AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

دمج البوتستراب الاحتمالي الموزون مع طريقة لارس الحصينة لاختيار المُتغيّرات في أنموذج الانحدار الخطّي بوجود مشكلتي الأبعاد العالية والقيم الشاذّة.

https://doi.org/10.29124/kjeas.1547.19

زينه حكمت عبد المنعم الأمير [1]
جامعة واسط / كلّية الإدارة والاقتصاد

أ. د. باسم شليبه مسلم[2]
الجامعة العراقية / كلّية الإدارة والاقتصاد

المستخلص

تـمّ فـي هـذا البحـث اقتـراح خوارزميـة جديـدة لاختيـار المُتغيّـرات المهمّـة فـي أنمـوذج الانحـدار بوجـود مشـكلتي الأبعـاد العاليـة والقـيم الشـاذّة، مـن خـلال توظيـف ودمـج أسـلوب البوتسـتراب الاحتمـالي المـوزون فـي طريقـة لارس الحصـينة Weighted Bootstrap probability - Robust Least Angle Regression Selecting واختصـاراً (WBP-LARS)، ومقارنتهـا مـع طريقـة اختيـار أخـرى هـي طريقـة لارس الحصـينة المعتمـدة علـى أسـلوب البوتسـتراب العـادي والمعروفـة اختصـاراً (B-LARS) تجريبيـاً بالمحاكـاة وتطبيقيـاً بالاعتمـاد علـى بيانـات حقيقيـة تتعلّـق بالقيمـة السـوقية لـبعض المصـارف الأهليـة فـي سـوق الأوراق الماليـة للمـدّة الزمنيـة 2010-2017.وقـد تضـمّنَتْ المقارنـة فـي المحاكـاة حـالتين لعـدد المُتغيّـرات التوضـيحية المطلـوب اختيارهـا (K=5, 7) فضـلاً عـن حـالتين عنـدما ($n < P$) ($n > P$) وأحجـام عيّنـات (50، 70، 20، 26) وبقيمـة ارتبـاط 0.95 وبنسـب تلـوّث مختلفـة ($\alpha = (0.05, 0.10, 0.15$) وقـد خَلُـصَ البحـث إلـى اسـتنتاجات أهمّهـا تحديـد عـدد المُتغيّـرات المهمّـة بـين الأعـداد 7-10 أفضـلية طريقـة (WBP-LARS) علـى طريقـة (B-LARS) عنـدما ($n > P$). فـي حـين ظهـرت أفضـلية بسـيطة لطريقـة (B-LARS) علـى الطريقـة المقترحـة عنـدما ($n < P$) وحجـم العيّنـة بعيـد عـن عـدد المُتغيّـرات الكلّيـة فـي الأنمـوذج، ولكـن تتقـارب الكفـاءة كلّمـا اقتـرب حجـم العيّنـة بشـكل كبيـر مـن عـدد المُتغيّـرات، ويمكـن أنْ يعتمـد عليهـا فـي عمليـة الاختيار للمُتغيّرات في هذه الحالة.

الكلمات المفتاحية : البوتستراب الاحتمالي الموزون، البوتستراب العادي ، الأبعاد العالية ، القيم الشاذّة ، طريقة LARS الحصينة.

Abstract

In this research, a new algorithm was proposed to select the important variables in the regression model with the presence of two problems of high dimensions and outliers by employing and integrating the weighted bootstrap probability - Robust Least Angle Regression Selecting (WBP-LARS) and comparing it with another selection method. It is a method of impregnable Lars based on the regular bootstrap method, known as (B-LARS), empirically simulated and applied, based on real data related to the market value of some private banks in the stock market for the period 2010-2017. The comparison in the simulation included two cases for the required number of explanatory variables Choosing (K = 5, K = 7) as well as two cases when (n>P) (n<P) and sample sizes (50, 70, 20, 26) with a correlation value of 0.95 and with different contamination rates $\alpha = (0.05, 0.10, 0.15)$ The research concluded with conclusions, the most important of which is determining the number of important variables between the numbers 7-10, the preference of the (WBP-LARS) method over the (B-LARS) method when (n>P), while a slight preference for the (B-LARS) method appeared over the proposed method when (n<P) and the sample size is far from the total number of variables in the model, but the efficiency converges whenever the sample size is very close to the number of variables and can be relied upon in the selection process for the variables in this case

Keywords: Weighted Bootstrap Probability, Classical Bootstrap, high dimensions, outliers, Robust LARS Method.

1- المقدّمة:

تمهيد 1-1

يواجـه الكثيـر مـن البـاحثين المعوّقـات والمشـكلات التـي مـن الممكـن أن تُصَـعّبَ مـن عمليـة التحليـل الإحصـائي المتضـمّنَة التقـدير Estimation واختبـار الفرضـيات Testing Hypothesis لبيانـات الظـاهرة تحـت الدراسـة. وتتنـوّع هـذه المشـكلات بحسـب ظهورهـا، فقـد تظهـر فـي بيانـات الظـاهرة نفسـها Data (أي بطبيعـة البيانـات) كمـا هـو الحـال بوجـود القـيم الشـاذّة Outliers التـي تظهـر فـي قـيم بعـض المُتغيّـرات أو كلّهـا - موضـوع الدراسـة- والتـي غالبـا مـا تكـون حاضـرة فـي الواقـع التطبيقـي ، وقـد تظهـر تلـك المشـكلات فـي هيكلّيـة (طبيعـة) الأنمـوذج الرياضـي Model (أي فـي مكوّناتـه)، والـذي يمثّـل تلـك البيانـات، كمـا هـو الحـال فـي أنمـوذج الانحـدار نتيجـة لخـرق أحـد شـروطه أو فرضـياته الأسـاس، ولا يقتصـر حـدوثها عنـد اسـتعمال أسـلوب تحليـل

الانحدار فقط ، بل قد تحدث عند استعمال أغلب الأساليب الإحصائية في عملية تحليل البيانات إحصائياً، ولا سيّما وجود القيم الشاذّة.

ومن الشروط المهمّة الواجب توفّرها عند التحليل الإحصائي لبيانات ظاهرة بأسلوب تحليل الانحدار الخطّي تحقّق شرط أنّ حجم العيّنة sample size يكون أكبر من عدد المُتغيّرات التوضيحية explanatory variables بما فيها المُتغيّر التوضيحي المرتبط بالحدّ الثابت ($n > P$)، أو بعبارة أخرى أنْ يكون حجم العيّنة أكبر من عدد المَعْلَمات في الأنموذج؛ لكي يتمّ تعريف معكوس المصفوفة $\acute{X}X$، ثُمّ تطبيق طريقة المربعات الصغرى الاعتيادية OLS بوصفها الطريقة الشائعة والمفضّلة لدى أغلب الباحثين مع تحقّق بقية الفرضيات الأساس الأخرى، ومنها: ثبات تجانس التباين للأخطاء العشوائية Error terms في الأنموذج، ما يعني خلوّ الأنموذج من مشكلة عدم تجانس التباين Heterosedasisticity، وكذلك عدم ترابط الأخطاء العشوائية للأنموذج فيما بينها ممّا يعني خلوّه من مشكلة الارتباط الذاتي Autocorrelation ، فضلاً عن عدم وجود مشاكل في البيانات نفسها، ومنها وجود القيم الشاذّة Outliers ، ما يعني الحصول على مقدّرات كفوءة لها صفة أفضل مقدّر خطّي غير متحيّز Best Linear Unbiased Estimator (BLUE)، والتي يمكن الاعتماد عليها في عملية التنبّؤ المستقبلي للظاهرة تحت الدراسة، ومن ثَمّ تسهل عملية اختيار المُتغيّرات التوضيحية المهمّة المؤثّرة فعلاً على مُتغيّر الاستجابة (المعتمد) بالاستعانة بأحد أساليب الاختيار الشائعة: (أسلوب اختيار الانحدار المتدرج(المتسلسل) Stepwise Regression ، أسلوب الاختيار الامامي Forward، وأسلوب الاختيار الخلفي Backward).

2-1 مشكلة البحث

في الواقع التطبيقي لا تتحقّق شروط الأنموذج كلّها General Linear Regression (GLR) جميعها فقد يكون ($n < P$)، ممّا يسبّب مشكلة الأبعاد العالية high dimensions ، فضلاً عن ذاك قد تكون بيانات الظاهرة المدروسة تعاني من وجود القيم الشاذّة outliers سواءً في مُتغيّر الاستجابة outliers أم النقاط الرافعة leverage points في المُتغيّرات التوضيحية أو وجودهما معا في البيانات ووجود هاتين المشكلتين في آنٍ واحدٍ.

يؤدّي إلى فقدان الطرائق الاعتيادية الشائعة كفاءتها لاختيار المُتغيّرات التوضيحية المهمّة في الأنموذج التي تؤثر على مُتغيّر الاستجابة Selection Variables والتقدير Estimation لمَعْلَماته، وبدوره يفقد الثقة بتلك المقدّرات؛ نتيجة عدم واقعيتها وابتعادها عن الحقيقة. ومن هنا تظهر مشكلة البحث فالباحث في موضوع الانحدار يهتم بعملية التنبّؤ المستقبلي للظاهرة تحت الدراسة، ممّا يفرض عليه البحث عن طرائق اختيار وتقدير مناسبة، تتّصِف بالدقّة والكفاءة بوجود مشكلتي الأبعاد العالية والقيم الشاذّة.

وقد نوقشت مشكلة البحث في كثير من البحوث المنشورة في خوارزميات الاختيار وطرائق التقدير لمَعْلَمات أنموذج الانحدار المتعدّد بوجود مشكلتي الأبعاد العالية والقيم الشاذّة، نذكر منها على سبيل المثال في عام 2005 قدّم الباحث khan وآخرون [14] بحثاً تضمّنَ أسلوبين مختلفين للحصول على مقدّرات LARS الحصين Robust؛ للتعامل مع مشكلتي الأبعاد والقيم الشاذّة في أنموذج الانحدار الخطّيGLR للحصول على أنموذج التنبّؤ الخطّي؛ لأنّ طريقة LARS تقليدية. وفي عام 2008 قدّم الباحثان Barrera & Aelst

خمسة بحوثٍ اقترحوا فيها أسلوباً تمهيدياً حصيناً للاختيار الحصين للمُتغيّرات المؤثّرة في الأنموذج بواسطة استعمال مقدّرات نقطية حصينة robust points estimators بوجود مشكلة الأبعاد العالية والقيم الشاذّة . وفي عام 2009 قدّمتْ الباحثة Midi وآخرون [17] بحثاً تضمّنَت خوارزمية بوتستراب (تمهيدية) حصينة robust bootstrap algorithm، والتي تعتمد على طريقة المربعات الصغرى المشذّبة Least LTS Trimmed squares ، والتي لا تتأثّر بالقيم الشاذّة، أُطلِقَ عليها تسمية Dynamic Robust Bootstrap-LTS (DRBLTS). وفي عام 2011 قدّمَ الباحث Alfons وآخرون [4] بحثاً لوصف أسلوب تحديد المُتغيّرات التوضيحية المؤثّرة على المُتغيّر المعتمد، وتمّ تسميته Bootstrap Robust Least Angle Regression Selected (BRLARS)، والتي اعتمدت على أسلوب بوتستراب العادي classical bootstrap . وفي عام 2016 أيضاً قدّمَ الباحث [11] Hettigoda اطروحة دكتوراه، ناقش فيها انحدار Least Angle Regression (LAR) بالتفصيل، وبيّنَ أنّه يشابه في أساس عمله لانحدار المرحلة الأمامي forward stagewise regression ولكنّه يعاني من صعوبات في العمليات الحسابية ، وفي عام 2017 قدّمَ الباحث السراي[1] اطروحة دكتوراه ركّز فيها على استعمال المقدّرات الحصينة الجزائية في عملية التقدير لمَعْلَمات أنموذج الانحدار الخطّي المتعدّد بوجود المشكلتين دون التركيز على طرائق عملية اختيار المُتغيّرات التوضيحية المهمّة. وفي عام 2021 قدّمَ الباحث Lindskou وآخرون خمسة عشرَ بحثاً تضمّنَ تطويراً لطريقة جديدة للكشف عن القيم الشاذّة outliers في البيانات ذات الأبعاد العالية high dimensional ، وقد استعانوا بالتقنيات الحاسوبية المتطوّرة لاستعمال نماذج الرسم .ويلاحظ أنّ أغلب البحوث لم تتطرّق بتركيز إلى عملية اختيار المُتغيّرات، بلْ ركّزتْ على عملية التقدير بطرائق جزائية حصينة فضلاً عن أنّها لم تتطرّق إلى أسلوب البوتستراب الاحتمالي الموزون واعتمدت البوتستراب العادي .

3-1 هدف البحث

يهدف البحث إلى اقتراح طريقة اختيار للمُتغيّرات المهمّة المؤثّرة على مُتغيّر الاستجابة في أنموذج الانحدار الخطّي المتعدّد بوجود مشكلتي الأبعاد العالية والقيم الشاذّة وبحالتين (n < P) (n > P)، من خلال الدمج بين أسلوب البوتستراب الاحتمالي الموزون Weighted Bootstrap Probability وطريقة لارس الحصينة RLARS، والتي يُرمزُ لها بـ WBP-RLARS و مقارنتها مع طريقة لارس الحصينة المدمجة مع أسلوب البوتستراب العادي B-RLARS بأسلوب المحاكاة فضلاً عن إجراء تطبيق عملي على بيانات حقيقية تتعلّق بالقيمة السوقية لبعض المصارف الأهلية في سوق الأوراق المالية للمدّة الزمنية 2010-2017

4-1 هيكلّية البحث

ومن أجل بلوغ البحث غايته فقد تَمّ تقسيمه على مباحث الأول منها تضمّنَ المقدّمة وأهمّية البحث، وكذلك مشكلة البحث وهدفه وبعض الاستعراض المرجعي لموضوع البحث. أمّا المبحث الثاني فتضمّنَ الجانب النظري للبحث متضمّناً بعض المفاهيم الأساس لمصطلحات موضوع البحث فضلاً عن استعراض طرائق الاختيار لمُتغيّرات أنموذج الانحدار الخطّي العام (المتعدّد) GLR المعتمدة على أسلوب البوتستراب Bootstrap وهي :

1- Bootstrap- Least Angle Regression Selection (B-LARS).
2- Bootstrap- Robust Least Angle Regression Selecting(B-RLARS).
3- Weighted Bootstrap probability - Robust Least Angle Regression Selecting (WB-RLARS).

وتضـمّنَ المبحـث أيضـاً طريقـة التقـدير LARS العاديـة، وطريقـة MM إحـدى أنـواع طريقـة M . فـي حـين تضمّنَ المبحث الثالث الجانب التجريبي والعملي. وأخيراً المبحث الرابع تضمّنَ الاستنتاجات والتوصيات.

2- الجانب النظري

1-2 مفاهيم أساس:

1-1-2 مشكلة الأبعاد العالية High dimensions:

عـادة تحـدث مشـكلة الأبعـاد العاليـة عنـد زيـادة عـدد المُتغيّـرات التوضـيحية فـي أنمـوذج الانحـدار، اي إنّ $(n \geq p)$ إذْ تظهـر ارتباطـات وهميـة بـين المُتغيّـرات التوضـيحية ومُتغيّـر الاسـتجابة، ممّـا يـؤدّي إلـى الاسـتدلّال الإحصـائي الخـاطئ .ويظهـر تـأثير المشـكلة علـى معكـوس المصـفوفة (XX) إذْ يكـون غيـر مُعـرّف، وعنـدها لا يوجد حلّ للنظام [1] [10].

2-1-2 القيم الشاذّة :Outliers

يُطلـقُ مصـطلح القـيم الشـاذّة values outliers علـى المشـاهدات أو القـيم التـي تختلـف بسـلوكها عـن أغلبيـة البيانـات، أيْ إنّ المشـاهدات التـي لا تشـبه بسـلوكها السـياق العـام لبيانـات العيّنـة تحـت الدراسـة. بعبـارة أخـرى أنّهـا تكـون مشـاهدات غيـر مُتّسـقة (inconsistent) مـع المشـاهدات الأخـرى ، وقـد تتواجـد هـذه القـيم فـي بيانات مُتغيّر الاستجابة أو في بيانات المُتغيّرات التوضيحية أو تتواجد في كليهما [16] [7] [9].

3-1-2 أسلوب البوتستراب : Bootstrap approach

يعـدّ أسـلوب البوتسـتراب مـن أسـاليب الإحصـاء المهمّـة الـذي يسـتند علـى فكـرة إعـادة المعاينـة لعـدد محـدّد مـن المرات بحسب قناعة الباحث وهو على نوعين2] [[16][12]:

أولا :البوتستراب التقليدي (البواقي) Classical Bootstrap

يعتمـد البوتسـتراب التقليـدي علـى تقـديرات OLS للحصـول علـى بقايـا البيانـات الأصـلية. ثُـمّ يـتمّ الحصـول علـى عيّنات bootstrap من خلال إعادة أخذ العيّنات المُتبقّية من الانحدار الأصلي. وخوارزميته كالآتي :

1- تطبيـق طريقـة OLS بالاعتمـاد علـى بيانـات العيّنـة الأصـلية للمشـاهدات للحصـول علـى مقـدّرات المَعْلَمـات $\hat{\beta}_{ols}$ ، ثُمّ إيجاد القيم التقديرية لبيانات مُتغيّر الاستجابة . $\hat{Y}_i = f(X_i, \hat{\beta}_{ols})$

2- استخرجْ الواقي $e_i = Y_i - \hat{Y}_i$ وإعطاء احتمال $\frac{1}{n}$ لِكُلّ قيمة e_i .

3- اسحبْ عيّنة عشوائية من bootstrap مع الاستبدال ، أي e_i^* مأخود من e_i ومرفق بها $\hat{Y}_i$ للحصول على قيم البواقي Y_i^{*b} أي إنّ :

$$Y_i^{*b} = f(X_i, \hat{\beta}_{ols}) + e_i^{*b}$$

4- تطبيق طريقة OLS بالاعتماد على بيانات bootstrap Y_i^{*b} على X الثابت للحصول عليها $\hat{\beta}_{ols}^{*b}$

5- كرّرْ الخطوتين 3 و 4 للحصول على مرات B $\hat{\beta}_{ols}^{*1}, \dots, \hat{\beta}_{ols}^{*B}$ إذْ B هي مكرّرات bootstrap.

ثانياً : بوتستراب الأزواج (المشاهدات) Pairs Bootstrap

تنصّ هذه الطريقة على أخذ ازواج من العيّنة الأصلية ذات الحجم n وهي :

$$(x_1, y_2), (x_2, y_2), \dots, (x_n, y_n)$$

ويحسب عدد B من العيّنات كأزواج من حجم العيّنة الأصلية نفسه بإرجاع ونحصل على عيّنات البوتستراب:

$$X^{*b} = (x_1^*, y_1^*), (x_2^*, y_2^*), \dots, (x_n^*, y_n^*) \quad , \ b = 1,2, \dots, B$$

وبقدر عدد العيّنات البوتسترابية B يتمّ إيجاد مقدّر المربعات الصغرى البوتسترابية لنحصل على :

$$b_b^* = b_1^*, b_2^*, \dots, b_B^*$$

إذْ يمثّل المتّجه b_b^* بصفه عامّة مقدّرات بوتستراب أزواج المشاهدات لمَعلَمات الانحدار الخطّي ، وتكوّن تلك المقدّرات ما يعرف بالتوزيع البوتستراپي والذي توقُّعه يمثّل تقدير لموجّه مَعْلَمات المجتمع . β

2-2 أنموذج الانحدار الخطّي بوجود الأبعاد العالية ومشكلة القيم الشاذّة.

يختلف أنموذج الانحدار الخطّي ذو الأبعاد العالية عن أنموذج الانحدار الخطّي العادي في عدد المُتغيّرات التوضيحية، إذْ يكون فيه عدد المُتغيّرات التوضيحية أكبر من حجم العيّنة، أيْ إنّ (p>n) عندها تظهر مشكلة الأبعاد العالية High Dimensional) (HD ويُعرّف كالآتي [10][7]:

$$Y = X\beta + \varepsilon \quad , \ i = 1,2, \dots, n \quad , j = 1,2, \dots, p \quad , p > n$$

وبغض النظر عن هذين النوعين من النماذج فإنّ اهتمام الباحث ينصبّ على مجموعة من الأهداف، وهي: التقدير، والتنبّؤ، واختيار المُتغيّرات variable selection [14]. ويعدّ اختيار المُتغيّرات التوضيحية المهمّة في مقدّمتها؛ لأنّه يحدّد المُتغيّرات المهمّة في الأنموذج، ويبقيها فيه ويهمل الأخرى، ممّا يؤدّي إلى خفض عدد المُتغيّرات التوضيحية في الأنموذج بأكمله، ما يعني اعتماد المقدّرات التي تدخل في المعادلّة التقديرية فقط لغرض إجراء التنبّؤ؛ ولذا نجد أنّ أغلب البحوث قد ركّزتْ على عملية التقدير دون طرائق اختيار المُتغيّرات مع إنّ طرائق الاختيار لِكُلّ منها خصائصها التي تتميّز بها عن الأخرى، والتي قد تفقدها عند وجود مشكلات أخرى، ومنها مشكلة وجود القيم الشاذّة مُتغيّر الاستجابة والنقاط الرافعة في المُتغيّرات

التوضـيحية أو فـي كليهمـا؛ لـذا اقترحـت أيضـاً طرائـق اختيـار مُتغيّـرات حصـينة للحفـاظ علـى قـوة الطريقـة والاختيـار، ومنهـا طريقـة لارس الحصـينة Robust LARS Method والتـي تُعَـدُّ مـن الطرائـق الجيـدة فـي عملية الاختيار لتقارن مع الطريقة المقترحة من قبل الباحثة بدمج أسلوب البوتستراب وهي كالآتي :

1-2-2 دمـج أسـلوب البوتسـتراب مـع طريقـة لارس الحصـينة Bootstrap-Robust LARS Method

للحصـول علـى طريقـة البوتسـتراب لارس الحصـينة B-RLARS لابُـدّ مـن عـرض طريقتـي LARS العاديـة وLARS الحصـينة فقـد اقتـرح Efron وآخـرون [8] فـي عـام (2004) طريقـة انحـدار الزاويـة الصـغرى Least Angle Regression وطريقـة LARS، والتـي هـي طريقـة اختيـار للمُتغيّـرات والتـي تـرتبط ارتباطـا وثيقـا بـإجراءات الاختيـار المرحلـي الأمـامي وطريقـة لاسـو LASSO (Tibshirani 1996) SFS ، إذ يـوفّر LARS ترتيبـا تـدخل فيـه المُتغيّـرات التوضـيحية أو يثبتهـا فـي أنمـوذج الانحـدار الخطّـي، وهـذا الترتيـب أو التسلسـل للمُتغيّـرات المهمّـة غالبـا مـا يكـون نفسـه فـي LASSO أو SFS، ولكـن يـتُمّ الحصـول عليـه بطريقـة فعّالـة حسـابيا، وهـذه ميـزة تحسـب لطريقـة LARS. وهنـاك ميـزة أخـرى تحسـب للطريقـة هـي أنّـه يمكـن إيجـاد الترتيـب أو التسلسـل للمُتغيّـرات التوضـيحية المهمّـة فـي أنمـوذج الانحـدار مـن مصـفوفة الارتبـاط للبيانـات دون الاعتمـاد علـى البيانـات نفسـها. وللتوضـيح نفـرض تـوفّر المُتغيّـرات Y، X_1 ، . . . ،. X_K وبشـكل معياريـة standardized بالاعتمـاد علـى الوسـط الحسـابي والانحـراف المعيـاري لِكُـلّ مُتغيّـر ، ولنفـرض أنّ r_{jY} يمثّـل معامـل الارتبـاط بـين مُتغيّـر الاسـتجابة Y والمُتغيّـر التوضـيحي X_j ، وخوارزميـة طريقـة LARS كـالآتي [8] [11]:

1- لـتكن B مجموعـة ستضـمّ المُتغيّـرات المهمّـة أو الفعّالـة وفـي أول الأمـر $B = \emptyset$وايضـاً لـيكن $S_B = \emptyset$ يمثّـل موجّه الإشارات للارتباطات بين المُتغيّرات التوضيحية .

2- نحـدّد قيمـة أكبـر قيمـة مطلقـة للارتباطـات بـن المُتغيّـرات التوضـيحية وإشـارتها وتحديـد قيمـة الارتبـاط r وكـالآتي :

$$X_m = \text{argmax}|r_j| \quad , s_m = \text{sign}(r_m) \quad , \; r = s_m r_m$$

3- ضـمّ المُتغيّـر X_m إلـى المجموعـة B فـي الخطـوة الأولـى، وكـذلك ضَـمّ إشـارة الارتبـاط إلـى الموجّـه S_B أيْ إنّ :

$$B \leftarrow B \cup \{X_m\} \quad , \quad S_B \leftarrow S_B \cup \{s_m\} .$$

4- حساب المقدار الآتي : $$b = [\Upsilon_B (H_B C_B H_B)^{-1} 1_B]^{\frac{-1}{2}}$$

إذْ إنّ:

1. موجّـه عناصـره كلّهـا مسـاوية إلـى الواحـد الصـحيح ومرتبتـه (عـدد المُتغيّـرات المهمّـة أو الفعّالـة مضـروبا بـ 1_B

H_B : : مصفوفة قطرية عناصرها عناصر الموجّه S_B أيْ إنّ $\mathrm{H_B = diag(S_B))}$ (.

C_B : : مصفوفة جزئية من مصفوفة الارتباطات C_X والتي تضُمّ ارتباطات المُتغيّرات المهمّة أو الفعّالة.

ويتُمّ في هذه الخطوة احتساب الآتي :

$$\theta_B = b(H_B C_B H_B)^{-1} 1_B \quad , \; b_j = \left(H_B r_{jB}\right)' \; \theta_B \, , \; j \in B^c \, ,$$

r_{jB} : موجّه الارتباطات بين المُتغيّر X_j والمُتغيّرات التوضيحية المهمّة أو الفعّالة.

B^c : المجموعة المكملة للمجوعة B وتضُمّ المُتغيّرات التوضيحية التي لم يتُمّ اختبار أهمّيتها بعد.

وتجدر الإشارة هنا أنّه في حالة وجود مُتغيّر مهمّ أو فعّال واحد فقط تَمّ اختياره فإنّه يكون :

$$b = 1 \; , \quad \theta = 1 \; , \quad b_j = r_{jm}$$

5- لِكُلّ مُتغيّر بتسلسل j إذْ إنْ $(X_j \in B^c)$ نقوم باحتساب المسافتين الآتيتين:

$$\delta_j^+ = \frac{r - r_{jY}}{b - b_j} \quad , \quad \delta_j^- = \frac{r + r_{jY}}{b + b_j}$$

ويتُمّ اختيار القيمة الأقلّ، أيْ إنّ : $\delta_j = \min(\delta^+, \delta^-)$

وبالاعتماد على قيمة δ_j يتمّ تحديد المسافة δ كالآتي: $\delta = \min(\delta_j, X_j \in B^c)$

فإذا قيمة δ الأقلّ مساوية إلى δ^+ نضع $s_{min} = +1$ ، وأي شيءٍ آخر نضع $s_{min} = -1.$

وفي هذه الخطوة يَتُمّ تعديل قيم الارتباط وكالآتي : $\forall X_j \in B^c$. $r_j = r_j - \delta b_j \quad , \quad r \leftarrow r - \delta b$

6- نعيد إجراء الخطوات 3، 4، 5 .

و يمكن كتابة مصفوفة الارتباطات للمُتغيّرات التوضيحية كالآتي :

$$C_X = [1 \; r_{12} \; \cdots \; r_{1K} \; \vdots \; \ddots \; \vdots \; r_{K1} \; r_{K2} \; \cdots \; 1 \;] \quad \ldots(1)$$

وبافتراض أنّ r_{mY} يملك القيمة المطلقة الأعلى من بين قيم الارتباطات بين المُتغيّرات التوضيحية ومُتغيّر الاستجابة، وايضاً نفرض أنّ s_m يمثّل إشارة الارتباط، أيْ إنّ $s_m = sign(r_{mY})$.

ووفقًا لما تُقدّم فإنّ المُتغيّر X_m يصبح المُتغيّر المهمّ أو الفعّال الأول الواجب بقاؤه في الأنموذج ومن ثمّ يكون التنبّؤ الحالي لمُتغيّر الاستجابة $\hat{\mu}$ وبعبارة أخرى $\hat{\mu} \to 0$ ، إذْ التنبّؤ يتطلّب تعديله عن طريق التحرّك على طول اتّجاه S_mX_m حتى مسافة معيّنة δ يمكن التعبير عنها من حيث الارتباطات بين المُتغيّرات ، وبتحديد قيمة المسافة δ تقوم طريقة LARS بتحديد المُتغيّر المهمّ الفعّال الثاني الواجب وجوده في الأنموذج، وبمجرد أنْ يكون لدينا أكثر من مُتغيّر مهمّ أو فعّال أو نشط ، يقوم LARS بتعديل التنبّؤ الحالي على طول الاتّجاه المتساوي ، أيْ الاتّجاه الذي له زاوية متساوية (ارتباط) مع المُتغيّرات المشتركة الفعّالة جميعها أو المهمّة الموجودة في الأنموذج. ومن هذه الآلية (اعتمادها على الزاوية المتساوية للمُتغيّرات المشتركة جميعها) اكتسبت الطريقة تسميتها [11] [8].

إنّ التحرّك على طول هذا الاتّجاه وفق الزاوية المتساوية يضمن أنّ الارتباط الحالي لِكُلّ مُتغيّر مُهمّ أو فعّال ينشط مع بقية المُتغيّرات الأخرى سينخفض بالتساوي، وقد وضّحَ الباحث Khan وآخرون [14] أسلوب الاشتقاق لتحديد المسافة δ لمُتغيّر توضيحي مهمّ أو فعّال في ملحق البحث تسلسل A (يمكن الاطّلاع عليه) وقيمتها كالآتي:

$$\delta(+X_j) = \frac{r - r_{jY}}{1 - r_{jm}} \quad \ldots(2)$$

$$\delta(-X_j) = \frac{r + r_{jY}}{1 + r_{jm}} \quad \ldots(3)$$

وعلى ضوء المعادلّتين (2) (3) يتمّ اختيار القيمة الأقلّ منهما، وكذلك الحدّ الأدنى على المُتغيّرات غير المهمّة أو الفعّالة جميعها، ثمّ ووفقاً لقيمة المسافة δ ، يُثمّ تحديد المُتغيّر المشترك الذي يدخل إلى الأنموذج مرافقاً للمُتغيّر السابق في الاختيار .

واعتماداً على ما تُقدّم أشار الباحث Khan وآخرون أنّ طريقة LARS بشكلها الحالي تتأثّر بتلوّث البيانات. وقد وضّحَ ذلك بمثال يمكن الاطّلاع عليه ، ولذا اقترحوا أسلوبين لزيادة الحصانة لطريقة LARS وسُمّي الأسلوبين: بالإضافة، والتنظيف plug- in and cleaning فأسلوب plug- in يستند إلى فكرة استبدال مكوّنات طريقة LARS غير الحصينة non robust وهي:(المتوسّط، والتباين، والارتباط) بمكوّنات حصينة robust، وأكثر مقاييس النزعة المركزية حصانة وأقلّ تأثيرا بالقيم الشاذّة هو الوسيط median (med)) في حين يُعدّ مقياس التشتّت (mad median absolute deviation) الأكثر حصانة. ولكن للحصول على الارتباط الحصين robust correlation يتوجّب توفّر على بيانات بقدر d من الأبعاد، وهذا يعني تطلّبه وقتاً طويلاً بعكس المقياسين الآخرين. وأشار الباحث Khan وآخرون أنّ الباحث Huber ذكر في كتابه عام 1981 أنّ أساليب الأزواج الحصينة robust pairwise ليست متكافئة، وهذا يعني أنّها حسّاسة للقيم المتطرّفة ثنائية الأبعاد، لذا يُعدُّ أحد الحلول للحصول على الارتباط الحصين باستعمال ارتباطات قوية مشتقّة من مقدّر التغاير المتساوي الأزواج. وهناك طريقتان لإيجاد الارتباط الحصين إحداهما طريقة مقدّر M ثنائي المُتغيّر، الذي عرّفه الباحث Maronna ، والثاني حساب الارتباط ثنائي المُتغيّر من بيانات Windsorized ثنائية المُتغيّر، وبقدر تعلّق البحث سيتُمّ الاعتماد على الطريقة الثانية.

تُعدّ مهمّة الحصول على مقدّر ارتباط حصين وبحساب سريع من أهمّ المعوّقات التي تواجه الباحثين بوجود بيانات الأبعاد العالية high dimensions. في 1981 قدّمَ الباحث Huber فكرة البيانات أحادية البعد وندزريشن one-dimensional Windsorization وبالاعتماد على البيانات المحوّلة إلى الصيغة القياسية يتُمّ حساب معاملات الارتباط التقليدية، ثُمّ جرت إعادة تدقيقية لأسلوب الباحث Huber لتقدير عناصر مصفوفة الارتباطات ذات الأبعاد العالية أو الكبيرة، إذْ إنّ البيانات المحوّلة إلى الصيغة القياسية لبعد واحد (مُتغيّر واحد بحجم عيّنة n) ليتُمّ الحصول عليها كالآتي :

$$u_i = \frac{f_c((X_i - med(X_i))}{mad(X_i)} \quad , i=1, 2..., n \tag{4}$$

$$f_c = \min[\max(-c , X_i), c] \tag{5}$$

إذْ إنّ قيمة c يتُمّ اختيارها من قبل الباحث .

ونستنتج من الصيغة (4) أنّ أسلوب (Huber one-dimensional Windsorization) يتميّز بسرعة الحسابات الرياضية لتحويل البيانات بشكل الصيغة القياسية، ولكن في البيانات ذات الأزواج .ولكن تفقد هذه الميزة في البيانات الثنائية bivariate data على الرغم من وجود هذه الميزة لكن لا يتغلّب على تأثير القيم الشاذّة بشكل نهائي، ونجد أنّ بعض قيم البيانات يتركها دون تغيير.

ولعلاج هذا القصور في التحويل Windsorization الأحادي يمكن تحويل البيانات وندزريشن الثنائي bivariate Windsorization اعتماداً على القطع الناقص الأولي للتسامح الأولي لغالبية البيانات أو التفاوت المسموح initial tolerance ellipse إذْ تمّ تقليص القيم الشاذّة إلى حدّ هذا القطع الناقص باستعمال التحويل ثنائي المُتغيّر الآتي:

$$u\left(\min\sqrt{\frac{c}{d(X)}} ,1\right)X \quad , \quad X = (X_1, X_2)' \tag{6}$$

$d(X)$: تمثّل مسافة مهلنوبس Mahalanobis distance التي يتمّ حسابها اعتماداً معامل الارتباط الثنائي الأولي في مصفوفة الارتباطات والذي يمثّل معامل الارتباط التقليدي كي يتمَّ حساب بيانات Windsorization المعدلّة adjusted Windsorization data والتي تحسب بالصيغة الآتية [9]:

$$d(X) = [(X_1 - X_2)' \ R_0 \ (X_1 - X_2)]^{0.5} \tag{7}$$

ما تقدّم يمثّل طريقة لارس الحصينة Robust LARS واختصاراً RLARS، ولكن يمكن توظيف أسلوب البوتستراب Bootstrap Approach في عملية اختيار المُتغيّرات التوضيحية المهمّة أو الفعّالة في أنموذج الانحدار الخطّي بوجود البيانات ذات الأبعاد العالية high dimensions من خلال الدمج مع طريقة RLARS للحصول على نتائج أكثر واقعية وموثوقية إذْ يتمّ توليد M من عيّنات البوتستراب (التمهيد) من البيانات الأصلية وفق النوع الثاني من أنواع البوتستراب ، ونستعمل بعد ذلك طريقة RLARS لتحديد التسلسل المقابل لِكُلّ من المُتغيّرات التوضيحية لِكُلّ عيّنة من عيّنات البوتستراب وتصنيف المُتغيّرات بشكل

رُتّبَ من التسلسل 1 إلى التسلسل V من حيث الأهمّية ، ثُمّ لِكُلّ مُتغيّر مشترك يمكن أنْ نأخذ المتوسّط على هذه الرتب M ، والمُتغيّرات المشتركة m التي تمتلك أصغر متوسّط الرتب تكون المجموعة المخفضة أو المجموعة المختزلة reduced set. وعليه يمكن وضع خوارزمية لتطبيق B-RLARS لاختيار المُتغيّرات في أنموذج الانحدار بوجود البيانات ذات الأبعاد العالية وكالآتي [14][11]:

1- توليد عيّنات بعدد M من عيّنات البوتستراب Bootstrap samples باختيار زوج البيانات مع الإرجاع من البيانات الأصلية، أي إنّ :

نفرض أنّ هناك بيانات للمُتغيّرات التوضيحية ومُتغيّر الاستجابة ($Y \quad , \quad X_1 \quad ,X_2,X_3,\ldots,X_K$) فتكون أزواج المشاهدات كالآتي[12] [17] :

$$(Y_i, X_{i1}, X_{i2}, \ldots X_{ik}) = (Y_i, X_i), \quad (X_i = X_{i1}, X_{i2}, \ldots X_{ik}) , i = 1,2, \ldots, n \qquad (8)$$

وبالاعتماد على تعريف البيانات في (8) ستكون البيانات الأصلية (أزواج البيانات)، التي سيتُمّ سحب عيّنات البوتستراب منه كالآتي:

$$(Y_1, X_1) , (Y_2, X_2), \ldots , (Y_n, X_n) \qquad (9)$$

وعلى ضوء بيانات الأزواج المعرفة في (9) يَتُمّ سحب عيّنات البوتستراب مع الإعادة وكالآتي:

$$(\dot{Y}_{1m}, \dot{X}_{1m}) , (\dot{Y}_{2m}, \dot{X}_{2m}), \ldots , (\dot{Y}_{nm}, \dot{X}_{nm}) \quad , \ m = 1,2, \ldots, M \qquad (10)$$

2- تطبيق طريقة RLARS على كُلّ عيّنة واحدة من عيّنات البوتستراب للحصول على تسلسل أو رتب المُتغيّرات التوضيحية بحسب الأهمّية.

3- ترتيب المُتغيّرات أو وضعها بتسلسل بحسب أهمّيتها بالاعتماد على نتائج الخطوة 2.

4- إعادة إجراء الخطوات 1,2,3 لـ M من عيّنات البوتستراب.

5- يتُمّ حساب الوسط الحسابي لرتب كُلّ مُتغيّر توضيحي، والتي حصل عليها من الخطوات 1,2,3,4 وهي بقدر M من عيّنات البوتستراب.

6- ترتيب المُتغيّرات التوضيحية بحسب أهمّيتها بحسب أقيام الأوساط الحسابية للرتب التي حصلت عليها من الخطوة 5 ،ويكون التسلسل الأفضل للمُتغيّر الذي يملك أصغر وسطٍ حسابيّ من بينها.

7- يتُمّ تحديد المجموعة المختزلة من المُتغيّرات التوضيحية المهمّة التي تَمّ ترتيبها في الخطوة 6 وبعدد 5 او 7 من المُتغيّرات و لا يتعدّى 10 مُتغيّرات.

1-2-2 الطريقة المقترحة لاختيار المُتغيّرات Suggest method for selection

تستند فكرة الطريقة المقترحة لاختيار المُتغيّرات المهمّة في أنموذج الانحدار، والحصول على المجموعة المختزلة بتوظيف أسلوب البوتستراب الموزون الاحتمالي، ودمجه مع طريقة لارس الحصينة Weighted Bootstrap with probability RLARS واختصاراً WBP-RLARS. ولذا فعملية كتابة الخوارزمية للطريقة المقترحة تستوجب عرض أسلوب البوتستراب (التمهيد) الموزون الاحتمالي Weighted Bootstrap with probability ، واختصاراً WBP ، ثُمّ دمجها مع خوارزمية RLARS وكالآتي [18] :

يعدّ هذا البوتستراب تعديلا على طريقة التشخيص قبل البوتستراب Diagnostic –Before Bootstrap المقدّم من قبل الباحث Imon وآخرون [12] لوقاية إجراء البوتستراب من القيم الشاذّة Outliers Values ، وفيها يتُمّ تحديد شيئين: الأول (تحديد أوزان للمشاهدات استعملت دالّة هامبل المرجحة Hamble's weighting Psi في تعيين الأوزان weights لنقاط البيانات كُلّها، والثاني (تحديد الاحتمالات لِكُلّ زوج من البيانات بالاعتماد على الأوزان المستخرجة)؛ ولذا فالأوزان تستخرج بحسب الصيغة الآتية:

$$\emptyset(u)_{Hample} \begin{cases} u & 0 \le abs(u) \le a \\ a\ sign(u) & a \le abs(u) \le b \\ a(c - abs(u))/(c-b)\ sign(u) & b \le abs(u) \le c \\ 0 & c \le abs(u) \end{cases} ,$$

$$\{ a = 1.31 ,\ b = 2.039 ,\ c = 4 \} \quad (11)$$

وبهذا يتوقّعُ أنْ تأخذ القيم الشاذّة outliers values أوزاناً تبعا لخطورتها .

ويتُمُّ حساب هذه الأوزان من البواقي القياسية لطريقة مربعات الوسيط الصغرى Least median squares (LMS) [19][4] وكالآتي :

$$u_i = \frac{r_i}{MAD(r_i)} \quad , i = 1,2, \dots n \quad (12)$$

واعتماداً على الصيغة (12) نستخرج الأوزان وكالآتي :

$$w_i(u_i) = \frac{\emptyset(u_i)}{u_i} \quad , \ i = 1,2, \dots, n \quad (13)$$

وبناءً على الصيغة (13) يتمّ استخراج الاحتمالات المقابلة لِكُلّ زوج من المشاهدات وفق الصيغة الآتية :

$$p_i = \frac{w_i}{\sum_{i=1}^{n} w_i} \qquad i = 1,2,\dots,n$$

(14)

واعتمادا على الاحتمالات المعرفة في المعادلّة 14 يتُمّ تخصيص كُلّ احتمال لزوج من البيانات، بمعنى أنّ الاحتمالات $p_1, p_2, \dots, p_n$ مُخصّصة لِكُلّ زوج من البيانات $(Y_1, X_1)\ , (Y_2, X_2), \dots, (Y_n, X_n)$ وتبعاً لقمة الاحتمالات التي قد تكون صفرا عندما تكون هناك قيم شاذة يتُمّ تقسيم البيانات على جزأين: البيانات الباقية R ، والبيانات المحذوفة D ، إذْ يجب أن يكون للبيانات الباقية قيم احتمالية أكبر من الصفر ووفقًا لذلك يمكن تعريف البيانات كالآتي :

$$X = [X^R\ X^D\ ]\ , Y = [Y^R\ Y^D\ ] \tag{15}$$

وبافتراض أنّ موجّه مقدّرات المَعْلَمات لأنموذج الانحدار هو $\hat{\beta}^D$ بعد حذف D من البيانات الشاذّة، والذي يستخرج بالاعتماد على البيانات الباقية لمُتغيّر الاستجابة Y^R والمُتغيّرات التوضيحية X^R، ويمكن توضيح خوارزمية طريقة **WBP** وكالآتي :

1- تقدير مَعْلَمات الأنموذج اعتمادا على البيانات الِكلّية من دون الحذف بطريقةً LMS للحصول على أوزان هامبل Hample ثُمّ تحديد الأوزان الأكبر من الصفر، والتي تقابل زوج البيانات الباقية؛ لكي نحصل على موجّه المقدّرات $\hat{\beta}^D$ ، ثُمّ استخراج القيم التنبّئية لمشاهدات مُتغيّر الاستجابة وكالآتي:

$$\hat{Y}_i^D = g(x_i, \hat{\beta}^D) \tag{16}$$

2- نستخرج البواقي residuals من خلال العلاقة الآتية :

$$\hat{\varepsilon}_i^D = Y_i - g(x_i, \hat{\beta}^D) \tag{17}$$

3- سحب مجموعة من البواقي الجديدة $\dot{\varepsilon}_i^D$ من البواقي المستخرجة من الخطوة 2 مع الإرجاع باستعمال البوتستراب، واستخراج الاحتمالات الجديدة وفقاً للصيغة 14 ، وباعتماد البواقي الجديدة، ومنها استخراج القيم التنبّئية لمُتغيّر الاستجابة الجديدة $\dot{Y}_i^D$ إذْ إنّ :

$$\dot{Y}_i^D = g(x_i, \hat{\beta}^D) + \dot{\varepsilon}_i^D \tag{18}$$

4- استخراج موجّه المقدّرات الجديد $\hat{\beta}^D$ بالاعتماد على بيانات البوتستراب المُتغيّر الاستجابة والمُتغيّرات التوضيحة المقابلة لها .

5- إعادة إجراء الخطوات 3,4، لعدد من المرات M من عيّنات البوتستراب للحصول على M من موجّهات المقدّرات المَعْلَمات . $\hat{\beta}_m^D, m = 1,2,\dots,M$

ووفقًا لخوارزمية WBP يمكن توظيفها مع طريقة RLARS للحصول على طريقة اختيار مقترحة حصينة أخرى تعتمد على أسلوب البوتستراب الاحتمالي with Probability Weighted Bootstrap تُسمّى WBP-RLARS وخوارزميتها هي خليط من خوارزمية طريقتي WBP و LARS وكالآتي :

1- توليد عيّنات بعدد M من عيّنات البوتستراب Bootstrap samples للحصول على زوج البيانات من المُتغيّر الاستجابة، والمُتغيّرات التوضيحية باتّباع خطوات خوارزمية WBP الموضّحَة سلفاً ، أيْ الحصول على الآتي :

$$\left(\dot{Y}_{1m}, \dot{X}_{1m}\right), \left(\dot{Y}_{2m}, \dot{X}_{2m}\right), \dots, \left(\dot{Y}_{nm}, \dot{X}_{nm}\right) \quad , \; m = 1,2, \dots, M \tag{19}$$

2- تطبيق طريقة RLARS على كُلّ عيّنة واحدة من عيّنات البوتستراب للحصول على تسلسل أو رتب المُتغيّرات التوضيحية بحسب الأهمّية.

3- ترتيب المُتغيّرات أو وضعها بتسلسل بحسب أهمّيتها بالاعتماد على نتائج الخطوة 2.

4- إعادة إجراء الخطوات 1,2,3 لـ M من عيّنات البوتستراب.

5- يتُمّ حساب الوسط الحسابي لرتب كُلّ مُتغيّر توضيحي والتي حصل عليها من الخطوات 1,2,3,4 وهي بقدر M من عيّنات البوتستراب.

6- ترتيب المُتغيّرات التوضيحية بحسب أهمّيتها بحسب أقيام الأوساط الحسابية للرتب التي حصلت عليها من الخطوة (5) ،ويكون التسلسل الأفضل للمُتغيّر الذي يملك أصغر وسطاً حسابياً من بينها.

7- يتُمُّ تحديد المجموعة المختزلة من المُتغيّرات التوضيحية المهمّة، التي تَمّ ترتيبها في الخطوة 6 وبعدد 5 أو 7 من المُتغيّرات ولا يتعدّى 10 مُتغيّرات.

3-2 طريقة التقدير MM

اقترح الباحث Yohai في عام 1987 تقديراً ،هو مزيج من تقدير قيمة الانهيار العالية high breakdown value estimation والتقدير الفعّال efficient estimation أُطلق عليه تسمية MM estimation ، وهي طريقة مطوّرة عن طريقة تقدير M ، ويكون مقدّر MM هو الحلّ للمعادلّة الآتية [19][4][21]:

$$\sum_{i=1}^{n} \; T(u_i) \left(\frac{Y_i - \sum_{j=0}^{K} X_{ij}\hat{\beta}_j}{s_{MM}}\right) X_{ij} = 0 \tag{20}$$

S : يمثّل الانحراف المعياري المستحصل عليه من بواقي تقدير s_{MM}

$T(u_i)$: تمثّل دالة توكي ثنائية الوزن Tukey's biweight function وتساوي :

$$T(u_i) = \begin{cases} \frac{u_i^2}{2} - \frac{u_i^4}{2(1.547)^2} + \frac{u_i^6}{6(1.547)^2} & , -1.547 \le u_i \le 1.547 \\ \frac{(1.547)^2}{6} & u_i < -1.547 \ or \ u_i > 1.547 \end{cases}$$

(21)

$$u_i = \frac{\varepsilon_i}{\sigma_s} \quad , \quad \hat{\sigma}_s = \frac{\text{median}\,|\varepsilon_i - \text{median}(\varepsilon_i)|}{0.6745}$$

(22)

ووفقًا لما تُقدّم يمكن وضع خوارزمية طريقة تقدير MM وكالآتي [21] :

1- إيجاد تقدير مَعْلَمات أنموذج الانحدار بطريقة OLS.

2- اختبار الفرضيات الأساس لأنموذج الانحدار الخطّي العادي Classical regression.

3- إجراء اختبار للكشف عن وجود القيم الشاذّة في البيانات.

4- حساب قيم البواقي residual بالاعتماد على مقدّرات S وفق المعادلّة $\varepsilon_i = Y_i - \hat{Y}_i$.

5- حساب قيمة . $\hat{\sigma}_s$.

6- حساب قيمة $u_i = \frac{\varepsilon_i}{\sigma_s}$ ،$i = 1,2,\dots,n$.

7- حساب قيمة الوزنw_i وفق الصيغة الآتية :

$$w_i = \begin{cases} \left[1 - \left(\frac{u_i}{4.685}\right)^2\right]^2 & , \quad |u_i| \le 4.685 \\ 0 & |u_i| > 4.685 \end{cases} , i = 1,2,\dots,n$$

(23)

8- حساب مقدّر مَعْلَمات أنموذج الانحدار $\hat{\beta}_{MM}$ بطريقة المربعات الصغرى الموزونة WLS بالاعتماد على الأوزان المحسوبة بالخطوة 7.

9- إعادة إجراء الخطوات من 5- 8 للحصول على تقارب بين قيم المقدّرات . $\hat{\beta}_{MM}$

10- إجراء اختبار معنوية المَعْلَمات؛ لتحديد أيّ من المُتغيّرات التوضيحية لها تأثير على مُتغيّر الاستجابة .

3- المحاكاة والتطبيق العملي

3-1 المحاكاة :

3-1-1 خطّة المحاكاة :

تـمّ تنفيـذ برنـامج المحاكـاة باسـتعمال لغـة R. إذْ تضـمّنَ البرنـامج توليـد البيانـات بالاعتمـاد علـى نمـوذج الانحـدار الخطّي المتعدّد الآتي :

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

إذْ أن X هـي مصـفوفة المُتغيّـرات التوضـيحية ذات بعـد $n \times p$ سـيتُمّ توليـدها فـي كُـلّ مـرّة خـلال توزيـع طبيعـي متعدد المُتغيّرات بمتوسّط قيمته صفراً، وتباين $\sigma = \rho^{|i-j|}$ ،أي إنّ :

$$X \sim N(0, \rho^{|i-j|}) \quad , \rho = 0.95 \text{ مع } (n = 50,70) \ , \ \rho = 0.50 \text{ مع } n = 20,26$$

أمّـا الأخطـاء العشـوائية ε فقد تـمّ توليـدها وفقـاً للتوزيـع الطبيعـي بمتوسّـط صـفر وتبـاين σ^2 ، والـذي تـمّ افتراضـه 1 .

وانّ القيم الأولية للمَعْلَمات كالآتي:

$$\beta = [3,0,3,0,0,3,0,0,0,3,0,0,0,0,3^{\frown 15}, 0,0, \dots, 0^{\frown 13}]_{1\times 28}$$

إذْ تـمّ اختيـار خمسـة مُتغيّـرات بقـيم لا صـفرية ، والمُتبقّـي مـن المُتغيّـرات التوضـيحية بقـيم صـفرية كمـا موضّـحَ فـي فيما تقدّم.

وقـد تـمّ تحديـد عـدد المُتغيّـرات التوضـيحية pبــ 28 مُتغيّـراً (مسـاوٍ لعـدد المُتغيّـرات التوضـيحية فـي التطبيـق العملـي). وأمّـا أحجـام العيّنـات n فكانـت بحـالات عـدّة منهـا مـا هـو أكبـر مـن عـدد المُتغيّـرات التوضـيحية $n > p$ كمـا فـي أحجـام (50,70) ، وبعضـها الآخـر كانـت أصـغر مـن عـدد المُتغيّـرات التوضـيحية $p > n$ كمـا فـي أحجـام (20,26) ، وأيضـاً لضـمان وجـود القـيم الشـاذّة والنقـاط الرافعـة تـمّ تثبيـت نسـب التلـوّث $\alpha = (0.05, 0.10, 0.15)$.

وقد تمّتْ عملية التلويث كالآتي:

1. تلويـث المُتغيّـرات اللاصـفرية $(X_1, X_3, X_6, X_{10}, X_{15})$فـي مصـفوفة المُتغيّـرات التوضـيحية بــ α مـن النقـاط الرافعة من خلال جمع القيم الأصلية بالقيمة 20.
2. تلويث المُتغيّر المعتمد بـ α من القيم الشاذّة تتّبع توزيع مربع كاي بدرجة حرية 50.
3. تلويث بيانات المُتغيّر الاستجابة والمُتغيّرات التوضيحية في آنٍ واحدٍ، ووفقاً للخطوتين المذكورين آنفًا. .

تكـرار توليـد البيانـات (100) مـرة و تخصـيص 50 دورة لعيّنـات البوتسـتراب مـع مراعـاة أحجـام العيّنـات المفترضة.

لاختبـار كفـاءة الطريقـة المقترحـة (WBP-RLARS) سـيتُمّ تشـغيل (100) دورة محاكـاة لغـرض مقارنتهـا مـع نتـائج الطريقـة B-RLARS ، ولأنّ الطـريقتين تقومـان بوضـع تسلسـل لأهمّيـة المُتغيّـرات التوضـيحية بالنسـبة

لمُتغيّـر الاسـتجابة؛ لـذا تـمّ اختيـار أول (5) مُتغيّـرات توضـيحية مـن كـلّ طريقـة علـى التـوالي، وذلـك بحسـاب معـدلّ عـدد المُتغيّـرات الصـحيحة (Right Selection)، التـي تـمّ اختيارهـا و معـدلّ عـدد المُتغيّـرات التـي تـمّ اختيارهـا خطـأ (False Selection) ، هـذا الأخيـر قـد يتضـمّن الـ Over fit و الـ Under fit. إنّ السـبب الـرئيس لاختيـار أول المُتغيّـرات الخمسـة أو السـبعة يعـود إلـى إنّنـا وضـعنا آليـة لاحتسـاب عـدد مُتغيّـرات الـ (Over fit) و الـ (Under fit) مـن عـدد المُتغيّـرات الـ (28) وفقـا لنسـب التلـوّث $\alpha = (0.05, 0.10, 0.15)$ ، وقـد اعتمـدت الباحثـة مقيـاس الاختيـار الخـاطِئ False Selection والاختيـار الصـائب Right Selection لطريقـة الاختيـار كمقيـاس للمقارنـة فـي تحديـد الأفضـل منهـا فـي عمليـة اختيـار المُتغيّـرات المهمّـة، والمـؤثّرة علـى مُتغيّـر الاسـتجابة، فالطريقـة الأفضـل هـي تلـك الطريقـة التـي تُظهـر أقـلّ اختياراً خاطئا و أكبر اختياراً صائباً لتحديد المُتغيّرات المؤثّرة في الأنموذج.

2-1-3 نتائج المحاكاة :

لإجـراء المحاكـاة تـمّ تكـوين حـالتين: الأولـى عنـدما يكـون حجـم العيّنـة أكبـر مـن عـدد المُتغيّـرات، والثانيـة عنـدما يكـون حجـم العيّنـة أصـغر مـن حجـم العيّنـة، وهـذه الحالـة تمثّـل مشـكلة الأبعـاد العاليـة high dimensions مـع قيمـة ارتبـاط 0.95 ونسـبة تلـوّث 0.15 للقـيم الشـاذّة والنقـاط الرافعـة معـاً ولخمـس مُتغيّـرات وسـبع مُتغيّـرات عنـد الاختيار وتفصيلات الحالتين كالآتي :

الحالة الأولى : ($\mathbf{n\ (50, 70) > p = 28}$)

الجـدول(1) يعـرض نتـائج الاختيـار الصـحيح و Right Selection والاختيـار الخـاطئ False Selection لطريقتي الاختيار ولحجمي (50 ،70) بحسب نسب التلوّث لأول خمس مُتغيّرات فقط.

$n > p = 28\ , \rho = 0.5\ ,\ ,\ \alpha = 0.05$						
قيم شـاذة ونقـاط رافعـة فـي ان واحد **Outliers & leverage points**		النقـاط الرافعـة فـي المُتغيّـرات $\mathbf{X^S}$ **leverage points**		القيم الشاذّة في المُتغيّر **Y** **Outliers**		مكان تلوّث البيانات
اختيـار خـاطئ للمُتغيّرات False Selection	اختيـار صـحيح للمُتغيّرات Right Selection	اختيـار خـاطئ للمُتغيّرات False Selection	اختيـار صـحيح للمُتغيّرات Right Selection	اختيـار خـاطئ للمُتغيّرات False Selection	اختيـار صـحيح للمُتغيّرات Right Selection	n طريقـة الاختيـار للمُتغيّرات Selection method

B-RLARS	50	**4.2**	**0.8**	**2.1**	**2.9**	**2.1**	**2.9**
WB-RLARS		**5**	**0**	**5**	**0**	**5**	**0**
B-RLARS	70	**5**	**0**	**2.1**	**2.9**	**2.1**	**2.9**
WBP-RLARS		**4.1**	**0.9**	**4.1**	**0.9**	**4.1**	**0.9**
$n > p = 28$, $\rho = 0.5$, , $\alpha = 0.10$							
B-RLARS	50	**4.1**	**0.9**	**1.1**	**3.9**	**2.1**	**2.9**
WB-RLARS		**5**	**0**	**4.1**	**0.9**	**5**	**0**
B-RLARS	70	**4.1**	**0.9**	**1.1**	**3.9**	**2.1**	**2.9**
WBP-RLARS		**4.1**	**0.9**	**4.1**	**0.9**	**4.1**	**0.9**
$n > p = 28$, $\rho = 0.5$, , $\alpha = 0.15$							
B-RLARS	50	**3.1**	**1.9**	**1**	**4**	**2.1**	**2.9**
WB-RLARS		**5**	**0**	**5**	**0**	**5**	**0**
B-RLARS	70	**4.1**	**0.9**	**1**	**4**	**2.9**	**2.1**
WBP-RLARS		**4.1**	**0.9**	**4.1**	**0.9**	**4.1**	**0.9**

يلاحظ أنّ الجزء الأول من الجدول (1) عندما يكون $n > p = 28 \ , \rho = 0.5 \ , \alpha = 0.05$ تفوّق الطريقة المقترحة WBP-RLARS على طريقة B-RLARS في الاختيار الصحيح للمُتغيّرات الخمسة المهمّة في مختلف أمكنة التلوّث سواء في القيم الشاذّة أم النقاط الرافعة أوكليهما مجتمعة عند حجم عيّنة 50 ، وينطبق الحال نفسه عند حجم عيّنة 70 باستثناء تفوّق بسيط لطريقة B-RLARS على طريقة WB-RLARS في مكان التلوّث في القيم الشاذّة لـ Y . أمّا الجزء الثاني من الجدول (1) عندما يكون $n > p = 28 \ , \rho = 0.5 \ , \alpha = 0.10$ تفوق الطريقة المقترحة WBP-RLARS على طريقة B-RLARS في الاختيار الصحيح للمُتغيّرات الخمسة في مختلف أمكنة التلوّث سواء في القيم الشاذّة أم النقاط الرافعة أوكليهما مجتمعة عند حجم عيّنة 50 ، وينطبق الحال نفسه عند حجم عيّنة 70 علماً أنّ بارتفاع نسبة التلوّث بالقيم الشاذّة للمُتغيّر الاستجابة Y حصراً، وازياد حجم العيّنة تتساوى كفاءة الطريقتين مع بعض. وبعبارة أخرى كلّما زاد حجم العيّنة مع زيادة نسبة التلوّث في القيم الشاذّة تتقارب أو تساوى كفاءة الطريقتين في هذه الحالة فقط. وقد تمّ التأكّد من ذلك مع أحجام عيّنات أكبر لم تذكر هنا في هذا البحث. أمّا الجزء الثالث من الجدول (1) عندما يكون $n > p = 28 \ , \rho = 0.5 \ , \alpha = 0.15$ ينطبق ما تَمّ التوصّل إليه في تحليل نتائج الجزء الثاني من الجدول (1) .

الحالة الثانية : $(\mathbf{n(20,26) < p = 28})$

الجدول(2) يعرض نتائج الاختيار الصحيح و Right Selection والاختيار الخاطئ False Selection لطريقتي الاختيار ولحجمي عيّنة (20 ،26) بحسب نسبة تلوّث 0.15 لأول خمسة وسبعة مُتغيّرات .

$n < p = 28 \ , \rho = 0.95 \ , \ , \ \alpha = 0.15$					
قيم شاذة ونقاط رافعة في ان واحد **Outliers & leverage points**					
اختيار خمسة مُتغيّرات **K=7**		اختيار خمسة مُتغيّرات **K=5**		مكان تلوّث البيانات	
اختيار خاطئ للمُتغيّرات False Selection	اختيار صحيح للمُتغيّرات Right Selection	اختيار خاطئ للمُتغيّرات False Selection	اختيار صحيح للمُتغيّرات Right Selection	n	طريقة الاختيار للمُتغيّرات Selection method
5	**2**	**3**	**2**	20	B-RLARS
6.7	**0.3**	**4.8**	**0.2**		WBP-RLARS

B-RLARS	26	**3.9**	**1.1**	**3.9**	**3.1**
WBP-RLARS		**2.2**	**2.8**	**3.1**	**3.9**

يلاحظ من الجدول (2) سوء أداء الطريقتين B-RLARS و WBP-RLARS في اختيار أول خمسة مُتغيّرات مهمة K=5 من بين p=28 بعد ترتيب أهمّيتها عند حجم عيّنة 20 مع أفضلية للطريقة B-RLARS ، ويرتفع أداؤهما عند حجم عيّنة 26 مع بقاء الأفضلية للطريقة نفسها، وبزيادة عدد أول المُتغيّرات المراد اختيارها من 5 إلى 7 يتحَسّن أداء الطريقتين ويتقاربا مع بعض عند حجم العيّنة 26 ، ولكن أداءهما يكون سيئاً عند حجم العيّنة 20 مع بقاء الأفضلية لطريقة B-RLARS . وتجدر الإشارة أنّ التجربة كانت أكثر من حالة تلوّث وكان تحليل النتائج متوافق لما تقدّم.

3-1-3 نتائج التطبيق العملي :

تمّ في هذا المبحث تحليل بيانات عن القيمة السوقية كمُتغيّر استجابة Y لبعض المصارف الأهلية، وعددها 13 مصرف وهي: (الأهلي، آشور، بابل، بغداد، إيلاف، الائتمان، الاستثمار، العراقي، الخليج، المنصور، سومر، التجاري والمصرف الوطني)، وهي عيّنة من عدد المصارف الكُلّي البالغ 48 ، وقد جمعت البيانات من النشرات المالية لسوق العراق للأوراق المالية المنشورة على الموقع الرسمي لسوق الأوراق المالية في العراق ،إذْ تمّ الحصول على 28 نسبة مالية وهي: (معدلّ دوران السهم (X_1) نسبة العائد على السهم (X_2) ، نسبة الملكية(X_3) ، مكرّر الأرباح(X_4)، نسبة التداول(X_5)، القيمة الدفترية(X_6)، سعر الإغلاق السنوي(X_7)، معدلّ السعر السنوي(X_8)، أعلى سعر نقد(X_9)، أدني سعر نقد(X_{10})، نسبة كفاية راس المال(X_{11})، نسبة السيولة(X_{12})، نسبة التشغيل(X_{13})، نسبة صافي الأرباح إلى راس المال(X_{14}) ، نسبة صافي الدخل إلى حقوق المساهمين(X_{15})، معدلّ دوران إجمالي الأصول(X_{16}) ، معدلّ دوران الأصول المتداولة(X_{17})، نسبة الإيرادات إلى إجمالي الودائع(X_{18})، نسبة المصروفات إلى الإيرادات(X_{19})،نسبة

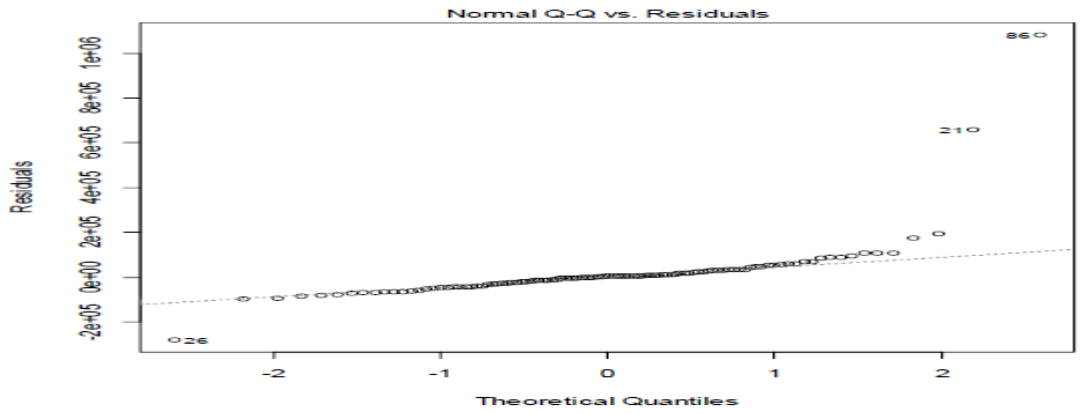
المصروفات إلى إجمالي الودائع(X_{20})، نسبة الودائع إلى (رأس المال + الاحتياطي) (X_{21})، نسبة (رأس المال + الاحتياطي) إلى الموجودات(X_{22})، مؤشّر السيولة (X_{23}) = (صافي رأس المال العامل / الموجودات) *100، مؤشّر السيولة(X_{24}) = (الأصول المتداولة / إجمالي الموجودات) *100، مؤشّر الرفع (X_{25}) = (حقوق المساهمين / الموجودات) *100، مؤشّر الربحية(X_{26}) = (صافي الربح قبل الضريبة/ إجمالي الموجودات) *100، مؤشّر الرفع (X_{27}) = (إجمالي الموجودات/ إجمالي المطلوبات) *100، مؤشّر الرفع (X_{28}) = (حقوق الملكية/ الموجودات الثابتة) *100)، ووفقًا لذلك يكون أنموذج الانحدار لعيّنة الدراسة كالآتي :

$$Y = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{28} X_{28} + \varepsilon$$

إذ إنّ : (ε) يمثّل حدّ الخطأ العشوائي الذي يتوزّع توزيعا طبيعياً.

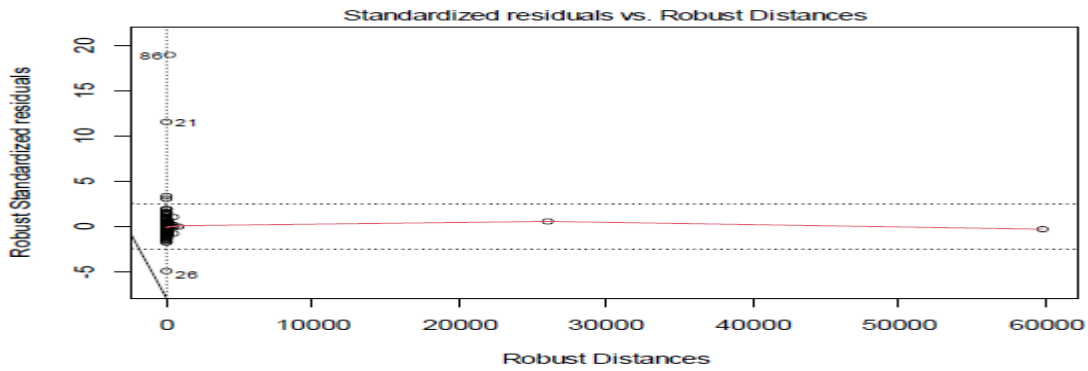
وبرسم البواقي يتّضح من الشكل (1) أنّ التوزيع الطبيعي لهذه البواقي تمّ خرقه ببعض القيم الشاذّة، أبرزها القيم المفهرسة (86، 21،26)، الأمر الذي يدفعنا إلى استعمال الطرائق الحصينة بدلًا من الطرائق التقليدية كما موضّحَ في الشكل الآتي:

الشكل (1) التمثيل البياني البواقي



واعتمادا على المحاكاة تبيّن أنّ تأثير النقاط الرافعة كان كبيرا على طريقة B-RLARS ، ولذلك لابُدّ من فحص وجود هذه النقاط في البيانات. والشكل (2) يُظهرُ العلاقة بين البواقي القياسية و مسافة مهالانوبيز الحصينة، وقد تبيّن أنّ هناك قيم شاذة outliers والظاهرة خارج الخطّين الأفقيين اللذان يمثّلان القيمتين (2.5، 2.5-)، كذلك وجود نقطتين رافعتين ظهرتا بعيدة جداً عن نقطة القطع، التي تمثّل الخطّ العمودي إلّا أنّهما ليستا الوحيدتين إذْ هناك ثَمّةَ نقاط رافعة أخرى تجاوزت خطَّ القطع العمودي هذا. وقد تَمّ استعمال Hat.matrix للكشف عن النقاط الرافعة والتي تبيّن أنّ نسبتها (15%) تقريبا وكما موضّحَ بالشكل الآتي :

الشكل (2) يبيّنُ وجود القيم الشاذّة و النقاط الرافعة في بيانات القيمة السوقية لبعض المصارف في العراق



ووفقا لحجم العيّنة الكُلّي البالغ 104 وبوجود 28 مُتغيّراً توضيحياً، وأيضاً نتائج المحاكاة تمّ استعمال طريقة WBP-RLARS؛ لأنّها الطريقة الأمثل في عملية اختيار المُتغيّرات عندما يكون حجم العيّنة أكبر من عدد المُتغيّرات التوضيحية وقد تمّ إجراء الآتي :

1- تحويل المُتغيّرات إلى مُتغيّرات قياسية حصينة

$$X_i = \frac{x_i - med(x_i)}{mad(x_i)} \qquad Y_i = \frac{y_i - med(y_i)}{mad(y_i)} \quad , \quad i = 1,2, \dots, n.$$

2- اختيار أول K من المُتغيّرات في تسلسل الأهمّية الناتج من الطريقة WBP-RLARS إذْ إنّ K=5و K=7 .

3- استخلاص مصفوفة فرعية $X_{(n\times k)}$ من مصفوفة المُتغيّرات التوضيحية $X_{(n\times p)}$، إذ إنّ k<p .

4- تقدير المَعْلَمات الحصينة باستعمال طريقة MM.

5- إيجاد قيمة p-value لِكُلّ معلمة مقدّرة بهدف اختبار الفرضيات الآتية :

$$H_0: \beta_j = 0 \qquad v.s \quad H_1: \beta_j \neq 0 \qquad , j = 1,2, \dots K$$

وقد كانت النتائج كالآتي :

الجدول (3) يوضّحَ اختيار الطريقة WBP-RLARS للمُتغيّرات التوضيحية المؤثّرة في القيمة السوقية لبعض المصارف الأهلية في العراق عندما K=5 و K=7

n=104 , p=28 , K=5				
المُتغيّر	$\hat{\beta}$	$Se(\hat{\beta})$	t-value	p-value
الحدّ الثابت	0.21	0.07	1.51	0.76
X_9	0.310	0.04	8.04	2.0E-12***
X_{28}	0.233	0.13	1.18	0.073
X_{24}	-0.231	0.08	-2.77	0.007**
X_7	0.004	0.03	0.20	0.85
X_{21}	-0.145	0.09	-1.60	0.11
n=104 , p=28 , K=7				
الحدّ الثابت	0.29	0.09	3.31	0.001**
X_9	0.310	0.04	8.17	2.0E-12***

X_{28}	0.249	0.08	3.24	0.002 **
X_{24}	-0.184	0.05	-3.52	0.0007***
X_7	-0.135	0.07	-1.83	0.0700
X_{21}	-0.231	0.09	-2.66	0.009 **
X_{27}	0.037	0.04	0.78	0.435
X_{19}	-0.014	0.01	-1.54	0.128

ويلاحظ من الجدول (3) أنّ الطريقة اختارت أول خمس مُتغيّرات هي $\{X_9, X_{28}, X_{24}, X_7, X_{21}\}$ من حيث الأهمّية عندما K=5 وتمثّل أعلى سعر نقد(X_9)، مؤشّر الرافعة المالية (X_{28})، مؤشّر السيولة (X_{24})، سعر الإغلاق السنوي(X_7)، نسبة الودائع إلى (رأس المال + الاحتياطي) (X_{21}) و فقا للنموذج النظري الآتي:

$$Y = \beta_0 + \beta_9 X_9 + \beta_{28} X_{28} + \beta_{24} X_{24} + \beta_7 X_7 + \beta_{21} X_{21} + \varepsilon$$

وقد تمّ تقدير المَعْلَمات الحصينة للنموذج المذكور آنفًا باستعمال طريقة MM وكما موضّحَ في الجدول ويتّضح من اختبار t وقيمة p-value رفض الفرضية الصفرية فيما يخصّ المَعْلَمات β_9و β_{24}

وعند توسيع عدد المُتغيّرات المراد اختيارها، أيْ جعل K=7 يتُمّ إضافة مُتغيّرين آخرين بحسب أهمّيتهما، ومن ثمّ تكون المُتغيّرات المختارة هي :$\{X_9, X_{28}, X_{24}, X_7, X_{21}, X_{27}, X_{19}\}$ ، أي إضافة نسبة المصروفات إلى الإيرادات(X_{19})، مؤشّر الرافعة المالية (X_{27}) الذي يساوي (إجمالي الموجودات/ إجمالي المطلوبات) *100. ليكون النموذج النظري الثاني كالآتي:

$$Y = \beta_0 + \beta_9 X_9 + \beta_{28} X_{28} + \beta_{24} X_{24} + \beta_7 X_7 + \beta_{21} X_{21} + \beta_{27} X_{27} + \beta_{19} X_{19} + \varepsilon$$

وقد أظهرتْ نتائج الاختبار وقيم p-value رفض الفرضية الصفرية للمَعْلَمات β_9و β_{24} و β_{28}و β_{21} ويلاحظ أنّه على الرغم من دخول المُتغيّرين X_{19} و X_{27} إلى الأنموذج لكنّهما لم يؤثّرا على مُتغيّر الاستجابة للقيمة السوقية للمصارف في عيّنة البحث، ولكن دخولهما في الأنموذج ساعد في الكشف عن مُتغيّرات مهمّة أخرى هي X_{28}و X_{21} لتكون المعادلّة التقدير النهائية التي توضّحَ العلاقة بين مُتغيّر الاستجابة القيمة السوقية وبعض المُتغيّرات التوضيحية المهمّة وفق الآتي :

$$\hat{Y} = 0.29 + 0.31 X_9 + 0.249\, X_{28} - 0.184\, X_{24} - 0.231 X_{21}$$

ولتأكيد نتائج تجربة المحاكاة عندما $n < p = 28$ تمّ اختزال المدة الزمنية للمصارف - عيّنة البحث - إلى سنتين فقط، بحيث يصبح حجم العيّنة 26 مشاهدة، وبتطبيق الطريقتين لاختيار المُتغيّرات عند اختيار كان اختيار طريقة WBP-RLARS للمُتغيّرات $\{X_{20}, X_4, X_{18}, X_{11}, X_{17}, X_6, X_{22}\}$ ، وهي مُتغيّرات

تختلف تماما عن التي اختارتها عندما كان عدد المشاهدات أكبر من عدد المُتغيّرات في التجارب السابقة. أمّا طريقة B-RLARS فقد اختارت $\{X_{13}, X_1, X_{26}, X_{21}, X_6, X_{11}, X_7\}$ وقد اشتركت مع اختيارات WBP-RLARS في المُتغيّرين (X_{21}، X_7) عندما $n > p = 28$ ، وهذا الاختيار يؤكّد نتائج المحاكاة وأفضلية طريقة B-RLARS على طريقة WBP-RLARS عندما تكون هناك مشكلة الأبعاد العالية high dimensions، أيْ $n > p$ وجود القيم الشاذّة outliers.

4- الاستنتاجات والتوصيات :

4-1 الاستنتاجات :

بالاعتماد على نتائج تجربة المحاكاة والتطبيق العملي يمكن استنتاج الآتي:

1- نستنتج أنّ لحجم العيّنة وعدد المُتغيّرات التوضيحية تأثيره على بيان أفضلية طريقة اختيار المُتغيّرات التوضيحية المهمّة والمؤثّرة فعلا على مُتغيّر الاستجابة، فضلاً عن عدد المُتغيّرات المراد اختيارها بحسب أهمّيتها إذ أبرزت نتائج الطريقتين عند K=7 ضماناً لاختيار المُتغيّرات المهمّة أكثر K=5 سواء عندما يكون حجم العيّنة أكبر أم أصغر من عدد المُتغيّرات التوضيحية في الأنموذج فنجد أنّ طريقة WBP-RLARS أثبتتْ كفاءتها على طريقة B-RLARS باختيار المُتغيّرات المهمّة في العيّنات من حجم 50.

2- تتقارب WBP-RLARS مع B-RLARS في كفاءتهما مع حجم العيّنات 70 وتبقى الأفضلية للطريقة WBP-RLARS مع ازدياد أحجام العيّنات، وهذا الحال هو نفسه عند K=7 ، ولكن هذا التقارب يظهر في وجود القيم الشاذّة فقط، أيْ في مُتغيّر الاستجابة حصراً ولكن الأفضلية تكون لطريقة WBP-RLARS بوجود النقاط الرافعة فقط أو وجود القيم الشاذّة والنقاط الرافعة معاً.

3- أفضلية طريقة B-RLARS عندما $n < p$ عند حالة اختيار K=5 وبحجم عيّنة 20 على طريقة WBP-RLARS التي يكون أداؤها في الاختيار سيئاً جداً في أحجام العيّنات الصغيرة ولكنّ أداءها يتُحَسّن كثيراً بتزايد حجم العيّنة كما حجم عيّنة 26 ، ويقترب كثيراً من أداء طريقة B-RLARS عند K=7.

4- اعتماداً على نتائج التطبيق العملي نستنتج أفضلية WBP-RLARS كانت مع حجم العيّنة n=104 الأكبر من عدد المُتغيّرات التوضيحية p=28 وعدد مُتغيّرات مختارة بحسب الأهمّية K=7 ولكن هذه الأفضلية تظهر لطريقة B-RLARS عند حجم عيّنةn=26 وعدد مُتغيّرات p=28 على حساب طريقة WBP-RLARS التي تعطي اختيارات مختلفة مع اختيارات الطريقة الأخرى التي تشترك معها ببعض الاختيارات، وهذا ما يتوافق مع نتائج المحاكاة وأنْ المُتغيّر X_9 والذي يمثّل أعلى سعر نقد للمصارف، هو المؤثر الأكثر تأثيراً ويأتي بالمرتبة الثانية المُتغيّر X_{28} ، الذي يمثّل مؤشّر الرفع للمصارف (حقوق المِلكِّية / الموجودات الثابتة) *100 وباتّجاه طردي. في حين جاء بالمرتبة الثالثة المُتغيّر X_{24} مؤشّر السيولة: (الأصول المتداولة /إجمالي الموجودات)*100 وبالمرتبة الرابعة X_{21} نسبة (رأس المال + الموجودات) إلى الموجودات وباتجاه عكسي.

2-4 التوصيات :

اعتماداً على الاستنتاجات توصي الباحثة بالآتي:

1- اعتماد عدد المُتغيّرات المهمّة من 7-10 حصراً في الأحوال كلّها؛ كونه يعطي ضماناً لاختيار المُتغيّرات المهمّة فعلاً.

2- استعمال خوارزمية WBP-RLARS في عملية الاختيار عندما $n > p$، والقيم الشاذّة في مُتغيّر الاستجابة بوجود النقاط الرافعة في المُتغيّرات التوضيحية أو القيم الشاذّة والنقاط الرافعة معاً في بيانات أنموذج الانحدار الخطّي المتعدّد، وبأي حجم عيّنة بما يحقّق الشرط المذكور آنفًا، وكما ويمكن استعمال طريقة B-RLARS في حالة وجود القيم الشاذّة في مُتغيّر الاستجابة فقط.

3- استعمال خوارزمية B-RLARS في حالة $n < p$ عند عملية اختيار المُتغيّرات المهمّة ولا سيّما عند أحجام العيّنات المبتعدّة من عدد المُتغيّرات التوضيحية حصراً ، ويمكن استعمال طريقة WBP-RLARS ايضاً لاختيار المُتغيّرات المهمّة في الأنموذج عندما يقترب حجم العيّنة من عدد المُتغيّرات التوضيحية في الأنموذج، إذ تقترب كفاءة الطريقتين من بعضهما عند هذه الحالة.

4- اهتمام المصارف الأهلية في العراق على المحافظة على أعلى سعر نقد، ومؤشّر الرفع لحقوق الملكية مقسوماً على الموجودات مرتفعاً؛ لأنّ له التأثير الكبير في رفع القيمة السوقية للمصارف في حين يجب عليها أنْ تحافظ على مستوى مناسب لمؤشّر السيولة فيها، وكذلك المُتغيّر X_{21} الذي يمثّل نسبة رأس المال + الاحتياطي) إلى الموجودات، كونهما يعملان على خفض القيمة السوقية للمصارف.

المصادر:-

أولا :- المصادر العربية

1- السراي ،علي حميد يوسف (2017) " استعمال المقدّرات الحصينة الجزائية لنموذج الانحدار الخطّي في ظلّ وجود مشكلتي الأبعاد والقيم الشاذّة مع تطبيق عملي " ، اطروحة دكتوراه ، كلّية الإدارة والاقتصاد – جامعة بغداد.

ثانياً :- المصار الاجنبية

1- Alamgir, Salahuddin, Ali Amjad (2013) " Split Sample Bootstrap Method" World Applied Sciences Journal 21 (7): 983-993.

2- Alfons, A., Ates ,N. Y. ,Groenen P J. F.(2022) " A Robust Bootstrap Test for Mediation Analysis" Organizational Research Methods, journals. sasepub. Com/home/otm, Vol. 25(3) 591–617.

3- Alfons, A., Baaske, W.E., Filzmoser, P., Mader W., Wieser, R. (2011)" Robust variable selection with application to quality of life research" 20(1), 65–82.

4- Barrera, M. S., Aelst S. V. (2008)" Robust model selection using fast and robust bootstrap" Computational Statistics and Data Analysis 52 , 5121–5135.

5- Blaine, Bruce E. (2018). "Winsorizing." The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation, 1817-1818. http://libguides.sjfc.edu/citations.

6- Buhlmann , Peter. Geer, Sara (2011) " Statistics for High Dimensional Data Methods, Theory and Applications". Springer.

7- Efron, B., Hastie, T., Johnstone I., Tibshirani, R. (2004)" LEAST ANGLE REGRESSION" The Annals of Statistics, vol. 32, No.2, 407-541.

8- Farnè, Matteo, Vouldis Angelos T. (2018)" A methodology for automatised outlier detection in high-dimensional datasets: an application to euro area banks' supervisory data" Working Paper Series, European Central Bank, No 2171.

9- Filzmoser, Peter, Nordhausen, Klaus, (2021) " Robust linear regression for high-dimensional data: An overview " WIREs Comput Stat.;13:e1524.

10- Hettigoda, Sandamala, (2016) "Computation of Least Angle Regression coefficient profiles and LASSO estimates.". Electronic Theses and Dissertations. Paper 2404. https://doi.org/10.18297/etd/240.

11- Imon, A. H. M. R, Ali M. M.(2005)" Bootstrapping Regression Residuals" Journal of Korean Data & Information Science Society, Vol. 16, No. 3, pp. 665-682.

12- Januaviani, T.M.A.,GusrianiN., Joebaedi, K., Supian, S., Subiyanto,(2019) " The Best Model of LASSO With The LARS (Least Angle Regression and Shrinkage) Algorithm Using Mallow's Cp" World Scientific News 116 , 245-252 .

13- Khan J. A., Aelst, S. V., Zamar, R. H.(2007) " Robust Linear Model Selection Based on Least Angle Regression " Journal of the American Statistical Association ,Vol. 102, No. 480, 1289-1299.

14- Lindskou, M., Tvedebrink, T., Eriksen, P.S., Morling, N. (2021)" Detecting Outliers in High-Dimensional Data with Mixed Variable Types Using Conditional Gaussian Regression Models"arXiv:2103.02366v3 [math.ST].

15- Mami, A. M., Jaber, A. M., Almabrouk, O. S. (2020)" Applying Bootstrap Robust Regression Method on Data with Outliers" International Journal of Sciences: Basic and Applied Research (IJSBAR) Volume 49, No 1, pp 143-160.

16- Midi. H., Uraibi, H.S., Talib, B.A. (2009)" Dynamic Robust Bootstrap Method Based on LTS Estimators " European Journal of Scientific Research, Vol.32 No.3, pp.277-287.

17- Mohamad, M, Ramli, N. M., Ghani, N. A. M.(2016) " Weighted Split Sample Bootstrap for Regression Models with High Dimensional Data" Indian Journal of Science and Technology, Vol 9(28) .

18- Ramli, Norazan M., Midi Habshah, Imon, A. H. M. R. (2009)" Estimating Regression Coefficients using Weighted Bootstrap with Probability" Wseas Transactions on Mathematics, Issue 7, Volume 8, 362-371.

19- Rana Sohel, Midi, Habshah, Imon, A. H. M. R. (2012) " Robust Wild Bootstrap for Stabilizing the Variance of Parameter Estimates in Heteroscedastic Regression Models in the Presence of Outliers" Mathematical Problems in Engineering Volume 2012, Article ID 730328, 14 pages.

20- Susanti, Yuliana, Pratiwi , Hasih, Sulistijowati, Sri, Liana ,Twenty (2014) " M ESTIMATION, S ESTIMATION, AND MM ESTIMATION IN ROBUST REGRESSION" International Journal of Pure and Applied Mathematics , Volume 91 No. 3, 349-360.