Journal of Economics and Administrative Sciences Vol.26 (NO. 124) 2020, pp. 493-505

Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq

التقدير الحصين لمعلمات الانحدار الخطي المتعدد بوجود مشكلة عدم تحقق تجانس تباين الخطأ وظهور القيم الشاذة

الباحث /شيماء إبراهيم خليل
كلية الإدارة والاقتصاد/جامعة بغداد/ قسم الاحصاء
shaimaa595@yahoo.com

أ.م.غفران إسماعيل كمال
كلية الإدارة والاقتصاد/جامعة بغداد/ قسم الإحصاء
ghufranka62@gmail.com

Received:19/7/2020 Accepted : 16/9/2020 Published :December / 2020

مستخلص البحث:

في كثير من الاحيان وخاصة في التطبيقات العملية يصعب الحصول على بيانات لا تشوبها مشكلة قد تتعلق بعدم تجانس تباين الخطأ او اي مشكلة اخرى تعوق استعمال الطرائق الاعتيادية المتمثلة بطريقة الـ(OLS) لإيجاد مقدرات معالم الانموذج الخطي المتعدد، لهذا يلتجئ الكثير من الاحصائيين الى استعمال التقديرات بالطرائق الحصينة وخاصة بوجود القيم الشاذة (الملوثة) الى جانب مشكلة عدم ثبات تباين الخطأ. فتم اعتماد طريقتين للحصانة هما طريقة المربعات الصغرى الموزونة الحصينة(RWLS) وطريقة المربعات الصغرى الحصينة ذات المرحلتين(TSRWLS) وتم التحقق من ادائهما من خلال تطبيقه للمحاكاة، واختيار افضل المقدرات الحصينة من خلال متوسط الخطأ النسبي المطلق (MAPE) معياراً احصائياً للمقارنة بينهم، وتوصلنا الى ان افضل طريقة للتقدير كانت طريقة ذات المرحلتين (TSRWLS).

المصطلحات الرئيسة للبحث: الانموذج الخطي المتعدد، عدم ثبات تباين الخطأ، القيم الشاذة (الملوثة)، طريقة المربعات الصغرى الموزونة الحصينة(RWLS)، طريقة المربعات الصغرى الموزونة ذات المرحلتين (TSRWLS)

- البحث مستل من رسالة ماجستير

1-المقدمة وهدف البحث

انموذج الانحدار الخطي المتعدد هـــو واحد مـن اهـم الاساليب الإحصائية والذي يبنى على هيئة معادلة مبينة العلاقة السببية بين عدة متغيرات توضيحية(X) و متغيـر استجابة (Y)واحد، وحيث قام العديد من الباحثين باستعمال إنموذج الانحدار الخطي المتعدد ويعتبر الباحثان (Merran&Maxewll) هما من اوائل من قدما مشكلة عدم تجانس التباين للخطأ وظهور القيم الشاذة معاً وعملا على استعمال اسلوب المحاكاة للمقارنة بينهم[8]، وفي العام (2002) قام الباحث (Colin Chen) بأستعمال اربع مقدرات حصينة هي (مقدرLTS)و(مقدر-M) و(مقدر -S)و(مقدر-MM) وهي الاكثر استعمالاً، في حالة احتواء البيانات على قيم شاذة، وتوصل الى استعمال اي مقدر من هذه المقدرات يجعل الانحدار حصين ويعتمد على النتائج التي يتم التوصل اليها[9]، وفي العام (2008) بين كل من (.Rana, Md) واخرون على ان وجود القيم الشاذة تجعل الكشف عن مشكلة عدم تجانس تباين الخطأ صعب وبطيء، وغالباً ما تجعل القيم الشاذة الانموذج المتجانس (المتماثل) الى إنموذج غير متجانس وتودي كذلك الى تشويه ادوات التشخيص، وعملوا على تعديل اكثر الاختبارات شيوعاً واستخداماً هو اختبار كولد فيلد[10].الهدف مـن هذا البحث هو تقدير معلمات انموذج الانحدار الخطي المتعدد باستعمال طريقة المربعات الصغرى الموزونة الحصينة(RWLS) وطريقة المربعات الصغرى الحصينة ذات المرحلتين(TSRWLS) والمقارنة بين هذه الطرائق لاختيار افضلها، وتم ذلك باستعمال المحاكاة ومن خلال معيار المقارنة متوسط الخطأ النسبي المطلق (MAPE).

2- أنموذج الانحدار الخطي المتعدد (Multiple Linear Regression Model)

انموذج الانحدار الخطي المتعدد هـــو واحد مـن اهـم الاساليب الإحصائية التي يتم مـن خلاله التأكد مـن دقـة الإستدلال والذي يبنى على هيئة معادلة مبينة العلاقة السببية بين عـدة متغيـرات توضيحية(X) و متغيـر استجابة(Y) واحد والتي يستدل منها على اهمية وقـوة تلك العلاقـة بين المتغيرات مـن خـلال تقدير معلماتها ، ثـم التنبؤ بالقـيم المـستقبلية واتخــاذ الـقـرارات المناسبة لحل الظـاهرة تحت البحث[3]

$$Y_i = \beta_0+\beta_1 X_{i1}+\beta_2 X_{i2} + \cdots + \beta_k X_{ik} + U_i \qquad \ldots(1)$$

$$i = 1, 2, \ldots, n \quad , \quad j = 1, 2, \ldots, k$$

حيث يكون المقدر التقليدي في تقدير هذا الانموذج هو مقدر المربعات الصغرى (OLS) ويمتلك هذا المقدر خاصية افضل مقدر خطي غير متحيز عند تحقق الافتراضات الخاصة بأنموذج الانحدار الخطي ولكن عند خرق الافتراض الخاص بثبات تباين قيم الخطأ .

$$\text{var}(u_i) = \sigma_u^2 \qquad \ldots(2)$$

سوف يعاني الانموذج من مشكلة عدم ثبات تجانس تباين الخطأ، في هذه الحالة فان استخدام مقدر (OLS) سوف يكون متحيزاً وغير كفوء، ويمكن التعبير عن المعادلة، (1) بأسلوب المصفوفات على النحو الاتي:-

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U} \qquad \ldots (3)$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{(n\times 1)} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}_{(n\times(k+1))} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{((k+1)\times 1)} + \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}_{(n\times 1)}$$

إذ أن :- [1]

Y: يمثل متجه متغير الاستجابة ذو مرتبة $(n \times 1)$.

X: تمثل مصفوفة للمتغيرات التوضيحية ذات مرتبة $(n \times (k+1))$ علما" بأن العمود الاول منها يمثل الحد الثابت .

β: يمثل متجه المعلمات المراد تقديرها ذو مرتبة $((k+1)\times 1)$ والعنصر الاول فيه يمثل الحد الثابت للإنموذج .
U: يمثل متجه الأخطاء العشوائية ذو مرتبة $(n \times 1)$.
k: عدد المتغيرات التوضيحية .
n : عدد المشاهدات في العينة.

3- مشكلة عدم تجانس تباين الخطأ العشوائي (Heteroscedasticity Problem)

احدى الفرضيات الاساسية والتي تعتبر ركيزة من الركائز التي يقوم عليها الانموذجان الخطيان (البسيط والعام) هو تجانس تباين الخطأ (ثبات التباين لحدود الخطأ) ويصبح الفرض فيهما كالآتي :[1]

في الإنموذج البسيط $E(U_i^2) = \sigma_u^2$

وفي الإنموذج العـام $E(U\grave{U}) = \sigma_u^2 In$

ولكن ما يحدث في الواقع التطبيقي والذي يواجهه اغلب الباحثين هو عدم تحقق الشرط اعلاه ، اي يصبح التباين غير ثابت لجميع المشاهدات ، ويكون الفرض هو كالآتي:[4]

$$E(U\grave{U}) \neq \sigma_u^2 ln \quad \dots (4)$$

وتظهر أغلب هـذه المشاكل بصورة خاصة في الدراســات التي تعتمد على البيانات المقطعية (Cross–Sectional data) كون إن كل مشاهدة فيها يختلف اختلافا" كبيرا" في قيم المتغيرات التوضيحية (X)الأمر الذي يؤثر على مشاهدات متغير الاستجابة(Y)[2]
ومن النتائج التي تترتب على مشكلة عدم تجانس تباين الخطأ هو صعوبة الاعتماد على نتائج مؤكدة لتأثير مشكلة الظاهرة قيد البحث، وتكون التباينات المتحصلة عليها كبيرة وبالتالي عدم تمتعها بخاصية اقل تباين ولن تكون ذات كفاءة سواء كان ذلك في العينات الصغيرة او الكبيرة ، وهذا يؤدي الى اتساع فترة الثقة للمعلمات ، وعدم القدرة بالتالي على التنبؤ.

4-القيم الشاذة (The Outliers)

إن وجود قيم غريبة في العينة المختارة قد اصطلح على تسميتها علميا" بالمشاهدات أو القيم الشاذة ، التي تؤثر بشكل واضح على تغيير نتائج التحليل المعتمد ويكون هذا التغيير كبيرا" كلما زادت اعداد القيم الشاذة، وتعتبر البيانات التي لا تحتوي على شواذ حالة استثنائية في الواقع العملي والتطبيقي .
وقد تظهر المشاهدات الشاذة في مجموعة من البيانات نتيجة لتوزيعات غير متماثلة (التواء عالٍ نحو اليمين أو نحو اليسار) ، وقد يكون سبب حدوثها نتيجة لأسباب أخرى منها أخطاء يقع فيها الباحث عند تسجيل القياسات ، أو نتيجة إلى وجود خلل في أجهزة القياسات وخاصةً في التجارب المختبرية ، أو نتيجة أخطاء في الحسابات [6]

5- طرائق التقدير الحصينة (Robust Estimation Methods)

- **طريقة المربعات الصغرى الموزونة الحصينة (Robust Weighted Least Squares(RWLS))**

تعد طريقة المربعات الصغرى الموزونة (WLS) من أكثر الطرائق استعمالاً في تقدير أنموذج الانحدار الخطي المتعدد بوجود مشكلة عدم ثبات تجانس تباين الخطأ، لكن قد تعاني مقدرات طريقة الـ (WLS) من ضعف بظهور بعض القيم الشاذة. ولمعالجة ذلك لابد من استعمال طريقة بديلة يتم من خلالها معالجة مشكلة عدم ثبات تباين الخطأ وظهور القيم الشاذة في البيانات، وتتألف هذه الطريقة من الخطوات الاتية [5]:

<u>الخطوة الاولى :</u> تحديد المشاهدات القريبة للمتغيرات التوضيحية (X) مع بعضها (الجار الاقرب)، وعدد هذه المجاميع (الجار الاقرب) يشير لها بـ g من المجاميع .
<u>الخطوة الثانية :</u> اعتماد وحساب الوسيط لكل مجموعة من المجموعات في المتغيرات التوضيحية (X) .

$$M_{ed}(X_{(L)}) \quad , L = 1, 2, \dots, g$$

وكذلك يتم حساب الانحرافات المطلقة للوسيط (Median absolute deviation(MAD)) لقيم متغيرالاستجابة (Y) اي :

$$(Y_L) = Median\{|Y_L - Median(Y_L)|\} \quad \dots (5)$$

لـ (g) من المجاميع لمتغير الاستجابة المقابلة لمجاميع المتغيرات التوضيحية X.
<u>الخطوة الثالثة :</u>حساب انحدار $\{MAD(Y_l)\}^2$ على $Med(X_l)$ للانموذج الخطي المتعدد، باستعمال طريقة المربعات الصغرى المشذبة (LTS)، وحساب معلماته.

الخطوة الرابعة : باستعمال المعلمات المقدرة في الخطوة الثالثة يتم الحصول منها على انحدار المربعات الصغرى المشذبة لـY على X'_S.

الخطوة الخامسة : الاوزان الاولية ستكون معكوس القيم المطلقة للقيم التقديرية يشار لها (W_{1i}) ، والاوزان النهائية ستكون بتوظيف دوال الوزن الحصينة كما في دالة الوزن المقترحة من قبل Huber عام (1973)والمعرفة كالاتي:

$$W_{2i} = \begin{cases} 1 & |e_i| \leq Ð \\ \dfrac{1.345}{|e_i|} & |e_i| > Ð \end{cases} \qquad \dots(6)$$

إذ ان :

Ð: يدعى ثابت الضبط او التناغم وتكون قيمته هي (1.345) .

e_i: تمثل البواقي القياسية للمربعات الصغرى المشذبة LTS والتي تم الحصول عليها في الخطوة الثالثة.

وبضرب الاوزان الاولية W_{1i} في الاوزان النهائية W_{2i} نحصل على الاوزان الحصينة W_i.

الخطوة السادسة : نطبق طريقة المربعات الصغرى الموزونة باستعمال الاوزان الحصينة W_i و الحصول على معاملات الانحدار التقديرية في ظل وجود مشكلتي عدم تجانس تباين الخطأ والقيم الشاذة، وكما في الصيغة التالية :-

$$b_{Rwls} = (X'W_iX)^{-1}X'W_iY \qquad \dots(7)$$

- **طريقة المربعات الصغرى الحصينة ذات المرحلتين Two-Step Robust Weighted Least (Squares (TSRWLS))**

تعد طريقة المربعات الصغرى الموزونة الحصينة ذات المرحلتين من الطرائق المهمة في التقدير عند وجود مشكلة عدم تجانس تباين الخطأ وظهور القيم الشاذة، وتعطي تقديرات متسقة وغير متحيزة ، تتألف هذه الطريقة من الخطوتين الأتيتين : [7]

الخطوة الاولى: تتضمن تحديد اوزان اولية.

والخطوة الثانية: تتضمن تحديد اوزان نهائية .

وكما يأتي :

الخطوة الاولى :

1. ايجاد القيم التقديرية ($\hat{Y}_i$) من انموذج الانحدار الخطي المتعدد حسب الصيغة (3) باستعمال طريقة المربعات الصغرى المشذبة (LTS)، ثم ايجاد قيم البواقي (e_i).
2. حساب القيم المطلقة للبواقي ، si=| e_i |.
3. ايجاد القيم التقديرية لـ s_i اي ايجاد ، $\widehat{Si} = \frac{|e_i e_i|}{n-p}$.
4. حساب الاوزان الاولية الحصينة من خلال صيغة معكوس مربعات القيم المقدرة لـ $(\widehat{Si})$ حسب الصيغة التالية:-

$$w_{1i} = \frac{1}{(\hat{s}_i)^2} \qquad \dots(8)$$

الخطوة الثانية : لتحديد الاوزان النهائية لابد من استعمال دوال اوزان حقيقية كدالة هوبر Huber function او دالة Bisquare function، وهنا سوف يتم استعمال Huber function كدالة اوزان والتي تأخذ الصيغة الاتية :

$$w_{2i} = \begin{cases} 1 & |e_i| \leq Ð \\ \dfrac{1.345}{|e_i|} & |e_i| > Ð \end{cases} \qquad \dots(9)$$

حيث ان :

Ð : يمثل ثابت التناغم وتكون نسبته هي 1.345.

e_i : البواقي القياسية التي تم الحصول عليها من طريقة LTS بالخطوة الاولى (1).

وبضرب الاوزان الاولية w_{1i} بالأوزان w_{2i} لإيجاد الاوزان النهائية W_i اي :

$$W_i = w_{1i} * w_{2i} \quad ...(10)$$

واخيراً سوف يكون اداء المربعات الصغرى الموزونة WLS باستعمال الاوزان النهائية جيد في تقدير معلمات انموذج الانحدار الخطي المتعدد العام بوجود مشكلتي عدم التجانس والقيم الشاذة.

$$b_{Tswls} = (X'W_iX)^{-1}X'W_iY \quad ...(11)$$

6- اختبار (كولد فيلد كوانت) المعدل Modification of the Goldfeld –Quandt Test

هو اختبار يعد بمثابة تعديل لاختبار Goldfeld-Quandt التقليدي (الذي يتم تطبيق فيه طريقة (OLS) والتي تكون كما هو معروف حساسة تجاه القيم الشاذة وتعطي تقديرات متحيزه)، ولذلك تم تعديله باستبدال واستعمال مقدرات لا تتأثر بالمشاهدات الشاذة مثل مقدر طريقة المربعات الصغرى المشذبة (Least Trimmed Squares Method) (LTS) وسمي هذا الاختبار اختبار كولد فليد كوانت المعدل (MGQ)، ويتضمن هذا الاختبار الخطوات الاتية :-[10]

الخطوة الاولى : يتم ترتيب القيم حسب قيم مصدر الاختلاف بشكل تصاعدي او تنازلي.

الخطوة الثانية : يتم تحديد واستبعاد (c) من مشاهدات (حذف المشاهدات الوسطية) المركز في حدود ربع المشاهدات الكلية:

$$\mathrm{C} \cong \frac{1}{4} * \mathrm{n} \quad ...(12)$$

والمتبقي من العينة (n – c) يتم تقسيمه الى مجموعتين متساويتان، كل منهما تشتمل على $(\frac{n-c}{2})$ من المشاهدات .

الخطوة الثالثة : إستعمال أحد اساليب الانحدار الحصين (طريقة المربعات الصغرى المشذبة (Least Trimmed Squares Method) (LTS) المقترحة من قبل (Rousseeuw and Leroy) لتوفيق خط الانحـــــدار، لكون هذا النوع من المقدرات يمتلك حصانة ضد القيم الشاذة.

الخطوة الرابعة : أستخراج البواقي المحذوفة ، لكلا المجموعتين الاولى والثانية على التوالي ، وبعد ذلك حذف القيم الشاذة الموجودة في البيانات لكلا العينتين الجزئيتين ،ثم حساب الوسيط لمربعات البواقي المحذوفة (Median of the Squared Deletion Residuals) (MSDR) من خلال الصيغة الاتية :-

$$MGQ = \frac{\mathrm{MSDR}_2}{\mathrm{MSDR}_1} \quad ...(13)$$

$MSDR_2$و$MSDR_1$ هما الوسيطيين لقيم مربعات البواقي على التوالي ، لأصغر واكبر تباين للمجموعة على التوالي ، تحت افتراض التوزيع الطبيعي الاحصاءة (MGQ) تتبع توزيع (F) بدرجات الحرية لكل من البسط والمقام $\frac{(n-c-2K)}{2}$.واختبار صحة فرضية العدم بوجود مشكلة عدم التجانس اذا كانت قيمة (F)المحتسبة اقل من قيمة (F) الحرجة في الجداول الاحصائية.

7- معيار المقارنة (Compare Criterion)

لغرض التوصل للمقدر الاكفأ وجدت عدة معايير (مقاييس) للمقارنة بين طرائق التقدير ، وإن افضلها من يمتلك اقل خطأ ممكن وهذا يقودنا بالتالي للأنموذج الافضل لغرض التقدير والتنبؤ للظاهرة تحت البحث، وأحد هذه المعايير هو متوسط الخطأ النسبي المطلق (MAPE) المعتمد للمقارنة في البحث هذا ، ويمكن حسابه بشكل نسبي من خلال الصيغة التالية:[11]

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - y_i}{Y_i}\right| \times 100 \quad ...(14)$$

Y_i : تمثل قيم متغير الاستجابة.

y_i : يمثل قيم المتنبأ بها لمتغير الاستجابة Y_i.

8- مراحل وصف تطبيق تجربة المحاكاة

سيتم الاعتماد على أنموذج الانحدار الخطي المتعدد حسب صيغة انموذج الانحدار الخطي المتعدد(1)، لوصف المراحل التجريبية للمحاكاة والمقارنة بين الطرائق الحصينة لاختيار افضلها حسب معيار متوسط الخطأ النسبي المطلق
.(Mean Absolute Percentage Error)(MAPE)
اذ يتم وصف تجارب المحاكاة من خلال المراحل والخطوات الآتية :

- تم اختيار حجوم عينات مختلفة هي (n=50,99,150) للمتغير المعتمد (Y) وفق أنموذج الانحدار الخطي المتعدد حسب الصيغة (1) .
- ويتوزع الخطأ $e_i \sim N\,(0,1) + couchy(0,10)$
- لتوليد انموذج الانحدار الغير متجانس يتم من خلال الصيغة التالية:

$$\sigma_i^2 = \sigma^2 \operatorname{Exp}(a x_{1i} + a\, x^2{}_{2i} + a\, x^3{}_{3i} + a\, x^4{}_{4i} + a\, x^5{}_{5i}) \qquad ...(15)$$

حيث ان $\sigma^2 = 1$.
a: يمثل ثابت الاضطراب .

$$\sigma = \max(\sigma_i^2) / \min(\sigma_i^2) \qquad ...(16)$$

وتم تحديد نسبة عدم تجانس تباين الخطأ ولكل حجم عينة { a=0 و a=1.5 وa=2.1 }، فعندما تكون قيمة (a=0) يكون (σ=1) وهي النسبة التي تكون عندها البيانات متجانسة ، وعند(a=1.5 , a=2.1) على التوالي يكون (σ=2.8 و σ=3.8) وهي النسب التي تدل على وجود مشكلة عدم تجانس تباين الخطأ، ويتم هذا كله بهدف ايجاد النموذج الغير متجانس وبعد تكرار التجربة (1000) مرة، ولحجوم العينات المختلفة (n=50,99,150)، وبتطبيق الصيغ (15)و(16) كانت النتيجة للبيانات هي (σ =2.8)، والتي دلت في البحث هذا على وجود مشكلة عدم تجانس التباين وبنسبة وجود للقيم الشاذة (10%) وحجم عينة (n=99) .

- تحديد واختيار القيم الافتراضية للمعالم بالاعتماد على المعالم الحقيقية (علماً انه تم اختيار قيم معلمات البحث من خلال تجربة نفذت على بيانات حقيقية لعسرة المياه تم الحصول عليها من امانة بغداد) ،وتعد هذه المرحلة من اهم المراحل التي يعتمد عليها لاحقاً.
- توليد المتغيرات التوضيحية (X_{ij}) الخمسة، وتم هذا بالاعتماد على توزيعها في البيانات الحقيقية، وباستعمال الدوال الجاهزة في برنامج (R) وكما يلي :

$X_1 \sim N(0.0925,0.01)$

$X_2 \sim N(0.0737,0.005)$

$X_3 \sim N(0.065,0.034)$
$X_4 \sim N(0.168,0.02)$
$X_5 \sim N(0.542,0.01)$

- تعويض المتغيرات التي تم توليدها في اعلاه لأجل الحصول على متغير الاستجابة (Y) .
- يتم حساب إحصاءه الطرائق الحصينة والمقارنة بينهم عن طريق معيار (MAPE)، وبتكرار (1000) مرة.

9 - نتائج تجربة المحاكاة :

سوف يتم استعرض نتائج تجارب المحاكاة وتحليلها وحسب القيم الافتراضية للمعلمات وذلك من خلال طريقتي (RWLS) و(TSRWLS) ولقيم مختلفة لكل من تباين الخطأ العشوائي والقيم الشاذة .
ملخص الجداول التي سوف يتم تناولها لتقدير معلمات الأنموذج: تم اعتماد حجوم عينات مختلفة هي (n =50, 99, 150) وعند قيم مختلفة لتباين الخطأ العشوائي $(\sigma = 1, 2.8, 3.8)$ ولنسب شواذ هي $(q = 0\%, 10\%, 20\%)$، ولأجل اجراء المقارنة سيتم الاعتماد على معيار متوسط الخطأ النسبي المطلق (MAPE) للمقارنة بين الطرائق الحصينة ، وفيما يلي النتائج الموضحة في الجداول التالية:

جدول (1)

يبين مقدرات المعلمات وقيم (MAPE) للانموذج الخطي (n=50)و(σ=1)

المعلمات	القيم الافتراضية	نسبة القيم الشاذة					
		%0		%10		%20	
		RWLS	TSRWLS	RWLS	TSRWLS	RWLS	TSRWLS
b_0	-0.38967	-0.43019	-0.38386	-0.44384	-0.37299	-0.43747	-0.38480
b_1	0.75359	0.82198	0.76668	0.75964	0.74951	0.74526	0.74508
b_2	3.54956	3.70728	3.51932	3.93372	3.55103	3.98747	3.63170
b_3	-0.07543	-0.08102	-0.07254	-0.08136	-0.05987	-0.07260	-0.07574
b_4	2.62533	2.75182	2.63013	2.87401	2.59476	2.93790	2.62836
b_5	0.21939	0.23322	0.22300	0.24526	0.21661	0.24206	0.22492
MAPE		0.10436	0.10207	0.36293	0.12441	0.65156	0.11039

جدول (2)

يبين مقدرات المعلمات وقيم (MAPE) للانموذج الخطي (n=99)و(σ=1)

المعلمات	القيم الافتراضية	نسبة القيم الشاذة					
		%0		%10		%20	
		RWLS	TSRWLS	RWLS	TSRWLS	RWLS	TSRWLS
b_0	-0.38967	-0.42152	-0.37651	-0.44221	-0.38019	-0.45326	-0.37778
b_1	0.75359	0.73211	0.70280	0.82862	0.75126	0.85307	0.73698
b_2	3.54956	3.73551	3.57182	3.85885	3.59906	4.00367	3.50597
b_3	-0.07543	-0.07652	-0.07831	-0.08420	-0.07582	-0.08295	-0.07292
b_4	2.62533	2.75541	2.62437	2.85157	2.61815	2.96032	2.62999
b_5	0.21939	0.23042	0.22169	0.24033	0.22123	0.24559	0.21933
MAPE		0.07911	0.06751	0.27211	0.10695	0.49317	0.10896

جدول (3)
يبين مقدرات المعلمات وقيم (MAPE) للانموذج الخطي (n=150)و(σ=1)

المعلمات	القيم الافتراضية	نسبة القيم الشاذة					
		%0		%10		%20	
		RWLS	TSRWLS	RWLS	TSRWLS	RWLS	TSRWLS
b_0	-0.38967	-0.42545	-0.37604	-0.43519	-0.38343	-0.46592	-0.37795
b_1	0.75359	0.76180	0.73519	0.81664	0.76140	0.87264	0.73630
b_2	3.54956	3.82540	3.55616	3.83751	3.59967	4.11819	3.60315
b_3	-0.07543	-0.08251	-0.07835	-0.09717	-0.07214	-0.08189	-0.07094
b_4	2.62533	2.76323	2.61550	2.83256	2.63003	2.97050	2.61807
b_5	0.21939	0.22855	0.21959	0.23852	0.22115	0.25583	0.21873
MAPE		0.06341	0.05180	0.21609	0.08309	0.40287	0.10708

جدول (4)
يبين مقدرات المعلمات وقيم (MAPE) للانموذج الخطي (n=50)و(σ=2.8)

المعلمات	القيم الافتراضية	نسبة القيم الشاذة					
		%0		%10		%20	
		RWLS	TSRWLS	RWLS	TSRWLS	RWLS	TSRWLS
b_0	-0.58967	-0.10208	-0.59185	-0.12551	-0.61115	-0.11783	-0.57671
b_1	0.55359	0.12981	0.50757	0.15413	0.61617	0.10658	0.53301
b_2	3.00000	0.51007	3.17669	0.72971	3.19235	0.80688	2.98886
b_3	-0.09543	-0.00156	-0.04524	-0.00062	-0.03819	0.02090	-0.06755
b_4	2.00000	0.33023	2.05630	0.39695	2.05531	0.39496	1.92411
b_5	0.21939	0.23322	0.22300	0.24526	0.21661	0.24206	0.22492
MAPE		2.33002	0.72808	3.31722	0.72065	4.63232	0.70951

جدول (5)

يبين مقدرات المعلمات وقيم (MAPE) للانموذج الخطي (n=99)و(σ=2.8)

المعلمات	القيم الافتراضية	نسبة القيم الشاذة					
		%0		%10		%20	
		RWLS	TSRWLS	RWLS	TSRWLS	RWLS	TSRWLS
b_0	-0.58967	-0.09976	-0.59528	-0.10604	-0.60050	-0.11532	-0.59025
b_1	0.55359	0.11093	0.63683	0.11843	0.61536	0.18518	0.57009
b_2	3.00000	0.53019	3.20829	0.46186	3.32015	0.21430	3.15428
b_3	-0.09543	0.00221	-0.03689	-0.00228	-0.03351	-0.00134	-0.08774
b_4	2.00000	0.32362	2.01194	0.35368	2.04586	0.43763	1.98837
b_5	0.21939	0.03089	0.16399	0.03549	0.16195	0.02794	0.18399
MAPE		1.46551	0.56917	2.09326	0.57257	5.51793	0.61093

جدول (6)

يبين مقدرات المعلمات وقيم (MAPE) للانموذج الخطي (n=150)و(σ=2.8)

المعلمات	القيم الافتراضية	نسبة القيم الشاذة					
		%0		%10		%20	
		RWLS	TSRWLS	RWLS	TSRWLS	RWLS	TSRWLS
b_0	-0.58967	-0.09099	-0.61052	-0.09842	-0.61265	-0.11443	-0.60053
b_1	0.55359	0.10212	0.71024	0.09659	0.56555	0.20390	0.52555
b_2	3.00000	0.39380	2.84984	0.43459	3.16054	1.21381	3.22972
b_3	-0.09543	0.00222	-0.01274	0.00115	0.00152	-0.06961	-0.04153
b_4	2.00000	0.29288	2.04889	0.33081	2.03977	0.21812	2.03027
b_5	0.21939	0.03143	0.17677	0.03326	0.19438	0.05921	0.18483
MAPE		1.33867	0.48855	1.97953	0.43993	4.13394	0.48350

جدول (7)

يبين مقدرات المعلمات وقيم (MAPE) للانموذج الخطي (n=150)و(σ=3.8)

المعلمات	القيم الافتراضية	نسبة القيم الشاذة					
		%0		%10		%20	
		RWLS	TSRWLS	RWLS	TSRWLS	RWLS	TSRWLS
b_0	-0.18967	0.35757	-0.24937	0.06478	-0.17958	0.36072	0.14542
b_1	0.95399	0.50426	1.30312	0.69668	1.15106	-0.37492	1.04403
b_2	4.00000	1.91911	3.60429	3.67058	3.46570	5.36812	3.86291
b_3	-0.05543	-0.93359	0.06166	-0.62556	0.06786	-0.77626	0.27101
b_4	3.20000	1.85683	3.33551	2.80344	3.58650	1.94425	2.94753
b_5	0.31939	-0.54491	0.44426	-0.43289	0.53742	-0.56751	0.45229
MAPE		2.06686	1.72484	17.18343	1.91832	2.12114	0.96556

جدول (8)

يبين مقدرات المعلمات وقيم (MAPE) للانموذج الخطي (n=150)و(σ=3.8)

المعلمات	القيم الافتراضية	نسبة القيم الشاذة					
		%0		%10		%20	
		RWLS	TSRWLS	RWLS	TSRWLS	RWLS	TSRWLS
b_0	-0.18967	0.21806	-0.27633	0.22743	-0.15619	0.31284	-0.27430
b_1	0.95399	-0.01445	1.22266	0.54050	1.33031	0.13210	1.21104
b_2	4.00000	2.87842	4.11505	4.15535	4.07317	0.36426	4.51434
b_3	-0.05543	-0.75787	0.13183	-0.74349	0.05021	-0.70793	0.05838
b_4	3.20000	2.08079	3.24600	1.99884	3.48980	2.41249	3.29584
b_5	0.31939	-0.39628	0.50154	-0.51032	0.47713	-0.54425	0.48244
MAPE		1.17404	0.92528	2.34808	1.68164	2.80392	1.72045

جدول (9)
يبين مقدرات المعلمات وقيم (MAPE) للانموذج الخطي (n=150)و(σ=3.8)

المعلمات	القيم الافتراضية	نسبة القيم الشاذة					
		%0		%10		%20	
		RWLS	TSRWLS	RWLS	TSRWLS	RWLS	TSRWLS
b_0	-0.18967	0.24593	-0.25828	0.32740	-0.16219	0.36215	-0.29584
b_1	0.95399	0.14783	0.91580	-0.32794	1.33200	-0.52350	1.28217
b_2	4.00000	3.23435	4.21258	3.27610	4.65786	2.39555	4.16496
b_3	-0.05543	-0.99775	0.08797	-0.75890	0.10379	-0.80602	0.15596
b_4	3.20000	2.02001	3.32695	2.08507	3.36510	2.17344	3.39537
b_5	0.31939	-0.46044	0.49521	-0.55791	0.49113	-0.56479	0.48095
MAPE		1.28137	1.17512	2.12253	1.56968	2.11414	1.29731

10- **التحليل العملي لنتائج تجربة المحاكاة** :-

- **من الجداول اعلاه اظهرت نتائج المحاكاة التي تم تكرارها (1000) مرة بأن افضل طريقة حصينة هي (TSRWLS) كونها اعطت اقل متوسط للخطأ المطلق النسبي (MAPE)"عندما تقترب قيمة المقدر من القيمة الحقيقية للمعالم عند زيادة حجم العينة، وهذا يدل على صحة العمل الاحصائي، وتليها الطريقة الحصينة (RWLS).**

11- **الاستنتاجات** (Conclusions)

ان اختبار (Goldfeld-Quadt) المعدل الحصين ، قدم تحسينات كبيرة على الاختبارات الشائعة الاستعمال(الذي يتم فيه تطبيق طريقة (OLS) والتي تكون كما هو معروف حساسة تجاه القيم الشاذة وتعطي تقديرات متحيزه)، واظهر اداء رائع في الكشف عن مشكلة عدم تجانس التباين للخطأ في ظل وجود القيم الشاذة، أثبتت طريقة (TSRWLS) كفاءة عالية في تقدير معلمات إنموذج الانحدار الخطي المتعدد، في حالة وجود مشكلة عدم تجانس التباين وظهور القيم لشاذة،

12-**التوصيات**

1- استعمال اساليب احصائية حصينة مثل مقدرات (MM,LTS)، عند ظهور قيم شاذة في بيانات البحث كونها لاتتأثر بها.

2- يتم الاحتفاظ بسجل جيد للتجارب التي يتم تكرارها بكافة النسب المعتمدة في البحث ، ومحاولة ايجاد افضل تفسير ممكن لها والبحث عن افضل طريقة من بين كافة الطرائق عند وجود مشكلة عدم تجانس تباين الخطأ وظهور القيم الشاذة.

3- الاعتماد على طريقة (TSRWLS) في تقدير معلمات إنموذج الانحدار الخطي المتعدد في حالة كون البيانات تعاني من مشكلة عدم تجانس تباين الخطأ واحتوائها على القيم الشاذة ، لكونها حققت اقل قيمة لـ(MAPE) من بين الطرائق الحصينة للتقدير .

المصادر

1- **Amori,H.K.,& AL-Qaisi,(2002), "Advanced Econom Measurement theory & practice",Spectrum press, Baghdad, Iraq.**

2- **Ibrahim, Bassam Younis,(2002),"Econometrics",Dar AL-azza for publishing &distribution Khartoum, Sudan.**

3- **AL-Rawi, Kasha Mahmoud,(1987),"The entrance to the regression analysis, Dar AL-kutub printing & publishing directorate, Mosul university, Iraq.**

4- **AL-tayeb ezz eldin malek,(2008),:"of the introduction to econometrics", first edition, J town press, Khartoum, Sudan .**

5- **M. Habshah, S. Rana, A. H. M. R. Imon,(2008), "The Performance of Robust Weighted Least Squares in the Presence of Outliers and Heteroscedastic," *WSEAS Transition of Mathematics*, 8 (2008), pp.351 – 360.**

6- **Bross, I.D.J. (1961), "Outliers in Patternend Experiments: strategic Re-Appraisal". Technimetrics. 3,pp. 19-102.**

7- **Midi, H. A. B. S. H. A. H., Rana, S. O. H. E. L., & Imon, A. H. M. R. (2013),"On a Robust Estimator in Heteroscedastic Regression Model in the Presence of Outliers", Proceedings of the World Congress on Engineering 2013 Vol I, WCE 2013, July 3 - 5, , London, U.K..**

8- **Evans, Merran A., and Maxwell L. King. (1988),"A further class of tests for heteroscedasticity." Journal of Econometrics 37.2: pp.265-276.**

9- **Chen, C. (2002)." robust regression and outlier detection with the robustreg procedure". In Proceedings of the Proceedings of the Twenty-Seventh Annual SAS Users Group International Conference. pp. 265-27**

10- **Midi, H. A. B. S. H. A. H., Rana, S., & Imon, A. (2008)." A Robust Modification of the Gqldfeld-Quandt test for the detecation of Heteroscedasticity in the Presence of Outliers". In WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering (No. 4). World Scientific and Engineering Academy and Society. pp.277-283**

11- **Abdul M.Hamza. AL-nasser& Safaa younis, 2005, "Acomparison between the usual & holistic capabilities of mixed dual time series models from the lower ranks", Iraqi Journal of statistical sciences5.8,pp.1-19.**

Robust estimation of multiple linear regression parameters in the presence of a problem of heterogeneity of variance and outliers values

Shaimaa ibraheem khaleel
shaimaa595@yahoo.com

ghufran ismaeel kamal
ghufranka62@gmail.com

Received:19/7/2020 **Accepted : 16/9/2020** **Published :December / 2020**

Abstract:

Often Times, especially in practical applications, it is difficult to obtain data that is not tainted by a problem that may be related to the inconsistency of the variance of error or any other problem that impedes the use of the usual methods represented by the method of the ordinary least squares (OLS), To find the capabilities of the features of the multiple linear model, This is why many statisticians resort to the use of estimates by immune methods Especially with the presence of outliers, as well as the problem of error Variance instability, Two methods of horsepower were adopted, they are the robust weighted least square(RWLS)& the two step robust weighted least square method(TSRWLS), and their performance was verified by applying it to simulate & selection the best methods for estimation by using measures mean absolute percentage error (MAPE) to compare them, the results show the method of (TSRWLS) is the best.

Keywords: **the multiple linear model, Heteroscedasticity, outliers, robust weighted least square , the two step robust weighted least square method.**