Al-Ghary Journal of Economic and Administrative Sciences Vol. 20 (No.2) 2024 PP. 607-622

مجلة الغري للعلوم الاقتصادية والادارية

مجلد (20) عدد (2) 2024

الانحدار الكمي الثنائي البيزي باستخدام دالة جزاء معكوس لاسو (Lasso) (مع تطبيق عملي)

Bayesian Binary reciprocal LASSO quantile regression (with practical application)

م . محمد طالب خنجر
جامعة الكوفة/كلية علوم الحاسوب والرياضيات
Mohammedt.azooz@uokufa.edu.iq

أ. د . احمد نعيم فليح
جامعة القادسية / كلية الادارة والاقتصاد
ahmad.flaih@qu.edu.iq

الملخص

في هذه الورقة البحثية تم دراسة الإنحدار الكمي لما له من اهمية كبيرة في التطبيق في مجالات العلوم المختلفة . هنالك العديد من الباحثين الذين تناولوا موضوع الانحدار الكمي مع البيانات الثنائية اي عندما يأخذ متغير الاستجابة (y) قيمتان فقط اما (0) او (1) باستخدام الطرق الكلاسيكية وكذلك الطرق البيزية . في هذه الدراسة الباحث يحاول بناء نموذج لتحليل البيانات الثنائية باستخدام طريقة الانحدار الكمي (Quintile Regression) اذ تم اعتماد الاسلوب البيزي وهو من الاساليب التي اخذت صدى واسعاً في الآونة الأخيرة لما لها من الدقة وخصوصاً في العينات الصغيرة .بالإضافة الى ذلك فقد قيد الباحث الدراسة بإستخدام دوال الجزاء المعكوسة لإختيار المتغيرات التوضيحية (explanatory variables) التي لها تأثير على المتغير المعتمد (dependent variable) فعلاً واستبعاد المتغيرات غير المؤثرة باستخدام طرق اختيار المتغيرات . **الكلمات المفتاحية** : الانحدار الكمي ــ اختيار المتغيرات ــ الانحدار الكمي الثنائي

Abstract

In this research paper, quantitative regression was studied because of its great importance in application in various fields of science. There are many researchers who have dealt with the topic of quantile regression with binary data, that is, when the response variable (y) takes only two values, either (0) or (1), using classical methods as well as Bayesian methods. In this study, the researcher is trying to build a model for analyzing binary data using the Quintile Regression method. The Bayesian method was adopted, which is one of the methods that has gained wide resonance recently because of its accuracy, especially in small samples. In addition, the researcher restricted the study by using

inverse penalty functions to select explanatory variables that actually have an effect on the dependent variable, and to exclude ineffective variables using variable selection methods..

1- المقدمة Introduction

يعتبرالانحدار الكمي (Quantile Regression(QR) هو احد نماذج الانحدار الذي يقيس اثر المتغير المستقل على المتغير المعتمد بخطوط انحدار غير منتهية عند تقسيمات معينه محصورة مابين (0,1) ويعتبر هذا النوع من نماذج الانحدار الحصينه لانه لايتأثر بالقيم الشاذه او القيم المتطرفة وذلك لمرور احد خطوط الانحدار بالقرب من هذه القيم [5 ,6].

ويمكن التعبير عن نموذج الانحدار الكمي كما في المعادلة التالية[5] :

$$y_i = x_i' \beta_p + \varepsilon_i$$

y_i : قيمة المشاهدة الحقيقية للمتغير الذي ترتيبه i

x_i : متجه من درجة (n×p)

وان : β_p تمثل متجه المعلمات عند الكمية (p) , $0 < p < 1$

β_p : متجه معالم النموذج من درجة (p×1)

وان عملية تقدير معالم نموذج الانحدار الكمي β_p يكون بتصغير دالة الخسارة (Loss function or check function) التالية :

$$\min_{\beta_p} \sum_{i=1}^{n} \rho_p (y_i - x_i' \beta_p)$$

حيث ان :

$$\rho_p (y_i - x_i' \beta_p) = \begin{cases} p(y_i - x_i' \beta_p) & if \quad y_i \geq x_i' \beta_p \\ -(1-p)(y_i - x_i' \beta_p) & if \quad y_i < x_i' \beta_p \end{cases}$$

في دراستنا هذه سوف ندرس اسلوب بيز في تقدير نموذج الانحدار الكمي (Quantile Regression modle في حالة ان متغير الاستجابة يكون ثنائي (Binary)الصفة وسوف

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (20) عدد (2) 2024

نعتمد على دوال جزاء معكوسة (reciprocal) تدخل في معادلة انحدار معالم النموذج المدروس . ويعتبر التعشيق بين نظرية بيز ونماذج الانحدار من الاساليب الاحصائية الواسعة الانتشار في التحليلات الاحصائية . ويمكن ان يعبر عنه بالصيغة الرياضية لنموذج الانحدار الكمي الثنائي وكما يلي :

$$y_i^* = x_i'\beta_p + \varepsilon_i$$

$$y_i = \begin{cases} 1 \quad if & y_i^* \geq 0 \\ 0 \quad if & \text{otherwise} \end{cases}$$

y_i : تمثل متغير الاستجابة المشاهدة والتي تحدد قيمته من قبل المتغير الاستجابة الكامنة (الغير مشاهدة) .

2- هدف البحث

الهدف من البحث هو اقتراح صيغة تقدير جديدة من خلال توظيف توزيع مسبق مقترح للوصول الى افضل المقدرات لاختيار اهم المتغيرات بطريقة تقدير ذات نتائج عالية الدقة وبناء خوارزمية سهلة وجذابة في اختصار وقت التقدير .

3- الجانب النظري

1-3 الانحدار الكمي (Quantile Regression)

الانحدار الكمي هو طريقة من الطرائق المهمة في تحليل الانحدار وهو يعتبر امتداد للانحدار الخطي في حالة عدم تحقق شروط الإنحدار الخطي .

وبفرض ان نموذج الانحدار الخطي الاعتيادي والمعطى حسب الصيغة الرياضية التالية[9] :

$$y = X\beta + \varepsilon$$

y : يمثل متجه الاستجابات من الدرجة $(n \times 1)$.

X : تمثل مصفوفة من الدرجة $(n \times p)$.

ε: يمثل متجه الاخطاء العشوائية من الدرجة $(n \times 1)$.

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

ان المقدر الكلاسيكي في الانحدار الخطي هو مقدر المربعات الصغرى الاعتيادية (OLS) في حالة توفر شروط هذه الطريقة حيث ان:

$$\{\hat{\beta}_{ols} = (X'X)^{-1}X'Y\}$$

والتي نحصل عليها من خلال تصغير مجموع مربعات الاخطاء العشوائية اي ان:

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (20) عدد (2) 2024

$$RSS = (Y - X\beta)'(Y - X\beta)$$

ومن المعروف او الملاحظ بان مقدر (OLS) يكون غير مستقر في حالة وجود مشكلة التعدد الخطي (Multicollinearity), وايضا عندما تكون $(p > n)$ فانه يترتب على ذلك انتاج مقدراً غير وحيد لان رتبة مصفوفة الانحدار تكون اقل من الرتبة الكاملة (Less than full mark) وان طريقة المربعات الصغرى (Least square ; LS) تكون حساسة في حالة وجود بيانات شاذه او متطرفة او عندما يكون الخطأ العشوائي لايتوزع توزيعاً طبيعياً وفي هذه الحالة لايمكن الضمان بتقديراتها او الاعتماد عليها بالتنبؤ ولهذا يكون الانحدار الكمي الخطي كإمتداد أو ملحق مفيد للانحدارالمتوسط الخطي القياسي[4] حيث انها لاتفترض بتوزيع الخطأ العشوائي توزيعاً طبيعياً على عكس النماذج الخطية وايضا لاتتأثر بالبيانات الشاذة او المتطرفة(extreme observations) لان هناك خطوط انحدار تمر بالقرب من هذه القيم , وبالتالي يمكن القول بانه يوفر نموذجاً احصائياً اكثر شمولا من النماذج الكلاسيكية(OLS), لذا فانها توفرحلول اكثر حصانة في كثير من التطبيقات الحقيقة, ولهذه الاسباب فان نموذج الانحدار الكمي يعتبر من النماذج الحصينة . وبفرض ان $\{(x_i, y_i); i = 1,2, \ldots\ldots, n\}$ حيث ان y_i تشير الى المتغير المعتمد (response variable) وان $(x_i \in R^p)$ تمثل متجه المتغيرات التفسيرية (المتغيرات المستقلة) . لذا فان نموذج الانحدار الكمي يعبر عنه كما في المعادلة التالية[5] :

$$y_i = x_i'\beta_p + \varepsilon_i \quad ; \quad (i = 1,2, \ldots\ldots, n) \ldots\ldots\ldots\ldots\ldots.\ldots\ldots (3-1)$$

وان : β_p تمثل متجه المعلمات عند الكمية (p) , $0 < p < 1$

3 - 2 استخدام طريقة معكوس (Lasso) في الانحدار الخطي :

تم تقديم بعض الاقتراحات والتحسينات من قبل مجموعة من الباحثين على مرالسنين[13,10,12,3,2] (Eforn et al.2004;Fan and Li,2001;Meinshausen,2007;Radchenko and James,2008;song and Laing,2015, among others) على طريقة تقدير (Lasso)المقترحة من قبل الباحث[14] Tibshirani ; 1996 في تقدير نماذج الانحدار الخطي . وفي السنوات الاخيرة قدم كل من الباحثان[12] (Song and Laing, 2015) بعض التطورات او التحسينات على طريقة التقدير (Lasso) حيث سميت الصيغة الجديدة بطريقة معكوس لاسو (the reciprocal Lasso) والتي تختصر بـ (rLasso) , وان تقدير معالم نموذج الانحدار الخطي باستخدام معكوس لاسو (rLasso) يكون بتصغير المعادلة الرياضية التالية :

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (20) عدد (2) 2024

$$\min_{\beta}\sum_{i=1}^{n}(y_i - x_i'\beta)^2 + \lambda\sum_{j=1}^{k}\frac{1}{|\beta_j|}\ \ I(\beta_j \neq 0) \ldots\ldots\ldots\ldots\ldots(3-2)$$

وان لدالة جزاء معكوس لاسو (rLasso)بعض الخصائص منها[11]

- دالة جزاء (rLasso) متناقصة في الفترة $(0, \infty)$.
- الدالة غير مستمرة عند الصفر .
- الدالة تقترب من المالانهاية عندما معاملات الانحدار تقترب من الصفر .
- معاملات الانحدار الكبيرة تحصل على جزاءات (penalties) صغيرة والعكس بالعكس
- دالة جزاء معكوس لاسو (rLasso) تميل الى اختيار نموذج اكثر شحاً (parsimonious) يحتوي على خطأ تنبؤي مشابه او افضل بالمقارنة مع دالة جزاء Lasso.

ولتقدير معالم نموذج الانحدار الكمي عند استخدام معكوس لاسو يكون من خلال تصغير المعادلة التالية وكما يلي[12] :

$$Q(B) = \min_{\beta}\sum_{i=1}^{n}\rho_p\left(y_i - x_i'\beta_p\right) + \lambda\sum_{j=1}^{k}\frac{1}{|\beta_j|}\ldots\ldots\ldots\ldots(3-3)$$

$(\lambda > 0)$ تمثل معلمة الضبط (Regularization)

$p \in (\,0\,,1\,)$

y_i : يمثل تمثل قيمة المشاهدة الحقيقية للمتغير الذي ترتيبه i

x_i : متجه من درجة (n×p)

β_p : متجه معالم النموذج من درجة (p×1)

3-3 الانحدار الكمي البيزي الثنائي باستخدام معكوس لاسو (Binary reciprocal lasso quantile regression) :

في هذا الجانب نقوم بتوسيع الانحدار الكمي البيزي الذي اشار اليه الباحث[8] (Li et al.(2010)) الى الانحدار الكمي البيزي باستخدام دالة جزاء معكوس لاسو (reciprocal lasso) لبيانات الاستجابة ثنائية التفرع . نعتقد ان هذا النهج ذو أهمية , لانه من خلال القيام

بذلك فاننا نستفيد معاً من الخصائص المرغوبة للانحدار الكمي الثنائي بالاضافة الى الخصائص الممتازة لدالة جزاء معكوس لاسو (rlasso) , وايضا يعتبر الانحدار الكمي الثنائي أداة مناسبة لتصنيف العينات التي تنتمي الى واحدة من فئتين مختلفتين[7 , 1] . وان الصيغة الرياضية لنموذج الانحدار الكمي الثنائي هي :

$$y_i^* = x_i'\beta_p + \varepsilon_i$$

$$y_i = \begin{cases} 1 \quad if & \quad\quad\quad\quad y_i^* \geq 0 \\ 0 \quad if & \quad\quad\quad\quad \text{otherwise} \end{cases}$$

حيث ان y_i تمثل متغير الاستجابة المشاهدة والتي تحدد قيمته من قبل المتغير الاستجابة الكامنة (الغير مشاهدة) .

وان تقديرمعلمات نموذج الانحدار الكمي باستخدام معكوس لاسو (rlasso) يكون حسب المعادلة المرقمة $(3-3)$.

ومن الميزات المهمة للانحدار الكمي هو انه قادر على استيعاب توزيع الاخطاء العشوائية غير الطبيعية .

اقترح الباحث[11] Mallick وآخرون في سنة 2021 مقترحاً لتمثيل التوزيع السابق (prior) للمعلمة β , حيث كان هذا المقترح يفترض ان توزيع معكوس لابلاس للمعلمة β يمكن تمثيله بالشكل الآتي :

$$\frac{\lambda}{2\beta^2} e^{-\frac{\lambda}{|\beta|}} = \int_0^\infty \frac{\eta}{2\beta^2} I\{|\beta| > \eta\} \frac{\lambda^2}{\Gamma 2} \eta^{-2-1} e^{-\frac{\lambda}{\eta}}\, d\eta \ldots\ldots\ldots (3-4)$$

في هذه البحث ولتسهيل العمليات الحسابية لخوارزمية Gibbs sampler تم اجراء التحويل التالي على التمثيل المذكور في المعادلة $(3-4)$:

لنفرض ان

$$Z = \frac{\lambda}{\eta} \implies \eta = \frac{\lambda}{Z}$$

$$\implies d\eta = \lambda\left(\frac{-1}{Z^2}\right) dz$$

فاذا كان $\eta = 0 \Longrightarrow Z = \infty$, واذا كانت $\eta = \infty \Longrightarrow Z = 0$

وبالتالي يمكن قلب حدود التكامل والغاء الاشارة السالبة في المشتقة أعلاه ونحصل على ما يلي :

$$\frac{\lambda}{2\beta^2}e^{-\frac{\lambda}{|\beta|}} = \int_0^\infty \frac{\frac{\lambda}{Z}}{2\beta^2} I\left\{|\beta| > \frac{\lambda}{Z}\right\} . \frac{\lambda^2}{\Gamma(2)}\left(\frac{\lambda}{Z}\right)^{-2-1} . e^{-Z}\, \lambda \left(\frac{-1}{Z^2}\right) dz$$

$$= \int_0^\infty \frac{\lambda}{2\beta^2} I\left\{|\beta| > \frac{\lambda}{Z}\right\} . e^{-Z} dz$$

حيث ان:

$$Z \sim standard\ exponential$$

$$\beta \sim Inverse\ uniform\ \left(-\frac{1}{\lambda}, \frac{1}{\lambda}\right) \ldots\ldots..(3-5)$$

وهذا يعني ان :

$$\beta \sim Inverse\ uniform\ (-\frac{1}{\lambda}\ , \frac{1}{\lambda}\ )$$

يمتلك دالة الكثافة الاحتمالية الآتية :

$$f(\beta|\lambda) = \frac{1}{\beta^2}\ Uniform\ \left(-\frac{1}{\lambda}, \frac{1}{\lambda}\right)$$

$$= \frac{1}{\beta^2} . \frac{1}{1/_\lambda + 1/_\lambda} = \frac{1}{\beta^2}\ \frac{1}{2/_\lambda} = \frac{\lambda}{2\beta^2}$$

وهو توزيع معكوس التوزيع المنتظم او يدعى بتوزيع باريتو المضاعف (double Pareto). وبالتالي من العلاقة (5 – 3) يمكن ملاحظة ان توزيع معكوس التوزيع المنتظم هو عبارة عن دالة التوزيع المنتظم $[-\frac{1}{\lambda}, \frac{1}{\lambda}]$ مضروب بالمقدار $\frac{1}{\beta^2}$, وبالتالي يمكن ربط النموذج الهرمي للتوزيعات المسبقة التالي (معكوس لاسو) مع النموذج الهرمي للتوزيعات المسبقة بوجود التوزيع المسبق لتوزيع لابلاس (طريقة انحدار لاسو)[9] (Mallick and Yi, 2014).

<u>4-3 النموذج الهرمي البيزي المقترح(Bayesian hierarchical model) للتوزيعات السابقة :</u>

يمكن افتراض النموذج الهرمي للتوزيعات المسبقة كما يلي :

$$y^* = X_i'\beta + e_i ,$$

$$,y_i = \begin{cases} 1 & if \quad y_i^* \geq 0 \\ 0 & if \quad y_i^* < 0 \quad i = 1, \ldots, n \end{cases}$$

$$y^*_{n\times 1}|X,\beta,\sigma^2 \sim N_n(X\beta\,,\sigma^2 I_n),$$

$$\beta^{p\times 1}|\,\lambda \sim \prod_{j=1}^{p} Inverse\ uniform\ \left(-\frac{1}{\lambda}\,,\frac{1}{\lambda}\right),$$

$$\lambda|\theta \sim \prod_{j=1}^{p} Inverse\ Gamma\ (2\,,\theta),$$

$$\sigma^2 \sim Inverse\ Gamma\ (c\,,d\,),$$

$$Z \sim standard\ exponential$$

5-3 التوزيعات الشرطية اللاحقة الكاملة (The full conditional posterior distribution) :

وبالاعتماد على النموذج الهرمي للتوزيعات المسبقة المذكور في الفقرة (3-4) يمكن الآن اشتقاق التوزيعات اللاحقة posterior ولكل متغير وكما موضح في ادناه :

أولاً يجب كتابة التوزيع المشترك الكامل full joint لجميع المتغيرات :

$$f(y^*|\beta,X,\sigma^2).\pi(\sigma^2).\pi(\lambda).\prod_{j=1}^{p}\pi\,(\beta_j|\lambda_j).\pi(z_j)I\left\{|\beta_j| > \frac{\lambda_j}{Z_j}\right\}$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}}\,(y^* - X\beta)'(y^* - X\beta) * \frac{d^c}{\Gamma(c)}(\sigma^2)^{-c-1}e^{-\frac{d}{\sigma^2}} * \prod_{j=1}^{p}\frac{\lambda_j}{2\beta^2}$$

$$* \frac{\theta^2}{\Gamma(2)}(\lambda)^{-2-1}e^{-\frac{\theta}{\lambda}} * \prod_{j=1}^{p} e^{-Z_j}$$

$$* \prod_{j=1}^{p} I\{\,|\beta_j| > \frac{\lambda_j}{Z_j}\} \ldots \ldots \ldots \ldots \ldots \ldots (3-6)$$

الآن يمكن الحصول على النموذج اللاحق المشترك لكل متغير (معلمة) وكما يلي :

1- التوزيع الشرطي الكامل للمتغير y_i^* هو توزيع طبيعي :

$$y_i^*|y_i,\beta,\sigma^2 = \begin{cases} \mathrm{N}(x_i'\beta\,,\sigma^2 I_n)\mathrm{I}\{y_i^*>0\}, \text{if } y_i=1 \\ \mathrm{N}(x_i'\beta\,,\sigma^2 I_n)\mathrm{I}\{y_i^*\le 0\}, \text{if } y_i=0 \end{cases}$$

2- التوزيع الشرطي اللاحق الكامل لـ (β)هو :

طالما ان خوارزمية *Gibbs Sampler* لاتحتاج الى اكثر من التوزيع اللاحق النسبي *proportional* او *unnormalized* الذي يعبر عن حاصل ضرب دالة الامكان بدالة التوزيع المسبق . لذلك سوف نحتاج فقط الأجزاء التي تتناسب او التي تحتوي على المتغير β من التوزيع المشترك في المعادلة $(3-6)$, وسوف نحذف اي جزء لايحتوي على β وبالاستعانة بالعلاقة التي اشرنا اليها من خلال الربط بين اسلوب لاسو واسوب معكوس لاسو, وسوف نحصل على :

$$\pi(\beta|y^*,X,\lambda\,,\sigma^2) \propto \pi(y^*|X,\beta\,,\sigma^2).\pi(\beta|\lambda)$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}(y^*-X\beta)'(y^*-X\beta)\right\}\prod_{j=1}^{p} I\{|\beta_j|>\frac{\lambda_j}{Z_j}\}$$

$$\propto \exp\{-\frac{1}{2\sigma^2}[\beta'(X'X)\beta-2y^*X\beta+y^*y^{*\prime}]\prod_{j=1}^{p} I\{|\beta_j|>\frac{\lambda_j}{Z_j}\}$$

$$\propto \exp\{-\frac{1}{2\sigma^2}[\beta'A\beta-2y^*X\beta+y^*y^{*\prime}]\ldots\ldots\ldots(3-7)$$

حيث ان $A=(X'X)$ ومن العلاقة $(2-13)$ يمكن استخدام المقدار التربيعي الآتي

$$(\beta-A^{-1}X'y^*)'A(\beta-A^{-1}X'y^*)$$
$$=\beta'A\beta-2y^*X\beta+y^{*\prime}(XA^{-1}X')y^*$$

وبالتالي يمكن كتابة العلاقة $(3-7)$ بالشكل الاتي :

$$\propto \exp\{-\frac{1}{2\sigma^2}[(\beta-A^{-1}X'y^*)'A(\beta-A^{-1}X'y^*)$$
$$+y^{*\prime}(I_n-XA^{-1}X')y^*]\}\ldots\ldots(3-8)$$

ان الجزء الثاني من العلاقة $(3-8)$ لايحتوي على β لذلك فان التوزيع المشترك الذي يحتوي على β سيتم اختصاره بالشكل الاتي :

$$\propto \exp\{-\frac{1}{2\sigma^2}[(\beta-A^{-1}X'y^*)'A(\beta-A^{-1}X'y^*)\}$$

وبالرجوع الى التوزيع الطبيعي متعدد المتغيرات $X \sim N(M, \Sigma)$, اي ان :

$$f(x; M, \Sigma) = \frac{e^{-\frac{1}{2}(X-M)'\Sigma^{-1}(X-M)}}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}$$

فانه يمكن الاستنتاج بان β تملك توزيع طبيعي متعدد المتغيرات بمتوسط $A^{-1}X'y^*$ وتباين $\sigma^2 A^{-1}$.

3- التوزيع الشرطي اللاحق الكامل لـ σ^2 هو :

سيتم توليد عينات للمتغير σ^2 باستخدام خوارزمية *Gibbs Sampler* من خلال اخذ جميع الاجزاء التي تضم σ^2 في التوزيع المشترك $(3-6)$ اي ان :

$$\pi(\sigma^2|y^*, x, \beta) \propto \pi(y^*|x, \beta, \sigma^2).\pi(\sigma^2)$$

$$\propto \frac{1}{\sigma^{2\frac{n-1}{2}}}[-\frac{1}{2\sigma^2}(y^* - X\beta)'(y^* - X\beta) * \frac{d^c}{\Gamma(c)}(\sigma^2)^{-c-1} * e^{-\frac{d}{\sigma^2}}$$

$$\propto \sigma^{2^{-\left(\frac{n-1}{2}\right)-c-1}}\exp[-\frac{1}{2\sigma^2}(y^* - X\beta)'(y^* - X\beta) - \frac{d}{\sigma^2}$$

$$\ldots (3-9)$$

وبالاستعانة بتوزيع معكوس كاما نجد ان :

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{-\alpha-1}\exp\left\{-\frac{\beta}{x}\right\} \ldots \ldots \ldots \ldots (3-10)$$

حيث α تمثل معلمة الشكل و β تمثل معلمة القياس , وباستخدام الدالة $(3-10)$ مع العلاقة $(3-9)$ يمكن الاستنتاج بان σ^2 يمتلك توزيع معكوس كاما بمعلمة شكل $(\frac{n-1}{2} + c)$ ومعلمة قياس $(y^* - X\beta)'(y^* - X\beta) + d$.

4- التوزيع الشرطي اللاحق الكامل لـ λ هو :

سيتم توليد عينات للمتغير λ من التوزيع المشترك في العلاقة $(3-6)$ بعد اخذ الحدود التي تضم فقط المتغير λ وباستخدام خوارزمية *Gibbs sampler* وكما يلي :

$$\pi(\lambda|\theta, \beta) \propto \pi(\beta|\lambda) * \pi(\lambda)$$

$$\propto \prod_{j=1}^{p} \frac{\lambda_j}{2\beta^2} \; \frac{\theta^2}{\Gamma(2)} (\lambda)^{-2-1} e^{-\frac{\theta}{\lambda}} . I\{|\beta_j| > \frac{\lambda_j}{Z_j}\}$$

$$\propto \lambda^{-(p+2)-1} e^{-\frac{\theta}{\lambda}} . I\{\lambda_j < Z_j|\beta_j|\} \ldots \ldots \ldots . (3-11)$$

حيث يمكن القول ان العلاقة $(3-11)$ انتجت توزيع المتغير λ ليكون

$. \; inverse \; gamma(p+2, \theta)$

5- ان التوزيع اللاحق الشرطي لـ Z هو :

$$\pi(z|\,\lambda\,) \propto \pi(z) * \pi(\beta|z)$$

$$\propto \prod_{j=1}^{p} e^{-z_j} . I\left\{Z_j > \frac{\lambda_j}{|\beta_j|}\right\} \propto \prod_{j=1}^{p} exponential(1) . I\left\{Z_j > \frac{\lambda_j}{|\beta_j|}\right\}$$

4 - الجانب التطبيقي

في هذا الجانب تم الاعتماد على بيانات حقيقة مأخوذة من احدى المستشفيات من القسم الخاص بتشخيص مرض كورونا (Covid - 19) من اجل تحليل العلاقة بين حالة المريض عند خروجه من المستشفى (المتغير المعتمد) ومجموعة من التشخيصات (المتغيرات المستقلة) المحددة من قبل اطباء الاختصاص لحالة المريض , حيث كانت مجموعة البيانت التي تم الحصول عليها من المستشفى من قبل الباحث هي 250 مشاهدة حيث كان عدد المتغيرات المستقلة مساوي الى 25 متغير وبيانات التدريب (training data) تساوي 20 وبيانات الاختبار (testing data) تساوي 230 . وتم اختيار الــ **(quantile)** مساوي الى (p=0.50) و(p=0.95) في تحليل هذه العلاقة . وسبب اختيار الباحث لهذه النسب هو انه دائماً البيانات تتمركز حول خط الوسط اي عندما (p=0.50) , واما بالنسبة للقيمة الاخرى وهي قيمة متطرفة للــ **(quantile)** اي عندما (p=0.95) وسبب اختيارنا لهذه القيمة لانه اذا كان خط الانحدار جيد عند هذه القيمة فهو مؤكد ان يكون جيد في قيم الـ **(quantile)** مادون ذلك او بشكل متقارب منه

1-4 توصيف البيانات

حيث صنفت المتغيرات وكما يلي :-

Y : حالة خروج المريض والتي تمثل المتغير المعتمد , X_1 : نوع الجنس ,X_2 : التحصيل الدراسي ,X_3 : الوظيفة ,X_4 : العمر بالسنوات ,X_5 : الحالة الزوجية ,X_6 : محل الاقامة ,X_7 : سكن المريض ,X_8 : مدة رقود المريض بالايام ,X_9 : حالة التدخين ,X_{10} : نسبة السكر في الدم

X_{11} : ضغط الدم, X_{12} : وزن الشخص, X_{13} : نسبة اليوريا في الدم, X_{14} : نسبة الكرياتين في الدم, X_{15} : (LDH)انزيم مرتبط بالرئة, X_{16} : (CRP)تحليل يختص بوجود الالتهابات, X_{17} : تحليل يمثل مخزون الحديد في الدم, X_{18} : (ESR) تحليل يختص بوجود الالتهابات, X_{19} : (HGB) يمثل نسبة الدم, X_{20} : (WBC)نسبة كريات الدم البيضاء, X_{21} : NEU) يمثل تحليل يكتشف المناعة عند الانسان, X_{22} : (LYM) تحليل يمثل وجود فايروس في الجسم, X_{23} : (PLT) تحليل يمثل عدد صفيحات الدم التي تفيد التخثر, X_{24} : (D-DIMER) تحليل يكتشف التخثر في الدم, X_{25} : (SPO2) تحليل يشير الى نسبة تركيز الاكسجين في الدم

2-4 نتائج تنفيذ الخوارزمية

وعند تنفيذ الخوارزمية التي تم كتابتها من قبل الباحث حيث كانت النتائج كما يلي :-

جدول رقم (4-1) يوضح نتائج الخوارزمية عند ($p = 0.5$)

	$BLBqr$	$BrLBqr$
$Intercept$	0	−1.828
$x1$	0.013	−6.864
$x2$	0.036	−0.144
$x3$	−0.004	2.673
$x4$	−0.002	−0.719
$x5$	0.013	2.911
$x6$	0.005	−1.981
$x7$	0.036	−4.649
$x8$	−0.02	−0.244
$x9$	0.009	−3.516
$x10$	−0.005	−5.468
$x11$	0.005	−0.606
$x12$	−0.018	−1.388
$x13$	−0.003	0.107
$x14$	−0.005	−1.829
$x15$	−0.003	−0.05
$x16$	−0.085	−1.154
$x17$	0	0.017
$x18$	0.005	0.268
$x19$	0.01	1.354
$x20$	0.002	0.492
$x21$	0.001	1.967
$x22$	0.001	2.348
$x23$	0	0.046
$x24$	−0.017	3.482
$x25$	0.026	−0.674

جدول رقم (2-4) يوضح نتائج الخوارزمية عند ($p = 0.95$)

	$BLBqr$	$BrLBqr$
$Intercept$	-0.002	15.317
$x1$	-0.014	-2.536
$x2$	0.006	0.671
$x3$	0.015	3.149
$x4$	0.008	0.195
$x5$	0.03	7.044
$x6$	-0.014	3.347
$x7$	0.036	1.398
$x8$	-0.009	0.239
$x9$	-0.009	4.053
$x10$	-0.025	0.829
$x11$	-0.02	0.611
$x12$	-0.001	-0.53
$x13$	-0.004	-0.015
$x14$	0.019	-1.299
$x15$	-0.001	-0.024
$x16$	-0.215	0.196
$x17$	0	0
$x18$	-0.005	-0.085
$x19$	-0.01	-2.339
$x20$	-0.009	-0.144
$x21$	0.011	0.058
$x22$	0.014	-0.164
$x23$	0.002	0.019
$x24$	-0.128	0.498
$x25$	0.062	0.392

3-4 تفسير النتائج

1- اظهرت نتائج الجدول رقم (1-4) تفوق في اداء الطريقة المقترحة $BrLBqr$ $Bayesian\ reciprocal\ Lasso\ Binary\ quantile\ regression$ عند المستوى ($p = 0.5$) حيث حصلت الطريقة المقترحة $BrLBqr$ على 217 نتيجة من التصنيف الصحيح من مجموع المشاهدات 230 , بينما الطريقة الاخرى

$BLBqr$ $Bayesian$ $Lasso$ $Binary$ $quantile$ $regression$ حصلت على 199 من التصنيف الصحيح من مجموع المشاهدات الكلي 230 .

2- افرزت نتائج الجدول رقم (2-4) ايضا تفوق في اداء الطريقة المقترحة $BrLBqr$ عند المستوى ($p = 0.95$) حيث حصلت على 205 من التصنيف الصحيح من المجموع الكلي للمشاهدات 230 , بينما الطرق الاخرى $BLBqr$ حصلت على 181 من التصنيف الصحيح من مجموع المشاهدات 230 .

5- الاستنتاجات والتوصيات

5-1 الاستنتاجات

في هذه البحث قدمنا نهجاً بايزياً للانحدار الكمي الثنائي مقترناً بتقنية اختيار المتغيرات (Selection variables) , اي الانحدار الكمي الثنائي البيزي مع دالة جزاء معكوس لاسو (Binary reciprocal lasso quantile regression). وان المزايا الرئيسية لهذا النهج تمثلت باجراء تقدير معالم النموذج واختيار المتغيرات (Selection variables) التنبوية المؤثرة على المتغير التابع دون التحسس بالقيم الشاذة على العكس من الطرائق الاخرى مثل طريقة المربعات الصغرى الاعتيادية (OLS) وغيرها من الطرق , وبالتالي يعتبر هذا النهج من الطرق الحصينة . وايضاً يمكن لهذا الاسلوب تحديد المتغيرات التي تعد تنبؤات مهمة لكميات مختلفة لتوزيع استجابة المتغير التابع (response variable) . تم تطبيق الطريقة المقترحة (BrLBqr) على بيانات حقيقية ومقارنتها مع طريقة (BLBqr). وباستخدام خوارزمية (Gibbs sampler) اظهرت النتائج بتفوق اداء الطريقة المقترحة (BrLBqr) مقارنةً بالطريقة الاخرى (BLBqr) .

5-2 التوصيات

يوصي الباحث باستخدام النموذج الهرمي التنظيمي (Regularization) المقترح في ظل التمثيل المختلط الجديد لتوزيع معكوس لابلاس (Asymmetric Laplace distribution ALD) مع انواع مختلفة من نماذج الانحدار مثل Tobit qauntile و Elastic net وغيرها من الطرق التنظيمية الاخرى . وايضاً نوصي باستخدام النموذج المقترح في مجالات اخرى مثل البحوث الاقتصادية , البحوث الاجتماعية والبحوث الرياضية . لذلك نوصي باستخدام النموذج المقترح لانه يوفر نماذج (Sparse) يمكن ان تكون مفيدة في حقول البيانات المختلفة .

References

[1] Benoit DF, Van den Poel D (2012) Binary Quantile Regression: A Bayesian Approach Based on the Asymmetric Laplace Distribution. J App Econom 27:1174{1188.

[2] Efron, B., T. Hastie, I. Johnstone, R. Tibshirani, et al. (2004). Least angle regression. The Annals of statistics 32 (2), 407–499.

[3] Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association 96 (456), 1348–1360.

[4] Koenker, R., Bassett G. J., 1978, “Regression quantiles”, Econometrica, 46, pp.33-50.

[5] Keming Yu& Rana A.Moyeed (2001)" Bayesian Quantile Regression" Statistical & Probability Letters 54:437-447.

[6] Keming Yu, Zudi Lu and Julian Stander (2003) “Quantile regression :applications and current research areas” The Statistician 52, Part 3, pp. 331-350.

[7] Kordas G (2006) Smoothed Binary Regression Quantiles. J Appl Econom 21:387- 407.

[8] Li Q, Xi R, Lin N (2010) Bayesian Regularized Quantile Regression. Bayesian Analysis 5:1-24.

[9] Mallick, H., Yi, N., 2014, “A new Bayesian lasso”, Statistics and its interface, 7: pp. 571-582.

[10] Meinshausen, N. (2007). Relaxed lasso. Computational Statistics & Data Analysis 52 (1), 374–393.

[11] Mallick, H., Alhamzawi, R., Paul, E., & Svetnik, V. (2021). The reciprocal Bayesian lasso. Statistics in Medicine, 40(22), 4830-4849.

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (20) عدد (2) 2024

[12] Qifan Song and Faming Liang. High-dimensional variable selection with reciprocal 1-Regularization.Jornal of the American Statistical Association, 110(512):1607-1620,2015.

[13] Radchenko, P. and G. M. James (2008). Variable inclusion and shrinkage algorithms. Journal of the American Statistical Association 103 (483), 1304–1315.

[14] Tibshirani (1996)"Regression Shrinkage and selection via lasso" J. R.Statis.Soc.58,No.1,pp. 267-288.