**المجلة العراقية للعلوم الاحصائية** (16) 2009
**ص ص** [219–242]

# تحسين أسلوب M الحصين في تقدير معلمات نموذج الانحدار الخطي

**صالح مؤيد شاكر***

---

**الملخص :**

تتلخص فكرة البحث في إيجاد تركيبة حصينة موزونة لمقـدرات معلمـات نموذج الانحدار الخطي بدمج أسلوب M الحصين وطريقة المربعـات الـصغرى الموزونة لمعالجة اثر القيم الشاذة والتي لا تنسجم ونمط البيانات العـام ، وبـالاتي الحصول على مقدرات جديدة ، وبالاعتماد على معيار متوسط مربعات الخطأ تـم فحص كفاءة المقدرات الجديدة والتي تبين أن نتائجها ذات دقة عالية مقارنـة مـع طريقة المربعات الصغرى الاعتيادية و المربعات الصغرى الموزونة وأسـلوب M الحصين.

**الكلمات الدالة:القيم الشاذة ، مقدرات M الحصين ، المربعات الصغرى الموزونة.**

## Modified Robust M-approach to estimate the parameters of linear Regression model

**Key Words: Outliers , Robust M Estimators , Weighted Least Squares Estimators**

Abstract :

The idea of this research is to find construction Weighted Robust for the estimate parameters of the linear regression model by blend Robust M- approach and Weighted Least Squares method to handle the influence of Outlier Values, which don't harmonies with the system of data set, in order to get new estimators, The mean square error criterion is used to check the efficiency of the new method with OLS , WLS and Robust M-estimators.

---

* مدرس مساعد – كلية علوم الحاسبات والرياضيات / قسم الإحصاء والمعلوماتية
تاريخ التسلم :2009/12/13 ———————— تاريخ القبول : 22 /3 /2010

## 1- المقدمة:

تعتمد الطرائق الإحصائية على مجموعة مـن الفرضـيات المهمـة فـي الحصول على نموذج الانحدار الدقيق ،وتعد معلومية التوزيع الاحتمـالي للبيانـات إحدى أهم هذه الفرضيات التي هي قيد الدراسة والتي عادةً ما تتوزع وفق التوزيع الطبيعي ،كما تعد هذه الفرضية إطارا لكل الطرائق الكلاسيكية في مجـال تحليـل التباين وتحليل متعدد المتغيرات.

يتصف النموذج الطبيعي للبيانات غالباً بتوزيعهِ توزيعاً طبيعياً، إلا انه أحيانـاً قـد تأخذ البيانات الموزعة نمطاً مختلفاً أو لا تتمثل بنمط معين من التوزيعات، والسبب أحيانا يعود إلى وجود القيم الشاذة (Outliers) إذ أن لهذه القيم التأثير السلبي على نتائج الطرائق الإحصائية ، وقد وضعت أساليب عديدة لمعالجة هذه القيم عند تقدير معلمات نموذج الانحدار أهمها أسلوب M الحصين إلا أن هذا الأسلوب يعتمد على معالجة متجه الأخطاء لطريقة المربعات الصغرى الاعتيادية فقـط **[Doddy and Ibnu, 2007]**، **[عبد الأحد ،2004]** دون معالجة الخلـل الموجـود فـي المتغيـرات التوضيحية لذا وانطلاقا من هذا قمنا بدمج أسلوب تقدير المعلمات بطريقة المربعات الصغرى الموزونة بأسلوب M الحصين إذ أن أسلوب تقـدير المعلمـات بطريقـة المربعات الصغرى الموزونة يعمل بمبدأ معالجة بيانـات المتغيـرات التوضـيحية **[John and Nico,1991]** وذلك باستعمال مصفوفة الأوزان ومن ثم تقدير معلمات النموذج.

تضمن هذا البحث جانبين,الأول تمثل بتعريف القـيم الـشاذة وأثرهـا وأسـاليب معالجتها،وعرض أسلوب M الحصين في معالجة هذه القيم وتوضيح مبدأ طريقـة المربعات الصغرى الموزونة وأخيرا طرح فكرة دمج أسـلوب M الحـصين مـع طريقة المربعات الصغرى الموزونة وإيجاد صيغة خاصة لتقدير معلمات النمـوذج لهذه الطريقة .أما الجانب الثاني فتمثل بتطبيق الطرائق الثلاثة على بيانات حقيقيـة وبيانات مولدة باستعمال أسلوب المحاكاة تحت عدة حالات من التطرف.وتلخـصت النتائج في أن أسلوب الدمج بين خواص أسلوب M الحصين وطريقـة المربعـات

الصغرى الموزونة قد أعطت نتائج ذات دقة عالية مقارنة بنتائج طريقة المربعـات الصغرى الاعتيادية والمربعات الصغرى الموزونة وأسلوب M الحصين.

## 2- بعض طرائق الكشف عن المشاهدات الشاذة

إن طريقة المربعات الصغرى الاعتيادية التي تأخذ العلاقة :

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y \quad \text{.........................................................(1)}$$

تشترط توافر عدد من الشروط من اجل الحصول على مقدرات دقيقة لمعلمات نموذج والذي يتمثل بالمعادلة الآتية:

$$Y = X\beta + e \quad \text{.......................................................................(2)}$$

إذ أن:

Y: متجه متغير الاستجابة وهو ذو سعة (n×1).

X: مصفوفة المتغيرات التوضيحية وهي ذات سعة [n×(m+1) ].

$\beta$:متجه معلمات الانحدار وهو ذو سعة [1× (m+1) ].

e:متجه الأخطاء العشوائية وهو ذو سعة (n×1).

m:عدد المتغيرات التوضيحية.

وإحدى هذه الشروط هي التوزيع الطبيعي لمتجه الأخطاء وتجانس تباين الأخطاء (Heteroscedastic) إلا إن هذه الشروط لا تتحقق في كثير من الحالات والسبب يعود إلى وجود القيم المتطرفة في متجه الأخطاء أو في قيم المتغيرات التوضيحية ، لذا فان تشخيص هذه القيم ومعالجتها قبل تحليل البيانات أمر في غاية الأهمية مـن اجل الحصول على نتائج واقعية ودقيقة **[Maronna & Martin,2006]** لقد وضعت أساليب عديدة للكشف عن هذه القيم نذكر منها :

1-الرسم الصندوقي :وهو احد الأساليب الرسومية التي توضح القيم ذات التطـرف القـوي (Extreme Outliers ) والقـيم ذات التطـرف المعتـدل ( Middle Outliers).

2-فحص عناصر قطـر المـصفوفة [Hat –Matrix (H)] **[Alan & David, 2003]** التي تسمى قيم قوة الرفع (Leverage Values) وعادة ما تستعمل للكشف عن القيم الشاذة في المتغيرات التوضيحية،إذ أن

$$H_{(n\times n)} = X(X'X)^{-1}X' = \begin{bmatrix} h_{11} & h_{12} & \ldots\ldots & h_{1n} \\ h_{21} & h_{22} & \ldots\ldots & h_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ h_{n1} & h_{n2} & \ldots\ldots & h_{nn} \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

فإذا كانت :

$h_{ii} > \frac{3m}{n}$ or $h_{ii} > \frac{2m}{n}$ فهذا يدل على أن المشاهدة (i) هي مشاهدة متطرفة.

وقد ذكر **[Belsly & Welsch,1980]** بأنه في حالة كون $h_{ii} > 0.2$ فمن الممكن أن تكون المشاهدة متطرفة أيضاً.

3-مقياس مربع مسافة كوك Cook's Squared Distance $CD^2_{[i]}$ **[Belsley & Kuh,1980]**

يستعمل هذا المقياس لاختبار تأثير حذف المشاهدة [i] على المعلمـات المقـدرة إذ أن:

$$CD^2_{[i]} = \frac{[\hat{\beta} - \hat{\beta}_{[i]}]' Z [\hat{\beta} - \hat{\beta}_{[i]}]'}{k}$$

$\hat{\beta}$:متجه المعلمات المقدرة قبل حذف المشاهدة [i].

$\hat{\beta}_{[i]}$: متجه المعلمات المقدرة بعد حذف المشاهدة [i].

$$k = mS_r^2 \ , \ Z = X'X$$

علما أن المشاهدة [i] في هذا الأسلوب تتمثل بأي صف من صفوف المـصفوفة U إذ أن :

$$U_i = [Y_i : X_{i1} \ \ X_{i2} \ldots\ldots\ldots\ldots X_{im}]$$

فإذا كان

$$CD^2_{[i]} > F_{tab}[(\alpha),(m,n-m)]$$

فهذا يدل على أن المشاهدة (i) هي مشاهدة متطرفة ولها تأثير على المعلمة $\hat{\beta}$.

4- مقياس DF Deferent fit when Excluding the ith case

**[Rcusseeum & Leroy,1987]**

يستعمل هذا المقياس في الكشف عن وجود القيم المتطرفة في متجه الأخطاء (e) إذ أن:

$$DF = \frac{e_i (h_{ii})^{\frac{1}{2}}}{S_{[i]}(1 - h_{ii})}$$

$S_{[i]}$: الخطأ القياسي بعد حذف المشاهدة [i].

فإذا كانت

$$DF > 2\sqrt{\frac{m}{n}}$$

فهذا يدل على أن المشاهدة [i] متطرفة ولها تأثير على تباين الخطأ.

## 3- مقدرات M الحصينة

بعد الكشف عن القيم المتطرفة وتشخيص تأثيرها تتم في الخطوة اللاحقة معالجة هذه القيم وذلك باستعمال الأساليب الحصينة وإحدى أهم هذه الأساليب هو أسلوب (M) الحصين (R.M) إذ تتلخص هذه الطريقة في تحجيم تأثير قيم البواقي الكبير باستعمال معادلة $\phi^*_{(e_i)}$ والتي تعرف بالصيغة الآتية **[Huber, 1964]**:

$$\phi^*_{(e_i)} = \frac{\Psi(e_i)}{e_i} \qquad \text{................(4)}$$

**إذ أن:**

$$\psi_{e_i} = \max\{-c, \min(e_i, c)\}, \quad c > 0 \quad \text{.......................................(5)}$$

وان: $c = 1.5$ or $c = 1.7$ وان $e_i = y_i - \hat{y}_i$

وبعد الحصول على قيم $\psi_{(e_i)}$ تتم موازنة قيمها من خلال العلاقة (4).

ومن المعادلة (4) يمكننا ملاحظة انه عندما تكون $|e_i| \to \infty$ فـان قيمـة $\phi^*_{(e_i)}$ تكون صغيرة وتقترب من الصفر $[\phi^*_{(e_i)} \to 0]$ وباستعمال مبدأ طريقة المربعـات الصغرى الاعتيادية والقاضي بتصغير مجموع مربعات الخطأ إلى اقل مـا يمكـن

للحصول على مقدرات M الحصينة (R.M) الاعتيادية وحسب العلاقة الآتية: **[Dorestt & Gunst, 1983]**

$$\hat{B}_M = (X'\phi X)^{-1} X'\phi Y \quad ..................................(6)$$

إذ أن :

$$\phi = \begin{bmatrix} \phi^*(e_1) & 0 & 0 & 0 \\ 0 & \phi^*(e_2) & 0 & 0 \\ 0 & 0 & . & 0 \\ 0 & 0 & 0 & \phi^*(e_n) \end{bmatrix}$$

## 4- مقدرات M الحصينة الموزونة

Robust M- Weighted Estimators (R.M.W)

يختص أسلوب M الحصين بمعالجة القيم المتطرفة الموجودة في متجه الأخطاء إلا انه في حالة وجود مثل هذه القيم في مصفوفة المتغيرات التوضيحية ، يصبح هذا الأسلوب غير قادر على معالجة هذه القيم والتخلص من أثرها ، لذا نقترح أولاً تعديل القيم المتطرفة الموجودة في مصفوفة المتغيرات التوضيحية باستعمال مصفوفة الأوزان لطريقة المربعات الصغرى الموزونة Weighted Least Squares (W.L.S) ومن ثم معالجة القيم المتطرفة الموجودة في متجه متغير الاستجابة من خلال استخدام متجه الأخطاء (متجه أخطاء المربعات الصغرى الموزونة) باستعمال أسلوب M الحصين ثانيا، وأخيرا إيجاد المقدرات الجديدة بعد التعديل الأخير وهذه المقدرات سيطلق عليها اسم مقدرات M الحصينة الموزونة(المقترحة)

Robust M- Weighted Estimator(R.M.W)

ويمكننا توضيح مراحل الوصول إلى هذه المقدرات على النحو كالآتي:

$$[Y_i : X_{ij}] = [Y_i : X_{i0}, X_{i1}, X_{i2}, .............., Xim], \quad i = 1, 2, ......., n$$

$$\hat{Y}_i = \hat{\beta}_0^{(w)} X_{i0} + \hat{\beta}_1^{(w)} X_{i1} + \hat{\beta}_2^{(w)} X_{i2} + ........... + \hat{\beta}_m^{(w)} X_{im}$$

إن المبدأ المستند عليه في إيجاد مقدرات نموذج الانحدار هـو تـصغير مجمـوع مربعات الخطأ إلى اقل ما يمكن، بذلك فان مقدرات المربعات الصغرى الموزونـة ستكون كالآتي:

$$\text{minimize} \sum_{i=1}^{n} \rho(e_{i(w)})$$

إذ أن

$$\rho(e_{i(w)}) = e_{i(w)}^{2} \quad , \quad \psi_{e_{i(w)}} = \frac{\partial \rho(e_{i(w)})}{\partial e_{i(w)}}$$

إذ أن $\hat{\beta}^{(w)}$ هو متجه مقدرات المربعات الصغرى الموزونة والتي تأخـذ الـصيغة الآتية:

$$\hat{\beta}^{(w)} = (X'WX)^{-1} X'WY$$

بنفس الطريقة يمكن إيجاد مقدرات M الحصين الموزون المقترح وكما يلي:

متجه الأخطاء بشكل عام سيكون $e_{i(w)}$ وذلك لأننا استعملنا $\hat{\beta}^{(w)}$ بـدل $\hat{\beta}_{OLS}$, أي أن:

$$e_{i(w)} = Y_i - \hat{Y}_i = Y_i - X_i' \hat{\beta}^{(w)}$$

$$\psi_{e_{i(w)}} = \max\{-c, \min(e_{i(w)}, c\} \quad , \quad c = 1.5 \ , \ c = 1.7$$

$$\phi_{e_{i(w)}} = \frac{\psi_{e_{i(w)}}}{e_{i(w)}}$$

$$\sum_{i=1}^{n} X_{ij} \phi_{e_{i(w)}} e_{i(w)} \, , \quad j = 1,2,\ldots\ldots, m$$

$$\sum_{i=1}^{n} X_{ij} \phi_{e_{i(w)}} (Y_i - X_i' \hat{B}) = 0$$

$$\sum_{i=1}^{n} X_{ij} \phi_{e_{i(w)}} Y_i - \sum_{i=1}^{n} X_{ij} \phi_{e_{i(w)}} X_i' \hat{B} = 0$$

الحد الأول:

$$\sum_{i=1}^{n} X_{ij} \phi_{e_{i(w)}} Y_i = X_j' \phi_{(w)} Y$$

إذ أن $X_j'$ هو مدور العمود j.

الحد الثاني:

$$\sum_{i=1}^{n} X_{ij}\,\phi_{e_{i(w)}}\,X_i'\,\hat{\beta} = (X_{1j}\,\phi_{e_{1(w)}}\,X_1' + \dots\dots\dots + X_{nj}\,\phi_{e_{n(w)}}\,X_n')\,\hat{\beta}$$

$$\sum_{i=1}^{n} X_{ij}\,\phi_{e_{i(w)}}\,X_i'\,\hat{\beta} = X_j'\,\phi_{(w)}\,X\,\hat{\beta}$$

$$\therefore \sum_{i=1}^{n} X_{ij}\,\phi_{e_{i(w)}}\,Y_i - \sum_{i=1}^{n} X_{ij}\,\phi_{e_{i(w)}}\,X_i'\,\hat{\beta} = X_j'\,\phi_{(w)}\,Y - X_j'\,\phi_{(w)}\,X\,\hat{\beta}$$

بضرب السابق لطرفي المعادلة أعلاه بالعمود $X_j$ نحصل على

$$X_j X_j'\,\phi_{(w)}\,Y - X_j X_j'\,\phi_{(w)}\,X\hat{\beta} = 0$$

إن $X_j$ هو متجه عمودي، وبما أن $X_j X_j' = I_M$ فان

$$\phi_{(w)}\,Y - \phi_{(w)}\,X\hat{\beta} = 0$$

وبالضرب السابق لطرفي المعادلة الأخيرة بـ 'X نحصل على

$$X'\phi_{(w)}\,Y - X'\phi_{(w)}X\hat{\beta} = 0$$

بضرب طرفي المعادلة أعلاه بـ $(X'\phi_{(w)}X)^{-1}$ نحصل على

$$(X'\phi_{(w)}X)^{-1}X'\phi_{(w)}Y - (X'\phi_{(w)}X)(X'\phi_{(w)}X)^{-1}\hat{\beta} = 0$$

$$\therefore \hat{\beta}_M^{(w)} = (X'\phi_{(w)}X)^{-1}X'\phi_{(w)}Y \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(7)$$

إن المعادلة (7) تمثل صيغة أسلوب M الحصين المـوزون المقتـرح(R.W.M) والذي يتم بوساطته معالجة التطرف الموجود في المتغيرات التوضيحية أو متغيـر الاستجابة أو كليهما معاً.

لقد وضعت أساليب عديدة لاختيار أوزان المصفوفة W أو تقـديرها ،وفـي هـذا البحث تم استخدام الأسلوب الذي وضعه كل من **[Dorsett & Gunst,(1983)]** وهو عبارة عن أسلوب تكراري لتقدير قيمة أوزان المشاهدات للمتغيـرات التوضـيحية وخطوات إيجاد هذه المصفوفة هي كالآتي:

1. افتراض قيمة أولية للأوزان وهي $W_i = 1$
2. حساب المقدار $h_i(w)$ كما يأتي

$$h_{i(w)} = W_i^2 X_i'(X'WX)^{-1} X_i$$
$$= W_i^2 d^2(X_i) \quad \text{..............................................(8)}$$

**Where** $d^2(x_i) = X_i'(X'WX)^{-1} X_i$

إن المقدار $h_i(w_i)$ في المعادلة (8) هو المقدار نفسه في المعادلة (3) ولكن هنا(المقدار $h_i(w_i)$) يعود لمقدرات المربعات الصغرى الموزونة.

3. حساب الوزن $W_i$ من المعادلة

$$W_i = \min(1, \eta / d(X_i)) \text{........................}(9)$$

تكرار الخطوة (2) و (3) حتى يحصل تقارب في قيم $h_i(w_i)$ .

وقد حدد قيمة $\eta^2$ التي تمثل الحد الفاصل بين قيمة المشاهدة i المقابلة للقيمة $h_i(w_i)$ في مصفوفة H بوصفها قيمة متطرفة أم لا.

وهنا لابد لنا أن نذكر أن الخطأ في حالة المربعات الصغرى الموزونة سيكون بالصيغة الآتية $e_i = Y_i - WX\hat{\beta}$ وهذا ما أثبته **[Dorestt & Gunst, 1983]** وبذلك فان الخطأ بأسلوب M الحصين الموزون سيكون بالصيغة الآتية $e_i = Y_i - \phi_{(w)}X\hat{\beta}$ وهذا هو الفرق بين الخطأ بطريقة المربعات الصغرى الموزونة والخطأ لأسلوب M الحصين الموزون المقترح الذي توصلنا إليه.

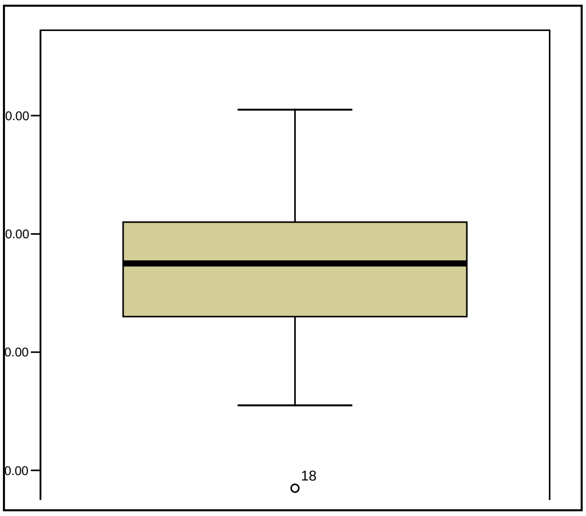
## 5- الجانب التطبيقي.

في هذا الجانب من البحث تم تطبيق ما ورد في الجانب النظـري علـى بيانـات لدراسة تتمثل بتأثير عمر الطفل بالشهور (X) على درجات كيـسل المكيفـة (Y) **[Micky and Clark,(1967)]** .

في الخطوة الأولى تم اختبار وجود القيم المتطرفة فـي كـلا المتغيـرين(متغيـر الاستجابة والمتغير التوضيحي) باستعمال الرسم الصندوقي كما هو موضـح فـي الشكلين (1) و (2) أدناه:

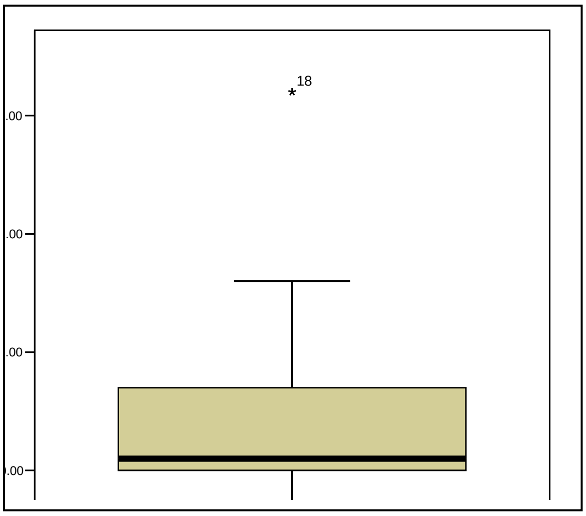
الشكل (1)

يوضح الرسم الصندوقي لمتغير الاستجابة (Y)



الشكل (2)

يوضح الرسم الصندوقي للمتغير التوضيحي (X)



من خلال ملاحظة الشكلين أعلاه نجد أن المتغير (Y) يحتوي على قيمـة معتدلـة التطرف وهي المشاهدة ($Y_{18}$) أما المتغير (X) فيحتوي على قيمة متطرفة وهـي ($x_{18}$) .

أما تأثير القيم المتطرفة ومقاييس الكشف عنها في الطرائق الأربعة المعتمـدة فـي البحث فكانت كالآتي:

**الالجدول (1) يوضح مقاييس الكشف عن التطرف لبيانات Mickey & Clark**

| | O.L.S | | | W.L.S | | | R.M | | | R.M.W | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cutt-off | 4.3 | 0.09 | 0.436 | 4.3 | 0.09 | 0.436 | 4.3 | 0.09 | 0.436 | 4.3 | 0.09 | 0.436 |
| ت | $CD^2$ | hii | DF | $CD^2$ | hii | DF | $CD^2$ | hii | DF | $CD^2$ | hii | DF |
| 2 | | 0.107 | | | 0.156 | | | * | | | * | |
| 5 | | | 0.455 | | | 0.505 | | | * | | | * |
| 6 | | | | | | | | 0.096 | | | | |
| 10 | | | 0.522 | | | 0.535 | | | * | | | 0.667 |
| 11 | | | 1.099 | | | 1.454 | | | * | | | 0.439 |
| 17 | | | 0.476 | | | 0.484 | | | * | | | * |
| 18 | | 0.604 | | | * | | | 0.569 | | | 0.370 | |
| 19 | | | 1.697 | | | 2.507 | | | * | | | * |

القيمة * :تمثل مقياس الكشف الذي يظهر معالجة التطرف لهذه القيمة إذ يكون اقل من نقطة القطع للمشاهدة المتطرفة بذلك تم حذفها ووضع العلامة (*).

الخليـــة باللون الفاتح : هي قيمة مقياس الكشف الذي يظهـر تخفيف حجم التطرف مقارنة بالطرائق السابقة ولكن أعلى من نقطة القطع للمشاهدة المتطرفة.

القيمة المضللة باللون الغامق : هي قيمة مقياس الكشف الذي يكشف عن وجـود تطرف جديد لم يكن موجود في الطريقة السابقة (أي لم تكن موجودة في المربعات الصغرى مثلا وظهرت بطريقة المربعـات الـصغرى الموزونـة أو أسـلوب M الحصين)فتم تظليلها باللون الغامق لتحديدها.

من ملاحظة النتائج في الالجدول (1) أعلاه يمكننا القول بان R.M.W قـد عـالج نسبة كبيرة من القيم المتطرفة مقارنة بكل من W.L.S و R.M ،وهذا واضح مـن خلال القيم ذات اللون الفاتح ورمز * .

ولإيجاد معلمات ومتوسط مربعات خطأ لكل نموذج من النمـاذج المدروسـة تـم استخدام البرنامج الجاهز SPSS إذ كان عدد التكرارات أو الدورات هي 6 دورات إلى أن تم الحصول على تقارب في قيم $h_i(w)$ والنتائج موضحة في الالجدول (2) الاتي:

**الجدول (2) يوضح قيم مقدرات الأساليب الأربعة لبيانات Micky & Clark**

| المقدرات | O.L.S | W.L.S | R.M | R.M.W |
|---|---|---|---|---|
| $\hat{\beta}_0$ | 109.874 | 106.171 | 112.422 | 110.067 |
| $\hat{\beta}_1$ | -1.127 | -0.814 | -1.269 | -1.060 |
| MSE | 121.505 | 113.971 | 120.127 | 13.176 |

من ملاحظة نتائج متوسط مربعات الخطأ لكل أسلوب نجد أن الأسـلوب المقتـرح R.M.W قد أعطى نتائج ذات دقة عالية مقارنة بالطرائق الأخرى الباقية وهذا مـا يعزز ما توصلنا إليه في الالجدول (1) السابق.

## 6- الجانب التجريبي:

في هذا الجانب تم استخدام أسلوب المحاكاة لتوليـد بيانـات لمتغيـرات عـشوائية ولحالتين وكما يلي:

### الحالة الأولى:

تمثلت هنا الدراسة بتوليد متغير توضيحي واحد وفق التوزيع الطبيعـي بمتوسـط واحد وتباين متغير كما تم توليد متجه أخطاء وفق التوزيع الطبيعي بمتوسط صـفر وتباين مختلف من اجل توفير قيم شاذة بعد ذلك يتم إيجاد مصفوفة الارتبـاط $X'X$ وإيجاد الجذور المميزة لها والتي ستكون قيم بديلة للمعلمات $\beta_1$ $\beta_0$ ومن ثم إيجاد y حسب العلاقة

$$y = \beta_0 + \beta_1 X_1 + e_i$$

تم توليد ثلاث أحجام عينات (n=50 و n=100 و n=500) وفي كـل عينـة تـم إجراء عملية التلويث لقيم Xi و ei وكما هو موضح في الالجدول (3).

**العينة الأولى**: n=50

1- عدم تلويث قيم (X) و (Y).

2- تلويث قيم X فقط بنسبة 10% إذ ولدت 45 مشاهدة غير شـاذة تـسلك وفـق التوزيع الطبيعي بمتوسط (1) وتباين (1) ومن ثم توليد 5 مشاهدات شـاذة تـسلك وفق التوزيع الطبيعي بمتوسط (1) وتباين (50).كما ولدت 50 مشاهدة غير شـاذة لمتجه الأخطاء تسلك وفق التوزيع الطبيعي بمتوسط (0) وتباين (1).

3- تلويث قيم $(e_i)$ فقط بنسبة 10% ،إذ ولدت 45 مـشاهدة غيـر شـاذة لمتجـه الأخطاء تسلك وفق التوزيع الطبيعي بمتوسط (0) وتباين (1) وتوليد 5 مـشاهدات شاذة تسلك وفق التوزيع الطبيعي بمتوسط (0) وتباين (50)،أما مشاهدات المتغيـر التوضيحي (X) فقد ولدت وفق التوزيع الطبيعي بمتوسط (1) وتباين (1).

4- تلويث قيم (X) وقيم $(e_i)$ بنسبة 20% لكل منهما ،إذ ولدت 40 مـشاهدة غيـر شاذة للمتغير X تسلك وفق التوزيـع الطبيعـي بمتوسـط (1) وتبـاين (1) و 10 مشاهدات شاذة وفق التوزيع الطبيعي بمتوسط (1) وتباين (5) ، أما متجه الأخطاء فقد ولدى 40 مشاهدة غير شاذة تسلك وفق التوزيع الطبيعي بمتوسط (0) وتبـاين (1) و10 مشاهدات شاذة بمتوسط (0) وتباين (5).

وعملية التلويث للعينـات الثلاثـة (n=50 و n=100 و n=500) ملخـصة فـي الالجدول(3).

بعد حالات التلويث تم إيجاد مقاييس الكشف ومعالجة القـيم المتطرفـة للطرائـق المعتمدة في البحث ولكثرة المشاهدات المتطرفة والمشاهدات التي تم معالجتها فقـد لخصت النتائج على شكل نسب مئوية والموضحة في الجدول (4).

من ملاحظة حقل مجموع المعالجة في الالجدول (4) أعـلاه يمكننـا القـول بـان R.M.W قد عالج نسبة كبيرة من القيم المتطرفة مقارنـة بكـل مـن W.L.S و R.M.

أما عدد التكرارات أو الدورات لكل عينة ولكل حالة تطرف حتى تم الحصول على تقارب في قيم $h_i(w)$ كانت كما موضحة في الالجدول (5) أدناه:

**الجدول (5) يوضح عدد الدورات لكل عينة ولكل حالة تطرف بتوليد متغير توضيحي واحد(نموذج بسيط)**

| حجم العينة | تطرف X | تطرف y | تطرف X y |
|---|---|---|---|
| n=50 | 7 دورة | 6 دورة | 5 دورة |
| n=100 | 6 دورة | 4 دورة | 5 دورة |
| n=500 | 5 دورة | 4 دورة | 4 دورة |

أما المعلمات المقدرة ومتوسط مربعات الخطأ للطرائق المعتمدة ولكل عينة وحالـة تطرف فهي موضحة في الجدول (6) الأتي:

من ملاحظة نتائج التجربة الأولى نجد أن هناك تقارب كبير فـي نتـائج متوسـط مربعات الخطأ في كل من الطرائق الأربعة المعتمدة في البحث في حالة عدم وجود تطرف، وفرق واضح في نتائج متوسط مربعات الخطأ للطرائق الأربعة عند وجود القيم المتطرفة في كل من المتغير التوضيحي أو متجه الخطأ أو في كليهما إذ أثبتت الطريقة المقترحة R.M.W بأنها أعلى الطرائق البقية وذلك لامتلاكها أقل متوسـط مربعات خطأ.

**الحالة الثانية:**

تمثلت هذه الحالة في توليد ثلاثة متغيرات توضيحية ومتجه أخطاء وفـق التوزيـع الطبيعي وبثلاثة أنواع من العينات وهي (n=50 و n=100 و n=500) وفي كـل عينة تم إجراء عملية التلويث لقيم Xi و ei وكما هو موضح في الالجدول (7).

بعد حالات التلويث تم إيجاد مقاييس الكشف ومعالجة القـيم المتطرفـة للطرائـق المعتمدة في البحث ولكثرة المشاهدات المتطرفة والمشاهدات التي تم معالجتها فقـد لخصت النتائج على شكل نسب مئوية والموضحة في الالجدول (8).

من ملاحظة حقل مجموع المعالجة في الجدول (8) أعـلاه يمكننـا القـول بـان R.M.W قد عالج نسبة كبيرة من القيم المتطرفة مقارنـة بكـل مـن W.L.S و R.M.

أما عدد التكرارات أو الدورات لكل عينة ولكل حالة تطرف حتى تم الحصول على تقارب في قيم $h_i(w)$ كانت كما موضحة في الجدول (9) أدناه:

**الجدول (9) يوضح عدد دورات لكل عينة ولكل حالة تطرف بتوليد ثلاث متغيرات توضيحية(نموذج متعدد)**

| حجم العينة | تطرف X | تطرف y | تطرف X y |
|---|---|---|---|
| n=50 | 6 دورة | 5 دورة | 6 دورة |
| n=100 | 7 دورة | 3 دورة | 5 دورة |
| n=500 | 5 دورة | 3 دورة | 4 دورة |

أما المعلمات المقدرة ومتوسط مربعات الخطأ للطرائق المعتمدة ولكل عينة وحالـة تطرف فهي موضحة في الجدول (10) الأتي:

من ملاحظة نتائج التجربة الثانية نجد فرقاً واضحاً في متوسـط مربعـات الخطـأ للطرائق الأربعة عند وجود القيم المتطرفة فـي المتغيـر التوضـيحي أو متغيـر الاستجابة أو في كليهما مع بقاء دقة الطريقة المقترحة R.M.W أعلى من الطرائق البقية وذلك لامتلاكها أقل متوسط مربعات خطأ.

## 7- الاستنتاجات

1- عدم وجود فروق في دقة نتائج الطرائق المدروسة عند عدم وجود القيم المتطرفة.

2- لا يمكن الاعتماد على مقدرات أسلوب M الحصين في حالة وجود تطرف قوي في مشاهدات المتغير (المتغيرات التوضيحية) التوضيحي أو حتى في متغير الاستجابة.

3- من نتائج دراسة الحالة التطبيقية والتجربتين يمكن القول بان أسلوب M الحصين الموزون R.M.W المقترح قد اثبت كفاءة في تقدير معلمات نموذج الانحدار وبدقة عالية عند وجود التطرف في قيم المتغير (المتغيرات) التوضيحية أو في المتغير المعتمد أو في كليهما.

## 8- المصادر


1. عبد الأحد ،مناهل دانيال،(2004)،"التقدير الحصين في نموذج الانحدار الذاتي من الرتبة الأولى"،رسالة ماجستير مقدمة إلى مجلس كلية علوم الحاسبات والرياضيات،جامعة الموصل،غير منشورة.
2. Alan D. Chare and David J. Thomson ,(2003),"A Bounded Influence Regression Estimator based on the Statistics of the Hat Matrix ,Appl. Statist., 52 , Part 3,pp.(307-322).
3. Belsley , D. A. Kuh , E. and Welsch , R. E. ,(1980),"Regression Diagnostics :Identifying Influential Data and Sources of Collinearity. Wiley ,New York.
4. Dovalee Dorsett and Richard F. Gunst ,(1982),"Bounded-Leverage Weights for Robust Regression Estimators", Technical Report 171.Southern Methodist University ,Dept. of Statistics.
5. Huber , P. ,(1964),"Robust Estimation of Location Parameter ", The Annals of Mathematical Statistics , 35 ,PP.(73-101).
6. Johan C. Akkerboom and Nico J. Nieuwenbroek, Neth,( 1991), Representative Bounded-Leverage Regression Estimation for finite Populatins, Center. Bur. of Stat. Akkerboom.

7. Doddy Sutarno and Ibnu Fatrio,(2007),"Robust M-Estimation of CSAMT Impedance Functions ",Indonesian Journal of Physics, Vol.18 No.3 ,PP.(81-85).
8. Mickey , M. R. Dunn , O. J. and Clark, V. ,(1967),"Note on the use of Stepwise Regression in Detecting Outliers", Computers and Biomedical Research ,1 ,pp.(105-111)
9. Ricardo A. Maronna , R. Douglas Martin and Victor J. Yohai ,(2006),"Robust Statistics Theory and Methods ", John Wiley and Sons, Ltd.
10. Rousseeum ,R. J. and A. M. Leroy, (1987),"Robust Regression and Outlier Detection , New York: Wiley.