Website : www.uoajournal.com
E-mail : info@uoajournal.com

مجلة كلية المعارف الجامعة
Journal of AlMaarif University College

A Comparison of Some Robust and Non-Robust Methods for Estimating Regression Models with Practical Application

Mathil Kamil Thamer
College of Administration & Economy, University of Anbar, Iraq
mathilthamir@yahoo.com

ABSTRACT:

The method of linear regression models is an important topic in statistical analysis, when two or more explanatory variables are related to the regression relationship. A set of problems occurs in which a violation of one of the basic assumptions of the method of ordinary least squares leads to obtaining biased estimates , There are several methods suggested to address these problems, including the robust methods of regression analysis, that the ordinary least squares (OLS) method is sensitive to the presence of outlier values in the data as the break As the breakdown point is very low and leads to inaccuracy of the results for the estimated parameters, where the estimated features are characterized by inefficiency, therefore robust methods of least squares are used to address the problem of the presence of outliers values and leverage points in the data and other problems. In this research, a theoretical and practical comparison between the OLS method and the robust methods of estimating linear regression models (M estimators, MM estimators, and S estimators) was presented with a review of the most important methods of finding outliers values in the data weights functions that accompany the estimation process through The iterative weighted least squares method. As for the applied side, real data taken from the Central Statistical Organization for the statistical group for the year 2017 in Iraq which pertains to natural conditions has been analyzed in order to estimate a set of linear relationships in the usual way and robust methods and it has been proven that robust S estimators data provided us with the best model according to statistical standards using the program (Eviews 10).

Keywords: Linear Regression Models; Ordinary Least Squares Method; Iterative Weighted Least Squares Method; Robust Regression; Robust Methods; Outlier Values.

مقارنة ما بين بعض الطرائق الحصينة وغير الحصينة لتقدير نماذج الانحدار مع تطبيق عملي

م. ماثل كامل ثامر

كلية الإدارة والاقتصاد، جامعة الأنبار – العراق

mathilthamir@yahoo.com

ملخص البحث

ان اسلوب بناء نماذج الانحدار الخطي من المواضيع المهمة في التحليل الاحصائي ، عندما يرتبط اثنان او اكثر من المتغيرات التوضيحية في علاقة الانحدار تحدث مجموعة من المشاكل والتي يكون فيها خرق لأحد الفروض الأساسية لطريقة المربعات الصغرى الاعتيادية مما يؤدي الى الحصول على تقديرات متحيزة او اقل دقة. هنالك عده طرق اقترحت لمعالجة هذه المشاكل نذكر منها الطرائق الحصينة في تحليل الانحدار، ان طريقة المربعات الصغرى الاعتيادية (**OLS**) حساسة لوجود القيم الشاذة في البيانات حيث ان نقطة الانهيار منخفضة جدا وتؤدي الى عدم دقة النتائج للمعالم المقدرة حيث تمتاز المعالم المقدرة بعدم الكفاءة لذلك يتم استعمال الطرائق الحصينة للمربعات الصغرى لمعالجة مشكلة وجود القيم الشاذة ونقاط الرفع في البيانات وغيرها من المشاكل. وقد جرى في هذا البحث مقارنة نظرية وتطبيقية بين طريقة المربعات الصغرى الاعتيادية (**OLS**) مع الطرائق الحصينة لتقدير نماذج الانحدار الخطي وتمثلت (مقدرات **M**، مقدرات **MM**، ومقدرات **S**) مع استعراض اهم طرق ايجاد القيم الشاذة في البيانات ودوال الاوزان التي ترافق عملية التقدير من خلال طريقة المربعات الصغرى الموزونة التكرارية اما فيما يتعلق بالجانب التطبيقي فقد تم تحليل بيانات حقيقية اخذت من الجهاز المركزي للاحصاء المجموعة الاحصائية لعام ٢٠١٧ والتي تخص الاحوال الطبيعية ، وذلك لتقدير مجموعة من العلاقات الخطية بالطريقة الاعتيادية والطرق الحصينة واثبت ان مقدرات **S** الحصينة زودتنا بأفضل نموذج وفقا للمقاييس الاحصائية باستخدام البرنامج .(**Eviews 10**)

الكلمات المفتاحية: نماذج الانحدار الخطي، طريقة المربعات الصغرى الاعنيادية، طريقة المربعات الصغرى الموزونة التكرارية، الانحدار الحصين، الطرائق الحصينة، القيم الشاذة.

١- المقدمة

تقابلنا كثيرا في الحياة العلمية مواقف تتضمن متغيرين وأكثر ويكون المطلوب معرفة ما اذا كان هناك علاقة بين هذه المتغيرات وما هو شكل هذه العلاقة وايضا كيفية التنبؤ باحد هذين المتغيرين في حالة معرفتنا بالمتغير الاخر وبما ان الاهتمام منصب على وصف بيانات الظواهر فسيكون الوصف اكثر دقة وواقعية لو استطعنا معرفة قوة وشكل هذه العلاقة وبالتالي يمكن استخدامها في تقدير قيم بعض الظواهر (المتغيرات) بمعرفة قيم الظواهر الاخرى حيث ان العلاقة بين المتغيرات قد تكون قوية أو ضعيفة وقد تكون طردية أو عكسية .

٢ :- منهجية البحث

٢-١ :- أهداف البحث

ان الهدف من هذا البحث هو المقارنة بين طرائق التقدير غير الحصينة مثل طريقة المربعات الصغرى والتي تكون عرضة الى تاثير القيم الشاذه والطرائق الحصينة من خلال دراسة وتطبيق انواع من طرائق التقدير ومنها (مقدرات M ، مقدرات MM ، ومقدرات S) ومحاولة تطبيقها على بعض المؤشرات في مجال الاحوال الطبيعية في العراق .

٢-٢ :- اهمية البحث

تأتي اهمية دراسة الطرائق الحصينة لتقدير النماذج لأنها لا تتأثر بالقيم الشاذة وبالتالي يؤدي الى حصولنا على مقدرات كفوءة يمكن الاستدلال بها و نظرا لان النماذج تعد الاساس لدراسة وتحليل العلاقات السببية ما بين المتغيرات ولما لها من اهمية كبيرة للتأكد من وجود علاقة التأثير المنطقية بينها مع امكانية اجراء الاختبارات الاحصائية لاكتشاف فيما اذا كانت هناك فروق ذات دلالة احصائية معنوية لمعاملات النموذج ومعاملاته ما بين المتغيرات .

٢-٣ :- مشكلة البحث

تكمن مشكلة البحث في حساب العلاقة الاحصائية ما بين بعض متغيرات الاحوال الطبيعية لمحافظات العراق وبالاخص محاولة بناء نموذج استدلالي لعدد الاقضية من علاقته بعدد سكان المحافظة ومساحة المحافظة من خلال النماذج الحصينة لمعالجة القصور في البيانات وطريقة التقدير الاعتيادية ذات نقطة الانهيار المنخفضة وهي طريقة المربعات الصغرى .

٣- الجانب النظري

يتضمن الجانب النظري نموذج الانحدار الخطي مع تحديد طريقة التقدير من خلال المربعات الصغرى مع انواع القيم الشاذه في بيانات انموذج الانحدار، و المفاهيم الأساسية للقيم الشاذة وقيم الرفع ويتمثل ايضا بوجود مشكلة عدم التجانس في التباين في انموذج الانحدار حيث تم توضيح المفاهيم الأساسية لهذه المشكلة واثارها وكيفية التعامل مع مشكلة عدم التجانس عندما تتواجد القيم الشاذة في النموذج ، كما تطرق البحث الى طرائق التقدير الحصينة مع اهم دوال الاوزان المرافقة لعملية التقدير لنموذج الانحدار الحصين .

٣-١ :- نموذج الانحدار الخطي

ان انموذج الانحدار الخطي المتعدد يدرس تاثير مجموعة من المتغيرات المستقله على المتغير المعتمد ويتم تقدير انموذج الانحدار الخطي المتعدد بطريقة المربعات الصغرى (OLS) لغرض توفيق أفضل معادلة انحدار خطية متعددة المتغيرات مع الاختبارات الاحصائية المرافقة لها ، وان خط الانحدار يهدف الى تصغير مجموع مربعات الانحرافات عن الخط المستقيم الى أدنى حد ممكن ، وفق الصيغة الاحصائية الاتية للنموذج : (الراوي ، خاشع محمود، ١٩٨٧)

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i \ldots\ldots\ldots\ldots\ldots \; 1$$

$$i = 1.2.3. \ldots \ldots \ldots \ldots . n$$

ومن الملاحظ على النموذج الاخير انه يتعامل مع متغير توضيحي واحد x_{1i} وبالامكان ان نحصل على نموذج انحدار متعدد المتغيرات المستقلة من خلال اعادة صياغة النموذج اعلاه .

٣-٢ :- القيم الشاذة في البيانات

ان القيمة الشاذة توصف من خلال سلوك البيانات و تعرف بانها القيمة المرتفعة او المنخفضة للمشاهدة و التي تبتعد عن سلوك باقي البيانات او بعبارة أخرى هي المشاهدة التي تكون غير متسقة من حيث القيمة مع بقية المشاهدات وهي بذلك تبتعد عن مركز البيانات. ان وجود القيم الشاذة في بيانات انموذج الانحدار يمكن ان تكون في المتغير المعتمد (متغير الاستجابة) ويطلق عليها (Outlier) ، او وجود القيم الشاذة في المتغيرات التوضيحية وعندئذ فأنها تسمى بنقاط الرفع (Leverage points) وهي على نوعين اما نقاط رفع جيدة والتي لا توثر على تقدير معادلة الانحدار او غير جيدة توثر على معالم النموذج ، كما ان وجود القيم الشاذة قد تكون في المتغير المعتمد وكذلك وجودها في المتغيرات التوضيحية معاً مما يفاقم من مشكلة التقدير ، فطرق التقدير لنموذج الانحدار و منها طريقة المربعات الصغرى تكون غير كفوءة حيث ان اخطاء النموذج لا تتبع التوزيع الطبيعي وهذا الامر يتطلب استخدام طرائق الانحدار الحصين لتلافي هذه المشاكل والحصول على مقدرات كفوءة للمعالم. (Rousseeuw, P.J., and A. Leroy, 1987) (Campell,N.A.,et al,1998)

٣-٣ :- مشكلة عدم تجانس التباين (Hetroscedasticity)

من اهم فرضيات نموذج الانحدار هي فرضية تجانس التباين وعند مخالفة الفرضية يكون التباين غير متجانس وتعني ان التوقع لمربعات الاخطاء يكون $E\,\underline{\epsilon}\,\underline{\epsilon}' = \sigma^2\,\Omega$ ومن هنا فان مقدرات المربعات الصغرى تكون متحيزة وغير متسقة مع انعدام الكفاءة وهي بذلك ليست افضل مقدر خطي غير متحيز ، و يمكن معالجة مشكلة عدم التجانس في التباين و ذلك باستخدام طريقة المربعات الصغرى الموزونة (Weighted Least Square) لتقدير معلمات الانموذج وذلك وفق الصيغة الاتية: (الراوي ، خاشع محمود ،١٩٨٧)

$$\hat{\beta}_{WLS} = (x'\,\Omega^{-1}\,x)^{-1} x'\,\Omega^{-1} y \qquad \ldots\ldots\ldots\ldots\ldots 2$$

اذ ان مصفوفة Ω^{-1} تمثل مصفوفة قطرية تحتوي على اوزان وتوصف بالاتي :

$$\Omega^{-1} = diag(W_i)$$

وفي اغلب الاحيان تكون الاوزان غير معلومة عندها تستخدم طريقة المربعات الصغرى الموزونة (Feasible Weighted Least Square) ، بعد تقدير اوزان النموذج .

٣-٤ :- بعض مقاييس تحديد نوع المشاهدة

البواقي الحصينة (Robust residuals)

اذا كانت البواقي r_i اذ ان i=1,2,....,n والمستحصل عليها من التقديرات الحصينة للانحدار فان مؤشر تحديد القيم الشاذة يمكن تعريفة بالشكل الاتي: (Huber ,P.J. 1964)

$$\text{Outlier}= \begin{cases} 0 & if\ |\epsilon| \leq k\sigma \\ 1 & other\ wise \end{cases} \quad \ldots\ldots\ldots\ldots 3$$

اذ ان ϵ البواقي القياسية (standardized residuals) و $k\sigma$ حدود البواقي الحصينة.

مسافة مهلا نوبس Mahalinobes distant

يستخدم لاكتشاف نقاط الرفع في المتغيرات المستقلة والتي تؤثر على كفاءة معالم نموذج الانحدار المقدرة والمؤشر يكشف مدى ابتعاد المشاهدات عن مركزها ويوصف بالصيغة التالية : (Huber ,P.J. 1964)

$$md_i = \sqrt{(x_i - \hat{\mu})\, \Sigma^{-1} (x_i - \hat{\mu})^t} \quad \ldots\ldots\ldots\ldots\ldots 4$$

اذ ان: Σ تمثل مصفوفة التباين المشترك التقليدية المحتسبة من العينة .

مصفوفة القبعة hat matrix

توصف المصفوفة بالاعتماد على قيم المتغيرات المستقلة : (Huber ,P.J. 1964)

$$H = x(x^T x)^{-1} x^T \ldots\ldots\ldots\ldots\ldots\ldots 5$$

وتستخدم للكشف عن المشاهدات غير الجيدة والمؤثرة في المتغيرات التوضيحية .

المسافة الحصينة (Robust distant)

ويرمز لها بـ (RD) والتي تقوم على أساس تقديرات لمؤشرات الموقع ومصفوفة تباين مشترك حصينة والصيغة الخاصة باحتساب المسافة الحصينة هي :

$$RD_i = \sqrt{(x_i - t(X))^T C(x)^{-1} (x_i - t(X))} \quad \ldots\ldots\ldots\ldots\ldots\ldots 6$$

اذ ان t(x)، C(x) هي تقديرات معلمة الموقع ، ومصفوفة التباين المشترك على التوالي الحصينة المحتسبة على أساس MCD. فاذا كانت $c(p)$ هو قيمة حد القطع ، والذي يحسب بالشكل الاتي: (Maronna, R. A.,et al,2006)

$$c(p) = \sqrt{\chi 2_{p;(1-\propto)}} \ldots\ldots\ldots\ldots\ldots\ldots 7$$

p عدد المتغيرات التوضيحية بالأنموذج و$\propto$ مستوى المعنوية وعليه يمكن تحديد ان المشاهدة في البيانات من نوع الرفع Leverage عند تحقق الاتي

$$leverage = \begin{cases} 0 & if\ RD_i \leq c(p) \\ 1 & other\ wise \end{cases} \ldots\ldots\ldots\ldots 8$$

٣-٥ :- الانحدار الحصين: (Robust Regression)

ان تقديرات طريقة المربعات الصغرى لأنموذج الانحدار تكون حساسة وبشكل كبير جداً لوجود القيم الشاذة حيث ان نقطة الانهيار التي تمثل النقاط التي تجعل معالم نموذج الانحدار غير كفوءة هي عبارة عن نسبة التلوث بالبيانات تعتبر مؤشر لمدى حصانة المقدرات لوجود القيم الشاذة بهذه النسبة، وعلى هذا الأساس تم وضع العديد من الطرائق الحصينة التي تهدف الى الحصول على نقاط انهيار عالية . (Maronna, R. A.,et al,2006)

نقطة الانهيار Breakdown Point

تعتبر نقطة الانهيار من المعايير المهمة التي تقيس حصانة (Robustness) المقدرات فهي اصغر نسبة من البيانات الملوثة والتي تجعل التقدير ينهار وان افضل نقطة انهيار هي ٥٠٪ اما في حالة تجاوز هذه النسبة يصبح من المستحيل ان نميز بين الجزء الجيد وغير الجيد (الشاذ) من بيانات العينة . (Maronna, R. A.,et al,2006)

٣-٦ :- طرائق التقدير الحصينة

٣-٦-١ :- مقدرات M (M-Estimation):

ان هذه الطريقة من أكثر الطرائق الحصينة استخداماً وذلك للكفاءة العالية للمقدرات المستحصلة مع طريقة المربعات الصغرى وهي طريقة حصينة لوجود القيم الشاذة في قيم المتغير المعتمد (متغير الاستجابة) الا انها غير حصينة لوجود نقاط الرفع ، ولقد وصفت مقدرات M على انها امتداد لمقدرات الامكان الاعظم .

ان نموذج الانحدار الخطي المتعدد يوصف بالعلاقة التالية :

$$y_i = \beta_0 + \beta_1 \; x_{1i} + \beta_2 \; x_{2i} + \dots\dots + \beta_k \, x_{ki} + \varepsilon_i \dots\dots\dots\dots 9$$

$$i = 1.2.3. \dots\dots\dots\dots . n$$

وتعتمد عملية تقدير النموذج بطريقة المربعات الصغرى وفق اسلوب المصفوفات وبالعلاقة التالية :

$$\hat{\beta} = (x^T x)^{-1} \; x^T y \quad \dots\dots\dots\dots\dots\dots 10$$

حيث ان

$\hat{\beta}$ متجه لمعالم النموذج المقدر بابعاد $px1$. ، x مصفوفة المتغيرات المستقلة عمودها الاول مساوي للواحد وتكون بابعاد $n \; x \; p$ ، y متجه المتغيرالمعتمد وتكون بابعاد $n \; x \; 1$. ، ان مقدر M العام يصغر دالة الهدف (objective function) موصوفة لاخطاء النموذج وبالصيغة التالية :

$$\sum_{i=1}^{n} \rho(r_i) = \sum_{i=1}^{n} \rho(\, y_i - X_i' \; b) \quad \dots\dots\dots\dots\dots\dots\dots 10$$

ففي حالة اسلوب المربعات الصغرى لتقدير معالم النموذج فان دالة الاخطاء توصف $\rho(r_i) = {r_i}^2$ وان هذه الدالة قابلة للاشتقاق وفق $\rho' = \psi$، وان اشتقاق دالة الهدف يكون بالنسبة الى معلمات النموذج وبوضع المشتقات الجزئية مساوية الى الصفر، ينتج عنه منظومة من المعادلات التقديرية لمعلمات النموذج :

$$\sum_{i=1}^{n} \psi \left(y_i - X_i' \; \underline{b}\right) X_i' = 0 \quad \dots\dots\dots\dots\dots\dots\dots 11$$

ولتكن دالة الوزن $\mathrm{W}(\mathrm{r}_i) = \psi(\mathrm{r}_i) / \mathrm{r}_i$ وهي دالة وزن باخطاء النموذج، ولتكن $\mathrm{W}_i = \mathrm{W}(\mathrm{r}_i)$. وعليه فان المعادلات التقديرية يمكن كتابتها بدلالة الاوزان بالشكل الاتي:

$$\sum_{i=1}^{n} \mathrm{W}_i \left(y_i - X_i' \; \underline{b}\right) X_i' = 0 \quad \dots\dots\dots\dots\dots\dots 12$$

ان حل المعادلات التقديرية يكون من خلال اسلوب المربعات الصغرى الموزونة التكرارية (iteratively reweighted least-squares,) ويرمز له (IRLS)، ويتطلب تنفيذ عملية التقدير بمجموعة من الخطوات التكرارية وفق الاتي :

$$\underline{b}^{(t)} = [x' w^{(t-1)} x]^{-1} x' w^{(t-1)} y \quad \dots\dots\dots\dots\dots\dots\dots 13$$

بعد اعتماد تقديرات المربعات الصغرى كقيم اولية للمعلمات $\underline{b}^{(0)}$ وبالاعتماد على التكرار يتم تعديل الاوزان وفق دوال مختارة متعددة اي ان حساب مصفوفة الاوزان القطرية يكون كالاتي :

$$w^{(t-1)} = diag\{W_i^{(t-1)}\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots 14$$

(HUBER, P.J. 1981) (Montgomery, D. C., et al , 2001)

٣-٦-٢:- مقدرات المربعات المشذبة الصغرى الحصينة (Robust Least Trimmed Squares Estimation)

اقترحت الطريقة من قبل الباحث (Rousseeuw) عام (١٩٨٤) ولنموذج الانحدار الخطي المتعدد حيث يتم تصغير مجموع مربعات الاخطاء التالي :

$$Q_{LTS}(\theta) = \sum_{i=1}^{h} r_{(i)}^2 \qquad \ldots\ldots\ldots\ldots\ldots 15$$

اذ ان: $r_{(1)}^2 \leq r_{(2)}^2 \leq \ldots\ldots \leq r_{(n)}^2$

وتمثل مربعات اخطاء النموذج المرتبة بالاضافة الى ان اعلى حد مختار لمجموع الاخطاء يكون ضمن المتباينة التالية :-

$$\frac{n}{2} + 1 \leq h \leq \frac{3n+p+1}{4} \ldots\ldots\ldots\ldots\ldots 16$$

حيث p عدد المتغيرات المستقلة ، وان نقطة الانهيار $Bp = \frac{n-h}{n}$ ، يتصف تقدير LTS بكونه يمتلك نقطة انهيار عالية ويساوي 0.50 .

ولقياس كفاءة النموذج وفق طريقة التقدير يتم حساب معامل التحديد وفق العلاقة الاتية : (Chun Yu, (Weixin Yao, and Xue Bai , 2014

$$R_{LTS}^2 = 1 - \frac{s_{LTS}^2(X,y)}{s_{LTS}^2(1,y)} \qquad \ldots\ldots\ldots\ldots\ldots 17$$

٣-٦-٣:- مقدرات S (S-Estimation):

وصفت مقدرات s من قبل (Rousseeuw , yohai) عام ١٩٨٤ بالصيغة الاتية :

$$\hat{\theta}_S = \arg\min_{\theta} S(\theta) \qquad \text{.........................}18$$

حيث ان $S(\theta)$ تمثل التشتت وبدلالة المعالم وتوصف

$$\frac{1}{n-p}\sum_{i=1}^{n}\chi\left(\frac{y_i - x_i^T\theta}{S}\right) = \beta \qquad \text{.................................}19$$

ان نقطة القطع الى $S(\theta)$ تقدر بالعلاقة

$$\frac{\beta}{\max_s \chi(s)} \qquad \text{....................}20$$

حيث توجد عدة دوال لاجل اختيار χ منها دالة توكي الاتية

$$\chi_{k_0}(s) = \begin{cases} 3(\frac{s}{k_0})^2 - 3(\frac{s}{k_0})^4 + (\frac{s}{k_0})^6, & \text{if } |s| \le k_0 \\ 1 & \text{otherwise} \end{cases} \qquad \text{.................}21$$

حيث ان k_0 تمثل نقطة القطع وتبلغ قيمتها الاولية 2.9366 وتسيطرعلى طبيعة مقدرات S .

(Maronna, R. A.,et al,2006) (Habshah Midi,et al,2009)

3–٧:- بعض دوال الاوزان للمقدرات الحصينة

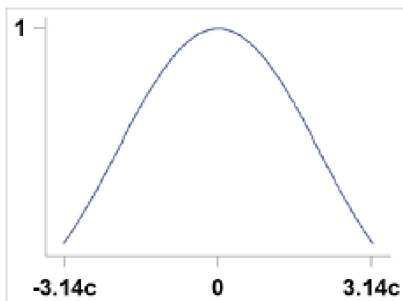
وهي مجموعة من الدوال التي تحدد في ضوءها الاوزان المرافقة الى المشاهدات ومنها :

دالة اندريو (Andrews function)

وتاخذ الصيغة التالية :

$$W(x,c)=\begin{cases}\frac{\sin(\frac{x}{c})}{\frac{x}{c}} & \text{if } |x|\le \pi c\\ 0 & \text{otherwise}\end{cases} \qquad \text{......................} 22$$

وشكلها البياني

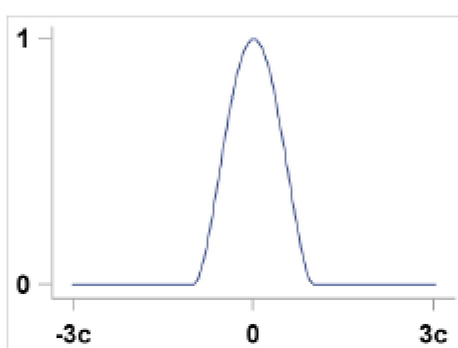


دالة (Bisquare function)

وتاخذ الصيغة التالية

$$W(x,c)=\begin{cases}(1-(\frac{x}{c})^2)^2 & \text{if } |x|< c\\ 0 & \text{otherwise}\end{cases} \qquad \text{......................} 23$$

وشكلها البياني

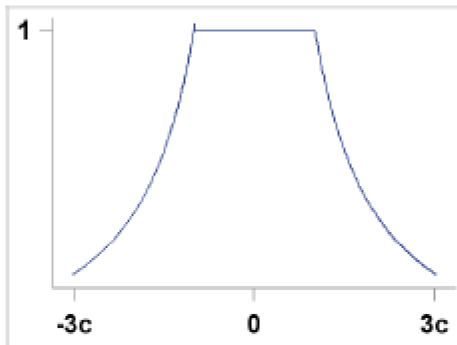


دالة (Huber function)

وتاخذ الصيغة التالية

$$W(x,c) = \begin{cases} 1 & \text{if } |x| < c \\ \frac{c}{|x|} & \text{otherwise} \end{cases} \qquad \text{..........................24}$$

وشكلها البياني



مع العلم انه توجد دوال اوزان اخرى تستخدم لاجل الحصول على المقدرات الحصينة .
(HUBER, P.J. 1981) (Habshah Midi,et al,2009)

4- الجانب التطبيقي:

تم الاعتماد على بيانات عدد الاقضية لمحافظات العراق كمتغير معتمد ومتغيرات مستقلة تمثلت بعدد السكان (نسمة) في كل محافظة ومساحة كل محافظة (كم ٢) حيث استخرجت النتائج من خلال برنامج Eviews 10، بالاعتماد على بيانات الجهاز المركزي للاحصاء – المجموعة الاحصائية السنوية لعام ٢٠١٧ والتي وصفت في جدول (١) التالي :

جدول (١) يبين عدد الاقضية والمساحة وعدد السكان في محافظات العراق لعام ٢٠١٧

عدد الاقضية	المساحة	السكان
10	37323	3633648
4	9679	1556618
6	17685	1594942
8	137808	1725914
10	4555	7916847
6	5119	2011706
3	5034	1187245
6	17153	1343125
9	24363	1554037
4	28824	1433583
4	8153	1257689
5	51740	793343
11	12900	2041066
6	15072	1083937
7	19070	2833375
7	6553	1805871
16	17023	1259150
9	15074	2106423

تم تقدير عدد الاقضية في العراق من علاقتها بالمساحة وعدد السكان باستخدام طريقة المربعات الصغرى ولوجود القيم الشاذة في متغيرات النموذج فان المعادلة المقدرة غير جيدة وغير معنوية بالاضافة الى عدم معنوية معالم النموذج وانخفاض معامل التحديد والذي بلغ ١١٪ .

Dependent Variable: Y
Method: Least Squares
Date: 12/13/19 Time: 21:45
Sample: 1 18
Included observations: 18

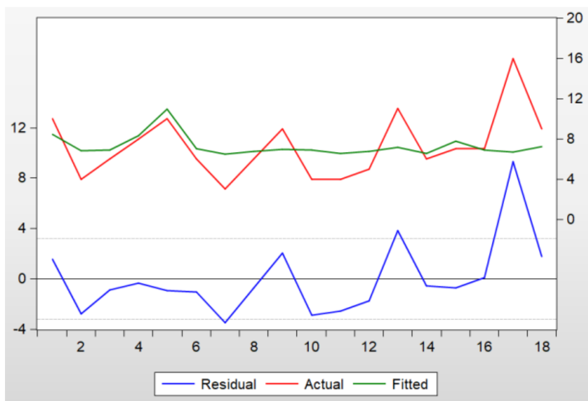
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.644459	1.455415	3.878248	0.0015
X1	1.11E-05	2.54E-05	0.438675	0.6671
X2	6.62E-07	4.88E-07	1.355153	0.1954

R-squared	0.112779	Mean dependent var	7.277778
Adjusted R-squared	-0.005517	S.D. dependent var	3.195687
S.E. of regression	3.204491	Akaike info criterion	5.317995
Sum squared resid	154.0314	Schwarz criterion	5.466390
Log likelihood	-44.86196	Hannan-Quinn criter.	5.338457
F-statistic	0.953362	Durbin-Watson stat	1.710860
Prob(F-statistic)	0.407604		

وبمؤشرات الكفاءة للقيم التنبؤيه التالي

Forecast: YF	
Actual: Y	
Forecast sample: 1 18	
Included observations: 18	
Root Mean Squared Error	2.925286
Mean Absolute Error	2.074032
Mean Abs. Percent Error	32.63284
Theil Inequality Coefficient	0.191635
Bias Proportion	0.000000
Variance Proportion	0.497201
Covariance Proportion	0.502799

وقد تم رسم القيم الاصلية والتنبوية مع الاخطاء للنموذج المقدر وكانت كالاتي



الشكل (١) القيم الاصلية والتنبؤية مع الاخطاء للنموذج المقدر بطريقة المربعات الصغرى

ثم بعد ذلك تم تقدير النموذج بالطرق الحصينة ومنها :

٤-١ مقدرات m الحصينة :

من خلال الطريقة تم استخدام دالة وزن هي (bisquare) ودالة هوبر من النوع الثاني للخطا المعياري والتغاير كانت نتائج تقدير النموذج مع مقاييس الكفاءة وفق الجدول التالي :

Dependent Variable: Y
Method: Robust Least Squares
Date: 12/13/19 Time: 22:11
Sample: 1 18
Included observations: 18
Method: M-estimation
M settings: weight=Bisquare, tuning=4.685, scale=MAD (median centered)
Huber Type I Standard Errors & Covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	4.384558	1.006299	4.357114	0.0000
X1	1.78E-05	1.75E-05	1.015607	0.3098
X2	8.60E-07	3.38E-07	2.547110	0.0109

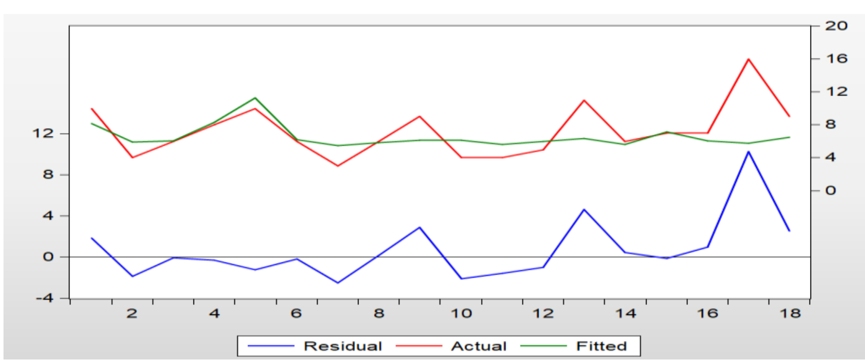
Robust Statistics

R-squared	0.264411	Adjusted R-squared	0.166332
Rw-squared	0.410226	Adjust Rw-squared	0.410226
Akaike info criterion	25.30654	Schwarz criterion	29.64920
Deviance	81.15086	Scale	1.966816
Rn-squared statistic	6.965267	Prob(Rn-squared stat.)	0.030726

Non-robust Statistics

Mean dependent var	7.277778	S.D. dependent var	3.195687
S.E. of regression	3.314176	Sum squared resid	164.7564

حيث تم ارتفاع معامل التحديد الى 26٪ والذي رافقه ارتفاع معامل التحديد الموزون بالمقارنة بالنموذج السابق ووصل الى 41٪ مع معنوية الحد الثابت ومتغير عدد السكان كعامل مؤثر في تحديد عدد الاقضية في العراق . وقد تم رسم القيم الاصلية والتنبؤية مع الاخطاء للنموذج المقدر وكانت كما في الشكل الاتي :



الشكل (٢) القيم الاصلية والتنبؤية مع الاخطاء للنموذج المقدر بطريقة مقدرات m الحصينة

٤- ٢مقدرات S الحصينة :

من خلال الطريقة تم تحديد نقطة انهيار 0.5 وبعدد محاولات للتقدير ٢٠٠ ودالة هوبر من النوع الثاني للخطا المعياري والتغاير كانت نتائج تقدير النموذج مع مقاييس الكفاءة وفق الجدول التالي :

Dependent Variable: Y
Method: Robust Least Squares
Date: 12/13/19 Time: 22:15
Sample: 1 18
Included observations: 18
Method: S-estimation
S settings: tuning=1.547645, breakdown=0.5, trials=200, subsmpl=3,
refine=2, compare=5
Random number generator: rng=kn, seed=953563277
Huber Type I Standard Errors & Covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	2.165918	1.100240	1.968586	0.0490
X1	1.99E-05	1.92E-05	1.038986	0.2988
X2	1.87E-06	3.69E-07	5.054468	0.0000

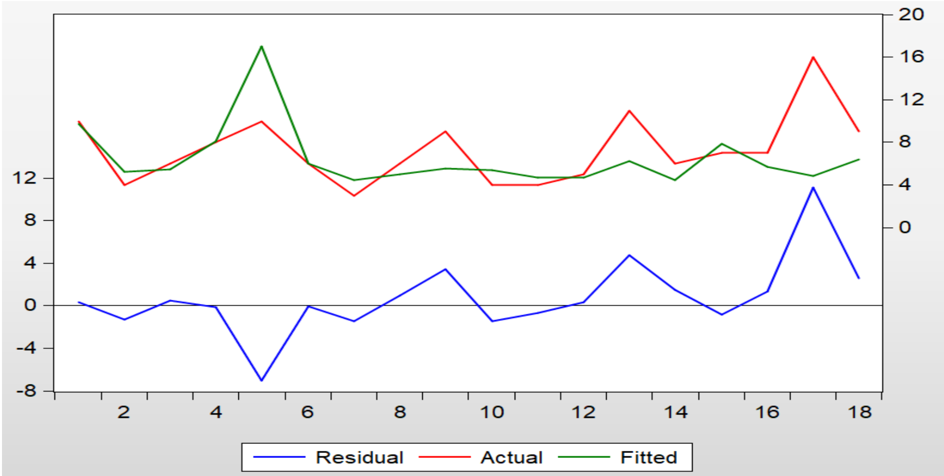
Robust Statistics

R-squared	0.338505	Adjusted R-squared	0.250306
Scale	2.285358	Deviance	5.222863
Rn-squared statistic	25.69706	Prob(Rn-squared stat.)	0.000003

Non-robust Statistics

Mean dependent var	7.277778	S.D. dependent var	3.195687
S.E. of regression	3.894357	Sum squared resid	227.4902

اذ بلغ معامل التحديد 33٪ والذي رافقه معنوية الحد الثابت ومعنوية عالية لمتغير عدد السكان كعامل مؤثر في تحديد عدد الاقضية في العراق . وقد تم رسم القيم الاصلية والتنبؤية مع الاخطاء للنموذج المقدر وكانت كما في الشكل الاتي :



الشكل (٣) القيم الاصلية والتنبؤية مع الاخطاء للنموذج المقدر بطريقة مقدرات s الحصينة

٤- ٣ مقدرات mm الحصينة:

من خلال الطريقة تم تحديد نقطة انهيار 0.5 وبعدد محاولات للتقدير ٢٠٠ تم استخدام دالة وزن هي (bisquare) ودالة هوبر من النوع الثاني للخطا المعياري والتغاير وكانت نتائج تقدير النموذج مع مقاييس الكفاءة وفق الجدول التالي :

Dependent Variable: Y
Method: Robust Least Squares
Date: 12/13/19 Time: 22:17
Sample: 1 18
Included observations: 18
Method: MM-estimation
S settings: tuning=1.547645, breakdown=0.5, trials=200, subsmpl=3,
refine=2, compare=5
M settings: weight=Bisquare, tuning=4.684
Random number generator: rng=kn, seed=953563277
Huber Type I Standard Errors & Covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	4.440582	1.027192	4.323028	0.0000
X1	1.75E-05	1.79E-05	0.975808	0.3292
X2	8.58E-07	3.45E-07	2.489040	0.0128

Robust Statistics

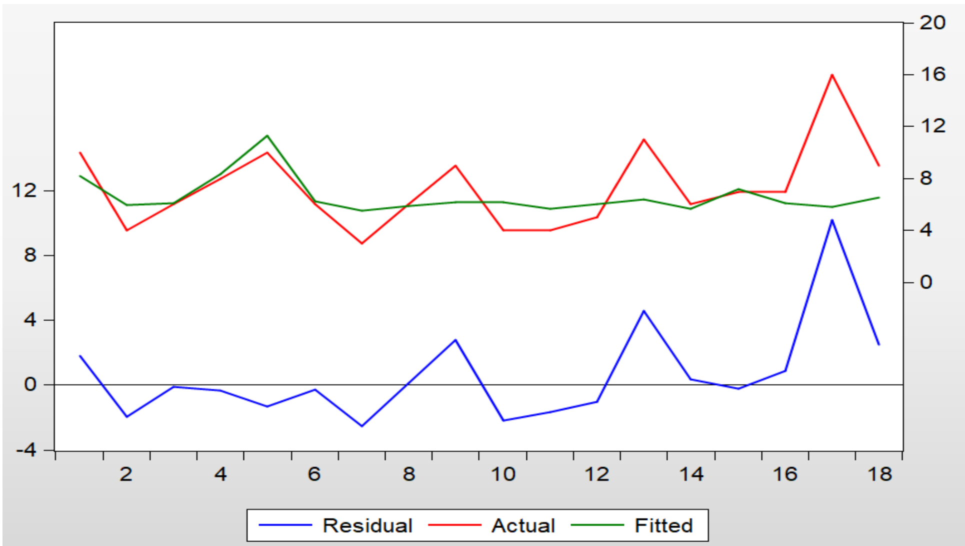
R-squared	0.241876	Adjusted R-squared	0.140793
Rw-squared	0.388842	Adjust Rw-squared	0.388842
Akaike info criterion	21.25235	Schwarz criterion	26.43498
Deviance	92.77827	Scale	2.285358
Rn-squared statistic	6.628907	Prob(Rn-squared stat.)	0.036354

Non-robust Statistics

Mean dependent var	7.277778	S.D. dependent var	3.195687
S.E. of regression	3.302749	Sum squared resid	163.6223

حيث تم ارتفاع معامل التحديد الى 24٪ والذي رافقه حساب معامل التحديد الموزون حيث وصل الى 38٪ مع معنوية الحد الثابت ومتغير عدد السكان كعامل مؤثر في تحديد عدد الاقضية في العراق .

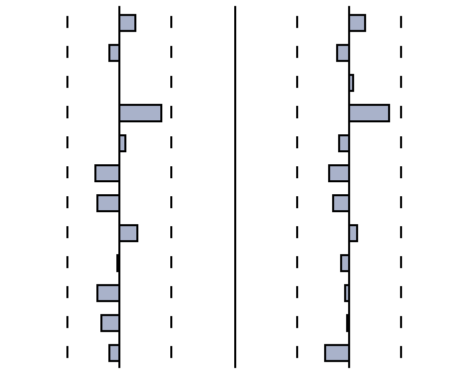
وقد تم رسم القيم الاصلية والتنبوية مع الاخطاء للنموذج المقدر وكانت كما في الشكل الاتي :



الشكل (٤) القيم الاصلية والتنبؤية مع الاخطاء للنموذج المقدر بطريقة مقدرات mm الحصينة

وبالمقارنة ما بين نتائج النماذج الموفقة الحصينة وغير الحصينة نجد ان افضل نموذج تم الحصول علية من خلال مقدرات S الحصينة لافضلية المقاييس الاحصائية .وقد تم فحص النموذج المختار من خلال حساب معاملات لارتباط الذاتي والارتباط الذاتي الجزئي لاخطاء النموذج ورسمها واظهرت وقوعها ضمن حدود الثقة وهي اشارة الى عدم وجود مشكلة ارتباط ذاتي في الاخطاء بالاضافة الى اختبارها عند كل ازاحة وكانت النتائج كما يلي :

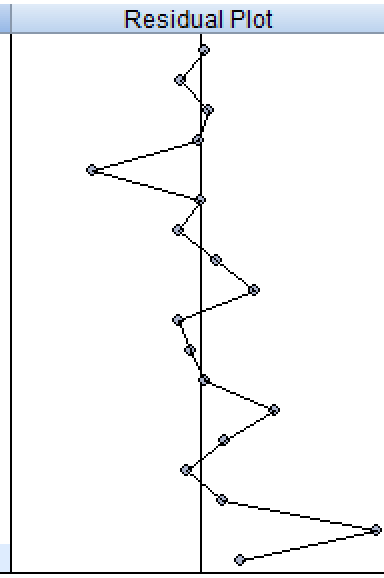
Date: 12/13/19 Time: 22:32
Sample: 1 18
Included observations: 18



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.154	0.154	0.4999	0.480
		2	-0.085	-0.111	0.6608	0.719
		3	0.008	0.041	0.6623	0.882
		4	0.370	0.366	4.1801	0.382
		5	0.044	-0.087	4.2338	0.516
		6	-0.221	-0.187	5.6951	0.458
		7	-0.196	-0.153	6.9498	0.434
		8	0.156	0.074	7.8249	0.451
		9	-0.013	-0.066	7.8316	0.551
		10	-0.195	-0.048	9.5471	0.481
		11	-0.163	-0.010	10.905	0.451
		12	-0.086	-0.228	11.345	0.500

وتم ايضا تقدير القيم الموفقة للنموذج الحصين المختار وكانت النتائج كما يلي :

obs	Actual	Fitted	Residual	Residual Plot
1	10.0000	9.69023	0.30977	
2	4.00000	5.26350	-1.26350	
3	6.00000	5.49458	0.50542	
4	8.00000	8.13316	-0.13316	
5	10.0000	17.0296	-7.02963	
6	6.00000	6.02181	-0.02181	
7	3.00000	4.48166	-1.48166	
8	6.00000	5.01408	0.98592	
9	9.00000	5.55135	3.44865	
10	4.00000	5.41549	-1.41549	
11	4.00000	4.67528	-0.67528	
12	5.00000	4.67754	0.32246	
13	11.0000	6.23168	4.76832	
14	6.00000	4.48896	1.51104	
15	7.00000	7.83311	-0.83311	
16	7.00000	5.66630	1.33370	
17	16.0000	4.85479	11.1452	
18	9.00000	6.39697	2.60303	



٥- الاستنتاجات والتوصيات

١- توجد علاقة انحدار خطي ضعيفة بين عدد الاقضية والمساحة وعدد السكان عند تقدير النموذج بالطرق غير الحصينة لوجود القيم الشاذة في المتغيرات ، وبذلك يشير النموذج الى عدم صلاحية طريقة المربعات الصغرى لتقدير مثل هذه البيانات .

٢- ان تقدير النموذج الخطي المتعدد باستخدام الطرق الحصينة قد حسن من كفاءة النموذج ومقدراته وان افضل طريقة حصينة للتقدير تكون من خلال استخدام مقدرات S الحصينة .

٣- نوصي باستخدام طرائق التقدير الحصينة لتقدير النماذج غير الخطية .

المصادر

١- الراوي ، خاشع محمود،(١٩٨٧) م" المدخل إلى تحليل الانحدار " مديرية دار الكتب للطباعة والنشر ، جامعة الموصل .

٢- Campell, N.A., Lopuhaa, H.P. and Rousseeuw, P.J., (19٩٨),"On the calculation of a robust S-Estimator of acovariance matrix", statistics in medicine,17. pp.2685-2695 .

٣- Chun Yu, Weixin Yao, and Xue Bai , 2014 , (Robust Linear Regression: A Review and Comparison) , Kansas State University, Manhattan, Kansas, USA 66506-0802 .

٤- Habshah Midi, Md. Sohel Rana, A. H. M. Rahmatullah Imon, (2009), “The Performance of Robust Weighted Least Squares in the Presence of

٥- Huber ,P.J.(1964)."Robust Estimation of location parameter .Ann.Math.Statist., 35,73-101 .

٦- HUBER, P.J. (1981) .Robust Statistics. Wiely, New York.

٧- Maronna, R. A., Martin, R. D. and Yohai, V. J. (2006), Robust Statistics. John Wiley.

٨- Montgomery, D. C. , Peck, E. A. and Vining, G.G . ,(2001) ,”Introduction to Linear Regression Analysis” , 3rd ed, Wiley, New York .

٩- Rousseeuw, P.J., and A. Leroy, 1987, *Robust Regression and Outlier Detection*, Wiley, New York, 1987.