



جمهورية العراق  
وزارة التعليم العالي والبحث العلمي  
جامعة كربلاء  
كلية الادارة والاقتصاد  
قسم الاحصاء

## الانحدار الكمي الثنائي البيزي المقيد بدوال الجزاء المعكوسة ( مع تطبيق عملي )

اطروحة مقدمة الى مجلس كلية الادارة والاقتصاد — جامعة كربلاء وهي جزء  
من متطلبات نيل درجة دكتوراه فلسفة في علوم الاحصاء

تقدم بها

محمد طالب خنجر

بإشراف

أ . د . أحمد نعيم فليح

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

(وَيَسْأَلُونَكَ عَنِ الرُّوحِ قُلِ

الرُّوحُ مِنْ أَمْرِ رَبِّي وَمَا


أُوتِيتُمْ مِنَ الْعِلْمِ إِلَّا قَلِيلًا)

صدق الله العلي العظيم

(الاسراء الاية 85)

## إقرار المشرف

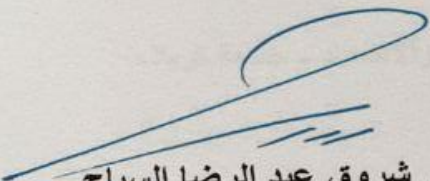
أشهد بأن إعداد هذه الأطروحة الموسومة ( الانحدار الكمي الثنائي البيزي المقيد بدوال الجزاء المعكوسة مع تطبيق عملي ) والتي تقدم بها الطالب " محمد طالب خنجر " قد جرى بإشرافي في قسم الاحصاء - كلية الادارة والاقتصاد - جامعة كربلاء ، وهي جزء من متطلبات نيل درجة الدكتوراه علوم في الاحصاء .

  
أ.د. احمد نعيم فليح

التاريخ: / / ٢٠٢٢

## توصية رئيس قسم الإحصاء

بناءً على توصية الاستاذ المشرف ، أرشح الأطروحة للمناقشة.

  
أ.د. شروق عبد الرضا السباح

رئيس قسم الاحصاء

التاريخ: / / ٢٠٢٢

## إقرار الخبير اللغوي

أشهد بأن الاطروحة الموسومة (الانحدار الكمي الثنائي البيزي المقيد بدوال الجزاء المعكوسة مع تطبيق عملي) قد جرى مراجعتها من الناحية اللغوية تحت اشرافي اذ أصبحت خالية من الاخطاء اللغوية ولأجله وقعت.

الخبير اللغوي

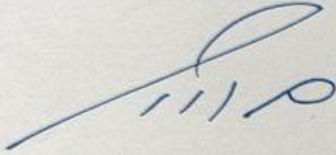
م. صلاح مهدي جابر

كلية الادارة والاقتصاد - جامعة كربلاء



## إقرار لجنة المناقشة

نشهد نحن أعضاء لجنة المناقشة بأننا قد اطلعنا على الاطروحة الموسومة  
(الاتحاد الكمي الثنائي البيزي المقيد بدوال الجزاء المعكوسة مع تطبيق عملي )  
والمقدمة من قبل الطالب "محمد طالب خنجر " وناقشنا الطالب في محتوياتها  
وفيما له علاقة بها، ووجدنا بأنها جديرة بنيل درجة الدكتوراه فلسفة في علوم  
الإحصاء .



أ.د. مهدي وهاب نعمة

عضواً

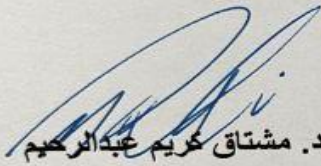
٢٠٢٢ / ١١ /



أ.د. عواد كاظم شعلان

رئيساً


٢٠٢٢ / ١١ /



أ.م. د. مشتاق كريم عبدالرحيم

عضواً

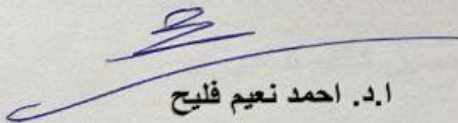
٢٠٢٢ / ١١ /



أ.م.د. صباح حسيب حسن

عضواً

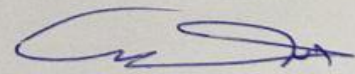
٢٠٢٢ / ١١ /



أ.د. احمد نعيم فليح

عضواً ومشرفاً

٢٠٢٢ / ١١ /



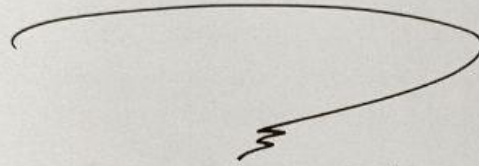
أ.م.د. ايناس عبدالحافظ محمد

عضواً

٢٠٢٢ / ١١ /

### إقرار رئيس لجنة الدراسات العليا

بناء على اقرار المشرف العلمي والخبير اللغوي على اطروحة  
الدكتوراه للطالب " محمد طالب خنجر " الموسومة بـ (الانحدار الكمي الثاني  
البيزي المقيد بدوال الجزاء المعكوسة مع تطبيق عملي ) ارشح هذه الاطروحة  
للمناقشة.



أ.د محمد حسين كاظم الجبوري

رئيس لجنة الدراسات العليا

معاون العميد للشؤون العلمية والدراسات العليا

### مصادقة مجلس الكلية

صادق مجلس كلية الادارة والاقتصاد/ جامعة كربلاء على قرار لجنة المناقشة.



أ.د.محمد حسين كاظم الجبوري

عميد كلية الادارة والاقتصاد- جامعة كربلاء

وكالة

## .... شكر وامتنان ....

الحمد والشكر لله رب العالمين والصلاة والسلام على سيد الخلق محمد (صلى الله عليه وآله وسلم) . أتقدم بخالص شكري وتقديري الى استاذي الفاضل الاستاذ الدكتور أحمد نعيم فليح الذي تفضل بالإشراف على هذه الأطروحة وشملها برعايته العلمية وتوجيهاته القيمة وآرائه السديدة في إتمام هذا البحث .

كما اتقدم بالشكر الجزيل الى الاساتذة الفضلاء رئيس لجنة المناقشة وعضائها المحترمين لتفضلهم بمناقشة اطروحتي واغنائها بملاحظاتهم القيمة ، كما اتقدم بالشكر والتقدير والامتنان الى كل من قدم يد العون لي خلال مدة كتابة هذه الاطروحة .

الباحث

## الإهداء

إلى أساتذتي وزملائي كافة ..... عرفاناً وامتناناً

إلى زوجتي وأولادي ..... إخلاصاً وحناناً

الباحث



## المحتويات

الصفحة	الموضوع	الفصل
I , II	المحتويات	
III	المختصرات والرموز	
IV , V	الجدول والمخططات	
VI	الملخص	
12 - 1	<b>المقدمة وهدف البحث والاستعراض المرجعي</b>	<b>الفصل الاول</b>
4-1	المقدمة	1-1
5	هدف البحث	2-1
12-5	الاستعراض المرجعي	3-1
32-13	<b>الجانب النظري</b>	<b>الفصل الثاني</b>
14-13	المقدمة	1-2
15-14	توزيع لابلاس	2-2
16-15	اختيار المتغيرات	3-2
18-16	مشكلة التعدد الخطي	4-2
20-18	الانحدار الكمي	5-2
21-20	استخدام طريق (Lasso) في الانحدار الخطي	6-2
22	استخدام طريق معكوس (Lasso) في الانحدار الخطي	7-2
28-26	الانحدار الكمي البيزي الثنائي باستخدام معكوس لاسو (Lasso)	8-2
28	النموذج الهرمي البيزي المقترح للتوزيعات السابقة	9-2
32-29	التوزيعات الشرطية اللاحقة الكاملة	10-2
48-33	<b>الجانب التجريبي</b>	<b>الفصل الثالث</b>
34-33	المقدمة	1-3
35	توليد الخطأ العشوائي	2-3
42-34	نتائج المحاكاة	3-3

48-43	الرسومات التوضيحية	4-3
52-49	<b>الجانب التطبيقي</b>	<b>الفصل الرابع</b>
49	المقدمة	1-4
51-50	توصيف البيانات	2-4
53-52	نتاج تنفيذ الخوارزمية	3-4
54	تفسير النتائج	4-4
56-55	<b>الاستنتاجات والتوصيات</b>	<b>الفصل الخامس</b>
56-55	الاستنتاجات	1-5
56	التوصيات	2-5
61-57	<b>المصادر</b>	

## المختصرات والرموز

Symbol	Term
QR	Quantile Regressin
Lasso	Least absolute shrinkage and selection operator
MMAD	Median of Mean Absolute Deviation
MCMC	Markov Chain Monte Carlo
LD	Laplace Distribution
AL	Asymmetric Laplace
OLS	Ordinary Least Square
BLUE	Best Linear Unbiased Estimator
LAD	least Absolute Deviation
RRM	Ridge Regression Method
RQR	Regularization Quantile Regression
BRQR	Bayesian Regularization Quantile Regression
BQR	Bayesian Quantile Regression
SM	Scale mixture
BrLBqr	Bayesian reciprocal Lasso Binary quantile regression
BLBqr	Bayesian Lasso Binary quantile regression
BBqr	Bayesian Binary quantile regression

## الجدول

الصفحة	اسم الجدول	رقم الجدول
35	يوضح استخدام الطرق مع كل محاكاة ومع كل حالة	1-3
37	قيم المعايير (MMAD) و (Sd) عندما (p=0.5) ، المحاكاة الاولى	2-3
38	قيم المعايير (MMAD) و (Sd) عندما (p=0.5) ، المحاكاة الثانية	3-3
39	قيم المعايير (MMAD) و (Sd) عندما (p=0.5) ، المحاكاة الثالثة	4-3
40	قيم المعايير (MMAD) و (Sd) عندما (p=0.95) ، المحاكاة الاولى	5-3
41	قيم المعايير (MMAD) و (Sd) عندما (p=0.95) ، المحاكاة الثانية	6-3
42	قيم المعايير (MMAD) و (Sd) عندما (p=0.95) ، المحاكاة الثالثة	7-3
52	يوضح نتائج الخوارزمية عند ( p=0.5 )	1-4
53	يوضح نتائج الخوارزمية عند ( p=0.95 )	2-4

## المخططات

الصفحة	اسم المخطط	رقم المخطط
20	يمثل خطوط الانحدار التي يتم تقديرها عبر نموذج الانحدار الكلاسيكي ونموذج الانحدار الكمي (QR)	1-2
22	يوضح رسم دالة جزاء لاسو	2-2
23	يوضح رسم دالة جزاء معكوس لاسو	3-2
24	يوضح رسم دالة جزاء لاسو مع دالة جزاء معكوس لاسو	4-2

43	رسم مخطط ( trace plot ) للمعالم المقدرة ( $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$ ) مع المحاكاة الاولى	1-3
44	المدرج التكراري ( Histogram ) لكل معلمة من المعالم المقدرة ( $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$ ) مع المحاكاة الاولى	2-3
45	رسم مخطط ( trace plot ) للمعالم المقدرة ( $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$ ) مع المحاكاة الثانية	3-3
46	المدرج التكراري ( Histogram ) لكل معلمة من المعالم المقدرة ( $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$ ) مع المحاكاة الثانية	4-3
47	رسم مخطط ( trace plot ) للمعالم المقدرة ( $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$ ) مع المحاكاة الثالثة	5-3
48	المدرج التكراري ( Histogram ) لكل معلمة من المعالم المقدرة ( $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$ ) مع المحاكاة الثالثة	6-3



## الملخص

يعد الانحدار الكمي من الطرق التي اخذت حيزاً واسعاً في التطبيق في العقدين السابقين لما تتمتع به هذه الطرق من ميزات جاذبة للباحثين فهو لا يتأثر بالقيم الشاذة اي انه يعتبر من الطرق الحصينة ، كما وانه يعطي تفاصيل اكثر لتأثير المتغيرات التوضيحية على المتغير المعتمد .

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 (Quantile Regression) اذ تم اعتماد الاسلوب البيزي وهو من الاساليب التي اخذت صدى واسعاً في الآونة الأخيرة لما لها من الدقة وخصوصاً في العينات الصغيرة .

في هذا البحث تم اقتراح نموذج هرمي بيبي لاختيار المتغير وتقديره في سياق الانحدار الكمي الثنائي . تعتبر المناهج الحالية لاختيار المتغير في سياق التصنيف الثنائي حساسة للقيم المتطرفة او الغير متجانسة وغيرها من الحالات الشاذة . الطريقة المقترحة في هذه الدراسة تتغلب على هذه المشاكل بطريقة جذابة ومباشرة .

بالإضافة الى ذلك فقد قيد الباحث الدراسة باستخدام 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 .

وباستخدام خوارزمية (Gibbs Sampler) الفعالة لتقدير معالم النموذج التي تتفوق على خوارزمية (Metropolis) المستخدمة في الدراسات السابقة . أظهرت نتائج كل من دراسة المحاكاة وتحليل البيانات الحقيقية الى ان الطريقة المقترحة (BrLBqr) تتفوق في الأداء من حيث تقدير معالم النموذج واختيار المتغيرات التوضيحية ذات الصلة او المؤثرة بمتغير الإستجابة مقارنةً بالطرق الاخرى (BLBqr) و (BBqr) .

# الفصل الاول

المقدمة وهدف البحث  
والاستعراض المرجعي

## 1-1 المقدمة ( Introduction ):

يعتبر تحليل الانحدار اداة احصائية مهمة في هيكله ونمذجة العلاقة بين متغير الاستجابة او المتغير المعتمد والمتغيرات التفسيرية ( explanatory variables ) ويعتبر ايضا من المواضيع المهمة في كثير من المجالات والبحوث العلمية ، ان انحدار المتوسط ( Mean Regression ) يعتبر احد الطرق الاحصائية الواسعة الانتشار في تحليل العلاقة مابين المتغيرات التنبؤية ( المتغيرات التوضيحية ) ومتغير الاستجابة ويكون ذلك التحليل من نقطة المتوسط في توزيع المتغير المعتمد ، ومن الواضح ان تحليل انحدار المتوسط يفترض عدد من الشروط الواجب تحققها حتى تتم عملية تقدير النموذج ، ومن هذه الشروط وابرزها ان الاخطاء العشوائية تكون مستقلة وتوزع توزيعاً طبيعياً بمتوسط مقداره صفر وتباين مقداره ثابت هو  $(\sigma_{\varepsilon}^2)$  اي ان :

$$\varepsilon_i \sim N(0, \sigma_{\varepsilon}^2)$$

ان الهدف من بناء نموذج الانحدار قد يكون لغرض تفسير العلاقة بين المتغيرات او لغرض التنبؤ ( prediction ) او لاغراض اختبار المتغيرات الاكثر اهمية من بين المتغيرات التفسيرية على متغير الاستجابة ، وتعتبر طريقة المربعات الصغرى من اهم واشهر طرق تقدير معالم نموذج الانحدار الخطي وخصوصا عندما يكون عدد المتغيرات التفسيرية ( p ) اقل من حجم العينة المدروسة ( n ) اي انه في حالة ( n > p ) فان مقدرات ( ols ) تكون افضل تقدير خطي غير متحيز BLUE .

ولكن في بعض الاحيان تصادف الباحث مشكلة وجود ( p > n ) او ( p ) قريبة من ( n ) فان هذا ينعكس على اداء طريقة التقدير ( ols ) في تقدير المعالم حيث تعطي تقديرات غير كفوءة و متحيزة وذات تباينات كبيرة .

كذلك فان وجود مشكلة ( p > n ) قد تولد لنا ترابطات ( اعتمادات خطية ) بين المتغيرات التفسيرية او التنبؤية مما يعني صعوبة حساب مصفوفة  $(X'X)^{-1}$  لتقدير المقدرات (  $\beta$  ) لنموذج الانحدار وبالتالي تكون تقديرات *ols* غير دقيقة ولايعتمد عليها في التنبؤ.

مما يعني ان حالة وجود ( p > n ) تولد نماذج صعبة التفسير ، لذلك نلجأ الى طرق تقدير اخرى تتغلب على هذه المشكلة مثل طريقة Ridge التي اقترحها ( Kennard Hoerl and ) (1970)<sup>(19)</sup> وكما في الصيغة الرياضية التالية:

$$Q(\beta) = (y - X\beta)'(y - X\beta) + \lambda \|\beta\|_2^2 ; \lambda \geq 0$$

وطريقة Lasso التي اقترحها الباحث ((Tibshirani , 1996))<sup>(45)</sup> كما في المعادلة الآتية :

$$Q(\beta) = (y - X\beta)'(y - X\beta) + \lambda \|\beta\|_1 ; \lambda \geq 0$$

وطرق تقدير اخرى من خلالها يتم فرض شرط جزائي على مجموع مربعات الخطأ بهدف الحصول على مقدرات ذات معنى لمعالم نموذج الانحدار.

هناك عدد كبير من البحوث العلمية الحديثة اثبتت ان التنظيم او الضبط (Regularization) تكون لها اهمية كبيره في نماذج الانحدار عند وجود مشكلة التعدد الخطي ، فمثلاً تقنية ( Ridge Regression) والتي قدمها كل من الباحثين ((Arthur & Robert ( 1970))<sup>(1)</sup> حيث كان لها دور مهم في التغلب على هذه المشكلة حيث تقوم هذه التقنية (R.R) بالتضحية بقدر بسيط من التحيز مقابل تقليل التباين (Variance) ، وايضاً طريقة لاسو (Lasso) المكتشفة من قبل الباحث ((Tibshirani (1996))<sup>(45)</sup> كان لها دوراً مهماً في اختيار المتغيرات المؤثرة وتقدير المعالم في آن واحد (Variable Selection & Parameter estimation) ، أما طريقة التقدير ( Elastic Net ) التي اكتشفت من قبل الباحثان (Zou & Hastie (2005))<sup>(21)</sup> وهي تعتبر كتطوير لطريقة لاسو (Lasso) حيث كان لها دورا بارزاً عندما تكون الارتباطات بين المتغيرات التفسيرية عالية الابعاد.

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

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

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

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

ان بداية استخدام الطريقة البيزية في نموذج الانحدار الكمي (QR) كانت من قبل الباحثين ( Yu (2001) & Moyeed (23) حيث تم استخدامها في نموذج الانحدار الكمي مفترضين ان دالة الامكان (Likelihood Function) لتقدير المعالم المدروسة هي دالة لابلاس الملتوي (Asymmetric Laplace distribution).

ويعتبر الانحدار الكمي (Quantile Regression (QR) هو احد نماذج الانحدار الذي يقيس اثر المتغير المستقل على المتغير المعتمد بخطوط انحدار غير منتهية عند تقسيمات معينه محصورة ما بين (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 (23,24).

ويمكن التعبير عن نموذج الانحدار الكمي كما في المعادلة التالية (23) :

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

$y_i$  : يمثل تمثل قيمة المشاهدة الحقيقية للمتغير الذي ترتيبه  $i$

$x_i$  : متجه من درجة (n×p)

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

$\beta_p$  : متجه معالم النموذج من درجة (p×1)

وان عملية تقدير معالم نموذج الانحدار الكمي  $\beta_p$  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 :

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

كذلك يعتبر اسلوب الانحدار الكمي من الطرق المهمة في كثير من الدراسات الاحصائية في الأونة الاخيرة لما لها من القدرة او الامكانية في تحري العلاقة بين المتغيرات التنبؤية )



المتغيرات التوضيحية) ومتغير الاستجابة وفي كل التوزيع الشرطي ( conditional distribution) لمتغير الاستجابة وذلك من خلال التجزيئات او الكميات الشرطية (conditional quantiles)  $Q_p(Y/X), p \in (0,1)$  المختلفة بدلاً من الاختصار على تقدير المتوسط او التوقع الشرطي  $E(Y/X)$  وهي طريقة انحدار المتوسط (mean regression) .

ان الكوانتايل (quantile) ليس تعبيراً جديداً في علم الاحصاء فعلى سبيل المثال مقياس الوسيط يعتبر حالة خاصة من (quantile) لانه يقسم مجموعة البيانات الى جزئين او مجموعتين متساويتين اي ان  $F(y) = 0.5$  ، وايضا هناك مقاييس اخرى تعتبر حالة خاصة مثل الربيعات والعشيرات والمننيات

ان اهمية الانحدار الكمي تظهر عندما يكون المطلوب هو تقدير العلاقة عند مستويات مختلفة غير نقطة المتوسط للتوزيع الشرطي لمتغير الاستجابة حيث ان كثير من البحوث والدراسات الحديثة بينت ان تحليل العلاقة لبعض الظواهر من نقطة المتوسط للتوزيع الشرطي لمتغير الاستجابة كما في الانحدار المتوسط (mean regression) ربما لاتعطي واقعية العلاقة بشكل صحيح بين المتغيرات التفسيرية ومتغير الاستجابة ، في حين ان نموذج الانحدار الكمي له القدرة والامكانية على دراسة هذه العلاقة في كل اقسام التوزيع .

ونظراً لاهمية (QR) في تحليل العلاقة بين المتغيرات التفسيرية ومتغير الاستجابة في عدد كبير من البحوث الحديثة وللصفات التي يتميز بها عن انحدار المتوسط ( الانحدار الاعتيادي) نلاحظ ان البحوث العلمية الحديثة تعطي اهتماماً كبيراً لهذه النماذج وفي مجالات كثيرة ، كما انها في كثير من المجالات التطبيقية تكون مناسبة لكثير من البيانات لانها لاتتشرط التوزيع الطبيعي للبيانات او التماثل في التوزيع وكذلك لاتفترض بتوزيع معين للاخطاء العشوائية على العكس من الطرق التقليدية الاخرى مثل طريقة (ols)<sup>(40)</sup> التي تضع شروطاً معينة يجب تحققها .

وتعتبر دراسة اسلوب بيز بوجود دوال جزاء معكوسة ( reciprocal Penalized ) في نموذج الانحدار الكمي الثنائي ( Binary Quantile ) موضوع جديد وقليل الدراسات ولم يتطرق احد مسبقاً لدمج دوال جزاء معكوسة في الانحدار الكمي الثنائي ( Binary Quantile ) .

وسوف نطبق النموذج المقترح على بيانات حقيقية ومقارنة النتائج مع طرق تقدير اخرى للوقوف على اداء نموذجنا .

## 2-1 الهدف من البحث

تهدف هذه الأطروحة الى 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 .

## 3-1 الاستعراض المرجعي Literature Review

في عام 1972 قدم الباحثان<sup>(30)</sup> ( Lindley & Smith ) بحثاً يتناول عملية تقدير معالم النموذج الإحصائي باستخدام الطريقة البيزية وتم تطبيقها على نموذج انحدار خطي متعدد ، وإشارة نتائج الدراسة بان الطريقة البيزية تعطي تقديرات غير متحيزة لمعالم النموذج .

في عام 1978 قدم الباحثين<sup>(22)</sup> ( Roger Koenker ; Gilbert Bassett, Jr ) طريقة الانحدار الكمي (QR) ( Quantile regression ) كاسلوب احصائي شامل لتحليل العلاقة بين المتغير المعتمد والمتغيرات التنبؤية ( التوضيحية ) وفي جميع التوزيع الشرطي لمتغير الاستجابة ( response variable ) وذلك بفرض عدة تقسيمات شرطية  $Q_p(Y/X)$  بدلا عن الطريقة الكلاسيكية في ايجاد تقدير المتوسط الشرطي  $E(Y|X)$  كما في تقدير الانحدار الاعتيادي ، حيث تم استخدام طريقة least Absolute Deviation method ( LAD ) في تقدير معالم نموذج الانحدار بدلاً من الطريقة المستخدمة في الانحدار الاعتيادي وهي دالة الخسارة التربيعية ، وايضا بين الباحثين اهمية الانحدار الكمي بسبب مرونته وقلة افتراضاته ومنها عدم افتراض التوزيع الطبيعي للخطأ العشوائي .

وفي عام 1986 قام الباحث<sup>(37)</sup> ( Powell, J. ) بتقديم بحثاً يتضمن تقدير معالم نموذج الانحدار المراقب وعند تقسيمات شرطية مختلفة للـ ( Quantile ) ومستخدماً فيها طريقة التقدير ( LAD ) ، وقد ايضاً بان التقديرات الناتجة بهذه الطريقة تكون تقديرات متنسقة ( Consistent ) ولها توزيع طبيعي تقاربي ( Asymptotically normally distributed ) .

وفي عام 1987 ، قام الباحثان<sup>(28)</sup> ( Koenker & D'orey ) بتطوير وتعديل خوارزمية تم استخدامها في ايجاد تقدير اقل الاخطاء العشوائية المطلقة في الانحدار الخطي لحساب معالم نموذج الانحدار الكمي من خلال تصغير المعادلة الرياضية التالية :

$$\hat{\beta}_{QR} = \operatorname{argmin} \sum_{i=1}^n \rho_p(y_i - x_i' \beta_p)$$

حيث ان  $\rho_p(\cdot)$  تمثل دالة الخسارة في الانحدار الكمي (QR) وتكون حسب الصيغة الرياضية التالية :

$$\rho_p(\varepsilon) = \frac{|\varepsilon| + (2p - 1)\varepsilon}{2}$$

في عام 1996 قدم الباحث<sup>(45)</sup> (Tibshirani) مقترح طريقة لاسو (Lasso) في تقدير معالم نموذج الانحدار الخطي واستخدم فيها دالة الجزاء النوع الاول (L<sub>1</sub> penalty) بدلا من دالة الجزاء النوع الثاني (L<sub>2</sub> penalty) والذي استخدم في نموذج انحدار الحرف (Ridge regression method) ، وبكلام اخر انه فرض قيداً بتصغير مجموع مربعات البواقي ، وان هذا القيد يفرض على مجموع القيم المطلقة لمعاملات النموذج ان تكون أقل من ثابت معين وبالصيغة الرياضية يكون كما يلي :

$$\sum_{k=1}^p |\beta_k| \leq t \quad ; \quad t \geq 0$$

وبين الباحث ان لهذا الشرط اهمية كبيرة لانه يجعل بعض معاملات النموذج مساوياً تماماً الى الصفر ، وبالتالي فان هذه الطريقة لها دور كبير في اختيار المتغيرات المهمة وجعل معاملات المتغيرات الغير مهمة مساوية الى الصفر ، اي انها تعمل في اختيار المتغيرات (Variable selection) وجعل نماذج الانحدار اكثر دقة في التفسير (interpretable models) .

وفي العام نفسه اقترح الباحث<sup>(45)</sup> (Tibshirani) طريقة لاسو (Lasso) كأداة تنظيم للتغلب على مشكلة البعد العالي (the high dimensional) ومشكلة التعدد الخطي (Multicollinearity) ، وان هذه الاداة استخدمها لعملية اختيار المتغيرات ويتم تعريف مقدر لاسو على النحو التالي :

$$\hat{\beta}_{lasso} = \operatorname{argmin} \|y - x'\beta\|_2^2 + \lambda \|\beta\|_1$$

حيث ان  $0 \leq \lambda$  ، وهي تمثل معلمة الانكماش .

وفي العام نفسه ايضا اقترح الباحث<sup>(45)</sup> (Tibshirani) طريقة التقدير البيزية باستخدام دالة جزاء لاسو the Bayes estimate in Lasso method مفترضاً التوزيع الاولي للمعالم هو double exponential distribution او مايعرف بتوزيع لابلاس (Laplace distribution).

في عام 2001 قدم الباحثين<sup>(23)</sup> (Yu & Moyeed) مقترح الطريقة البيزية في تقدير معالم نموذج الانحدار الكمي الخطي (Linear QR) وبافتراض ان الخطأ العشوائي له توزيع لابلاس الغير متماثل (او الملتوي) (Asymmetric Laplace (AL)) من اجل ربط الطريقة البيزية بالطريقة التقليدية في ايجاد تقدير معالم النموذج وبملاحظة دالة الخسارة التريبيعية التي تم استخدامها في الاسلوب الكلاسيكي فان تصغير هذه الدالة هو مكافئ او مساوي الى تعظيم دالة الامكان عند استخدام توزيع لابلاس الغير متماثل (AL) للخطأ العشوائي .

وفي عام 2003 قام الباحث<sup>(46)</sup> (Tsiionas) بتقديم بحثاً يتضمن استخدام الطريقة البيزية في ايجاد تقدير معالم نموذج الانحدار الكمي (QR) مستخدماً فيها تمثيلاً مختلط لتوزيع لابلاس الغير متماثل او الملتوي لتوزيع الخطأ العشوائي ، وان سبب استخدام التمثيل المختلط لهذا التوزيع (AL) من اجل تسهيل وتسريع عمل الخوارزمية في اجراء المعاينة من التوزيع الشرطي اللاحق وبهذا يكون له افضلية على الاستخدام من قبل الباحثين (Yu & Moyeed<sup>(23)</sup>) ، وان الباحث استخدم في بحثه خوارزمية (Gibbs sampler) في طريقة الاستدلال البيزية . ان الـ quantile الذي استخدمه الباحث هو (p=0.50) حيث اظهرت النتائج بان لها الافضلية على طريقة (Yu & Moyeed 2001)<sup>(23)</sup> .

وفي عام 2004 استخدم الباحث<sup>(26)</sup> (Koenker) اسلوب الضبط او التنظيم (Regularization) المحاولة الاولى في نماذج (Mixed effect Quantile regression models) عند تطبيق تقنية او طريقة لاسو (Lasso) حيث كان لها الدور المهم والاداء الجيد في اختيار المتغير (variable selection) المهم وجعل معاملات المتغيرات الغير مهمة مساوية تماما الى الصفر .

في عام 2005 صرح الباحث<sup>(25)</sup> (Koenker) بانه يمكن النظر الى الانحدار الكمي (QR) كنموذج انحدار رصين يعطي تأثيرات مختلفة للمتغير المتنبئ به على مستوى مختلف من الكميات (Quantile) تحت دالة تنظيم  $L_1$

وفي عام 2007 قدم الباحثين<sup>(27)</sup> ( Yu & Stander ) مناقشة حول عملية ايجاد تقدير لمعالم نموذج الانحدار الكمي من وجهة المنظور البيزي وافترض الباحثان بان توزيع الخطأ العشوائي هو توزيع لابلاس الملتوي ( AL ) واستخدموا خوارزمية ( Markov chain Monte Carlo MCMC )) في ايجاد توزيعات المعاينة من التوزيعات الشرطية اللاحقة ، كما بينت النتائج الاداء الجيد للطريقة المستخدمة عند مقارنتها مع الطرائق الاعتيادية لتقدير معالم نموذج الانحدار في عام 2008 قدم الباحثين<sup>(33)</sup> ( Li & Zhu ) بحثاً يتضمن استخدام طريقة لاسو ( Lasso ) في الطريقة التقليدية لتقدير معالم النموذج ، حيث بين الباحثان بان استخدام هذه التقنية في التقدير لها اهمية كبيرة في اختيار المتغيرات المهمة ( variable selection ) وجعل معالم المتغيرات الغير مهمة مساوية تماما الى الصفر وبالتالي الوصول الى نموذج انحدار تفسيري جيد يمكن التنبؤ به .

في عام 2010 قدم الباحثين<sup>(40)</sup> ( Li, Xi & Lin ) بحثاً يتضمن دراسة الضبط او التنظيم Regularization في نموذج الانحدار الكمي الخطي ( Linear QR ) استخدموا فيها للمرة الاولى الطريقة البيزية الهرمية ( Bayesian Hierarchical method ) مع تقنيات او اساليب الضبط ( group lasso & elastic net penalties ) فضلاً عن اسلوب او تقنية لاسو ( Lasso ) لتقدير معالم نموذج الانحدار ( Linear QR ) ، حيث استخدموا خوارزمية ( Gibbs sampler & Metropolis Hasting ) في دراسة العلاقة لبيانات محاكاة وبيانات واقعية حيث اظهرت النتائج بان الطريقة البيزية التي تتضمن تقنية الضبط او التنظيم ( BRQR ) هي افضل من الطريقة التي لا تتضمن على تقنية التنظيم او الضبط ( BQR ) وايضا تعتبر افضل من الطرائق التي تستخدم تقنية التنظيم ولا تستخدم الطريقة البيزية Regularization (Quantile regression) (RQR).

في عام 2011 قدم الباحثين<sup>(20)</sup> ( Kozumi & Kobayshi ) دراسة تتضمن استخدام الطريقة البيزية في تقدير معالم نموذج الانحدار الكمي وبافتراض ان التوزيع الاولي لمتجه المعالم هو توزيع لابلاس مستخدمين فيها تمثيلاً مختلطاً لتوزيع لابلاس الغير متماثل (AL) ، حيث كان المقترح للتمثيل المختلط يتمثل بالتوزيع الطبيعي القياسي مع التوزيع الاسي القياسي وذلك لتسهيل وتسريع عمل الخوارزمية .

في عام 2013 قدم الباحثين<sup>(10)</sup> ( Benoit, D. F., Alhamzawi, R., & Yu, K. ) اقتراح نموذج هرمي بيزي لاختيار المتغيرات ( variable selection ) وتقدير معالم النموذج في



سياق الانحدار الكمي الثنائي و باستخدام دالة جزاء لاسو (Lasso) ، وذكر الباحثين بان اسلوب اختيار المتغير في سياق التصنيف الثنائي يكون حساس عند القيم المتطرفة ، الطريقة المقترحة في هذه الدراسة تتغلب على هذه المشكلة بطريقة جذابة ومباشرة . افترض الباحثين بان التوزيع الاولي للمعالم هو توزيع لابلاس . وتمت عملية التقدير باستخدام خوارزمية (MCMC) حيث اشارت نتائج المحاكاة والبيانات الحقيقية الى ان الطريقة المقترحة تعمل بشكل جيد مقارنة بالطرق الاخرى .

في عام 2014 قدم الباحثين<sup>(33)</sup> ( Mallick & Yi ) تمثيلاً جديداً مختلط لدالة توزيع لابلاس ( التوزيع الاولي المفترض الى معالم نموذج الانحدار ومع استخدام دالة جزاء لاسو ( Lasso ) ) حيث كان المقترح من قبل الباحثين للتمثيل المختلط لتوزيع لابلاس هو ( Scale mixture of uniform(SMU) ) ، بدلاً من الاقتراح المقدم من قبل الباحثين<sup>(2)</sup> ( Andrews, D.F. & Mallows(1974) ) للـ ( Scale mixture (SM) Normal & exponential ) اي انه يمثل ( Scale mixture of normal(SMN) ) .

حيث افرزت طريقتهم المقترحة كفاءة عالية في تقدير معالم نموذج الانحدار الخطي الإعتيادي واختيار المتغيرات .

في عام 2014 قدم الباحث<sup>(44)</sup> ( Song, Q ) اطروحة الدكتوراه تناول فيها طريقة اختيار المتغيرات (Selection variables) المهمة والمؤثرة بالمتغير التابع باستخدام الاسلوب البيزي في الانحدار الخطي وبوجود بيانات عالية الابعاد استناداً الى استراتيجية الانقسام والدمج . حيث كان النهج المقترح يتكون من مرحلتين ، المرحلة الاولى تقسيم البيانات فائقة الابعاد الى عدد من المجموعات الفرعية ( Subset ) ذات الابعاد المنخفضة وتحديد المتغيرات ذات الصلة ( المهمة ) من كل مجموعة فرعية ، اما المرحلة الثانية فانها تتضمن تجميع المتغيرات المحددة من كل مجموعة فرعية ثم تحديد المتغيرات المهمة من المتغيرات التي تم الحصول عليها من المرحلة الاولى .

في عام 2015 قدم الباحثين<sup>(43)</sup> ( Song & Liang ) بحثاً يتناول دراسة تقدير نموذج الانحدار باستخدام دالة جزاء معكوس لاسو ( Reciprocal Lasso ) وذلك عن طريق تصغير المعادلة التالية :

$$\min_{\beta} \sum_{i=1}^n (y_i - x_i' \beta)^2 + \lambda \sum_{k=1}^p \frac{1}{|\beta_k|} , \quad I(\beta_k \neq 0)$$

حيث ان  $I(.)$  تمثل دالة المؤشر . و  $0 \leq \lambda$  تمثل معلمة الجزاء .

في عام 2017 قدم الباحث<sup>(7)</sup> (Alhousseini , F) بحثاً يتضمن عملية تقدير معالم نموذج الانحدار الكمي البيزي وبافتراض التمثيل المختلط للتوزيع الاولي لمعالم النموذج الذي يتبع توزيع لابلاس كخليط من التوزيع المنتظم (Uniform distribution) مع توزيع كاما بمعالم  $(2, \lambda)$  ، حيث اقترح الباحث نموذج للانحدار الكمي البيزي مقترناً بدالة جزاء لاسو (Lasso) لغرض اجراء اختيار المتغيرات التنبؤية المهمة عبر تطوير خوارزمية Gibbs sampler . تم تطبيق النموذج المقترح باستخدام المحاكاة وتحليل البيانات الحقيقية ، حيث اشارة النتائج بتفوق اداء النموذج المقترح عند مقارنتها مع الطرائق الاخرى .

في عام 2017 قدم الباحث<sup>(6)</sup> (Alhousseini , F) بحثاً يتضمن تقدير واختيار المتغيرات التنبؤية المهمة باستخدام الاسلوب البيزي مع تقنية الضبط (Lasso) ، وتم اعتماده على التمثيل المختلط لتوزيع لابلاس (SMU) المقترح من قبل (Mallick & Yi. 2014)<sup>(33)</sup> كتوزيع اولي لمعالم النموذج . وبالاعتماد عليه تم انشاء النموذج الهرمي السابق للمعالم وبشكل اكثر تطوراً ، كما تم ايضا استخدام التمثيل المختلط لتوزيع الخطأ (AL) والمقترح من قبل الباحثين<sup>(20)</sup> (Kozumi & Kobayshi) . وباستخدام خوارزمية (Gibbs sampler) اظهرت النتائج الاداء الافضل لطريق البحث عند مقارنته مع الطرق الاعتيادية .

في عام 2018 قدم الباحث<sup>(8)</sup> (Alhousseini , F) بحثاً تناول فيه الاسوب البيزي مقترناً بتقنية الضبط لاسو (Lasso) في تقدير نموذج الانحدار الكمي لتوبت المركب (Composite Tobit quantile regression). ان نموذج الانحدار المقترح يدرس الانحدار الكمي المركب استنادا الى نظرية بيز لاختيار المتغيرات التنبؤية المؤثرة على متغير الاستجابة . حيث اقترح الباحث خوارزمية MCMC عبر اشتقاق صيغة للتوزيع اللاحق معتمداً على التمثيل المختلط من التوزيع الطبيعي (Normal distribution) مع دالة الكثافة الاحتمالية الاسية (Exponential distribution) ، وتم تطبيق النموذج المقترح من قبل الباحث في جانب المحاكاة والبيانات الحقيقية واظهرت النتائج كفاءة تقديرات المقدرات المدروسة .

في عام 2020 قدم الباحث<sup>(15)</sup> (Flaih et al) ورقة بحثية مهمة بدراسة اختيار المتغيرات التوضيحية المهمة لنموذج الانحدار الخطي المتعدد المقترن بدالة جزاء لاسو (Lasso) وفق الطريقة البيزية بناءً على اقتراح ان توزيع لابلاس (LD) هو التوزيع المسبق لمعالم النموذج ، والذي تم تمثيله بخليط من التوزيع الطبيعي مع دالة الكثافة الاحتمالية لتوزيع رايلي وكما في المعادلة التالية :

$$\frac{\lambda}{2\sigma^2} \exp\left[\frac{-\lambda|\beta|}{\sigma^2}\right] = \int_0^{\infty} \frac{1}{\sqrt{2\pi\sigma^2\tau}} \exp\left[-\frac{\beta^2}{2\sigma^2\tau}\right] \frac{\lambda}{2} \exp\left(-\frac{\lambda}{2}\tau\right) \tau d\tau$$

وبالاعتماد على ذلك التمثيل تم اشتقاق النموذج الهرمي (hierarchical model) للتوزيعات الاولية للمعالم المدروسة في النموذج واشتقاق التوزيعات اللاحقة . وتم تطبيق النموذج المقترح في جانب المحاكاة وتحليل البيانات الحقيقية واطهرت النتائج كفاءة تقديرات المقدرات المدروسة .

في عام 2020 اقترح الباحثين<sup>(5)</sup> (Alhamzawi, R., & Mallick, H) طريقة جديدة لتقدير نموذج الانحدار الكمي البيزي باستخدام دالة جزاء معكوس لاسو (Reciprocal Lasso) ، وذلك عن طريق تصغير المعادلة التالية :

$$\min_{\beta} \sum_{i=1}^n \rho_{\alpha}(y_i - x_i'\beta) + \lambda \sum_{k=1}^p \frac{1}{|\beta_k|} , \quad I(\beta_k \neq 0)$$

مفترضين ان التوزيع الاولي لمعلمات النموذج هو توزيع معكوس لابلاس (inverse Laplace distributio) ، حيث قدمت هذه الدراسة في الانحدار الكمي من منظور بيزي . وتم تطوير خوارزمية (Gibbs sampler) السهلة والفعالة للاستدلال اللاحق وباستخدام (a scale mixture of inverse uniforms) ، وايضا قام الباحثان بنشر هذه الخوارزمية لحل مشكلة معكوس لاسو وتوسيعها لتشمل ايضا طرق اخرى ذات صلة مثل reciprocal adaptive LASSO, reciprocal bridge, and reciprocal adaptive bridge . اظهرت نتائج دراسة المحاكاة والبيانات الحقيقية الى ان الطريقة المقترحة تؤدي أداءً جيداً مقارنةً بطرق اخرى

في عام 2021 قدم الباحثون<sup>(36)</sup> (Mallick, H., Alhamzawi, R., Paul, E., & Svetnik, V) بحثاً يتضمن طريقة بيزية جديدة باستخدام دالة جزاء معكوس لاسو في تقدير معالم نموذج الانحدار الاعتيادي وذلك عن طريق تقليل المعادلة الرياضية التالية :

$$Q(\beta) = \min_{\beta} (Y - X\beta)'(Y - X\beta) + \lambda \sum_{k=1}^p \frac{1}{|\beta_k|}, \quad I(\beta_k \neq 0)$$

حيث ان  $\lambda \geq 0$  وتمثل معلمة الجزاء .

وافترض الباحثون ان التوزيع الاولي لمعالم نموذج الانحدار هو توزيع معكوس لابلاس (ALD) وكما في الصيغة الرياضية الآتية :

$$\pi(\beta) = \prod_{j=1}^p \frac{\lambda}{2\beta_j^2} \exp\left(-\frac{\lambda}{|\beta_j|}\right), \quad I(\beta_j \neq 0)$$

واقترح الباحثون تمثيلاً مختلطاً لهذا التوزيع هو ( a scale mixture of truncated normal) من اجل تسهيل عمل الخوارزمية ، وتمت عملية التقدير من خلال خوارزمية (Gibbs Sampler) حيث اظهرت نتائج المحاكاة وتحليل البيانات الحقيقية بتفوق أداء الطريقة البيزية المقترحة في تقدير معالم النموذج وباستخدام ( a scale mixture of truncated normal) مقارنة مع الطرق الكلاسيكية في التقدير والتبؤ واختيار المتغيرات .

# الفصل الثاني

الجانب النظري



## 1-2 المقدمة ( Introduction ):-

لتحسين دقة التنبؤ في انحدار المربعات الصغرى الاعتيادية (OLS) في ظل وجود مشكلة التعدد الخطي ومشكلة وجود متغيرات متنبا بها أكبر من حجم العينة ، تم فرض دوال جزائية على مجموع مربعات الخطأ (دالة الخسارة) لغرض الحصول على دقة تنبؤية افضل ، ومن هذه الدوال دالة انحدار (Rigde) التي اقترحها (Kennard 1970 Hoerl and)<sup>(19)</sup> وكما في الصيغة الرياضية التالية

$$Q(\beta) = (y - X\beta)'(y - X\beta) + \lambda \|\beta\|_2^2 ; \lambda \geq 0$$

والتي تعمل على تصغير (Residual sum of squares , RSS) استناداً الى القيد  $\sum_{j=1}^p |\beta_j|^2 \leq t$  وبالتالي فان انحدار ريدج (Rigde regression) يعطي دقة تنبؤية افضل عن طريق انكماش (shrinking) او تقليص معاملات انحدار (OLS) ولكن هذه الطريقة لاتعمل بألية تحديد او اختيار المتغيرات (variable selection) بمعنى انها تحتفظ بجميع المتغيرات التفسيرية . وايضا قدم كل من الباحثان (Frank and Friedman , 1993)<sup>(14)</sup> طريقة تحسين لـ (OLS) والتي تسمى بانحدار بريدج (bridge regression) استناداً الى القيد  $\sum_{j=1}^p |\beta_j|^q \leq t$  حيث ان  $q \geq 0$  وان هذه الطريقة تعتبر الحالة العامة والمتمثلة في الصيغة التالية:

$$Q(\beta) = (y - X\beta)'(y - X\beta) + \lambda \sum_{j=1}^p |\beta_j|^q ; \lambda \geq 0$$

وعندما تكون  $q = 2$  هي حالة خاصة والتي تعبر عن انحدار ريدج .

ومن بين التطويرات او التحسينات الاخرى اقترح كل من (Fan and Li , 2001)<sup>(13)</sup> طريقة تقدير جزائية اخرى تسمى ( , the Smoothly Clipped Absolute Deviation penalty ) وايضا قدم الباحث (Zhang , 2010)<sup>(48)</sup> طريقة تحسين جزائية تسمى ( the SCAD ) وتعتبر تقديرات كل من الطريقتين أعلاه هي تقديرات متسقة ، ومستمرة وذات حلول صفرية (sparse and continuous)<sup>(13,48)</sup> .

ومن بين تقنيات طرائق الانحدار الجزائية ، وربما تكون الاكثر استخداماً في الاديبيات الاحصائية هي طريقة انحدار لاسو (Lasso regression) وهي تعتبر حالة خاصة من انحدار بريدج

(brigde regression) عندما ( $q = 1$ ) وان تقديرات معالم النموذج نحصل عليها من خلال تصغير المعادلة التالية<sup>(45)</sup> :

$$Q(\beta) = (y - X\beta)'(y - X\beta) + \lambda \|\beta\|_1 ; \lambda \geq 0$$

وبمقارنة انحدار ريدج (Rigde regression) مع انحدار لاسو (Lasso regression) نجد ان الاخير يتمتع بخاصية رائعة وهي انكماش (shrink) بعض المعاملات الى الصفر تماماً ، مما يسهل من عملية الاختيار التلقائي للمتغيرات (variable selection) ، مما يعني الحصول على نموذج يمتلك خصائص جيدة للمعالم المقدرة وبالتالي زيادة القوة التفسيرية للنموذج وكذلك زيادة دقة التنبؤ .

وفي الأونة الاخير تم انجاز الكثير من الاعمال من قبل الباحثين في الاتجاه البيزي وذلك بالاعتماد على الفكرة التي طرحها الباحث (Tibshirani , 1996)<sup>(45)</sup> حيث اقترح ان تقديرات لاسو بإمكانها ان تفسر توزيع المعاينة للتوزيع اللاحق عندما تكون معاملات الانحدار مستقلة ومتطابقة التوزيع (independent and identical distribution, i.i.d) وتمتلك توزيع لابلاس (Laplace distribution) .

وفي هذا الفصل سوف نتطرق ايضا عن نموذج الانحدار الكمي الثنائي وطريقة معكوس لاسو المقترنة مع الاسلوب البيزي في تقدير معالم هذا النموذج وباقتراض ان الاخطاء العشوائية لها توزيع لابلاس الملتوي (ALD) والتوزيع السابق لمعالم النموذج يكون توزيع معكوس لابلاس (Inverse Laplace distribution) وكذلك تم التطرق الى بعض التوزيعات المختلطة التي تمثل توزيع لابلاس وذلك لحاجة الباحث لبعض من هذه التوزيعات المختلطة في تبسيط عمليات اجراء المعاينة من التوزيعات اللاحقة .

ونظراً للحاجة لاستخدام توزيع لابلاس الملتوي في هذه الاطروحة سنتطرق الى هذا التوزيع وكما موضح في الفقرة التالية .

## **2-2 توزيع لابلاس (Laplace distribution)<sup>(42)</sup> :**

يعتبر العالم الفرنسي ( Pierre-Simon Laplace ) هو اول من اقترح هذا التوزيع في عام (1774) وهو يعتبر من التوزيعات الاحتمالية المستمرة .

ان لهذا التوزيع وبأنواعه (الملتوي والمتماثلة) وقوانينه الرياضية المختلفة تطبيقات في كثير من البحوث العلمية وكذلك في مختلف الدراسات والابحاث الاقتصادية والطبية والهندسية وغيرها ، كما ان لهذا التوزيع درجة عالية من الاهمية لدى الكثير من الباحثين وخصوصاً في الدراسات والبحوث الحصينة والذي يعتبر كبديل للتوزيع الطبيعي (Normal distribution) .

ويعرف توزيع لابلاس (Laplace distribution) كما في الصيغة الرياضية الآتي<sup>(38)</sup>:

$$f(x|\theta, \tau) = \frac{1}{2\tau} \exp\left(-\frac{|x-\theta|}{\tau}\right) \dots \dots \dots (2 - 1)$$

$$-\infty < x < \infty \quad ; \quad -\infty < \theta < \infty \quad ; \quad \tau > 0$$

حيث ان:

$\theta$  : تمثل معلمة الموقع (Location parameter)

$\tau$  : تمثل معلمة القياس (Scale parameter)

ويعرف توزيع لابلاس الملتوي (Asymmetric Laplace distribution) كما في الصيغة الرياضية الآتية<sup>(38)</sup>:

$$f_p(x) = p(1 - p)\exp\{-\rho_p(x)\}$$

حيث ان الدالة  $\rho_p(x)$  يمكن التعبير عنها بالصيغة التالية<sup>(38)</sup>

$$\rho_p(x) = \begin{cases} px & \text{if } x \geq 0 \\ -(1-p)x & \text{if } x < 0 \end{cases}$$

### **3-2 : اختيار المتغيرات<sup>(3، 45)</sup> ( Selection variables )**

في العديد من التطبيقات، اصبح اختيار المتغيرات المهمة من النقاط الاساسية في بناء النموذج، وهي تمثل عملية اختيار مجموعة فرعية من المتغيرات التوضيحية المهمة لاستخدامها في بناء نموذج الانحدار، حيث انها توفر نموذجاً تنبؤي جيد بالاضافة الى ابراز المتغيرات المهمة في ملائمة النموذج<sup>(17)</sup> .

ان الافتراض الرئيس عند استخدام (Selection variables) بأن البيانات تحتوي على العديد من المتغيرات الغير مهمة ، لذلك هناك هدفان رئيسيان يجب التمييز بينهما عند انشاء نموذج

الانحدار . الهدف الاول هو انشاء نموذج انحدار يمتلك القدرة على التنبؤ بشكل جيد وذات قوة تفسيرية عالية ، اما الهدف الثاني يتمثل بمحاولة تحديد المتغيرات التوضيحية المؤثرة والحصول على نظرة ثاقبة حول الصلة بين المتغيرات التوضيحية ومتغير الاستجابة عبر هيكل نموذج الانحدار .

ان عملية اختيار المتغيرات المهمة تعتبر احد المتطلبات المهمة عند بناء نموذج الانحدار ، لذلك فان استبعاد المتغيرات المستقلة الغير مهمة تكون من مصلحة النموذج مع الاحتفاظ بالمتغيرات المهمة في النموذج وهذا يحسن الدقة التنبؤية له . ومن التقنيات المهمة في استخدام هذه العملية (عملية اختيار المتغير ) في تقدير نماذج الانحدار هي تقنية لاسو ( Lasso ) التي تم اقتراحها من قبل الباحث<sup>(45)</sup> ( Tibshirani ; 1996 ) والتي حصلت على اهتماماً واسعاً في السنوات الاخيرة ، حيث تعمل على اجبار معاملات المتغيرات الغير مهمة مساوية الى الصفر تماماً وتحفظ بالمتغيرات المهمة .

ونذكر بعض الاسباب المهمة والمتعلقة في تحديد المتغيرات الاساسية ومنها :

- عادة مايكون للتخلص من مشكلة الابعاد العالية (high dimensional) .
- التحسين في دقة التنبؤ (Improving prediction accuracy) .
- لعدم الوقوع في مشكلة (Over fitting) والتي بدورها تؤثر على دقة التنبؤ .
- بناء نموذج يتضمن اقل عدد من المتغيرات المستقلة (المتغيرات المهمة ) وبالتالي يكون قابل للتفسير بسهولة بمعنى انه بالامكان معرفة تأثير كل متغير من المتغيرات التوضيحية على المتغير التابع بشكل منفصل عن تأثيرات المتغيرات الأخرى

## 4-2 مشكلة التعدد الخطي (Multicollinearity) (18، 16، 9)

عادةً ما تحدث مشكلة الارتباط الخطي (Multicollinearity) في نماذج الانحدار الخطي المتعدد ، اي في النماذج التي تحتوي على أكثر من متغير مستقل ولذلك فهي من المشاكل التي تختص بالمتغيرات المستقلة .

كثيراً ما تواجه الباحثين في العديد من الدراسات والبحوث هذه المشكلة (Multicollinearity) نتيجةً لارتباط بعض من هذه المتغيرات ( المتغيرات المستقلة ) فيما بينها او جميعها تكون مرتبطة بقيم ارتباط عاليه ، وبالتالي يكون من الصعب فصل تأثير كل متغير مستقل على المتغير التابع . ومن الاسباب الاخرى ايضاً لحدوث هذه المشكلة عندما تكون قيم احد هذه المتغيرات

المستقلة متساوية لجميع المشاهدات او اعتماد قيمة بعض من هذه المتغيرات على قيمة او اكثر من المتغيرات الاخرى في النموذج ، ولذلك عند تقدير معالم نموذج الانحدار بوجود احد هذه الاسباب يؤدي الى تقديرات متحيزة وغير كفوءة وذات تباينات عالية لمعاملات الانحدار ، وبالتالي تكون القوة التفسيرية لنموذج الانحدار التنبؤي ضعيفة وغير دقيقة .

ان مشكلة التعدد الخطي بين المتغيرات التوضيحية تصنف الى نوعين النوع الاول هو التعدد الخطي التام حيث تحدث هذه المشكلة عندما تكون المصفوفة  $(X'X)$  ذات رتبة غير كاملة ولذلك تحت هذا النوع لايمكن ايجاد مقدرات لمعالم نموذج الانحدار باستخدام طريقة المربعات الصغرى  $\hat{\beta} = (X'X)^{-1}X'Y$  بسبب ان محدد المصفوفة  $(X'X)$  يكون مساوياً للصفر وبالتالي لايمكن ايجاد معكوس المصفوفة  $(X'X)$  ، اما النوع الثاني يسمى بالتعدد الخطي الغير تام وهي الاكثر مصادفة في الدراسات التطبيقية وتحدث هذه المشكلة عندما يكون احد المتغيرات المستقلة دالة بالمتغيرات الاخرى ، او يكون عدد المتغيرات التوضيحية اكبر من عدد المشاهدات  $(p > n)$  وعندها يكون محدد المصفوفة  $(X'X)$  قريباً من الصفر مما يؤدي الى تضخم في تباين المقدرات التنبؤية لمعالم نموذج الانحدار وبالتالي تكون طريقة (OLS) غير مجدية . ولذلك يكون النموذج غير قادر على اظهار اثر المتغيرات المستقلة بشكل منفصل على المتغير المعتمد بسبب الاعتمادية او الارتباط فيما بينهما .

ولذلك تم ايجاد طرق مختلفة لتحسين كفاءة طريقة المربعات الصغرى (OLS) عندما تكون نتائجها غير مرضية ومن هذه الطرق طريقة ( stepwise ) والتي تعمل على حذف المتغيرات الغير مهمة وتحفظ بالمتغيرات المهمة في النموذج ، اما الاسلوب الآخر هو اضافة حد الجزاء الى مجموع مربعات الخطأ ، وهناك طرق اخرى تعمل على حذف بعض المتغيرات الغير مهمة من خلال جعل معاملاتها تساوى الى الصفر تماماً حيث تعتبر هذه الطريقة من الطرق التي يتم من خلالها اختيار المتغيرات<sup>(45)</sup> (variable selection) حيث يتم حساب قيمة المقدر من الصيغة التالية :

$$\hat{\beta} = \operatorname{argmin} \sum_{i=1}^n (y_i - x_i' \beta)^2 + \lambda * (\text{دالة الجزاء})$$

ومن الباحثين الذين افترضوا دالة الجزاء منهم الباحث<sup>(45)</sup> (Tibshirani , 1996) والتي تسمى بطريقة لاسو (Lasso) وكما في الصيغة الرياضية التالية :

$$\hat{\beta} = \min_{\beta} \sum_{i=1}^n (y_i - x_i' \beta)^2 + \lambda \sum_{k=1}^p |\beta_k|$$

كذلك قام الباحثان<sup>(21)</sup> (Zou & Hastie ; 2005) بتطوير طريقة لاختيار المتغيرات والتي تسمى بطريقة (elastic net) . وايضا تم اقتراح طريقة (adaptive lasso) لاختيار المتغيرات من قبل الباحث<sup>(47)</sup> Zou عام 2006

## 5-2 : الانحدار الكمي (Quantile Regression)

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

وبفرض ان نموذج الانحدار الخطي الاعتيادي والمعطى حسب الصيغة الرياضية التالية<sup>(33)</sup> :

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

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

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

$\varepsilon$  : 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:

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

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 مقدرًا 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 شاذة 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 للانحدار المتوسط الخطي القياسي<sup>(22)</sup> حيث انها لا تفترض بتوزيع الخطأ العشوائي توزيعاً طبيعياً على عكس النماذج الخطية وايضا لا تتأثر 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 توفر حلول اكثر حصانة في كثير من التطبيقات الحقيقية، ولهذه الاسباب فان نموذج الانحدار الكمي يعتبر من النماذج الحصينة . وبفرض ان  $\{(x_i, y_i); i = 1, 2, \dots, n\}$  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<sup>(23)</sup> :

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

وان  $\beta_p$  تمثل متجه المعلمات عند الكمية  $(p)$  ،  $0 < p < 1$  ،

وبافتراض ان توزيع الخطأ يكون توزيع (AL) اي ان<sup>(23)</sup> :

$$f_p(\varepsilon_i) = p(1 - p)\exp\{-\rho_p(\varepsilon_i)\} \dots \dots \dots (2 - 3)$$

$\varepsilon_i$  : تمثل حد الخطأ العشوائي (البواقي) والتي لها توزيع مقيد<sup>(20)</sup>

$$:\int_{-\infty}^0 f_p(\varepsilon_i) d\varepsilon_i = p$$

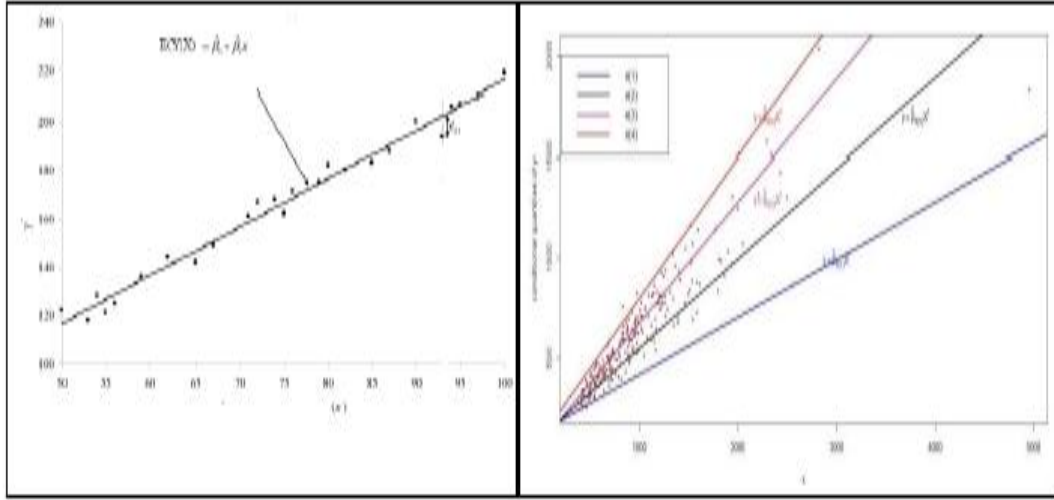
$$Or F_{\varepsilon}^{-1}(p) = 0$$

وان عملية تقدير معالم نموذج الانحدار الكمي  $\beta_p$  يكون بتصغير دالة الخسارة ( Loss

(function or check function) التالية :

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

الشكل (1-2) يمثل خطوط الانحدار التي يتم تقديرها عبر نموذج الانحدار الكلاسيكي ونموذج الانحدار الكمي (QR)



ومن ملاحظة الشكل اعلاه رقم (1-2) نلاحظ ان الشكل الايسر من الرسم يوضح خط الانحدار المقدر بطريقة نموذج الانحدار الكلاسيكي ، ولكن في بعض الاحيان هذا النموذج لا يمكن ان يوفر لنا معلومات كاملة حول العلاقة بين المتغير التابع والمتغيرات التوضيحية لوجود احد الاسباب التي تم التطرق اليها اعلاه ، بينما الانحدار الكمي يعطينا صورته واضحة عن العلاقة بين متغير الاستجابة والمتغيرات التوضيحية لانه يتم تقدير العديد من خطوط الانحدار وعند كميات (Quantile) مختلفة وكما موضح في الشكل الايمن من الرسم اعلاه حيث هناك اربعة خطوط مقدره بطريقة الانحدار الكمي (QR) عند اربعة مستويات مختلفة .

## **6-2 : استخدام طريقة (Lasso) في الانحدار الخطي<sup>(45)</sup> :**

اقترح الباحث (Tibshirani ; 1996) طريقة تقدير في نماذج الانحدار الخطي والتي تعرف بطريقة (Lasso) في التقدير وهي تعني مختصر لـ (Least absolute shrinkage and selection operator) والتي حصلت على اهتمامات كبيرة في السنوات الاخيرة والتي بدورها تعتبر كطريقة بديلة لطريقة المربعات الصغرى في حالة وجود بعض المشاكل على سبيل المثال عندما يكون عدد المتغيرات كبير او عدد المتغيرات اكبر من عدد المشاهدات ( $p > n$ ) ، وبوجود هذه المشاكل فان طريقة المربعات الصغرى تكون نتائجها غير دقيقة وتقديراتها تكون متحيزة وذات تباينات عالية لذا فان طريقة (Lasso) تعتبر من الطرق المهمة في معالجة هذه



المشاكل حيث تعطي تقديرات بتحيز قليل ولكن بتباينات قليلة . وان دالة الخسارة باستخدام طريقة (Lasso) يكون بتصغير الصيغة الرياضية التالية :

$$\min_{\beta} \sum_{i=1}^n (y_i - x_i' \beta)^2 + \lambda \sum_{k=1}^p |\beta_k|$$

حيث ان :  $\lambda \geq 0$  وهي تمثل معلمة الضبط (Regularization) والتي تعمل على الموازنة ما بين التحيز والتباين للمعاملات .

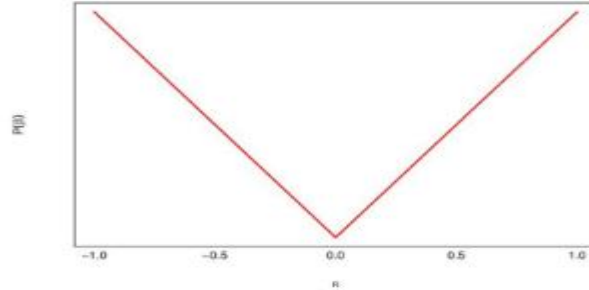
### خصائص دالة جزاء (Lasso) <sup>(36,34)</sup>

- دالة متماثلة (Symmetric) حول الصفر.
- دالة مستمرة (Continuous) و متزايدة (increasing) في الفترة  $(0, \infty)$ .
- معاملات الانحدار الكبيرة تحصل على جزاءات كبيرة والعكس بالعكس وذلك لتنظيم او تقييد المعاملات .
- الكفاءة الحسابية (Computationally efficient).
- انكماش او تقليص معاملات الانحدار للمتغيرات المستقلة الغير مهمة في الدراسة الى الصفر تماماً وبالتالي يتم تلقائياً اختيار المتغيرات المهمة والمؤثرة في متغير الاستجابة (Response variable) فقط .
- حصلت كثير من الاضافات (extensions) والتحسينات (improvements) على مر السنين لايجاد طرائق اخرى في تقدير نماذج الانحدار عندما تتوقف عندها طريقة المربعات الصغرى مثل (elastic net , adaptive lasso , MCP , scad) ولكن كل هذه الطرق متماثلة من حيث الاساس الرياضي .

الشكل رقم (2-2) يوضح رسم دالة جزاء لاسو

$$Q(\beta) = (y - X\beta)'(y - X\beta) + \lambda \|\beta\|_1$$

### Least Absolute Shrinkage and Selection Operator



ولتقدير معالم نموذج الانحدار الكمي عند استخدام طريقة لاسو يكون من خلال تصغير المعادلة التالية وكما يلي (39) :

$$Q(B) = \min_{\beta} \sum_{i=1}^n \rho_p(y_i - x'_i \beta_p) + \lambda \sum_{j=1}^k |\beta_j|$$

### 7-2: استخدام طريقة معكوس (Lasso) في الانحدار الخطي :

تم تقديم بعض الاقتراحات والتحسينات من قبل مجموعة من الباحثين على مرالسنين (39,41,34,13,12) Fan and Efor et al.2004; Li,2001;Meinshausen,2007;Radchenko and James,2008;song and Laing,2015, among others) على طريقة تقدير (Lasso) المقترحة من قبل الباحث (45) Tibshirani ; 1996 في تقدير نماذج الانحدار الخطي . وفي السنوات الاخيرة قدم كل من الباحثان (39) (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 :

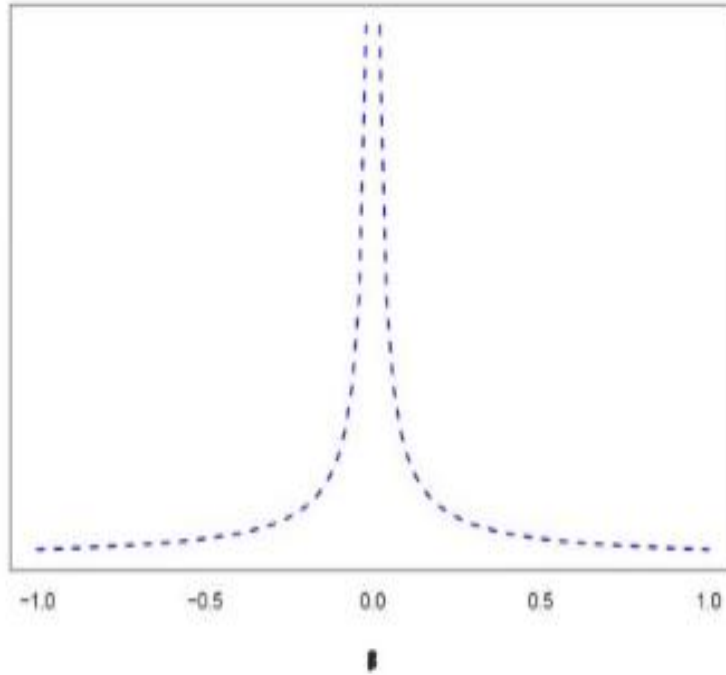
$$\min_{\beta} \sum_{i=1}^n (y_i - x'_i \beta)^2 + \lambda \sum_{j=1}^k \frac{1}{|\beta_j|} I(\beta_j \neq 0) \dots \dots \dots (2 - 4)$$

### خصائص دالة جزاء معكوس لاسو (rLasso)<sup>(36)</sup>

- 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.

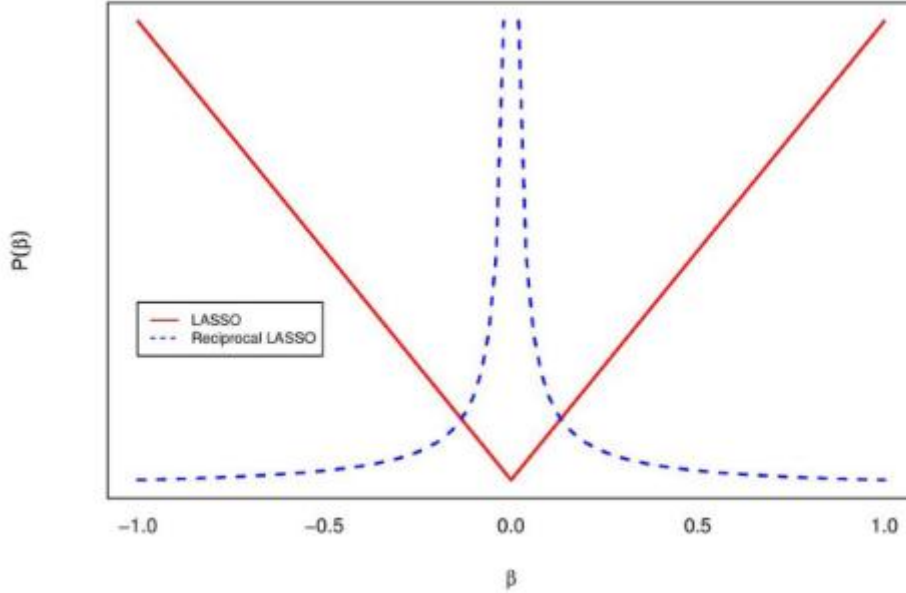
الشكل رقم (2-3) يوضح رسم دالة جزاء معكوس لاسو

$$Q(\beta) = \min_{\beta} (y - X\beta)'(y - X\beta) + \lambda \sum_{j=1}^p \frac{1}{|\beta_j|} I\{\beta_j \neq 0\}.$$



الشكل رقم (2-4) يوضح رسم دالة جزاء لاسو ذو اللون الاحمر مع دالة جزاء معكوس لاسو ذو اللون الازرق

## LASSO and Reciprocal LASSO



ولتقدير معالم نموذج الانحدار الكمي عند استخدام معكوس لاسو يكون من خلال تصغير المعادلة التالية وكما يلي<sup>(39)</sup> :

$$Q(B) = \min_{\beta} \sum_{i=1}^n \rho_p(y_i - x_i' \beta_p) + \lambda \sum_{j=1}^k \frac{1}{|\beta_j|} \dots \dots \dots (2 - 5)$$

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

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

$y_i$  : يمثل تمثّل قيمة المشاهدة الحقيقية للمتغير الذي ترتيبه  $i$

$x_i$  : متجه من درجة  $(n \times p)$

$\beta_p$  : متجه معالم النموذج من درجة  $(p \times 1)$

ومن الجدير بالذكر ان كثير من الباحثين استخدموا التمثيل المختلط لتوزيع لابلاس الملتوي (AL) عند اقترانه بالطريقة البيزية وذلك لما له من اهمية كبيرة في تسهيل العمليات الحسابية في التوزيعات اللاحقة لتقدير المعالم ، وقد اقترح كل من الباحثان ( Kozumi and Kobayashi ) تمثيلاً مختلطاً لهذا التوزيع وكما مبين في ادناه<sup>(20)</sup> :

بفرض ان  $u$  متغير عشوائي يتوزع توزيع طبيعي قياسي اي ان :

$$f(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) ; \quad -\infty < u < \infty \dots\dots\dots (2 - 6)$$

وبفرض ان  $u$  متغير عشوائي يتوزع توزيع اسي قياسي اي ان :

$$f(z) = \exp(-z) ; \quad z > 0 \dots\dots\dots (2 - 7)$$

فاذا المتغير العشوائي ( $\varepsilon$ ) يتبع توزيع لابلاس الملتوي ( Asymmetric Laplace distribution ) بدالة كثافة احتمالية كما مشار اليها في المعادلة (2-3) ، فانه بالامكان تمثيله بـ ( Location-scale mixture of normal ) وكما يلي :

$$\varepsilon = \theta z + \tau\sqrt{z}u \dots\dots\dots (2 - 8)$$

$$\theta = \frac{1-2p}{p(1-p)} ; \quad \tau = 2/p(1-p) \quad \text{حيث ان :}$$

ومن النتيجة اعلاه يمكن اعادة صياغة نموذج الانحدار في المعادلة (2-11) وكما يلي :

$$y_i = x_i' B_p + \theta z_i + \tau\sqrt{z_i}u_i ; \quad i = 1, 2, \dots, n$$

$$z_i \sim \varepsilon(1) \quad \text{and} \quad u_i \sim N(0,1) \quad \text{حيث ان :}$$

ولذلك فان التوزيع الشرطي لمتغير الاستجابة ( $y_i$ ) بمعلومية المتغير ( $z_i$ ) يكون توزيعاً طبيعياً بمتوسط مقداره  $x_i' B_p + \theta z_i$  وتباين  $\tau^2 z_i$  ، وكما في المعادلة التالية :

$$f(y/B_p, z) \propto \left( \prod_{i=1}^n z_i^{-\frac{1}{2}} \right) \exp\left( -\sum_{i=1}^n \frac{(y_i - x_i' B_p + \theta z_i)^2}{2\tau^2 z_i} \right) \dots (2 - 9)$$

$$y = (y_1, y_2, \dots, y_n)' ; \quad z = (z_1, z_2, \dots, z_n)'$$

حيث ان :

## 8-2 الانحدار الكمي البيزي الثنائي باستخدام معكوس لاسو ( Binary reciprocal lasso )

في هذا الفصل نقوم بتوسيع الانحدار الكمي البيزي الذي اشار اليه الباحث<sup>(31)</sup> 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<sup>(11 و 29)</sup> . وان الصيغة الرياضية لنموذج الانحدار الكمي الثنائي هي :

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

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

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

وان تقدير معاملات نموذج الانحدار الكمي باستخدام معكوس لاسو ( rlasso ) يكون حسب المعادلة المرقمة ( 5 – 2 ) .

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

اقترح الباحث<sup>(36)</sup> Mallick وآخرون في سنة 2021 مقترحاً لتمثيل التوزيع السابق ( prior ) للمعلمة  $\beta$  ، حيث كان هذا المقترح يفترض ان توزيع معكوس لابلاس للمعلمة  $\beta$  يمكن تمثيله بالشكل الآتي :

$$\frac{\lambda}{2\beta^2} e^{-\frac{\lambda}{|\beta|}} = \int_0^{\infty} \frac{\eta}{2\beta^2} I\{|\beta| > \eta\} \frac{\lambda^2}{\Gamma 2} \eta^{-2-1} e^{-\frac{\lambda}{\eta}} d\eta \dots \dots \dots (2 - 10)$$

في هذه الإطروحة ولتسهيل العمليات الحسابية لخوارزمية Gibbs sampler تم اجراء التحويل التالي على التمثيل المذكور في المعادلة (10 - 2) :

لنفرض ان

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

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

فاذا كان  $Z = 0 \iff \eta = \infty$  ، واذا كانت  $Z = \infty \iff \eta = 0$

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

$$\begin{aligned} \frac{\lambda}{2\beta^2} e^{-\frac{\lambda}{|\beta|}} &= \int_0^{\infty} \frac{\lambda}{2\beta^2} I\left\{|\beta| > \frac{\lambda}{Z}\right\} \cdot \frac{\lambda^2}{\Gamma(2)} \left(\frac{\lambda}{Z}\right)^{-2-1} \cdot e^{-Z} \lambda \left(\frac{-1}{Z^2}\right) dz \\ &= \int_0^{\infty} \frac{\lambda}{2\beta^2} I\left\{|\beta| > \frac{\lambda}{Z}\right\} \cdot e^{-Z} dz \end{aligned}$$

حيث ان:

$Z \sim \text{standard exponential}$

$$\beta \sim \text{Inverse uniform} \left( -\frac{1}{\lambda}, \frac{1}{\lambda} \right) \dots \dots \dots (2 - 11)$$

وهذا يعني ان :

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

يمتلك دالة الكثافة الاحتمالية الآتية :

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

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

وهو توزيع معكوس التوزيع المنتظم او يدعى بتوزيع باريتو المضاعف (double Pareto). وبالتالي من العلاقة (11 – 2) 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)<sup>(33)</sup> (Mallick and Yi, 2014).

## 9-2 النموذج الهرمي البيزي المقترح (Bayesian hierarchical model) للتوزيعات السابقة :

يمكن افتراض النموذج الهرمي للتوزيعات المسبقة كما يلي :

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

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

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

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

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

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

$$Z \sim \text{standard exponential}$$



## 10-2 التوزيعات الشرطية اللاحقة الكاملة ( The full conditional )

### : (posterior distribution)

وبالاعتماد على النموذج الهرمي للتوزيعات المسبقة المذكور في الفقرة (9-2) يمكن الآن اشتقاق التوزيعات اللاحقة posterior ولكل متغير وكما موضح في ادناه :

أولاً يجب كتابة التوزيع المشترك الكامل full joint لجميع المتغيرات :

$$\begin{aligned}
 & f(y^*|\beta, X, \sigma^2) \cdot \pi(\sigma^2) \cdot \pi(\lambda) \cdot \prod_{j=1}^p \pi(\beta_j|\lambda_j) \cdot \pi(z_j) I\left\{|\beta_j| > \frac{\lambda_j}{z_j}\right\} \\
 &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}(y^* - X\beta)'(y^* - X\beta)} * \frac{d^c}{\Gamma(c)} (\sigma^2)^{-c-1} e^{-\frac{d}{\sigma^2}} * \prod_{j=1}^p \frac{\lambda_j}{2\beta^2} \\
 & \quad * \frac{\theta^2}{\Gamma(2)} (\lambda)^{-2-1} e^{-\frac{\theta}{\lambda}} * \prod_{j=1}^p e^{-z_j} \\
 & \quad * \prod_{j=1}^p I\left\{|\beta_j| > \frac{\lambda_j}{z_j}\right\} \dots\dots\dots (2 - 12)
 \end{aligned}$$

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

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

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

2- التوزيع الشرطي اللاحق الكامل لـ  $(\beta)$  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  $\beta$  من التوزيع المشترك في المعادلة (2 - 12) ، وسوف نحذف اي جزء

لا يحتوي على  $\beta$  وبالإستعانة بالعلاقة التي اشرنا اليها من خلال الربط بين اسلوب لاسو واسوب معكوس لاسو، وسوف نحصل على :

$$\begin{aligned} \pi(\beta|y^*, X, \lambda, \sigma^2) &\propto \pi(y^*|X, \beta, \sigma^2) \cdot \pi(\beta|\lambda) \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}(y^* - X\beta)'(y^* - X\beta)\right\} \prod_{j=1}^p I\{|\beta_j| > \frac{\lambda_j}{Z_j}\} \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}[\beta'(X'X)\beta - 2y^*X\beta + y^*y^{*'}]\right\} \prod_{j=1}^p I\{|\beta_j| > \frac{\lambda_j}{Z_j}\} \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}[\beta'A\beta - 2y^*X\beta + y^*y^{*'}]\right\} \dots \dots \dots (2 - 13) \end{aligned}$$

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

$$\begin{aligned} &(\beta - A^{-1}X'y^*)'A(\beta - A^{-1}X'y^*) \\ &= \beta'A\beta - 2y^*X\beta + y^{*'}(XA^{-1}X')y^* \end{aligned}$$

وبالتالي يمكن كتابة العلاقة (2 - 13) بالشكل الآتي :

$$\begin{aligned} &\propto \exp\left\{-\frac{1}{2\sigma^2}[(\beta - A^{-1}X'y^*)'A(\beta - A^{-1}X'y^*)\right. \\ &\quad \left.+ y^{*'}(I_n - XA^{-1}X')y^*]\right\} \dots \dots \dots (2 - 14) \end{aligned}$$

ان الجزء الثاني من العلاقة (2 - 14) لا يحتوي على  $\beta$  لذلك فان التوزيع المشترك الذي يحتوي على  $\beta$  سيتم اختصاره بالشكل الآتي :

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

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  $\beta$  تملك توزيع طبيعي متعدد المتغيرات بمتوسط  $A^{-1}X'y^*$  وتباين  $\sigma^2 A^{-1}$ .

3- التوزيع الشرطي اللاحق الكامل لـ  $\sigma^2$  هو :

سيتم توليد عينات للمتغير  $\sigma^2$  باستخدام خوارزمية *Gibbs Sampler* من خلال اخذ جميع الاجزاء التي تضم  $\sigma^2$  في التوزيع المشترك (12 - 2) اي ان :

$$\begin{aligned} \pi(\sigma^2 | y^*, x, \beta) &\propto \pi(y^* | x, \beta, \sigma^2) \cdot \pi(\sigma^2) \\ &\propto \frac{1}{\sigma^{\frac{n-1}{2}}} \left[ -\frac{1}{2\sigma^2} (y^* - X\beta)'(y^* - X\beta) * \frac{d^c}{\Gamma(c)} (\sigma^2)^{-c-1} * e^{-\frac{d}{\sigma^2}} \right] \\ &\propto \sigma^{2 - \left(\frac{n-1}{2}\right) - c - 1} \exp\left[-\frac{1}{2\sigma^2} (y^* - X\beta)'(y^* - X\beta) - \frac{d}{\sigma^2}\right] \end{aligned} \quad \dots (2 - 15)$$

وبالاستعانة بتوزيع معكوس كما نجد ان :

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left\{-\frac{\beta}{x}\right\} \dots \dots \dots (2 - 16)$$

حيث  $\alpha$  تمثل معلمة الشكل و  $\beta$  تمثل معلمة القياس ، وباستخدام الدالة (16 - 2) مع العلاقة (15 - 2) يمكن الاستنتاج بان  $\sigma^2$  يمتلك توزيع معكوس كما بمعلمة شكل

$$\cdot (y^* - X\beta)'(y^* - X\beta) + d \text{ ومعلمة قياس } \left(\frac{n-1}{2} + c\right)$$

4- التوزيع الشرطي اللاحق الكامل لـ  $\lambda$  هو :

سيتم توليد عينات للمتغير  $\lambda$  من التوزيع المشترك في العلاقة (12 - 2) بعد اخذ الحدود التي تضم فقط المتغير  $\lambda$  وباستخدام خوارزمية *Gibbs sampler* وكما يلي :

$$\begin{aligned} \pi(\lambda | \theta, \beta) &\propto \pi(\beta | \lambda) * \pi(\lambda) \\ &\propto \prod_{j=1}^p \frac{\lambda_j}{2\beta^2} \frac{\theta^2}{\Gamma(2)} (\lambda)^{-2-1} e^{-\frac{\theta}{\lambda}} \cdot I\{|\beta_j| > \frac{\lambda_j}{Z_j}\} \\ &\propto \lambda^{-(p+2)-1} e^{-\frac{\theta}{\lambda}} \cdot I\{\lambda_j < Z_j |\beta_j|\} \dots \dots \dots (2 - 17) \end{aligned}$$

حيث يمكن القول ان العلاقة (17 - 2) انتجت توزيع المتغير  $\lambda$  ليكون

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

5- ان التوزيع اللاحق الشرطي لـ  $Z$  هو :

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

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

# الفصل الثالث

الجانب التجريبي

### 3- الجانب التجريبي :

#### 1-3 : المقدمة :

في هذا الفصل تم اجراء اسلوب المحاكاة لاختبار طريقتنا المقترحة ( الطريقة البيزية باستخدام التمثيل المختلط لتوزيع معكوس لابلاس Bayesian reciprocal Lasso binary quantile regression (BrLBqr) ) من قبل الباحث ومقارنتها مع طريقة الانحدار الكمي الثنائي البيزي (Bayesian binary) quantile regression (BBqr) وطريقة الانحدار الكمي الثنائي البيزي مع استخدام دالة جزاء لاسو ( Bayesian Lasso binary quantile regression (BLBqr) )<sup>(4)</sup> (Alhamzawi and Ali , 2020) في تقدير معالم نموذج الانحدار الكمي البيزي .

حيث تم استخدام اسلوب Markov Chain Monte Carlo (MCMC) في اجراء المعاينة من التوزيع الشرطي اللاحق ، ولقد تم استخدام خوارزمية (Gibbs Sampler) للمعاينة من قبل الباحث وفي برنامج لغة R . ولأجل توضيح ماتوصل اليه الباحث من اختبار للطريقة المقترحة وبمقارنتها مع الطرق التقديرية الاخرى ، حيث تم افتراض مجموعة من التوزيعات للخطأ العشوائي منها (  $\varepsilon \sim N(0,9)$  ،  $\varepsilon \sim t(3)$  ،  $\varepsilon \sim \chi^2_{(3)}$  ) ، وايضا لعدة متجهات مختلفة للمعلمات وفي (Quantile) مختلف للتوزيع الشرطي للمتغير التابع ( متغير الاستجابة ) وعند مستويين (p= 0.5 و p= 0.95) .

اما بالنسبة للمعايير التي تم استخدامها من اجل المقارنة بين الطرق المختلفة لايجاد الافضل منها هي :

1- Median of Mean Absolute Deviation (MMAD) : وهو يعتبر أحد المعايير المهمة والمستخدم في كثير من البحوث الحديثة للمقارنة بين مجموعة من طرائق التقدير من خلال استخدام الخطأ العشوائي ( خطأ التنبؤ )<sup>(40)</sup> .

$$MMAD = \text{median}(\text{mean} |x'_i \hat{\beta} - x'_i \beta|)$$

2- الانحراف المعياري للمعيار أعلاه Standard Deviation of Mean Absolute Deviation (Sd(MAD) Deviation) .

وتم توليد البيانات في تجارب المحاكاة استناداً الى نموذج الانحدار التالي<sup>(4)</sup> : ( Alhamzawi and Ali , 2020)

$$Y = X'\beta + \varepsilon$$

حيث ان المتغيرات المتنبأ بها تكون مرتبطة فيما بينها من خلال علاقة ارتباط باستخدام معامل الانحدار:

$$Cor(X_i , X_j) = \rho^{|i-j|}$$

علماً ان  $X \sim N(0, 1)$

طالما ان المتغيرات المتنبأ بها تمتلك تباين مقداره واحد فان مصفوفة مشاهدات

المتغير المتنبأ بها تتوزع توزيعاً طبيعياً بمتوسط صفر وتباين  $(\Sigma)$  حيث ان :

$$\Sigma_{ij} = \rho^{|i-j|}$$

الحالة الاولى (case I) :- مصفوفة احادية ذات حجم  $p$ . (Identity matrix of size  $p$ ).

الحالة الثانية (case II) :- مصفوفة متماثلة مركبة (Compound symmetry matrix)

بحيث ان :

$$\Sigma_{ij} = 0.5, \text{ if } i \neq j \text{ and } \Sigma_{ii} = 1, \text{ for } i = 1, \dots, n.$$

الحالة الثالثة (case III) :- مصفوفة الارتباط الانحدار الذاتي (Autoregressive

correlated matrix) بحيث ان

$$\Sigma_{ij} = 0.5^{|i-j|} \text{ for all } 1 \leq i \leq j \leq p$$

ولقد تم افتراض ثلاث متجهات حقيقية لمعالم الانحدار في تجارب المحاكاة والمستخدمه من

قبل الباحث (1996) Tibshirani. من اجل اختبار طريقتنا المقترحة (BrLBqr) عند مقارنتها

مع بقية الطرائق الاخرى (BBqr) و (BLBqr) والمقترحة من قبل (Alhamzawi and Ali , )

(2020) . حيث ان المتجهات الثلاثة المقترضة هي :

• المحاكاة الاولى (simulation 1) :  $\beta = (2,1,0,0,3,0,0,0,0,0)$  والذي يسمى

بمتجه المعلمات بـ sparse case وهذا يعني :

$$f(x) = \beta_1 x_1 + \beta_2 x_2 + \beta_5 x_5$$

• المحاكاة الثانية (simulation 2) :  $\beta = (3,0,0,0,0,0,0,0,0,0)$  والذي يسمى

بمتجه المعلمات بـ highly sparse case ، اي ان :

$$f(x) = \beta_1 x_1$$

• المحاكاة الثالثة (simulation 3) :  $\beta = (1,1,1,1,1,1,1,1,1,1)$  والذي يسمى

بمتجه المعلمات بـ dense case ، وذلك يعني :

$$f(x) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10}$$

### 2-3 توليد الخطأ العشوائي

تم افتراض ثلاث حالات لتوزيع الخطأ العشوائي  $\varepsilon$  في تجربة المحاكاة وهي :

- 1- الخطأ العشوائي يتوزع توزيعاً طبيعياً  $\varepsilon \sim N(0,9)$  .
  - 2- الخطأ العشوائي يتوزع توزيع  $t$  بدرجة حرية تساوي 3 اي ان  $\varepsilon \sim t_{(3)}$  .
  - 3- الخطأ العشوائي يتوزع توزيع  $\chi^2$  بدرجة حرية تساوي 3 اي ان  $\varepsilon \sim \chi^2_{(3)}$  .
- والجدول رقم ( 1-3 ) يوضح ماترقتنا اليه سابقا مع استخدام كل محاكاة ومع كل حالة وكما يلي :

Method	Case	Error
BBqr	I	$N(0,9)$
BLBqr	I	$N(0,9)$
BrLBqr	I	$N(0,9)$
BBqr	II	$N(0,9)$
BLBqr	II	$N(0,9)$
BrLBqr	II	$N(0,9)$
BBqr	III	$N(0,9)$
BLBqr	III	$N(0,9)$
BrLBqr	III	$N(0,9)$
BBqr	I	$t(a)$
BLBqr	I	$t(a)$
BrLBqr	I	$t(a)$
BBqr	II	$t(a)$
BLBqr	II	$t(a)$
BrLBqr	II	$t(a)$
BBqr	III	$t(a)$
BLBqr	III	$t(a)$
BrLBqr	III	$t(a)$
BBqr	I	$\chi^2(a)$
BLBqr	I	$\chi^2(a)$
BrLBqr	I	$\chi^2(a)$
BBqr	II	$\chi^2(a)$
BLBqr	II	$\chi^2(a)$
BrLBqr	II	$\chi^2(a)$
BBqr	III	$\chi^2(a)$
BLBqr	III	$\chi^2(a)$
BrLBqr	III	$\chi^2(a)$



### 3-3 نتائج المحاكاة

ومن خلال تجارب المحاكاة ولكافة الطرائق مع الطريقة المقترحة وباستخدام التوزيعات المختلفة للخطأ العشوائي المذكورة أعلاه ، تم استخدام الخوارزمية Gibbs Sampler وبتكرار التجربة بـ 150 مرة . وعند كل تكرار في تجربة المحاكاة تم الافتراض بان مجموعة بيانات التدريب ( training set ) تساوي 20 مشاهدة وبيانات الاختبار ( testing set ) تساوي 200 مشاهدة من اجل الوصول الى افضل التقديرات لمتجه المعالم ، وتم ايضا افتراض التقسيمات عند (p=0.5) و (p=0.95) ، وتم تنفيذ الخوارزمية Gibbs Sampler ( 13000 iteration ) عند ( 3000 burn ) ، علما بان الخوارزمية تم تنفيذها ببرنامج لغة R . واعتمادا على المعايير اعلاه في بيان افضلية الطرائق فقد تبين تفوق الطريقة المقترحة (BrLBqr) على الطرق الاخرى والمقترحة من قبل ( Alhamzawi and Ali , 2020 ) وهي (BBqr) و (BLBqr) .

والجداول في ادناه توضح نتائج تجربة المحاكاة عند (p=0.5) و (p=0.95) :

الجدول رقم (2-3) قيم المعايير (MMAD) و (Sd) عندما (p=0.5)

Method	Case	Error	MMAD	Sd(MAD)
BBqr	I	$N(0,9)$	2.1156	1.6175
BLBqr	I	$N(0,9)$	2.4232	1.8804
BrlBqr	I	$N(0,9)$	2.095	1.6051
BBqr	II	$N(0,9)$	2.6265	2.0657
BLBqr	II	$N(0,9)$	2.8592	2.1776
BrlBqr	II	$N(0,9)$	2.6047	2.0408
BBqr	III	$N(0,9)$	2.4599	1.9456
BLBqr	III	$N(0,9)$	2.7014	2.0963
BrlBqr	III	$N(0,9)$	2.4681	1.9362
BBqr	I	$t_{(3)}$	2.3735	1.7678
BLBqr	I	$t_{(3)}$	2.6148	1.996
BrlBqr	I	$t_{(3)}$	2.3585	1.7558
BBqr	II	$t_{(3)}$	2.4599	1.9456
BLBqr	II	$t_{(3)}$	2.7014	2.0963
BrlBqr	II	$t_{(3)}$	2.4681	1.9362
BBqr	III	$t_{(3)}$	2.4161	1.8676
BLBqr	III	$t_{(3)}$	2.6773	2.0495
BrlBqr	III	$t_{(3)}$	2.3886	1.843
BBqr	I	$\chi^2_{(3)}$	2.3571	1.7613
BLBqr	I	$\chi^2_{(3)}$	2.6025	1.9983
BrlBqr	I	$\chi^2_{(3)}$	2.3209	1.7461
BBqr	II	$\chi^2_{(3)}$	2.4187	1.8676
BLBqr	II	$\chi^2_{(3)}$	2.6773	2.0523
BrlBqr	II	$\chi^2_{(3)}$	2.3958	1.8456
BBqr	III	$\chi^2_{(3)}$	2.4194	1.8413
BLBqr	III	$\chi^2_{(3)}$	2.6902	2.0424
BrlBqr	III	$\chi^2_{(3)}$	2.4029	1.8407

اظهرت نتائج الجدول رقم (2-3) تفوق في اداء الطريقة المقترحة وتحت كل نوع من انواع الخطأ وذلك من خلال قيم المعايير (MMAD) و (Sd) . وهذا يعكس قدرة الطريقة المقترحة في تنفيذ اجراء اختيار المتغيرات ( Variable Selection ) مما يعني ان طريقتنا المقترحة تفوقت على الطرق الاخرى من خلال جودت قيم تقديرات المعالم وتطابقها مع متجه المعالم الحقيقية  $\beta = (2,1,0,0,3,0,0,0,0,0)$  عند خط الانحدار . (p=0.5)

الجدول رقم (3-3) قيم المعايير (MMAD) و (Sd) عندما (p=0.5)

Method	Case	Error	MMAD	Sd(MAD)
BBqr	I	$N(0,9)$	2.3911	1.7723
BLBqr	I	$N(0,9)$	2.6395	1.9975
BrLBqr	I	$N(0,9)$	2.3668	1.7545
BBqr	II	$N(0,9)$	2.3315	1.7334
BLBqr	II	$N(0,9)$	2.5959	1.9578
BrLBqr	II	$N(0,9)$	2.294	1.7227
BBqr	III	$N(0,9)$	2.2814	1.703
BLBqr	III	$N(0,9)$	2.5747	1.9277
BrLBqr	III	$N(0,9)$	2.2604	1.6784
BBqr	I	$t_{(3)}$	2.2175	1.6725
BLBqr	I	$t_{(3)}$	2.5212	1.899
BrLBqr	I	$t_{(3)}$	2.1881	1.6451
BBqr	II	$t_{(3)}$	2.1644	1.6171
BLBqr	II	$t_{(3)}$	2.4854	1.8716
BrLBqr	II	$t_{(3)}$	2.1353	1.5989
BBqr	III	$t_{(3)}$	2.1131	1.5895
BLBqr	III	$t_{(3)}$	2.4232	1.8307
BrLBqr	III	$t_{(3)}$	2.0888	1.5683
BBqr	I	$x^2_{(3)}$	2.0739	1.5605
BLBqr	I	$x^2_{(3)}$	2.3336	1.7757
BrLBqr	I	$x^2_{(3)}$	2.0553	1.5337
BBqr	II	$x^2_{(3)}$	2.0064	1.5088
BLBqr	II	$x^2_{(3)}$	2.293	1.7279
BrLBqr	II	$x^2_{(3)}$	1.9689	1.4925
BBqr	III	$x^2_{(3)}$	1.9781	1.4838
BLBqr	III	$x^2_{(3)}$	2.248	1.6875
BrLBqr	III	$x^2_{(3)}$	1.9482	1.4712

نلاحظ من خلال نتائج الجدول رقم (3-3) أعلاه تفوق في أداء الطريقة المقترحة وتحت كل نوع من أنواع الخطأ وذلك من خلال قيم المعايير (MMAD) و (Sd) . وهذا دلالة على قدرة الطريقة المقترحة في تنفيذ اجراء اختيار المتغيرات ( Variable Selection ) وهذا يعني ان طريقتنا المقترحة تفوقت على الطرائق الاخرى من خلال جودت قيم تقديرات المعالم وتطابقها مع متجه المعالم الحقيقية  $\beta = (3,0,0,0,0,0,0,0,0)$  عند خط الانحدار (p=0.5)

الجدول رقم (3-4) قيم المعايير (MMAD) و (Sd) عندما (p=0.5)

Method	Case	Error	MMAD	Sd(MAD)
BBqr	I	$N(0,9)$	1.9532	1.4731
BLBqr	I	$N(0,9)$	2.1968	1.68
BrLBqr	I	$N(0,9)$	1.9266	1.4573
BBqr	II	$N(0,9)$	1.9885	1.4893
BLBqr	II	$N(0,9)$	2.2713	1.7129
BrLBqr	II	$N(0,9)$	1.9635	1.4813
BBqr	III	$N(0,9)$	2.0571	1.5508
BLBqr	III	$N(0,9)$	2.3034	1.7432
BrLBqr	III	$N(0,9)$	2.0375	1.5216
BBqr	I	$t_{(3)}$	1.9885	1.4893
BLBqr	I	$t_{(3)}$	2.2713	1.7129
BrLBqr	I	$t_{(3)}$	1.9635	1.4813
BBqr	II	$t_{(3)}$	2.0571	1.5508
BLBqr	II	$t_{(3)}$	2.3034	1.7432
BrLBqr	II	$t_{(3)}$	2.0375	1.5216
BBqr	III	$t_{(3)}$	2.0974	1.5744
BLBqr	III	$t_{(3)}$	2.3396	1.7801
BrLBqr	III	$t_{(3)}$	2.0675	1.5532
BBqr	I	$x^2_{(3)}$	2.0595	1.5533
BLBqr	I	$x^2_{(3)}$	2.3034	1.7442
BrLBqr	I	$x^2_{(3)}$	2.0393	1.5216
BBqr	II	$x^2_{(3)}$	2.0974	1.5802
BLBqr	II	$x^2_{(3)}$	2.3396	1.7801
BrLBqr	II	$x^2_{(3)}$	2.0675	1.5572
BBqr	III	$x^2_{(3)}$	2.161	1.6022
BLBqr	III	$x^2_{(3)}$	2.4232	1.8307
BrLBqr	III	$x^2_{(3)}$	2.1208	1.5946

افرزت نتائج الجدول رقم (3-4) أعلاه تفوق ملحوظ في اداء طريقتنا المقترحة و لكل نوع من انواع الخطأ وذلك من خلال قيم المعايير (MMAD) و (Sd) . وهذا يعني قدرة الطريقة المقترحة في تنفيذ اجراء اختيار المتغيرات ( Variable Selection ) مما يعني ان الطريقة المقترحة تفوقت في اداءها على الطرائق الاخرى من خلال جودت قيم تقديرات المعالم وتطابقها مع متجه المعالم الحقيقية  $\beta = (1,1,1,1,1,1,1,1,1)$  عند خط الانحدار (p=0.5)

الجدول رقم (3-5) قيم المعايير (MMAD) و (Sd) عندما (p=0.95)

Method	Case	Error	MMAD	Sd(MAD)
BBqr	I	$N(0,9)$	2.0111	1.5162
BLBqr	I	$N(0,9)$	2.0974	1.6186
BrLBqr	I	$N(0,9)$	1.9817	1.4892
BBqr	II	$N(0,9)$	2.2326	1.7639
BLBqr	II	$N(0,9)$	2.415	1.7654
BrLBqr	II	$N(0,9)$	2.2105	1.7574
BBqr	III	$N(0,9)$	2.2326	1.7392
BLBqr	III	$N(0,9)$	2.4366	1.7654
BrLBqr	III	$N(0,9)$	2.2029	1.6997
BBqr	I	$t_{(3)}$	2.0913	1.5662
BLBqr	I	$t_{(3)}$	2.2358	1.6701
BrLBqr	I	$t_{(3)}$	1.9697	1.5221
BBqr	II	$t_{(3)}$	2.1465	1.6257
BLBqr	II	$t_{(3)}$	2.3286	1.7335
BrLBqr	II	$t_{(3)}$	2.055	1.5611
BBqr	III	$t_{(3)}$	2.0913	1.5504
BLBqr	III	$t_{(3)}$	2.1976	1.6404
BrLBqr	III	$t_{(3)}$	1.9697	1.4856
BBqr	I	$x^2_{(3)}$	2.0515	1.5093
BLBqr	I	$x^2_{(3)}$	2.1417	1.6227
BrLBqr	I	$x^2_{(3)}$	1.9466	1.4384
BBqr	II	$x^2_{(3)}$	2.0913	1.5735
BLBqr	II	$x^2_{(3)}$	2.1976	1.6443
BrLBqr	II	$x^2_{(3)}$	1.9679	1.4993
BBqr	III	$x^2_{(3)}$	2.0913	1.5662
BLBqr	III	$x^2_{(3)}$	2.1469	1.6338
BrLBqr	III	$x^2_{(3)}$	1.9679	1.4976

من خلال ملاحظة نتائج الجدول رقم (3-5) نلاحظ تفوق واضح في اداء طريقتنا المقترحة ولكل نوع من انواع الخطأ وذلك من خلال قيم المعايير (MMAD) و (Sd) . وهذا يبين قدرة الطريقة المقترحة في تنفيذ اجراء اختيار المتغيرات ( Variable Selection ) مما يعني ان طريقتنا المقترحة تفوقت في اداءها على الطرق الاخرى من خلال جودت قيم تقديرات المعالم وتطابقها مع متجه المعالم الحقيقية  $\beta = (2,1,0,0,3,0,0,0,0,0)$  عند خط الانحدار (p=0.95)

الجدول رقم (6-3) قيم المعايير (MMAD) و (Sd) عندما (p=0.95)

Method	Case	Error	MMAD	Sd(MAD)
BBqr	I	$N(0,9)$	2.0322	1.5214
BLBqr	I	$N(0,9)$	2.1173	1.6172
BrLBqr	I	$N(0,9)$	1.9505	1.479
BBqr	II	$N(0,9)$	2.0045	1.5016
BLBqr	II	$N(0,9)$	2.0699	1.5517
BrLBqr	II	$N(0,9)$	1.9329	1.4599
BBqr	III	$N(0,9)$	1.9829	1.4715
BLBqr	III	$N(0,9)$	2.0351	1.5074
BrLBqr	III	$N(0,9)$	1.9091	1.4384
BBqr	I	$t_{(3)}$	1.9344	1.4465
BLBqr	I	$t_{(3)}$	1.9905	1.4894
BrLBqr	I	$t_{(3)}$	1.866	1.4119
BBqr	II	$t_{(3)}$	1.8857	1.4204
BLBqr	II	$t_{(3)}$	1.9291	1.4696
BrLBqr	II	$t_{(3)}$	1.8512	1.3988
BBqr	III	$t_{(3)}$	1.8391	1.4008
BLBqr	III	$t_{(3)}$	1.8687	1.4147
BrLBqr	III	$t_{(3)}$	1.8379	1.3691
BBqr	I	$\chi^2_{(3)}$	1.8333	1.4008
BLBqr	I	$\chi^2_{(3)}$	1.844	1.3932
BrLBqr	I	$\chi^2_{(3)}$	1.8379	1.3691
BBqr	II	$\chi^2_{(3)}$	1.8262	1.3923
BLBqr	II	$\chi^2_{(3)}$	1.8311	1.3872
BrLBqr	II	$\chi^2_{(3)}$	1.8316	1.3669
BBqr	III	$\chi^2_{(3)}$	1.7831	1.3422
BLBqr	III	$\chi^2_{(3)}$	1.7986	1.3759
BrLBqr	III	$\chi^2_{(3)}$	1.8216	1.3605

افرزت نتائج الجدول رقم (6-3) تفوقاً في اداء طريقتنا المقترحة وتحت كل نوع من انواع الخطأ وذلك من خلال قيم المعايير (MMAD) و (Sd) . وهذا دلالة على قدرة طريقتنا المقترحة في تنفيذ اجراء اختيار المتغيرات ( Variable Selection ) وهذا يعني ان طريقتنا المقترحة تفوقت في اداءها على الطرائق الاخرى من خلال جودت قيم تقديرات المعالم وتطابقها مع متجه المعالم الحقيقية  $\beta = (3,0,0,0,0,0,0,0,0)$  عند خط الانحدار (p=0.95)

الجدول رقم (7-3) قيم المعايير (MMAD) و (Sd) عندما (p=0.95)

Method	Case	Error	MMAD	Sd(MAD)
BBqr	I	$N(0,9)$	1.7788	1.3395
BLBqr	I	$N(0,9)$	1.8118	1.3813
BrLBqr	I	$N(0,9)$	1.8162	1.3605
BBqr	II	$N(0,9)$	1.8333	1.4008
BLBqr	II	$N(0,9)$	1.8426	1.3936
BrLBqr	II	$N(0,9)$	1.8418	1.3744
BBqr	III	$N(0,9)$	1.8743	1.4179
BLBqr	III	$N(0,9)$	1.8687	1.4163
BrLBqr	III	$N(0,9)$	1.8572	1.4119
BBqr	I	$t_{(3)}$	1.8333	1.4008
BLBqr	I	$t_{(3)}$	1.844	1.3941
BrLBqr	I	$t_{(3)}$	1.8418	1.3744
BBqr	II	$t_{(3)}$	1.8743	1.4179
BLBqr	II	$t_{(3)}$	1.8807	1.4163
BrLBqr	II	$t_{(3)}$	1.8572	1.4119
BBqr	III	$t_{(3)}$	1.9137	1.4351
BLBqr	III	$t_{(3)}$	1.9291	1.4438
BrLBqr	III	$t_{(3)}$	1.8931	1.4384
BBqr	I	$x^2_{(3)}$	1.8822	1.4253
BLBqr	I	$x^2_{(3)}$	1.8951	1.4217
BrLBqr	I	$x^2_{(3)}$	1.8734	1.4229
BBqr	II	$x^2_{(3)}$	1.9237	1.4465
BLBqr	II	$x^2_{(3)}$	1.9379	1.455
BrLBqr	II	$x^2_{(3)}$	1.9425	1.4592
BBqr	III	$x^2_{(3)}$	1.9611	1.473
BLBqr	III	$x^2_{(3)}$	1.9773	1.4821
BrLBqr	III	$x^2_{(3)}$	1.9535	1.479

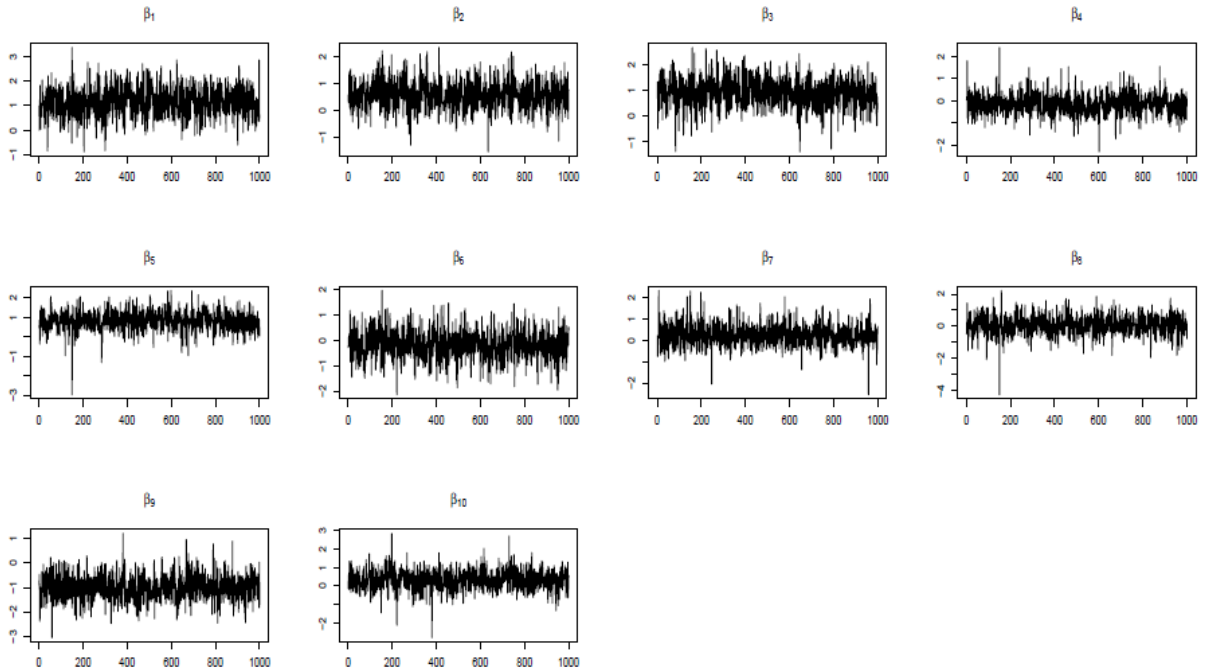
اظهرت نتائج الجدول رقم (7-3) تفوقاً ملحوظاً في الاداء للطريقة المقترحة وتحت كل نوع من انواع الخطأ وذلك من خلال قيم المعايير (MMAD) و (Sd) . وهذا يعني ان الطريقة المقترحة لها القدرة في تنفيذ اجراء اختيار المتغيرات ( Variable Selection ) وهذا يدل على ان الطريقة المقترحة تفوقت على الطرائق الاخرى من خلال جودت قيم تقديرات المعالم وتطابقها مع متجه المعالم الحقيقية  $\beta = (1,1,1,1,1,1,1,1,1)$  عند خط الانحدار (p=0.95)

### 4-3 الرسومات التوضيحية :

الاشكال التوضيحية في ادناه تبين رسم (trace plot) حيث يستخدم هذا الرسم لمعرفة كفاءة عمل خوارزمية لكل توزيع من التوزيعات اللاحقة (posterior distributions) ، كذلك من خلال هذا الرسم يمكن تتبع التقارب في الخوارزمية لتوليدها لقيم المعالم .

ان جودت أداء الخوارزمية يمكن تتبعه في حالة عدم وجود فراغات في الرسم البياني ، وهذا يعني سرعة تنفيذ الخوارزمية وعدم تباطؤها في توليد قيم المعالم ، كذلك يتبين من رسم (trace plot) عدم وجود قيم بعيدة عن خط الوسط . ورسم ( Histogram ) المدرج التكراري الذي عادةً ما يستخدم لاختبار توزيع البيانات .

شكل رقم ( 1-3 ) رسم مخطط (trace plot) للمعالم المقدره  
(  $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$  )

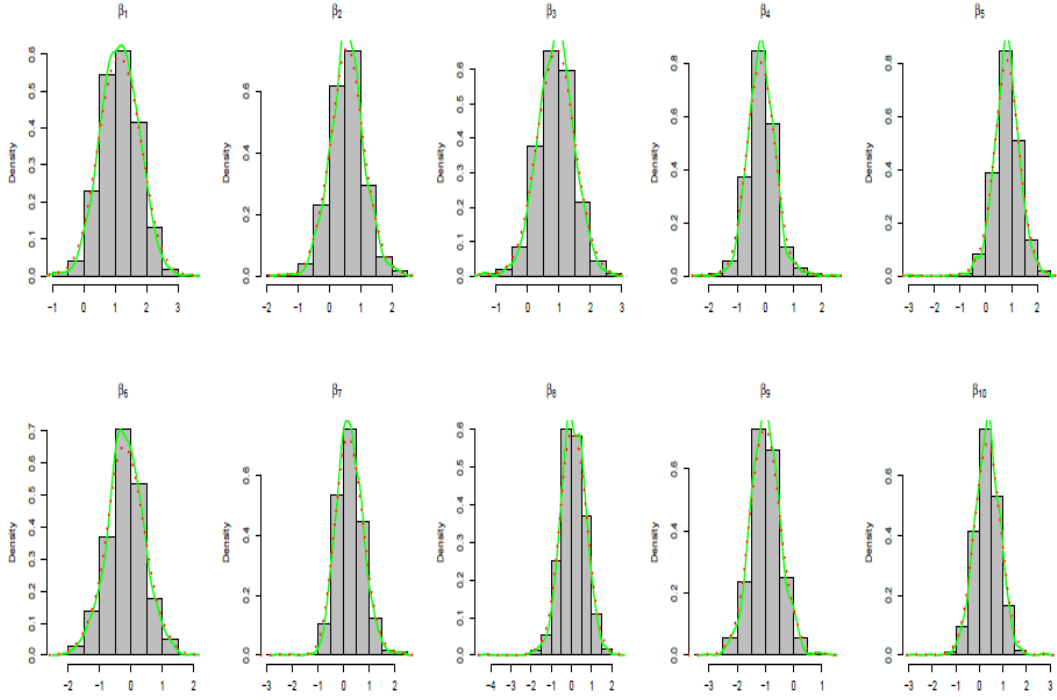


يظهر من الشكل رقم ( 1-3 ) الذي يوضح رسم (trace plot) كفاءة تقارب خوارزمية (Gibbs Sampler) للتوزيع اللاحق (posterior distribution) لكل معلمة من المعالم المقدره (  $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$  ) .



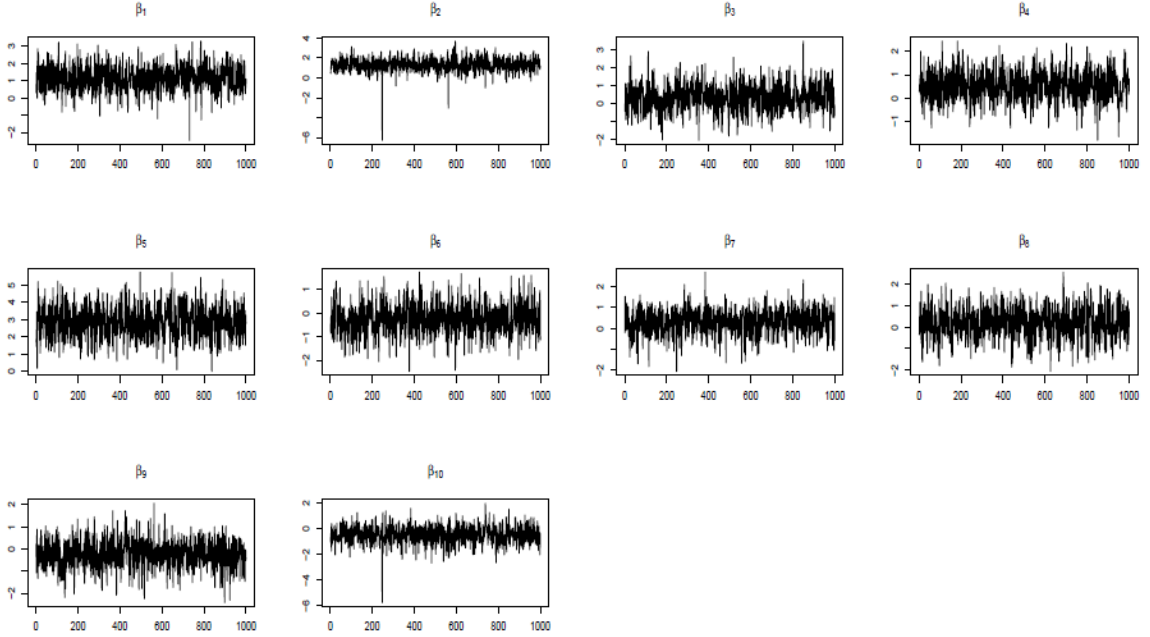
شكل رقم ( 2-3 ) المدرج التكراري ( Histogram ) لكل معلمة من المعالم المقدرة

$$(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10})$$



نلاحظ من اشكال المدرج التكراري ولكل معلمة من المعالم المقدرة  $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10})$  انها تتبع التوزيع الطبيعي وهذا يتطابق مع الجانب النظري حيث تم افتراض ان المعلمة  $\beta$  تتبع التوزيع الطبيعي كتوزيع لاحق لها .

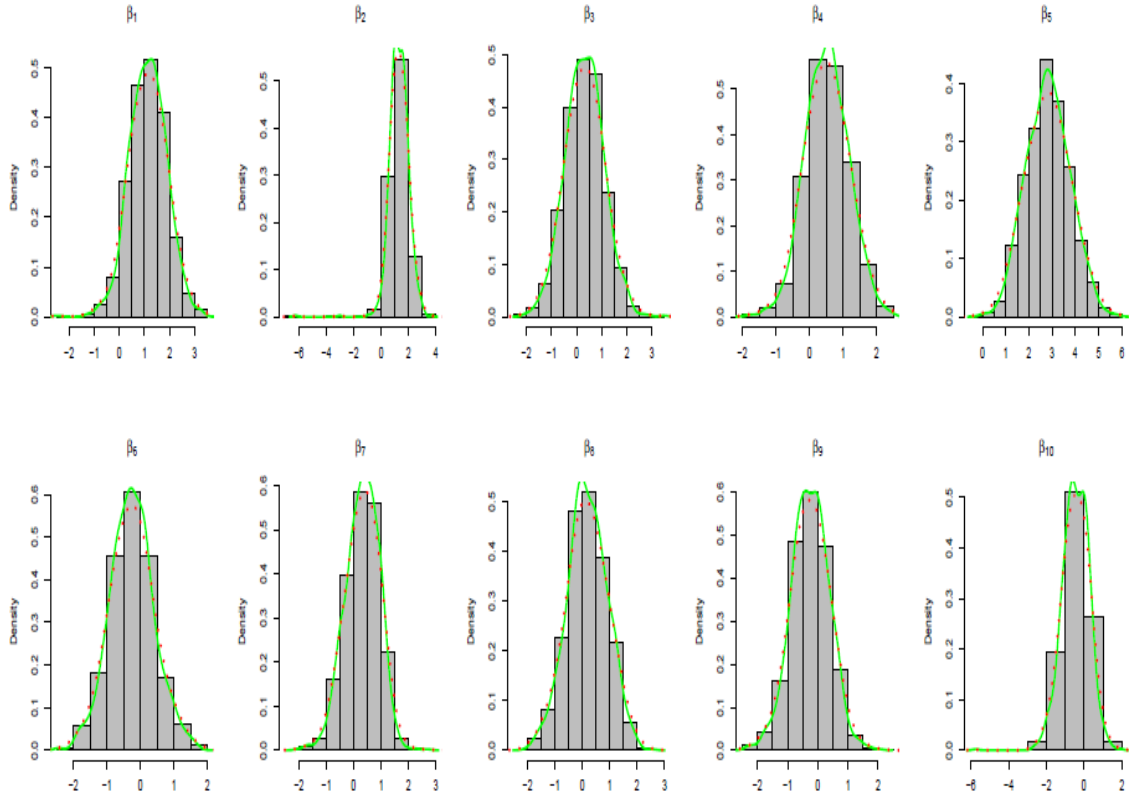
شكل رقم ( 3-3 ) رسم مخطط ( trace plot ) للمعالم المقدرة  
(  $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$  )



نلاحظ من الشكل رقم ( 3-3 ) الذي يوضح رسم ( trace plot ) الاداء الجيد في تقارب خوارزمية (Gibbs Sampler) للتوزيع اللاحق (posterior distribution) لكل معلمة من المعالم المقدرة (  $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$  ).

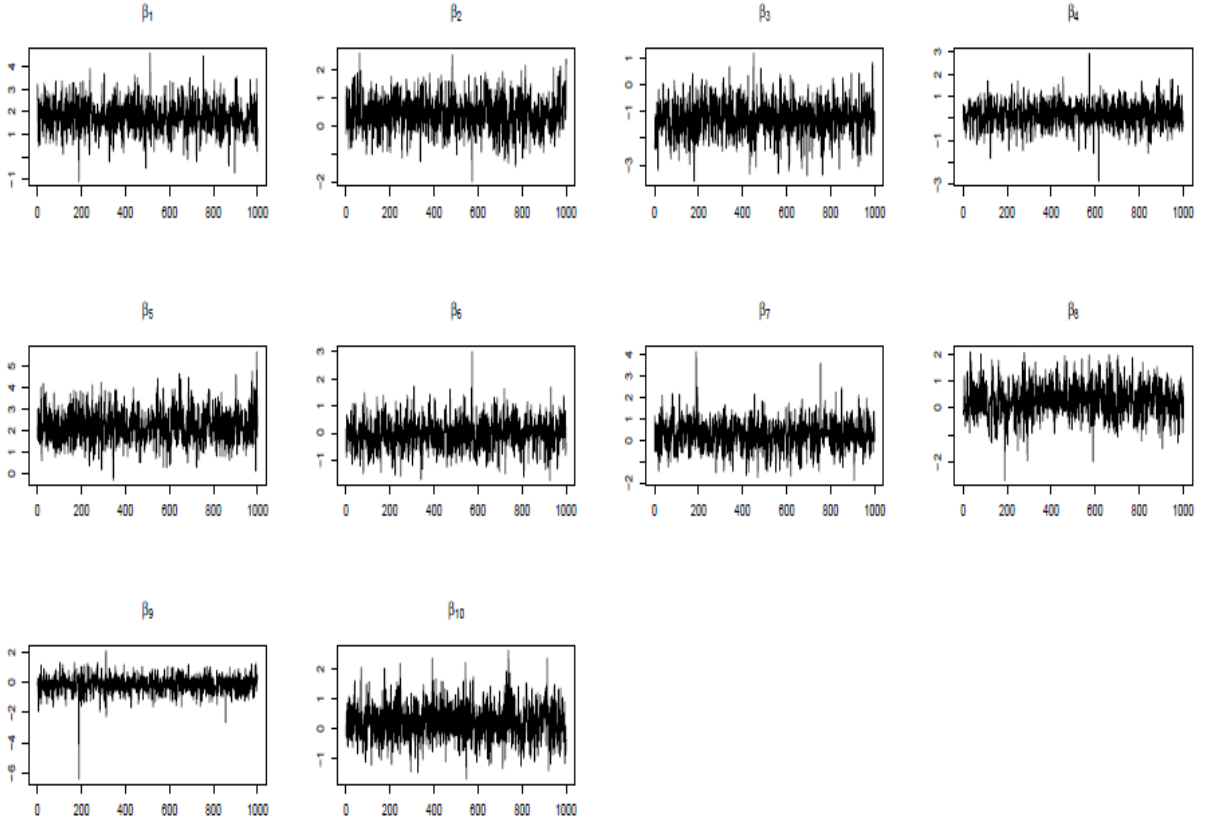
شكل رقم ( 3-4 ) المدرج التكراري ( Histogram ) لكل معلمة من المعالم المقدره

$$(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10})$$



ومن ملاحظة اشكال المدرج التكراري اعلاه ولكل معلمة من المعالم  $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10})$  نجد انها تتبع التوزيع الطبيعي وهذا دلالة على التطابق مع الجانب النظري حيث تم افتراض ان المعلمة  $\beta$  تتبع التوزيع الطبيعي كتوزيع لاحق لها .

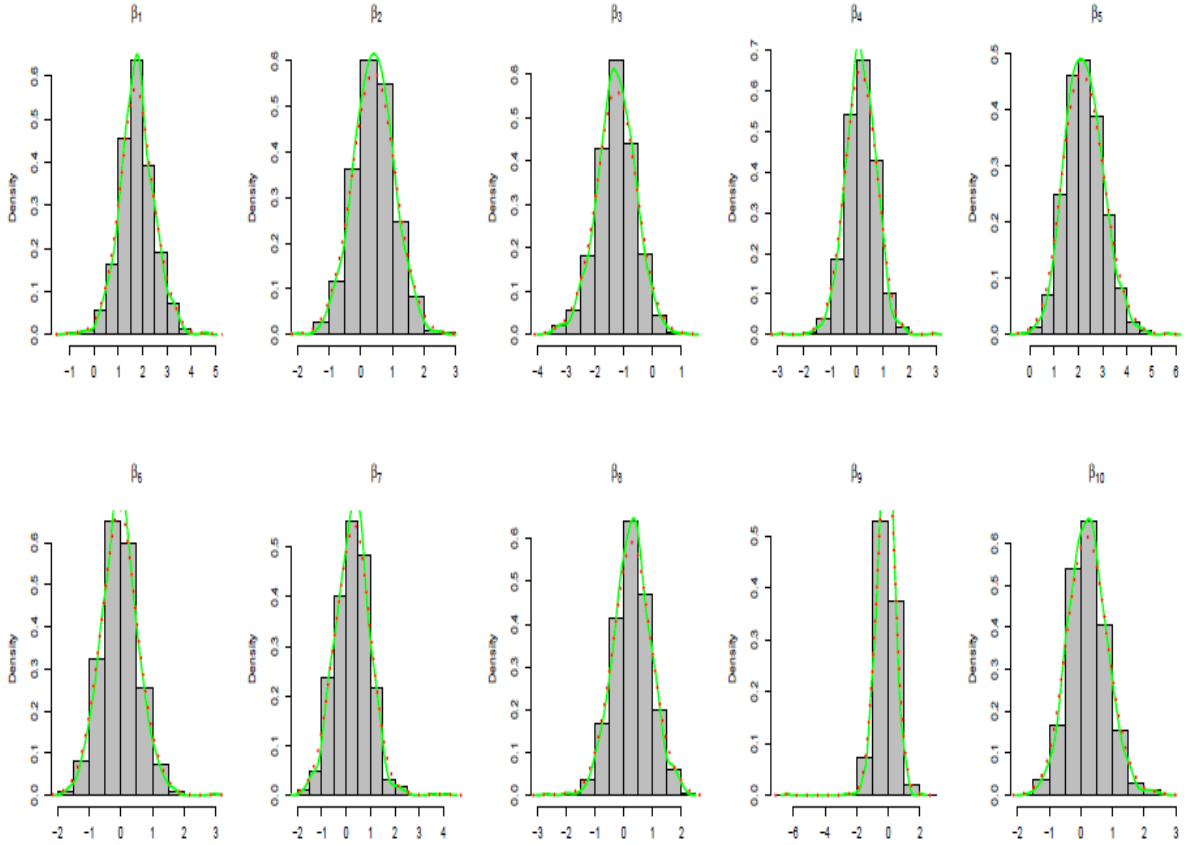
شكل رقم ( 5-3 ) رسم مخطط ( trace plot ) للمعالم المقدرة  
(  $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$  )



يلاحظ من الشكل رقم ( 5-3 ) الذي يوضح رسم ( trace plot ) كفاءة تقارب خوارزمية (Gibbs Sampler) للتوزيع اللاحق ( posterior distribution ) لكل معلمة من المعالم المقدرة (  $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$  ).

شكل رقم ( 3-6 ) المدرج التكراري ( Histogram ) لكل معلمة من المعالم المقدره

$$(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10})$$



نلاحظ من اشكال المدرج التكراري اعلاه ولكل معلمة من المعالم المقدره  $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10})$  انها تتبع التوزيع الطبيعي وهذا يدل على التطابق مع الجانب النظري حيث تم افتراض ان المعلمة  $\beta$  تتبع التوزيع الطبيعي كتوزيع لاحق لها .

# الفصل الرابع

الجانب التطبيقي

## 4- الجانب التطبيقي

### 1-4 المقدمة

تعد الفحوصات الدورية للمريض من الاجراءات المهمة والضرورية اللازمة من اجل ديمومة الصحة والسلامة للانسان ، ولقد تم تطبيق الطريقة المقترحة وتوظيفها في هذا الفصل 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 حتى يتمكن الاطباء من الوقوف عند العلاج اللازم والمفيد من اجل تحسين حالة المريض .

وفي هذا البحث تناولنا مجموعة من المتغيرات التوضيحية ( تشخيصات اطباء الاختصاص ) والتي تم الحصول عليها من مستشفى الحكيم في النجف الاشرف والتي تختص بـ بمرض كورونا ( Covid - 19 ) ومن هذه التشخيصات نسبة اليوريا ونسبة الكرياتينين وبعض الانزيمات التي تحدد بدورها نسبة الالتهاب في الجسم والبعض الاخر من التحليلات التي تقيس نسبة الدم في الجسم ونسبة كريات الدم البيضاء وقياس المناعة واخرى تحدد صفيحات الدم التي تفيد التخثر وهناك ايضا متغيرات توضيحية اخرى تم تضمينها في هذا البحث والتي سيتم التطرق اليها لاحقاً .

في هذا الجانب تم الاعتماد على بيانات حقيقة مأخوذة من مستشفى الحكيم في النجف الاشرف 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 البيانات 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 الـ (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 للـ (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 . بمعنى آخر ان خط الانحدار عند اي (quantile) اخر سواء كان عند (  $p=0.25$  ) او (  $p=0.75$  ) فهو يكون خط انحدار جيد للتنبؤ والتقدير

## 2-4 توصيف البيانات

تم الحصول على هذه البيانات من احد مستشفيات النجف الاشرف قسم تشخيص مرض كورونا ( Covid - 19 ) ، وقد تمت الاستعانة بمجموعة من اطباء الاختصاص لتصنيف اهم العوامل التي تكون مسببة لهذا المرض .

حيث صنفنا المتغيرات وكما يلي :-

Y : حالة خروج المريض والتي تمثل المتغير المعتمد

X<sub>1</sub> : نوع الجنس

X<sub>2</sub> : التحصيل الدراسي

X<sub>3</sub> : الوظيفة

X<sub>4</sub> : العمر بالسنوات

X<sub>5</sub> : الحالة الزوجية

X<sub>6</sub> : محل الإقامة

X<sub>7</sub> : سكن المريض

X<sub>8</sub> : مدة رقود المريض بالايام

X<sub>9</sub> : حالة التدخين

X<sub>10</sub> : نسبة السكر في الدم

X<sub>11</sub> : ضغط الدم

X<sub>12</sub> : وزن الشخص

X<sub>13</sub> : نسبة اليوريا في الدم

X<sub>14</sub> : نسبة الكرياتين في الدم

X<sub>15</sub> : ( LDH ) انزيم مرتبط بالرئة

X<sub>16</sub> : ( CRP ) تحليل يختص بوجود الالتهابات

X<sub>17</sub> : تحليل يمثل مخزون الحديد في الدم

X<sub>18</sub> : ( ESR ) تحليل يختص بوجود الالتهابات

X<sub>19</sub> : ( HGB ) يمثل نسبة الدم



( WBC ) : X<sub>20</sub> نسبة كريات الدم البيضاء

( NEU ) : X<sub>21</sub> يمثل تحليل يكتشف المناعة عند الانسان

( LYM ) : X<sub>22</sub> تحليل يمثل وجود فايروس في الجسم

( PLT ) : X<sub>23</sub> تحليل يمثل عدد صفيحات الدم التي تفيد التخثر

( D-DIMER ) : X<sub>24</sub> تحليل يكتشف التخثر في الدم

( SPO2 ) : X<sub>25</sub> تحليل يشير الى نسبة تركيز الاكسجين في الدم

### 3-4 نتائج تنفيذ الخوارزمية

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

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

20 training data 230 testing data tau=0.5			
	BBqr	BLBqr	BrLBqr
Intercept	0.009	0	-1.828
x1	0.232	0.013	-6.864
x2	0.237	0.036	-0.144
x3	-0.006	-0.004	2.673
x4	0.018	-0.002	-0.719
x5	0.302	0.013	2.911
x6	0.022	0.005	-1.981
x7	0.093	0.036	-4.649
x8	-0.015	-0.02	-0.244
x9	0.211	0.009	-3.516
x10	-0.107	-0.005	-5.468
x11	-0.008	0.005	-0.606
x12	-0.007	-0.018	-1.388
x13	0.005	-0.003	0.107
x14	-0.031	-0.005	-1.829
x15	-0.002	-0.003	-0.05
x16	-0.122	-0.085	-1.154
x17	0	0	0.017
x18	-0.004	0.005	0.268
x19	0.056	0.01	1.354
x20	-0.009	0.002	0.492
x21	-0.049	0.001	1.967
x22	-0.048	0.001	2.348
x23	0	0	0.046
x24	-0.054	-0.017	3.482
x25	0.039	0.026	-0.674

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

20 training data 230 testing data			
tau=0.95			
	BBqr	BLBqr	BrLBqr
Intercept	0.457	-0.002	15.317
x1	-0.031	-0.014	-2.536
x2	0.072	0.006	0.671
x3	-0.064	0.015	3.149
x4	0.01	0.008	0.195
x5	0.237	0.03	7.044
x6	-0.066	-0.014	3.347
x7	0.063	0.036	1.398
x8	0.005	-0.009	0.239
x9	0.039	-0.009	4.053
x10	-0.367	-0.025	0.829
x11	-0.102	-0.02	0.611
x12	-0.001	-0.001	-0.53
x13	-0.006	-0.004	-0.015
x14	0.353	0.019	-1.299
x15	-0.001	-0.001	-0.024
x16	-0.284	-0.215	0.196
x17	0	0	0
x18	-0.003	-0.005	-0.085
x19	-0.027	-0.01	-2.339
x20	-0.007	-0.009	-0.144
x21	0.03	0.011	0.058
x22	0.028	0.014	-0.164
x23	0.002	0.002	0.019
x24	-1.077	-0.128	0.498
x25	0.05	0.062	0.392

#### 4-4 تفسير النتائج

1- اظهرت نتائج الجدول رقم (4-1) تفوق في اداء الطريقة المقترحة BrLBq عند المستوى (  $p=0.5$  ) حيث حصلت الطريقة المقترحة BrLBq على 217 نتيجة من التصنيف الصحيح من مجموع المشاهدات 230 ، بينما الطرق الاخرى BLBq و BBq حصلت على 199 و 202 من التصنيف الصحيح من مجموع المشاهدات الكلي 230 ، على التوالي .

2- افرزت نتائج الجدول رقم (4-2) ايضا تفوق في اداء الطريقة المقترحة BrLBq عند المستوى (  $p=0.95$  ) حيث حصلت على 205 من التصنيف الصحيح من المجموع الكلي للمشاهدات 230 ، بينما الطرق الاخرى BLBq و BBq حصلت على 181 ، 194 من التصنيف الصحيح من مجموع المشاهدات 230 ، على التوالي .

# الفصل الخامس

الاستنتاجات والتوصيات

## الاستنتاجات والتوصيات

### 1-5 الاستنتاجات

الاهتمام الرئيسي لهذه الاطروحة هو ادخال طريقة التنظيم البيزي في نموذج الانحدار الكمي الثنائي من اجل الحصول على نموذج تقديري ذات قوة تفسيرية عالية وله القابلية على التنافس مع النماذج الاخرى . حيث تم اقتراح نموذج هرمي بيبي جديد للانحدار الكمي الثنائي من خلال اقتراح تمثيل مختلط للتوزيع السابق لمعالم النموذج .

في هذه البحث قدمنا نهجاً بايزياً للانحدار الكمي الثنائي مقترناً بتقنية اختيار المتغيرات (Selection variables) ، اي الانحدار الكمي الثنائي البيبي مع دالة جزاء معكوس لاسو ( Binary reciprocal lasso quantile regression ). وان المزايا الرئيسية لهذا النهج تمثلت باجراء تقدير معالم النموذج واختيار المتغيرات (Selection variables) التنبؤية 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) .

ايضاً قدمنا في هذا البحث طريقة تنظيم (Regularization) معكوس لاسو ( Reciprocal Lasso ) للانحدار الكمي الثنائي البيبي. حيث كان الاسلوب البيبي لهذه المشكلة هو وضع توزيع معكوس لابلاس ((Inverse Laplace distribution)) للتوزيع السابق (Prior) لمعالم الانحدار .

تم تطبيق الطريقة المقترحة (BrLBqr) في كل من بيانات المحاكاة والبيانات الحقيقية . وباستخدام خوارزمية (Gibbs sampler) اظهرت النتائج بتفوق اداء الطريقة المقترحة (BrLBqr) وفي كلا الجانبين (جانب المحاكاة والجانب التطبيقي ) مقارنةً بالطرائق الاخرى (BLBqr) و (BBqr) وبالاعتماد على معايير المقارنة وهما متوسط الانحراف المطلق ( MMAD) و الانحراف المعياري ( Sd) لقياس جودة الملائمة في نماذج الانحدار المختلفة، حيث كانت الطريقة المقترحة تمتلك اقل القيم ولكلا المعيارين، اي ان جزاء معكوس لاسو اظهر الاداء الافضل للطريقة المقترحة .

وأخيراً ، غالباً ما يعتمد الباحثون على نماذج لوغارتمية او احتمالية معروفة بانها متحيزة عند وجود بيانات متطرفة وفي حالة عدم تجانس التباين وما الى ذلك . في حين ان الطريقة المقترحة لاتحتوي على هذا القصور ، وبالتالي يمكن ان يكون منهج قيم في هذا الاعداد البحثي . ومع ذلك ، نحن مقتنعون انه في العديد من المجالات الاخرى ايضاً يمكن للباحثين الاستفادة من الخصائص الجذابة لمعكوس لاسو البيزي جنباً الى جنب مع الانحدار الكمي الثنائي .

## **2-5 التوصيات**

يوصي الباحث باستخدام النموذج الهرمي التنظيمي (Regularization) المقترح في ظل التمثيل المختلط الجديد لتوزيع معكوس لابلاس (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 والبحاث الرياضية . لذلك نوصي باستخدام النموذج المقترح لانه يوفر نماذج (Sparse) يمكن ان تكون مفيدة في حقول البيانات المختلفة .

# المصادر



## **References**

- [1] Arthur E. Hoerl and Robert W. Kennard (1970) ." Ridge regression: Biased estimation for nonorthogonal problems". *Technometrics*, 12, 69-82.
- [2] Andrews, D. F., & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(1), 99-102 .
- [3] Alhamzawi, R., & Yu, K. (2012). Variable selection in quantile regression via Gibbs sampling. *Journal of Applied Statistics*, 39(4), 799-813.
- [4] Alhamzawi, R., Ali, H.T.M. Brq: an R package for Bayesian quantile regression. *METRON* 78, 313{328 (2020).<https://doi.org/10.1007/s40300-020-00190-6> .
- [5] Alhamzawi, R., & Mallick, H. (2020). Bayesian reciprocal LASSO quantile regression. *Communications in Statistics-Simulation and Computation*, 1-16.
- [6] Alhusseini, F. (2017) " New Bayesian Lasso in Tobit Quantile Regression" *Revista Română de Statistică - Supplement nr. 6 / 2017* .pp.213-229.
- [7] Alhusseini, F. H. H. (2017). Bayesian Quantile Regression with Scale Mixture of uniform prior Distribution, University of Craiova, Romania.
- [8] Alhusseini, F. & Vasile Georgescu (2018)" Bayesian Composite Tobit quantile Regression" *Journal of Applied Statistics*, Pages. Volume 45, Issue 4, pp. - 238 222 .
- [9] Alin, A. (2010). Multicollinearity. *Wiley interdisciplinary reviews: computational statistics*, 2(3), 370-374.

- [10] Benoit, D. F., Alhamzawi, R., & Yu, K. (2013). Bayesian lasso binary quantile regression. *Computational Statistics*, 28(6), 2861-2873.
- [11] 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.
- [12] Efron, B., T. Hastie, I. Johnstone, R. Tibshirani, et al. (2004). Least angle regression. *The Annals of statistics* 32 (2), 407–499.
- [13] 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.
- [14] Frank I, Friedman JH. A statistical view of some chemometrics regression tools (with discussion). *Technometrics*. 1993; 35:109–135.
- [15] Flaih, A.N., Alshaybawee, T., Alhusseini, F. H. H. (2020). Sparsity via new Bayesian Lasso. *Periodicals of engineering and Natural Sciences* 8(1).
- [16] Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics*, 92-107.
- [17] Griffin, J. E. and Brown P. J. Brown(2010). Bayesian adaptive Lassos with non-convex penalization. Technical report. Institute of Mathematics, Statistics and Actuarial Science, University of Kent.
- [18] Haitovsky, Y. (1969). Multicollinearity in regression analysis: Comment. *The Review of economics and statistics*, 486-489.
- [19] Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*. 1970; 12:55–67.

- [20] Hideo Kozumi and G. Kobayashi (2011). "Gibbs sampling methods for Bayesian quantile regression". *Journal of Statistical Computation and Simulation*, 81, 1565–1578.
- [21] Hui Zou and Trevor Hastie (2005) "Regularization and variable selection via the elastic net"*J. R. Statist. Soc. B* (2005) 67, Part 2, pp. 301–320.
- [22] Koenker, R., Bassett G. J., 1978, "Regression quantiles", *Econometrica*, 46, pp.33-50.
- [23] Keming Yu & Rana A.Moyeed (2001)" Bayesian Quantile Regression" *Statistical & Probability Letters* 54:437-447.
- [24] Keming Yu, Zudi Lu and Julian Stander (2003) "Quantile regression :applications and current research areas" *The Statistician* 52, Part 3, pp. 331-350.
- [25] Koenker, R. (2005) "Quantile Regression".*Cambridge Books,Cambrdgr University Press , London .*
- [26] Koenker, R. (2004). "Quantile Regression for Longitudinal Data." *Journal of Multivariate Analysis*, 91: 74-89
- [27] Keming Yu & Julian Stander (2007) "Bayesian analysis of a Tobit quantile regression model" . *Econometrics*.vol.137,issue 1,pp. 260-276.
- [28] Koenker, R. and V. D'orey (1987). Algorithm AS 229: Computing regression quantile. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 36, 383-393.
- [29] Kordas G (2006) Smoothed Binary Regression Quantiles. *J Appl Econom* 21:387- 407.

- [30] Lindley, D. V., & Smith, A. F. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(1), 1-18.
- [31] Li Q, Xi R, Lin N (2010) Bayesian Regularized Quantile Regression. *Bayesian Analysis* 5:1-24.
- [32] Li, Y. and Zhu, J. (2008). "L1-Norm Quantile Regression." *Journal of Computational and Graphical Statistics*, 17: 163-185
- [33] Mallick, H., Yi, N., 2014, "A new Bayesian lasso", *Statistics and its interface*, 7: pp. 571-582.
- [34] Meinshausen, N. (2007). Relaxed lasso. *Computational Statistics & Data Analysis* 52 (1), 374–393.
- [35] Manski CF (1975) Maximum Score Estimation of the Stochastic Utility Model of Choice. *J Econometrics* 3:205-228.
- [36] Mallick, H., Alhamzawi, R., Paul, E., & Svetnik, V. (2021). The reciprocal Bayesian lasso. *Statistics in Medicine*, 40(22), 4830-4849.
- [37] Powell, J., 1986, "Censored regression quantiles", *Journal of Econometrics*, 32, pp. 143-155.
- [38] Qing Li, Rubin Xi & Nan Lin "Bayesian Regularized Quantile regression" *Bayesian Analysis* (2010) ,5.Number 3,pp. 533-556.
- [39] Qifan Song and Faming Liang. High-dimensional variable selection with reciprocal  $\ell_1$ -Regularization. *Journal of the American Statistical Association*, 110(512):1607-1620,2015.
- [40] Qing Li, Rubin Xi & Nan Lin "Bayesian Regularized Quantile regression" *Bayesian Analysis* (2010) ,5.Number 3,pp. 533-556.

- [41] Radchenko, P. and G. M. James (2008). Variable inclusion and shrinkage algorithms. *Journal of the American Statistical Association* 103 (483), 1304–1315.
- [42] Samuel Kotz, Tomasz J. Kozubowski, Krzysztof Podgórski(2001) “The Laplace Distribution and Generalizations” A Revisit with New Applications to Communications ,Economics,Boston:Engineering and Finance Birkauser
- [43] Song, Q. and F. Liang (2015). High-dimensional variable selection with reciprocal L1-regularization. *Journal of the American Statistical Association* 110 (512), 1607-1620.
- [44] Song, Q. (2014). Variable Selection for Ultra High Dimensional Data (Doctoral dissertation).
- [45] Tibshirani (1996)"Regression Shrinkage and selection via lasso" *J. R.Statis.Soc.*58,No.1,pp. 267-288
- [46] Tsioans , E. (2003) “Bayesian quantile inference” *Journal of statistical computation and simulation*, 73:659-674.
- [47] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476), 1418-1429.
- [48] Zhang C. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*.2010; 38(2):894–942.

## **Abstract**

Quantile regression is one of the methods that has taken a wide space in application in the previous two decades because of the attractive features of these methods to researchers, as it is not affected by outliers values, meaning that it is considered one of the robust methods, and it gives more details of the effect of explanatory variables on the dependent variable.

There are many researchers who have addressed the issue of quantile regression with binary data, that is, when the response variable ( $y$ ) takes only two values, either (0) or (1) using classical methods and also Bayesian methods. In this study, the researcher is trying to build a model for analyzing binary data using the quantile regression method (QR), as the Bayesian method was adopted, and it is one of the methods that has received wide resonance in recent times because of its accuracy, especially in small samples.

In this paper, a Bayesian hierarchical model for variable selection and estimation in the context of binary quantitative regression is proposed. Current approaches to variable selection in the context of binary classification are sensitive to outliers, heteroskedasticity values, or other anomalies. The proposed method in this study overcomes these problems in an attractive and straightforward way.

In addition, the researcher restricted the study by using inverse penalty functions to choose the explanatory variables that have an effect on the truly dependent variable and to exclude the unaffected variables using the methods of selecting variables.

And using the effective (Gibbs Sampler) algorithm to estimate the parameters of the model that outperforms on the (Metropolis) algorithm used in previous studies. The results of both the simulation study and the analysis of real data showed that the proposed method (BrLBqr) outperforms in terms of estimating the parameters of the model and selecting explanatory variables related to or affecting the response variable compared to the other methods (BLBqr) and (BBqr).

Republic of Iraq  
Ministry of Higher Education  
and Scientific Research  
Karbala University  
college of Management and Economics  
Department of Statistics



**Bayesian Binary Quantile Regression restricted by  
inverse penalty functions (with practical application)**

**A thesis**

submitted to the council of the College of Administration and  
Economics at the University of Karbala as partial fulfillment of the  
requirements for the degree Doctorate of Philosophy in Statistics

**By**

**Mohammed Taleb kahnger**

**Under Supervision**

**Prof. Dr. Ahmad Naeem Flaih**