مجلة الكوت للعلوم الاقتصادية والادارية تصدر عن كلية الادارة والاقتصاد جامعة واسط العدد(28) الجزء الاول كانون الاول 2017

مقارنة بين مقدري Huber Lasso و Huber Elastic Net باستخدام المحاكاة(*)

comparison between Huber Lasso and Huber Elastic Net estimators by using simulation

أ.م.د. عماد حازم عبودي   م.د. علي حميد يوسف

جامعة بغداد – كلية الإدارة والاقتصاد   جامعة واسط – كلية الإدارة والاقتصاد

الملخص

ان طريقة المربعات الصغرى الجزائية تتمتع بخصائص منها ضمان الحصول على تنبؤ عالي الدقة وكذلك قيامها بعملية التقدير واختيار المتغيرات في ان واحد ، حيث انها تعطي نموذج متبعثر (Model Sparse) اي النموذج الذي يتضمن اقل عدد ممكن من المتغيرات وبالتالي يكون قابل للتفسير بسهولة. ان طريقة المربعات الصغرى الجزائية تكون حساسة جداً في حالة وجود نسبة من الشواذ في البيانات ،وللتعامل مع هذه المشكلة يتم استعمال دالة خسارة حصينة ليتم الحصول على طريقة المربعات الصغرى الجزائية الحصينة.

وفي هذا البحث تمت عملية المقارنة بين مقدري (Huber Lasso) و(Huber Elastic Net) باستعمال المحاكاة وقد تم التوصل الى افضلية مقدر (Huber Lasso)) في معظم التجارب وذلك بالاعتماد على معيار متوسط مربعات الخطأ ، معدل الايجابية الزائف ومعدل السلبية الزائف.

Abstract

The penalized least square method have many properties such that given high prediction accuracy and making estimation and variables selection At once. The penalized least square method gives a sparse model ,that meaning a model with small variables so that can be interpreted easily .The penalized least square is very sensitive to the presence of outlying observation , to deal with this problem, we can used a robust loss function to get the robust penalized least square method . In this paper a compression had been made Huber Elastic Net estimator and Huber Lasso estimator by using simulation and the simulation results show that the Huber Lasso estimator is best for every experiments, Depending on the criteria for the Mean Square Error, False Positive Rate and False negative Rate .

1- المقدمة

يعتبر الانحدار الخطي من اكثر الاساليب الاحصائية استخداماً في دراسة وتحليل العلاقة بين المتغيرات المدروسة ولكافة العلوم . وتعتبر طريقة المربعات الصغرى الاعتيادية من الطرائق الاحصائية الشائعة الاستعمال في تقدير معلمات نموذج الانحدار الخطي والتي تستند على مبدأ تصغير مجموع مربعات الخطأ . ومع ذلك فأن لطريقة المربعات الصغرى الاعتيادية تحديين رئيسيين .

(*) مستل من اطروحة دكتوراه للباحث الثاني .

التحدي الاول هو دقة التنبؤ ، حيث ان تقديرات المربعات الصغرى الاعتيادية لها تحيز قليل وتباين عالي ، وان دقة التنبؤ من الممكن ان يحسن من خلال التقليص او وضع بعض المعاملات للصفر وبهذه العملية سنضحي بقليل من التحيز لتقليل تباين القيم التنبؤية وبالتالي تحسين دقة التنبؤ ككل .

اما التحدي الثاني هو التفسير ، مع الزيادة في عدد المتغيرات التوضيحية فأننا نرغب في تحديد اصغر مجموعة جزئية بحيث من الممكن ان تحقق تأثيراً عالياً .

كما ان نماذج الانحدار الخطية التي تتضمن عدداً كبيراً من المتغيرات التوضيحية تكون ذات أداء ضعيف ، وذلك بسبب كبر التباين فضلاً عن ذلك فإنها تكون صعبة التفسير . وفي العديد من الدراسات يكون فيها عدد المتغيرات التوضيحية اكبر من حجم العينة الأمر الذي يؤدي إلى فشل النموذج التقليدي في التعامل مع البيانات عالية الإبعاد .

ان إحدى المسائل المهمة في الإحصاء هي اختيار المتغيرات في الانحدار ، ومع الزيادة في عدد المتغيرات التوضيحية مع صغر حجم العينة يصبح هنالك تحدياً رئيسياً في عملية تقدير المعلمات واختيار المتغيرات في النموذج .

ففي نموذج الانحدار التقليدي فأن أسلوب اختيار المتغيرات (الاختبارات الأمامية ، الاختبارات الخلفية ، الانحدار المتدرج ..الخ) هي غير ملائمة في التعامل مع البيانات ذات الابعاد العالية .

وفضلاً عن ذلك فان طريقة المربعات الصغرى الاعتيادية (Ordinary Least Square) والتي هي شائعة الاستخدام في الانحدار تكون غير ملائمة في التعامل مع البيانات ذات الابعاد العالية ، اذ إنه لا يمكن ان تكون مصفوفة المعلومات ذات رتبة كاملة الأمر الذي يؤدي الى عدم الحصول على حل وحيد .

ان الاسلوب الشائع للتعامل مع البيانات ذات الابعاد العالية هو طريقة المربعات الصغرى الجزائية والتي تستند على مبدأ تصغير مجموع مربعات الخطأ وفقاً لقيد معين على المعلمات .

ان من ضمن المزايا التي تتمتع بها طريقة المربعات الصغرى الجزائية هي ضمان الحصول على تنبؤ عالي الدقة وكذلك قيامها بعملية التقدير واختيار المتغيرات في ان واحد ، حيث تقوم بتقليص بعض المعاملات وجعل الاخرى مساوية للصفر . حيث انها تعطي نموذج متبعثر (Sparse Model) أي النموذج الذي يتضمن اقل عدد ممكن من المتغيرات وبالتالي يكون قابل للتفسير بسهوله.

كما ان طريقة المربعات الصغرى الجزائية هي غير حصينة بمعنى تتأثر بالقيم الشاذة ، وللتغلب على هذه المشكلة يتم استبدال دالة خسارة المربعات الصغرى الجزائية بدالة خسارة حصينة لنحصل على طريقة المربعات الصغرى الجزائية الحصينة ، ويكون المقدر الناتج يدعى بالمقدر الجزائي الحصين الذي يستطيع التعامل مع مشكلتي الابعاد والقيم الشاذة .

2- هدف البحث

يهدف هذا البحث الى المقارنة بين مقدري Huber Lasso و Huber Elastic Net لنموذج الانحدار والحصول على افضل مقدر من بين المقدرات الاخرى وبالتالي الحصول على افضل تقدير باستعمال المحاكاة وذلك بالاعتماد على معيار متوسط مربعات الخطأ ، معدل الايجابية الزائف ومعدل السلبية الزائف.

3- الجانب النظري

1-3 طريقة المربعات الصغرى الاعتيادية[2][1] Ordinary Least Square Method

في الانحدار الخطي يتم تحديد العلاقة الخطية بين متغيرين او اكثر من المتغيرات بحيث انه من الممكن ان يتم التنبؤ بأحد المتغيرات عن طريق الاخر . حيث المتغير المراد التنبؤ به يدعى بمتغير الاستجابة (Response Variable) او المتغير المعتمد(Dependent Variable) . اما المتغير التنبؤي يدعى بالمتغير التوضيحي (explanatory Variable) او المتغير المستقل (Independent Variable) .

ان نموذج الانحدار الخطي العام يكتب وفق الصيغة الاتية :-

$$Y = X\beta + \varepsilon \qquad (1)$$

حيث ان :-

Y: متجه المتغير التابع من البعد $(nx1)$

X : مصفوفة المتغيرات التوضيحية من البعد (nxp)

β : متجه المعلمات المراد تقديرها من البعد $(Px1)$

ε : متجه حد الخطأ العشوائي من البعد $(nx1)$ ، وحيث ان موجه الأخطاء يفترض ان يتوزع توزيع طبيعي بمتوسط صفر وتباين ثابت $\sigma^2 I_n$ ، بمعنى اخر ممكن كتابته كالاتي :-

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

ان مقدر المربعات الصغرى يكون على وفق للصيغة الاتية :-

$$\hat{\beta}_{LS} = \left(\acute{X}X\right)^{-1}\acute{X}Y \qquad (2)$$

اما مصفوفة التباين والتباين المشترك للمقدر $(\hat{\beta}_{LS})$ فأنها تعطى بالصيغة الاتية :-

$$Var - Cov\left(\hat{\beta}_{LS}\right) = \sigma^2\left(\acute{X}X\right)^{-1} \qquad (3)$$

ان موجه معلمات نموذج الانحدار الخطي المقدر بطريقة المربعات الصغرى (OLS) يمتلك أفضل تقدير خطي غير متحيز (Best Linear Unbiased Estimator) .

ان من ضمن شرط تطبيق المربعات الصغرى الاعتيادية ان لا يكون هنالك تعدد ارتباط عالي بين المتغيرات التوضيحية (التعدد الخطي) ، كما ينبغي ان يكون عدد المشاهدات اكبر من عدد المعلمات المطلوب تقديرها $p < n$ ، وهذا يعني ان رتبة المصفوفة X في النموذج (1) يجب ان تكون اقل من عدد المشاهدات اي ان :-

$$rank(X) = p < n$$

2-3 طريقة المربعات الصغرى الجزائية [1][4]

Penalized Least Square Method

تعد طريقة المربعات الصغرى الجزائية طريقة ملائمة وشائعة للتعامل مع البيانات عالية الابعاد ، اي التي يكون فيها عدد المتغيرات التوضيحية اكبر من حجم العينة ، حيث انه لا يمكن في هذه الحالة استخدام طريقة المربعات الصغرى الاعتيادية .

ان طريقة المربعات الصغرى الجزائية تستند على مبدأ تصغير مجموع مربعات الخطأ مع بعض القيود على المعلمات ، حيث يتم الحصول على تقديرات المربعات الصغرى الجزائية من خلال تقليل دالة الهدف (Object Function) والتي تتألف من جزئين هما دالة الخسارة (loss function) ودالة الجزاء (penalty function) والتي تكون وفقاً للصيغة الاتية :-

$$p_{ls}(\lambda, \beta) = (y - X\beta)' (y - X\beta) + n\sum_{j=1}^{p} p_\lambda\left(|\beta_j|\right) \qquad (4)$$

حيث ان:-

$p(\cdot)$: تمثل دالة الجزاء (penalty function).

λ : تمثل معلمة الجزاء (penalty parameter)

عليه فأن المقدر الجزائي يتم الحصول عليه وفقاً للصيغة الاتية :-

$$\hat{\beta} = \operatorname{argmin}\{p_{ls}(\lambda, \beta)\} \qquad (5)$$

ان طريقة المربعات الصغرى الجزائية تقوم بعملية التقدير واختيار المتغيرات في ان واحد .

ان دالة الجزاء الجيدة يجب ان تؤدي إلى ان مقدر المربعات الصغرى الجزائية يكون له ثلاث من الخصائص والتي تم اقتراحها من قبل (Fan &Li) وهي :-

1)**عدم التحيز (Unbiasedness) :** مقدر المربعات الصغرى الجزائية يكون غير متحيز تقريبا عندما تكون المعلمة الحقيقية المجهولة كبيرة لتجنب تحيز النمذجة غير الضرورية .

2)**التبعثر(Sparsity) :** مقدر المربعات الصغرى الجزائية يكون قاعدة مستوى العتبة والتي تضع التقديرات ذات المعاملات الصغيرة إلى الصفر .

3) **الاستمرارية (Continuity) :** مقدر المربعات الصغرى الجزائية يجب ان يكون دالة مستمرة في البيانات لتجنب عدم الاستقرارية في تنبؤ النموذج .

1-2-3 مقدر Lasso [5][6]

اقترح (Tibshirani) عام 1996 دالة جزاء لنموذج الانحدار الخطي تعرف ب (Lasso) وهي مختصر ل (Least Absolute Shrinkage and Selection Operator) لتقدير معلمات نموذج الانحدار الخطي وإجراء اختيار المتغير (Variable Selection) بشكل اني .

ان مبدأ طريقة (Lasso) هو تصغير مجموع مربعات البواقي وفقآ إلى قيد يمثل المجموع المطلق للمعاملات والتي تكون اصغر من ثابت معين . فلنموذج الانحدار الخطي (1) ، فأن مقدر (Lasso) ، $\hat{\beta}_{\text{Lasso}} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$ يتم الحصول عليه وفقاً للصيغة الاتية :-

$$\hat{\beta}_{\text{Lasso}} = \operatorname{argmin}_{\beta} \left\{ \sum_{i=1}^{n} (y_i - x_i'\beta)^2 + n\lambda \sum_{j=1}^{p} |\beta_j| \right\} \qquad (6)$$

حيث ان :-

λ : تمثل معلمة الجزاء (Penalty Parameter) و تدعى بمعلمة الضبط (Regularization Parameter) .

$\lambda \sum_{j=1}^{p} |\beta_j|$: تدعى بدالة الجزاء (Penalty Function) وتدعى ايضاً بدالة الضبط (Regularization Function) ويرمز لها بالرمز L_1 norm

ان طريقة (Lasso) تعتبر أكثر جاذبية من حيث اختيار المتغير لاحتفاظها بخصائص جيدة حيث يتم من خلالها وضع بعض معاملات الانحدار بالضبط للصفر وتقليص الأخرى بمقدار معين مع التقليل من دالة الخسارة ومن خلال ذلك يعني ان تقديرات (Lasso) يمكن ان تنتج مجموعة مبعثرة (Sparse Set) من معاملات الانحداروبذلك تعطينا نموذج أكثر تفسيرا .

ان مقدر (Lasso) في حالة كون المصفوفة X متعامدة (Orthogonal) يكون كما في الصيغة الاتية :-

$$\hat{\beta}_{\text{Lasso}} = \operatorname{sign}(\hat{\beta}_j^0)\left(|\hat{\beta}_j^0| - \lambda\right)_+ \quad ; \ j = 1,2,..,p \qquad (7)$$

حيث ان :-

+: تشير الى الجزء الموجب داخل القوسين

2-2-3 مقدر Elastic Net [3][8]

اقترح الباحثان (Zou& Hastie) عام 2005 دالة جزاء تدعى (Elastic Net Penalty) .

وان المقدر الناتج يدعى بمقدر (Elastic Net) .

ان مقدر(Elastic Net) هو مشابه إلى مقدر (Lasso) يعمل على اختيار المتغير والتقليص المستمر بشكل اني . وهو بإمكانه اختيار المتغيرات المصنفة (Group) المرتبطة.

ان دالة الجزاء (Elastic Net) تعرف كما في الصيغة الاتية :-

$$P_{\lambda\alpha}(\beta) = \lambda \sum_{j=1}^{p} \left[\frac{1}{2}(1-\alpha)\beta_j^2 + \alpha|\beta_j| \right] \qquad (8)$$

حيث ان :-

$P_{\lambda\alpha}(\beta)$: تمثل دالة الجزاء (Penalty Function)

λ :- تمثل معلمة الجزاء(Penalty Parameter) وان $\lambda > 0$

عليه فأن مسألة المربعات الصغرى الجزائية مع دالة الجزاء (Elastic Net) ستكون وفقا للصيغة الاتية:-

$$\frac{1}{2n}\sum_{i=1}^{n}(y_i - \acute{x}_i\beta)^2 + P_{\lambda\alpha}(\beta) \qquad (9)$$

ومن خلال اشتقاق المعادلة (9) الى β ومساواة المشتقة الاولى للصفر نحصل على مقدر المربعات الصغرى الجزائية مع دالة الجزاء (Elastic Net) وكما في الصيغة الاتية :-

$$\hat{\beta}_{\textbf{Elastic Net}} = \operatorname{argmin}_{\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}(y_i - \acute{x}_i\beta)^2 + P_{\lambda\alpha}(\beta)\right\} \qquad (10)$$

ولو عوضنا عن$\left(P_{\lambda\alpha}(\beta)\right)$ بما يساويها فأن المعادلة (10) تكون وفقاً للصيغة الاتية :-

$$\hat{\beta}_{\textbf{Elastic Net}} = \operatorname{argmin}_{\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}(y_i - \acute{x}_i\beta)^2 + \lambda\sum_{j=1}^{p}\left[\frac{1}{2}(1-\alpha)\beta_j^2 + \alpha|\beta_j|\right]\right\} \quad (11)$$

وعندما تكون قيمة $(\alpha = 1)$ فأن مقدر (Elastic Net) يختزل الى مقدر (Lasso) .

3-3 مقدر Huber Elastic-Net [7][8]

ان طريقة المربعات الصغرى الجزائية مع دالة الجزاء (Elastic-Net) تكون غير حصينة بمعنى تتأثر بالقيم الشاذة . ولذلك فقد اقترح (Yi& Huang) عام 2016 استبدال دالة خسارة المربعات الصغرى الجزائية مع دالة الجزاء (Elastic-Net) بدالة الخسارة (Huber) الحصينة لنحصل على طريقة (Huber Elastic-Net) وان المقدر الناتج من هذه الطريقة يدعى بمقدر (Huber Elastic-Net) .

ان دالة الهدف مؤلفة من حدين هما دالة الخسارة (Huber) ودالة الجزاء (Elastic-Net) والتي تكون وفقاً للصيغة الاتية :-

$$\frac{1}{n}\sum_{i=1}^{n}h_{\gamma}(y_i - \beta_0 - x_i'\beta) + P_{\lambda\alpha}(\beta) \qquad (\ 12\ )$$

حيث ان :-

$h_{\gamma}(.)$: تمثل دالة خسارة (Huber) والتي تكون وفقاً للصيغة الاتية :-

$$h_{\gamma} = \begin{cases} \dfrac{t^2}{2\gamma} & \text{if } |t| \le \gamma \\ |t| - \dfrac{\gamma}{2} & \text{if } |t| > \gamma \end{cases} \qquad (13)$$

γ : يمثل ثابت.

λ: تمثل معلمة الجزاء .

$P_{\lambda\alpha}(\beta)$: تمثل دالة الجزاء Elastic-Net والتي تكون وفقاً للصيغة الاتية :-

$$P_{\lambda\alpha}(\beta) = \lambda\sum_{j=1}^{p}\left[\frac{1}{2}(1-\alpha)\beta_j^2 + \alpha|\beta_j|\right] \quad ;\ 0 \le \alpha \le 1 \qquad (14)$$

ومن خلال اشتقاق المعادلة (12) بالنسبة الى β ومساواة المشتقة الاولى للصفر نحصل على مقدر (Huber Elastic-Net) ووفقاً للصيغة التالية :-

$$\left(\hat{\beta}_0, \hat{\beta}\right) = \mathrm{argmin}_{(\beta_0,\beta)} \left\{ \frac{1}{n} \sum_{i=}^{n} h_\gamma(y_i - \beta_0 - x_i'\beta) + P_{\lambda\alpha}(\beta) \right\} \qquad (15)$$

ان دالة الهدف في (15) تكون محدبة(convex) ، كما ان $\left(\hat{\beta}_0, \hat{\beta}\right)$ تحقق الشروط الكافية والضرورية (Karush Kuhn Tucker) والذي يرمز لها بالرمز (KKT) .

ليكن $\partial|t|$ تشير الى (sub gradients) لدالة القيمة المطلقة | .| عند t ، فأنه من الممكن ان نبين بأن

$$s \in \partial|t| \quad \text{iff} \quad t = S(t+s) \qquad (16)$$

حيث ان

S: يمثل عامل مستوى العتبة (soft- thresholding operator) بالعتبة 1

وبمعنى اخر فأن

$$S(z) = \mathrm{sgn}(z)(|z|-1)_+ \qquad (17)$$

عليه فأن شروط (KKT) للمعادلة (15) يمكن كتابتها كما في المعادلات الاتية :-

$$\begin{cases} -\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \acute{h}_\gamma\left(y_i - \hat{\beta}_0 - x_i'\hat{\beta}\right) = 0 \\ -\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \acute{h}_\gamma\left(y_i - \hat{\beta}_0 - x_i'\hat{\beta}\right)x_{ij} + \lambda\alpha\,\hat{s}_j + \lambda(1-\alpha)\hat{\beta}_j = 0 \\ \hat{\beta}_j - S\left(\hat{\beta}_j + \hat{s}_j\right) = 0 \ \ , j = 1,2,\dots,p \end{cases} \qquad (18)$$

اذ ان

$$\hat{s}_j \in \partial\left|\hat{\beta}\right|$$

وان

$\acute{h}_\gamma(.)$: تمثل المشتقة للدالة $h_\alpha(.)$ والتي تكون وفقاً للصيغة الاتية :-

$$\acute{h}_\alpha(t) = \begin{cases} \dfrac{t}{\gamma} & \text{if } |t| \le \gamma \\ \mathrm{sgn}(.) & \text{if } |t| > \gamma \end{cases} \qquad (19)$$

وقد اقترح الباحثان (Yi& Huang) خوارزمية لحل المعادلات (18) تدعى بإحداثية نيوتن شبه الممهدة (Semi smooth Newton Coordinate Descent) ويرمز لها بالرمز (SNCD). ويتم التدوير من خلال (β_o, β, s) عند كل خطوة ، اي ان الزوج (β_j, s_j) و $(\beta_o$ مع نفسها) وتتم عملية التحديث من خلال حل الجزء المناظر ل(18) ، بينما المتغيرات الأخرى تكون ثابتة عند القيم الحالية $\tilde{\beta}_k, \tilde{s}_k, k \neq j$. و يتم حل المعادلات التالية لكل خطوة وكالاتي:-

❖ بالنسبة الى (β_j, s_j) :

$$\begin{cases} -\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \acute{h}_\gamma\left(\tilde{r}_i + x_{ij}\tilde{\beta}_j - x_{ij}\beta_j\right)x_{ij} + \lambda\alpha\, s_j + \lambda(1-\alpha)\beta_j = 0 \\ \beta_j - S\left(\beta_j + s_j\right) = 0 \end{cases} \qquad (20)$$

❖ النسبة الى (β_o) :

$$-\frac{1}{n}\sum_{i=1}^{n}\acute{h}_{\gamma}\left(\tilde{r}_i+\tilde{\beta}_0-\beta_o\right)=0 \qquad (21)$$

حيث ان :-

$$\tilde{r}_i = y_i-\tilde{\beta}_o-x_i'\tilde{\beta} \quad , \quad i=1.2,\dots,n$$

ولتكن

$$\varphi_{\gamma}(t)=\frac{1}{\gamma}\ I(|t|\le\gamma) \qquad (22)$$

وان

$$\varphi_{\gamma}\in\nabla_N\acute{h}_{\gamma}(t)\ ,\forall t\in R.$$

عليه فأن الخوارزمية (SNCD) التكرارية تكون وفقاً للخطوات الاتية :-

(1) يتم تجديد β_0 .

ليكن

$$F_0\left(z;\tilde{\beta}\right)=-\frac{1}{n}\sum_{i=}^{n}\acute{h}_{\gamma}\left(\tilde{r}_i+\tilde{\beta}_0-z\right)$$

بحيث ان :-

$$H_0(z)=\frac{1}{n}\sum_{i=1}^{n}\varphi_{\gamma}\left(\tilde{r}_i+\tilde{\beta}_0-z\right)\in\nabla_N\left(F_0(z)\right)$$

ويتم تجديد (β_0) لكي نصل على المعادلة الاتية :-

$$\hat{\beta}_0=\tilde{\beta}_0+\frac{\sum_{i=1}^{n}\acute{h}_{\gamma}(\tilde{r}_i)}{\sum_{i=1}^{n}\sum_{i=1}^{n}\varphi_{\gamma}(\tilde{r}_i)} \qquad (23)$$

2) يتم تجديد $\left(z;\tilde{\beta}\right)$.

ليكن

$$F_j\left(z;\tilde{\beta}\right)=\begin{bmatrix}-\frac{1}{n}\sum_{i=}^{n}\acute{h}_{\gamma}\left(\tilde{r}_i+x_{ij}\tilde{\beta}_j-x_{ij}z_1\right)x_{ij}+\lambda\alpha\, z_2+\lambda(1-\alpha)z_1=0\\ z_1-S(z_1+z_2)\end{bmatrix}$$

حيث ان

$$z=(z_1,z_2)'$$

بحيث :-

ان

$$z_1-S(z_1+z_2)=\begin{cases}-z_2+\mathrm{sgn}(z_1+z_2) & \text{if } |z_1+z_2|>1\\ z_1 & \text{if } |z_1+z_2|\le 1\end{cases} \qquad (24)$$

ويتم ايجاد الحل ل(β_j,s_j) من (20) في نوعين من التجديدات :-

(a) في حالة كون $\left|\tilde{\beta}_j+\tilde{s}_j\right|>1$

بالنسبة الى z و $|z_1+z_2|>1$ ، فأن مشتقة (Newton) لF_j عند z تكون وفقاً للصيغة التالية :-

$$H_j(z) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \varphi_\gamma\left((\tilde{r}_i - x_{ij}\tilde{\beta}_j - x_{ij}z_1)x_{ij}^2 + \lambda(1-\alpha) \right. & \left. \lambda\alpha \right) \\ 0 & -1 \end{bmatrix} \in \nabla_N F_j(z) \quad (25)$$

وعليه فأن التجديد يكون كالاتي :-

$$\begin{bmatrix} \hat{\beta}_j \\ s_j \end{bmatrix} = \begin{bmatrix} \tilde{\beta}_j + \dfrac{\frac{1}{n}\sum_{i=1}^{n} \acute{h}_\gamma(\tilde{r}_i)x_{ij} - \lambda\alpha\, \mathrm{sgn}(\tilde{\beta}_j + \tilde{s}_j) - \lambda(1-\alpha)\tilde{\beta}_j}{\frac{1}{n}\sum_{i=1}^{n} \varphi_\gamma(\tilde{r}_i)x_{ij}^2 + \lambda(1-\alpha)} \\ \mathrm{sgn}(\tilde{\beta}_j + \tilde{s}_j) \end{bmatrix} \quad (26)$$

(b) في حالة كون $|\tilde{\beta}_j + \tilde{s}_j| \le 1$

بالنسبة ل z مع $|z_1 + z_2| \le 1$ ، فأن مشتقة (Newton) ل F_j عند z تكون وفقاً للصيغة التالية :-

$$H_j(z) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \varphi_\gamma\left((\tilde{r}_i - x_{ij}\tilde{\beta}_j - x_{ij}z_1)x_{ij}^2 + \lambda(1-\alpha) \right. & \left. \lambda\alpha \right) \\ 1 & 0 \end{bmatrix} \in \nabla_N F_j(z) \; (27)$$

وعليه فأن التجديد يكون وفقاً للصيغة الاتية :-

$$\begin{bmatrix} \hat{\beta}_j \\ s_j \end{bmatrix} = \begin{bmatrix} 0 \\ \dfrac{\frac{1}{n}\sum_{i=1}^{n} \acute{h}_\gamma(\tilde{r}_i)x_{ij} + \tilde{\beta}_j \cdot \frac{1}{n}\sum_{i=1}^{n} \varphi_\gamma(\tilde{r}_i)x_{ij}^2}{\lambda\alpha} \end{bmatrix} \quad (28)$$

ويتم الحصول على معلمة الجزاء (λ) من خلال الاعتماد معيار العبور الشرعي الحصين .

4-3 مقدر Huber Lasso [7][8]

ان طريقة المربعات الصغرى الجزائية مع دالة الجزاء (Lasso) تكون غير حصينة بمعنى تتأثر بالقيم الشاذة . ولذلك فقد اقترح (Yi& Huang) عام 2016 استبدال دالة الخسارة المربعات الصغرى الجزائية مع دالة الجزاء (Lasso) بدالة الخسارة (Huber) الحصينة لنحصل على طريقة (Huber Lasso) ، وان المقدر الناتج يدعى بمقدر(Huber Lasso) .

ان دالة الهدف مؤلفة من حدين هما دالة الخسارة (Huber) ودالة الجزاء (Lasso) والتي تكون وفقاً للصيغة الاتية :-

$$\frac{1}{n}\sum_{i=1}^{n} h_\gamma(y_i - \beta_0 - x_i'\beta) + \lambda\sum_{j=1}^{p}|\beta_j| \quad (29)$$

ومن خلال اشتقاق المعادلة (29) بالنسبة الى β ومساواة المشتقة الاولى للصفر نحصل على مقدر (Huber Lasso) ووفقاً للصيغة الاتية :-

$$(\hat{\beta}_0, \hat{\beta}) = \mathrm{argmin}_{(\beta_0,\beta)} \left\{ \frac{1}{n}\sum_{i=1}^{n} h_\gamma(y_i - \beta_0 - x_i'\beta) + \lambda\sum_{j=1}^{p}|\beta_j| \right\} \quad (30)$$

عليه فأن شروط (KKT) ل(30) يمكن كتابتها كما في المعادلات التالية :-

$$\begin{cases} -\frac{1}{n}\sum_{i=1}^{n}\acute{h}_{\gamma}\left(y_i-\hat{\beta}_0-x_i'\hat{\beta}\right)=0 \\ -\frac{1}{n}\sum_{i=1}^{n}\acute{h}_{\gamma}\left(y_i-\hat{\beta}_0-x_i'\hat{\beta}\right)x_{ij}+\lambda\,\hat{s}_j=0 \\ \hat{\beta}_j-S\left(\hat{\beta}_j+\hat{s}_j\right)=0\ \ ,j=1,2,\dots,p \end{cases} \qquad (31)$$

اذ ان

$$\hat{s}_j\in\partial\left|\hat{\beta}\right|$$

وان

$\acute{h}_{\gamma}(.)$: تمثل المشتقة للدالة $h_{\alpha}(.)$

وبتطبيق خوارزمية (SNCD) واستعمال نفس الخطوات التي تم توضيحه في (Huber Elastic-Net) فأن خوارزمية (SNCD)يتم تجديدها وكالاتي :-

(1) بالنسبة الى β_o :

$$\hat{\beta}_0=\tilde{\beta}_0+\frac{\sum_{i=1}^{n}\acute{h}_{\gamma}(\tilde{r}_i)}{\sum_{i=1}^{n}\sum_{i=1}^{n}\varphi_{\gamma}\,(\tilde{r}_i)} \qquad (32)$$

(2) بالنسبة الى (β_j,s_j) :

(a)اذا كان $\left|\tilde{\beta}_j+\tilde{s}_j\right|>1$ فأن :

$$\begin{bmatrix}\hat{\beta}_j\\ s_j\end{bmatrix}=\begin{bmatrix}\tilde{\beta}_j+\dfrac{\frac{1}{n}\sum_{i=1}^{n}\acute{h}_{\gamma}(\tilde{r}_i)x_{ij}-\lambda\,\mathrm{sgn}\left(\tilde{\beta}_j+\tilde{s}_j\right)}{\frac{1}{n}\sum_{i=1}^{n}\varphi_{\gamma}(\tilde{r}_i)x_{ij}^2}\\ \mathrm{sgn}\left(\tilde{\beta}_j+\tilde{s}_j\right)\end{bmatrix} \qquad (33)$$

(b) اذا كان $\left|\tilde{\beta}_j+\tilde{s}_j\right|\le 1$ فأن :

$$\begin{bmatrix}\hat{\beta}_j\\ s_j\end{bmatrix}=\begin{bmatrix}0\\ \dfrac{\frac{1}{n}\sum_{i=1}^{n}\acute{h}_{\gamma}(\tilde{r}_i)x_{ij}+\tilde{\beta}_j\,.\frac{1}{n}\sum_{i=1}^{n}\varphi_{\gamma}(\tilde{r}_i)x_{ij}^2}{\lambda}\end{bmatrix} \qquad (34)$$

ويتم الحصول على معلمة الجزاء (λ) من خلال الاعتماد معيار العبور الشرعي الحصين .

4-الجانب التجريبي

من اجل المقارنة بين المقدرات الجزائية الحصينة الواردة في الجانب النظري تم الاعتماد على اسلوب المحاكاة (Simulation) بطريقة (Monte Carlo) بغية الحصول على عدد كبير من التجارب والتي تكون اكثر شمولية من حيث حجوم العينات والقيم الاولية للمعلمات ، وقد تم الاعتماد على المعايير الاحصائية متوسط مربعات الخطأ (MSE) ،معدل الايجابية الكاذب (*FPR*) ومعدل السلبية الكاذب(*FNR*) والتي توصف كالاتي :-

1) متوسط مربعات الخطأ Mean Square Error (MSE)

ان احدى اهداف تقدير النموذج المتبعثر (*Sparse Model Estimation*) هو تحسين اداء التنبؤ ، والمقدرات المختلفة يتم تقيمها من خلال متوسط مربعات الخطأ ويرمز له بالرمز (*MSE*) والذي يكون وفقاً للصيغة الاتية :-

$$MSE(\hat{\beta}) = \frac{1}{p}\sum_{j=1}^{p}\left(\hat{\beta}_j - \beta_j\right)^2 \qquad (35)$$

ويكون المقدر الافضل هو الذي يعطي اقل (MSE) .

2) فيما يتعلق بالتبعثر ، يتم تقييم النماذج المقدرة من خلال معدل الايجابية الزائف (*False Positive Rate*) ويرمز له بالرمز (*FPR*) ومعدل السلبية الزائف (*False Negative Rate*) ويرمز له بالرمز(*FNR*)

ان معدل الايجابية الزائف هو المعاملات المساوية للصفر في النموذج الحقيقي ، ولكن تكون غير صفرية في التقدير . اما معدل السلبية الزائف هي المعاملات غير الصفرية في النموذج الحقيقي ولكن تكون صفرية في التقدير.

$$FPR(\hat{\beta}) = \frac{\left|\left\{j \in \{1,\dots,p\}: \hat{\beta}_j \neq 0 \cap \beta_j = 0\right\}\right|}{\left|\left\{j \in \{1,\dots,p\}: \beta_j = 0\right\}\right|} \qquad (36)$$

$$FNR(\hat{\beta}) = \frac{\left|\left\{j \in \{1,\dots,p\}: \hat{\beta}_j = 0 \cap \beta_j \neq 0\right\}\right|}{\left|\left\{j \in \{1,\dots,p\}: \beta_j \neq 0\right\}\right|} \qquad (37)$$

1-4 خطوات اجراء المحاكاة

*1-1-4*توليد المتغيرات التوضيحية

يتم توليد المتغيرات التوضيحية وفق التوزيع الطبيعي متعدد المتغيرات ووفق الصيغة الاتية:-

$X \sim MN(0,\sum)$

حيث ان :-

$\sum_{ij} = \rho^{|i-j|} \quad , \quad \rho = 0.5$

*2-1-4*توليد الاخطاء العشوائية

يتم توليد الاخطاء العشوائية وفقاً للتوزيع الطبيعي بمتوسط (صفر) وتباين (1) اي ان :-

$\epsilon_i \sim N(o,1) \quad , i = 1,2,\dots,n$

اما نسب التلويث فهي (*0%,10%,20%*) وان عملية التلويث تمت عن طريق تلويث توزيع حد الخطأ ب$N(20,1)$ بدل $N(o,1)$

3-1-4 وصف التجارب والقيم الافتراضية

يتم وصف القيم الافتراضية للمعلمات وحجوم العينات كالاتي :-

التجربة الاولى :- *p=40 ,n=50,100*

$\beta = (3,1.5,0,0,2,0,\dots,0)$

التجربة الثانية :- *p=150 ,n=50,100*

$\beta = (3,3,3,3,3,0,\dots,0)$

جدول(1)

يبين نتائج التجربة الاولى ولجميع المقدرات الجزائية الحصينة

n	نسب التلويث	Methods	MSE	FPR	FNR
	0%	Elastic Net	0.00967	0.22	0
		Huber Elastic Net	0.00926	0.22	0
		Huber Lasso	0.00807	0.18	0
		Elastic Net	0.27384	0.17	0.21

50	10%	Huber Elastic Net	0.01542	0.17	0
		Huber Lasso	0.01316	0.14	0
	20%	Elastic Net	0.37831	0.15	0.36
		Huber Elastic Net	0.03688	0.16	0.01
		Huber Lasso	0.03061	0.12	0.01
100	0%	Elastic Net	0.00326	0.19	0
		Huber Elastic Net	0.00386	0.20	0
		Huber Lasso	0.00330	0.16	0
	10%	Elastic Net	0.11908	0.19	0.06
		Huber Elastic Net	0.00609	0.19	0
		Huber Lasso	0.00513	0.15	0
	20%	Elastic Net	0.19007	0.17	0.17
		Huber Elastic Net	0.01124	0.18	0
		Huber Lasso	0.00937	0.14	0

جدول(2)
يبين نتائج التجربة الثانية ولجميع المقدرات الجزائية الحصينة

n	نسب التلويث	Methods	MSE	FPR	FNR
50	0%	Elastic Net	0.00321	0.09	0
		Huber Elastic Net	0.00417	0.11	0
		Huber Lasso	0.00339	0.08	0
	10%	Elastic Net	0.11743	0.09	0.06
		Huber Elastic Net	0.00929	0.07	0
		Huber Lasso	0.00696	0.05	0
	20%	Elastic Net	0.17996	0.09	0.13
		Huber Elastic Net	0.05169	0.07	0.03
		Huber Lasso	0.04814	0.05	0.04
100	0%	Elastic Net	0.00136	0.09	0
		Huber Elastic Net	0.00169	0.11	0
		Huber Lasso	0.00139	0.08	0
	10%	Elastic Net	0.04928	0.09	0
		Huber Elastic Net	0.00255	0.08	0
		Huber Lasso	0.00208	0.06	0
	20%	Elastic Net	0.09244	0.10	0.02
		Huber Elastic Net	0.00593	0.07	0
		Huber Lasso	0.00456	0.05	0

5-الاستنتاجات والتوصيات

1) من خلال تحليل نتائج التجربة الاولى فأن مقدر (Huber Lasso) هو الافضل بشكل مطلق في حالة وجود الشواذ في البيانات كونه يعطي اقل قيمة ل (MSE) ولمختلف حجوم العينات وكذلك نسب التلويث

2) من خلال تحليل نتائج التجربة الاولى فأن مقدر (Huber Lasso) هو الافضل بشكل مطلق من ناحية اختيار المتغيرات كونه يعطي اقل قيمة ل(FPR) وكذلك (FNR) .

3) من خلال تحليل نتائج التجربة الثانية فأن مقدر (Huber Lasso) هو الافضل بشكل مطلق في حالة وجود الشواذ في البيانات كونه يعطي اقل قيمة ل (MSE) ولمختلف حجوم العينات وكذلك نسب التلويث .

4) من خلال تحليل نتائج التجربة الثانية فأن مقدر (Huber Lasso) هو الافضل بشكل مطلق من ناحية اختيار المتغيرات كونه يعطي اقل قيمة ل(FPR) وكذلك (FNR).

5) نوصي باعتماد طريقة (Huber Lasso) في حالة وجود الشواذ في البيانات .

6) نوصي باستعمال دوال جزاء اخرى في عملية التقدير.

المصادر

[1]Buhlmann , Peter. Geer, Sara (2013) " Statistics for High Dimensional Data Methods, Theory and Applications". Springer

[2]Coster,Jamie De (2003) "Notes on Applied Linear Regression". Department of Social Psychology Free University Amsterdam*137–152*

[3]Friedman, Jerome. Hastie, Trevor. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical software 33*.

[4] Fu, Wenjiang J (1998), "Penalized Regressions: The Bridge Versus the Lasso", *Journal of Computational and Graphical Statistics, Volume 7, Number 3, Pages 397–416*

[5] Hesterberg, T., Choi, N.H., Meier, L. and Fraley, C. (2008)"Least angle and ℓ1 penalized regression:A review. Statistics Surveys, 2, 61-93.

[6] Tibshirani, R. (1996)"Regression shrinkage and selection via the lasso" *J. Royal. Statist. Soc B., vol. 58, no. 1, pp. 267–288*

[7] Yi ,Congrui . Huang , Jian (2016)" Semismooth Newton Coordinate Descent Algorithm for Elastic-Net Penalized Huber Loss Regression and Quantile Regression .

[8]Zou, Hui.& Hastie, Trevor. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society B, 67*, 301–320.