونأخذ التوقع للناتج فنحصل على الصيغة الاتية :

$$\begin{aligned} \gamma_k &= \mathrm{E}(Z_t Z_{t-k}) \\ &= \mathrm{E}\left[\left(a_t - \theta_1 a_{t-1} - \cdots\cdots - \theta_q a_{t-q}\right)\left(a_{t-k} - \theta_1 a_{t-k-1} - \cdots\cdots - \theta_q a_{t-k-q}\right)\right] \end{aligned} \qquad \text{.....2-8}$$

ومنها نحصل على

$$\gamma_0 = (1 + \theta_1^2 + \cdots\cdots + \theta_q^2)\sigma_a^2 \qquad \text{.....2-9}$$

$$\gamma_k = \begin{cases} (-\theta_k + \theta_1\theta_{k-1} + \cdots\cdots + \theta_q\theta_{k-q})\sigma_a^2 & ,k = 1,2,\ldots\ldots,q \\ 0 & ,k \geq q+1 \end{cases} \qquad \text{.....2-10}$$

ويمكن الحصول على صيغة الارتباط الذاتي للنموذج بقسمة ٢-١٠ على ٢-٩ لنحصل على :

$$\rho_k = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k-1} + \cdots\cdots + \theta_q\theta_{k-q}}{1 + \theta_1^2 + \cdots\cdots + \theta_q^2} & ,k = 1,2,\ldots\ldots,q \\ 0 & ,k \geq q+1 \end{cases} \qquad \text{.....2-11}$$

وان الدالة التباين المشترك للنموذج MA(1) هي:

$$\gamma_k = \begin{cases} (1+\theta_1^2)\sigma_a^2 & ,k = 0 \\ -\theta_1\sigma_a^2 & ,k = 1 \\ 0 & ,k > 1 \end{cases} \qquad \text{.....2-12}$$

اما دالة الارتباط الذاتي فتكون

$$\rho_k = \begin{cases} \dfrac{-\theta_1}{1-\theta_1^2} & ,k=1 \\ 0 & ,k>1 \end{cases} \qquad \text{.....2-13}$$

٢-٤ النموذج المختلط للارتباط الذاتي والأوساط المتحركة

The Mixed Autoregressive Moving Average Model

قد يكون اقرب تعبير عن السلاسل الزمنية ذو اقل عدد من المعلمات ، هو بالنماذج المختلطة (Mixed) لنماذج الانحدار الذاتي مع نماذج الأوساط المتحركة ، هو النموذج الذي يدعى (ARMA Model) ويرمز له بالرمز ARMA(p,q) ، ويعبر عنه وفق الصيغة التالية :

$$Z_t = \phi_1 Z_{t-1} + \cdots\cdots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \cdots\cdots - \theta_q a_{t-q} \qquad \text{.....2-14}$$

أو

$$\phi(B) Z_t = \theta(B) a_t$$

إذ أن

$\phi(B)$: هي متعددة حدود في (B) لمعالم نموذج الانحدار الذاتي $(\phi_1, \cdots\cdots, \phi_p)$

$\theta(B)$: هي متعددة حدود في (B) لمعالم نموذج الأوساط المتحركة $(\theta_1, \cdots\cdots, \theta_q)$

ولكي تتوفر الاستقرارية في هذا النموذج يجب إن تكون جذور المعادلة $\phi_p(B) = 0$ هي خارج دائرة الوحدة ، وكذلك بالنسبة إلى القابلية للانعكاس يجب إن تكون جذور المعادلة $\theta_q(B) = 0$ هي خارج دائرة الوحدة أيضا[١] .

وتكون دالتي التباين المشترك والارتباط الذاتي للنموذج ARMA(p,q) للصيغة ٢-١٤ كما في الشكل التالي :

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots\cdots + \phi_p \gamma_{k-p} \qquad ,k \geq q+1 \qquad \text{.....2-15}$$

$$\rho_k = \phi_1 \rho_{k-1} + \cdots\cdots\cdots\cdots + \phi_p \rho_{k-p} \qquad ,k \geq (q+1) \qquad \text{.....2-16}$$

اما بالنسبة إلى ARMA(1,1) فان دالة التباين والارتباط يكونان :

$$\gamma_0 = \frac{(1+\theta_1^2 - 2\phi_1\theta_1)}{(1-\phi_1^2)} \sigma_a^2 \qquad \text{.....2-17}$$

$$\gamma_1 = \phi_1 \gamma_0 - \phi_1 \sigma_a^2$$

$$= \frac{(\phi_1 - \theta_1)(1-\phi_1\theta_1)}{(1-\theta_1^2)} \sigma_a^2 \qquad \text{.....2-18}$$

$$\rho_k = \begin{cases} 1 & ,k=0 \\ \dfrac{(\phi_1 - \theta_1)(1-\phi_1\theta_1)}{1-\theta_1^2 - 2\phi_1\theta_1} & ,k=1 \\ \phi_1 \rho_{k-1} & ,k \geq 2 \end{cases} \qquad \text{.....2-19}$$

٢-٥ الاستقرارية Stationary

يقصد بالاستقرارية إن تميز العملية بأن خواصها الإحصائية لا تتغير مع تغير الزمن ، أو تكون العزوم المشتركة موجودة ومساوية إلى العزوم المشتركة حتى الرتبة m .

تعريف (١) :[٦]

تكون العملية (X_t) مستقرة بشكل كامل إذا وجدت ($t_1 , t_2 , ... , t_n$) وأي (k) إذ أن التوزيع الاحتمالي المشترك إلى $\left(X_{t_1}, X_{t_2}, \cdots\cdots, X_{t_n}\right)$ يتماثل مع التوزيع الاحتمالي المشترك إلى $X_{(t_1+k)}, X_{(t_2+k)}, \cdots\cdots, X_{(t_n+k)}$

أي إن

$$F_{X_{t_1},\cdots;X_{t_n}}\left(X_1,\cdots\cdots;X_n\right)\equiv F_{X_{t_1+k},\cdots;X_{t_n+k}}\left(X_1,\cdots\cdots;X_n\right)$$

تعريف (٢) :[٦]

تكون العملية (X_t) مستقرة حتى الرتبة (m) إذا وجدت ($t_1 , t_2 , ... , t_n$) وأي (k) فإن العزوم المشتركة حتى الرتبة (m) لـ $\left(X_{t_1}, X_{t_2}, \cdots\cdots, X_{t_n}\right)$ تكون موجودة ومساوية إلى العزوم المشتركة حتى الرتبة (m) لـ ($X_{(t_1+k)}, X_{(t_2+k)}, \cdots\cdots, X_{(t_n+k)}$) أي بعد الإزاحة (k) فتكون كما في الصيغة التالية :

$$\mathrm{E}\left|\left\{X_{t_1}\right\}^{m_1}\left\{X_{t_2}\right\}^{m_2}\cdots\cdots\left\{X_{t_n}\right\}^{m_n}\right| = \mathrm{E}\left|\left\{X_{t_1+k}\right\}^{m_1}\left\{X_{t_2+k}\right\}^{m_2}\cdots\cdots\left\{X_{t_n+k}\right\}^{m_n}\right|$$

ولأي (k) وكل الأعداد الصحيحة $m_1 , m_2 , ... , m_n$

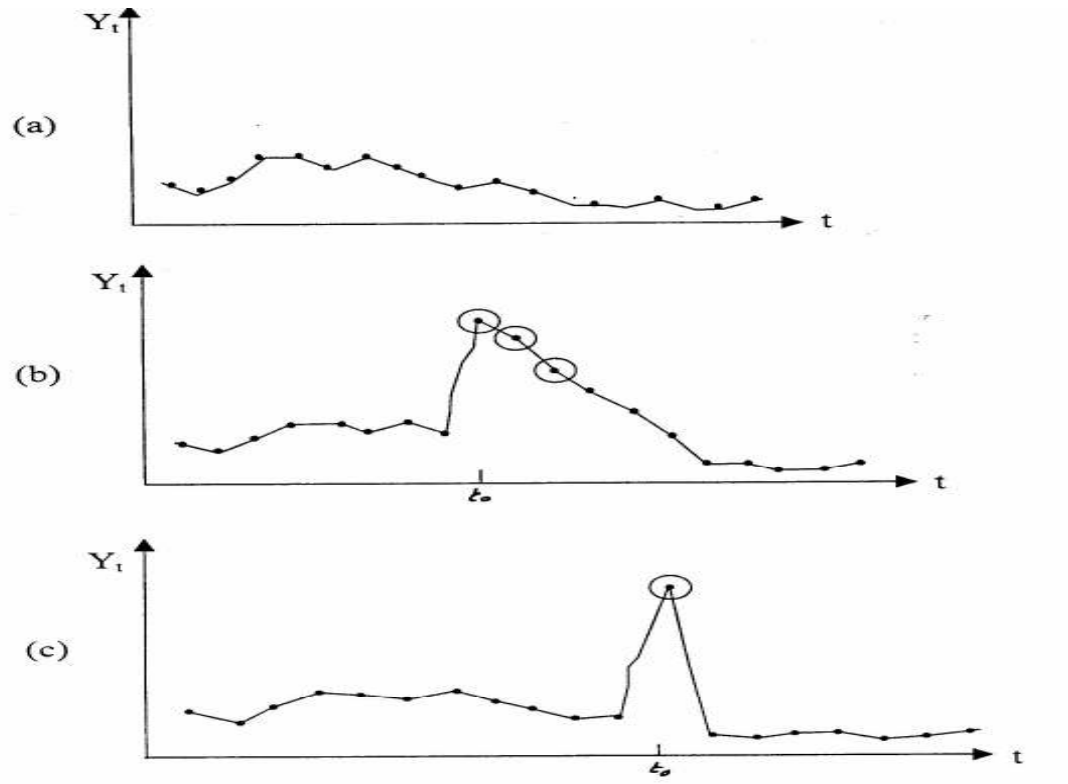
ويمكن إن نقول النموذج يمتلك الاستقرارية وذلك عندما يكون بوسط ثابت وتباين محدد .

٢-٦ الشوارد في السلاسل الزمنية

تتأثر السلسلة الزمنية أحيانا" ببعض المتغيرات الخارجية التي لها تأثيرا" فعالا" على بناء النموذج ، منها ما يتعلق بالضربات العسكرية ونشوب الحروب ، والتغيرات الفجائية لأسعار البضائع في الأسواق ، والطفرات الوراثية في الدراسات الحياتية وغيرها الأمر الذي يؤدي إلى جعل السلسلة الزمنية تتضمن قيما" شاردة عن النسق الذي تتبعه بقية قيم هذه السلسلة . ولذلك من المهم تحديد إن كان للسلسلة شوارد ام لا وما هي طبيعتها قبل القيام بتحليل السلسلة أو النموذج . ومعرفة طرق معالجتها أو إزالة تأثيرها بعد معرفة موقعها .

وقد حدد (Fox – 1972)[٧] ومن ثم (Donbg & Martin – 1979)[٨] نوعين من الشوارد وهي الشوارد النمطية (Innovation outliers) والشوارد المضافة (Additive outliers) .

الشوارد النمطية (Io) يتحقق هذا النوع من الشوارد عندما يكون توزيع الأخطاء ثقيل الاطراف بالنسبة إلى التوزيع الطبيعي . بمفهوم أخر إن الشوارد النمطية سوف تؤثر على بقية القيم التي تلي موقعها ، وهذا يعني إن الشوارد النمطية تعرف على يمين أول قيمة شاردة . كما في الشكل (1-b) الشوارد المضافة (Ao) يتحقق هذا النوع من الشوارد عندما لا تحقق السلسلة الزمنية (Z_t) النموذج المفترض لقيمها المشاهدة . بمفهوم أخر إن أحدى القيم تكون خارجة عن النسق العام . كما في الشكل (1-c)



الشكل (١) يوضح انواع الشوارد في السلاسل الزمنية اذ ان (a) : عدم وجود شوارد (b) : شوارد نوع (Io) ، (c) : شوارد نوع (Ao)

٢-٦-١ الكشف عن الشوارد في السلاسل الزمنية

إن عملية الكشف عن الشوارد وتحديد طبيعتها هي المرحلة الأكثر أهمية في عملية معالجة تأثيراتها السلبية على مقدرات معالم النماذج المختلفة . ويعد مقياس نسبة الامكان من المقاييس المهمة والفعالة في هذا المجال إذ يتم استخدامه في الحالتين التاليتين :

الحالة الأولى : عندما تكون معلمات النموذج وكذلك التباين معلومة[٩] :

في هذه الحالة يتم تقدير التأثير (ω) للنموذج (Io) حسب الصيغة التالية :

$$y_t = Z_t + \frac{\theta(B)}{\phi(B)}\omega\zeta^{(T)} \qquad .....2-20$$

اما للنموذج (Ao) فتكون الصيغة

$$y_t = \frac{\theta(B)}{\phi(B)}a_t + \omega\zeta^{(T)} \qquad .....2-21$$

إذ أن المعلومات حول وجود (Io) ستكون متوفرة في سلسلة الأخطاء e_t على يمين النقطة T ، في حين إن (Ao) سوف تكون منتشرة في سلسلة الأخطاء ..... , e_{t+1} , e_t ويكون النموذج

$$Z_t = \psi(B)\zeta_t^T \qquad .....2-22$$

وعليه يكون

$$Z_t = \begin{cases} 0 & for \quad t < T \\ 1 & t = T \\ -\psi_{t-T} & t > T \end{cases}$$

وباستخدام المربعات الصغرى (Ls) يمكن تقدير (ω) كما يأتي :

$$\hat{\omega}_{Io} = e_T$$

$$\hat{\omega}_{A0} = \rho^2 \psi(F) e_t$$

$$= \rho^2 \left(1 - \psi_1 F - \psi_2 F^2 \cdots\cdots\cdots \psi_{n-T} F^{n-T}\right) e_t$$

إذ أن

$$\rho 2 = (1 + \psi_1^2 + \cdots\cdots\cdots + \psi_{n-T}^2)^{-1}$$

و F هو معامل الارتداد الأمامي أي إن ($Fe_t=e_{t+1}$) كما إن التباين لهذه المقدرات يكون وفق الصيغة التالية :

$$Var(\hat{\omega}_{Ao}) = \rho^2 \sigma_a^2 \quad و \quad Var(\hat{\omega}_{Io}) = \sigma_a^2$$

وإذا وضعت الفرضية $[H_0 : \omega = 0]$ في الصيغة (٢٠-٢) والصيغة (٢١-٢) ضد الفرضية $H_1 : \omega \neq 0$ فإن مقياس نسبة الامكان لاختبار فرضية ضد أخرى سيكون كمايلي :

$$H_0 \quad V-S \quad H_1 : \lambda_{1,T} = \hat{\omega}_{Io} / \sigma_a \quad \sim N(0,1)$$

$$H_0 \quad V-S \quad H_2 : \lambda_{2,T} = \hat{\omega}_{Ao} / \rho\sigma_a \quad \sim N(0,1) \quad .....2-23$$

$$H_1 \quad V-S \quad H_3 : \lambda_{3,T} = \frac{\rho^{-2}\hat{\omega}_{Ao}^2 - \hat{\omega}_{Io}^2}{2\sigma_a^2 (1-\rho^2)^{1/2}}$$

وعليه فأن نسبة الامكان تحتاج لكي تحدد وجود قيم شاردة في السلسلة ومن أي نوع ، إلى $\max|\lambda_{1,T}|$ في حالة وجود شوارد من النوع (Io) و $\max|\lambda_{2,T}|$ في حالة وجود شوارد من نوع (Ao) .

الحالة الثانية : عندما تكون معلمات النموذج التباين مجهولة[٩] .

في اغلب التطبيقات العملية تكون هذه الحالة الأكثر شيوعا" ، ولكن يمكن الحصول على التباين والمعالم من خلال تقديرها مع (ω) بالنسبة لكل نموذج سواء" كان (Io) أو (Ao) ، وذلك باستخدام أسلوب (MLE (Maximum Likelihood Estimation) ومن إيجاد كل من $\hat{\lambda}_{1,T}$, $\hat{\lambda}_{2,T}$ و $\hat{\lambda}_{3,T}$ وكماياتي :

$$\hat{\lambda}_{1,T} = \hat{\omega}_{Io} / \hat{\sigma}_a$$

$$\hat{\lambda}_{2,T} = \hat{\omega}_{Ao} / \hat{\rho}\hat{\sigma}_a \quad .....2-24$$

$$\hat{\lambda}_{3,T} = \frac{\hat{\rho}^{-2}\hat{\omega}_{Ao}^2 - \hat{\omega}_{Io}^2}{2\hat{\sigma}_a^2 (1-\hat{\rho}^2)^{1/2}}$$

وان

$$\hat{\omega}_{Io} = \hat{e}_t$$
$$\hat{\omega}_{Ao} = \hat{\rho}^2(1-\hat{\psi}_1 F - \hat{\psi}_2 F^2 - \cdots\cdots\cdots - \hat{\psi}_{n-T} F^{n-1})\hat{e}_T$$
$$\hat{\rho}^2 = (1+\hat{\psi}_1^2 + \hat{\psi}_2^2 + \cdots\cdots\cdots + \hat{\psi}_{n-T}^2)^{-1}$$

وهذه في الحقيقة تكافي تقريبا" نسبة الامكان عندما تكون معلمات النموذج والتباين معلومة .

وللكشف عن (Io) و (Ao) ذات المواقع المجهولة فإنه يمكن اختبار ذلك عن طريق

$$\left.\begin{aligned}\hat{\eta}_{Io} &= \max|\lambda_{1,t}| > c \qquad ,t = 1,\ldots\ldots\ldots\ldots,n \\ \hat{\eta}_{Ao} &= \max|\lambda_{2,t}| > c \qquad ,t = 1,\ldots\ldots\ldots\ldots,n\end{aligned}\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots 2-25$$

إذ أن (c) عدد ثابت موجب اختياري .

وللتميز بين أنواع الشوارد Io و Ao يتم استخدام المتباينات التالية :

إذا كانت $\left|\hat{\lambda}_{1,T}\right| > \left|\hat{\lambda}_{2-T}\right|$ فهذا يعني وجود شوارد من النوع (Io)

إذا كانت $\left|\hat{\lambda}_{1,T}\right| \le \left|\hat{\lambda}_{2-T}\right|$ فهذا يعني وجود شوارد من النوع (Ao)

٢-٦-٢ الأسلوب المكرر لكشف الشوارد The iterative Outlier Detection Procedure

يتبع هذا الأسلوب خطوات محددة تعتمد على الصيغ الأولية في مقياس نسبة الامكان لكشف الشوارد والتخلص من تأثيرها على السلسلة الزمنية من دون الاستغناء عن أي مشاهدة فيها ، والذي يستبعد هنا هو فقط تأثير الشوارد التي يتم الكشف عنها ، ذلك باتباع الخطوات التالية :

١- يتم تقدير النموذج بإحدى الطرق التقليدية ، ومن ثم حساب الخطأ (e_t) و $\hat{\sigma}_a^2 = \sum_{i=0}^{n} \frac{\hat{e}_t^2}{n}$ كتقدير إلى $\hat{\sigma}_a^2$ ، وكتقدير حصين إلى ($\hat{\sigma}_a^2$) يمكن اعتماد الوسيط للقيم المطلقة للبواقي .

٢- حساب $\left(\hat{\lambda}_{1,T}\right)$ و $\left(\hat{\lambda}_{2,T}\right)$ كما في الصيغة (٢٤-٢) مع افتراض

$$\Delta_t = \max\left\{\left|\hat{\lambda}_{1,T}\right|,\left|\hat{\lambda}_{2,T}\right|\right\} \qquad ,t = 1,2,\ldots\ldots\ldots\ldots,n \qquad \ldots\ldots\ldots\ldots 2-26$$

فإذا كانت $\max \Delta t = \left|\hat{\lambda}_{1,T}\right| > c$ ، إذ أن (c) هي قيمة ثابتة موجبة تؤخذ عادة مابين ٣ و ٤ ، فإن ذلك يعني إمكانية وجود شاردة من نوع (Io) ، ويمكن تقدير تأثيره بواسطة $\left(\hat{\omega}_{Io}\right)$ ويتم حذفه بتعريف أخطاء جديدة عند الزمن (T) بالصيغة التالية :

$$\tilde{e}_T = \hat{e}_T - \hat{\omega}_{Io} = 0 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2-27$$

اما إذا كانت $\max \Delta t = \left|\hat{\lambda}_{2,T}\right| > c$ فإن ذلك يعني إمكانية وجود شاردة من نوع (Ao) عند الزمن (T) وأن تقدير تأثيره $\left(\hat{\omega}_{Ao}\right)$ يمكن ازالته بتعريف أخطاء جديدة وفق الصيغة التالية :

$$\tilde{e}_t = \tilde{e}_t - \hat{\omega}_{Ao}\hat{\psi}(B)\zeta^{(T)} \qquad , \qquad t \ge T \qquad \ldots\ldots\ldots\ldots 2-28$$

وبالتالي إيجاد $\left(\hat{\sigma}_a^2\right)$ للأخطاء المعدلة في كلا الحالتين ولتكن $\left(\tilde{\sigma}_a^2\right)$.

٣- إذا حدث أي من نوعي الشوارد (Io) أو (Ao) كما في الخطوة (٢) فإنه يعاد حساب $\left(\hat{\lambda}_{1,T}\right)$ و $\left(\hat{\lambda}_{2,T}\right)$ بالاعتماد على نفس المقدرات الأولية لمعلمات السلسلة الزمنية ولكن باستخدام سلسلة الأخطاء الجديدة المعدلة وكذلك $\left(\tilde{\sigma}_a\right)$ ومن ثم تكرار الخطوة الثانية مرة أخرى .

٤- يتم الاستمرار في تكرار الفقرتين (٢) و (٣) لحين اختفاء الشوارد وعدم وظهورها في مراحل الاختبار ، وبالتالي اختفاء تأثيرها وهذا يعني الوصول إلى مقدرات حصينة بالاعتماد على المقدرات التقليدية لمعلمات السلسلة الزمنية.

٣- <u>الجانب التجريبي</u> :

٣-١ <u>توليد الأعداد العشوائية :</u>

في تجارب المحاكاة تكون الخطوة الأولى هي توليد الأعداد العشوائية ، والتي تتبع التوزيع المنتظم المستمر (Continuous Uniform Distribution) والتي من المفترض تكون مستقلة . وتقوم الخطوة التالية بتهيئة الوسيلة الرياضية الخاصة بتحويل العدد العشوائي المنتظم إلى متغير عشوائي من توزيع يصف النموذج تحت التجربة ، والجدير بالذكر أن أفضل أسلوب للتحويل هو استخدام طريقة التحويل المعكوس كوسيلة لتوليد متغيرات عشوائية[١٠] وذلك بالاعتماد على دالة الكثافة التراكمية c.d.f :

$$y = F(x)$$
$$x = F^{-1}(y) \qquad \text{.....3-1}$$

وعلى سبيل المثال إذا كان (x) يمثل متغير عشوائي يتبع التوزيع الاسي $x \sim Exp(\theta)$ بدالة التوزيع c.d.f: $F(x) = 1 - e^{-\theta x}$ فانه بالامكان التعويض كالاتي :

$$u = 1 - e^{-\theta x}$$

إذ إن u تمثل متغير عشوائي منتظم ، وبذلك فان المتغيرات العشوائية ذات التوزيع الاسي يمكن توليدها ، وفق الصيغة التالية :

$$x = -\frac{1}{\theta} Log(1-u) \qquad \text{.....3-2}$$

ونظرا" لكون (1-u) يمثل متغيرا" عشوائيا" ذا توزيع منتظم أيضا ، وعليه يكون

$$x = -\frac{1}{\theta} Log(u) \qquad \text{.....3-3}$$

ونظرا" لأهمية التوزيع الطبيعي واستخداماته الشائعة ــ فان هناك طرق عدة لتوليد المتغير ، أهمها طريقة بوكس-ميلر (Box – Muller) عام ١٩٥٨ وصيغتها كماياتي[١٠] :

$$X_1 = \left[-2Lnu_1\right]^{1/2}\left[Sin(2\Pi u_2)\right]$$
$$X_2 = \left[-2Lnu_1\right]^{1/2}\left[Cos(2\Pi u_2)\right] \qquad \text{.....3-4}$$

إذ أن u_1 , u_2 تتبعان التوزيع المنتظم u(1,0) وان x_1 و x_2 تتبعان التوزيع الطبيعي القياسي .

٣-٢ تطبيق أسلوب المحاكاة

في تطبيق تجارب المحاكاة للسلاسل الزمنية وبوجود الشوارد فقد تم توليد الأخطاء لنموذج AR(1) والنموذج MA(1) بطريقة بوكس – ميلر كما في الصيغ (٤-٣) . كما تم استخدام عدد من المكررات واحجام مختلفة وقد كانت :

تجربة محاكاة للكشف عن الشوارد للنماذج AR(1) و MA(1) و AR(2) و ARMA(1,1) ولاحجام عينات مختلفة (١٥٠ ، ١٠٠ ، ٥٠) ، وعندما تكون هناك شاردة واحدة أو قيمتين شاردتين ومن كلا النوعين .

٣-٣ الكشف عن الشوارد

تم تصميم تجارب المحاكاة للكشف عن وجود الشوارد سواء كانت من نوع (Io) أو (Ao) على انفراد أو كلاهما معا" . ففي حالة وجود شوارد من نوع واحد فقد تم اختبار ٢٤ حالة ناتجة عن توافيق سلسلتين زمنيتين للمتغير x_1 هما AR(1) وذلك عندما $\phi = 0.6$ و $\sigma_a^2 = 1$ و MA(1) عندما $\theta = 0.6$ و $\sigma_a^2 = 1$ وذلك بوجود نوعين من الشوارد هما (Io) و (Ao) و ARMA(1,1) وذلك عندما $\phi = 0.6$ ، $\theta = 0.6$ و $\sigma_a^2 = 1$.

كما تم استخدام قيمتين مختلفتين للشوارد هما $\omega = 3\sigma_a^2$ و $\omega = 5\sigma_a^2$ وقد تم استخدام ثلاث احجام مختلفة للعينة وهي (١٥٠ ، ١٠٠ ، ٥٠) .

أن موقع القيمة الشاردة قد حدد في منتصف فترة المشاهدات عند وجود قيمة شاردة من نوع واحد . حيث كان موقع القيمة الشاردة عندما تكون حجم العينة (n=50) هو (T=26) وعندما تكون (n=100) فإن (T=51) وعندما تكون (n=150) فإن (T=76) .

وعند وجود قيمتين من الشوارد (Io) و (Ao) فقد تم دراسة النموذج AR(1) في حالة وجود ثلاثة أنواع منها وهي وجود قيمتين من الشوارد من نوع (Ao) وقيمتين من الشوارد من نوع (Io) وقيمة واحدة من (Io) وأخرى من نوع (Ao) وقد افترض لكل من نوعي الشوارد إن تكون قيمها $\omega = 3\sigma_a^2$ و $\omega = 5\sigma_a^2$ وذلك لثلاثة احجام العينات (50 , 100 , 150=n) اما مواقع الشوارد المفترضة فقد تم وضعها في الثلث الأول والثلث الثاني من المشاهدات أي (T_1=17) و (T_2=34) عندما (n=50) و (T_1=34) و (T_2=66) عندما (n=100) و (T_1=51) و (T_2=101) عندما (n=150) . وفي كلتا الحالتين السابقتين فقد تم توليد السلسلة الزمنية طبقا" لهذه المواصفات حيث افترض إن النموذج معلوم وانه ليس هناك شوارد في العينة . ومن الجدير بالذكر انه قد تم استخدام ثلاث قيم حرجة للثابت (c) وهي(٣ ، ٢,٥ ، ٢) كلا على حده والجداول (١-٣) و (٢-٣) و (٣-٣) تبين تكرارات الكشف والتصحيح لمواقع الشوارد باستخدام عدد من المكررات (١٠٠٠) والقيم الحرجة المذكورة .ومن خلال الجداول تبين مايلي :

١- كلما ازداد حجم العينة أدى ذلك إلى زيادة عدد مكررات الكشف للمواقع المفترضة للقيمة الشاردة . على سبيل المثال نلاحظ جدول (١-٣) كانت قيمتها (٥٦٧) عند حجم (٥٠) بينما تكون (٥٨١) عند حجم (١٥٠) .

٢- تزداد عدد المكررات المصححة عند زيادة القيمة الافتراضية للشوارد حيث إن القيم الافتراضية $3\sigma_a^2$ و $5\sigma_a^2$ فمثلا في جدول (١-٣) وعند حجم (٥٠) وتحت القيمة $3\sigma_a^2$ تكون (٥٦٧) اما عند القيمة $5\sigma_a^2$ فتكون (٨١٨) .

٣- بشكل عام إن عدد المكررات للمواقع المصححة في حالة وجود (Ao) هي اكبر بكثير من عدد المكررات في حالة وجود (Io) .

٤- يلاحظ أن القيم الشاردة في النموذج AR(1) والتي هي من نوع (Io) تكون أكثر تذبذبا" في عدد المكررات عند تزايد حجم العينة حيث نجد إن عدد المكررات عند حجم العينة (١٠٠) تساوي ٥٠٧ ثم بدأت المكررات بالتناقص لتكون قيمتها (٤٧٤) عندما يكون حجم العينة (١٥٠) وهذا يشير إلى أنه كلما ازداد حجم العينة فإن المعالجة للشوارد من نوع (Io) وللنماذج AR(1) تكون أقل آلية من غيرها .

٥- نجد أن قيم المكررات في الجداول الثلاثة تتناقص كلما ازدادت قيم (c) وعلى سبيل المثال إن قيمة المكرر للنموذج Io ١ و MA وعند حجم عينة تساوي (٢٢٩) عند المقارنة مع القيمة الحرجة (c=2) في الجدول رقم (١-٣) وتساوي (١٦١) عند المقارنة مع القيمة الحرجة في الجدول رقم (٢-٣) وتساوي ٧٠ عند المقارنة مع القيمة الحرجة (c=3) والموجودة في الجدول رقم (٣-٣) . وهذا يعد منطقيا حيث كلما ازدادت قيمة (c) انخفضت عدد المكررات لمواقع الشوارد .

٦- نلاحظ انه كلما ازدادت عدد الشوارد في النموذج نجد إن عدد المكررات تتزايد خصوصا" عندما يكون نوع القيمة الشاردة من نوع (Ao) فيما نجد إن هذا العدد يتقلص مع تزايد قيمة (c) ويتزايد مع ازدياد حجم العينة سواء كانت قيمة الشوارد $3\sigma_a^2$ أو $5\sigma_a^2$ فيما نجد إن هناك انخفاضا" في اعداد هذه المكررات عندما تكون الحالة ممثلة للشوارد من نوع (Io) .

٧- يلاحظ ان القيم الشاردة في النموذج ARMA(1,1) والتي من نوع (Ao) تكون عدد المكررات فيها اكثر من حالة وجود نوع (Io) .

٨- يلاحظ ان عدد التكرارات يزداد كلما زادت درجة النموذج في تصحيح الاخطاء ويكون واضحا من خلال المقارنة بين AR(1) و MA(1) وبين AR(2) و ARMA(1,1) .

جدول (١-٣): يبين تكرار كشف التصحيح لمواقع الشوارد عند c=2.0 وعدد المكررات (١٠٠٠)

ω / n / النموذج	$3\sigma_a^2$			$5\sigma_a^2$		
	٥٠	١٠٠	١٥٠	٥٠	١٠٠	١٥٠
AR , 1 Ao	٥٦٧	٥٧٤	٥٨١	٨١٨	٨٣٢	٨٤٤
AR , 1 Io	٢٥٥	٥٠٧	٧٤٧	٦١١	٦٠٢	٦٤١
MA , 1 Ao	٦١٢	٧٣٧	٧٠٨	٨٨٣	٨٩٢	٨٩٥
MA , 1 Io	٢٢٩	٣٦٣	٥٨٩	٧٢١	٧٣٢	٧٦٠
AR , 2 Ao's	٧٣٦	٨٣٠	٨٩٥	٩٤١	٩٦٠	٩٣٤
AR , 2 Io's	١٩٠	٢٨٦	١٦٩	٩٠٨	٩١٢	٩٣١
AR , Io Ao	٣٤٤	٤٥٨	٧٣٣	٧٢٦	٧٤١	٧٩٧
ARMA(1,1) Ao	٥٩٤	٦٥٥	٦٤٦	٨٣٩	٨٦١	٨٧٧
ARMA(1,1) Io	٣٤١	٥٢٣	٧٥٦	٧٨٨	٧٧٥	٧٥٨

جدول (٢-٣): يبين تكرار كشف التصحيح لمواقع الشوارد عند c=2.5 وعدد المكررات (١٠٠٠)

ω / n / النموذج	$3\sigma_a^2$			$5\sigma_a^2$		
	٥٠	١٠٠	١٥٠	٥٠	١٠٠	١٥٠
AR , 1 Ao	٤٧٥	٥٤٦	٥٨١	٧٩٩	٨٠٦	٨١٩
AR , 1 Io	١٧٥	٤٦٤	٤٥١	٦٠٣	٦١١	٦٠٨
MA , 1 Ao	٤٨٤	٧٠٤	٧٠٤	٨٥١	٨٥٤	٨٥٩
MA , 1 Io	١٦١	٣٢٠	٥٥٥	٦٩٦	٧٠٢	٧٠٦
AR , 2 Ao's	٩٧١	٨١٩	٨٩٠	٩٠٤	٩٤٠	٩٤٤
AR , 2 Io's	١٥٥	٢٨٠	٥٠٣	٨١١	٨١٩	٨٩٠

AR , Io Ao	٢٦٣	٤٠٦	٦٧٦	٦٩٣	٦٧٢	٦٨٥
ARMA(1,1) Ao	٥٧٩	٦٢٥	٦٤٦	٨٢٤	٨٣٢	٨٣٨
ARMA(1,1) Io	٣٠٧	٥٣٢	٦١٢	٧٦٧	٧٩٨	٧٩١

جدول (٣-٣): يبين تكرار كشف التصحيح لمواقع الشوارد عند c=3 وعدد المكررات (١٠٠٠)

ω / النموذج n	$3\sigma_a^2$			$5\sigma_a^2$		
	٥٠	١٠٠	١٥٠	٥٠	١٠٠	١٥٠
AR , 1 Ao	٣٣٥	٤١٤	٤٨٤	٦٢٥	٦٧٥	**٦٩٦**
AR , 1 Io	٧١	١٨٨	٢٧٣	٥١٦	٥١٨	**٥٣٢**
MA , 1 Ao	٣٢٣	٥٧٥	٥٥٦	٧٩٤	٧٧٦	**٨٢٤**
MA , 1 Io	٧٠	١٦٧	٢٦٩	٦١٣	٦٢١	**٦٥٨**
AR , 2 Ao's	٥١٧	٦٢٨	٧٢٠	٨٠٧	٨١٩	**٨٠٢**
AR , 2 Io's	٣٣	٩٢	٥٠٣	٧٠٨	٧١٩	**٧٠٨**
AR , Io Ao	٣٦	٨٠	١٩٧	٥٧٨	٥٨١	**٥٩٢**
ARMA(1,1) Ao	٣٣٢	٤٧٦	٥١٣	٦٧٢	٦٨٨	**٧٠٥**
ARMA(1,1) Io	١٦٦	٢٤٢	٣٣٤	٦٦٧	٧٢٢	**٧٨٣**

٤- الاستنتاجات والتوصيات:

الاستنتاجات :

من خلال ما تقدم من الدراسة النظرية والتجريبية يمكن التوصل إلى الاستنتاجات التالية :

١- أن استخدام أساليب الكشف عن الشوارد في السلاسل الزمنية يكون فعالا جدا" في حالة الدرجات الدنيا بينما تقل هذه الفعالية أو الجدوى كلما زادت درجة النموذج ، إذ إنها تكون بحاجة إلى عدد كبير جدا" من المكررات لتصحيح موقع واحد من مواقع الشوارد .

٢- إن الشوارد من نوع (Ao) تكون أكثر تأثيرا" على النموذج من النوع (Io) كما انها تحتاج إلى أعداد اكبر من المكررات لتصحيح مواقع الشوارد من هذا النوع .

التوصيات :

من خلال ما تقدم فإنه يمكن التوصية بمايأتي :

١- استخدام أساليب الكشف عن الشوارد في السلاسل الزمنية قبل البدء بعمليات التقديرات لأي نماذج كانت ، لغرض التأكد من عدم احتوائها عل أي شوارد تؤثر على مستوى كفاءة المقدرات والتنبؤ .

٢- تطبيق أسلوب الكشف عن الشوارد في السلاسل الزمنية على نماذج مختلفة ذات رتب أعلى مما تم تجريبه في هذه الدراسة .

المصادر :

1) Wei , W.W.S., *Time series Analysis : Univariate and Multivariate Methods* , Addison – Wesley publishing company , Inc. , New York, 1990 .

2) Wegman , Edward J., *Time Series Analysis : Theory , Data Analysis And Computation* , Addison – Wesley publishing company , Inc. , New York , 2000.
3) Huber , Peter J., *Robust Statistics* , John Wiley and Sons , New York , 1981.
4) Box , G.P. and Jenkins , G.M., *Time Series Analysis : Forecasting and Control* , Revised edition , Holden – day – Inc. , Sonfracisco , 1976.
5) Basawa , I.V. , Huggins , R.M. and staudte , R.G. , " Robust Tests For Time Series With An Application To First Autoregressive Processes " , *Biomtrika* , 72 , 559-571 , 1985 .
6) Priestley , M.B. , *Spectral Analysis And Time Series* , Academic press Harcort Brace Jovanovich , Publishers , London , San Diego , New York , Tokyo Toronto, 1981 .
7) Fox , A.J. , " Outliers In Time Series " , *J. of the Royal statistical society* , Ser.B , 43 , 350-363 , 1972 .
8) Donby , L. and Martin , R.D. , " Robust Estimation Of The First Order Autoregressive Parameter " , *JASA* , 74 , 365 , 140-146 , 1979 .
9) Guo , H.J. , "Robust Estimation Of The Coefficient Of The First Order Autoregressive Prosess " , *Commun – Statist - theory method* , 29 , 1 , 55-66 , 2000 .
10) GEORGE S.Fishman , *Concept And Methods In Discrete Event Digital Simulation* , John Wiley and Sons , New York , London , Sydney , Toronto , 1973 .