AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

دراسة مقارنة بين طريقتي GM6.IDRGP(RMVN) وطريقة GM6 لتحليل انموذج الانحدار الخطّي المتعدّد في ظلّ وجود القيم الشاذّة باستعمال أسلوب المحاكاة

A comparative study between the two methods GM6.IDRGP(RMVN) and the GM6 method for analyzing the multiple linear regression model in the presence of outliers using the simulation method.

https://doi.org/10.29124/kjeas.1549.16

فرحان يعقوب عذاب[1] **أ . م . د طارق عزيز صالح[2]**

كلّية الإدارة والاقتصاد/ جامعة واسط

المستخلص

يُعدّ الانحدار الخطّي المتعّدد أحد أكثر طرائق التحليل الإحصائي استعمالا في العديد من المجالات العلمية. إذْ يتَمّ تقدير معلماته بالاعتماد على طريقة المربّعات الصغرى الاعتيادية. التي تعطي أفضل تقدير خطي غير متحّيز في حالة تحقّق فرضياتها. من أهم هذه الفرضيات هو أن يكون لها توزيع طبيعي للخطأ بمتوسط صفر وتباين ثابت. إذْا كانت البيانات لا تفي ببعض الافتراضات، فقد تكون تقديرات العيّنة والنتائج مُضلّلة. إذْ يكون نموذج الانحدار الخطّي حساس لظهور القيم الشاذّة ونقاط الرافعة. لذلك تَمّ تطوير تقنيات إحصائية قادرة على التعامل مع القيم الشاذّة أو اكتشافها. ما أدّى إلى ظهور العديد من طرائق البديلة عن طريقة (OLS) مثل M-Huber وS وLTS وMM التي تتَمّتع بكفاءة ونقطة انهيار عالية لكنّها تتأثّر بالـ (HLPs) التي تنتج عنها مشكلة الاخفاء والغمر. إنّ طريقة GM6 هي من طرائق التي وُضِعَتْ من أجل معالجة مثل هكذا مشاكل من خلال استعمال دالة وزن إلاّ أنّ دالة الوزن تعتَمّد على التشخيص الفردي الذي يعطي نتائج غير دقيقة. من أجل التغلّب على هذه المشكلة تَمّ اقتراح طرائق التشخيص الكُلّي مثل طريقة DRGPوIDRGP.RMVN. لذلك اقترحنا في هذه الدراسة توظيف طريقة IDRGP.RMVN في خوارزمية طريقة GM6. ومقارنتها مع بعض طرائق الانحدار الحصين عن طريق دراسة محاكاة في تحديد أفضل الطرائق.

المصطلحات / الانحدار الخطّي المتعّدد، نقاط الرفع العالية، مقدّر DRGP و IDRGP.RMVN ،طريقة GM6 و GM6.IDRGP(RMVN).

Abstract

Multiple linear regression is one of the most widely used statistical analysis methods in many scientific fields. Its parameters are estimated based on the ordinary least squares method. Which gives the best unbiased linear estimate if its assumptions are met. The most important of these assumptions is that it has a normal distribution of error with a mean of zero and a constant variance. If the data does not meet certain assumptions, the sample estimates and results may be misleading. The linear regression model is sensitive to the appearance of outliers and leverage points. Therefore, statistical techniques have been developed capable of dealing with or detecting outliers. This led to the emergence of many alternative methods to the OLS method, such as M-Huber, s, LTS, and MM, which have high efficiency and breakdown points, but are affected by HLPs, which result in the problem of Masking and Swamping. The GM6 method is one of the methods that was developed in order to treat such problems through the use of a weight function, but the weight function depends on the individual diagnosis, which gives inaccurate results. In order to overcome this problem, comprehensive diagnostic methods have been proposed, such as the DRGP and IDRGP.RMVN methods. Therefore, in this study, we proposed to employ the IDRGP.RMVN method in the GM6 method algorithm. And comparing it with some hippocampal regression methods through a simulation study in determining the best methods.

Key wards: Multiple Linear Regression, Leverage Point, DRGP Estimator and IDRGP.RMVN، methods GM6 and GM6.IDRGP(RMVN).

1- المقدّمة **Introduction**

اهتمّ العديد من المؤلّفين والباحثين في مجال تقدير المعلمات بإيجاد أفضل مقدّر لهذه المعلمات. لذلك، تَمّ تطوير العديد من الطرائق لهذا الغرض. يُعدّ نموذج الانحدار الخطّي المتعّدد أحد النماذج الرياضية الأكثر استعمالاً لتحليل بيانات الحقيقي. ان طريقة المربّعات الصغرى الاعتيادية (OLS) هي طريقة شائعة الاستعمال في الانحدار الخطّي المتعّدد بسبب خصائصها المثلى. ومع ذلك، إذْا كانت البيانات تحتوي على قيم شاذّة، فإنّ تقديرات OLS تصبح غير فعّالة. يؤدّي هذا إلى استعمال طرائق تقدير بديلة أكثر دقة وكفاءة من التقدير (OLS). يطلق على هذه الأساليب بالطرائق الحصين (Robust estimation). الهدف من الإحصائيات الحصينة هو طرائق إحصائية جديدة لا تتأثر بالقيم الشاذّة وتكون أكثر عملية للبيانات

التي لا يتَمّ توزيعها بشكل طبيعي. يمكن العثور على العديد من طرائق التقدير الحصينة مثل M وMM وLMS وLTS ومع ذلك، فإن هذه الأساليب تعاني من تأثير القيم الشاذّة ونقاط الرافعة العالية (HLPs). التي تسبّب بظهور ظاهرة الاخفاء والغمر Masking and Swamping التـــي تجعل من تقـــديرات هذه الطـــريقة غيـــر دقيقة، ممّا يؤدّي إلى عدم قدرتها على اكتشاف أنواع القيم الشاذّة جميع و (HLPs) بدقة في مجموعة البيانات. الغمر (swamping) هو تشخيص قيمة نظيفة على انها قيم شاذّة لكن بالحقيقة هي ليست بالقيمة الشاذّة. اما ظاهرة الاخفاء (masking) هي تشخيص أحد القيمة الشاذّة وتطغي على بقيت القيم الشاذّة التي لا يمكن تشخيصها بسبب تأثير القيمة المشخصة (Maroona and Yohai, 2006). إذْ قدم Imon (2002) مقياسًا جديدًا للقضاء على الاخفاء والغمر. السمية بمقياس (GP). ومع ذلك، لم يتمكن مقياس GP من تحديد العدد الدقيق لنقاط الرافع العالية ولا يزال يعاني من تأثيرات الإخفاء والغمر. لذلك تَمّ اقترح خوارزمية جديدة من قبل Habshah و آخرون عام (2009) تسمى Diagnostic Generalized Robust Scale(DRGP).الغرض من هذا المقياس هو التشخيص الدقيق وتقليل تأثير هذه الظاهرة. لم تقضي الطريقة الأخيرة تماما على تأثيرات ظاهرة الإخفاء والغمر، ولكنها قللت من هذا التأثير قدر الإمكان، لاسيمّا مع أحجام العيّنات الصغيرة وزيادة معدلات تلوث مجموعات البيانات. لمعالجة هذا الوضع في عام 2015، محمد وآخرون اقترحوا تحسين الطريقة السابقة من خلال طريقة جديدة تسمى Improvised DRGP (IDRGP) التي رفعت معدل التشخيص والتقليل من معدل تأثير الإخفاء والغمر إلى أدنى حد ممكن لكنها لم تتمكن من التخلص منها بشكل نهائي. قام كُلٌّ من (Uraibi and Alhussieny, 2022) و بالاعتماد على طريقة (RMVN) المقترحة من Olive and Hawkins (2010) بتطوير طريقة DRGP من خلال دمج RMVN مع RMD بدلاً من MVE وكذلك استعمال فكرة Mohammad *et.al* (2015) بتحسين الطريقة (DRGP) من خلال طريقة جديدة أطلق عليها Improvised DRGP (IDRGP) سمية IDRGP.RMVN لتكون أكثر كفاءة في كشف ظاهر الاخفاء والغمر وتقليلها والحد من تثيرها. ممّا تقدّم يمكن تلخيص مشكلة البحث بوجود قيم شاذّة ونقاط رافعة عالية توثّر على نتائج التقدير أمّا هدف هذا البحث هو إيجاد طرائق بديلة تَمتلك كفاءة و نقطة انهيار عالية ولا التأثر بهذه القيم. في القسم 2 يتَمّ وصف القيم الشاذّة واهمّ الطرائق المستخدمة في تشخيصها اما في القسم 3 يتَمّ شرح للطرائق الحصينة وفي القسم 4 تكون دراسة المحاكات ونتائج المستحصلة عليها من المحاكاة.

2- القيم الشاذّة في انموذج الانحدار الخطّي المتعّدد وطرائق تشخيصها

وقد صَنَّفَ كُلَّ من (Rousseeuw & Zomeren 1990)[12] القيم الشاذّة على ثلاث لما يخصّ موقعها وتأثيرها مشاهدات الشاذّة لمتغيّر (Y) وهي القيم الشاذّة العمودية (Vertical Outliers) او (VO) وهي المشاهدات الشاذّة أاو البعيدة في تنسيق متغير المعتَمّد او متغير الاستجابة (y). مشاهدات الشاذّة لمتغير (X) نقاط التأثير المرتفعة (HLPs) [3] وهي المشاهدات الشاذّة التي تكون في قيم المتغيّر (X) او البعيدة عن مجموع المتغيّرات التوضيحية وهي قيم خارجية توثّر على المتغيّر(X) والتي تؤدّي إلى انعطاف ميل الانحدار نحوها وتكون مرتبطة بالمتغيّرات التوضيحية. وتكون على نوعين: الأولى تسمّى (Bad-Leverage) او (BLPs) نقاط تأثير سيئة وهي المشاهدات التي تكون بعيدة عن خطَّ الانحدار أو مركز بيانات المتغيّر التوضيحي وتكون ذات تأثير على القيم المحسوبة تقدير متغير المعتَمّد. أمّا الثانية فتُسمّى (Good-Leverage) او(GLPs) نقاط التأثير الجيدة وهي المشاهدات التي تكون بعيدة عن بيانات المتغيّر التوضيحي. وتبقى على مقربة من خطّ الانحدار وتُسهم في كفاءة التقدير وليس لها تأثير على تقديرات المتغيّر المعتَمّد. وهناك نوع ثالث يكون في

المتغيّر المعتَمَد (Y) والمتغيّرات التوضيحية (X) معا[9 11]. ان لتشخيص القيم الشاذّة دور مهم في معرفة اين يقع تأثير تلك القيم بالتحديد ليتّم التعامل معها بحرفية إذْ تَمّ وضع طرائق للكشف عنها سواء أكانت بالمحور العمودي ام الافقي. لذلك اهتمّ الباحثون كثيرا في قضية الكشف عن HLP، مثلا قدم (Hadi) عام (1992) طريقته المعروفة Hadi's Potential للكشف عن هذه المشاهدات المؤثّرة على التقديرات .

Hat Matrix -1-2

هي مصفوفات الوزن وبذلك تُسمّى بـ (Hat Matrix) [5] ويَتُمّ ايجاد هذه المصفوفة في تحديد صفوف المشاهدات لـ (X) التي تحتوي القيم الشاذّة وتُعرّف تحليل الانحدار كمقياس للكشف عن وجود القيم الشاذّة وتكون الصيغة لها كالآتي :

$$H = \mathrm{x}(\mathrm{x}'\mathrm{x})^{-1}\mathrm{x}' \qquad (1)$$

إذْ أنّ عناصر القطر الرئيس لمصفوفة الوزن (H) تكون صيغتها كالآتي :

$$\mathrm{h}_{ii} = \mathrm{x}_i(\mathrm{x}'\mathrm{x})^{-1}\mathrm{x}'_i \qquad (2)$$

لهذه العناصر خصائص مفيدة ولاسيّما عندما تكون قيمها تتراوح ما بين (1 و 0) وتكون مجموع (h_{ii}) يساوي (p). ومن الممكن إضافة الاثبات(h_{ii})المؤشر على احتوائها للقيم الشاذّة لـ (LP) للحالة (i^{th})إذْ تكون مقياس للمسافة بين قيم لـ (x) للحالات جميعها (n)، لذلك تشير القيمة الكبيرة لـ(h_{ii})إلى أنّ الحالة (i^{th})بعيد عن مركز مشاهدات المتغيّر (x) [11]

2-2-البواقي القياسية المحذوفة Deleted Studentized Residuals

طريقة أخرى للتشخيص العمودي هي بقايا ستيودنت المُتبقّي القياسي أي إنّ المتبقّي القياسي هو مجرد بقايا محذوفة مقسومة على الانحراف المعياري لأخطاء$\hat{\sigma}$ مضروب في $\sqrt{1-h_{ii}}$. h_{ii} مصفوفة القطرية لمصفوفة القبعة [14]

$$t_i = \frac{e_i}{\hat{\sigma}\sqrt{1-h_{ii}}} \qquad i = 1,2,\dots,n \qquad (3)$$

إذْ إنّ $|t_i|$ إذا كانت أكبر من 2.5 فتعدُّ قيمة شاذّة.

3-طرائق التقدير الحصينة

عمل كثير من الباحثين في مجال الانحدار على طرائق التقدير البديلة للمربّعات الصغرى الاعتيادية عندما لا تتحقّق طبيعة البيانات من الافتراضات الأساس. إذْ ناقش Alma (2011) [4] التقدير الحصين هو طريقة مهمّة لتحليل البيانات الملوّثة بالقيم الشاذّة، الهدف الرئيس هو توفير مقدّرات غير متحيّزة ومرغوبة بوجود تلك القيم. من خلال استعمال خصائص الكفاءة ونقاط

الانهيار ونقاط الرافعة العالية لتحديد أداء التقنيات الحصينة. يتمثّل أحد أهداف المقدّرات الحصينة في الحصول على نقطة انهيار عالية محدودة للعّينة والتي حدّدها Donoho & Huber ، في عام 1983 [7]. تَمّ تصميم طرائق التقدير الحصينة بإذْ لا تتأثر بشكل كبير بالقيم الشاذّة. سوف نتطرق لبعض طرائق التقدير الاكثر شيوعا بشيء من التفصيل.

3-1-نقطة الانهيار Break down Point

تُعدُّ نقطة الانهيار من المعايير المهمة التي تقيس حصانة المقدّرات. وتكون من الركائز الرئيسية في الحصانة إذْ تُعدُّ نقطه الفصل بين المقدّر ذي الفائدة من عديم الفائدة فكُلّما كانت نقطه انهيار المقدّر عالية كلما كانت المقاومة أكبر. وقد وصفها الباحث (Hampel) عام (1971) بأنّها (درجة حصانة المقدّر عندما تحتوي البيانات على قيم شاذّة). وعرّفها Donoho و Huberعام (1983)، Leroy,Rousseeuw عام (1987) [14] و Maronna و آخرون عام (2006) بأنّها هي أصغر كمّية من التلّوث التي قد تتسبب في أنْ يأخذ المقدّر قيما شاذّة كبيرة بشكل تعسّفي .كما عرّفها (Alma, 2011) [4] بأنّها النقطة أو النسبة المئوية المحدّدة للتلّوث في البيانات التي يَتَمّ عندها إغراق أي إحصائيات اختبار أولاً.

وعليه يمكن وصف نقطة الانهيار بأنّها الحدّ الذي يوضّح مدى مقاومة المقدّر للقيم الشاذّة ويمكن التعبير عنها بالشكل الرياضي الآتي التي تشير فيه كُلٌّ من T إلى مقدّرات الانحدار و Z إلى بيانات العيّنة و n إلى حجم العيّنة و m إلى النقاط الشاذّة :

$$BP\ (T,Z) = min\ \{\frac{m}{n} : bais\ (m;T,Z) = \infty\} \qquad \dots (4)$$

3-1- طريقة IDRGP(RMVN)

هذه الطريقة المقترحة من قبل (Uraibi;Haraj;2022) [10] تدعو إلى توظيف مصفوفة الموقع والقياس Reweighted Multivariate Normal Estimators (RMVN) بدلاً من MVE في خوارزمية DRGP(MVE). فضلً عن الحاق الخطوات التي أضافها Mohammad al et 2015 للتأكّد من صحة التشخيص. ان مصفوفة RMVN هي من اقتراح Olive and Hawkins (2010)الذي اقترح RMVN لإعادة ترجيح مقدّر متعدد المتغيّرات الطبيعي من خلال خوارزمية سريعة ومّتسقة وتملك نقطة انهيار عالية. إذْ قام الباحثان بالتركيز على خوارزمية ذات الخطوات الخمسة في المرحلتين الأولى و الثانية التي تتطلب حساب مصفوفة DGK للتباين والتباين المشترك الذي اقترحها(1981).Devlin al et و مصفوفة الموقع و القياس الـ Median Ball(MB) المقترحة من قبل Olive (2004). وبعدها يتَمّ حساب مقدّر Fast consistent and high breakdown (FCH) الذي اقترحه Olive and Hawkins (2010) وفي الأخير يتَمّ إعادة الترجيح مرة أخرى للحصول على المقدّر النهائي لهذه الطريقة. لقد وَظّفَ الباحثان (Uraibi;Haraj;2022) مقدّرات هذه الطريقة لاقتراح خوارزمية جديدة بعنوان DRGP(RMVN) التي يمكن وصفها كالآتي:

1- حساب الوسط الحسابي ومصفوفة التباين والتباين المشترك كمقدّرين تبدأ بهما الخوارزمية إذْ يرمز لهما بالرمزين $(T_{0,1}, C_{0,1})$ على التوالي.

2- استخراج مصفوفة التباين والتباين المشترك DGK: بعده يتّمّ حساب مسافة مهالانوبيز (MD) بالاعتماد على المقدّرات في الخطوة (1) ثم نرتب القيم MD تصاعديا للحصول على وسيط قيم MD كعتبة للحصول على مصفوفة جيدة من المشاهدات تناظر تلك الصفوف التي قيم MD لها أقل من قيمة العتبة. ونقوم بتكرار الخطوتين السابقتين خمس مرات وفي كُلّ مرة تحسب قيم MD لكامل مجموعة البيانات للحصول على $(T_{5,1}, C_{5,1})$ كمقدّرين للموضع والقياس حصينين.

3- إيجاد مقدّري الموضع والقياس لمصفوفة (MB) خوارزمية هذه المصفوفة تبدأ مع مقدّر الوسيط ومصفوفة الوحدة كمصفوفة للتباين والتباين المشترك كمراحل بداية لها. تختلف هذه المصفوفة عن سابقتها في أنّها تستعمل في حساب MD الوسيط وليس الوسط الحسابي ونعيد الخطوات الخمس السابقة نفسها للحصول على $(T_{5,2}, C_{5,2})$

4- مقدّرات FCH: الخطوة الأولى في خوارزمية هذه الطريقة هي حساب المسافة التقليدية بين مقدّري الموقع في الخطوتين السابقتين $(T_{5,1}, T_{5,2})$ على التوالي. فإذا كانت قيمة المسافة أقل من اخر قيمة للعتبة (وسيط مسافة مهالانوبيز في الخطوة الخامسة) في مصفوفة MB تبدأ FCH لاختيار أحد هذين المقدّرين كالآتي:

$$T_{FCH} \begin{cases} T_{5,2} & if \quad \sqrt{|C_{5,1}|} < \sqrt{|C_{5,2}|} \\ T_{5,2} & otherwise \end{cases} \qquad (5)$$

يتّمّ حساب مقدّر القياس أيضا بالاعتماد على الشرط نفسه ولكنّ المقدّر هنا يضرب بثابت أو معامل تصحيح كالآتي:

$$C_{FCH} \begin{cases} \dfrac{MED\left(MD_i^2(T_{5,1}, C_{5,1})\right)}{\chi^2_{(p,0.5)}} \times C_{5,1}, & if \sqrt{|C_{5,1}|} < \sqrt{|C_{5,2}|} \\ \dfrac{MED\left(MD_i^2(T_{5,2}, C_{5,2})\right)}{\chi^2_{(p,0.5)}} \times C_{5,2}, & otherwise \end{cases} \qquad (6)$$

5- مقدّر RFCH: بعد الحصول على المقدّرين $(T_{1,RFCH}, C_{1,RFCH})$ من الخطوة السابقة يَتُمّ حساب مسافة مهالانوبيز مرة أخرى لكامل المشاهدة ثُمّ نكرّر الخطوة السابقة لإيجاد $(T_{2,RFCH})$ مع مراعاة معامل التصحيح و كالآتي:

$$C_{2,RFCH} = \frac{MED\left(MD_i^2(T_{1,RFCH}, C_{1,RFCH})\right)}{\chi^2_{(p,0.975)}} \times C_{1,RFCH}, \qquad (7)$$

6- مقدّري الـ RMVN: خوارزمية هذه الطريقة تسعى أولا إلى إيجاد مصفوفة مشاهدات جديدة للمتغيرات بالاعتماد على مقدّر الموقع و القياس السابقة و كالآتي:

افرض ان $S^0 = \sum_{j=1}^{n_1} X_k$ إذْ ان

$$X_k = \left\{X_k : MD_i(T_{2,RFCH}, C_{2,RFCH}) \leq MED\left(MD_i(T_{2,RFCH}, C_{2,RFCH})\right)\right\}$$

Where $k = 1, \dots, n_1 \; and \; i = 1, \dots, n$

افرض ان $Q^{(1)} = min\{0.5 \times 0.975 \times n/s^0, 0.995\}$ للحصول على مقدّر القياس الأول لـ RMVN وكالآتي

$$C_{RMVN}^{(1)} = \frac{MED\left(D_i^2(T_{RFCH}, C_{RFCH})\right)}{\chi^2_{(p,Q^{(1)})}} \qquad (8)$$

و لتحسين أداء هذا المقدّر اقترح Olive and Hawkins (2010) إعادة ترجيح مقدّر القياس السابق (8) مع الجزء الثاني من البيانات n_2 للحصول على المقدّر النهائي $C_{RMVN}^{(2)}$.

ممّا تقدّم نصل إلى أنّ مقدّر معلمة الموقع $T_{RMVN} = T_{RFCH}$ وأنّ مقدّر معلمة القياس هو $C_{RMVN}^{(2)}$.

7- حساب مسافة مهالانوبيز الحصينة:

$$RMD_i(RMVN) = \sqrt{(X - T_{RMVN}(X))' {C_{RMVN}^{(2)}}^{-1} (X - T_{RMVN}(X))}\,, \qquad (9)$$

8- تحديد المشاهدات المشكوك بها بالمقارنة مع العتبة، $RMD_i(RMVN) > \sqrt{\chi^2_{p,0.95}}$.

و وضعها في المصفوفة D الموصوفة في طريقة GP ،أما بقية المشاهدات فتوضع في المصفوفة R.

9- نحسب p_{ii} كما في المعادلة (10) لمعرفة المشاهدات في المصفوفة D هل تحتوي على LP أم لا، فإذا تَمّ تشخيصها تبقى في المصفوفة D وبخلافه تنقل إلى المصفوفة R.

$$p_{ii} = \begin{cases} w_{ii}^{(-D)} & \forall i \in D\,, \\ \dfrac{w_{ii}^{(-D)}}{1 - w_{ii}^{(-D)}} & \forall i \in R\,, \end{cases} \qquad (10)$$

أمّا الخطوات التي اقترحها Mohammad 2015) قد استخدمها (Uraibi;Sawsan;2022) لتحسين أداء الطريقة المقترحة التي اطلق عليها IDRGP(RMVN) فلاحظ Mohammad 2015) أنّ تشخيص HLP في الخطوة الثانية من خوارزمية DRGP(RMVN) التي وضعها في مصفوفة جزئية أطلق عليها D لم يَتُمّ التحقق منها بشكل صحيح. لذلك اقترح إضافة خطوة جديدة إلى الخوارزمية من خلال التشخيص باستعمال hat matrix. ثُمّ مقاطعة التشخيصين للتحقق من التشخيص الأول، إذْ يقارن فيها بين ما تَمّ تشخيصه كـ HLP بوصفها نتيجة نهائية للخوارزمية الأولى و التي افترضها

بالمصفوفة الجزئية D_2 و قارنها بما تَمّ تشخيصه في الخطوة الثانية التي انتجت المصفوفة الجزئية D أي المقارنة بين D و D_2 كالآتي:

- الحالة الأولى: إذْا كانت المشاهدات المشخّصة كـ HLP هي ذاتها في D و D_2 فالخطوة الأخيرة للخوارزمية الإعلان عن هذا التشخيص وتتوقّف.
- الحالة الثانية: إذْا كان عدد المشاهدات المشخصة كـ HLP في D_2هي أكبر من عدد تلك المشخصة في D، عندئذ تعمل الخوارزمية على إعادة المشاهدات التي لم يَتُمّ تشخيصها في D_2 إلى المصفوفة R واحدة تلو الأخرى على ان يتَمّ حساب p_{ii} لكل واحدة منها في كل مرة و تقارن بقيمة نقطة القطع أو العتبة في المعادلة (11) فإذا تجاوزت قيمة العتبة تضاف إلى المشاهدة المشخصة و بخلافه تبقى مع المشاهدات النظيفة.

$$Median\,(P_{ii}) \pm 3MAD\,(P_{ii}) \qquad (11)$$

- الحالة الثالثة: إذْا كان عدد المشاهدات المشخّصة كـ HLP في D_2هي أقل من عدد تلك المشخّصة في D، أي تَمّ تشخيص مشاهدات جديدة لم تشخص سابقا، عندئذ تعمل الخوارزمية على الدمج بين D و D_2 ثُمّ يتُمّ حسابp_{ii} لكُلّ مشاهدة في المصفوفة الجديدة و تقارن بقيمة العتبة في المعادلة (2-23) فإذا تجاوزت قيمة العتبة تضاف إلى المشاهدة المشخصة و بخلافه تبقى مع المشاهدات النظيفة.

3-2- طريقة مقدّر GM6

ان طريقة GM6 من الطرائق الحصينة شاسعة الاستعمال لأنّها تحتوي على نقطة انهيار عالية تساوي تقريبًا 50%[2] وكفاءة عالية وتأثير محدود في النموذج العادي والتي تَمّ اقتراحها من قبل العالمان (Coakley و Hettmansperger) في عام (1993)[6] .و تستعمل المربّعات الأقل تشذبا كمقدّر اولي في خوارزميةGM6. يعرف مقدّر GM بأنه حل المعادلات العادية التي يَتُمّ تقديمها من خلال:

$$\sum_{i=1}^{n} \pi_i \Psi \left\{ \frac{y_i - x^t{}_i \hat{\beta}}{\hat{\sigma} \pi_i} \right\} \boldsymbol{x}_i = 0 \qquad \ldots (12)$$

إذْ $\psi = p'$ هو مشتق من دالة إعادة تنازلي (دالة الوزن).

π_i هو عنصر الوزن الأولي للمصفوفة القطرية W $i = 1,2,\ldots\ldots,n$

$\hat{\sigma}$ هو تقدير المقياس.

$\widehat{\beta}$ هو متجّه تقديرات المعلمات.

إنّ الهدف الرئيس لمقدّرات GM [8] يعتمد على تقليل نقاط الرفع العالية التي تحتوي على مخلّفات كبيرة. في هذا الصدد يجب اكتشاف القيم الشاذّة ونقاط الرفع العالية في بداية الامر باستعمال طرائق تشخيص HLPs وبمجرد اكتشافها. بجب تقليل اثارها لزيادة كفاءة مقدّرات GM. يستعمل GM6 مسافات مهالانوبيز الحصينة (RMD) استنادا إلى الحد الأدنى من حجم Ellipsoid(MVE) او الحد الأدنى من محدد التغاير (MCD) للكشف عن نقاط التأثير العالية.

قيمة النقطة الفاصلة كما اقترحها ((Rousseeuw and Leroy (1987) هي $\sqrt{\chi^2_{p,0.5}}$. سيتَمّ تحديد المشاهدة (i) التي تتوافق مع RMD_i والتي تتجاوز نقطة القطع كنقطة رفع عالية.

خوارزمية مقدّر GM6

الخطوة 1: احسب القيم المتبقية (r_i) بناءً على مقدّر المربّعات الأقل تشذيبًا (LTS).

الخطوة 2: حساب المقياس المقدّر (σ) للبواقي،

$s = (1.4826)(1 + \frac{5}{(n-p-1)})(median(|r_i|))$ إذْ يتَمّ الحصول على r_i من الخطوة 1.

الخطوة 3: حساب القيم المتبقية المعيارية (e_i). إذْ $e_i = \frac{r_i}{s}$.

الخطوة 4: نحسب الوزن الأولي ، والمشار إليه بالرمز π_i. إذْ

$\pi_i = min\left\{1, \frac{\chi^2{}_{0.95,k}}{RMD\,(MVE)}\right\}$، إذْ RMD وMVE هي مسافة مهالانوبيز القوية

الخطوة 5: حساب دالة التأثير المحدود، $t_i = \frac{e_i}{\pi_i}$.

الخطوة 6: احسب خطوة واحدة نيوتن رافسون للحصول على تقديرات GM6.

3-3- طريقة المقترحة GM6.IDRGP(RMVN)

من خلال اقتراح (Uraibi; Haraj) [10] في عام 2022 لطريقة IDRGP(RMVN) والتي استطاعت أنْ تقلّل تأثير ظاهرة الاخفاء والغمر في البيانات. نقوم بتوظيف هذه الطريقة في اطار خوارزمي واحد مع GM6 مع تعديل خوارزمية GM6 لتستوعب الاوزان المستخدمة كافة في الطريقة كالآتي :

1- حساب بواقي (r_i) الانحدار بناء على مقدّر حصين (MM).

2- حساب المقدّر (S) للبواقي.

$$s = (1.4826)(1 + \frac{5}{(n-p-1)})(median(|r_i|))$$

3- حساب معلمات المقدّر الحصين $\hat{\beta}_{Rob}$.

4- إيجاد معلمتي الموقع والقياس باستعمال RMVN.

5- حساب مسافة مهالانوبيز الحصينة RMD المعادلة (10) و تحديد المشاهدات المشكوك بها و وضعها في المصفوفة D الموصوفة في طريقةGP ،أمّا بقية المشاهدات فتوضع في المصفوفةR.

6- نحسب p_{ii} كما في المعادلة (9) لمعرفة المشاهدات في المصفوفة D هل تحتوي على LP أم لا، فإذا تَمّ تشخيصها تبقى في المصفوفة D و بخلافه تنقل إلى المصفوفةR.

7- أيّة قيمة من قيمp_{ii} تتجاوز قيمة القطع في معادلة (10) تعد نقطة رافعة.

8- تضرب المشاهدات كُلّها بوزن أولي ل π_i حيث:

$$\pi_i = min\left\{1, \frac{p_{ii}}{RMD(RMVN)}\right\},$$

9- نجد قيم المعلمات المقدّرة للبيانات المرجحة باستعمال طريقة المربّعات الصغرى $\hat{\beta}_{OLS}$

إذْا كانت $|\hat{\beta}_{Rob} - \hat{\beta}_{OLS}| \leq 0.0001$ تتوقف الخوارزمية وتكون $\hat{\beta}_{OLS}$ هي معلمة المقدّر الحصينة لطريقة ة لطريقة GM6.

10- إذا لم يتحقّق الشرط السابق نحسب بواقي الانحدار r_i^* و العودة إلى الخطوة 2.

4- دراسة أسلوب المحاكاة

ليكن لدينا نموذج الانحدار الخطّي المتعّدد الآتي :-

$$y = X\beta + e \qquad (15)$$

حيث أنّ X هي مصفوفة المتغيرات التوضيحية ذات بعد $n \times (p+1)$ المولّدة من توزيع طبيعي متعدّد المتغيّرات بمتوسّط قيمته صفراً وتباين $\sigma = \rho^{|i-j|}$ أيّ أنّ

$$X \sim MVTN(0, \rho^{|i-j|})$$

إذْ $p = \{4\}$ بمعنى (4) متغيّرات في دراسة المحاكاة. و$n = \{30,50,90,130\}$

و $\rho = 0.5$

β هي متجّه الوحدة لمعلمات هذا النموذج مع الحّد الثابت كالآتي:

$$\beta = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}_{px1}$$

أما e فهو متجّه الأخطاء العشوائية للنموذج بمتوسّط قيمته صفر وانحراف معياري قيمته 2.

قام الباحث بتلويث بيانات كل دراسة محاكاة بنسب مختلفة $(0.15\ ,0.10\ ,0.05)$ من القيم الشاذّة والنقاط الرافعة. إذْ تَمّ توليد عينة بحجم n مشاهدة وتلويث حد الخطأ العشوائي بـ(α) من القيم الشاذّة من خلال توليد مشاهدات تتبع توزيع χ^2 بدرجة حرية (25). اما مصفوفة المتغيّرات التوضيحية فقد تَمّ تلويثها بـα من BLP بوجود HLP واحدة. وذلك من خلال ضرب الصفوف الأولى من المتغيّر الثاني إلى المتغيّر الرابع بالرقم 10و بما يتوافق مع نسب التلويث، اخيرا تضرب أعظم قيمة للمتغيّر الأول وما يناظرها في المتغيّرy بالرقم 50.

استعمل الباحث عددا من المعايير التي ممكن توظيفها للمقارنة. المعيار الأول هو القيمة المطلقة لتحيز المعلمات المقدّرة عن المعلمات المفترضة في دراستنا هذه. المعيار الثاني هو معيار متوسط مربعات الخطأ MSE والثالث هو متوسّط انحرافات القيم المطلقة لبواقي الانحدار عن متوسّطها MAE وأخيرا الوسيط المعياري لانحرافات القيم المطلقة لبواقي الانحدار عن وسيطها MAD. عليه تكون الطريقة التي تجمع فيها أقل قيمة المعايير كُلّها هي الطريقة الأفضل.

الجدول رقم (1) متوسّط تقدير معلمات طرائق التقدير GM6,GM6.IDRGP.RMVN لهذه المعلمات لبيانات بحجم (30,50,90,130) مشاهدة ملوثة بـ (0.15,0.10,0.05) من القيم الشاذّة و النقاط الرافعة بوجود نقطة رافعة عالية جدا.

	α	n	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$

GM6	**0.05**	**30**	**1.817**	**3.755**	**-0.232**	**0.647**	**0.715**
GM6.IDRGP.RMVN			**0.98**	**0.907**	**0.976**	**0.977**	**0.974**
GM6		**50**	**1.614**	**2.798**	**0.291**	**0.683**	**0.717**
GM6.IDRGP.RMVN			**0.846**	**1.158**	**1.058**	**0.868**	**0.961**
GM6		**90**	**1.359**	**2.467**	**0.283**	**0.593**	**0.591**
GM6.IDRGP.RMVN			**0.76**	**0.904**	**1.11**	**0.88**	**0.998**
GM6		**130**	**1.139**	**1.902**	**0.299**	**0.623**	**0.621**
GM6.IDRGP.RMVN			**0.696**	**1.027**	**0.939**	**0.992**	**0.932**
GM6	**0.10**	**30**	**2.293**	**3.786**	**-0.207**	**0.695**	**0.288**
GM6.IDRGP.RMVN			**1.021**	**0.895**	**0.968**	**1.007**	**0.964**
GM6		**50**	**2.102**	**3.25**	**-0.044**	**0.576**	**0.339**
GM6.IDRGP.RMVN			**0.959**	**1.061**	**1.044**	**0.793**	**1.028**
GM6		**90**	**1.605**	**2.97**	**0.029**	**0.441**	**0.316**
GM6.IDRGP.RMVN			**0.851**	**1.122**	**0.931**	**0.986**	**0.931**

GM6		**130**	**1.5**	**2.437**	**0.073**	**0.352**	**0.318**
GM6.IDRGP. RMVN			**0.772**	**0.959**	**0.914**	**1.01**	**0.968**
GM6	**0.15**	**30**	**2.28**	**2.937**	**0.463**	**0.435**	**0.563**
GM6.IDRGP. RMVN			**1.028**	**1.042**	**0.866**	**0.997**	**0.908**
GM6		**50**	**1.881**	**1.948**	**0.288**	**0.501**	**0.307**
GM6.IDRGP. RMVN			**0.937**	**1.081**	**0.816**	**0.919**	**0.851**
GM6		**90**	**1.553**	**1.749**	**0.331**	**0.307**	**0.4**
GM6.IDRGP. RMVN			**0.869**	**1.027**	**0.982**	**0.921**	**1.004**
GM6		**130**	**1.416**	**2.014**	**0.322**	**0.381**	**0.337**
GM6.IDRGP. RMVN			**0.754**	**1.01**	**0.945**	**0.986**	**0.909**

الجدول (2) متوسّط التحيّز معلمات طرائق التقدير GM6,GM6.IDRGP.RMVN لبيانات بحجم (30,50,90,130) مشاهدة ملوثة بـ (0.15,0.10,0.05) من القيم الشاذّة و النقاط الرافعة بوجود نقطة رافعة عالية جدا.

	α	N	$\overline{bias}(\hat{\beta}_0)$	$\overline{bias}(\hat{\beta}_1)$	$\overline{bias}(\hat{\beta}_2)$	$\overline{bias}(\hat{\beta}_3)$	$\overline{bias}(\hat{\beta}_4)$
GM6	**0.05**	**30**	**2.514**	**29.019**	**4.738**	**1.622**	**1.862**
GM6.IDRGP. RMVN			**0.093**	**0.394**	**0.535**	**0.327**	**0.294**
GM6		**50**	**0.718**	**9.824**	**1.42**	**0.458**	**0.349**

GM6.IDRGP. RMVN			**0.189**	**0.478**	**0.302**	**0.3**	**0.198**
GM6		**90**	**0.231**	**4.87**	**0.914**	**0.343**	**0.316**
GM6.IDRGP. RMVN			**0.087**	**0.104**	**0.123**	**0.145**	**0.1**
GM6		**130**	**0.052**	**1.454**	**0.598**	**0.274**	**0.234**
GM6.IDRGP. RMVN			**0.11**	**0.091**	**0.084**	**0.104**	**0.069**
GM6	**0.10**	**30**	**2.185**	**22.34**	**3.634**	**1.65**	**1.278**
GM6.IDRGP. RMVN			**0.095**	**0.403**	**0.567**	**0.619**	**0.356**
GM6		**50**	**1.786**	**16.148**	**1.975**	**0.833**	**0.778**
GM6.IDRGP. RMVN			**0.058**	**0.283**	**0.29**	**0.359**	**0.186**
GM6		**90**	**0.477**	**9.813**	**1.308**	**0.481**	**0.542**
GM6.IDRGP. RMVN			**0.052**	**0.131**	**0.146**	**0.17**	**0.115**
GM6		**130**	**0.304**	**3.967**	**0.977**	**0.459**	**0.515**
GM6.IDRGP. RMVN			**0.124**	**0.773**	**0.588**	**0.238**	**0.276**
GM6	**0.15**	**30**	**4.133**	**22.238**	**0.714**	**0.609**	**0.597**
GM6.IDRGP. RMVN			**0.129**	**0.332**	**0.265**	**0.326**	**0.207**

GM6		**50**	**1.006**	**7.863**	**0.607**	**0.367**	**0.63**
GM6.IDRGP. RMVN			**0.063**	**0.263**	**0.191**	**0.139**	**0.229**
GM6		**90**	**0.354**	**2.506**	**0.507**	**0.543**	**0.429**
GM6.IDRGP. RMVN			**0.041**	**0.11**	**0.074**	**0.088**	**0.066**
GM6		**130**	**0.222**	**2.563**	**0.494**	**0.419**	**0.476**
GM6.IDRGP. RMVN			**0.081**	**0.098**	**0.067**	**0.066**	**0.071**

الجدول (3) المتوسّط العام لمتوسّط مربعات الخطأ معلمات طرائق التقدير GM6,GM6.IDRGP.RMVN لهذه المعلمات لبيانات بحجم (30,50,90,130) مشاهدة ملوثة بـ (0.15,0.10,0.05) من القيم الشاذّة و النقاط الرافعة بوجود نقطة رافعة عالية جدا.

	α	N	$\overline{Mse}(\hat{\beta}_0)$	$\overline{Mse}(\hat{\beta}_1)$	$\overline{Mse}(\hat{\beta}_2)$	$\overline{Mse}(\hat{\beta}_3)$	$\overline{Mse}(\hat{\beta}_4)$
GM6	**0.05**	**30**	**3.526**	**29.023**	**5.391**	**2.39**	**2.66**
GM6.IDRGP. RMVN			**0.174**	**0.59**	**0.743**	**0.52**	**0.455**
GM6		**50**	**1.097**	**9.827**	**1.633**	**0.652**	**0.531**
GM6.IDRGP. RMVN			**0.093**	**0.245**	**0.313**	**0.292**	**0.221**
GM6		**90**	**0.362**	**4.871**	**0.961**	**0.396**	**0.365**
GM6.IDRGP. RMVN			**0.098**	**0.134**	**0.156**	**0.179**	**0.126**
GM6		**13**	**0.107**	**1.454**	**0.615**	**0.294**	**0.252**

GM6.IDRGP. RMVN		**0**	**0.115**	**0.108**	**0.104**	**0.123**	**0.086**
GM6	**0.10**	**30**	**3.399**	**22.346**	**4.214**	**2.299**	**1.634**
GM6.IDRGP. RMVN			**0.185**	**0.597**	**0.805**	**0.841**	**0.532**
GM6		**50**	**2.34**	**16.15**	**2.17**	**1.025**	**0.949**
GM6.IDRGP. RMVN			**0.094**	**0.366**	**0.384**	**0.451**	**0.255**
GM6		**90**	**0.649**	**9.814**	**1.344**	**0.518**	**0.57**
GM6.IDRGP. RMVN			**0.065**	**0.166**	**0.183**	**0.205**	**0.144**
GM6		**130**	**0.387**	**3.968**	**0.991**	**0.475**	**0.527**
GM6.IDRGP. RMVN			**0.072**	**0.114**	**0.144**	**0.101**	**0.102**
GM6	**0.15**	**30**	**5.237**	**22.243**	**1.029**	**0.925**	**0.931**
GM6.IDRGP. RMVN			**0.225**	**0.512**	**0.408**	**0.476**	**0.335**
GM6		**50**	**1.375**	**7.864**	**0.714**	**0.445**	**0.708**
GM6.IDRGP. RMVN			**0.099**	**0.334**	**0.249**	**0.19**	**0.282**
GM6		**90**	**0.464**	**2.507**	**0.523**	**0.557**	**0.444**

GM6.IDRGP. RMVN			**0.055**	**0.141**	**0.1**	**0.112**	**0.09**
GM6		**130**	**0.292**	**2.563**	**0.503**	**0.427**	**0.484**
GM6.IDRGP. RMVN			**0.088**	**0.113**	**0.081**	**0.08**	**0.086**

الجدول(4) قيم MSE,MAE,MAD للطرائق GM6, GM6.IDRGP.RMVN لبيانات بحجم (30,50,90,130) مشاهدة ملّوثة بـ (0.15,0.10,0.050) من القيم الشاذّة و النقاط الرافعة بوجود نقطة رافعة عالية جدا.

Method	α	**n**	**MSE**	**MAE**	**MAD**
GM6	**0.05**	**30**	**23.224**	**2.807**	**3.024**
GM6.IDRGP.RMVN			**2.128**	**1.123**	**1.588**
GM6		**50**	**16.919**	**2.507**	**2.893**
GM6.IDRGP.RMVN			**1.607**	**1.045**	**1.557**
GM6		**90**	**8.936**	**1.8**	**1.961**
GM6.IDRGP.RMVN			**0.996**	**0.859**	**1.241**
GM6		**130**	**6.506**	**1.544**	**1.605**
GM6.IDRGP.RMVN			**0.71**	**0.732**	**1.037**
GM6	**0.10**	**30**	**33.138**	**3.453**	**4.068**
GM6.IDRGP.RMVN			**2.335**	**1.17**	**1.608**
GM6		**50**	**22.163**	**2.896**	**3.282**
GM6.IDRGP.RMVN			**1.7**	**1.076**	**1.549**

GM6		**90**	**11.044**	**2.147**	**2.215**
GM6.IDRGP.RMVN			**1.187**	**0.935**	**1.39**
GM6		**130**	**8.412**	**1.904**	**1.852**
GM6.IDRGP.RMVN			**0.904**	**0.82**	**1.242**
GM6	**0.15**	**30**	**59.576**	**4.712**	**5.274**
GM6.IDRGP.RMVN			**5.252**	**1.705**	**2.257**
GM6		**50**	**35.835**	**3.869**	**4.215**
GM6.IDRGP.RMVN			**2.177**	**1.201**	**1.68**
GM6		**90**	**15.088**	**2.697**	**2.631**
GM6.IDRGP.RMVN			**1.473**	**1.036**	**1.513**
GM6		**130**	**9.866**	**2.26**	**2.045**
GM6.IDRGP.RMVN			**1.191**	**0.938**	**1.381**

يعرض الجدول (3,2,1) نتائج المحاكاة لتقدير معلمات الطرائق GM6, GM6.IDRGP.RMVN لـ (30,50,90,130) مشاهدة ملوثة بنسب تلويث مختلفة (0.15,0.10,0.05) من النقاط الرافعة والقيم الشاذّة بوجود نقطة رافعة عالية. الملاحظ أنّ متوسّط مقدّرات طريقة المقترحة GM6.IDRGP.RMVN هي الأقرب لقيم المعلمات الأولية المستعملة في دراسة المحاكاة الا وهو الواحد الصحيح لكل معلمة، كما نلاحظ ان لطريقة المقترحة حققت أقل قيم من قيم $\overline{bias}$ أقل من الواحد الصحيح. حصول الطريقة المقترحة GM6.IDRGP.RMVN على قيم $\overline{Mse}(\hat{\beta})$ أقل بكثر من طريقة GM6. مما يشير إلى دقة تقدير معلمات الطريقة المقترحة هذا من جانب دقة تقدير المعلمات.

أمّا من جانب المعايير التي تخص المقارنة بين النماذج كما أسلفت فلجأ الباحث إلى حساب متوسّط القيمة المطلقة للتحيّز لكل طريقة وعرضها في الجداول (4) الذي بيّن أنّ MSE وMAE وMAD الطريقة المقترحة تتناقص عند ثبات نسبة التلويث وزياد حجم العيّنة. وأنّ الطريقة المقترحة حققت أقل قيم من قيم المعاير الثلاثة مما يوكّدُ تفّوق طريقة المقترحة GM6.IDRGP.RMVN على طريقة GM6.

الاستنتاجات: من خلال نتائج البحث تَمّ التوصل إلى ما يأتي:

1- نلاحظ أنّ الطريقة المقترحة (GM6.IDRGP.RMVN) كانت لها الأفضلية عند حجوم العيانات الصغيرة(30) والمتوسطة(90,50) والكبيرة(130) بالاعتماد على معاير المقارنة (MSE وMAE وMAD).

التوصيات: من خلال نتائج البحث نوصي بالآتي: -

1- يَتُمّ الاعتماد على الطريقة المقترحة (GM6.IDRGP.RMVN) لأنّ لها الأفضلية بالتقدير ومقاومة النقاط الرفع العالية HLPs ولحجوم العيّنات جميعاً.

المصادر

1- **A A. S. Hadi,(1992) "A new measure of overall potential influence in linear regression," Computational Statistics & Data Analysis, vol. 14, no. 1, pp. 1–27.**
2- **Ali, D. A and Habshah. M,(2020)" On the Robust Parameter Estimation Method for Linear Model with Autocorrelated Errors in the Presence of High Leverage Points and Outliers in the Y-Direction" Malaysian Journal of Mathematical Sciences 14(3): 505-517.**
3- **Bagheri, A., and Midi , H. (2009) "Robust Estimations as a Remedy for Multicollinearity Caused by Multiple High Leverage Points" Journal of Mathematics and Statistics,NO. 5(4), PP 311-321.**
4- **Alma, Ö. G. (2011) "Comparison of Robust Regression Methods in Linear Regression" Int. J. Contemp. Math. Sciences,NO. 6(9), PP 409-421.**
5- **Chave, A. D., and Thomson , D. J. (2003) "A Bounded Influence Regression Estimator Based on the Statistics of the Hat Matrix " Journal of the Royal Statistical Society: Series C (Applied Statistics),NO.52 (3),PP 307-322.**
6- **Coakley.Clint W and Thomas P. Hettamansperger(1993)" A Bounded Influence, High Breakdown, Efficient Regression Estimato" Journal of the American Statistical Association, Vol. 88, No. 423 , pp. 872- 880.**
7- **D.L. Donoho and P.J. Huber(1983), The notion of breakdown point. In: Bickel PJ, Doksum KA, Hodges JL Jr (Editors), A Festschrift for Erich L. LehmannWadsworth, Belmont,(pp 157-184).**
8- **Habshah. M ,Shelan.S, Jayanthl.T, & Mohammed.A(2021):" Simple and Fast Generalized - M (GM) Estimator and Its Application to Real Data Set" Sains Malaysiana 50(3) 859-867.**

9- Hassan S. Uraibi (2009) "DYNAMIC ROBUST BOOTSTRAP ALGORITHM FOR LINEAR MODEL SELECTION USING LEAST TRIMMED SQUARES"Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master Science.

10-Hassan Uraibi and Sawsan Haraj (2022)" GROUP DIAGNOSTIC MEASURES OF DIFFERENT TYPES OF OUTLIERS IN MULTIPLE LINEAR REGRESSION MODEL" Malaysian Journal of Science 41 (Special Issue 1): 23-33.

11-Lukman, A. F., Osowole , O. I. & Ayinde, K.(2015) "Two Stage Robust Ridge Method in a Linear Regression Model" Journal of Modern Applied Statistical Methods, NO. 14(2), PP 53-67.

12-Rousseeuw, P. J. and Leroy, A. M. (2005) Robust Regression and Outlier Detection, John wiley & sons(p218).

13- Rousseeuw, P.J., Van Zomeren, B.C., 1990. Unmasking multivariate outliers and leverage points. Journal of the American Statistical Association 85, 633–639.

14-Yohai , V.J. (1987) "High breakdown point and high efficiency robust estimates for regression" The Annals of Statistics, NO. 15 (20), PP 642-656.