

AL- Rafidain
University College

PISSN: (1681-6870); EISSN: (2790-2293)

مجلة كلية الرافدين الجامعة للعلوم

Available online at: <https://www.jrucs.iq>

JRUCS

Journal of AL-Rafidain
University College
for Sciences

ترجيح طريقة GM6 لتقدير معلمات عالية الحصانة لنموذج الانحدار الخطي المتعدد

نور علوان عوده	أ.م.د. حسن سامي عريبي
noor.alwan22@qu.edu.iq	hassan.uraibi@qu.edu.iq
قسم الاحصاء، كلية الادارة و الاقتصاد، جامعة القادسية، القادسية، العراق	

المستخلص

معلومات البحث

تواريخ البحث:

تاريخ تقديم البحث: 23/2/2024

تاريخ قبول البحث: 12/4/2024

تاريخ رفع البحث على الموقع:

31/12/2024

الكلمات المفتاحية:

GM6, MM, IDRGP, RMVN,
HLP

للمراسلة:

اسم الباحث : د. حسن سامي عريبي

hassan.uraibi@qu.edu.iqDOI: <https://doi.org/10.55562/jrucs.v56i1.3>

يعد نموذج الانحدار الخطي المتعدد أحد الاساليب الاحصائية الأكثر استعمالاً في العديد من المجالات العلمية. إذ يتم تقدير معلمات النموذج باستخدام طريقة المربعات الصغرى الاعتيادية التي تعطي أفضل تقدير خطي غير متحيز عند تحقق فرضياتها، أحد من هذه الفرضيات ان توزيع حد الخطأ العشوائي يكون طبيعياً بمتوسط صفر وتباين ثابت. ان وجود القيم الشاذة يؤدي الى خرق هذا الافتراض وعليه لا يمكن تطبيق طريقة المربعات الصغرى. لقد ظهرت الطرائق الحصينة كبديل مناسب عند ظهور القيم الشاذة مثل 'طريقة M، LMS، LTS ومقدر S الا ان ظهور النقاط الرافعة Leverage Points كنوع اخر من القيم الشاذة في فضاء المتغيرات التوضيحية أدى الى عدم دقة تقديرات هذه الطرائق. لقد اقترح في الادبيات الإحصائية طرائق مميزة جداً مثل طريقة مقدر MM وطريقة GM6 لكن ظهور مشكلة النقاط الرافعة العالية High Leverage Points التي يعزى سبب وجودها مصدراً لظهور ظاهريتي الـ (Masking & Swamping). هاتين الظاهرتين تشير الى عدم الدقة في تشخيص النقاط الرافعة مما يسبب اما الخسارة في المعلومات او عدم معالجة النقاط الرافعة الرديئة بالكامل. ان مقترح هذا البحث يسعى الى تحسين أداء طريقة GM6 حتى مع وجود هاتين الظاهرتين من خلال توظيف خوارزمية تشخيص عالية الدقة تدعى IDRGP(RMVN) من اجل الحصول على اوزان عالية الدقة لتقليل أثر النقاط الرافعة في طريقة GM6. لقد وضعنا التشخيص الدقيق وطريقة GM6 في إطار خوارزمي واحد أطلقنا عليه GM6.IDRGP(RMVN) التي أظهرت أداء مميزاً مقارنة بالطرق السابقة من خلال نتائج دراسة المحاكاة والبيانات الحقيقية.

1. المقدمة Introduction

أن الانحدار الخطي المتعدد الذي يقوم بالأساس في تفسير ودراسة مدى تأثير المتغيرات المستقلة على المتغير المعتمد يعتمد على مجموعة من الفرضيات الأساسية التي يجب توفرها. عند مخالفة إحدى هذه الفرضيات يكون التحليل غير دقيق كأن تكون الأخطاء العشوائية لا تتوزع طبيعياً بسبب وجود القيم الشاذة. القيمة الشاذة تعرف على انها القيمة التي تكون منحرفة عن باقي المشاهدات كما عرفها [1] وفي غالب تكون غير منسجمة و بعيدة عن مركز تجمع البيانات، و ان سبب عدم الانسجام قد يعود الى انها قيمة غير منطقية بسبب أخطاء في التسجيل [2] او انها تولدت من توزيع مختلف عن توزيع معظم البيانات كما وصفها [3]. تصنف القيم الشاذة في نماذج الانحدار الخطي المتعدد حسب موقع وجودها فتلك التي توجد في الخطأ العشوائي والمتغير المعتمد والتي يطلق عليها مصطلح (Outliers)، أما القيم الشاذة التي تظهر في مصفوفة المتغيرات التوضيحية يطلق عليها مصطلح Leverage Point (LP) النقاط الرافعة التي يمكن تصنيفها الى جيد و رديء كما ان هناك النقاط الرافعة العالية (HLP) High Leverage Point. لمزيد من المعلومات انظر [4].

ان المقدرات الحصينة يجب ان تتمتع بالكفاءة ونقطة انهيار عاليتان تثبت قدرتها على مقاومة القيم الشاذة، ولذلك لا تعد مقدرات طريقة المربعات الصغرى كفاءة عند وجود القيم الشاذة في مجموعة البيانات لان قيمة نقطة انهيارها تكون صفراً. لمعالجة مثل هذه المشكلة اقترح العديد من الباحثين عدداً من الطرائق الحصينة في الادبيات الإحصائية مثل [5] عام (1981) قدموا طريقة M-estimate التي استبدلت دالة الخسارة التربيعية للأخطاء بدالة حصينة و بأسلوب شبه لا معلمي يعتمد على إعادة ترجيح الأخطاء من خلال دالة Huber و بأسلوب تكراري، عام (1984) اقترح [6] طريقة وسيط المربعات الصغرى Least Median Squares (LMS) of Squares التي اختزلت احتساب حصانة المقدرات على نحو 50% من حجم بواقي الانحدار بعد اخضاعها للإحصاء

الترتيبي، و في عام (1987) اعلن عن طريقة المربعات الصغرى المشدبة (LTS) Least Trimmed of Squares [7] التي عملت على تقليل مجموع مربعات البواقي المشدبة بعد ان يتم ترتيبها تصاعديا او تنازليا. على الرغم من ان هذه الطرائق الحصينة تملك نقطة انهيار تصل الى 0.50 الا ان كفاءتها متباينة و قد تصل نقطة انهيارها الى الصفر عن ظهور نقطة رافعة رديئة واحدة على الأقل في مصفوفة المتغيرات التوضيحية. لذلك دعت الحاجة لاقتراح طرائق بديلة لها القدرة على مقاومة القيم الشاذة والنقاط الرافعة، لذلك اقترح [8] طريقة (MM) التي تمتع بكفاءة عالية تصل الى 95% و نقطة انهيار تبلغ 0.50 و طريقة GM التي وضع خطواتها الاولى [9] التي تعمل على ترجيح النقاط الرافعة بعد تشخيصها. تعد طريقة مسافة Mahalanobis للكشف على النقاط الرافعة العالية واحدة من اهم الطرق الا ان [10] اظهرا ان اعتماد هذه المسافة على الوسط الحسابي و مصفوفة التباين والتباين المشترك التقليدية قد يؤدي الى حدوث ظاهرتي الـ Masking & Swamping و لذلك اقترح استخدام مصفوفة الموقع و القياس الحصينة بدلا عنهما. اذ ان ظاهرة Masking تحدث عندما تشخص الطريقة قيمة شاذة واحدة على الأقل و لكن واقعا هذه القيمة تخفي و تغطي على قيمة او قيم شاذة أخرى. اما ظاهرة الـ Swamping فهي خطأ في التشخيص من خلال عد مشاهدة طبيعية واحدة على الأقل على انها شاذة مما تقدم يتضح ان أي خطأ في تشخيص النقاط الرافعة بسبب وجود الـ HLP يؤدي بالضرورة الى عدم الدقة في التقدير.

2. الكشف الكلي للنقاط الرافعة

لجأ الكثير من الباحثين الى تحسين مسافة Mahalanobis عن طريق ابدال مدخلاتها التي تعتمد على الوسط الحسابي والتباين بمعلمتي الموقع والقياس الحصينة للتخلص من ظاهرتي Masking & Swamping. لكن هذه الإجراءات لم تساعدهم من التخلص الكلي من هاتين الظاهرتين لانهم استخدموا ما يسمى بطرق التشخيص الفردي. لذلك قدم [11] طريقة GP (Generalized Potential measure) عام (2002) للتشخيص الكلي و هي تعميم لطريقة Potential measure التي اقترحها [12] عام (1992). حفز اقتراح [11] باحثين آخرين لتحسين أداء هذه الطريقة، لذلك قدمت Midi وآخرون عام (2009) [13] خوارزمية DRGP (MVE) كطريقة لتشخيص العدد الكلي للـ HLP من خلال الدمج بين النهجين السابقين. فالتشخيص الاول يكون باستخدام MD الحصينة بالاعتماد على Minimum Volume Ellipsoid (MVE) و من ثم وضع الصفوف المشكوك باحتوائها على النقاط الرافعة العالية في مصفوفة تدعى مصفوفة الحذف D ، اما الصفوف الغير مشكوك توضع في مصفوفة أخرى و تطبيق منهج GP ، و من ثم إعادة فحص صفوف مصفوفة الحذف فيما اذا كانت تحتوي على نقاط رافعة باستخدام منهج Potential measure .

إذ لاحظ [14] عام (2015) ان تأثير الـ Masking and Swamping مازال موجودا على طريقة Diagnostic Robust Generalized Potential based on MVE (DRGP(MVE)) عندما تكون نسبة HLP تتراوح بين 5% و 10%. و قد أعزى السبب إلى إن تشخيص HLP في الخطوة الثانية من خوارزمية DRGP(MVE) التي وضعتها في مصفوفة الحذف التي أطلق عليها D لم يتم التحقق منها بشكل صحيح. لذلك اقترح Improvised DRGP based on MVE (IDRGP(MVE)) من خلال إضافة خطوة جديدة الى خوارزمية التشخيص باستخدام hat matrix التي وضع المصفوفة المشكوك بها في مصفوفة الحذف D₂ ومن ثم أجرى مقاطعة للتشخيصين (D₁, D₂) للتحقق من التشخيص الأول D ، فيتم الاحتفاظ على ما تم التوافق عليه بين المصفوفتين من حيث التشخيص و ما لم يتم التوافق عليه يعاد الى مصفوفة المشاهدات النظيفة.

لقد اثبت [15] ان استخدام مصفوفة الموقع و القياس MVE مع مسافة مهالانوبيز غير كافي للتخلص من هذه المشكلة وان خطأ التشخيص في هذه المرحلة لا يصحح في مرحلة مصفوفة القبة Hat matrix لوجود علاقة ما بين مسافة مهالانوبيز و هذه المصفوفة، هذا بالإضافة الى إمكانية فقدان مصفوفة القبة الى دقة التشخيص لكون عتبة التشخيص للمصفوفة غير ثابتة و تناولتها الادبيات على ثلاث عتبات الامر الذي يجعل عدم التخلص التام من مشكلة Masking & Swamping امر منطقيا، لذلك اقترحا طريقة IDRGP(RMVN) من خلال وضع خوارزمية Re-weighted Multivariate Normal (RMVN) ذات التركيز العالي جدا في اطار خوارزمي واحد مع طريقة التشخيص الكلي :

3. الطريقة المقترحة GM6.IDRGP(RMVN)

تم اقتراح هذه الطريقة بناء على خوارزمية IDRGP(RMVN) المقترحة من قبل [15] إذ تتبع نفس الخطوات فقط يتم ابدال مصفوفة الموقع والقياس بـ RMVN مكان MVE لكي تحافظ على كفاءة الطريقة وتحسينها من تأثير نقاط الرفع العالية وايضا معالجة ظاهرتي Masking & Swamping وتخفيض تأثيرها بنسب عالية. إذ كانت الخوارزميات السابقة تعاني من هذه المشكلة على الرغم من تخفيض Masking & Swamping إلا أنها لم تتخلص من تأثيرها بشكل كافٍ. لذلك قمنا في الطريقة المقترحة تكوين تركيبة من خلال توظيف خوارزمية [15]، لتكوين دالة وزن حصينة لطريقة التقدير GM6 ويتم توضيح ذلك من خلال الخطوات الاتية:

1. توظيف الخوارزمية المقترحة من قبل (Uaribi&Haraje,2022) التي عملت على ابدال مصفوفة الموقع والقياس الـ MVE بـ RMVN لحساب مسافة مهالانوبيز الحصينة RMD(RMVN) عن طريق الآتي:

$$RMD(RMVN) = \sqrt{(X_i - T_{RMVN}(X))^T C^{-1}_{RMVN}(X) (X_i - T_{RMVN}(X))}$$

2. تكون الأوزان الأولية الجديدة لطريقة GM6 بعد توظيف خوارزمية IDRGP(RMVN) محسوبة بالشكل الاتي:

$$\pi_i = \min \left\{ 1, \frac{\chi^2_{(k,0.95)}}{RMD(RMVN)} \right\}$$

3. احسب المقدّر الأولي $\tilde{\beta}$ باستخدام طريقة المربعات الصغرى المشدّبة (LTS).

4. احسب البواقي (r_i) ومقدّر القياس $\tilde{\sigma}$ عن طريق الصيغة الآتية:

$$\tilde{\sigma} = (1.4826) \left[\left(1 + \frac{5}{n-k-1} \right) \text{med}|r_i| \right]$$

5. حساب دالة وزن لإيجاد مقدّر GM كالآتي

$$W_i = \pi_i \psi \left(\frac{r_i}{\pi_i \tilde{\sigma}} \right) X_i$$

6. احسب قيم جديدة لـ $\tilde{\beta}$ باستخدام طريقة المربعات الصغرى المشدّبة (LTS) مع الاوزان.

$$\tilde{\beta}_{GM} = (X'WX)^{-1}X'Wy$$

7. كرر الخطوات من (4-6) إلى أن يتم الوصول لمقدّر الطريقة المقترحة GM6.IDRGP(RMVN) الى التقارب.

4. المحاكاة Simulation

ليكن نموذج الانحدار الخطي المتعدد الآتي :

$$y = X\beta + e$$

حيث أن X هي مصفوفة المتغيرات التوضيحية ذات بعد $n \times (p + 1)$ المولدة من توزيع طبيعي متعدد المتغيرات بمتوسط قيمته صفراً وتباين $\sigma = \rho^{||i-j||}$ أي أن

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

حيث أن $p = 3$ و $n = \{25, 50\}$ و $\rho = 0.5$ و β هي متجه الوحدة لمعاملات هذا النموذج مع الحد الثابت $\beta = [1 \ 1 \ 1 \ 1]$ ، أما e فهو متجه الأخطاء العشوائية للنموذج يتوزع طبيعياً بمتوسط قيمته صفر وتباين ثابت. قام الباحث باختبار كفاءة طرق التقدير خلال الحالات الآتية:

✓ **الحالة الأولى:** توليد حد الخطأ العشوائي e من توزيع مختلط وفقاً للصيغة الآتية

$$e = [(1 - \alpha) \times n] \sim N(0, 2) + [\alpha \times n] \sim \chi^2_{(25, 0.95)}$$

حيث $\alpha = \{0.05\}$ وضرب أعظم قيمة في X_1 وأعظم قيمة في y بالرقم (50) لضمان وجود الـ HLP ومن ثم ضرب $[\alpha \times n]$ من قيم X_2 و X_3 بالرقم (10) لضمان وجود الـ LP's.

✓ **الحالة الثانية:** تختلف هذه الحالة عن الحالة الثانية بجعل توزيع الخطأ العشوائي توزيعاً طبيعياً مختلطاً بتضخيم الوسط الحسابي وجعل موقع القيم الشاذة على نفس صفوف النقاط الرافعة العليا المذكورة في الحالة الثانية كالآتي:

$$e = [\alpha \times n] \sim N(5, 2) + [(1 - \alpha) \times n] \sim N(0, 2)$$

✓ **الحالة الثالثة:** لا تختلف هذه الحالة عن الحالة الثالثة سوى أن الباحث رجع متوسط التوزيع الطبيعي الملوّث الى الصفر و ضخّم التباين الى الـ (25)، هذا بالإضافة جعل مواقع القيم الشاذة بمواقع لا تتناظر النقاط الرافعة العليا كالآتي:

$$e = [(1 - \alpha) \times n] \sim N(0, 2) + [\alpha \times n] \sim N(0, 25)$$

اخضع الباحث الحالات السابقة الى 5000 دورة محاكاة لكل نسبة تلوّث α التي ولدت n من المشاهدات لكل نموذج. تم في كل دورة محاكاة تقدير معاملات نموذج الانحدار $\hat{\beta}_j^*$ بوجود الحد الثابت حيث $j = 0, 1, 2, 3$. لقد اعتمدت الباحثة على متوسط مربعات خطأ تقدير المعلمات لأجمالي دورة المحاكاة كمعيار للمفاضلة بين طرق التقدير،

$$Mse(\hat{\beta}_j) = \frac{\sum_{j=1}^{5000} (\hat{\beta}_j^* - \beta_j)^2}{5000}$$

فالطريقة التي تحقق اقل قيمة لـ $Mse(\hat{\beta}_j)$ لكل معلمة مقدرة او لأغلب المعالم المقدرة بما فيها الحد الثابت. من ناحية أخرى قام الباحث برسم $Mse(\hat{\beta}_j)$ لحالات المحاكاة الأربعة كمخرجات لكل طريقة ذلك بهدف معرفة فيما اذا أداء الطرق ثابتاً مع تغيير حالات المحاكاة ام لا، مع ملاحظة ان الطريقة الأفضل هي التي ستظهر ان خط رسم $Mse(\hat{\beta}_j)$ لكل معلمة مقدرة يكون تحت خطوط التقدير لبقية الطرق.

لغرض قياس كفاءة الطريقة المقترحة تمت مقارنتها مع مقدري M لـ Huber و Bisquer و طريقتي LMS و LTS بالإضافة الى مقدري S و MM

(1) عندما $n=25$ ونسبة التلوّث في البيانات $\alpha=0.05$

جدول (1) $Mse(\hat{\beta})$ لطرق التقدير لـ 5000 دورة لكل حالة محاكاة عندما $n = 25$ ، $\alpha = 0.05$

Method	α	$Mse(\hat{\beta}_0)$	$Mse(\hat{\beta}_1)$	$Mse(\hat{\beta}_2)$	$Mse(\hat{\beta}_3)$
M_{hub}	Case 1	3.738	24.067	1.3066	0.957
M_{bisq}		2.705	24.093	1.316	0.957
LMS		2.008	1.341	1.141	1.053
LTS		1.910	1.262	1.115	1.004
S estimator		1.531	0.902	0.766	0.691
MM-Estimator		1.290	0.632	0.538	0.456
REWLSE		1.266	0.600	0.488	0.423
GM6.IDRGP(RMVE)		1.122	0.412	0.357	0.337
M_{hub}	Case 2	1.872	3.688	0.895	0.770
M_{bisq}		1.850	3.692	0.912	0.800
LMS		1.973	1.548	1.424	1.068
LTS		1.871	1.459	1.329	1.126
S estimator		1.628	0.960	0.939	0.742
MM-Estimator		1.360	0.663	0.559	0.490
REWLSE		1.325	0.601	0.534	0.449
GM6.IDRGP(RMVE)		0.823	0.354	0.321	0.316
M_{hub}	Case 3	3.444	21.777	8.277	5.194
M_{bisq}		3.176	21.831	7.493	5.521
LMS		1.919	1.290	1.283	1.164
LTS		1.819	1.311	1.166	1.105
S estimator		1.561	0.858	0.822	0.850
MM-Estimator		1.266	0.724	0.559	0.497
REWLSE		1.292	0.651	0.513	0.481
GM6.IDRGP(RMVE)		0.753	0.406	0.377	0.342

من النتائج المعروضة في الجدول (1) للحالات الأربعة للمحاكاة عندما عدد المشاهدات (25) و بوجود 5% من القيم الشاذة. لقد لجأت الباحثة الى أسلوب مناقشة النتائج للحالات الأربعة للمحاكاة لبقية المعالم المقدرة كالآتي:

- أولاً $Mse(\hat{\beta}_0)$: نلاحظ في جميع ان $Mse(\hat{\beta}_0)$ للمقترح هو اقل بكثير من بقية الطرق وانه تغيير الحالات من الأولى الى الثالثة اظهر أداء قويا و متسقا، اذ انخفضت قيمته من (0.792) في الحالة الأولى الى (0.585) في الحالة الثانية و أخير الى (0.546) في الحالة الثالثة. اما بقية الطرق لم تكن مستقرة فقط سجلت ارتفاعا ملحوظا في الحالة الثانية و من ثم سجلت انخفاضا جديدا في الحالة الثالثة للدلالة على عدم الاستقرار. ان اقرب الطرق لها هي طريقتي ، MM ، REWLSE.

- أولاً $Mse(\hat{\beta}_1)$: ان في كافة حالات المحاكاة X_1 تضمنت بياناته نقطة رفع عالية واحدة HLP تقع في موقع عشوائي ربما يكون مختلفا عن مواقع القيم الشاذة و النقاط الرافعة. هذا لا يعني ان الأسلوب العشوائي المتبع قد لا يجعل موقع HLP على نفس الصف المفهرس لمواقع القيم الشاذة و النقاط الرافعة. عموما لنناقش النتائج حسب الحالات كآتي:

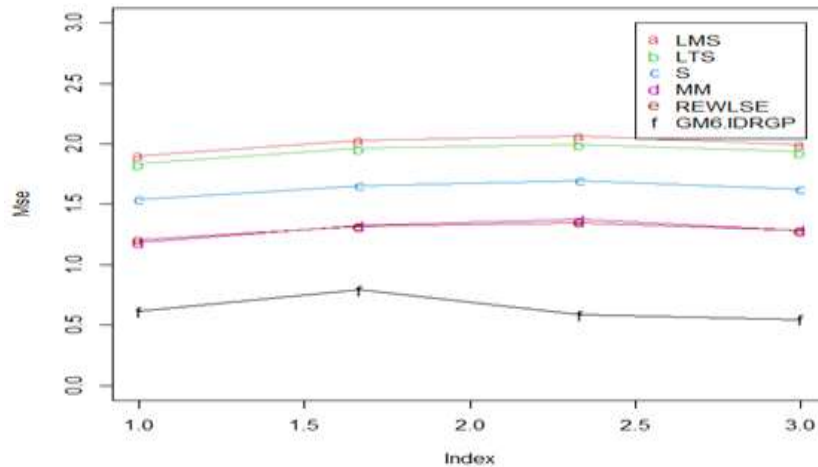
❖ Case 1: نلاحظ للمقترح GM6.IDRGP(RMVE) الذي اظهر مقاومة عالية ليحظى باقل قيمة لمتوسط مربعات الخطأ (0.43)، على العكس من طريقتي M_{hub} و M_{bisq} اللتان تتبعان أسلوب الترجيح لبواقي الانحدار قد تأثرتا تماما بالنقطة الرافعة العالية HLP من جانب و موقع القيم الشاذة في صف مفهرس مختلف عن الصف المفهرس للنقاط الرافعة من جانب اخر، لنتتج لنا متوسط مربعات خطأ لمعلمتهما يصل الى (20.883، 20.905). بالمقابل ان طريقتي LMS، LTS كان اقل تأثرا من سابقتها بسبب ان هاتين الطريقتين تستخدمان أسلوب الحذف او التشذيب، فطريقة LMS تتعامل مع 50% من البيانات كإحصاء ترتيبي، في حين LTS تعمل على تشذيب نسبة محددة من البيانات مكافئة لنسبة القيم الشاذة في بواقي الانحدار وإحصاء ترتيبي أيضا. بناء على ذلك فان هاتين الطريقتين شملت هذه الـ HLP النقاط الرافعة LP's بالحذف لكنها لم تتمكن من التخلص من تأثير الحذف الذي بدوره خفض درجات الحرية الامر الذي رفع التباين. بالتأكيد طرق التقدير الحصينة S، MM، REWLSE كانت اقل تأثرا من الطرق السابقة في حساب $Mse(\hat{\beta}_1)$ وان MM و REWLSE هما من الطرق التي لها القدرة على التعامل من كافة أنواع القيم الشاذة و النقاط الرافعة ولذلك اظهرتا أداء أفضل من طريقة S التي تتحسس للنقاط الرافعة و لكن بدرجة اقل من طرق الحذف و الترجيح السابقة.

❖ Case 2: نلاحظ ان قيمة $Mse(\hat{\beta}_1)$ في الحالة الثانية قد تغير بالنسبة لطرق الترجيح M_{hub} و M_{bisq} و أظهرت حساسية اقل لـ HLP و النقاط الرافعة و السبب يعود في ذلك الى ان القيم الشاذة و النقاط الرافعة وقعا على نفس الصفوف المفهرسة و

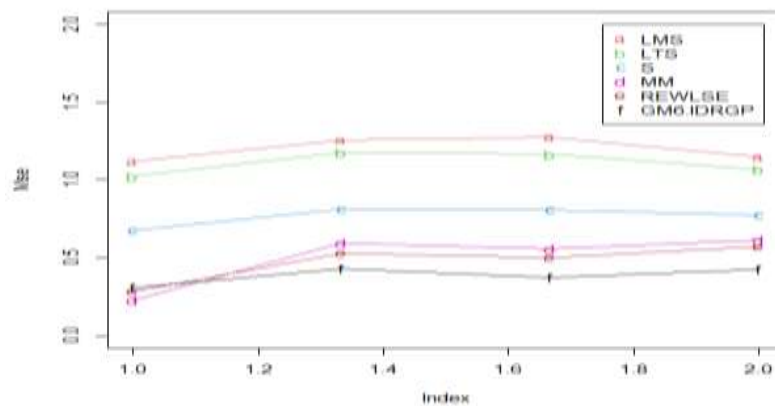
لذلك الترتيب شمل البواقي المفهرسة على نفس تلك الصفوف. بكلام آخر ان كل البيانات سوف تضرب بدالة ترجيح واحدة و عالية القيم الشاذة و النقاط الرافعة الواقعة المفهرسة على نفس الصف ستضرب بنفس القيمة المرجحة و التي بالتاكيد هدفها تقليل تأثير القيم الشاذة و لذلك ظهرت قيم متوسط مربعات الخطأ لهاتين الطريقتين (3.255 ، 3.256). اما طريقة LMS لم تتأثرا كثيرا بهذه التغيير لذلك اظهرت نتيجة مقاربة مع الحالة الأولى للمحاكاة. اما طريقة LTS أظهرت اداء افضل من ادائها في الحالة الأولى لان عملية التشذيب لم رفعت من درجات الحرية في الحالة الأولى لكون النقاط و الرافعة و القيم الشاذة وقعت على نفس الفهرس و هي في الاداء شأنها شأن تقديرات الطرق S، MM، REWLSE مع الحفاظ على ذات الافضلية السابقة. الملاحظ ان المقترح GM6.IDRGP(RMVE) قدم أداء افضل حتى من الحالة الأولى للمحاكاة حيث ان متوسط مربعات الخطأ انخفض من (0.43) الى (0.373) لتتفوق على الطرق الأخرى في الحالة الثانية أيضا.

❖ Case3 : ان ارتفاع قيمة متوسط مربعات الخطأ لطرق الترجيح M_{hub} و M_{bisq} الى (17.096، 17.188) ليس نتيجة لتضخم تباين للتوزيع الطبيعي الملوث للتوزيع الطبيعي لبواقي الانحدار و انما جاء بسبب اختلاف مواقع النقاط الرافعة عن القيم الشاذة و لذلك ظهر تأثير هذه الحالة على هاتين الطريقتين اسوة بتأثرهما بالحالة الأولى للمحاكاة، مع محافظة الطرق الأخرى على نفس الأداء الغير مستقر ما عدا المقترح GM6.IDRGP(RMVE) الذي اظهر قدرة عالية على التعامل مع الحالات الثلاث للمحاكاة بل بأداء افضل في الحالة الثالثة من ادائها في الحالة الأولى و الثانية على الرغم من تفوقه على الطرق الأخرى في جميع الحالات.

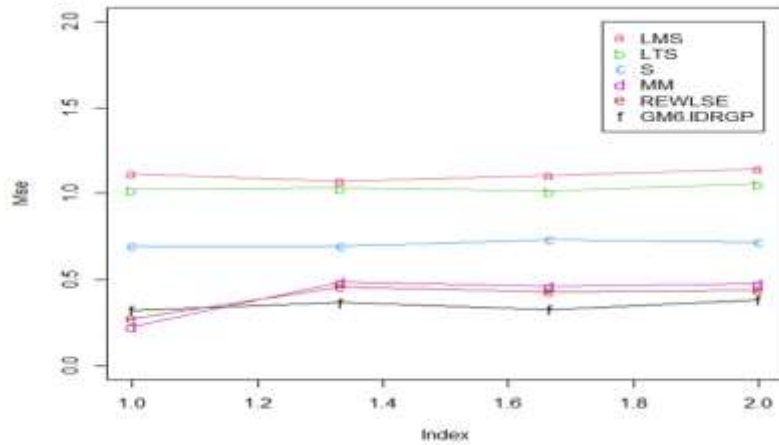
لقد سعينا الى رسم متوسط مربعات الأخطاء لطرق التقدير أولا من خلال بيانات نظيفة و خالية تماما من أي نوع من أنواع القيم الشاذة كحالة قياسية تضاف للحالات الثلاث لدراسة المحاكاة. من الشكل (1-3) الذي يظهر بما لا يقبل الشك ان $Mse(\hat{\beta}_0)$ لمقترح GM6.IDRGP(RMVN) يتفوق على الطرق الأخرى في الحالة القياسية و الحالات الثلاثة الأخرى للمحاكاة مع تقارب شبه تام في أداء كل من REWLSE، MM. الشكل (2) يعرض رسما لنتائج $Mse(\hat{\beta}_1)$ لطرائق الستة في الحالة القياسية و حالات المحاكاة الثلاث. من الواضح في الحالة القياسية لهذا النموذج هناك تفوق في أداء طريقتي MM و REWLSE و لكن هذا التفوق سرعان ما تراجع لصالح GM6.IDRGP(RMVE) في النماذج الأخرى للمحاكاة مع بقاء تفوق الطريقتين السابقتين على الطرق



شكل (1): $Mse(\hat{\beta}_0)$ لطرق التقدير الحصينة لكل حالات المحاكاة عندما $\alpha = 0.05, n = 25$

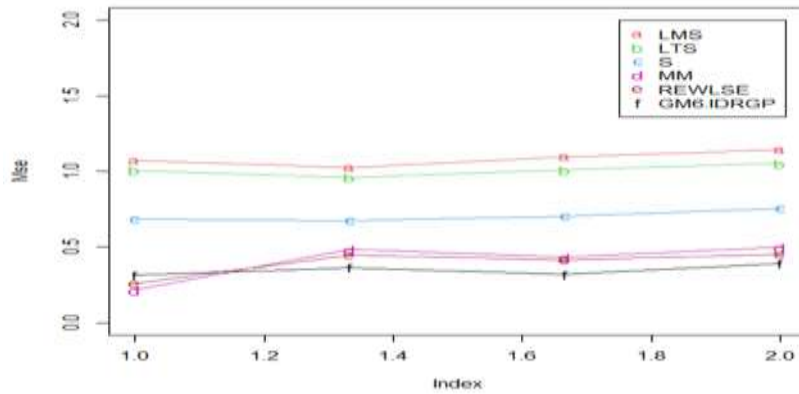


شكل (2): $Mse(\hat{\beta}_1)$ لطرق التقدير الحصينة لكل حالات المحاكاة عندما $\alpha = 0.05, n = 25$



شكل (3): طرق التقدير الحصينة لكل حالات المحاكاة عندما $n = 25$

هذا التفوق لمقترحنا استمر في تقدير $Mse(\hat{\beta}_2)$ و $Mse(\hat{\beta}_3)$ كما هو واضح في الشكل (3) و الشكل (4) مع أداء ثابت لكل الطرق تقريبا في الحالة القياسية او في حالات المحاكاة.



شكل (4): طرق التقدير الحصينة لكل حالات المحاكاة عندما $\alpha=0.05$, $n = 25$

(2) عندما يكون حجم العينة $n=50$ و نسبة التلوث في البيانات $\alpha=0.05$

جدول (2): $Mse(\hat{\beta})$ لطرق التقدير لـ 5000 دورة لكل حالة محاكاة عندما $\alpha=0.05$, $n = 50$

Method	α	$Mse(\hat{\beta}_0)$	$Mse(\hat{\beta}_1)$	$Mse(\hat{\beta}_2)$	$Mse(\hat{\beta}_3)$
M_{hub}	Case 1	2.275	11.146	0.662	0.581
M_{bisq}		1.619	11.155	0.656	0.575
LMS		1.493	0.581	0.529	0.533
LTS		1.459	0.561	0.534	0.539
S estimator		1.311	0.376	0.384	0.379
MM-Estimator		1.122	0.221	0.306	0.302
REWLSE		1.115	0.194	0.251	0.248
GM6.IDRGP(RMVN)		0.675	0.171	0.151	0.154
M_{hub}	Case 2	1.486	2.116	0.562	0.527
M_{bisq}		1.393	2.127	0.539	0.506
LMS		1.506	0.578	0.512	0.523
LTS		1.487	0.568	0.504	0.521
S estimator		1.340	0.381	0.364	0.355
MM-Estimator		1.153	0.214	0.235	0.221
REWLSE		1.138	0.184	0.218	0.197
GM6.IDRGP(RMVN)		0.527	0.159	0.149	0.136

M_{hub}	Case 3	1.773	11.937	2.199	1.754
M_{bisq}		1.607	11.697	1.663	1.363
LMS		1.511	0.555	0.533	0.518
LTS		1.481	0.557	0.538	0.521
S estimator		1.321	0.386	0.372	0.358
MM-Estimator		1.116	0.315	0.241	0.247
REWLSE		1.122	0.265	0.215	0.217
GM6.IDRGP(RMVN)		0.503	0.168	0.159	0.152

من النتائج المعروضة في الجدول (2) للحالات الأربعة للمحاكاة عندما عدد المشاهدات (50) و بوجود 0.05 من القيم الشاذة. لقد لجأت الى أسلوب مناقشة النتائج للحالات الأربعة للمحاكاة لبقية المعالم المقدرة كالآتي:

• **أولا $Mse(\beta_0)$:** عند مقارنة متوسط مربعات الاخطاء للمعلمة المقدرة لجميع الطرق الحصينة مع الطريقة المقترحة اعطت الطريقة المقترحة GM6.IDRGP(RMVN) اقل قيمة مقارنة مع الطرق الاخرى في ظل وجود القيم الشاذة لانها حافظت على كفاءتها وقدرتها على المقاومة وعدم الانهيار، حينما اعطت الحالة الثالثة افضل تقدير قوي متنسق بلغت قيمته (0.503) واوضحت استقرارها العالي مقارنة مع الحالات الاخرى اما الحالة الثانية التي اعطت نتيجة مقارنة الثالثة (0.527) الا ان الحالة الاولى سجلت قيمة مقارنة لطرق التقدير الاخرى. اعطت الطريقتان (REWLSE,MM) نتائج مقاربه اليها.

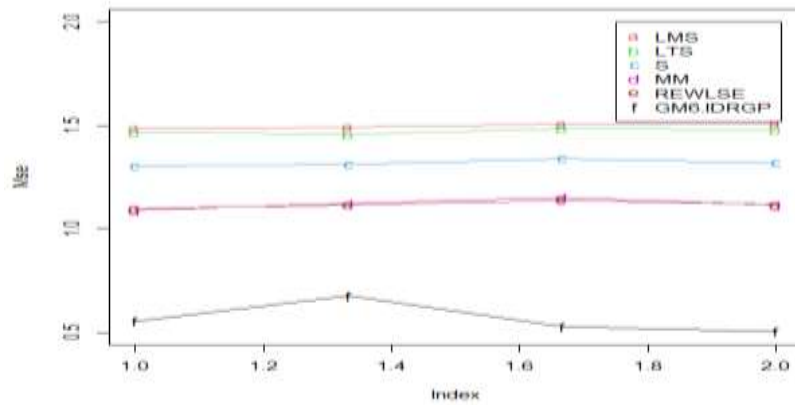
• **أولا $Mse(\beta_1)$:** يتم مقارنة متوسط مربعات الاخطاء للمعلمات المقدرة بالطرق التقدير الحصينة في ظل وجود نقاط الرفع والقيم الشاذة ونقاط الرفع العالية HLP, بالاعتماد على $Mse(\beta_1)$ حيث نلاحظ قيمتها عالية وهذا يدل على وجود نقاط الرفع والقيم الشاذة على مواقع مختلفة وهذا الامر لا يعني بان نقاط الرفع العالية HLP تكون في اتجاه بعيد عن موقع نقاط الرفع والشاذة في حالة كون الصف يحتوي على هذه القيم، لذلك يتم مناقشة النتائج حسب الحالات بالشكل الاتي:-

❖ **Case1:** بناء على نتائج المحاكاة في الجدول اعلاه لوحظ ان طريقتنا المقترحة GM6.IDRGP(RMVN) قد تفوقت على الطرق الاخرى باعطاء اقل متوسط مربعات اخطاء للمعلمات المقدرة واثبتت قدرتها العالية على مقاومة القيم الشاذة ونقاط الرفع ونقاط الرفع العالية HLP، حيث سجلت الحالة الاولى قيمة تصل الى (0.171) وكانت قيمة كل من الطريقتين (MM,S) قريبة في المقاومة لهكذا نوع من القيم حيث اعطت طريقة S قيمة تصل الى (0.376) بينما اعطت MM (0.221) بينما ابدت طريقة Mhub و Mbisq فشلها في مقاومة نقاط الرفع العالية HLP ونقاط الرفع على الرغم من مقاومتها للقيم الشاذة في الصف الذي يحتوي على هذه القيم.

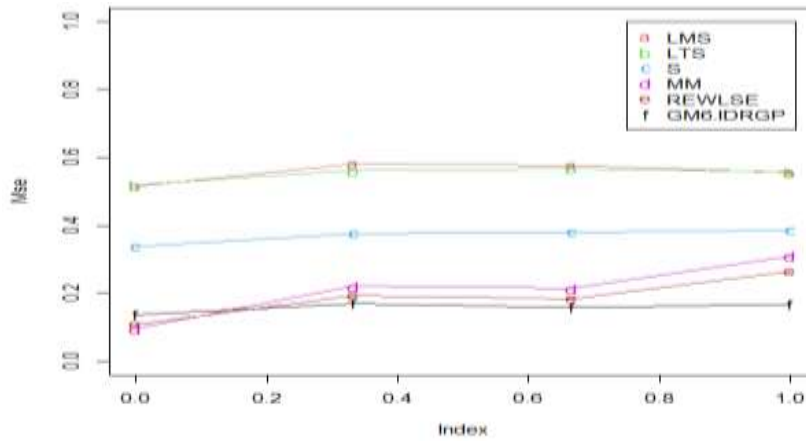
❖ **Case 2:** نلاحظ ان قيمة متوسط مربعات الاخطاء $Mse(\beta_1)$ في الحالة الثانية سجلت عند الطريقة المقترحة GM6.IDRGP(RMVN) افضل النتائج حيث اعطت (0.159)، وهذه القيمة ابدت مقاومتها لوجود القيم الشاذة ونقاط الرفع ونقاط الرفع العالية HLP. في الصف الذي يحتوي عليها كما سجلت كل من (MM,S) نتائج متقاربة من الطريقة المقترحة، حيث اعطت الطريقة الاولى قيمه مقدارها (0.381) وتعد جيدة في مقاومة تأثير القيم في صف مصفوفة المتغيرات المستقلة، بينما اعطت طريقة MM قيمة (0.214) وهي تعد طريقة مقاومة لتأثير القيم الشاذة ونقاط الرفع ولكنها لم تستطع مقاومة تأثير نقاط الرفع العالية HLP، ولكن طريقة تعد قريبة للطريقة المقترحة، حيث تعد افضل الطرق الحصينة لمقاومة هذه القيم الرديئة بينما بلغت كل من الطريقتين (LTS,LMS) قيم متقاربة في مقاومة القيم الشاذة ولكنها لم تظهر مقاومة لنقاط الرفع العالية HLP ونقاط الرفع حيث اعطت متوسط مربعات اخطاء بلغت طريقة LMS (0.578) حيث تعتمد هذه الطريقة على استخدام الوسيط كمعلمة للموقع بدلا من الوسط الحسابي، لكي تقلل من تأثير القيم الشاذة وفي نفس الوقت تحسن المعالم وعلى نفس النهج سجلت طريقة (LTS) التي تعمل على تشذيب نسب من البيانات لتقليل من تأثير القيم الشاذة وكذلك تعمل على ترتيب البيانات كما في طريقة LMS. اي اعطت قيمة (0.568) من متوسط مربعات الخطأ وتماشيا على ما تم ذكره سجلت الطريقتان (Mbisq,Mhub) نتائج متقاربة حيث اعطت Mhub قيمة (2.116) واعطت الطريقة الاخرى على نفس النهج قيمة (2.127) ويلاحظ من خلال هذه القيم ان الطريقتان تقاوم تأثير القيم الشاذة في البيانات الا انهما تفشلان في المقاومة امام وجود نقاط الرفع ونقاط الرفع العالية HLP في الصف الذي يحتوي على القيم الشاذة لذلك اعطت كلتا الطريقتان متوسط مربعات اخطاء عالي جدا مقارنة بالطرق التقدير الحصينة الاخرى التي سجلت قيم مقاومة لا بأس بها، ولكن طريقتنا المقترحة اثبت تفوقها في تقدير معلمات النموذج في ظل وجود تأثير القيم الشاذة ونقاط الرفع وكذلك نقاط الرفع العالية HLP واعطت مقدار كفوء من خلال تقليل متوسط مربعات الاخطاء.

❖ **Case3:** في الحالة الاخيرة من المحاكاة نلاحظ بقي الطريقة المقترحة GM6.IDRGP(RMVN) على مقاومتها وكفاءتها في اعطى افضل التقديرات وكذلك اقل متوسط مربعات اخطاء، اذا كانت قيمتها (0.168) وهي قيمة عالية المقاومة لتأثير القيم الشاذة ونقاط الرفع ونقاط الرفع العالية HLP، على الرغم من اختلاف مواقعها من النموذج مقارنة بالطرق الحصينة الاخرى وكذلك سجلت كل من (S-estimator,MM-Estimator,REWLSE) قيم متقاربه على التوالي في مقاومة هذه القيم حيث كانت القيم الاتية (0.386,0.315,0.265) لكل طريقة، في حين سجلت الطريقتان (Mbisq,Mhub) قيم (11.697,11.937) على التوالي وهي قيم عالية مقارنة بطريق الاخرى والطريقة المقترحة وهذا يعني اثبات ان الطريقتان ابدت فشلها في اعطى متوسط مربعات اخطاء منخفض وعدم قدرتها على تقدير معالم النموذج في ظل وجود القيم

الشاذة ونقاط الرفع ونقاط الرفع العالية. لقد تطرقنا الى رسم متوسط مربعات الأخطاء لطرق التقدير أولا من خلال بيانات نظيفة و خالية تماما من أي نوع من أنواع القيم الشاذة كحالة قياسية. من خلال الشكل (4) الذي يظهر بان الطريقة المقترحة GM6.IDRGP(RMVN) قد تفوقت على الطرق التقدير الحصينة الأخرى وهذا في الحالة القياسية وكذلك الحالات الأخرى للمحاكاة , مع ملاحظة وجود تقارب بعض نتائج المحاكاة للطرق الحصينة.

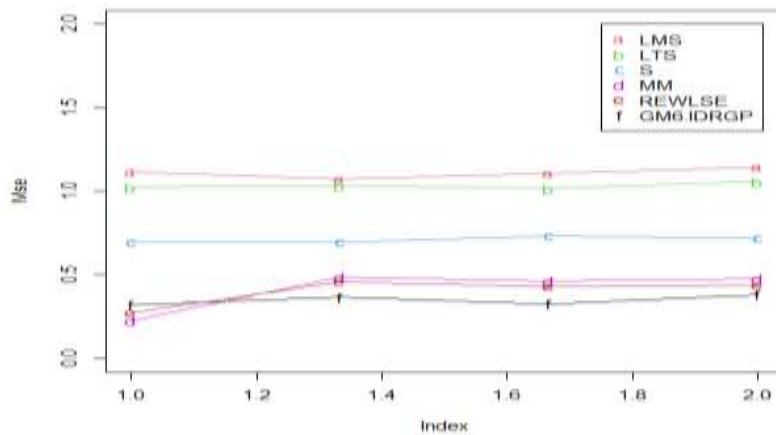


شكل (5): $Mse(\hat{\beta}_0)$ لطرق التقدير الحصينة لكل حالات المحاكاة عندما $n=50, \alpha=0.05$



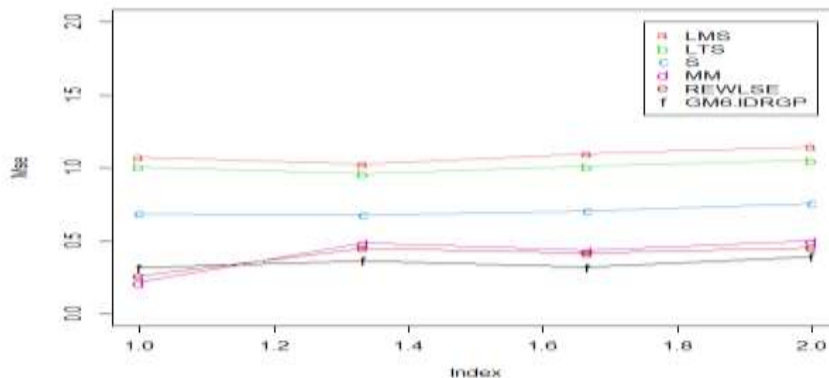
شكل (6): $Mse(\hat{\beta}_1)$ لطرق التقدير الحصينة لكل حالات المحاكاة عندما $\alpha=0.05, n=50$

الشكل (5) يوضح نتائج متوسط مربعات الأخطاء لطرق التقدير الحصينة الخمسة مع الطريقة المقترحة حيث ابدت طريقتنا المقترحة تفوقها واستقرارها في حالة المحاكاة , كما اعطت الطريقتان REWLSE,MM نتائج متقارب وتفوقها على الطرق الأخرى.



شكل (7): $Mse(\hat{\beta}_2)$ لطرق التقدير الحصينة لكل حالات المحاكاة عندما $\alpha=0.05, n=50$

من خلال الشكل (7) والشكل (8) نلاحظ استمرار تفوق الطريقة المقترحة بأخطاء اداء ثابت لكل حالات المحاكاة، من خلال تسجيلها أقل متوسط مربعات اخطاء وكما ان الطريقتان REWLSE,MM من خلال الشكلين ادت عمل متقارب وتفوقتا على الطرق الاخرى من خلال اعطى متوسط مربعات اخطاء كاد يكون قريب من الطريقة المقترحة، اما الطرق الحصينة الاخرى كانت بعيدة جدا في اداءها عن طريقتنا وهذا يلاحظ اعطى متوسط مربعات اخطاء عالي في نتائج وحالات المحاكاة.



شكل (8): $Mse(\hat{\beta}_3)$ لطرق التقدير الحصينة لكل حالات المحاكاة عندما $\alpha=0.05$, $n = 50$

5. بيانات القيمة السوقية:

تم الحصول على هذه البيانات من الموقع الرسمي لسوق العراق للأوراق المالية، إذ تم جمع البيانات السنوية للقيمة السوقية لتسعة من المصارف المحلية الأكثر تداولاً والمسجلة في سوق العراق للأوراق المالية في المدة (2011 – 2015). إذ تعد هذه البيانات هي شهرية وكما تضمنت أربعة متغيرات وهي (نسبة التداول، معدل سعر السهم السنوي، الموجودات والقيمة السوقية) وكان حجم العينة ($n=45$)، (Uraibi&Haraje,2022). إذ اعتبرنا ثلاثة من هذه المتغيرات توضيحية أو تفسر حجم القيمة السوقية وفقاً لأنموذج الانحدار الخطي المتعدد يمكن وصفه كآلاتي :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 +$$

حيث إن y : تمثل القيمة السوقية، x_1 تمثل نسبة التداول، x_2 : تمثل معدل سعر السهم السنوي، x_3 : تمثل الموجودات، e : حد الخطأ العشوائي.

نلاحظ من خلال الجدول (2) أعلاه مقارنة معاملات نموذج الانحدار الخطي المتعدد والتي تم تقدير معالمها بواسطة طريقة M_{hub} قد سجلت β_3 أقل قيمة من قيم معالم الأخرى، وعند مقارنة الخطأ المعياري للمعالم نلاحظ أن قيمة β_2 أعطت أقل قيمة عند مقارنتها مع الأخطاء المعيارية للمعالم الأخرى. إذ عند انتقال لطريقة M_{bisq} ومقارنة معالمها نلاحظ أن β_3 قد أعطت أقل قيمة، والخطأ المعياري لمعالم الأنموذج اتجه نحو قيمة β_3 إذ سجلت أقل قيمة من بين الأخطاء المعيارية الأخرى. في نفس الصدد أعطت طريقة LMS التي تعتمد على ترتيب المشاهدات تصاعدياً وتركز على مقياس الوسيط في تقدير معالم الأنموذج كون الوسيط الحسابي حساس للقيم الشاذة، وعند مقارنة قيم معالم لهذه الطريقة الحصينة سجلت β_3 أقل قيمة من المعالم الأخرى وأقل خطأ معياري من بين الأخطاء المعيارية الأخرى للمعالم. ويتم مقارنة جميع قيم معالم الأنموذج والأخطاء المعيارية لمعالم الأنموذج، بنفس الطريقة إذ إن قيمة المعلمة والأخطاء المعيارية لطرائق التقدير الحصينة لها تعطي أقل قيمة هي أفضل طريقة حصينة في تقدير معالم مقارنة بقيم معالم الأخرى. وفي هذا السياق نلاحظ من خلال مقارنة طرائق التقدير الحصينة على أساس قيمة للأخطاء المعيارية مقارنه مع الطرائق الحصينة الأخرى.

مقياس الخطأ المعياري لبواقي الأنموذج، إذ سجلت الطريقة الحصينة المقترحة (GM6.IDRGP(RMVN)) أقل وهذا يعني تفوق طريقتنا المقترحة على الطرائق الأخرى الحصينة أي أنها استطاعت أن تقاوم تأثير القيم الشاذة ونقاط الرفع العالية في البيانات وأعطت أفضل النتائج، في حين أن طريقة REWLSE أعطت أعلى خطأ معياري للبواقي في الأنموذج من بين الطرائق الأخرى وهذا يعني فشل هذه الطريقة من مقاومة تأثير القيم الشاذة ونقاط الرفع العالية في البيانات إذ إن هذه الطريقة لا يعول عليها في تقدير معالم الأنموذج. أما الطرائق الأخرى فكانت نسبة الأخطاء المعيارية فيها متقاربة ولكن لا تعد أنها جيدة من ناحية التقدير كونها تتأثر بوجود نقاط الرفع العالية والقيم الشاذة، على رغم من أن طريقة MM تعد من أكثر الطرائق الحصينة شيوعاً واستخداماً في الأبحاث العلمية إلا أنها لم تبدي أي مقاومة بتأثير نقاط الرفع العالية في البيانات. نلاحظ من خلال مقارنة قيم معالم التحديد لطرائق التقدير الحصينة مع الطريقة الحصينة المقترحة بوجود نقاط الرفع العالية والقيم الشاذة، وبلا شك أن طريقة التقدير الحصينة المقترحة أعطت أفضل قيمة لمعامل التحديد إذ إن قيمة معامل التحديد التي تقترب من الواحد هي أفضل قيمة من التي تقترب من الصفر وذلك يعني إن الأنموذج قادر على أن يفسر البيانات بشكل دقيق وقوي. وبالتالي يمكن استخدام الطريقة التي تعطي أكبر معامل تحديد في تحسين معاملات نموذج انحدار الخطي المتعدد بوجود نقاط الرفع العالية والقيم الشاذة.

6. الاستنتاجات Conclusions

من النتائج المعروضة في الجداول السابقة نستنتج بأن طريقة لعملية التشخيص الدقيق دور مهم في تحديد حصانة طريقة التقدير و ان عملية دمج طريقة التشخيص الكلي المقترحة (IDRGP(RMVN)) مع طريقة التقدير GM6 كان لها الأفضلية على

كافة الطرق الأخرى و تحديدا عند وجود النقاط الرافعة العالية و حدوث ظاهرتي (Masking & Swamping). من خلال الجانب التجريبي للمحاكاة وباستخدام معيار متوسط مربعات الاخطاء للمقارنة بين اداء الطرائق الحصينة والطريقة المقترحة وكما تم توضيح رسم المعالم المقدرة لكل طريقة بوجود نسبة تلوث في البيانات وبحجم عينة محدد، فقد لاحظنا تقدم اداء الطريقة IDRGP(RMVN) بتسجيل أقل متوسط مربعات اخطاء ولكافة حالات محاكاة بينما اعطت الطرائق الاخرى الحصينة نتائج متقاربة في المقاومة لتأثير القيم الشاذة ونقاط الرفع العالية HLP. من جانب اخر أظهر الجانب التطبيقي للبيانات الحقيقية (بيانات القيمة السوقية) نتائج جيدة للطريقة المقترحة GM6.IDRGP(RMVN) من خلال استخدام مجموعه من المعايير للمقارنة في ما بين طريقتنا والطرائق الاخرى كمعامل التحديد والخطأ المعياري لبواقي النموذج، وبحجم عينة 45 وبوجود النقاط الرافعة في البيانات حيث كان اداء الطريقة المقترحة متفوق في معرفة مدى تأثير المتغيرات على القيمة السوقية.

الجدول (2) يبين قيم $\hat{\beta}$ ، $Se(\hat{\beta})$ ، الخطأ المعياري لبواقي لآتمودج Se. Model ومعامل التحديد R^2

Method	$\hat{\beta}$	$Se(\hat{\beta})$	Se. Model	R^2
M_{hub}	-0.221	0.039	0.193	0.18
	0.215	0.046		
	0.261	0.051		
	0.434	0.05		
M_{bisq}	-0.261	0.035	0.198	0.126
	0.151	0.041		
	0.275	0.046		
	0.415	0.045		
LMS	-0.265	0.047	0.139	0.175
	0.215	0.055		
	0.212	0.061		
	0.422	0.061		
LTS	-0.162	0.05	0.134	0.168
	0.223	0.058		
	0.175	0.066		
	0.655	0.065		
S	-0.294	0.065	0.190	0.153
	0.197	0.075		
	0.209	0.084		
	0.41	0.083		
MM	-0.296	0.035	0.193	0.516
	0.133	0.04		
	0.278	0.043		
	0.374	0.045		
REWLSE	-0.293	0.036	0.199	0.518
	0.134	0.041		
	0.276	0.044		
	0.376	0.046		
GM6.IDRGP(RMVN)	-0.113	0.021	0.104	0.884
	0.261	0.034		
	0.196	0.038		
	0.668	0.049		

References

- [1] Bross, I.D.J. (1961), "Outliers in Patterned Experiments: strategic Re-Appraisal". Technometrics. 3,19-102.
- [2] Barnett, Vic, & Lewis, Toby. (1994). Outliers in statistical data (Vol. 3): Wiley New York.
- [3] Freeman, R. B. (1980). Unionism and the Dispersion of Wages. ILR Review, 34(1).
- [4] Keller, G. & Brian Warrack, (2000),"Statistic For Management and Economics, "5th edition, Duxbury, Thomson Learning, U.S.A.

- [5] Huber, P. J., & Ronchetti, E. M. (1981). Robust statistics, ser. Wiley Series in Probability & Mathematical Statistics. New York, NY, USA, Wiley-IEEE, 52.
- [6] Rousseeuw, P. J. (1984). Least median of squares regression. Journal of the American statistical association, 79(388), 871-880..
- [7] Rousseeuw, P. J. and Zomeren B. Van, (1990). Unmasking multivariate outliers and leverage points, J. AM. STAT. ASSOC., 85, 633-639
- [8] Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. The Annals of statistics, 642-656.
- [9] Coakley, C.W. and Hettmansperger, T.P. (1993). A bounded influence, high breakdown, efficient regression estimator. JASA, 88(423), 872-880.
- [10] Rousseeuw P.J. and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. Technometrics. 41:212–223
- [11] Imon, A.H.M.R. (2002). Identifying multiple high leverage points in linear regression. Journal of Statistical Studies. Special Volume in Honour of Professor Mir Masoom Ali. 3: 207–218.
- [12] Hadi, A. S. (1992). A new measure of overall potential influence in linear regression. Computational and Statistical Data Analysis. 14:1-27.
- [13] Midi, H., Ramli, N and Imon, A.H.M.R. (2009). The performance of Diagnostic-Robust Generalized Potentials for the identification of multiple high leverage points in linear regression. Journal of Applied Statistics. 36(5): 507-520.
- [14] Mohammad A. Mohammad (2015), Rbust Estimation Methods and Robust Multicollinearity Diagnostics For Multiple Regression Model in The Presence of High Leverage Collinearity -Influential Observation, Thesis submitted to the School of Graduate Studies, UPM
- [15] Uraibi, H. S., & Haraj, S. A. A. (2022). Group Diagnostic Measures of Different Types of Outliers in Multiple Linear Regression Model. Malaysian Journal of Science, 41(sp1), 23-33.

AL- Rafidain
University College

PISSN: (1681-6870); EISSN: (2790-2293)

**Journal of AL-Rafidain
University College for Sciences**Available online at: <https://www.jrucs.iq>**JRUCS**Journal of AL-Rafidain
University College
For Sciences**Weighting GM6 Method to Estimate High Robustness Coefficients for
a Multiple Regression Model****Hassan S. Uraibi**hassan.uraibi@eq.edu.iq**Noor A. Odeh**noor.alwan22@qu.edu.iqStatistics Department, College of Administration and Economics , University of Al-Qadisiyah
Al-Qadisiyah , Iraq**Article Information****Article History:**

Received: February, 23, 2024

Accepted: April, 12, 2024

Available Online: December,
31, 2024**Keywords:**GM6, MM, IDRGP, RMVN,
HLP**Abstract**

The multiple linear regression model is one of the most widely used statistical methods in many scientific fields. The model parameters are estimated using the ordinary least squares method, which gives the best unbiased linear estimate when its assumptions are met. One of these assumptions is that the distribution of the random error term is normal with a mean of zero and a constant variance. The presence of outliers leads to a violation of this assumption, and therefore the least squares method cannot be applied. Robust methods have emerged as a suitable alternative when outliers appear, such as the M method, LMS, LTS, and the S estimator. However, the appearance of leverage points as another type of outlier in the space of explanatory variables has led to inaccuracy in the estimates of these methods. Very distinctive methods have been proposed in the statistical literature, such as the MM estimator method and the GM6 method, but the emergence of the problem of high leverage points, the reason for whose existence is attributed to the emergence of the phenomena of (Masking and Swamping). These two phenomena indicate inaccuracy in diagnosing leverage points, which causes either a loss of information or failure to completely treat bad leverage points. The proposal of this research seeks to improve the performance of the GM6 method, even in the presence of these two phenomena, by employing a high-precision diagnostic algorithm called IDRGP (RMVN) in order to obtain high-precision weights to reduce the impact of leverage points in the GM6 method. We have put the accurate diagnosis and GM6 method into a single algorithmic framework that we called GM6.IDRGP(RMVN), which showed outstanding performance compared to previous methods through the results of simulation studies and real data.

Correspondence:

Hassan S. Uraibi

hassan.uraibi@qu.edu.iqDOI: <https://doi.org/10.55562/jrucs.v56i1.3>