الخوارزمية الديناميكية (DRBLTS) و الموزونة احتماليا (WBP) لتقدير معلمات الانحدار الحصين باستخدام تقنية الـ Bootstrap (دراسة مقارنة)

م. م. طه حسين علي
كلية الإدارة والاقتصاد- جامعة القادسية
قسم الإحصاء

المستخلص

الـ Bootstrap واحدة من أهم تقنيات إعادة المعاينة التي جذبت انتباه الباحثين مؤخرا، وجود القيم الشاذة في مجموعة البيانات الأصلية هي المشكلة الأساسية في عينات Bootstrap ، خصوصا عندما يتجاوز عددها نسبة القيم الشاذة في العينة الأصلية . العديد من الطرق اقترحت لحل هذه المشكلة مثل DRBLTS و WBP .

هذا البحث هو محاولة لبيان دقة المعالم المقدرة بمقارنة نتائج كلا الطريقتين. bias و MSE وكذلك RMSE اعتمدت كمعايير للمقارنة بين هذه المقدرات، معيار الدقة يعتمد على RMSE ، فالطريقة الأفضل هي الطريقة التي تزودنا بأقل قيمة لجذر متوسط مربعات الخطأ.
لهذا الغرض الدراسة استخدمت بيانات حقيقية، النتائج بينت ان مقدرات DRBLTS أكثر دقة من مقدرات الطريقة الثانية.

Abstract

Bootstrap is one of an important re-sampling technique which has given the attention of researches recently. The presence of outliers in the original data set may cause serious problem to the classical bootstrap when the percentage of outliers are higher than the original one. Many methods are proposed to overcome this problem such Dynamic Robust Bootstrap for LTS (DRBLTS) and Weighted Bootstrap with Probability (WBP). This paper try to show the accuracy of parameters estimation by comparison the results of both methods. The bias , MSE and RMSE are considered. The criterion of the accuracy is based on the RMSE value since the method that provide us RMSE value smaller than other is consider better. For this purpose the study used real data. The results show the (DRBLTS) estimators are more accuracy than other.

١. المقدمة

تعد طريقة المربعات الصغرى LS من أفضل الطرق لتقدير معادلة خط الانحدار ، إذ أنها أغنت البحث العلمي لسنوات عديدة لما تتمتع به من خصائص ومزايا جيدة لتحقيق فرضيات الانحدار الخطي ولأنها وكما هو معروف الطريقة الأمثل في تصغير مجموع مربعات الانحرافات بين القيم المشاهدة والقيم المتوقعة فهي تعطي أفضل مقدرات خطية غير متحيزة (BLUE)، لكن من المعروف ان دقة التقديرات يتأثر بحجم العينة فكلما كان حجم العينة اكبر كانت التقديرات أدق والعكس صحيح أي أن هنالك علاقة طردية بين حجم العينة ودقة التقديرات وهذا أمر منطقي جدا، اقترح (Efron 1979)[3] تقنية Bootstrap لإعادة المعاينة وهي طريقة تكرارية استخدمت لمعالجة دقة التقديرات في العينات الصغيرة حيث تقوم هذه التقنية على مبدأ إيجاد تقديرات غير متحيزة من مجموعة من التقديرات المتحيزة وذلك بتوليد مجموعة كبيرة من العينات المسحوبة بشكل عشوائي من نفس بيانات العينة مع الإرجاع وبنفس حجم العينة الأصلي.

لقد حاول الباحث Efron بهذه الطريقة حل مشكلة صغر حجم العينة بواسطة إعادة المعاينة وتوليد عدد كبير من العينات ، لكن هنالك سؤال يطرح هو ماذا لو كانت هذه البيانات لا تحقق فروض المربعات الصغرى، مثلا خرق فرضية التوزيع الطبيعي للأخطاء والمتأتية من وجود قيمة او مجموعة قيم شاذة **Outlier** ضمن البيانات مما ينتج عنه مقدرات لا تحقق خصائص مقدرات المربعات الصغرى، ويشير إلى ذلك[4] **Huber (1981)** حيث يقول "حتى القيمة الشاذة الواحدة قد تهدم المزايا الجيدة لمقدرات المربعات الصغرى"، إذ تعرف القيمة الشاذة بأنها القيمة التي تنحرف بشكل ملحوظ عن بقية المشاهدات والتي قد تنشا من توزيعات ثقيلة الذيل Heavy Tailed Distribution او مشاهدة رديئة ناتجة من أخطاء في عملية إدخال البيانات.

وقد صنف كل من [12] (Rousseeuw & Zomeren 1990) القيم الشاذة إلى ثلاثة أصناف تبعا لموقعها وتأثيرها وهي كما يأتي:-

١- مشاهدات مؤثرة Influence observation وهي التي يؤثر وجودها على اتجاه خط الانحدار. وقد يظهر بعض من هذه المشاهدات ضمن قيم المتغير المعتمد.
٢- مشاهدات متناقصة Disc repancy observation تظهر هذه المشاهدات ضمن المتغير المستقل وتبتعد بشكل واضح عن خط الانحدار وقد تسمى **Bad leverage points**.
٣- نقاط رافعة Leverage points وهي البيانات التي يمكن اعتبارها نقاط رافعة جيدة أي أنها تقع على امتداد خط الانحدار والتي عندها يختزل حجم حد الخطأ العشوائي وهذه النقطة تظهر بفعل قيمتين شاذتين الأولى ضمن المتغير المستقل والثانية ضمن المتغير المعتمد.

وعلية فان عملية استخدام طريقة المربعات الصغرى تكون طريقة غير كفوءة عندما تحتوي البيانات على قيمة شاذة حيث ان نقطة الانهيار Breakdown point مساو إلى(٠%) ، اذ تعرف نقطة الانهيار [1] (احمد ٢٠٠٥) بأنها الحد الذي يصف مقدار مقاومة المقدر للبيانات الملوثة قبل ان يصبح من دون فائدة. ويوصف المقدر بأنه مقدر مقاوم (Resistant Estimator) إذا كانت نقطة انهياره اكبر من الصفر.

لقد طرحت العديد من الطرق ونوقشت العديد من النظريات لمعالجة القيم الشاذة في عينات تقنية Bootstrap ، أننا في هذا البحث سوف نحاول تسليط الضوء على طريقتان ونقارن بينهما، الطريقة الأولى هي تقنية Bootstrap الموزونة احتماليا "WBP" (Weighted Bootstrap with Probability) والتي اقترحت من قبل [9] (Ramli 2008) والتي تهدف إلى الحصول على تقديرات حصينة اذ تقوم فكرتها على تخصيص احتمال صغير جدا للقيم الشاذة بحيث تكون فرصة ظهور هذه القيم في عينات Bootstrap يكاد ان يكون معدوما، اذ اظهر تطبيق هذه التقنية على عدد من البيانات المختبرة من العديد من الباحثين في مجال الانحدار الحصين كفاءة عالية جدا عن سابقاتها. أما الطريقة الثانية فهي تقنية Bootstrap الديناميكية الحصينة لطريقة المربعات الصغرى المشذبة، (Dynamic Robust Bootstrap for Least Trimmed of squares) "DRBLTS" اقترحت من قبل [6] (Midi, Uraibi & Talib 2009) وتقوم فكرتها على التشخيص الديناميكي للقيم الشاذة داخل عينات Bootstrap ومن ثم تشذيبها بشرط ان لا تتجاوز القيم الشاذة قيمة نقطة الانهيار، اما في حالة تجاوزها فيتم توليد عينة Bootstrap أخرى. إذ وظف (Uraibi, Midi, Talib & Yousif 2009) [15] هذه التقنية في مجال اختيار النموذج الأفضل من بين العديد من نماذج الانحدار الخطي حيث أثبتت هذه التقنية كفاءة عالية بالوصول الى نماذج خطية معتبرة لنفس البيانات التي اعتمدت عليها [9] (Ramli 2008) في تقنية (WBP).

١.٢ هدف البحث

يهدف هذا البحث إلى أجراء مقارنة بين طريقتين تم استخدامهما لإيجاد مقدرات حصينة في عينات تقنية bootstrap الطريقة الأولى اعتمدت على أعطاء أوزان قليلة للقيم الشاذة يناظرها احتمال ضعيف في الظهور عند أعادة المعاينة وهي طريقة (WBP)، والثانية اعتمدت على تشخيص القيم الشاذة داخل عينات Bootstrap وتشذيبها وهي طريقة (DRBLTS) لذا فقد حاولت هذه الدراسة البحث في أي الطريقتين أفضل اذ اعتمدت على التحيز والجذر ألتربيعي MSE كمعايير للمقارنة بين مقدرات هاتين الطريقتين وبالتالي تشخيص الطريقة الأفضل.

٢. منهجية البحث

٢.١ تقنية Bootstrap :

ان تقنية Bootstrap لإعادة المعاينة والتي هي حالة خاصة من أساليب المحاكاة بطريقة المونت كارلو والتي اقترحت من قبل [3] (Efron 1979) وذلك للحصول على مقدرات أكثر دقة عندما يكون حجم العينة صغيرا، ان الفكرة الأساسية لتقنية Bootstrap تتضمن القيام بتوليد عدد كبير من العينات المسحوبة مع الإرجاع من بيانات العينة الأصلية لـ (B) من المرات ويتم حساب معاملات الانحدار لكل عينة من عينات Bootstrap ، وهي تقديرات متحيزة لكننا في النهاية نحصل على تقديرات غير متحيزة بأخذ الوسط الحسابي لهذه التقديرات المتحيزة، كذلك فان هذه الطريقة تزودنا بالخطأ المعياري للمقدرات بدون أن يتطلب ذلك حسابات نظرية وهذا يوفر الوقت والجهد.

لقد اعتمدت تقنية Bootstrap الكلاسيكية على طريقة المربعات الصغرى الاعتيادية Ordinary Least Square في تقديراتها ولكن في حالة احتواء العينة على قيمة شاذة واحدة او أكثر فان طريقة المربعات الصغرى تكون غير مجدية ، إذ ان السحب والإرجاع من عينة تحتوي على قيمة شاذة لـ (B) من المرات ، قد يجعل من عينة Bootstrap تحتوي على نسبة من القيم الشاذة تزيد عن نسبتها في العينة الأصلية وربما تتجاوز هذه النسبة نقطة الانهيار (Breakdown Point) وبالتالي ستكون استنتاجاتنا غير صحيحة وان التنبؤات المستقبلية تكون خاطئة.

٢.٢ طريقة المربعات الصغرى المشذبة (LTS) Least Trimmed Square

طريقة المربعات الصغرى المشذبة اقترحت من قبل الباحث [11] (Rousseeuw 1984) يتم إيجاد المقدرات بهذه الطريقة وذلك بتصغير دالة الهدف الى اقل ما يمكن

$$\min \sum_{i=1}^{h} r_{(i)}^2$$

اذ ان $r_{(i)}^2$ تمثل مربعات البواقي المرتبة و h يمكن تحديدها كالأتي:

$$h = [(1-\alpha)*n] + [\alpha(k+1)]$$

تمثل α النسبة المئوية للقيم الشاذة او هي نسبة تشذيب المشاهدات. أما n هي حجم العينة و k هي عدد معاملات نموذج الانحدار. [10] (Rousseeuw & Leroy 2003) ذكروا ان النقطة التي تنهار فيها هذا المقدرات هي : $BP = ([(n-p)/2]+1)/n$

حيث ان P هي عدد المعلمات للنموذج و n تمثل حجم العينة.

أما الخوارزمية هي كالأتي:

١- تشخيص القيم الشاذة و معرفة نسبتها المئوية من مجمل البيانات على ان لا تكون هذه النسبة اكبر من BP.

٢- تحديد حجم المجموعة الجزئية من خلال $h = [(1-\alpha)*n] + [\alpha(k+1)]$.

٣- إيجاد التوافيق لاختيار h من n مشاهدة (بمعنى بكم طريقة يمكن اختيار h من n مشاهدة، C_h^n).

٤- تقدير معلمات الانحدار باستخدام OLS لكل مجموعة جزئية h_i .

٥- احتساب MSE لكل مجموعة جزئية، المجموعة التي تحقق اقل MSE فان معلماتها المقدرة تكون حصينة.

٢.٣ طريقة وسيط المربعات الصغرى (LMS) Least Median Square

اقترحت طريقة LMS من قبل [11] (Rousseeuw 1984) لتقدير معلمات نموذج الانحدار عند احتواء البيانات على قيم شاذة اذ يتم بهذه الطريق تصغير وسيط مربعات الخطأ للمجموعة الجزئية المرتبة h حيث ان

$$h = [n/2] + [(k+1)/2]$$

حيث ان n تمثل حجم العينة و k تمثل عدد معلمات الانحدار وان $\alpha = \frac{1}{2}$ والتي تمثل نسبة القيم الشاذة المفترضة اذ يتم ترتيب بواقي الانحدار تصاعديا ومن ثم يتم اختيار مجموعة من البيانات بحجم h من n من مربعات البواقي r_i المرتبة تصاعديا لتحقيق دالة الهدف

$$Minimize\ Med\ r_i^2$$

وبذلك فان معلمات الانحدار للمجموعة الجزئية h التي تحقق دالة الهدف أعلاه تكون معلماتها حصينة اذ سيتم تقسيم مربعات البواقي المرتبة إلى مجموعتين وعلية ستكون البواقي الشاذة في المجموعة (n-h) يكون لدينا لكل مجموعة خط انحدار المجموعة التي بحجم h.

٢.٤ خوارزمية Bootstrap الموزونة احتماليا (WBP)

اقترحت هذه الطريقة من قبل [9] (Ramli 2008) وهي عملية تحسين على خوارزمية اقترحها [8] (Imon 2005) ، إذ اعتمدت على خوارزمية (Tille 2006) [13] Bootstrapping with Re-sampling probabilities والتي تقترح تخصيص احتمال لكل عينة من العينات الممكنة الظهور في عينات Bootstrap تبنى على قيمه عشوائية تتبع توزيع منتظم (٠,١) وبهذا يتم اختيار عينات دون سواها، كذلك فان [9] (Ramli 2008) اعتمدت بتشخيص القيم الشاذة على بواقي طريقة وسيط المربعات الصغرى LMS اذ تم ترجيح البواقي بالاعتماد على دالة (Hampel) للوزن وذلك لغرض تحديد الاحتمال لظهور كل مفردة ضمن بيانات العينة وذلك من خلال معرفة الوزن الترجيحي للمشاهدة إلى مجموع الأوزان الترجيحية .

مما تقدم نلاحظ انه تم تخصيص احتمال مرجح للمشاهدة، لذلك دالة Hampel ستخصص وزن ترجيحي (٠) للقيم الشاذة وعلية سيكون احتمال ظهور القيمة الشاذة مساو إلى الصفر عند أعادة المعاينة وبذلك تكون عينات Bootstrap غير ملوثة وبالتالي إجراء عملية التقدير. ويمكن تلخيص الخوارزمية المقترحة من قبل Ramli بالشكل التالي:

١- تقدير معلمات الانحدار لكامل العينة باستخدام طريقة LMS . ثم يتم تشخص القيم الشاذة باستخدام دالة Hampel الموزونة لبواقي LMS المقترحة من قبل [14] (Venables & Ripley 1999) .

٢- قدر معلمات نموذج الانحدار للمشاهدات الباقية بدون القيم الشاذة للحصول على $\hat{\beta}^{(-D)}$ ومن ثم الحصول على $\hat{y}_i^{b(-D)} = f(x_i, \hat{\beta}^{(-D)})$

٣- حساب البواقي وحسب المعادلة $\hat{r}_i^{(-D)} = y_i - f(x_i, \hat{\beta}^{(-D)})$

٤- يتم سحب عينة Bootstrap الأولى عشوائيا مع الإرجاع من بواقي الانحدار $\hat{r}_i^{(-D)}$ للحصول على $\hat{r}_i^{b(-D)}$ بحجم n باستخدام خوارزمية [13] (Tille 2006).

٥- أيجاد $\hat{y}_i^{b(-D)} = f(x_R, \hat{\beta}^{(-D)}) + r_i^{b(-D)}$

٦- حساب معاملات الانحدار للقيم المقدرة $\hat{y}_i^{b(-D)}$ باستخدام المربعات الصغرى الاعتيادية للحصول على $\hat{\beta}^{b(-D)}$.

٧- كرر الخطوات من ٤-٦ لـ B من المرات (حيث ان B عدد طبيعي اكبر من ٢).
٨- يتم حساب معاملات الانحدار النهائية كما يأتي

$$\hat{\beta}_{boot}^{(-D)} = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}^{b(-D)} \qquad b = 1,2,..........., B$$

٢.٥ تقنية Bootstrap الديناميكية الحصينة "DRBLTS"

اقترحت هذه الخوارزمية من قبل (Midi, Uraibi & Talib 2009) اذ أشاروا إلى ان عملية استخدام تقنية Bootstrap الحصينة المعتمدة على طريقة المربعات الصغرى المشذبة RBLTS هو بديل جيد لتقنية Bootstrap المعتمدة على طريقة المربعات الصغرى الاعتيادية BOLS في حالة وجود قيم شاذة في العينة، و ذلك لان القيم الشاذة قد تظهر في عينات Bootstrap بنسب مختلفة بسبب عملية السحب والإرجاع، اذ قد تقل عن نسبتها في العينة الأصل او قد تزيد عنها لدرجة قد تتجاوز نقطة الانهيار، بل ان هناك ثمة احتمال ان يكون بعض هذه العينات خالية من القيم الشاذة. عليه استخدام طريقة المربعات الصغرى المشذبة لتقدير معلمات الانحدار لهذه العينات المعادة وبنسبة تشذيب مفترضة مسبقا، قد يؤدي إلى خسارة جزء من بيانات العينة المعادة او يؤدي إلى تشذيب جزء من القيم الشاذة في بيانات عينات أخرى. حيث ان العينة المعادة الخالية من القيم الشاذة سوف تخسر بيانات غير شاذة تساو نسبة التشذيب المحددة مسبقا، اما العينة المعادة والتي تكون فيها نسبة القيم الشاذة اقل من النسبة المفترضة فإنها سوف تخسر بيانات غير شاذة تساوي الفرق ما بين النسبتين. من جانب أخر إذا كانت العينة المعادة تحتوي على قيم شاذة تزيد عن النسبة المفترضة فان هذه التقنية سوف تشذب ما هو مساو إلى النسبة المفترضة و هذا يبقي نسبة من الشواذ مكافئة للفرق بين النسبتين. بناء على ما تقدم التشذيب التام لكل القيم الشاذة يحدث عندما النسبة المفترضة تساوي نسبة الشواذ في العينة المعادة. على اية حال ان التأثير سيكون عكسيا على المقدرات وبالتالي الحصول على مقدرات غير كفوءة .

لقد اقترح الباحثون تعديل يتم بموجبة تحديد النسبة المضبوطة للقيم الشاذة (α) من البيانات الأصلية بعدها يتم استخدام طريقة المربعات الصغرى المشذبة LTS في تقدير معلمات النموذج ، وحساب البواقي ومنها حساب القيم المعيارية للأخطاء فتكون القيم الشاذة هي القيم التي مطلق القيمة المعيارية للخطأ لها اكبر من ثلاثة.

ان هذه الخوارزمية المقترحة تحاول جعل التقدير أكثر حصانة من خلال إجراء يشخص بدقة نسبة القيم الشاذة (α) ثم يقوم بتشذيبها تلقائيا.

وعلية يمكن كتابة خوارزمية هذه الطريقة بالشكل التالي:-

١- يتم تحديد القيم الشاذة في البيانات بالاعتماد على طريقة وسيط المربعات الصغرى LMS ، وبالتالي تحديد قيمة (α) والتي هي نسبة القيم الشاذة في البيانات .

٢- استخدام طريقة المربعات الصغرى المشذبة LTS لتقدير معلمات النموذج $\hat{\beta}_{LTS}$ وحساب قيم $\hat{y}_i = f(x_i, \hat{\beta})$ ، اذ يتم تحديد القيم الشاذة كما ذكر في النقطة(١).

٣- يتم حساب الأخطاء $\varepsilon_i = y_i - \hat{y}_i$ ويتم إعطاء احتمال مقداره $\frac{1}{n}$ لكل ε_i ، بعدها يتم حساب القيم المعيارية للأخطاء وبالتالي تحديد القيم الشاذة وهي القيم التي يكون مطلق خطأها المعياري اكبر من ثلاثة.

٤- يتم سحب عينة Bootstrap من الأخطاء ε_i بحجم (n) مع الإرجاع، ثم تضاف الأخطاء المسحوبة ε_i^{*b} الى الجزء الثابت من معادلة الانحدار لإيجاد قيمة $y_i^{*b} = f\left(x_i, \hat{\beta}\right) + \varepsilon_i^{*b}$.

٥- تطبق طريقة المربعات الصغرى المشذبة LTS على عينة Bootstrap للحصول على $\hat{\beta}^*$ ، بعد تشذيب نسب القيم الشاذة بالاعتماد على الخطوة رقم(٤).

٦- يتم إعادة الخطوة ٣,٤,٥ لـ B من المرات للحصول على $\hat{\beta}^{*1}, \hat{\beta}^{*2}, ..., \hat{\beta}^{*B}$ حيث ان B يمثل عدد العينات التي تم سحبها.

كـل عينـة Bootstrap نسـبة القـيم الشـاذة بهـا اكبـر مـن نقطـة الانـهيـار يمكـن (Breakdown Point)(BP) يتم استبدالها تلقائيا بعينة مسحوبة أخرى.

٧- $$\hat{\beta}_{(boot)} = \frac{\sum_{b=1}^{B} \hat{\beta}^{*b}}{B} .$$

٨- $$bootstrap\,bias = \hat{\beta}_{(boot)} - \hat{\beta}_{(LTS)}$$

٩- $$\mathrm{var}(\hat{\beta}_{(boot)}) = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\beta}^{*b} - \hat{\beta}_{(boot)}\right)\left(\hat{\beta}^{*b} - \hat{\beta}_{(boot)}\right)^{T}$$

يشير الباحثان كل من [7] (Imon &Ali 2005) إلى انه ليس هنالك عدد متفق علية بين الاحصـائيون لعينات Bootstrap المسحوبة فمن الممكن ان تكون(٢٥) ولكن لتقدير الأخطاء المعيارية بشكل جيد عادة مـا تؤخذ ضمن المدى(٢٥-٢٥٠) وللحصول على حدود ثقـة عاليـة يتطلب ذلك ان تكون العينـات بين (٥٠٠-١٠٠٠٠) عينـة .

٢.٦ معيار المقارنـة

وللمقارنـة بين هاتين الطريقتين أيهما أفضل سيتم الاعتمـاد علـى معيـارين للمقارنـة همـا التحيز والـذي صيغته هي

$$bootstrap\,bias = \hat{\beta}_{(boot)} - \hat{\beta}_{(LTS)}$$

والجذر ألتربيعي لمتوسط مربعات الخطأ حيث ان

$$MSE(\hat{\beta}_{boot}) = (bootstrapbias)^2 + \mathrm{var}(\hat{\beta}_{(boot)}) .$$

وعلية فان الجذر لمتوسط مربعات الخطأ يكون

$$\sqrt{MSE(\hat{\beta}_{boot})}$$

٣. مناقشة النتائج

في هذا البحث سيتم الاعتماد على بيانات اختبرت من قبل العديد من الباحثين وذلك لغرض إجراء المقارنة بين استخدام الطريقتين التي تم التطرق إليهما في الجانب النظري هذه البيانات هي بيانات (Hawkins) وكذلك بيانات (Stackloss Data) وهي كما يأتي:-

٣.١ بيانـات (Hawkins ,Bradu & Kass 1984) [5]

تتكون هذه البيانـات من ثلاثة متغيرات تنبؤية اصطناعية لمجموعة من ٧٥ مشاهدة ، أشار الخبراء إلى وجود ١٠ قيم شاذة على كلا الجانبين (x,y) هي الحالات من (١-١٠)، و ٤ قيم شاذة ضمن مشاهدات المتغير المعتمد y هي المشاهدات (١١-١٤) ، والتي يمكن تمثيلها بالنموذج التالي:-

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$$

وبعد إجـراء عمليـة التحليـل لـهذه البيانـات بواسـطة برنـامج بلغـة **S-plus** تـم كتابتـه لـهذا الغرض أي لحساب النتائج من هذه البيانات تم التوصل إلى النتائج التالية :

جدول رقم (١)

يبين التقديرات التي تم التوصل إليها باستخدام طريقة (WBP) وطريقة (DRBLTS)

Estimators	Methods	
	WBP	DRBLTS
$\hat{\beta}_0$	-0.18118	-0.9647507
$\hat{\beta}_1$	0.077715	0.141532
$\hat{\beta}_2$	-0.053225	0.2036738
$\hat{\beta}_3$	0.0000	0.1807521
Bias $\hat{\beta}_0$	0.8514907	0.06792006
Bias $\hat{\beta}_1$	-0.07614485	-0.01232784
Bias $\hat{\beta}_2$	-0.07614485	0.00273005
Bias $\hat{\beta}_3$	-0.1798329	0.0009192287
RMSE $\hat{\beta}_0$	0.8514907	0.06792006
RMSE $\hat{\beta}_1$	0.07614485	0.01232784
RMSE $\hat{\beta}_2$	0.2541687	0.00273005
RMSE $\hat{\beta}_3$	0.2541687	0.0009192287

من الجدول رقم (١) والذي يوضح التقديرات لمعاملات النموذج بالاعتماد على الطريقتين WBP و DRBLTS ، نلاحظ من النتائج التي تم التوصل أليها ان مطلق التحيز لمعالم النموذج المقدرة بطريقة DRBLTS اقل من التحيز لمعالم النموذج المقدرة بطريقة WBP، كذلك فان جذر متوسط مربعات الخطأ لمعالم النموذج المقدرة بطريقة DRBLTS اقل منه لمعالم النموذج المقدرة بطريقة WBP وعلية يمكن القول أن طريقة DRBLTS أفضل في التقدير لهذه البيانات في عينات Bootstrap .

٣.٢ بيانات (Stackloss for Brownlee 1965) [2]

تصف هذه البيانات عملية تسميد نبات معين بنسبة من أوكسيد الامونيا و حامض النتريك والتي تتكون من ٢١ مشاهدة رباعية الأبعاد حيث متغير الاستجابة (y) الخسارة في ارتفاع النبات، أما x1 يمثل نسبة العملية أي عملية التسميد ، $x2$ درجة الحرارة المدخلة لتبريد الماء ، و$x3$ تمثل التركيز ألحامضي. بعض الباحثين أشاروا إلى أن هذه البيانات تحتوي على أربع أو خمس قيم شاذة.أما النموذج فتم تمثيله بالشكل التالي

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$$

وبعد إجراء عملية التحليل لهذه البيانات تم التوصل إلى النتائج التالية

جدول رقم (٢)

يبين التقديرات التي تم التوصل إليها باستخدام طريقة (WBP) وطريقة (DRBLTS)

Method	WBP	DRBLTS
$\hat{\beta}_0$	-37.57667	-38.0577
$\hat{\beta}_1$	0.794605	0.8811702
$\hat{\beta}_2$	-0.06474	0.6860433
$\hat{\beta}_3$	0.000	-0.1384689
Bias $\hat{\beta}_0$	1.245076	0.7640472
Bias $\hat{\beta}_1$	-0.08539765	0.001167585
Bias $\hat{\beta}_2$	-0.7338712	0.01691206
Bias $\hat{\beta}_3$	0.1284749	-0.009994068
RMSE $\hat{\beta}_0$	1.245076	0.7640472
RMSE $\hat{\beta}_1$	0.08539765	0.001167585
RMSE $\hat{\beta}_2$	0.7338712	0.01691206
RMSE $\hat{\beta}_3$	0.1284749	0.009994068

من خلال الجدول رقم (٢) والذي يوضح التقديرات التي تم التوصل أليها من الطريقتين التي تم استخدامهما نلاحظ أن مطلق التحيز لمعلمات النموذج المقدرة باستخدام طريقة DRBLTS كانت اقل من التحيز لمعلمات النموذج المقدر بطريقة WBP ، كذلك أن جذر متوسط مربعات الخطأ المقدرة بطريقة DRBLTS اقل من جذر متوسط مربعات الخطأ المقدرة بطريقة WBP وعليه فأننا نستنتج بان طريقة DRBLTS أفضل من طريقة WBP في تقدير معلمات النموذج لعينات Bootstrapالتي تحوي قيم شاذة .

٤. الاستنتاجات والتوصيات

٤.١ الاستنتاجات

لقد توصل الباحث إلى مجموعة من النتائج أهمها

١- من الاستعراض الذي تم التطرق إليه في هذا البحث نستطيع ان نستنتج انـه يجب على الباحث وقبل تحليل البيانات معرفة فيما أذا كانت تحتوي على قيم شاذة أم لا وعلى اثر ذلك يتم تحديد الطريقة المثلى في عملية التقدير ، لأنه ان تم التقدير بالطرق العادية مع وجود قيم شاذة فان هذه التقديرات سوف لن تعطينا تقديرات صحيحة للبيانات وبالتالي الحصول على تنبؤات خاطئة.

٢- نستنتج ان استخدام طريقة DRBLTS أعطت تحيز اقل وجذر تربيعي لمتوسط مربعات خطأ اقل ممـا أعطتـه طريقة WBP وعلية فأنها تقنية جيدة في تقدير معلمات الانحدار الحصين باستخدام تقنية Bootstrap.

٤.٢ التوصيات

بناء على ما تقدم يوصي الباحث بما يأتي:

١- استخدام تقنية DRBLTS لتقدير معالم انحدار حصينة.

٢- اختبار كلا التقنيتين في تقدير تنبؤات حصينة في موضوع السلاسل الزمنية.

المصادر

١- صالح، احمد مهدي (٢٠٠٥) " التقديرات الحصينة للجرعة المؤثرة الوسيطية ED50 في التجارب الحياتية ثنائية الاستجابة" رسالة ماجستير مقدمة الى مجلس كلية الادارة والاقتصاد الجامعة المستنصرية.

2- Brownlee, K.A. 1965. “Statistical Theory and Methodology in Science and Engineering”, 2nd Edn.,New York: John Wiley and Sons.

3- Efron, B. (1979). Bootstrap Methods. Another Look at the Jackknife. The Annals of Statistics, 7: 1–26.

4- Huber, P. J.(1981), Robust Statistics, John Wiley & Sons, New York, NY, USA,

5- Hawkins, D.M., Bradu, D., and Kass, G.V. (1984) Location of several outliers in multiple regression data using elemental sets. *Technometrics,* 26, 197–208

6- Hassan S.Uraibi, Habshah Midi,and Bashar A. Talib (2009). “Dynamic Robust Bootstrap Method Based on LTS Estimators ”, European Journal of Scientific Research”, 32 3,277-287

7- Imon, A.H.M.R, and Ali.M.M, 2005. “Bootstrapping regression residuals”, *J.Korean Data Inform.Sci*.Soc.,16, 665-682.

8- Imon A.H.M.R (2005) A Stepwise Procedure for the Identification of Multiple Outliers and High Leverage Points in Linear Regression. Pakistan Journal of Statistics, 21, 71-86.

9- Ramli,N.M.(2008)"weighted maximum median likelihood Estimation for Parameters in multiple linear regression model" Ph.D thesis , University Putra Malaysia (UPM) , Malaysia.

10- Rousseeuw P. J. and Leroy. M.A ., 2003. “Robust Regression and Outlier detection”, Illustrated Edn., New York:Wiley-IEEE.

11- Rousseeuw P. J. (1984) Least median of squares regression, *J. Am. Stat. Assoc.,* 79, 871-880.

12- Rousseeuw P. J. & Zomeren, B. C. V. (1990) Unmasking Multivariate Outliers and Leverage Points (with discussion). Journal of the American Statistical Association, 85, 633-651

13- Tille,Y. (2006)"Sampling and Algorithms" Springer Verlag New York ,LLC

14- Venables,W.N.& Ripley B.D. (1999) " Modern Applied Statistics with S-PLUS " Third Edition , New York, Springer.

15- Uraibi, S. Hassan, Midi, Habshah , Talib A. Bashar & Yousif H. Jabar. (2009) Linear Regression Model Selection Based on Robust Bootstrapping Technique, American Journal of Applied Sciences 6 (6): 1191-1198,