

دراسة تأثير القيم الشاذة على طريقتي بوكس-جينكز وفورير

م.م فاضل حميد هادي

جامعة القادسية

كلية الإدارة والاقتصاد

قسم الإحصاء

الخلاصة

كل المقاييس الإحصائية تعتمد بشكل مباشر على البيانات المتوفرة، فكلما كانت هذه البيانات نقية كانت نتائجها دقيقة، لكن عندما تكون بيانات تلك المقاييس ملوثة بنسب معينة من البيانات الشاذة فإن نتائج تلك المقاييس تكون غير دقيقة. لذا فلسلاسل الزمنية هي الأخرى التي تتأثر بالقيم الشاذة لكن بصورة متفاوتة تختلف من طريقة تنبؤ إلى الأخرى، حيث في هذا البحث تم التركيز على طريقتي التنبؤ (طريقة بوكس-جينكز، وطريقة فورير) إذ تم المقارنة مابين هاتين الطريقتين عند ثلاث مراحل هي عندما لا توجد هناك بيانات شاذة (أي عندما تكون نسبة التلوث ٠%) والمرحلة الثانية من المقارنة هي عندما تكون نسبة التلوث بالبيانات الشاذة ٥% والمرحلة الثالثة من المقارنة هي عندما تكون نسبة التلوث ١٠% (لنموذج الموسمي ثابت، ونموذج موسمي تضاعفي) حيث تم التوصل إلى مجموعة من الاستنتاجات كما هي موضح من خلال هذا البحث.

المقدمة

قد تحتوي بيانات بعض الظواهر على بيانات شاذة (OUT LIARS) مما يؤدي إلى الوصول إلى نتائج غير دقيقة ومضللة، وبالتالي فإن هذه النتائج الغير جيدة ستؤدي إلى استنتاجات بعيدة عن واقع الظاهرة المدروسة، لذا لا بد من تشخيص هذه القيم الشاذة وكذلك التأكد من نسبتها قياساً بحجم العينة الكلية الممثلة للظاهرة المدروسة، إذ إن تأثيرها ومعرفة موقعها يختلف مع اختلاف طريقة التنبؤ وحجم العينة المستخدمة في السلاسل الزمنية. تعتبر السلاسل الزمنية من المواضيع الإحصائية الحساسة جداً إلى القيم الشاذة إذ تهتم السلاسل الزمنية بدراسة الظواهر التي تكون مشاهداتها مرتبة زمنياً، لذلك يجب أن يكون اختيارنا لأسلوب التحليل وطريقة التحليل مبنياً على أساس علمي ومعتمداً بشكل كبير على بيانات الظاهرة المدروسة. يوجد هناك أسلوبين في تحليل السلاسل الزمنية، الأول هو التحليل تحت مجال الزمن (Analysis under Time Domain) ومن الطرق الشائعة في هذا الأسلوب هي طرق بوكس جينكز وطرق التمهيد الأسي وطرق تحليل الانحدار أما الأسلوب الثاني هو التحليل تحت مجال الترددات (Analysis under Frequency Domain) من الطرق الشائعة في هذا الأسلوب هي التحليل الطيفي (Spectral analysis) وطرق فورير (Fourier methods). في هذا البحث تم إجراء المقارنة بين أسلوب التحليل في مجال الزمن (طريقة بوكس جينكز) وأسلوب التحليل في مجال الترددات (طريقة فورير) لبيانات موسمية ملوثة بنسب مختلفة من البيانات الشاذة (٠%، ٥%، ١٠%) وقد تم استخدام أسلوب المحاكاة في توليد البيانات الموسمية والأجراء المقارنة تم الاعتماد على مجموعة من أحجام العينات (٢٠٠، ١٥٠، ١٠٠، ٨٠، ٦٠، ٤٠، ٢٠)، وكذلك تم استخدام مجموعة من معايير دقة التنبؤات (Mse, Mae, Mape) بهدف إجراء المقارنة في ظل وجود وعدم وجود القيم الشاذة.

مشكلة الدراسة:

قد يكون وجود القيم الشاذة عائقاً كبيراً في الحصول على مقدرات دقيقة لمعالم طرق التنبؤ في السلاسل الزمنية طريقة (بوكس - جنكنز) و طريقة (طريقة فورير) وتكون نتائج تقدير هذه المعالم مضللة وغير دقيقة، وبالتالي فإن استخدام تنبؤات هذه الطرق ستعطي تنبؤات بعيدة عن الواقع ولحل هذه المشكلة لابد من استخدام طرق التقدير الحصينة في تقدير معالم الطرق المستخدمة في هذا البحث .

هدف البحث:

يهدف هذا البحث إلى إجراء مقارنة من حيث أفضلية التنبؤ مابين طريقة بوكس - جنكنز وطريقة فورير للتنبؤ بالسلاسل الزمنية الموسمية في حالة وجود نسب مختلفة من البيانات الشاذة، وكذلك تحديد أي الطريقتين أكثر تحسناً للبيانات الشاذة.

١ - الجانب النظري:

١-١ البيانات الشاذة (3) (2)

من المعروف ان دراسة الظواهر تعتمد على البيانات التي تخص تلك الظاهرة المدروسة ، فكلما كانت هذه البيانات نقية ودقيقة ستعكس نتائج ودراسات دقيقة وشاملة عن تلك الظاهرة المدروسة، لكن سرعان ما تواجه هذه الظواهر ظروفاً غير طبيعية تنتج عنها بيانات غير طبيعية ومن هذه البيانات الغير طبيعية هي البيانات الشاذة (out liars) . وقد عرفت البيانات الشاذة بمجموعة من التعاريف ومن ابسطها : وهي مجموعة من البيانات ذات توزيع يختلف عن توزيع باقي البيانات ضمن العينة الواحدة، وهناك عدة طرق لتحديد البيانات الشاذة ابسط هذه الطرق هي طريقة الدرجة المعيارية (التي تكون صيغتها الرياضية)

$$Z_i = \frac{X_i - \bar{X}}{s} \dots\dots\dots(1-1)$$

فإذا كانت قيمة (Z_i) ((للمشاهدة (i) حيث إن ($i=1,2,\dots,n$)) أكبر من (٣) فتعتبر هذه القيمة قيمة شاذة، لكن نجد ان هذه الطريقة غير مجدية نوعاً ما لأنها تعتمد على الوسط الحسابي الذي يتحسس للقيم الشاذة والمتطرفة ، لذا تم تحسين هذه الطريقة وذلك بالاعتماد على الوسيط الذي يعتبر مقياس غير متأثر بالقيم الشاذة أو المتطرفة ، وبالتالي فإن الصيغة الرياضية لهذه الطريقة تأخذ الشكل التالي :

$$Z_i = \frac{X_i - med(X_i)}{MAD(X_i)} \dots\dots\dots(1-2)$$

حيث إن

$$MAD(X_i) = med|X_i - med(X_i) |$$

كذلك اذا كانت قيمة (Z_i) ((للمشاهدة (i) حيث ان ($i=1,2,\dots,n$)) أكبر من (٣) فتعتبر هذه القيمة قيمة شاذة، ويوجد ه ناك مجموعة من الطرق لتشخيص القيم الشاذة ، وكذلك مجموعة من طرق معالجة البيانات الشاذة، ومن هذه الطرق طريقة البتر بعد تحديد نقطة البتر المثلى أو استخدام أسلوب التحويل، أو استخدام أسلوب التجزئة، لكن قد تتسبب طرق المعالجة هذه بفقدان كمية كبيرة من المعلومات وبالتالي فالإبقاء على القيم الشاذة أفضل من معالجتها ، كذلك فإن تأثير القيم الشاذة يختلف من مقياس إحصائي إلى آخر. وفيما يخص السلاسل الزمنية فإن تأثير القيم الشاذة يختلف من طريقة تنبؤ إلى أخرى.

٢-١ طريقة المربعات الصغرى المشذبة (LTS) (٢)(٣)

تعتبر طريقة المربعات الصغرى المشذبة من طرق التقدير الحصينة وتعمل هذه الطريقة كطريقة تقدير في ظل وجود او عدم وجود قيم شاذة (Out liars) حيث تعمل هذه الطريقة كطريقة تقدير اعتيادية في حالة غياب القيم الشاذة وتعمل كطريقة تقدير حصينة في ظل وجود القيم الشاذة. إذ يعتمد عمل هذه الطريقة على نسبة البيانات الشاذة الموجودة في البيانات الكلية . ويمكن إيجاد مقدرات هذه الطريقة من خلال تقليل مربعات البواقي إلى اقل ما يمكن

$$\text{Min} \sum_{i=1}^h r_{(i)}^2 \dots \dots \dots (1-3)$$

حيث إن

$r_{(i)}^2$ تمثل مربعات البواقي المرتبة

h هو حجم المجموعة الجزئية

ويمكن تحديدها كالتالي

$$h = [(1-a) \times n] + [a(k+1)]$$

حيث إن a هي النسبة المئوية للقيم الشاذة او هي نسبة تشذيب المشاهدات، اما n فهي حجم

العينة، و k هي عدد المعلمات

، و يمكن وصف خوارزمية هذه الطريقة ، كما يأتي

١- نحدد عدد القيم الشاذة (t)

٢- كذلك نحدد نسبة هذه القيم الشاذة من بين البيانات الكلية (n)

٣- نحسب التوافق للحصول على عدد المجموعات الجزئية ،

$$c_h^n = \frac{n!}{h!(n-h)!} = \dots \dots \dots (1-4)$$

اذ سوف يتم الحصول على عدد من المجموعات الجزئية يكون عددها (c_h^n) ،

٤- في هذه الحالة نقوم بحساب (ols) لكل مجموعة ثم نحسب (mse) لكل مجموعة تم تقدير

معالمها ، فالمجموعة التي تمتلك mse الأقل تعتبر مقدراتها حصينة . فسوف يتم الاعتماد على

طريقة التقدير هذه لتقدير معالم طريقتي التنبؤ بالسلاسل الزمنية الموسمية (بوكس - جنكنز)

وطريقة فوريير .

٣-١ طريقة بوكس جينكنز Box-Jenkins method (٦)(٤)

تعتبر طريقة بوكس- جينكنز من الطرائق المهمة والشائعة الاستخدام في تحليل السلاسل

الزمنية تحت مجال الزمن ($under\ time\ domain$) وهي أسلوب كفوء في الحصول على

التنبؤات المستقبلية للظواهر الم- دروسة في حالة تطبيقها بشكل صحيح وتمتاز هذه الطريقة بأنها

قادرة على معالجة السلاسل الزمنية التي تكون بياناتها توليفة من تأثيرات الاتجاه العام والدورية

والموسمية والعرضية وسوف نتناول في دراستنا احد نماذج بوكس- جينكنز وهو أنموذج الانحدار

الذاتي من الدرجة الثانية ($AR(2)$).

انموذج الانحدار الذاتي Autoregressive model (٧)(٦)(٥)

ان نماذج الانحدار الذاتي AR هي نماذج عشوائية تتحقق عندما توجد مجموعة من

الأوزان الغير مساوية للصفر حيث يمكن وصف النموذج بالصيغة التالية :

$$y_t = f_1 y_{t-1} + f_2 y_{t-2} + \dots + f_p y_{t-p} + a_t \dots \dots \dots (1-5)$$

حيث ان:

y_t هو المتغير المعتمد (الاستجابة) Response variable

$y_{t-1}, y_{t-2}, \dots, y_{t-p}$ المتغيرات التوضيحية

f_1, f_2, \dots, f_p معلمات الأنموذج وهي أوزان Weightes

a_t مركبة الخطأ العشوائي للنموذج تسمى الضوضاء البيضاء *white noise*
تعريف: إذا كانت العملية a_t تتوزع حسب التوزيع الطبيعي بمتوسط صفر وتباين محدد S_a^2
وكانت دالة الارتباط الذاتي (ACF) تحقق الفروض التالية :

$$r_0 = 1$$

$$r_k = r_{-k}$$

$$|r_k| < 1, \text{ for } k \neq 0$$

$$r_k \rightarrow 0, \text{ as } k \rightarrow \infty$$

فإن العملية a_t تدعى عملية الضوضاء البيضاء وهي السلاسل الزمنية ذات المتغيرات المتماثلة
ومستقلة وتتوزع طبيعياً (I.I.N.D)

يمكن وصف أنموذج الانحدار الذاتي من الدرجة P $AR(P)$ بالشكل التالي

$$f_p(b)y_t = a_t$$

$$\text{wher } f_p = (1 - f_1 b - f_2 b^2 - \dots - f_p b^p)$$

$$\text{and } b = y_{t-1}, b^2 = y_{t-2}, \dots, b^p = y_{t-p}$$

إن عدد المعلمات p هي التي تحدد درجة النموذج فإذا كانت $p=2$ فإن النموذج هو من الدرجة
الثانية $AR(2)$ ويكتب بالشكل التالي

$$y_t = f_1 y_{t-1} + f_2 y_{t-2} + a_t \quad \dots \dots \dots (1-6)$$

وان أنموذج $AR(2)$ الذي يعالج الموسمية هو

$$y_t = f_L y_{t-L} + f_{L+1} y_{t-L+1} + a_t$$

حيث إن L هو طول الموسم ، فإذا كانت السلسلة الزمنية شهرية فإن $L=12$

ويكون النموذج التنبؤي الموسمي حسب الصيغة التالية

$$\hat{y}_t = \hat{f}_L y_{t-L} + \hat{f}_{L+1} y_{t-L+1} \quad \dots \dots \dots (1-7)$$

حيث يتم تقدير معلمات النموذج باستخدام طريقة المربعات المشدبة (LTS)

٤-١ طريقة فوريير Fourier Method (١)(٧)(٤)

ان طريقة فوريير هي طريقة تنبؤية بالسلاسل الزمنية تعمل تحت مجال الترددات (under
frequency domain) وتعتمد بشكل أساسي على تحليل فوريير (Fourier
Analysis)

حيث يتم تحويل السلسلة الزمنية من مجال الزمن الى مجال الترددات بالاعتماد على مبدأ
فوريير ويتم وصف السلسلة الزمنية كتركيبية خطية من الدوال المثلثية (sine , cosine) والتي
تكون دورية (Periodic) ومتوافقة (Harmonic) ومتعامدة (Orthogonal) وفق النظام التالي :

$$y_t = \sum_{k=0}^{[n/2]} [a_k \text{Sin}(2pkt/n) + b_k \text{Cos}(2pkt/n)] , t = 1, 2, \dots, n \quad \dots \dots \dots (1-8)$$

ان المعادلة أعلاه تدعى متسلسلة فوريير في المتتابعة y_t ، وان a_k, b_k هي معاملات فوريير،
وباستخدام خاصية التعامد (orthogonality) للدوال المثلثية يمكن إيجاد المعاملات a_k, b_k ، وذلك

بضرب طرفي المعادلة (1-8) بالمقدارين $(\sin \frac{2pkt}{n}, \cos \frac{2pkt}{n})$ بالتعاقب ثم الجمع لكل t حيث $t = 1, 2, \dots, n$ ومن ثم نحصل على :

$$a_k = \begin{cases} \frac{1}{n} \sum_{t=1}^n y_t \cos(w_k) & , k = 0 \text{ and } k = \frac{n}{2} \text{ if } n \text{ is even} \\ \frac{2}{n} \sum_{t=1}^n y_t \cos(w_k) & , k = 1, 2, \dots, \frac{n-1}{2} \end{cases} \dots\dots(1-9)$$

$$b_k = \frac{2}{n} \sum_{t=1}^n y_t \sin(w_k) \quad , k = 1, 2, \dots, \frac{n-1}{2}$$

$$w_k = \frac{2pkt}{n} \quad , k = 0, 1, \dots, \left[\frac{n}{2} \right]$$

ان w تمثل ترددات الموجة الجيبية وتعتمد قيمتها على طبيعة المشاهدات فإذا كانت الدورة موسمية سنوية بمشاهدات شهرية فإن $w = \frac{2p}{12}$ ، أما إذا كانت المشاهدات فصلية فإن $w = \frac{2p}{4}$ وتكون قيمتها $w = \frac{2p}{52}$ إذا كانت البيانات اسبوعية .

فعلى فرض أن لدينا سلسلة زمنية بمشاهدات شهرية $(s_1, s_2, \dots, s_{12})$ فإن تحليل فورير يعرض المركبة الموسمية S للموسم j في هذه السلسلة للنقاط الاثنتي عشرة التي تقع جميعها ضمن منحنى واحد بالشكل الآتي:

$$S_j = 1 + \sum_{k=1}^6 (a_k \cos \frac{2pkj}{6} + b_k \sin \frac{2pkj}{6}) \dots\dots\dots(1-10)$$

والمعادلة (1-10) في أعلاه هي معادلة منحنى انحدار بمعاملات انحدار a_k, b_k أي ان :

$$S_j = 1 + a_1 \cos \frac{2pj}{6} + b_1 \sin \frac{2pj}{6} + a_2 \cos \frac{2pj}{6} + b_2 \sin \frac{2pj}{6} \dots\dots\dots (1-11)$$

ان النموذج في معادلة أعلاه هو أنموذج انحدار خطي بالمعاملات a_1, b_1, a_2, b_2 والذي يمكن ان يكتب بالصيغة التالية :

$$y_t = B_1 + B_2 t + B_3 \sin \frac{2pt}{c} + B_4 \cos \frac{2pt}{c} + e_t$$

حيث ان C هو طول الموسم وان النموذج التنبؤي هو

$$\hat{y}_{t+c} = \hat{b}_1 + \hat{b}_2(t) + \hat{b}_3 \sin \frac{2pt}{c} + \hat{b}_4 \cos \frac{2pt}{c} \dots\dots\dots (1-12)$$

حيث يتم تقدير معاملات النموذج كذلك باستخدام طريقة المربعات المشدبة (LTS)

٥-١ قياس دقة التنبؤ measuring forecast accuracy (١)(٨)

يقصد بمفهوم قياس دقة التنبؤ هو النموذج التنبؤي على توليد مشاهدات تتبع السلوك نفسه للسلسلة الزمنية . ان اهم ما يعتمد عليه في هذه المرحلة هو خطأ التنبؤ المفرد (*one – step forecast error*) إذ إن الدقة (*accuracy*) تعني بها الملاءمة الجيدة أو حسن المطابقة (*goodness of fit*) فإذا كان لدينا n من التنبؤات فسوف يكون لدينا n من الأخطاء المفردة وان اغلب مقاييس دقة التنبؤات تعتمد على أخطاء التنبؤ ، وسوف نتناول في دراستنا المقاييس التالية:

1- معيار متوسط مربعات الخطأ (*MSE*) إذ إن

$$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2 \quad \dots\dots\dots (1-13)$$

2- معيار متوسط نسبة الخطأ (*PE*) إذ إن

$$MpE = \sum_{t=1}^n \frac{(y_t - \hat{y}_t)}{y_t} \times 100\% \quad \dots\dots\dots (1-14)$$

3- واحصاء U (مقياس ثيل) (*U statistic*)

$$U = \sqrt{\frac{(\sum_{t=1}^{n-1} FPE_{t+1} - APE_{t+1})^2}{\sum_{t=1}^{n-1} (APE_{t+1})^2}} \quad \dots\dots\dots (1-15)$$

حيث أن

$$FPE_{t+1} = \frac{\hat{y}_{t+1} - y_t}{y_t}$$

$$APE_{t+1} = \frac{y_{t+1} - y_t}{y_t}$$

٢- الجانب التطبيقي

في هذا الجانب تم استخدام اسلوب المحاكاة لتوليد المشاهدات الخاصة بالدراسة وذلك لدراسة كفاءة طريقتي بوكس-جينكز وطريقة فورير في التنبؤ بالسلاسل الزمنية الموسمية في ظل وجود قيم شاذة حيث تمت المقارنة بين الطريقتين عن طريق مقاييس دقة التنبؤات المستخدمة في الدراسة (*U* , *MSE* , *MPE* , *statistic*).

٢-١ وصف تجربة المحاكاة

تم بناء تجربة المحاكاة بالشكل التالي:

- ١- توليد مشاهدات الخطأ العشوائي حسب التوزيع الطبيعي المعياري
- ٢- توليد سلاسل زمنية موسمية ببيانات شهرية (طول الموسم = ١٢) وحسب النماذج التالية

أ- النموذج الموسمي الثابت بالصيغة التالية

$$y_t = b_0 + S_t + e_t \quad t = 1, 2, \dots, n \quad \dots \dots \dots (2-1)$$

حيث إن

b_0 هي مركبة الحد الثابت

S_t هي مركبة الموسمية (Seasonality)

ب- النموذج الموسمي الاتجاهي بالصيغة التالية

$$y_t = (b_0 + b_1 t) S_t + e_t \quad , t = 1, 2, \dots, n \quad \dots \dots \dots (2-2)$$

حيث إن

b_1 هي مركبة الاتجاه العام (Trend)

٣- توليد بيانات السلاسل الزمنية لكل انموذج باستخدام احجام عينات مختلفة (n=20, 40, 60, 80, 100, 125, 150, 200)

و بنسب تلوث (0%، ٥٠%، ١٠٠%)

٤- تكرار تجربة المحاكاة ٥٠٠ مرة لضمان العشوائية

٥- تقدير معلمات الطريقتين باستخدام طريقة LTS وعند كل حجم عينة .

٦- تطبيق النماذج التنبؤية للطريقتين الموصوفة في المعادلات (٣) و (١٧)

٧- قياس دقة التنبؤات للطريقتين حسب المقاييس الموصوفة في المعادلات (١٣-١) و (١٤-١)

(١-١٥)

٨- اجراء المقارنات بين الطريقتين على ضوء قيم مقاييس دقة التنبؤات

٢-٢ نتائج تجربة المحاكاة

أولاً- تم تطبيق طريقتي الدراسة على النموذج الموسمي الثابت تم حساب مقاييس المقارنة لكل طريقة عند مجموعة من أحجام العينات المختلفة، والجدول التالية تبين نتائج التجربة.

جدول رقم (١)

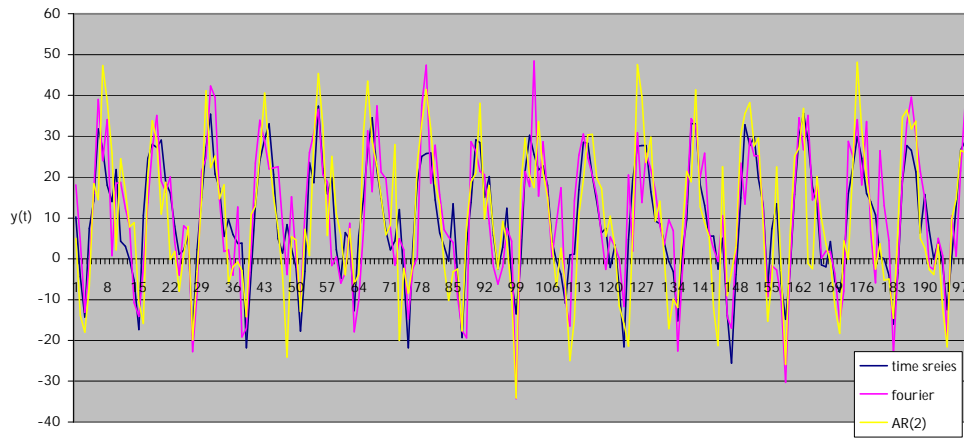
يوضح قيم معايير المقارنة لطريقتي فورير و بوكس-جينكز عند أحجام عينات مختلفة و عند نسبة تلوث (0%) لسلاسل زمنية ذي نموذج موسمي ثابت

نسبة التلوث	n	Box-Jenkins			Fourier		
		MSE	U-stst.	MPE	MSE	U-stst.	MPE
0 %	٢٠	1٥.508	4.572	1.58	1٧.864	6.742	1.٧12
	٤٠	13.٣58	4.146	1.212	1٣.068	4.344	1.52
	٨٠	11.٨48	3.852	1.022	١١.364	٤.174	١.878
	١٠٠	9.49	3.546	0.824	٨.٩16	2.936	0.722
	١٥٠	7.052	3.42	0.592	4.932	2.68	0.53
	٢٠٠	5.892	3.398	0.4942	4.396	2.572	0.4562

من النتائج المبينة في جدول رقم (١) وعند نسبة تلوث (0%) نلاحظ ان قيم معايير المقارنة وهي (متوسط مربعات الخطأ العشوائي واحصاءة U ومتوسط نسبة الخطأ المطلق) تنخفض قيمها مع زيادة حجم العينة ولكلا الطريقتين وهذا يدل على تحسن أداء الطريقتين مع زيادة حجوم العينات وحسب هذه المقاييس نلاحظ أن طريقة بوكس-جينكز هي أفضل من طريقة فورير وعند حجم

العينة (٢٠) لكن عند بقيت أحجام العينات كانت طريقة فورير أفضل من طريقة بوكس-جينكز (على الرغم من أن قيم معايير المقارنة كانت متقاربة نوعاً ما بالنسبة لطريقتي التنبؤ) وكما هو واضح من خلال قيم معايير المقارنة الثلاث، إذ تدل هذه النتائج على أن طريقة بوكس-جينكز تكون أكثر كفاءة من طريقة فورير عند أحجام العينات الصغيرة جداً، وبالعكس فإن طريقة فورير تكون أكثر كفاءة من طريقة بوكس-جينكز عند أحجام العينات الكبيرة (لنموذج موسمي ثابت). إذ نجد مثلاً أن قيمة MSE لطريقة بوكس-جينكز هي (15.508) عند الحجم $n=20$ بينما أصبحت قيمتها (5.892) عند الحجم $n=200$ بينما كانت قيمة MSE لطريقة فورير هي (١٧,٨٦٤) عند الحجم $n=20$ بينما أصبحت قيمتها (4.396) عند الحجم $n=200$. وكما هو مبين في الشكل البياني رقم (١)

شكل رقم (١) يمثل امنى التنبؤات لطريقة فورير و طريقة AR(2) لبيانات موسمية للموذج الثابت عند نسبة تلوث ٥%



جدول رقم (٢)

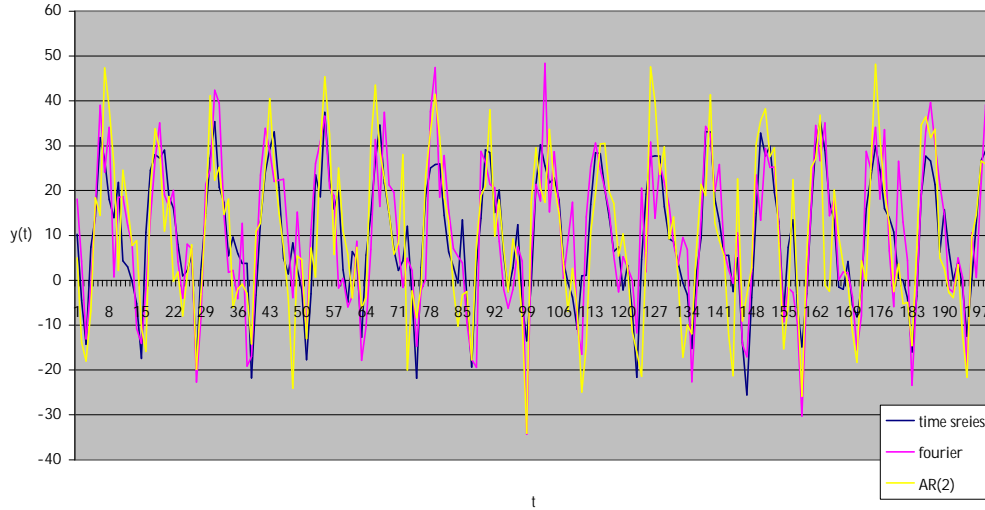
يوضح قيم معايير المقارنة لطريقتي فورير و بوكس-جينكز عند أحجام عينات مختلفة وعند نسبة تلوث (5%) لسلاسل زمنية ذي نموذج موسمي ثابت

نسبة التلوث	n	Box-Jenkins			Fourier		
		MSE	U-stst.	MPE	MSE	U-stst.	MPE
5 %	٢٠	34.241	11.477	5.231	٣٣.456	١٥.534	٩.656
	٤٠	27.112	11.001	6.213	٢٥.323	١١.654	٩.087
	٨٠	22.978	8.929	4.067	١٩.٩٩٧	٨.231	٣.345
	١٠٠	21.102	8.523	3.893	١٩.674	٦.046	٣.٩34
	١٥٠	15.317	6.218	3.684	١١.785	٣.023	١.764
	٢٠٠	12.666	5.214	1.921	٧.326	١.678	١.١24

من النتائج المبينة في جدول رقم (٢) وعند نسبة تلوث (5%) نلاحظ أن قيم معايير المقارنة قد ازدادت قياساً عما كانت عليه عند عدم وجود بيانات شاذة فهذا يدل على أن الطريقتين تحسنت للقيم الشاذة لكن بشكل متفاوت . نجد أن طريقة فورير أفضل من طريقة بوكس-جينكز عند حجوم العينات كافة هذا يدل على أن طريقة فورير هي أقل تحسناً بالقيم الشاذة قياساً بطريقة

بوكس-جينكز، وكذلك نجد أن هناك انخفاض معايير المقارنة لكلا الطريقتين مع زيادة حجم العينة مما يدل على أن تأثير القيم الشاذة (في السلاسل الزمنية) ينحسر تدريجياً عند ازدياد حجومات العينات. وكما هو واضح في الشكل البياني رقم (٢).

شكل رقم (٢) يمثل مبدئي تنبؤات طريقة فورير وطريقة AR(2) لسلاسل زمنية موسمية للنموذج الثابت عند نسبة تلوث ٥%



جدول رقم (٣)

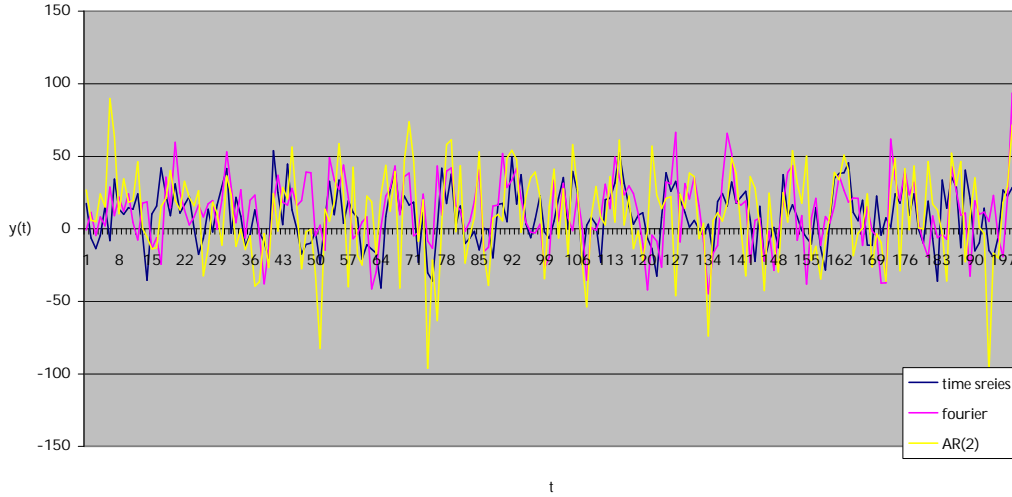
يوضح قيم معايير المقارنة لطريقتي فورير و بوكس-جينكز عند احجام عينات مختلفة وعند نسبة تلوث (10%) لسلاسل زمنية ذي نموذج موسمي ثابت

نسبة التلوث	n	Box-Jenkins			Fourier		
		MSE	U-stst.	MPE	MSE	U-stst.	MPE
10 %	٢٠	٤٦.231	1٧.789	6.342	٤٥.453	1٥.٩76	6.056
	٤٠	٤٢.678	14.453	8.267	32.653	12.879	6.627
	٨٠	37.905	14.342	3.237	26.675	11.542	4.378
	١٠٠	26.343	12.267	3.964	20.346	9.357	3.891
	١٥٠	22.657	9.678	2.096	17.987	5.798	3.356
	٢٠٠	18.672	5.987	1.867	14.435	3.478	1.903

من النتائج المبينة في جدول رقم (٣) وعند نسبة تلوث (10%) نلاحظ ان قيم معايير المقارنة قد ازدادت قياساً عما كانت عليه في نتائج الجدولين السابقين. وهذا يفسر ان التغير الذي حدث على طريقتي التنبؤ كان نتيجة لزيادة نسبة البيانات الشاذة، فهذا يدل على ان الطريقتين تحسستا للقيم الشاذة ولكن بمقادير متفاوتة، وكذلك نجد ان تحسس طريقة فورير بالقيم الشاذة اقل من طريق بوكس-جينكز عند حجومات العينات كافة، مما يدل هذا على ان طريقة فورير أكثر مقاومة للقيم الشاذة من طريقة بوكس-جينكز، وكذلك نجد أن كفاءة الطريقتين تزداد مع زيادة أحجام العينات، في ظل وجود نسبة التلوث بالقيم الشاذة. وذلك واضح من خلال انخفاض قيم معايير

المقارنة ولكلا الطريقتين ، مما يدل على أن تأثير القيم الشاذة ينحسر تدريجياً عند ازدياد حجم العينات . وكما هو مبين في الشكل البياني رقم (٣)

شكل رقم (٣) يمثل منحني تنبؤات طريقة فورير وطريقة AR(2) لبيانات موسمية للنموذج الثابته عند نسبة تلوث ٠%



من خلال الشكل رقم (١) والشكل رقم (٢) والشكل رقم (٣) نجد ان هناك سلوكاً موسمياً واضحاً للسلسلة الزمنية ولكنه أفقي بدون اتجاه وان كلا الطريقتين قد تنبأت بشكل جيد ببيانات السلسلة الزمنية ولكن منحني طريقة فورير كان أكثر مطابقة لبيانات السلسلة المدروسة حتى عند وجود النسب المختلفة من البيانات الشاذة ، مما يدل أن طريقة فورير أكثر تكيف من طريقة بوكس-جينكز.

ثانياً- تطبيق طريقتي الدراسة على النموذج الموسمي لاتجاهي الموصوف بالصيغة (1-2) وتم حساب مقاييس المقارنة لكل طريقة وعند أحجام عينات مختلفة، والنتائج كما هي موضحة في الجدول أدناه.

جدول رقم (٤)

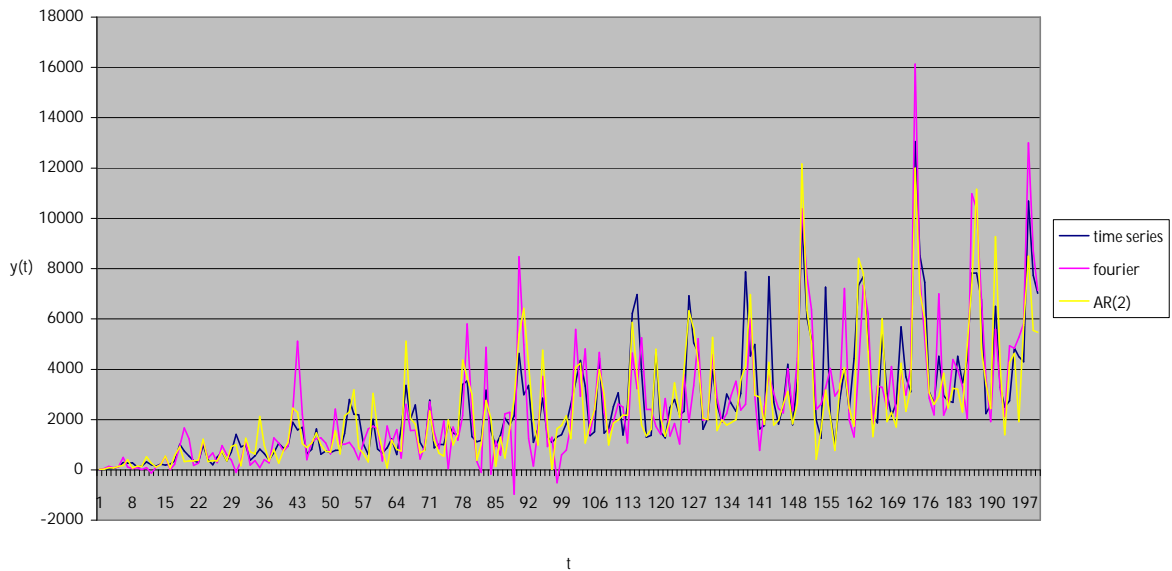
يوضح قيم معايير المقارنة لطريقتي فورير و بوكس-جينكز عند أحجام عينات مختلفة وعند نسبة تلوث (0 %) لسلاسل زمنية ذي نموذج موسمي اتجاهي

نسبة التلوث	n	Box-Jenkins			Fourier		
		MSE	U-stst.	MPE	MSE	U-stst.	MPE
0 %	٢٠	228.824	43.816	26.234	320.284	22.532	22.046
	٤٠	207.725	42.752	21.012	298.031	42.082	25.152
	٨٠	137.497	28.408	13.66	265.972	40.514	24.515
	١٠٠	76.812	18.012	8.758	160.052	28.494	19.784
	١٥٠	29.190	14.134	4.118	77.918	19.828	18.458
	٢٠٠	14.310	9.122	3.684	٥٤.328	13.652	10.488

من خلال النتائج المبينة في الجدول أعلاه والخاصة بالنموذج الموسمي الاتجاهي وعند نسبة تلوث (0%) نلاحظ ان قيم معايير المقارنة وهي (متوسط مربعات الخطأ العشوائي واحصاءة U ومتوسط نسبة الخطأ المطلق) تنخفض قيمها مع زيادة حجم العينة ولكلا الطريقتين وهذا يدل على تحسن أداء الطريقتين مع زيادة حجم العينات وحسب هذه المقاييس نلاحظ أن طريقة بوكس-جينكز هي أفضل من طريقة فورير عند كافة أحجام العينات ، مما يدل على أن

طريقة بوكس- جينكز هي طريقة التنبؤ الأفضل بنسبة للنموذج الموسمي الاتجاهي قياساً بطريقة فورير، وهذه النتيجة واضحة من خلال قيم معايير المقارنة المستخدمة، حيث نجد ان قيمة (MSE) الخاص بطريقة بوكس-جينكز عند حجم العينة (٢٠) هي (٢٢٨,٨٢٤) بينما كانت قيمة (MSE) الخاص بطريقة فورير عند نفس الحجم هي (٣٢٠,٢٨٤) مما يدل على أن طريقة بوكس- جينكز أكثر قدرة تنبؤية قياساً بطريقة فورير (عند النموذج الموسمي الاتجاهي)، وكذلك نجد ان قيمة (MSE) الخاص بطريقة بوكس-جينكز عند حجم العينة (٢٠٠) هي (١٤,٣١٠) بينما كانت قيمة (MSE) الخاصة بطريقة فورير عند نفس الحجم هي (٥٤,٣٢٨) مما يدل على ان طريقة بوكس-جينكز بقيت أكثر قدرة تنبؤية من طريقة فورير حتى عند حجوم العينات الكبيرة، لكن نجد ان هناك تحسناً ملحوظاً للطريقتين عند حجوم العينات الكبيرة وهذا واضح من خلال انخفاض قيم معايير المقارنة الثلاث ولكلا الطريقتين. وكما موضح في الشكل رقم (٤)

شكل رقم (٤) يوضح منحني تنبؤاه لطريقة فورير وطريقة AR2 لنموذج الموسمي الاتجاهي عند نسبة تلوه ٥%



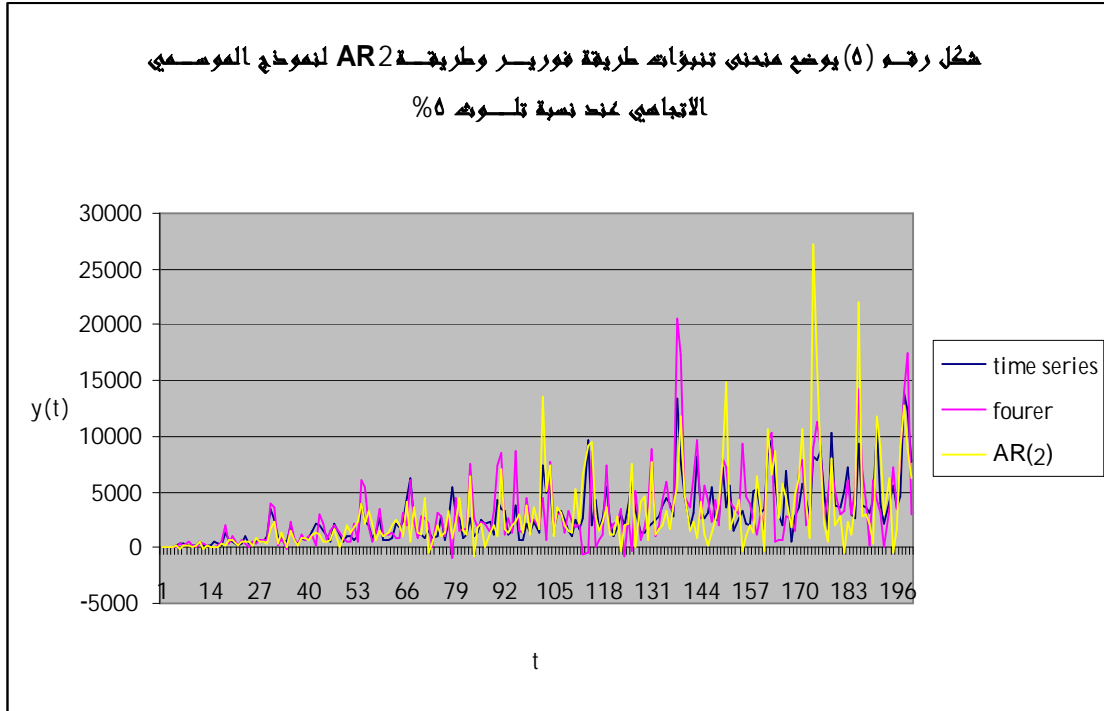
جدول رقم (٥)

يوضح قيم معايير المقارنة لطريقتي فورير و بوكس-جينكز عند أحجام عينات مختلفة وعند نسبة تلوث (5 %) لسلاسل زمنية ذي نموذج موسمي اتجاهي

نسبة التلوث	n	Box-Jenkins			Fourier		
		MSE	U-stst.	MPE	MSE	U-stst.	MPE
5 %	٢٠	423.768	86.654	26.543	612.945	150.345	87.076
	٤٠	512.456	87.823	21.675	583.342	111.094	66.763
	٨٠	372.423	77.545	13.456	424.235	102.124	40.785
	١٠٠	306.976	40.007	8.945	450.095	78.457	29.324
	١٥٠	220.034	39.456	4.123	205.213	37.177	3.211
	٢٠٠	162.003	22.534	3.457	١٥٤.457	20.653	2.934

من خلال النتائج المعروضة في الجدول أعلاه والخاصة بالنموذج الموسمي الاتجاهي وعند نسبة تلوث (5%)، نلاحظ ان قيم معايير المقارنة قد ازدادت قياساً عن ما كانت عليه عند عدم وجود بيانات شاذة فهذا يدل على ان الطريقتين تحسنت للقيم الشاذة لكن بدرجات

متفاوتة حيث نجد ان طريقة بوكس-جينكز هي الأقل تحسس للقيم الشاذة وذلك عند حجم العينات (١٠٠، ٨٠، ٤٠، ٢٠)، لكن عند حجم العينات (٢٠٠، ١٥٠) نفاجا ان طريقة فورير أصبحت أفضل من طريقة بوكس-جينكز (عند النموذج الموسمي الاتجاهي) قياساً بنتائج الجدول السابق ، كذلك نلاحظ ان قيم معايير المقارنة تنخفض قيمها مع زيادة حجم العينة ولكلا الطريقتين وهذا يدل على تحسن أداء الطريقتين مع زيادة حجم العينات من جهة ، ومن جهة أخرى نجد ان تأثير القيم الشاذة يتلاشى تدريجياً عند زيادة حجم العينة ، وهذا واضح من خلال قيم معايير المقارنة الثلاث ، وكما هو مبين في الجدول أدناه.



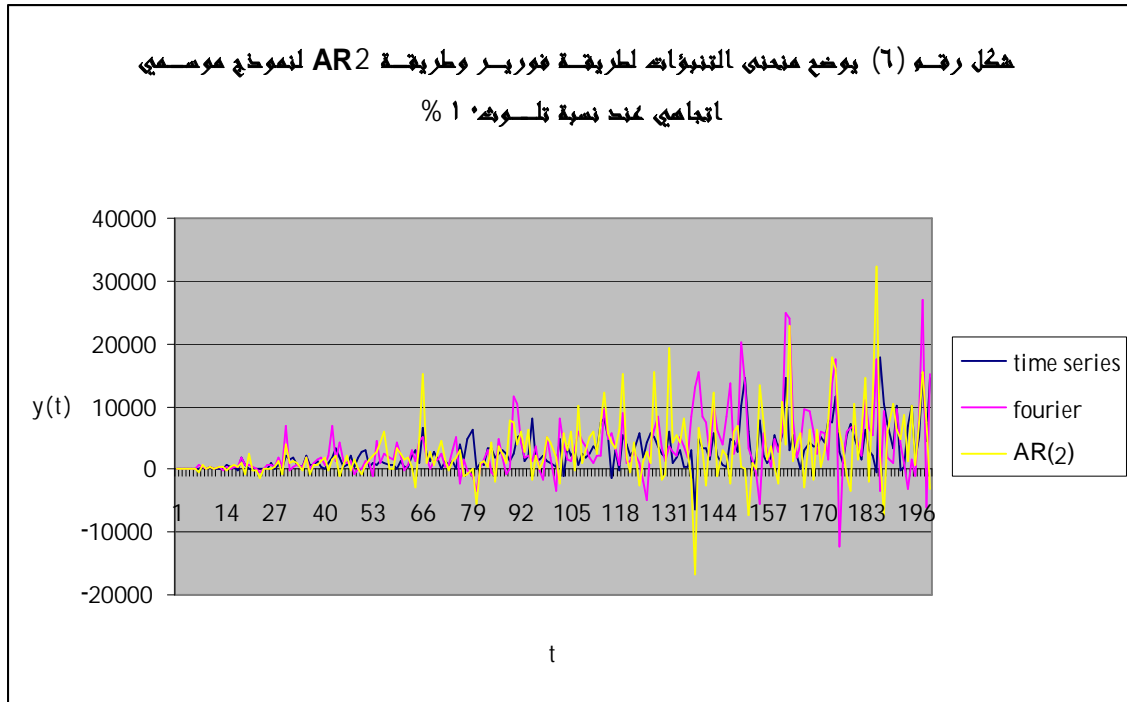
جدول رقم (٦)

يوضح قيم معايير المقارنة لطريقتي فورير و بوكس-جينكز عند احجام عينات مختلفة وعند نسبة تلوث (10%) لسلاسل زمنية ذات نموذج موسمي اتجاهي.

نسبة التلوث	n	Box-Jenkins			Fourier		
		MSE	U-stst.	MPE	MSE	U-stst.	MPE
10 %	٢٠	١٠٦٧.453	127.454	102.235	919.28	144.532	113.046
	٤٠	٩٦٧.453	101.665	77.012	900.03	99.082	70.152
	٨٠	٨٨٨.342	83.445	59.454	798.97	72.514	40.515
	١٠٠	٦٤٠.653	42.644	32.233	520.05	38.49	29.784
	١٥٠	511.077	39.324	27.118	312.91	32.828	23.458
	٢٠٠	386.652	31.443	23.767	٢٨٨.32	21.65	19.488

من خلال النتائج المعروضة في الجدول أعلاه والخاصة بالنموذج الموسمي الاتجاهي وعند نسبة تلوث (10%) ، نلاحظ ان قيم معايير المقارنة قد ازدادت قياساً عما كانت عليه في نتائج الجدولين السابقين وهذا يفسر ان التغيير الذي طرأ على قيم معايير المقارنة الثلاث كان نتيجة لزيادة نسبة البيانات الشاذة فهذا يدل على ان الطريقتين تحسنت للقيم الشاذة بصورة واضحة ، وكذلك نجد وعند نسبة (10%) من القيم الشاذة كانت طريقة فورير أفضل من طريقة بوكس-جينكز عند كافة أحجام العينات ، مما يدل على ان طريقة بوكس-جينكز تنهار

أمام النسب العالية من القيم الشاذة، لكن طريقة فورير تبدو أكثر تماسكاً أمام النسب العالية من القيم الشاذة (عند النموذج الموسمي الاتجاهي) وهذا واضح من خلال قيم معايير المقارنة. وكما هو واضح من خلال أشكال أدناه.



من خلال الأشكال رقم (٤) والشكل رقم (٦) والشكل رقم (٦) نلاحظ ان نمط السلسلة الزمنية هو موسمي متضاعف اتجاهي أي ان المركبة الموسمية تتضاعف بنسبة معينة مع الزمن مع وجود الاتجاه نحو الأعلى، وبشكل عام نلاحظ ان أداء الطريقتين يسوء مع زيادة نسبة القيم الشاذة، لكن نجد ان طريقة بوكس-جينكز أكثر تحسناً للقيم الشاذة من طريقة فورير حيث كان أداءها أفضل من أداء طريقة فورير عند عدم وجود قيم شاذة بينما كانت أداء طريقة فورير أفضل عندما تكون البيانات ملوثة

الاستنتاجات

- ١- بصورة عامة ولكلا النموذجين (النموذج موسمي ثابت، والنموذج موسمي اتجاهي) نجد ان لطريقتي التنبؤ بوكس-جينكز وطريقة فورير يتأثرا بصورة واضحة بالقيم الشاذة.
- ٢- بصورة عامة ولكلا النموذجين (النموذج موسمي ثابت، والنموذج موسمي اتجاهي) نجد أن أداء الطريقتين يتحسن كلما ازدادت حجوم العينات، هذا يعني ان هناك علاقة طردية ما بين تحسن أداء الطريقتين وحجوم العينات المستخدمة، وهذا واضح حتى عند وجود نسبة من البيانات الشاذة.
- ٣- في النموذج الموسمي الثابت نجد أن طريقة بوكس-جينكز أفضل من طريقة فورير عند أحجام العينات الصغيرة جداً، ونجد عكس هذا الشيء عند حجوم العينات الكبير حيث تكون طريقة فورير أكثر كفاء من طريقة بوكس-جينكز.
- ٤- نجد ان طريقة فورير طريقة تنبؤية جيدة في ظل وجود القيم الشاذة، وهذا واضح من انخفاض قيم معايير المقارنة الخاصة بطريقة فورير قياساً بطريقة بوكس-جينكز مع ازدياد نسبة القيم الشاذة مما يدل على ان طريقة فورير أكثر مقاومة للقيم الشاذة (عند النموذج الموسمي الثابت)

- ٥- بصورة عامة ولكلا النموذجين (النموذج موسمي ثابت، والنموذج موسمي اتجاهي) نجد ان تأثير القيم الشاذة يتلاشى تدريجياً مع زيادة حجوم العينات .
- ٦- نجد أن طريقة بوكس-جينكز طريقة تنبؤية جيدة للنموذج الموسمي الاتجاهي (وذلك في حالة عدم وجود قيم شاذة) قياساً بطريقة فورير لكن سرعان ما تتحول هذه النتيجة مع ظهور نسب مختلفة من القيم الشاذة .
- ٧- في النموذج الموسمي الاتجاهي وعند زيادة نسبة البيانات الشاذة نجد ان طريقة بوكس-جينكز يضعف أدائها قياساً بطريقة فورير. فنستنتج ان طريقة فورير أكثر ملاءمة للتنبؤ عند وجود قيم شاذة.

التوصيات

- ١- تحسين نموذج فورير المستخدم في التنبؤ بالسلاسل الزمنية الموسمية المتضاعفة وذلك باستخدام نماذج أعلى رتبة، بإضافة أزواج (sin , cos) جديدة الى النموذج .
- ٢- في حالة النموذج الموسمي الثابت لابد من استخدام طريقة بوكس-جينكز عند أحجام العينات الصغيرة واستخدام طريقة فورير عند أحجام العينات الكبير .
- ٣- قد يتسبب معالجة البيانات الشاذة (الموجودة في السلاسل الزمنية) خسارة كبيرة في كمية المعلومات، لذا لابد من استخدام طريقة فورير للتنبؤ عند وجود نسبة من البيانات الشاذة. لان هذه الطريقة وحسب النتائج المبينة أعلاه تكون اقل تحسناً للقيم الشاذة.

المصادر

- ١- عبد الحسين محمد (٢٠٠٨) "استخدام طريقة المربعات الصغرى المتناقصة في تقدير معلمات أنموذج فورير للسلاسل الزمنية" رسالة ماجستير مقدمة إلى قسم الإحصاء، كلية الإدارة والاقتصاد، جامعة المستنصرية.
- 2- Habshah Midi and Hassan S. Uraibi and Bashar A. Talib(2009) " Dynamic Robust Bootstrap Method Based on LTS Estimators" European Journal of Scientific Research , ISSN 1450-216X Vol.32 No.3 pp.277-287.
- 3- Huber M,Rousseeuw P.j (1996) "Robust regression with both categorical co variables" Springer-Verlag, New Yourk.
- 4-Jonathan D.Cryer and Kung-sik Chan(2008) "time series Analysis with applications in R" Second Edition University of Iowa .USA.
- 5-Montgomery, D.C and Johanson, L.A. (1976) , "Forecasting and time series analysis ",Mc Gram Hill, New York.
- 6-Montgomery, D.C. (1990), "Forecasting and Time Series Analysis", McGraw-Hil Inc,Singapore.
- 7- Richard Harris and Roles Solli(s2003)" Applied Time Series Forecasting" University of Durham. Modeling and
- 8-Robert A.Yaffee and Monnie McGee (2004) "Introduction to time series Analysis and forecasting with application of sAs and Spss" Academic press, INC, INC University of New York.
- 9- Rousseeuw P. J. and Van Driessen K., 1999. "Computing LTS regression for Large Data Sets",Data Mining Knowledge Discovery 12, pp.29-45.