Comparison of some robust methods in the presence of problems of multicollinearity and high leverage points

المقارنة بين بعض الطرائق الحصينة في ظل وجود مشكلتي تعدد العلاقة الخطية ونقاط الانعطاف العالية

ا.م. غفران اسماعيل كمال/ كلية الادارة والاقتصاد/ جامعة بغداد
ghufran62@gmail.com
الباحث / سيف الامام سعدي خزعل / كلية الادارة والاقتصاد/ جامعة بغداد saif.alemams@gmail.com

P - ISSN 2518 - 5764
E - ISSN 2227 - 703X

OPEN ACCESS
BY NC SA

Received:27/11/2018
Accepted: 26/12/2018

مستخلص البحث

يعد أنموذج الانحدار الخطي المتعدد من نماذج الانحدار المهمة والمستعملة في تحليل البيانات لمختلف مجالات العلم وعلى نطاق واسع مثل الاعمال والاقتصاد والطب والعلوم الاجتماعية، ان تعدد العلاقة الخطية مشكلة كبيرة في الانحدار الخطي المتعدد اذ تؤدي في ابسط حالتها الى ابتعاد معلمات الأنموذج المقدرة على خصائصها العلمية وغالباً ما تعطي استنتاجات مظللة، ايضاً هناك مشكلة هامة في تحليل الانحدار هو وجود نقاط الانعطاف العالية في البيانات مما تؤدي الى تأثيرات غير مرغوب بها على نتائج التحليل .

نستعرض في هذا البحث بعض الطرائق الحصينة في أنموذج الانحدار الخطي المتعدد ومن هذه الطرائق طريقتي انحدار الحرف لمقدر ال جاكنايف (Jackknife Ridge Regression) بالاعتماد على مقدر (MM-estimator) (MM) ومقدر (Modified Generalized M-estimator) (GM2)، ومن خلال استعمال المحاكاة بأسلوب مونت كارلو تمت اجراء المقارنة بين هاتين الطريقتين وفق معيار المقارنة متوسط مربعات الخطأ (MSE) ولحجوم عينات (n=20،n=50،n=100) ونسب تلوث مختلفة ($\tau = 5\% , 15\%$)، واتضح من خلال المقارنة ان طريقة (RJGM2) هي الافضل في تقدير معلمات أنموذج الانحدار الخطي المتعدد و يمتلك اقل قيمة لمتوسط مربعات خطأ (MSE) مقارنة مع بقية المقدرات الأخرى.

المصطلحات الرئيسة للبحث/ الانحدار الخطي المتعدد، تعدد العلاقة الخطية، نقاط الانعطاف العالية، مقدر MM ، مقدر GM2، انحدار الحرف للـ جاكنايف .

Journal of Economics and Administrative Sciences
2019; Vol. 25, No.114
Pages: 523- 538

*البحث مستل من رسالة ماجستير

1- المقدمة Introduction

يستعمل أنموذج الانحدار الخطي المتعدد(Multiple Linear Regression)(MLR) في العديد من مجالات الدراسة مثل الاعمال والاقتصاد والطب والعلوم الاجتماعية، فعند دراسة أي ظاهرة يجب تحديد المتغيرات المؤثرة في تلك الظاهرة وصياغة العلاقة بين تلك المتغيرات على هيئة أنموذج، ان العلاقة الخطية بين عدة متغيرات احدهما متغير معتمد (Dependent Variable) والباقي متغيرات توضيحية (Explanatory variables) يطلق عليها أنموذج الانحدار الخطي المتعدد (MLR)، أن مشكلة تعدد العلاقة الخطية أصبحت معروفة لدى العديد من الباحثين الإحصائيين و تمثل حالة انعدام الاستقلالية بين المتغيرات التوضيحية مما يسبب وجودها تأثيرات على تقديرات وتباينات المعاملات عند تطبيق طريقة المربعات الصغرى الاعتيادية (OLS) وبالتالي تظهر النتائج غير دقيقة، مشكلة مهمة أخرى في تحليل الانحدار هي وجود نقاط الانعطاف العالية (High Leverage Points) (HLPs) والتي تمثل (القياسات غير طبيعية في قيم المتغير x) والتي تسبب في ميل خط الانحدار نحوها بعيداً عن موقعة الحقيقي، وهي على نوعين الجيدة التي تتفق مع خط الانحدار وتساهم في كفاءة التقدير ومنها السيئة التي تؤثر على القيم المحسوبة لمختلف التقديرات . (Midi & Mohammed, 2015:P147) [1]

ان نقاط الانعطاف العالية (HLPs) لها تأثيرات كبيرة على الأنموذج الخطي منها التسبب في تضليل البيانات وفشل التحليل كونها تعتبر مصدر اخر لمصادر تعدد العلاقة الخطية واظهرت نتائج دراسة المحاكاة لكل من الباحثان (Imon & Kamruzzaman) عام (2002 م) ان نقاط الانعطاف العالية يمكن ان تولد تعدداً خطيا كبيراً وأن درجته قد تزداد مع زيادة نسبتها في البيانات .
(Kamruzzaman & Imon, 2002:P445) [3]

مما يستوجب استعمال طرائق بديلة عن الطرائق التقليدية لمعالجة تلك المشاكل تكون اكثر كفاءة في التقدير والتي تدعى (بالطرائق التقدير الحصينة) (Robust estimation methods) حيث تتصف انها قليلة الحساسية تجاه نقاط الانعطاف العالية (HLPs) اذ يتم الحصول عليها من خلال مقدرات حصينة تمتلك كفاءة عالية .

2- مشكلة البحث The problem of the research

أن أنموذج الانحدار الخطي المتعدد (MLR) يتضمن عدد من المتغيرات التوضيحية التي تكون ذات أداء ضعيف للمقدرات لوجود علاقة خطية بين المتغيرات التوضيحية مما يؤدي لعواقب سيئة مثل زيادة في تباينات المقدرات وانخفاض معامل الدقة وغالباً ما تكون النتائج مربكة وتعطي استنتاجات مظللة، بالإضافة الى ذلك وجود نقاط الانعطاف العالية (HLPs) في البيانات التي ستضاعف من المشكلة .

3- هدف البحث The Aim of the Research

الهدف من البحث هو الحصول على افضل تقدير لمعلمات أنموذج الانحدار الخطي المتعدد (MLR) عند وجود مشكلة تعدد العلاقة الخطية ونقاط الانعطاف العالية (HLPs) من خلال استعمال معيار المقارنة متوسط مربعات الخطأ (MSE) للحصول على أفضل مقدر بين المقدرات الاخرى .

4- الجانب النظري

يعطي تحليل الانحدار الاجابات على الاسئلة المتعلقة بالعلاقة الوظيفية للمتغير المعتمد (Y) مع واحد او اكثر من المتغيرات التوضيحية ($X's$)، يضاف مصطلح الخطأ العشوائي الى الأنموذج الاحصائي لحساب الفروق الفردية، أن الدالة الخطية بين Y و X تسمى أنموذج الانحدار الخطي المتعدد (MLR) ويمكن تعريفها على أنها :

$$Y_i = \beta_0+\beta_1 X_{1i}+\beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \qquad \ldots (1)$$

i =1، 2، …، n ، j =0,1، …، K

الأنموذج اعلاه يمكن كتابته بشكل مختصر وكالاتي :-

$$Y_i = \sum_{j=0}^{k} B_j\, X_{ij} + \varepsilon_i \qquad \dots (2)$$

وبدلالة المصفوفات يمكن صياغة أنموذج الخطي العام كما في الشكل الاتي :

$$\underline{y} = X\underline{B} + \underline{\varepsilon} \qquad \dots (3)$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1j} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2j} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{i1} & X_{i2} & \dots & X_{ij} & \cdots & X_{ik} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nj} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Y : موجه مشاهدات المتغير المعتمد من الدرجة $(n * 1)$.

X : مصفوفة من الدرجة $(n * (k + 1))$ وتمثل مشاهدات المتغيرات التوضيحية علماً بأن العمود الاول من هذه المصفوفة يمثل الحد الثابت .

β : موجه المعالم المطلوب تقديرها من الدرجة $(k + 1) * 1$.

k : عدد المتغيرات التوضيحية .

n : عدد المشاهدات .

ε : موجه الاخطاء العشوائية من الدرجة $(n * 1)$.

لتقدير معلمات أنموذج الانحدار الخطي المتعدد (MLR) يتم الاعتماد على بعض المقدرات الحصينة وهي : (كاظم ومسلم، 2002 : 50) [3]

5- بعض مقدرات الانحدار الحصينة Regression Estimates Some of Robust

5-1- مقدر MM MM- estimator

أن مقدر(MM) في التسمية يشير الى استعمال لأكثر من عملية تقدير لمقدر M، اقترح من الباحث (Yohai) عام (1987 م) أن هذا المقدر له العديد من الخصائص الجيدة منها أن لها كفاءة عالية في حالة التوزيع الطبيعي للأخطاء مع نقطة الانهيار العالية، تعد طريقة (MM) واحدة من أكثر المقدرات شيوعاً التي تستعمل في مجال الانحدار الحصين ويكون وفقاً للخطوات التالية :-

1- يتم تحديد مقدر أولي ذو نقطة انهيار عالية، ولكن ليس بالضرورة أن يكون كفوء، نرمز له بالرمز $(\hat{\beta}_s)$ وباستخدامه يتم حساب البواقي الاولية وفقاً للصيغة التالية :

$$r_i(\hat{\beta}_s) = y_i - x_i'\hat{\beta}_s \qquad , \quad 1 < i < n \qquad \dots (4)$$

2- يتم حساب مقدرM القياس (σ_n) للبواقي الاولية $r_i(\hat{\beta}_s)$ وفق معادلة M التقديرية لمعلمة القياس وبالشكل الاتي :

$$\frac{1}{n}\sum_{i=1}^{n} p\left(\frac{r_i(\hat{\beta}_s)}{\sigma}\right) = 0{,}5 \qquad \dots (5)$$

3- مقدر MM يعرف كمقدر M لـ β باستعمال دالة (re-descending score) :

$$\psi_1(u) = \frac{\partial\, p_1(u)}{\partial u} \qquad \dots (6)$$

عليه فأن مقدر MM والذي نرمز له بالرمز $\hat{\beta}_{MM}$ يكون الحل للمعادلة التالية :

$$\sum_{i=1}^{n} X_{ij}\psi_1\left(\frac{y_i - x_i'\beta}{\sigma_n}\right) = 0 \quad , \quad j = 1,2,\dots,k \qquad \dots (7)$$

أن تقدير القياس σ_n يتم أيجاده في الخطوة (2) .
اذ أن :

$$r_i(\beta) = y_i - x_i'\beta$$

[5] (Lukman, et.al, 2015:P58)

6- **مقدر** (GM2) GM2-estimator

اقترح من قبل الباحثة (Bagheri) عام (2011 م)، أن فكرة هذا المقدر تقوم على اساس التعديل لمقدر(Modified generalized M-estimator) (GM) للحصول على مقدر حصين ذو خصائص عالية الكفاءة وأكثر مقاومة لنقاط الانعطاف العالية (HLPs)، يتم تلخيص الطريقة من خلال عدة خطوات على النحو الاتي :

1- يتم حساب البواقي الاولية (r_i) من مقدر S، ومن ثم أيجاد مقياس للبواقي ($\hat{\tau}$) وكالاتي :

$$r_i = y_i - \hat{y}_i \quad ، \quad i=1,2\dots,n \qquad \dots (8)$$

$$\hat{\tau} = 1.4826(1 + 5/(n-p))\ \text{Median}\ |r_i| \qquad \dots (9)$$

2 - يتم أيجاد مصفوفة الاوزان القطرية (W) ، وعناصر القطر هي الاوزان (W_i) من خلال المعادلة الاتية :

$$w_i = \min[1, \{\frac{x^2_{(0.95,\ p+1)}}{RMD^2}\}] \quad ، \quad i=1,2\dots,n \qquad \dots (10)$$

RMD : تمثل مسافة (مهن لوبس) (Mahalanobis Distance) الحصينة بالاعتماد على المقدار الأدنى للمقياس الإهليجي (MVE) .

3- حساب دالة التأثير Ψ^* للبواق المعيارية، وذلك بحل المعادلة الاتية :

$$A = diag\ \Psi^*\left(\frac{r_i}{\hat{\tau} * w_i}\right) \qquad \dots (11)$$

Ψ^* : مشتقة من دالة التأثير لـ (Huber's) .

4- عليه فأن مقدر (GM2) والذي نرمز له بالرمز ($\hat{\beta}_{GM2}$) يمكننا الحصول عليه عن طريق اشتقاق خطوة واحدة لطريقة (Newton Raphson) ويكون بالصيغة الاتية :

$$\tilde{\alpha}_{GM2} = \hat{\beta}_0 + (X'A\,X)^{-1}X'W\psi\left(\frac{r_i}{w_i\hat{\tau}}\right)\hat{\tau} \qquad \dots (12)$$

[1] (Alguraibawi, et.al, 2015:P311)

7- **طرائق التقدير**:
لتقدير معلمات أنموذج الانحدار الخطي المتعدد سيتم الاعتماد في هذا البحث على بعض الطرائق الحصينة :

7-1- **طريقة** (RJRR) :
A-Robust Jackknife Ridge Regression based on MM- estimator(RJRMM)
B -Robust Jackknife Ridge Regression based on GM2-estimator(RJRGM2)
نظرياً اقترحت طريقة ال جاكنايف (Jackknife method) من الباحث (Quenouille) عام (1949 م) لتقليل التحيز والتباين في التقدير, وان المبدأ الاساسي لهذه الطريقة هو استبعاد جزء من البيانات عند أجراء كل عملية تقدير بحيث أن الجزء المستبعد اما يهمل او يؤخذ بنظر العناية عند التقدير من خلال احتسابه، في عام (1986 م) أقترح (Singh) وأخرون أسلوبا للتحايل على التحيز في انحدار الحرف بالاعتماد على تقنية ال جاكنايف من خلال الصيغة الاتية :

$$y_{-i} = X_{-i}\beta + \varepsilon^* \qquad \dots (13)$$

من خلال أنموذج الانحدار الخطي أعلاه بصيغة ال جاكنايف يتم استبعاد قيمة من المتجه y_{-i} يقابلها استبعاد صف بالكامل من المصفوفة X_{-i} حيث ليس من الضروري ان تكون المصفوفة X_{-i} كاملة الرتبة للأعمدة، أما بالنسبة لــ β تمثل متجه المعلمات لانموذج الانحدار .

ε^* يمثل حد الخطأ العشوائي حيث ان :

$$E(\varepsilon^*) = 0$$
$$\mathrm{Cov}(\varepsilon^*) = \sigma^2 I_{n-1}$$

هنا يكون الأنموذج المختزل لانحدار الحرف الاعتيادي (RR) لـ (β) بالصيغة الاتية :

$$\hat{\beta}_{RR(-i)} = (X'_{-i}X_{-i} + kI_p)^{-1}X'_{-i}y_{-i} \qquad \dots (14)$$

X_{-i} : الجزء او الصف i المستبعد من المصفوفة X .
y_{-i} : الجزء او القيمة i المستبعدة من المتجه y .
بإمكاننا أعادة كتابة المعادلة (14) للأنموذج المختزل لانحدار الحرف الاعتيادي (RR) بالشكل القانوني (Canonical form) لـ (α) بالشكل الاتي :

$$\hat{\alpha}_{RR(-i)} = (Z'_{-i}Z_{-i} + kI_p)^{-1}Z'_{-i}Y_{-i} \qquad \dots (15)$$

نأخذ كل من z_i ، y_i لمتجه الاعمدة Z و Y بالتنسيق، تكون الصيغة بالشكل الاتي :

$$\hat{\alpha}_{RR(-i)} = (Z'Z - z_i'z_i + kI)^{-1}(Z'y - z_i y_i) \qquad \dots (16)$$

[1] **(Alguraibawi, et.al, 2015:PP308-309)**

ومن خلال تطبيق نظرية (Binomial Inverse Theorem) نحصل على :

$$=\left[B^{-1}+\frac{B^{-1}z_i'z_iB^{-1}}{1-z_i'B^{-1}z_i}\right]\times(Z'y-z_iy_i)$$

$$=(B^{-1}Z'y-B^{-1}z_iy_i+\frac{B^{-1}z_i'z_iB^{-1}}{1-z_i'B^{-1}z_i}Z'y-\frac{B^{-1}z_i'z_iB^{-1}}{1-z_i'B^{-1}z_i}z_iy_i$$

$$=\hat{\alpha}_{RR}-B^{-1}z_iy_i\left(1+\frac{z_i'B^{-1}z_i}{1-z_i'B^{-1}z_i}\right)+\frac{B^{-1}z_i'z_i\hat{\alpha}_{RR}}{1-z_i'B^{-1}z_i}$$

$$=\hat{\alpha}_{RR}-\frac{B^{-1}z_i\left(Y_i-z_i'\hat{\alpha}\right)}{1-z_i'B^{-1}z_i}$$

$$\hat{\alpha}_{RR(-i)}=\hat{\alpha}_{RR}-\frac{B^{-1}z_ie_i}{1-h_i} \qquad \dots(17)$$

[2] (Duran & Akdeniz, 2010:P268)

أذ أن

$$h_i=z_i'(Z'Z)^{-1}z_i \qquad ، e_i=(Y_i-z_i'\hat{\alpha})$$

e_i : موجة البواقي .

عندها مقدر ال جاكنايف يكون وفق الشكل الأتي :

$$p_i=n\hat{\alpha}_{RR}-(n-1)\hat{\alpha}_{RR(-i)} \qquad \dots(18)$$

المعادلة أعلاه تسمى بالقيم غير الأكيدة (الوهمية) (Pseudo – values) لتقدير ال جاكنايف ، من المعادلة (17) و (18) يكون الوسط الحسابي لهذه التقديرات بالصيغة الاتية :

$$\bar{P}=\frac{1}{n}\sum_{i=1}^{n}P_i=\hat{\alpha}_{RR}+\frac{(n-1)}{n}B^{-1}\sum_{i=1}^{n}\frac{z_ie_i}{(1-h_i)} \qquad \dots(19)$$

ان القيم غير الأكيدة في المعادلة (18) تعرف انها متماثلة فيما يتعلق بالمشاهدات ، في حين ان الأنموذج غير متوازن بشكل عام وينعكس نقص التوازن في المسافات (h_i)، بالإضافة الى ان التباين ($\hat{\alpha}_{RR(-i)}-\hat{\alpha}_{RR}$) هو دالة متزايدة لـ h_i، يمكننا تعريف القيم غير الأكيدة الموزونة كما في الصيغة الاتية :

$$Q_i = \hat{\alpha}_{RR} + n(1-h_i)\left(\hat{\alpha}_{RR} - \hat{\alpha}_{RR(-i)}\right) \qquad \dots(20)$$

هنا تكون تقديرات **(JRR)(Jackknife Ridge Regression)** متماثلة، اي مقدر $\hat{\alpha}_{JRR}(k)$ يعطى وفق بالصيغة الاتية :

$$\hat{\alpha}_{JRR}(k) = \bar{Q} = \frac{1}{n}\sum Q_i = \hat{\alpha}_{RR} + B^{-1}\sum z_i e_i \qquad \dots(21)$$

[7] **(Singh, et.al, 1986:344)**

يمكن تبسيط الأنموذج (JRR) بالصيغة الاتية :

$$\hat{\alpha}_{JRR}(k) = (I + kB^{-1})\hat{\alpha}_{JRR}(k) \qquad \dots(22)$$
$$= (I - k^2B^{-2})\hat{\alpha}$$

علما ان معاملات الانحدار الاصلية لـ (JRR) هي :

$$\hat{\beta}_{JRR} = \gamma\hat{\alpha}_{JRR}(k)$$

وان مقدر (JRR) متحيز بالنسبة للمعلمة α ومقدار التحيز هو :

$$Bias\,(\hat{\alpha}_{JRR}(k)) = -k^2B^{-2}\,\hat{\alpha} \qquad \dots(23)$$

ومصفوفة التباين لمقدر (JRR) كالاتي :

$$\mathrm{Var}\,(\hat{\alpha}_{JRR}(k)) = \sigma^2(I - k^2B^{-2})\,\Lambda^{-1}(I - k^2B^{-2})' \qquad \dots(24)$$

ومصفوفة متوسط مربعات الخطأ لمقدر (JRR) بالشكل الأتي :

$$MSE\left(\hat{\alpha}_{JRR}(k)\right) = Var(\hat{\alpha}_{JRR}(k)) + [Bias(\hat{\alpha}_{JRR}(k))][Bias(\hat{\alpha}_{JRR}(k))]'$$
$$= \sigma^2(I - k^2B^{-1})\,\Lambda^{-1}(I - k^2B^{-1})' + k^4B^{-2}\hat{\alpha}\,\hat{\alpha}'B^{-2} \qquad \dots(25)$$

ان الصيغة العامة لمقدر (RJRR) (Robust Jackknife Ridge Regression) لــ (α) يكون كالاتي :

$$\hat{\alpha}_{RJRR}(k) = [I + kB]\tilde{\alpha}_{RRR} \qquad \dots(26)$$
$$= [I + kB^{-1}][I - kB^{-1}]\tilde{\alpha}$$
$$= (I - k^2B^{-2})\tilde{\alpha}$$

و ان :

$$k = \frac{P\tilde{\sigma}^2}{\tilde{B}'\,\tilde{B}} \qquad \dots(27)$$

وان :

$$\tilde{\sigma}^2 = \frac{(Y - X\tilde{B})'(Y - X\tilde{B})}{\tilde{B}'\,\tilde{B}} \qquad \dots(28)$$

علما إن معاملات الانحدار الاصلية لمقدر (RJRR) لـــ (α) تعطى بالشكل الأتي :

$$\hat{\beta}_{RJRR} = \gamma\hat{\alpha}_{RJRR}(k)$$

أن مقدر (RJRR) متحيز للمعلمة الأصلية، ومقدار التحيز هو :

$$\begin{aligned} Bias\,(\hat{\alpha}_{RJRR}(k)) &= E[\hat{\alpha}_{RJRR}(k)] - \alpha \\ &= E[(\mathrm{I} - k^2B^{-2})\tilde{\alpha}] - \alpha \\ &= (\mathrm{I} - k^2B^{-2})E[\tilde{\alpha}] - \alpha \\ &= (\mathrm{I} - k^2B^{-2})\alpha - \alpha \\ &= -k^2B^{-2}\,\alpha \qquad \dots (29) \end{aligned}$$

والتباين لمقدر (RJRR) يعطى بالشكل الاتي :

$$\begin{aligned} \mathrm{Var}(\hat{\alpha}_{RJRR}(k)) &= E[\hat{\alpha}_{RJRR}(k) - E(\hat{\alpha}_{RJRR}(k))][\hat{\alpha}_{RJRR}(k) - E(\hat{\alpha}_{RJRR}(k))]' \\ &= E[\hat{\alpha}_{RJRR}(k) - \alpha\,][(\hat{\alpha}_{RJRR}(k) - \alpha]' \qquad \dots (30) \\ &= (\mathrm{I} - k^2B^{-2})\,\Omega\,(\mathrm{I} - k^2B^{-2})' \end{aligned}$$

ومتوسط مربعات الخطأ (MSE) لمقدر (RJRR) يكون بالشكل الأتي :

$$MSE(\hat{\alpha}_{RJRR}(k)) = \mathrm{Var}(\hat{\alpha}_{RJRR}(k)) + [Bias(\hat{\alpha}_{RJRR}(k))][Bias(\hat{\alpha}_{RJRR}(k))]'$$

$$= (\mathrm{I} - k^2B^{-2})\,\Omega\,(\mathrm{I} - k^2B^{-2})' + k^4B^{-2}\,\alpha\alpha' B^{-2} \qquad \dots (31)$$

مـن الصـيغة (26) نعـوض المعلمــة ($\tilde{\alpha}$) بالمقـدر (MM) الحصـين حيـث تكـون الصـيغة النهائيــة لمقـدر (RJMM) كالاتي :

$$\begin{aligned} \hat{\alpha}_{\mathrm{RJMM}}(k) &= [\mathrm{I} + k\mathrm{B}]\tilde{\alpha}_{MM} \\ &= [\mathrm{I} + k\mathrm{B}^{-1}][\mathrm{I} - k\mathrm{B}^{-1}]\tilde{\alpha}_{MM} \\ &= (\mathrm{I} - k^2\mathrm{B}^{-2})\tilde{\alpha}_{MM} \qquad \dots (32) \end{aligned}$$

وايضا من للمعادلة (26) نعوض المعلمة ($\tilde{\alpha}$) بالمقدر الحصين (GM2) و تكون الصيغة النهائية للمقدر (RJGM2) كالاتي :

$$\begin{aligned} \hat{\alpha}_{\mathrm{RJGM2}}(k) &= [\mathrm{I} + k\mathrm{B}]\tilde{\alpha}_{GM2} \\ &= [\mathrm{I} + k\mathrm{B}^{-1}][\mathrm{I} - k\mathrm{B}^{-1}]\tilde{\alpha}_{GM2} \\ &= (\mathrm{I} - k^2\mathrm{B}^{-2})\tilde{\alpha}_{GM2} \qquad \dots (33) \end{aligned}$$

[1] **(Alguraibawi, et.al, 2015:PP309-313)**

8- الجانب التجريبي

1-8- مفهوم المحاكاة The Concept of Simulation

يعد اسلوب المحاكاة من الاساليب العلمية الرصينة باعتباره اسلوبًا للاختبار يقوم على اساس اعطاء صورة طبق الاصل لظاهرة حقيقة قبل تطبيق التجربة على بيانات واقعية، لغرض الاستفادة من هذه الصورة في دراسة خواص تلك الظاهرة ومميزاتها، حيث يعتمد اسلوب المحاكاة على اثبات البرهان الرياضي نظرياً . [1] (الجشعمي، 2007 :34)

ان اسلوب المحاكاة هو عملية تمثيل سلوك الظاهرة الحقيقية قيد الدراسة بشكل تمثل الانموذج المدروس للواقع الحقيقي ، و يمكن ان يُعتمد أسلوب المحاكاة في اثبات صحة طريقة معينة يكون من الصعب اثبات صحتها نظرياً، وكلما كانت النتائج دقيقة دل ذلك على ان اسلوب المحاكاة يكون اكثر قرباً للواقع المدروس. [2] (شهاب، 2017 :43)

8-2- مراحل تطبيق تجارب المحاكاة

Stages of the application of simulation experiments

لقد تضمنت تجارب المحاكاة لهذا البحث كتابة عدد من البرامج بلغة (R) في توليد البيانات، واقترحت ثلاث حجوم للعينة (n=20،n=50،n=100) ونسب تلوث مختلفة ($\tau = 5\%, 15\%$)، وقد تم استعمال الصيغة الاتية في توليد المتغيرات التوضيحية :-

$$X_{ij} = p\, v_{ik} + (1-p^2)^{1/2} v_{ij} \quad , \qquad i = 1,2,\dots,n \ , \qquad j = 1,2,\dots,k$$

v_{ij} : الإعداد العشوائية المولدة والتي تتبع التوزيع الطبيعي القياسي .

v_{ik} : يمثل قيم العمود الاخير من اعمدة المتغيرات المولدة .

k : يمثل عدد المتغيرات المرتبطة .

i : يمثل عدد المشاهدات .

ρ : يمثل قيمة الارتباط بين المتغيرات التوضيحية في الأنموذج المدروس، الذي يأخذ القيم ($\rho = 0.90, 0.95, 0.99$) .

(Alguraibawi, et.al, 2015:P313) [1]

ويكون الأنموذج كالاتي :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i \qquad ..(34)$$

$$i = 1,2,..,n$$

يتم توليد الاخطاء العشوائية وفقاً للتوزيع الطبيعي بمتوسط (صفر) وتباين (σ^2) اي ان :

$$\varepsilon_i \sim (0, \sigma^2) \quad , \quad i = 1,2,\dots,n$$

وبالنسبة لقيم الانحراف المعياري فهي ($\sigma = 0.5, 1, 1.5$) .

يتم تحديد القيم الافتراضية للمعلمات من خلال دراسات سابقة وبافتراض أن عدد المعالم هي $k = 6$ وفقاً للقيم التالية :-

$$\beta_1 = 5\ ,\ \beta_2 = 3\ ,\ \beta_3 = \sqrt{6}\ ,\ \beta_4 = 0\ ,\ \beta_5 = 0\ ,\ \beta_0 = 0$$

(Kan, et.al, 2013:P648) [4]

ثم يتم تقدير معلمات انموذج الانحدار الخطي المتعدد (MLR) وفق طرائق التقدير التي تم عرضها في الجانب النظري من البحث وهي كما يلي :-

1- طريقة انحدار الحرف (RR) لمقدر ال جاكنايف بالاعتماد على مقدر (MM) (RJMM) .

2- طريقة انحدار الحرف لمقدر ال جاكنايف بالاعتماد على مقدر(GM2) (RJGM2) .

وقد تم الاعتماد على المقياس الاحصائي متوسط مربعات الخطأ(MSE) للأنموذج لمعرفة اي النماذج افضل في تمثيل البيانات وكان عدد مرات تكرار التجربة (R=1000) مرة .

8—3 تحليل نتائج تجربة المحاكاة

Analyzing the results of the simulation experiment

جدول (1) تقدير المعلمات و(MSE) للطريقتين ، ولكافة حجوم العينـات عندما تكون نسبة التلوث (τ=5%) وقيمة الانحراف المعياري ($\sigma = 1$) .

Coffe	N	Parameters / Methods	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	MSE
$\rho = 0.90$	n=20	RJMM	0.213876	0.20999	0.20863	0.10212	0.12416	0.014047
		RJGM2	0.31949	0.27230	0.24319	0.06064	0.06622	0.01249
	n=50	RJMM	0.25065	0.17109	0.05731	0.10252	0.09237	0.00957
		RJGM2	0.48167	0.29385	-0.0299	0.00786	0.02436	0.00753
	n=100	RJMM	0.12598	0.11411	0.11846	0.10658	0.10446	0.00582
		RJGM2	0.21053	0.15107	0.16763	0.09859	0.09065	0.00551
$\rho = 0.95$	n=20	RJMM	0.19441	0.21181	0.21690	0.11222	0.13002	0.01186
		RJGM2	0.29125	0.26430	0.24253	0.07102	0.07328	0.01119
	n=50	RJMM	0.24999	0.17245	0.06736	0.11688	0.10058	0.00883
		RJGM2	0.50334	0.30061	-0.0621	0.01252	0.02111	0.00728
	n=100	RJMM	0.12718	0.11955	0.12334	0.11595	0.11441	0.00541
		RJGM2	0.19522	0.14513	0.16643	0.10926	0.10178	0.00522
$\rho = 0.99$	n=20	RJMM	0.16647	0.23724	0.25986	0.10714	0.12521	0.01026
		RJGM2	0.289963	0.26008	0.26526	0.03962	0.07908	0.01018
	n=50	RJMM	0.28012	0.18168	0.05966	0.12924	0.09955	0.00810
		RJGM2	0.66635	0.37027	-0.2073	-0.02308	-0.02941	0.00711
	n=100	RJMM	0.138178	0.13343	0.13719	0.13260	0.13108	0.00499
		RJGM2	0.18417	0.13747	0.17326	0.11980	0.10954	0.00496

جدول(2) تقدير المعلمات و(MSE) للطريقتين، ولكافة حجوم العينات عندما تكون نسبة التلوث (τ=5%) وقيمة الانحراف المعياري ($\sigma = 1.5$) .

Coffe	N	Parameters / Methods	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\beta}_3$	$\tilde{\beta}_4$	$\tilde{\beta}_5$	MSE
$\rho = 0.90$	n=20	RJMM	0.18473	0.19654	0.20110	0.07738	0.10546	0.01520
		RJGM2	0.28340	0.25953	0.23804	0.04137	0.05015	0.01266
	n=50	RJMM	0.12111	0.10701	0.10109	0.09288	0.09013	0.01596
		RJGM2	0.22294	0.15039	0.12809	0.09257	0.08784	0.01593
	n=100	RJMM	0.08740	0.07964	0.08395	0.07675	0.07484	0.00753
		RJGM2	0.15491	0.10867	0.13112	0.07943	0.07174	0.00711
$\rho = 0.95$	n=20	RJMM	0.17208	0.20816	0.22058	0.08088	0.10686	0.01611
		RJGM2	0.26637	0.26073	0.24855	0.04025	0.04742	0.01383
	n=50	RJMM	0.12702	0.11308	0.10811	0.10324	0.10132	0.01221
		RJGM2	0.22407	0.14417	0.12259	0.09836	0.09658	0.01156
	n=100	RJMM	0.09509	0.08913	0.09298	0.08756	0.08596	0.00673
		RJGM2	0.15100	0.10812	0.13389	0.08668	0.07878	0.00642
$\rho = 0.99$	n=20	RJMM	0.14250	0.25549	0.29369	0.05336	0.08750	0.01938
		RJGM2	0.28771	0.27179	0.30409	- 0.03643	0.03812	0.01960
	n=50	RJMM	0.30866	0.16922	-0.0168	0.08828	0.04639	0.01259
		RJGM2	0.79406	0.42444	-0.3773	- 0.11089	- 0.11054	0.01111
	n=100	RJMM	0.10998	0.10489	0.10967	0.10491	0.10274	0.00695
		RJGM2	0.15097	0.10279	0.14727	0.09297	0.07936	0.00693

جـدول (3) تقـدير المعلمـات و(MSE) للطـريقتين ، ولكافـة حجـوم العينـات عنـدما تكـون نسـبة التلـوث (τ=15%) وقيمة الانحراف المعياري ($\sigma = 1$) .

Coffe	N	Parameters / Methods	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	MSE
$\rho = 0.90$	n=20	RJMM	0.21387	0.20999	0.20863	0.10212	0.12416	0.01404
		RJGM2	0.31949	0.27230	0.24319	0.06064	0.06622	0.01249
	n=50	RJMM	0.25065	0.17109	0.05731	0.10252	0.09237	0.00957
		RJGM2	0.48167	0.29385	-0.0299	0.00786	0.02436	0.00753
	n=100	RJMM	0.12598	0.11411	0.11846	0.10658	0.10446	0.00582
		RJGM2	0.21053	0.15107	0.16763	0.09859	0.09065	0.00551
$\rho = 0.95$	n=20	RJMM	0.19441	0.21181	0.21690	0.11222	0.13002	0.01186
		RJGM2	0.29125	0.26430	0.24253	0.07102	0.07328	0.01119
	n=50	RJMM	0.24999	0.17245	0.06736	0.11688	0.10058	0.00883
		RJGM2	0.50334	0.30061	-0.0621	0.01252	0.02111	0.00728
	n=100	RJMM	0.12718	0.11955	0.12334	0.11595	0.11441	0.00541
		RJGM2	0.19522	0.14513	0.16643	0.10926	0.10178	0.00522
$\rho = 0.99$	n=20	RJMM	0.16647	0.23724	0.25987	0.10714	0.12521	0.01026
		RJGM2	0.28996	0.26008	0.26526	0.03962	0.07908	0.01018
	n=50	RJMM	0.28012	0.18168	0.05966	0.12924	0.09955	0.00810
		RJGM2	0.66635	0.37027	-0.2073	-0.02308	-0.02941	0.00711
	n=100	RJMM	0.13817	0.13343	0.13719	0.13260	0.13108	0.00499
		RJGM2	0.18417	0.13747	0.17326	0.11980	0.10954	0.00496

جـدول (4) تقـدير المعلمــات و(MSE) للطــريقتين ، ولكافــة حجــوم العينــات عنــدما تكــون نســبة التلــوث (τ=15%) وقيمة الانحراف المعياري ($\sigma = 1.5$) .

Coffe	N	Parameters / Methods	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	MSE
$\rho = 0.90$	n=20	RJMM	0.18504	0.19528	0.20032	0.07817	0.10580	0.02321
		RJGM2	0.28380	0.25765	0.23790	0.04353	0.04924	0.02215
	n=50	RJMM	0.21261	0.13888	0.02280	0.07539	0.06197	0.01385
		RJGM2	0.45060	0.26984	-0.0915	-0.01845	-0.00559	0.01145
	n=100	RJMM	0.08733	0.07976	0.08373	0.07653	0.07480	0.00776
		RJGM2	0.15466	0.10977	0.12987	0.12987	0.07952	0.00747
$\rho = 0.95$	n=20	RJMM	0.17286	0.20711	0.21896	0.08197	0.10722	0.02049
		RJGM2	0.26706	0.25900	0.24766	0.04342	0.04619	0.02019
	n=50	RJMM	0.23127	0.14732	0.02410	0.08570	0.06555	0.013187
		RJGM2	0.50844	0.29377	-0.1412	-0.02800	-0.01964	0.01123
	n=100	RJMM	0.09504	0.08914	0.09292	0.08743	0.08588	0.00735
		RJGM2	0.15087	0.10857	0.13355	0.08652	0.07867	0.00719
$\rho = 0.99$	n=20	RJMM	0.14458	0.25377	0.28908	0.05508	0.08939	0.01822
		RJGM2	0.29000	0.27218	0.29967	-0.03164	0.03520	0.01854
	n=50	RJMM	0.30642	0.16823	-0.0143	0.08929	0.04773	0.01259
		RJGM2	0.78643	0.41721	-0.3690	-0.10590	-0.10728	0.01113
	n=100	RJMM	0.10998	0.10489	0.10967	0.10491	0.10274	0.00695
		RJGM2	0.15097	0.10279	0.14727	0.09297	0.07936	0.00693

9- **مناقشة تجارب المحاكاة في حالة وجود تلوث** ($\tau = 5\%$ ،15%)

من خلال النتائج المبينة بالجداول (1) (2) (3) (4) نلاحظ ما يلي :-

1- يتضح لنا ان قيمة متوسط مربعات الخطأ (MSE) للأنموذج تتناقص كلما زاد حجم العينة ولجميع طرائق التقدير المدروسة .

2- في حالة في حالة وجود تلوث (τ =5%) وتوزيع الاخطاء التوزيع الطبيعي بمتوسط (0) وانحراف معياري ($\sigma = 1$) نلاحظ من خلال مقياس المقارنة (MSE) للجدول (1) ان طريقة (RJGM2) افضل طريقة عند احجام العينات (n=50،n=100، n=20) وعندما تكون قيمة معامل الارتباط ($p = 0.90, 0.95, 0.99$) وذلك لا نها حققت اقل قيمة ل (MSE) للأنموذج.

3- وايضاً عند وجود تلوث(τ =5%) وتوزع الاخطاء التوزيع الطبيعي بمتوسط (0) وانحراف معياري ($\sigma = 1.5$) للجدول (2) ومن خلال المقياس (MSE) للأنموذج نلاحظ ان طريقة (RJGM2) هي افضل طريقة عند احجام العينات (n=100، n=50، n=20) عندما تكون قيمة معامل الارتباط ($p = 0.90$,0.95)، ايضاً حققت طريقة (RJGM2) الأفضلية من خلال معيار المقارنة (MSE) للأنموذج وعند احجام العينات (n=50،n=100) عندما تكون قيمة معامل الارتباط ($p = 0.99$)، بينما حققت طريقة (RJMM) اقل قيمة ل (MSE) عند حجم العينة (n=20) .

4- وفي حالة وجود التلوث (τ =15%) وتوزيع الاخطاء بالتوزيع الطبيعي بمتوسط (0) وانحراف معياري ($\sigma = 1$) نلاحظ من خلال مقياس المقارنة (MSE) للجدول (3) ان طريقة (RJGM2) افضل طريقة عند احجام العينات (n=100،n=20 ,n=50) عندما تكون قيمة معامل الارتباط ($p = 0.99$,$0.95, 0.90$) وذلك لا نها حققت اقل قيمة ل (MSE) للأنموذج، ايضاً حققت طريقة (RJGM2) للجدول (4) الأفضلية لا قل قيمة لمعيار المقارنة(MSE) عند احجام العينات (n=100،n=50، n=20) عند قيمة معامل الارتباط ($p = 0.90$,0.95)، وايضاً الافضلية عندما تكون قيمة معامل الارتباط ($p = 0.99$) لا حجام العينات (n=50،n=100) ، بينما حققت طريقة (RJMM) الافضلية عند حجم العينة (n=20) كونها حققت اقل قيمة لمتوسط مربعات الخطأ (MSE) .

10-**الاستنتاجات والتوصيات**

10-1- **الاستنتاجات**

1- اظهرت نتائج المحاكاة لهذا البحث ان طريقة (RJGM2) هي الافضل، و تمتلك اقل قيمة لمتوسط مربعات الخطأ (MSE) مقارنة مع الطريقة الاخرى .

2- ايضاً اثبتت طريقة (RJGM2) اكثر كفاءة في حالة زيادة حجوم العينات ونسب التلوث العالية .

3- اثبتت طريقة (RJMM) كفاءتها عند حجوم العينات الصغيرة .

10-2- **التوصيات**

1- استعمال طريقة (RJGM2) في تقدير معلمات أنموذج الانحدار الخطي المتعدد (MLR) وباختلاف احجام العينات لما تبديه من كفاءة ومرونة في التطبيق .

2- اعتماد طريقة (RJGM2) في تقدير معلمات أنموذج الانحدار الخطي المتعدد (MLR) في حالة احجام العينات الكبيرة ونسب تلوث العالية .

3- اعتماد طريقة (RJMM) عند حجوم العينات الصغيرة .

11- المصادر العربية

1. الجشعمي, حسين علي عبد الله 2007 م، " مقارنة لبعض المقدرات الحصينة لمعالم النماذج اللاخطية " اطروحة دكتوراه في الاحصاء, كلية الادارة والاقتصاد, الجامعة المستنصرية .

2. شهاب , ضمياء حامد (2017) م, " مقارنة بعض طرائق التقدير الحصينة مع أسلوب بيز في تقدير دالة الانحدار اللوجستي مع تطبيق عملي ", اطروحة دكتوراه في الاحصاء, كلية الادارة والاقتصاد, الجامعة المستنصرية.

3. كاظم، أموري هادي و مسلم، باسم شليبة 2002م، " القياس الاقتصادي المتقدم النظرية والتطبيق"، مكتبة دنيا الامل ، بغداد.

12- المصادر الاجنبية

1.Alguraibawi، M., Midi، H., & Rana، L. S. (2015) "Robust Jackknife Ridge Regression to Combat Multicollinearity and High Leverage Points in Multiple Linear Regressions" Economic Computation and Economic Cybernetics Studies and Research, NO. 4، PP 305-322.

2.Duran, E. A ., and Akdeniz, F. (2010) "Efficiency of the Modified Jackknifed Liu-Type Estimator" Statistical Papers, NO. 53, PP 265-280.

3.Kamruzzaman, MD, and Imon, A.H.M.R. (2002) "High leverage point, another source of multicollinearity" Pakistanian Journal of Statistics،NO. 18,PP 435-448.

4.Kan، B., Alpu, Ö., & Yazıcı B. (2013) "Robust Ridge and Robust Liu Estimator for Regression Based on the LTS Estimator" Journal of Applied Statistics, NO. 40 (3),PP 644-655.

5.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.

6.Midi, H., and Mohammed .M.A (2015) "The Identification of Good and Bad High Leverage Points in Multiple Linear Regression Model" Mathematical Methods and System in Science and Engineering, PP 147-153.

7.Singh, B., Chaubey, Y. P., & Dwivedi, T. D. (1986) "An almost unbiased ridge estimator" The Indian Journal of Statistics, NO.48 (3), PP 342-346.

Comparison of some robust methods in the presence of problems of multicollinearity and high leverage points

Ghufran Ismail Kamal / Assistant Prof- College Of Administration & Economics- University Of Baghdad-Statistics section - ghufran62@gmail.com .
Researcher Saif alemam saadi khazaal / College Of Administration & Economics- University Of Baghdad-Statistics section- saif.alemams@gmail.com

Abstract

The multiple linear regression model of the important regression models used in the analysis for different fields of science Such as business, economics, medicine and social sciences high in data has undesirable effects on analysis results . The multicollinearity is a major problem in multiple linear regression. In its simplest state, it leads to the departure of the model parameter that is capable of its scientific properties, Also there is an important problem in regression analysis is the presence of high leverage points in the data have undesirable effects on the results of the analysis , In this research , we present some of the robust methods in the multiple linear regression model These methods include the (Jackknife Ridge regression) methods based on the (MM) estimator and the (GM2) estimator (Modified Generalized M-estimator) . Using the Monte Carlo simulation, the two methods were compared in accordance with the comparison criterion, the mean squares error (MSE) and sample sizes (n = 20, n = 50, n = 100) and different pollution ratios (τ = 5%, 15%) , The comparison shows that (RJGM2) is the best method for estimating the parameters of the multiple linear regression model, which has the lowest value for mean squares error (MSE) compared with the rest of the other estimations.

Keywords : Multiple Linear Regression , Multicollinearity, high leverage point, Jackknife ridge regression, MM-estimator, GM2-estimator.