Journal of Techniques, ISSN: 2708-8383, Vol. 5, No. 4, 2023, Pages 220-227
DOI: https://doi.org/10.51173/jt.v5i4.1329

JOURNAL OF TECHNIQUES

Journal homepage: *http://journal.mtu.edu.iq*

RESEARCH ARTICLE - MANAGEMENT

Estimation of Fuzzy Lasso Regression Model

Rawya Emad Kareem[1*], Mohammed Jasim Mohammed[1]

[1]Department of Statistics, College of Administration & Economics, University of Baghdad, Baghdad, Iraq

[*] Corresponding author E-mail: rawya.emad1101a@coadec.uobaghdad.edu.iq

Article Info.

Article history:

Received
17 March 2023

Accepted
13 June 2023

Publishing
31 December 2023

Abstract

Estimating the fuzzy lasso regression model when the data suffers from the problem of multicollinearity because of the failure of using the fuzzy least squares method in estimating the model by using the fuzzy theory using a triangular function, as the fuzzy lasso regression gives better results by comparing the results with the squares method fuzzy minimum using the mean square error (MSE) criterion The simulation results showed that the fuzzy lasso regression method gave less MSE than the fuzzy least squares method in the presence of the problem of multicollinearity in the data.

Publisher: Middle Technical University

Keywords: Fuzzy Least Squares; Fuzzy Regression; Fuzzy Lasso Regression; Fuzzy Numbers; Triangular Function.

تقدير انموذج انحدار لاسو الضبابي

راويه عماد كريم[1*]، محمد جاسم محمد[1]

[1] جامعة بغداد - كلية الادارة والاقتصاد - قسم الاحصاء

* البريد الإلكتروني: rawya.emad1101a@coadec.uobaghdad.edu.iq

معلومات المقالة

تاريخ الاستلام
16 اذار 2023

تاريخ القبول
13 حزيران 2023

تاريخ النشر
31 كانون الاول 2023

الخلاصة

تقدير انموذج انحدار لاسو الضبابي عندما تعاني البيانات من مشكلة التعدد خطي لفشل استخدام طريقة المربعات الصغرى الضبابية في تقدير الانموذج عن طريقة استخدام نظرية المجموعة الضبابية باستخدام دالة مثلثية حيث ان انحدار لاسو الضبابي (Fuzzy Lasso regression) يعطي نتائج افضل من خلال مقارنة النتائج مع طريقة المربعات الصغرى الضبابية باستخدام معيار متوسط مربعات الخطأ MSE وقد اظهرت النتائج باستخدام المحاكاة ان طريقة انحدار لاسو الضبابي قد اعطى اقل MSE من طريقة المربعات الصغرى الضبابية في ظل وجود مشكلة التعدد الخطي في البيانات.

الكلمات المفتاحية: المربعات الصغرى الضبابية؛ الانحدار الضبابي؛ انحدار لاسو الضبابي؛ المجموعة الضبابية؛ الدالة المثلثية.

1. المقدمة

أن الفكرة الأساسية في الإحصاء هو أيجاد مقدرات لمعلمات المجتمع على أساس معلومات صحيحة داخل العينة، أن اغلب بيانات البلدان النامية تعاني من عدم دقتها أو فقر تقدير مقاييسها بسبب التخلف الاجتماعي ونقص التسجيل وقلة الوعي مما جعل نوعية واستخدام هذه البيانات محدود جداً، ومن هنا جاء السؤال هل من الممكن استغلال مصادر عدم الدقة في مثل هذا النوع من البيانات لغرض بناء وتحليل نماذج انحدار تستخدم للتقدير والتنبؤ بشكل كفوء عوضا عن إهمالها، وتكمن أهمية البحث إلى تقدير معلمات انموذج الانحدار الضبابي باستخدام الدالة المثلثية بطريقة المربعات الصغرى الضبابية وبما ان هناك مشكلة التعدد الخطي في البيانات سيتم اللجوء الى انحدار لاسو الضبابي للمعالجة المشكلة وبناء انموذج انحدار ضبابي باستخدام المحاكاة إذ يكون متغير الاستجابة أرقاماً ضبابية مثلثية والمعلمات ضبابية اما المتغيرات المستقلة ستكون غير ضبابية ، ويتم استخدام معيار متوسط مربعات الخطأ Mean Square Error(MSE) للمقارنة بين نتائج طريقة تقدير المربعات الصغرى الضبابية وطريقة التقدير باستخدام انحدار لاسو الضبابي Fuzzy Lasso regression [1].

1.1. *مشكلة الدراسة*

لغرض تقدير أنموذج انحدار خطي ضبابي لا بد من تمتع الانموذج بخصائص معينة تعتمد على افتراضات عدة وفي حال **غياب** احد تلك الافتراضات فإن الانموذج سيعاني جملة من المشكلات التي تجعل عملية التقدير خاطئة او في بعض الاحيان غير ممكنة ومن هذه المشاكل هي مشكلة التعدد الخطي اي عدم وجود استقلالية بين المتغيرات التوضيحية إذ تكون البيانات مترابطة فيما بينها ولعلاج هذه المشكلة تم استعمال طريقة انحدار لاسو(LASSO) للتقدير معلمات أنموذج الانحدار الخطي الضبابي المتعدد.

1.2. هدف الدراسة

تعدّ نماذج الانحدار الخطي الضبابي من النماذج التي تستخدم في الجانب التطبيقي، إذ تكون احدى متغيرات الانموذج أرقاماً ضبابية، **وللاستفادة** من هذه النماذج لا بد من تقدير معلماتها، لذلك هدفنا في هذه الدراسة إلى تقديم طرائق لتقدير معلمات مثل هذا النوع من النماذج في حالة وجود مشكلة تعدد الخطي في البيانات والتغلب على مشكلة التعدد الخطي وفهم اعمق للظاهرة وتناول طرق الكشف عنها واهم المظاهر الدالة على وجودها في نموذج الانحدار الخطي الضبابي ودراسة طريقة انحدار Lasso لمشكلة التعدد الخطي في البيانات للحصول على مقدرات اكثر دقة والمقارنة بينها للحصول على افضل تقدير للمعلمات أنموذج الانحدار الخطي المضبب .

2. الدراسات السابقة

استخدمت (الغنام) في العام 2009 الانحدار المضبب في دراسة المتغيرات المضببة بالاعتماد على نظرية المجموعة الضبابية ودراسة الانحدار المتعدد المضبب وتقدير نموذج معلمات الانحدار المضبب باستخدام طريقة المربعات الصغرى المضببة [2].

في عام 2017 قام الباحث (احمد فاروق عباس) بعمل بحث علمي حديث عن الانحدار الضبابي اذا قدر أنموذج الانحدار الضبابي المكيف باستخدام دالة الأنتروبي وهدفت الدراسة الى تطوير وتوسعة عدد من نماذج الانحدار الضبابي في حالة اعتماد مصدر عدم الدقة في البيانات مصدرا لعدم التأكد ولكن بدلاً عن السائد في تحليل الانحدار الخطي الضبابي تم اعتماد طريقة تضبيب حديثة التي تعتمد على الموقع ودوال الأنتروبي لرقمين ضبابيين ثلاثيين ورباعيين ذاتا شكل منحني انتماء مختلف بدلاً عن استخدام دوال الانتماء التي تعتمد على مجموعة المركز والانتشار، واستعمل معامل التحديد المصحح الضبابي ومتوسط الفرق المطلق ومخطط تايلور كمقاييس لتقيم الأداء، وأظهرت نتائج الدراسة كفاءة استخدام الموقع ودوال الأنتروبي لوصف الأرقام الضبابية وتفوقها على استخدام دوال الانتماء [3].

في عام 2019 قام الباحثان Mohammad Ghasem Akbari & Gholamreza Hesamian بتحليل الانحدار الضبابي متعدد المتغيرات إلى نمذجة العلاقة بين مجموعة من الاستجابات الضبابية ومجموعة من المتغيرات التوضيحية غير الضبابية. واستخدام طريقة انحدار لاسو لنموذج الانحدار الخطي المتعدد الذي يمتلك متغيرات تفسيرية غير ضبابية واستجابات غير واضحة. ان طريقة Lasso قادرة على زيادة قابلية تفسير الانموذج عن طريق إزالة المتغيرات غير ذات الصلة بمتغيرات الاستجابة الضبابية. أظهرت النتائج العددية بوضوح دقة أعلى لطريقة لاسو الضبابية المقترحة مقارنة بنماذج الانحدار المتعدد الغامض الأخرى الموجودة في تحديد المتغيرات التوضيحية غير المعلوماتية. ومن ثم، يمكن تطبيق نموذج الانحدار الضبابي Lasso المقترح بنجاح لتحسين دقة التنبؤ وقابلية تفسير نماذج الانحدار المتعددة الغامضة لتطبيقات الحياة الواقعية في الأنظمة الخبيرة [4].

ومن كل ما سبق من الدراسات التي تم عرضها، لوحظ ندرة الدراسات العربية التي تناولت موضوع انحدار لاسو المضبب في معالجة مشكلة التعدد الخطي لإنه من الطرائق الحديثة، في هذا البحث سيتم تسليط الضوء عليه في تقدير انموذج الانحدار الضبابي عندما تكون الضبابية في المخرجات والمعالم وتكون المدخلات حقيقية.

3. المفاهيم النظرية

3.1. المجموعة الضبابية Fuzzy sets

يعدّ الباحث (Lotfi zadeh) في عام (1965) هو أول من طرح فكرة المجموعة الضبابية. فمن المعروف في نظرية المجموعات الاعتيادية بان العنصر في المجموعة يأخذ قيمتين هما{0,1}فإذا كانت لدينا المجموعة الشاملة X وكانت Aهي مجموعة جزئية من المجموعة الشاملة X فإن اي عنصر من عناصر المجموعة الشاملة أما ينتمي إلى المجموعة الجزئية A أو لا ينتمي إلى المجموعة الجزئية A [5].

أما في نظرية المجموعات الضبابية فإن العنصر في المجموعة يأخذ مجموعة قيم محصورة بين [0,1] مع درجة عضوية معينة , أي أن العنصر يحدد من خلال درجة عضويته.

لذا فإن المجموعة الضبابية تتصف بوجود دالة الانتماء(Membership function) أو دالة درجة العضوية.

$A=\{[\ x_i, \mu\widetilde{A}_{\ (x_i)}\ ]\}$ $\qquad\qquad$ $X=\{x_i, i=1,2,3\ldots,n\}$

إذ إن $\mu\widetilde{A}_{\ (x_i)}$دالة انتماء ودرجة عضوية العنصر x_i في المجموعة A.

فعندما يأخذ العنصر درجة عضوية (1) فهذا يعني ان العنصر ينتمي بالتمام إلى المجموعة الضبابية وعندما تكون درجة العضوية (0) فهذا يعني أن العنصر لا ينتمي أطلاقا إلى المجموعة والدرجات الأخرى تتفاوت بين الصفر والواحد فعندما تكون درجة العضوية (0.5) فهذا يعني إن العنصر ينتمي بنسبة (0.5) إلى المجموعة الضبابية ولا ينتمي إلى المجموعة بالنسبة نفسها ويدعى هذا العنصر بنقطة التوازن(Equilibrium point) وعندما تكون درجة العضوية (0.9) فهذا يعني أن العنصر ينتمي إلى المجموعة الضبابية بنسبة (0.9) ولا ينتمي بنسبة (0.1) وهذا اقرب إلى العضوية من عدمه [6].

لذا فإن نظرية المجموعة الضبابية هي توسيع لنظرية المجاميع الكلاسيكية (الاعتيادية)(Crisp set) وان نظرية المجاميع الاعتيادية هي حالة خاصة من نظرية المجاميع الضبابية.

قدم العديد من الباحثين تعاريف حول المجموعة الضبابية فقد عرفها الباحث kaufamm في عام 1975بأنها تلك المجموعة التي لا يكون فيها حدود واضحة بدقة بين تلك العناصر التي تنتمي وتلك التي لا تنتمي إليها [7].

في حين عرفها الباحث Zimmerman التي تعدّ أكثر التعاريف دقة عام 1988 وهو كالاتي:

اذا كانت X هي مجموعة ضبابية من العناصر يرمز لها عموما بالرمز x فإن المجموعة الضبابية A في X هي مجموعة من الأزواج المرتبة.

$$A=\{x,\mu\widetilde{A}_{\ (x_i)}/x\epsilon X\} \qquad (1)$$

إذا إن $\mu\widetilde{A}_{\ (x_i)}$هي دالة العضوية إلى x في $\widetilde{A}$ التي هي دالة من X إلى M إذا إن M مجال العضوية المستمر في المدة المغلقة (0,1).

3.2. الأعداد الضبابية (Fuzzy numbers)

يعرف العدد الضبابي $\widetilde{A}$ بانه عبارة عن مجموعة ضبابية على خط الأعداد الحقيقي R ولا بد أن يحقق الشروط التالية:

- يوجد على الأقل عنصر واحد X_0 R∈ إذ إن $\mu_{\widetilde{A}}=1$.
- $\mu\widetilde{A}_{\ (X)}$هو زوج مرتب مستمر.
- $\widetilde{A}$لا بد أن يكون طبيعيا ومقعرا او (محدبا) [8].

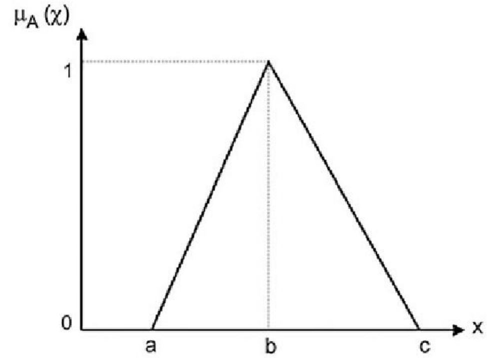
3.3. دالة الانتماء(The membership function)

وهي دالة تعبر عن درجة الانتماء او درجة العضوية التي تكون أعداد حقيقية ضمن المدة المغلقة [0,1] ويعبر عنها بدرجة العضوية أو الانتماء ($M(F_{(X)})$ التي تمثل درجة انتماء العنصر من المتغير xإلى مجموعة المضببة وتمتلك مجموعة من الدوال الخطية وسيتم في بحثنا هذا الاعتماد على استخدام الدالة المثلثية:

3.3.1. الدالة المثلثية (Triangular function)

هي من أكثر دوال الانتماء وأكثرها شيوعا واستخداما وتمتلك هذه الدالة ثلاث معلمات أساسية (c ،b ،a). اي عدد ضبابي مثلثي يمكن تمثيله بواسطة ثلاثة إعداد حقيقية [9], ويمكن تمثيل الدالة المثلثة كما في الصيغة الاتية , لاحظ الشكل (1):

$$\mu\widetilde{A}_{\,(x_i)} = \begin{cases} \dfrac{(x-a)}{(b-a)} & a \le x \le b \\ 1 & x = b \\ \dfrac{(c-x)}{(c-b)} & b \le x \le c \end{cases} \quad (2)$$



إذ إن: a) الحد الأدنى b) الحد الوسط(المركز) c) الحد الأعلى

الشكل (1) دالة الانتماء المثلثية

3.3.2. انموذج الانحدار الخطي الضبابي(المضبب) (Fuzzy Liner Regression)

يستخدم أنموذج الانحدار المضبب الخطي لتقدير العلاقة الدالة بين متغير الاستجابة والمتغيرات المفسرة (التوضيحية) في محيط مضبب مع دالة خطية وبذلك سمي بالانحدار الخطي المضبب (FLR). وهناك ثلاثة اصناف للانحدار المضبب, في نماذج الانحدار التقليدي يكون هنالك عدم تأكد يكون ناتج من العشوائية (Randomness) ولكن في حالة إن عدم التأكد ناتج من الضبابية (Fuzziness) فإن النظرية الاحتمالية لا يمكن استخدامها وإنما يتم استخدام نظرية المجموعات الضبابية.

ينتج عدم التأكد في الانحدار الضبابي في حال كانت العلاقة بين المتغيرات التوضيحية والمتغير المعتمد ضبابية، أو في حال إن البيانات نفسها ضبابية هذان النوعان يقودان إلى الأنواع الآتية من الانحدار الضبابي [10].

- العلاقة بين المتغيرات ضبابية (المعلمات ضبابية) [11].
- البيانات تكون ضبابية والمعلمات قطعية وتقدر المعلمات بطريقة المربعات الصغرى الضبابية (Fuzzy Least Squares Method) (FLSM) [12] .
- البيانات مضببة والمعلمات مضببة.

وسيتم الاعتماد في دراستنا على الانموذج الضبابي حيث تكون المدخلات حقيقة والمخرجات والمعلمات ضبابية.

$$\tilde{Y} = \quad f(x،\tilde{\beta}) = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{\beta}_2\, x_2 \ldots + \tilde{\beta}_n \mathrm{x}_n + \varepsilon \qquad\qquad i = 1,2,\ldots,n \quad (3)$$

إذ إن $(\tilde{\beta})$: هي معلمات الأنموذج الضبابية ، $\tilde{Y}$: المتغير المعتمد الضبابي، $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$: المتغيرات التوضيحية (المستقلة)غير ضبابية.

3.4. تقدير انموذج الانحدار الخطي الضبابي باستخدام المربعات الصغرى المضببة (Fuzzy Least Squares) (FLS)

لتقدير انموذج انحدار خطي ضبابي يتم الاعتماد على طريقة المربعات الصغرى وبالاعتماد على الانموذج الضبابي يتكون انموذج الانحدار الضبابي الخطي للحصول على تقدير لمعلمات الانموذج الضبابي من خلال الخطوات التالية [13]:

$$\hat{\tilde{\beta}} = \mathrm{Min} \textstyle\sum_{i=1}^{n}(\tilde{y}_i - \tilde{\beta}_0 - \sum_{j=1}^{p} \tilde{\beta}_j x_{ij})^2 \qquad ,\tilde{\beta} = (\tilde{\beta}_0, \tilde{\beta}_1, \ldots, \tilde{\beta}_n) \quad (4)$$

وباشتقاق المعادلة اعلاه بالنسبة الى $\tilde{\beta}$ ومساواة المشتقة بالصفر نحصل على:

$$\hat{\tilde{\beta}}^{FLSE} \;= \arg\min \textstyle\sum_{i=1}^{n} d^2(\tilde{y}_i - \tilde{y}_i{}^*) \quad (5)$$

$$d_1 = (\tilde{y}_i،\tilde{y}_i{}^*) = \{[(y_i{}^C - y_i{}^{*C})^2 + [(y_i{}^C - x_i{}^L) - (y_i{}^{*C} - y_i{}^{*L})]^2 + [(y_i{}^C - y_i{}^U) - (y_i{}^{*C} - y_i{}^{*U})]^2]\}^{1l2} \quad (6)$$

بعدها يتم الحصول على مقدرات انموذج الانحدار الضبابي التي سيتم استخدامها في تقدير انموذج الانحدار الخطي الضبابي.

3.5. مشكلة التعدد الخطي (Multicollinearity Problem)

إن أول من لاحظ مشكلة التعدد الخطي هو الإحصائي النرويجي (Frisch) عند تحليله لبيانات السلسلة الزمنية ، إذ لاحظ أن معظم الحالات توجد فيها درجة من التداخل بين المتغيرات المستقلة، إن تحليل البيانات لسلاسل زمنية خاصة بالمتغيرات الاقتصادية يظهر أن بعض المتغيرات المستقلة قد تتطور خلال مدة زمنية معينة لتتأثر بعوامل اقتصادية أخرى مما يؤدي إلى

التداخل الخطي، علما أن ظاهرة التعدد الخطي خاصة بالنموذج الخطي المتعدد لأنها تدرس العلاقات بين المتغيرات التوضيحية، كذلك من الفروض الأساسية التي يقوم عليها أنموذج الانحدار الخطي المتعدد التي يتأثر بها انموذج الانحدار الضبابي المتعدد عدم وجود علاقة بين المتغيرات المستقلة أو بين متغير مستقل وتركيبة خطية من المتغيرات المستقلة الأخرى ، أي إن هذه الفرضية تدل على غياب التداخل الخطي المتعدد [14].

التداخل أو الارتباط الخطي المتعدد أو الارتباط الخطي المتعدد هو مصطلح مركب من (Multi) متعدد و (co) مشترك أو متداخل أو مرتبط و خطي (linearity) ، تحصل مشكلة التعدد الخطي عندما يرتبط اثنان او اكثر من المتغيرات المستقلة في علاقة خطية قوية جداً إذ يصبح من الصعب فصل اثر كل متغير عن المتغير المعتمد في الواقع التطبيقي.

3.6. طريقة الكشف عن تعدد العلاقة الخطية

توجد مقاييس عديدة تستخدم للكشف عن وجود تعدد العلاقة الخطية بين المتغيرات المستقلة وسيتم استخدام معامل تضخم التباين للكشف عن وجود مشكلة التعدد الخطي في البيانات الضبابية.

- معامل تضخم التباين(VIF) (Variance Inflation Factor)

يستخدم عامل تضخم التباين كمعيار للكشف عن التداخل الخطي وتحديد المتغير التوضيحي المسؤول عن ذلك.

اقترح هذا المقياس من الباحثين Farrar & Glauber(1967) وقد أطلق عليه (بعوامل تضخم التباين Variance Inflation Factors ويرمز له بالرمز (VIF) وعوامل تضخم التباين يمكن التعبير عنها بصيغة رياضية تأخذ الشكل الآتي[15]:

$$VIF_j = \frac{1}{1-R_j^2} \qquad j = 1,2,3,\dots,p \qquad (7)$$

p : تمثل عدد المتغيرات التوضيحية.

R_j^2 : معامل التحديد لنموذج انحدار المتغير المستقل x_j على بقية المتغيرات التوضيحية وأن $0 \le R_j^2 \le 1$.

$$R_j^2 = \frac{\left\| \hat{x}_{ij} - \bar{x}_j \right\|^2}{\sum_{i=1}^{n} \left\| x_{ij} - \bar{x}_j \right\|^2} \qquad (8)$$

من اجل تطوير (VIF) لتطبيقه مع انموذج الانحدار الضبابي للمتغيرات الضبابية يتم حسابه كما في الصيغة التالية:

$$R_{Gj}^2 = \frac{\sum_{i=1}^{n} d^2(\hat{x}_{ij}, \bar{x}_j)}{\sum_{i=1}^{n} d^2(x_{ij}, \bar{x}_j)} \qquad (9)$$

$$VIF_{Gj} = \frac{1}{1-R_{Gj}^2} \qquad (10)$$

وقد اقترح (Gunst & Mason ،1980)، إنه إذا كان : $VIF_j > 5$ فذلك يعمي احتمال وجود مشكلة التعدد الخطي بين المتغيرات التوضيحية، وهذا سيكون سبباً كافيا لإهمال المتغير x_jمن التحليل أو استخدام طريقة أخرى بديلة عن المربعات الصغرى الضبابية في التقدير، أما إذا كان المتغير x_j مستقلاً عن بقية المتغيرات التنبؤية الأخرى فإن (R^2= 0) وبذلك تكون قيمة (VIF= 1) ويعني هذا عدم وجود مشكلة التعدد الخطي بين المتغيرات التنبؤية الذي سيتم اعتمادة كمقياس للكشف عن وجود تعدد خطي في البيانات الضبابية [16].

3.6.1. طريقة انحدار لاسو الضبابي Fuzzy LASSO Regression

قدمت هذه الطريقة لأول مرة في الادب الجيوفيزياوي في عام (1982) ، ثم اعيد اكتشافها بشكل مستقل في عام (1996) من الباحث روبرت تيبشراني Ropert Tibshirani إذ قام بصياغة هذه الطريقة وقدم الكثير من الافكار حول ادائها "[17].

ومصطلح لاسو يمثل الحروف الأولى لمفهوم(Least Absolute Shrinkage and Selection Operator) ، وهي دالة جزاء لأنموذج الانحدار الخطي وهي طريقة لتقدير معلمات أنموذج الانحدار وكذلك لاختيار وتنظيم المتغيرات الداخلة في الأنموذج لزيادة الثقة التوضيحية لنماذج الانحدار المستخدمة في تحليل الظاهرة محل الدراسة من خلال عمليات ملاءمة الأنموذج لاختيار مجموعة فرعية من المتغيرات المشتركة في الأنموذج النهائي بدلا من استخدامها كلها، ففي طريقة لاسو يتم تصغير مجموع مربعات الاخطاء العشوائية لأعلى حد مع مجموع القيم المطلقة لمعاملات انموذج الانحدار، صممت طريقة LASSO اصلا لنماذج المربعات الصغرى Least squares models ، كما في الانحدار الضبابي الخطي انه لا يجب ان تكون تقديرات معامل لاسو وحيدة اذا كانت المتغيرات التوضيحية تعاني من مشكلة التعدد الخطي، وان طريقة لاسو لها القدرة على اختيار مجموعة جزئية تعتمد على صيغة القيد، وعلى الرغم من انه تم تعريف لاسو للمربعات الصغرى الضبابية الا انه يمكن بسهولة استعمال طريقة لاسو في مجموعة واسعة في كثير من النماذج الاحصائية ويمكن ان تستعمل لاسو في كثير من المجالات مثل الهندسة والاحصاءات البيزية.

3.6.2. مبدأ انحدار لاسو LASSO Regression Principle

ان مبدأ طريقة انحدار لاسو هو تصغير مجموع مربعات البواقي على وفق قيد يمثل المجموع المطلق للمعاملات التي تكون أصغر من ثابت معين، اذ يضع لاسو قيداً على مجموع القيم المطلقة لمعاملات Coefficients الأنموذج ، إذ يجب ان يكون المجموع اقل من قيمة ثابتة (الحد الأعلى) من أجل القيام بذلك تطبق لاسو عملية التقلص (تنظيم)، اذ انها تقوم بعمل جزاء لمعاملات الانحدار وتقلص بعضها إلى الصفر، واثناء عملية اختيار المتغيرات سيتم تحديد المتغيرات التي تكون معاملاتها غير صفرية بعد عملية التقلص Shrinkage وستكون جزءاً من الأنموذج والهدف من هذه العملية هو تقليل خطأ التنبؤ.

ان الفائدة من معلمة طريقة لاسو λ التي تمثل معلمة تتحكم في قوة الجزاء (الانكماش) على معلمات الانحدار.

فعندما تكون معلمة الضبط كبيرة بشكل كافٍ سوف تصبح المعاملات مساوية للصفر، وهذه مفيدة في تقليل المتغيرات في الأنموذج، اي بمعنى كلما كانت قيمة معلمة الضبط كبيرة معناه عدد أكبر من المعاملات المساوية للصفر. وإذا كانت معلمة الضبط مساوية للصفر، سنحصل على انحدار المربعات الصغرى الضبابية [18].

3.6.3. مميزات انحدار لاسو الضبابي LASSO Regression Advantages

هناك العديد من المميزات في استخدام طريقة لاسو ومنها الآتي:

- يمكن أن توفر طريقة لاسو دقة تنبؤية جيدة للغاية لان تقليص وازالة المتغيرات يمكن أن يخفض التباين دون زيادة كبيرة في التحيز، وهذا مفيد بشكل خاص عندما يكون لدينا عدد قليل من المشاهدات وعدد كبير من المتغيرات.
- تساعد طريقة لاسو على زيادة امكانية تفسير الأنموذج من خلال حذف المتغيرات غير ذات الصلة التي لا ترتبط بمتغير الاستجابة.

فبذلك تعدّ طريقة لاسو طريقة لاختيار وتنظيم المتغيرات الداخلة في نموذج الانحدار [19].

3.6.4. صيغة انحدار لاسو LASSO Regression Formula

يتم تقدير معلمات انحدار لاسو طبقا لمبدأ المربعات الصغرى الضبابية وكالآتي:

$$\hat{\tilde{\beta}}^{Lasso} = \sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \lambda\sum_{i=1}^{p}|\beta_j| \tag{11}$$

إذ إن $||\beta||_1 \leq t$

t : تمثل معلمة حرة تحدد مسبقاً التي تعين مقدار التسوية (التقليص).

λ هي المعلمة التي تتحكم في قوة الجزاء (الانكماش) على معلمات الانحدار.

بالاعتماد على انموذج الانحدار الضبابي الذي يمثل مدخلا تغير ضبابية ام الملمات والمخرجات ضبابية وباستخدام دالة مثلثية.

$$\tilde{y}_i = \oplus_{j=1}^{p}\left(\tilde{\beta}_j \otimes x_{ij}\right) \oplus \tilde{\varepsilon}_i \tag{12}$$

$$\tilde{y}^*{}_i = \oplus_{j=1}^{p}\left(\tilde{\beta}_j \otimes x_{ij}\right)$$

وفي حالة المتغير المعتمد ياخذ الرقم الضبابي المثلي اي ان :

$$y_i^* = \sum_{j=1}^{n} x_{ij}\beta_j \qquad \tilde{y}^*{}_i = (\tilde{y}^*{}_i، l_{y_i^*}، r_{y_i^*})_T \qquad \tilde{y}_i = (y_i; l_{y_i}، r_{y_i})_T$$

$$l_{y_i^*} = \sum_{j=1}^{n}(s_{ij}x_{ij}l_{\beta_j} - (1-s_{ij})x_{ij}r_{\beta_j})$$

$$r_{y_i^*} = \sum_{j=1}^{n}\left(s_{ij}x_{ij}r_{\beta_j} - (1-s_{ij})x_{ij}l_{\beta_j}\right)$$

$$j = 1,2,\dots,p, \qquad \text{And } s_{ij}=\text{I}(x_{ij} \geq 0) \qquad for\ \ i = 1,2,\dots,n$$

وان معلمات الانموذج $\tilde{\beta}_0، \tilde{\beta}_1، \dots ، \tilde{\beta}_p$

$$\tilde{\beta} = (\beta، l_\beta، r_\beta)_T$$

$$r_\beta = (r_{\beta_0}\ r_{\beta_1} \cdots\cdots r_{\beta_p})، \qquad l_\beta = (l_{\beta_0}\ l_{\beta_1} \cdots\cdots l_{\beta_p}) \qquad \tilde{\beta} = (\beta_0\ \ \beta_1\ \cdots\cdots \beta_p)$$

وبتقدير معلمات انموذج انحدار لاسو الضبابي تصبح كالتالي:

$$\tilde{\hat{\beta}}(\emptyset_1، \emptyset_2، \emptyset_3)=\arg\ min_{\beta\in(F(R))^{p+1}} \quad ,L_{\beta\in(F(R^+))^{p+1}}\ , \quad R_{\beta\in(F(R^+))^{p+1}} \sum_{i=1}^{n} D_2^2(\tilde{y}_i، y_i^*) \tag{13}$$

إذ إن $\emptyset_1، \emptyset_2، \emptyset_{31}$ هي ثلاثة عوامل لضبط الانموذج [20].

على وفق القيود التالية:

$\sum_{j=1}^{p}|\beta_j| \leq \emptyset_1$

$\sum_{j=1}^{p}|\beta_j| \leq \emptyset_2$

$\sum_{j=1}^{p}|\beta_j| \leq \emptyset_3$

4. الجانب التجريبي

تم استعمال لغة البرمجة الإحصائية R 4.2.1 لكتابة برنامج المحاكاة وهي تتضمن أربع مراحل أساسية لتقدير أنموذج الانحدار المضبب، وكما يأتي:

4.1. المرحلة الأولى: مرحلة تحديد القيم الافتراضية

إذ يتم في هذه المرحلة اختيار القيم الافتراضية للمعلمات، وكما يأتي:

- تم اختيار 6 متغيرات توضيحية واختيار قيم افتراضية للمعلمات.
- اختيرت القيم الافتراضية المختلفة للانحراف المعيار للأخطاء العشوائية ($\sigma = 0.5$، 2).
- تم اختيار قيم مختلفة لمعلمة الارتباط بين المتغيرات التوضيحية ($\rho = 0.8$، 0.9، 0.95، 0.99).
- تم اختيار أحجام مختلفة للعينات (n=20, 40, 80, 160).
- تم تكرار كل تجربة 1000 مرة.

4.2. المرحلة الثانية: توليد البيانات

وهي مرحلة مهمة جداً لاعتماد الخطوات التي تليها عليها، إذ يتم فيها توليد المتغيرات التوضيحية التي بينها ارتباط يساوي ρ، والخطأ العشوائي الذي يتبع التوزيع الطبيعي بمتوسط صفر، وتباين σ^2، وبعد ذلك إيجاد قيم المتغير التابع بناءً على القيمة المولدة لكل من المتغيرات التوضيحية والخطأ العشوائي.

4.3. المرحلة الثالثة: التقدير

يتم في هذه المرحلة إجراء عملية التقدير لمعلمات الانحدار وذلك باستعمال طرائق التقدير محل اهتمام البحث، وكما يأتي:

- طريقة انحدار المربعات الصغرى الضبابية الضبابي (FLS).
- طريقة انحدار لاسو الضبابي (FLR).

4.4. المرحلة الرابعة: مرحلة المقارنة بين الطرائق

لغرض المقارنة بين طرائق التقدير المختلفة للمعلمات وإيجاد أفضل المقدرات تم استعمال معيارا متوسطا لمربعات الخطأ Mean Squared Error (MSE) وكما يأتي:

$$MSE = \frac{1}{R}\sum_{I=1}^{R}\left[\left(\hat{\beta}_l - \beta\right)'\left(\hat{\beta}_l - \beta\right) + \left(\hat{\beta}_m - \beta\right)'\left(\hat{\beta}_m - \beta\right) + \left(\hat{\beta}_r - \beta\right)'\left(\hat{\beta}_r - \beta\right)\right] \quad (14)$$

R: تمثل عدد تكرار التجربة التي تساوي 1000 مرة

وقد تم الحصول على النتائج كما في الجدول (1), من خلال جدول (1) نلحظ ان اقل قيمة MSE هي التي تمثل الطريقة الافضل ومن خلال المقارنة تبين ان طريقة انحدار لاسو الضبابي تغلبت على طريقة تقدير المربعات الصغرى الضبابية.

ويتضح من الجدول (1) و(2) و(3) و(4) ما يأتي:

- أفضلية طريقة انحدار لاسو الضبابي (FLR) على طريقة انحدار المربعات الصغرى الضبابية (FLS) وفي الحالات جميعاً.
- نقصان قيمة MSE بزيادة حجم العينة، مما يدل على امتلاك مقدرات الطريقتين خاصية الاتساق.
- زيادة قيمة MSE بزيادة الانحراف المعياري.
- زيادة قيمة MSE بزيادة معامل الارتباط بين المتغيرات التوضيحية.
- تقارب سلوك الطريقتين عند زيادة حجم العينة.

ومن خلال الشكل (2) نلاحظ انه كلما كبر حجم العينة كلما قلت قيمة MSE مما يدل على اتساق طريقة المربعات الصغرى الضبابية مع طريقة انحدار لاسو الضبابي وكلما زاد حجم العينة تقارب النتائج.

جدول (1) قيم MSE للطرائق المختلفة عندما (ρ=0.8)

σ	σ=0.5			σ=1			σ=2		
n	FLS	FLR	الافضل	FLS	FLR	الأفضل	FLS	FLR	الافضل
n=20	1.8141	1.7647	FLR	7.41099	6.95203	FLR	30.13145	25.87478	FLR
n=40	0.79141	0.78404	FLR	3.07079	2.98895	FLR	12.48215	11.70657	FLR
n=80	0.41254	0.41079	FLR	1.65997	1.64197	FLR	6.57295	6.37581	FLR
n=160	0.25887	0.25849	FLR	1.03236	1.0275	FLR	4.18727	4.13478	FLR

جدول (2) قيم MSE للطرائق المختلفة عندما (ρ=0.9)

σ	σ=0.5		σ=1		σ=2	
n	FLS	FLR	FLS	FLR	FLS	FLR
n=20	3.5102	3.3559	13.775	12.298	55.825	45.435
n=40	1.3925	1.369	5.517	5.2704	22	19.621
n=80	0.6952	0.6898	2.6378	2.5838	11.194	10.564
n=160	0.4014	0.3996	1.5404	1.524	6.3419	6.162

جدول (3) قيم MSE للطرائق المختلفة عندما (ρ=0.95)

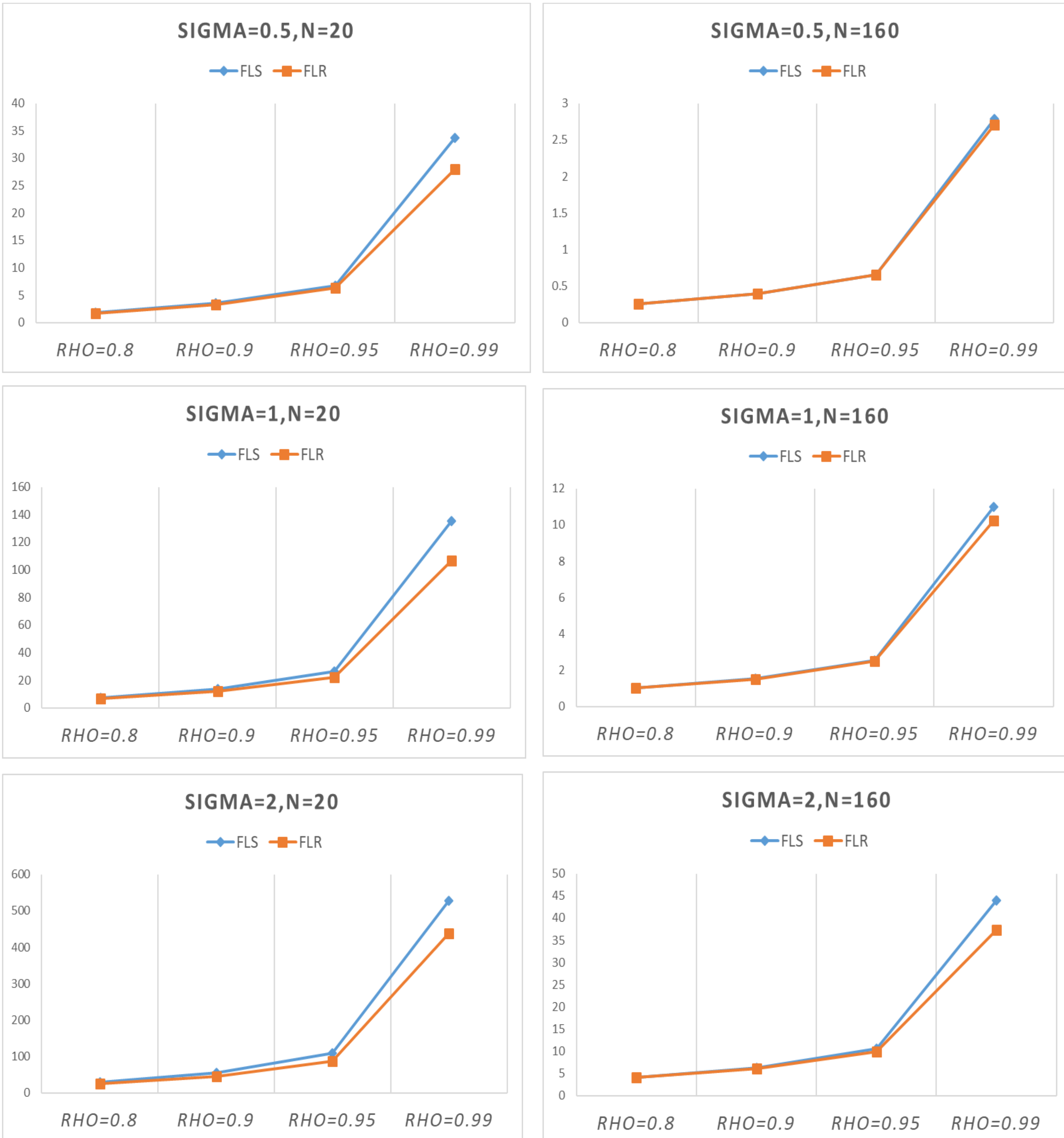
σ	σ=0.5		σ=1		σ=2	
n	FLS	FLR	FLS	FLR	FLS	FLR
n=20	6.75908	6.29646	26.493	22.362	108.92	86.748
n=40	2.57606	2.50462	10.436	9.6303	43.32	36.824
n=80	1.27784	1.25922	5.0399	4.8361	19.397	17.53
n=160	0.66073	0.65554	2.5439	2.4967	10.547	9.9929

جدول (4) قيم MSE للطرائق المختلفة عندما (ρ=0.99)

σ	σ=0.5		σ=1		σ=2	
n	FLS	FLR	FLS	FLR	FLS	FLR
n=20	33.674	27.96	135.42	106.93	528.32	437.91
n=40	12.73	11.558	50.62	42.035	194.71	156.71
n=80	5.4321	5.1482	22.865	20.291	88.069	71.297
n=160	2.7909	2.7113	11.002	10.223	43.953	37.369

5. الاستنتاجات

- من خلال استخدام المحاكاة وبالاعتماد على معيار متوسط مربعات الخطأ تبين أن طريقة انحدار لاسو الضبابي أفضل من طريقة المربعات الصغرى الضبابية.
- كلما زاد قيمة الارتباط كلما زادة كفاءة طريقة لاسو.
- تقارب سلوك الطريقتين عند زيادة حجم العينة كما في الشكل (2).



شكل (2) رسم قيم MSE بزيادة قيمة معامل الارتباط

References

[1] Mohammed, Mohammed J. "Semi Parametric Logistic Regression Model with the Outputs Representing Trapezoidal Intuitionistic Fuzzy Number." Journal of Economics and Administrative Sciences 28.133 (2022): 70-81, https://doi.org/10.33095/jeas.v28i133.2350.

[2] Al-Ghannam, H. Attia, H. Muhammad, "A study in fuzzy variables and fuzzy multiple regression" Tikrit Journal of Administration and Economics Sciences, Volume 5, Issue 14, P 166-180, 2009.

[3] Mohammed Jasim Mohammed, Ahmad Farouk Abbas, "A Comparisons between Memberships function And Entropy Function in Fuzzy Adaptive Linear Regression", Journal of Economics and Administrative Sciences, 24(103), 453. https://doi.org/10.33095/jeas.v24i103.123.

[4] Hesamian, Gholamreza, Mohammad Ghasem Akbari. "Fuzzy Lasso regression model with exact explanatory variables and fuzzy responses." International Journal of Approximate Reasoning, 115, 290-300, 2019, https://doi.org/10.1016/j.ijar.2019.10.007.

[5] Zimmermann, H-J., Lotfi Asker Zadeh, and Brian R. Gaines. "Fuzzy sets and decision analysis", studies in the management sciences, 1984.

[6] Effati, S., H. Sadoghi, and Z. Saberi. "First Joint Congress on Fuzzy and Intelligent Systems" Ferdowsi University of Mashhad, Iran 29-31 Aug 2007.
[7] Hadi Basirzadeh, "An Approach for Solving Fuzzy Transportation problem", Applied Mathematical sciences, Vo1.5, No 32, pp 1549-1566, 2011.
[8] Gebray, Gebru, and B. Krishna Reddy. "Fuzzy set field and fuzzy metric", Advances in Fuzzy Systems, 201, https://doi.org/10.1155/2014/968405.
[9] Nareshkumar, S., and S. Ghuru. "Solving Fuzzy Transportation Problem Using Symmetric Triangular Fuzzy Number." International Journal of Advanced Research in mathematics and applications, 1(1), 74-83, 2014.
[10] Taheri, S. Mahmoud. "Trends in fuzzy statistics." Austrian journal of statistics, 32, 3, 239-257, 2003.
[11] Nassif, A. B., Azzeh, M., Idri, A., & Abran, A. "Software development effort estimation using regression fuzzy models." Computational intelligence and neuroscience, 2019, https://doi.org/10.1155/2019/8367214.
[12] Gorgees, Hazim Mansoor, and Fatimah Assim Mahdi. "The Comparison Between Different Approaches to Overcome the Multicollinearity Problem in Linear Regression Models." Ibn AL-Haitham Journal For Pure and Applied Sciences, 31(1), 212-221, 2018, https://doi.org/10.30526/31.1.1841.
[13] Choi, Seung Hoe, and Jin Hee Yoon. "General fuzzy regression using least squares method." International Journal of Systems Science, 41(5), 477-485, 2010, https://doi.org/10.1080/00207720902774813.
[14] Kalnins, Arturs. "Multicollinearity: How common factors cause Type 1 errors in multivariate regression", Strategic Management Journal, 39(8), 2362-2385, 2018, https://doi.org/10.1002/smj.2783.
[15] E. Farrar Donald, R. Glauber Robert, "Multicollinearity in regression analysis: The problem revisited", The Review of Economics and Statistics, Vol. 49, No. 1, pp. 92-107, 1967.
[16] Gunst, R. F. & Mason, R. L, "Regression Analysis and its Application", Marcel Dekker Inc. New York, U.S.A, p.114, 1980.
[17] R. Tibshirani, "Regression penalized and selection via the LASSO". J. R. Stat. Soc. B, 58, 267–288, 1996.
[18] Zeng, Wenyi, Qilei Feng, and Junhong Li. "Fuzzy least absolute linear regression", Applied Soft Computing, 52, 1009-101, 2017, https://doi.org/10.1016/j.asoc.2016.09.029.
[19] Chachi, Jalal. "A weighted least squares fuzzy regression for crisp input-fuzzy output data", IEEE Transactions on Fuzzy Systems, 27(4), 739-748, 2018, https://doi.org/10.1109/TFUZZ.2018.2868554.
[20] Chukhrova, Nataliya, and Arne Johannssen. "Fuzzy regression analysis: systematic review and bibliography", Applied Soft Computing, 84, 105708, 2019, https://doi.org/10.1016/j.asoc.2019.105708.