



مقارنة بعض التقديرات الحصينة لمعاملات نموذج انحدار العتبة ذي الأبعاد العالية<sup>1</sup>

## Comparison of some robust estimates

### of high-dimensional threshold regression model parameters

إ.د. باسم شليبه مسلم القيسي الدلفي<sup>(2)</sup>

غادة أيمن محمد حسن عباس<sup>(1)</sup>

[basim.sh@uokerbala.edu.iq](mailto:basim.sh@uokerbala.edu.iq)

[ghadah.a@s.uokerbala.edu.iq](mailto:ghadah.a@s.uokerbala.edu.iq)

قسم الاحصاء - كلية الادارة والاقتصاد - جامعة كربلاء

المستخلص

في هذا البحث، قُدرت معاملات نموذج انحدار العتبة Threshold Regression بوجود مشكلتي القيم الشاذة والأبعاد العالية بطرائق التقدير الحصينة الجزائية عن طريق دمج، أو إضافة دالة الخسارة الحصينة Loss Robust Function (خسارة Huber) مع دالة الجزاء  $L_1$  دالة جزاء (لاسو) Lasso، فضلاً عن اقتراح استعمال دالة الخسارة الحصينة Pseudo Tukey وبذلك يُضمن إيجاد التقدير للمعاملات واختيار المتغيرات المؤثرة في المتغير المعتمد لأنموذج انحدار العتبة في وقتٍ واحد. وقد أجريت محاكاة لغرض مقارنة مقدرات المعلمات على وفق دوال الخسارة الحصينة لحالتين افتراضيتين الأولى منها تساوي حجم العينة مع عدد المتغيرات التوضيحية في الانموذج والثانية حالة ان يكون حجم العينة اصغر من عدد المتغيرات التوضيحية وباقتراض نسبتيين من التلوث في البيانات 10% و 15% فضلاً عن افتراض توزيع الاخطاء بتوزيعات هي الطبيعي وتوزيع t بدرجة حرية 3 كتوزيع طويل الذيل وتوزيع طبيعي ملوث، وقد تم مقارنة اداء الطريقتين على وفق وجود التلوث من عدمه باستعمال معايير مقارنة واهمها معيار اداء حساسية الاختيار Sensitivity of Selection ومعيار اداء خصوصية الاختيار Specificity of Selection، وقد اظهرت نتائج المحاكاة تفوق طريقة التقدير Pseudo Tukey- Lasso (PT-L) على الطريقة الأخرى Huber-Lasso H-L في التقدير والاختيار بوجود التلوث وبكلتا النسبتيين من التلوث وايضاً بعدمه؛ فضلاً عن نوع توزيع الاخطاء سواء كان طبيعياً ام توزيعاً ذي ذبيلاً طويلاً أو طبيعياً ملوثاً. وقد اجري تطبيقاً عملياً على بيانات حقيقية لظاهرة اقتصادية، وقد طُبقت الطريقتين PT-L و H-L في

<sup>1</sup> بحث مستل من رسالة ماجستير في علوم الاحصاء

الجانب التطبيقي، واطهرت نتائجه تفوق PT-L في التقدير واختيار المتغيرات المؤثرة معدل الناتج المحلي في العراق وكان متغيري النفط والقطاع الزراعي في صدارتها في التأثير.

الكلمات المفتاحية : الانحدار غير المستمر ،انحدار العتبة ، التقدير الحصين الجزائي، دالة الخسارة ابيديو توكي .

## Abstract

In this research, the parameters of the Threshold Regression model were estimated in the presence of the problems of outliers and high dimensions by the methods of robust penalty estimation by combining or adding types of robust Loss Function Robust (absolute loss, Huber loss, Tukey loss) with the L1 penalty function, the Lasso penalty function, as well as proposing the use of the Pseudo Tukey robust loss function, and thus ensuring finding the estimate of the parameters and selecting the variables affecting the dependent variable of the Threshold Regression model at the same time. A simulation was conducted to compare the parameter estimates according to robust loss functions for several hypothetical cases, including the sample size being equal to the number of explanatory variables in the model, and also cases where the sample size is smaller than the number of explanatory variables. All cases were taken with two contamination rates in the data, 10% and 15%, as well as several distributions, namely the normal distribution and the t-distribution with 3 degrees of freedom as a long-tailed distribution and a contaminated normal distribution. The parameters were estimated and the performance of the methods was compared according to the presence or absence of contamination, according to comparative criteria, the most important of which are the Sensitivity of Selection performance criterion and the Specificity of Selection performance criterion. The simulation results showed the superiority of the Pseudo Tukey-Lasso (PT-L) estimation method over the other methods in estimation and selection with or without contamination and at both contamination rates, as well as the type of error distribution, whether it was a normal distribution, a long-tailed distribution, or a contaminated normal distribution. A practical application was carried out on real data of an economic phenomenon, and the first three methods in order were applied (Pseudo Tukey-Lasso, Tukey-Lasso and Huber-Lasso), and its results showed the superiority of Pseudo Tukey-Lasso in estimating and selecting the variables affecting the rate of GDP in Iraq, and the oil and agricultural sector variables were at the forefront of its influence.

Key Word: Discontinuous Regression, Threshold Regression, penalized robust estimation, Pseudo Tukey loss.

## Introduction

-1 المقدمة

إنَّ التحليل الاحصائي لبيانات اي ظاهرة حقيقية بمختلف المجالات يستوجب جمع البيانات data عن متغيراتها فضلاً عن تحديد الانموذج الرياضي المناسب لتمثيل تلك العلاقة بين متغيراتها ولكن يرافق مهمة جمعها بعض التحديات على وفق حالتين الأولى منها: يظهر التحدي في وجود خلل في البيانات نفسها وفقدان نقاوتها عن طريق تتضمنها قيماً أو مشاهدات تُعد شاذة outliers ، كما في بيانات المتغير المعتمد، أو قد توجد بصورة قيماً رافعة Leverage points في بيانات كل أو بعض المتغيرات التوضيحية قد توجد في كليهما وجميعها يندرج تحت بند ( القيم الشاذة ) فضلاً عن وجود توزيعات للأخطاء ذات ذيول طويلة أو ملوثة . اما في الحالة الثانية فهي بروز تحدي تحديد طبيعة العلاقة أوكلها التي يمثلها انموذج الانحدار من حيث انها علاقة خطية linearity بمسمى انموذج الانحدار الخطي Linear Regression مكتسباً صفة الاستمرارية يندرج في ضوئها ضمن نماذج الانحدار المستمر Continuous Regression أو تكون علاقة لا خطية non linearity بمسمى انموذج الانحدار اللاخطي Non-Linear Regression قد يفقد معها هذا الانموذج صفة الاستمرارية لامتلاك متغيراته التوضيحية نقاط عتبة هذه الصفة ويندرج في ضوئها ضمن نماذج الانحدار غير المستمر Discontinuous Regression كما هو الحال في انموذج انحدار العتبة Threshold Regression ، ولكل نوع اسلوبه الخاص في التحليل الاحصائي ومن ناحية أخرى وضمن الحالة الثانية ايضاً قد يبرز تحدي آخر بأن لا تتوفر مشاهدات كافية لمتغيرات الظاهرة تحت الدراسة عند جمعها لاي سبب كان بحيث يصبح عدد المتغيرات التوضيحية في الانموذج أكبر من حجم العينة  $p > n$  يسبب مشكلة الابعاد العالية high dimensional ، وفي كلا الحالتين لا يمكن استعمال طرائق التقدير التقليدية ومنها طريقة المربعات الصغرى الاعتيادية إذ تفقد كفاءتها بوجود هاتين المشكلتين، بسبب تأثرها الكبير بهما ولاسيما بالقيم الشاذة كونه يؤدي الى الحصول على مقدرات غير كفوءة وهذا يُعد مشكلة في التحليل الاحصائي لبيانات اي ظاهرة تستوجب البحث والتقصي.

يُعد الباحث Tong [17] أول من قدم مفهوم نقاط العتبة في عام 1978 في موضوع السلاسل الزمنية في حين يُعد الباحث Hansen [9] أول من قدم بحثاً في انموذج الانحدار العتبة Threshed Regression Model في 1999، وعده نوعاً مهماً من انواع الانحدار غير المستمر، ثم تلتها بحوثاً في هذا المجال هي Hansen [10] في 2000 والباحثان Caner & Hansen [4] في عام 2002 ، والباحث Gannaz [8] في عام 2007. اما موضوعات القيم الشاذة والابعاد العالية فقد نوقشت بشكل منفرد ابتدأت من الباحث [5] Caner ببحثه عن انموذج انحدار العتبة بوجود القيم الشاذة outliers وفي عام 2016 قدم الباحث Lee واخرون بحثاً يُعد أول محاولة بحثية تتناول انموذج انحدار العتبة (نقاط تغير) change point بوجود البيانات العالية الابعاد high dimensional . ولكن في عام 2020 قدم الباحث Li واخرون [13] بحثاً يُعد أول محاولة لتقدير معالم انموذج انحدار العتبة TRM بوجود القيم الشاذة والبيانات العالية في ان واحد ، وفي العام 2023 قدم الباحثان Abdelmonim & Msallam [3] بحثاً تضمن طريقة مقترحة لاختيار المتغيرات التوضيحية في انموذج الانحدار الخطي عند وجود مشكلتي القيم الشاذة والابعاد العالية. وفي عام 2024 قدم الباحث Liu واخرون [14] بحثاً تضمن اقتراحاً لانموذج

انحدار العتبة حصيناً بوجود بيانات ذات الابعاد العالية ، وفي العراق قدم الباحث يوسف [2] عام 2017 اطروحته في الدكتوراه مستعملاً المقدرات الحصينة الجزائية لتقدير معلمات الانحدار التقليدي بوجود مشكلتي القيم الشاذة والبيانات العالية الابعاد، وفي عام 2023 قدمت عبد المنعم [1] رسالتها في الماجستير التي استعملت طريقة مقترحة لاختيار المتغيرات التوضيحية في الانموذج بوجود مشكلة البيانات ذات الابعاد العالية. اما في هذا البحث سيتم تقدير معلمات انموذج انحدار العتبة TRM باستعمال طرائق تقدير حصينة جزائية عن طريق اقتراح استعمال دالتي خسارة ابيدو توكي Pseudo Tukey Loss ودالة هوبر Huber Loss ودمجهما مع دالة الجزاء (لاسو Lasso) ومن ثم مقارنتهما مع بعض لتحديد الافضل منها في تقدير معلمات TRM .

وفي ضوء ذلك سيتم تقسيم البحث لأربعة اجزاء يتضمن الأول منها مقدمة عامة تتضمن مشكلة وهدف البحث والثاني يتضمن الجانب النظري عرضاً لدوال الخسارة الحصينة فضلاً عن دالة جزاء لاسو ، وتضمن الجزء الثالث المحاكاة ، ورابعاً يتضمن جانباً تطبيقياً .

## 2- الجانب النظري

### Threshold Regression Model

### 1-2 انموذج انحدار العتبة

يكتب انموذج انحدار العتبة TR بالشكل الشائع كما يأتي [18] [10] :

$$Y_i = \begin{cases} X_i' \beta_1 + \epsilon_i & X_{ij} \leq \gamma \\ X_i' \beta_2 + \epsilon_i & X_{ij} > \gamma \end{cases}, i = 1, 2, \dots, n \quad (1)$$

إذ ان :

$Y_i$  : تمثل القيمة  $i$  من مشاهدات المتغير المعتمد وفقاً للمشاهدة  $i$  من المتغيرات التوضيحية.

$$X_i' = [1 \quad X_{i1} \quad \dots \quad X_{ij} \quad \dots \quad X_{ik}]_{1 \times p}, \quad p = k + 1$$

$X_j$  : يمثل متغير العتبة threshold variable في الانموذج

$\beta_2, \beta_1$  : يمثلان معلمات الانموذج قبل العتبة وبعدها على الترتيب من مرتبة  $1 \times p$  لكل منهما .

$\gamma$  : معلمة العتبة أو نقطة العتبة المجهولة .

$\epsilon_i' / \sigma$  : الاخطاء العشوائية التي تتوزع توزيعاً عشوائياً مستقلاً ومتماثلاً لكل منها ، وتجدر الاشارة إلى ان تبايناتها متساوية لكل منها ايضاً .

وعليه، يمكن اعادة الانموذج (1) بدلالة دالة المؤشر indicator function وكما يأتي [11]:

$$Y_i = X_i' \beta_1 + X_i' \delta 1\{X_{ij} < \gamma\} + \epsilon_i \quad (2)$$

$\delta = \beta_2 - \beta_1$  : إذ ان :

وكلاهما  $\delta, \beta_1$  معاملات الانموذج من مرتبة  $px1$  ، وان  $\{X_{ij} < \gamma\}$  تشير الى دالة المؤشر indicator function .

من اهم الفرضيات التي يقوم عليها انموذج انحدار العتبة ان يكون الخطأ العشوائي يتبع توزيعاً متماثلاً بمتوسط صفر وتباين ثابت لجميع الاخطاء العشوائية وغالبا ما يُتفترض التوزيع الطبيعي ليتم تطبيق طريقة المربعات الصغرى الاعتيادية على وفق البحث الشبكي، وهي تتصف بكفاءة التقدير مادام توزيع الاخطاء توزيعاً طبيعياً ولكنها تفقد هذه الصفة عندما لا يكون التوزيع طبيعياً ولا سيما بالتوزيعات التي تمتلك ذيولا طويلة heavy tails . ويمكن عرض خوارزمية لتطبيقها وكما يأتي [9]:

- 1- تحديد شبكة grid من القيم لمعلمة العتبة ضمن مجموعة G: ويتم باختيار مجموعة من القيم التي يمكن ان تكون نقاط عتبة في المتغير التوضيحي  $X_j$  اي ان  $d = 1, 2, \dots, m$  ،  $\gamma_{jd}$  .
- 2- تستثنى القيم المتطرفة (الحدود العليا والدنيا) من الشبكة المستحصل عليه في الخطوة 1 وبنسبة 10 % أو 15 % بعد ترتيبها تصاعدياً لكي لا يكون القطع بالأطراف.
- 3- تقسيم الانموذج الانحدار الى أنموذجي انحدار فرعيين وفقاً لنقطة العتبة المختارة من قيم الشبكة، أحدهما قبل نقطة العتبة والاخر بعدها، كما معرف بالصيغة (1)، ويتم تطبيق طريقة المربعات الصغرى الاعتيادية لكل انحدار فرعي لتقدير معاملات الانموذج كما يأتي :

$$\hat{\beta}_1 = \arg. \min_{\beta_1} \sum_{X_{ij} \leq \gamma_{jd}} (Y_i - X_i' \beta_1)^2 \quad (3)$$

$$\hat{\beta}_2 = \arg. \min_{\beta_2} \sum_{X_{ij} > \gamma_{jd}} (Y_i - X_i' \beta_2)^2 \quad (4)$$

- 4- حساب مجموع مربعات الخطأ لانموذج انحدار العتبة عند قيمة العتبة  $\gamma_{jd}$  على وفق الاتي :

$$SSE(\gamma_{jd}) = \sum_{X_{ij} \leq \gamma_{jd}} (Y_i - X_i' \hat{\beta}_1)^2 + \sum_{X_{ij} > \gamma_{jd}} (Y_i - X_i' \hat{\beta}_2)^2 \quad (5)$$

- 5- تحديد قيمة العتبة المثلى  $\hat{\gamma}$  من خلال الاتي :

$$\hat{\gamma} = \arg \min_{\gamma_{jd} \in G} SSE(\gamma_{jd}) \quad (6)$$

والخطوات من 1-5 تمثل عملية البحث الشبكي grid search لتحديد قيمة العتبة المثلى.

- 6- تطبيق بطريقة المربعات الصغرى الاعتيادية OLS اعتماداً على القيمة المثلى للعتبة  $\hat{\gamma}$  للحصول على المقدرات النهائية  $(\hat{\beta}_1, \hat{\beta}_2)$  لمعاملات انموذج انحدار العتبة TR.

تعد هذه الطريقة الاكثر شيوعاً لتقدير لمعاملات انموذج انحدار العتبة TR [9][10] ، وهي طريقة كفوءة ولكن ما يعيها ان تتأثر بأي مشكلة ومنها وجود مشكلة الابعاد العالية high dimensional وايضاً وجود القيم الشاذة outliers ولذا لا بد من البحث عن طرائق تقدير بديلة تأخذ بنظر الاعتبار وجود تلك المشكلات.

## 2-2 التقدير بالطرائق الجزائية في نماذج الانحدار

### Estimation via Penalized Methods in Regression Models

ان التقدير بالطرائق الجزائية في نماذج الانحدار عامة، يكون كما يأتي [16] [15]:

$$\text{argmin}_{\beta} \underbrace{\text{Objective Function}}_{\text{Panelized Loss Function}} = \text{argmin}_{\beta} \left( \underbrace{L(Y, \hat{Y})}_{\text{Loss Function}} + \underbrace{\lambda P(\beta)}_{\text{Penlnty Function}} \right) \quad (7)$$

وتمثل دالة الهدف Objective Function (7) الصيغة العامة للطرائق الجزائية Panelized methods التي تُستعمل لتقدير معاملات نموذج الانحدار على وفق الانحدار المنتظم Regularized Regression والتي تتكون من مجموع دالتين رئيسيتين هما دالة الخسارة Loss Function ودالة الجزاء Penalty Function .  
3-2 التقديرات الحصينة الجزائية لمعاملات نموذج انحدار العتبة ذي الابعاد العالية

### Panelized -Robust Estimate Parameters of High Dimensional Threshold Regression

ترتكز الطرائق الحصينة Robust Methods في عملية التقدير على فكرة إعادة الصياغة لعملية التقدير عن طريق ابدال دالة الخسارة التربيعية Squared Loss/عادية أو اي دالة خسارة تقليدية بأخرى تعتمد الأوزان التي تُسهم في تقليل دور المشاهدات أو القيم الشاذة في دالة الهدف، إذ يتم الحصول على مقدرات معاملات نموذج انحدار العتبة باعتماد الطرائق الجزائية الحصينة Robust Penalized Methods التي يمكن كتابتها بشكل عام كما يأتي :

$$\text{argmin}_{\beta} \underbrace{\text{Objective Function}}_{\text{Robust Panelized Loss Function}} = \text{argmin}_{\beta} \left( \underbrace{L(Y, \hat{Y})}_{\text{Robust Loss Function}} + \underbrace{\lambda P(\beta)}_{\text{Penlnty Function}} \right) \quad (8)$$

وبتطبيق الصيغة العامة للطرائق الجزائية الحصينة على نموذج انحدار العتبة TR عن طريق تعويض دالة الجزاء لاسو Lasso في الصيغة (8) والتي تُعرف كما يأتي :

$$P(\beta) = \sum_{j=1}^k |\beta_j| \quad , (\beta = \beta_1, \dots, \beta_k), (j = 1, \dots, k) \quad (9)$$

إنَّ سبب تسميتها Lasso -  $L_1$  تميزاً؛ لأنها تستعمل كدالة جزاء مع طريقة التقصص والاختيار بالقيمة المطلقة الصغرى Least Absolute Shrinkage and Selection Operator واختصاراً Lasso ، ولها تسمية أخرى هي  $L_1$  - norm ، وعليه تصبح الدالة (8) كما يأتي:

$$\hat{\beta}_{R-Lasso} = \text{argmin}_{\beta=\beta_1, \delta} \left( \frac{1}{n} \sum_{i=1}^n L(Y_i - X_i' \beta_1 - X_i' \delta 1\{X_{ij} < \gamma\}) + \lambda \sum_{j=1}^{2k} |\beta_j| \right) \quad (10)$$

إذ ان:

$\lambda$  : يمثل معلمة الضبط Tuning Parameter .

ونلاحظ من الصيغة (10) أن مقدرات (لاسو) الحصينة R- Lasso لمعلمات نموذج HDTR تعتمد على دالة الخسارة الحصينة  $L(\cdot)$  المستعملة في الصيغة ، وبما ان هناك انواعاً من هذه الدالة ،لذا تنتوع مقدرات لاسو الحصينة ايضاً تبعاً لنوع الدالة المستعملة وتتصف هذه المقدرات بانها مقدرات تجمع بين الحصانة في التقدير واختيار المتغيرات التوضيحية المؤثرة في النموذج انحدار العتبة TR في أن واحد ، إذ تكون مقدرات حصينة مبعثرة Sparse Robust Estimators ، وللحصول على هذه المقدرات الحصينة سيتم تقديم عدداً من هذه الدوال ويجاد مقدرات (لاسو) الحصينة وفقاً لها وكما يأتي :

### Huber-Lasso مقدر 1-3-2

وهو من أهم المقدرات الحصينة الشائعة ايضاً واختصاراً Huber -Lasso ، ويعتمد هذا المقدر على في احتسابه على دالة خسارة هابر Huber Loss والتي تُعرف كما يأتي [13] :

$$L_{\text{Huber Loss}}(\cdot) = H_a(r_i) = \begin{cases} \frac{1}{2}r_i^2 & \text{if } |r_i| \leq a \\ a|r_i| - \frac{1}{2}a^2 & \text{if } |r_i| > a \end{cases} \quad (11)$$

$$r_i = Y_i - X_i'\beta_1 - X_i'\delta 1\{X_{ij} < \gamma\} \quad \text{إذ ان :}$$

وان :

a: قيمة أو معامل الضبط tuning parameter في دالة خسارة Huber التي على ضوءها يتم التعامل مع البواقي ، فعند تحقق الشرط الأول (البواقي صغيرة) يكون التعامل تربيعياً quadratic ومثل الانحدار التقليدي على وفق طريقة OLS ، وعند تحقق الشرط الثاني (البواقي كبيرة) يتم التعامل معها خطياً linear لتقليل أو تخفيض تأثير القيم الشاذة outliers ، واقترح الباحث Huber استعمال  $a=1.345$  ، لتكون مقدرات Huber قريبة من مقدرات OLS بكفاءة 95% من التقارب للتوزيع الطبيعي عندما لا توجد هناك قيماً شاذة.

وبتعويض صيغة Huber Loss في صيغة دالة الهدف الجزئية (8) يتم الحصول على مقدر Huber -Lasso عن طريق الاتي:

$$\hat{\beta}_{\text{Huber-Lasso}} = \operatorname{argmin}_{\beta} \left( \frac{1}{n} \sum_{i=1}^n H_a(r_i) + \lambda \sum_{j=1}^{2k} |\beta_j| \right) \quad (12)$$

ونلاحظ من الصيغة (12) ، ان مقدر Huber -Lasso يرتبط بألية التقدير للانحدار الحصين Robust Regression باستعمال طريقة M- Estimation .

وعليه، يمكن حلّ الصيغة (12) والحصول على مُقَدَّر Huber -Lasso باستعمال خوارزمية المربعات الصغرى الموزونة التكرارية Iterative Reweighted Least Squares وقبل توضيح خطواتها هناك اجراءات تسبقها هي افتراض أن تكون  $\lambda$  ثابتة fixed ، ويجاد مشتقة دالة  $H_a(r_i)$  والتي عن طريقها تعرف دالة الوزن weight function التي تستعمل في الخوارزمية ولكل مشاهدة من مشاهدات العينة  $i = 1, 2, \dots, n$  وكما يأتي:

$$\underset{\text{Huber Loss}}{L'}(\cdot) = H'_a(r_i) = \begin{cases} r_i & \text{if } |r_i| \leq a \\ a \operatorname{sign}(r_i) & \text{if } |r_i| > a \end{cases} \quad (13)$$

$$\text{weight}(r_i) = w(r_i) = w_i = \frac{H'_a(r_i)}{r_i} = \begin{cases} 1 & \text{if } |r_i| \leq a \\ \frac{a}{|r_i|} & \text{if } |r_i| > a \end{cases} \quad (14)$$

وعليه فان دالة الوزن weight function تعرف بشكل عام كما يأتي :

$$w(r) = \frac{H'_a(r)}{r} \quad (15)$$

واعتماداً على الصيغة (14) يمكن توضيح خطوات خوارزمية المربعات الصغرى الموزونة التكرارية

IRLS كما يأتي :

1- اختر مقدرات ابتدائية  $\hat{\beta}^{(0)}$  لموجه المعلمات  $\beta$  كأن تكون مقدرات طريقة OLS أو اي مقدرات أخرى

2- احسب  $r_i^{(t)}$  باستعمال الصيغة الاتية :

$$r_i^{(t)} = Y_i - X_i' \hat{\beta}_1^{(t)} - X_i' \hat{\delta}^{(t)} 1\{X_{ij} < \gamma\}, t = 0, 1, 2, \dots \quad (16)$$

إذ ان :

$$r_i^{(0)} = Y_i - X_i' \hat{\beta}_1^{(0)} - X_i' \hat{\delta}^{(0)} 1\{X_{ij} < \gamma\}$$

3- احسب  $w_i^{(t)}$  كما يأتي:

$$w_i^{(t)} = \begin{cases} 1 & \text{if } |r_i| \leq 1.345 \\ \frac{1.345}{|r_i^{(t)}|} & \text{if } |r_i| > 1.345 \end{cases}, t = 0, 1, 2, \dots \quad (17)$$

ويمكن افتراض ان :  $w_i^{(0)} = 1$  كما ورد في المصدر [ 13 ] .

4- احسب :

$$\underset{\text{Huber-Lasso}}{\hat{\beta}^{(t+1)}} = \operatorname{argmin}_{\beta} \left( \frac{1}{n} \sum_{i=1}^n w_i^{(t)} (Y_i - X_i' \hat{\beta}_1^{(t)} - X_i' \hat{\delta}^{(t)} 1\{X_{ij} < \gamma\}) + \lambda \sum_{j=1}^{2k} |\hat{\beta}_j^{(t)}| \right)$$

$$t = 0, 1, 2, \dots \quad (18)$$

5- تحقق من الشرط الاتي :

$$\left| \underset{\text{Huber-Lasso}}{\hat{\beta}^{(t+1)}} - \underset{\text{Huber-Lasso}}{\hat{\beta}^{(t)}} \right| < 10^{-6} \quad (19)$$

6- اعد تطبيق الخطوات من 2- 5 في حالة عدم تحقق الشرط .

من مساوي تطبيق دالة خسارة Huber المعرفة بالصيغتين (11) على البواقي residuals الاصلية الخام  $r_i$  انها قد تؤدي الى فقدان ثبات المقياس بمعنى انه اي تغير في وحدة المقياس سوف يغير  $|r_i|$  وبدوره يغير الأوزان والحصانة للمقياس وهذا بسبب اعتماد دالة الخسارة على فرضية التباين الموحد لكل مشاهدات متغير الاستجابة، وللتغلب على هذه المشكلة يتم تطبيقها على البواقي القياسية scaled residuals ، وتكتب دالة الخسارة Huber القياسية كما يأتي:

$$L_{\text{Huber Loss}}(\cdot) = H_a(u_i) = \begin{cases} \frac{1}{2}u_i^2 & \text{if } |u_i| \leq a \\ a|u_i| - \frac{1}{2}a^2 & \text{if } |u_i| > a \end{cases} \quad (20)$$

إذ ان :

$$u_i = \frac{r_i}{\sigma}$$

$$\hat{\sigma} = \frac{\text{MAD}}{0.6745} , \text{MAD} = \text{median}(r_i - \text{median}(r_i)) \quad (21)$$

وعليه باستبدال دالة خسارة Huber المعتمدة على البواقي الاصلية المعرفة بالصيغة (11) بدالة خسارة Huber القياسية المعرفة بالصيغة (20) في الصيغة (12) لتصبح كما يأتي:

$$\hat{\beta}_{\text{Huber-Lasso}} = \text{argmin}_{\beta} \left( \frac{1}{n} \sum_{i=1}^n H_a(u_i) + \lambda \sum_{j=1}^{2k} |\beta_j| \right) \quad (22)$$

إذ ان :

$$H_a(u_i) = H_a\left(\frac{Y_i - X_i' \beta_1 - X_i' \delta \mathbb{1}\{X_{ij} < \gamma\}}{\sigma}\right)$$

ويمكن اعادة تطبيق الخطوات نفسها على الصيغة (22) والتي تم شرحها انفاً على الصيغة (12) لنحصل على مقدر Huber-Lasso المعتمد على دالة خسارة Huber القياسية ، والمقدر الاخير المعرف بالصيغة (22) هو المعتمد في البحث.

### 2-3-2 مقدر Pseudo Tukey-Lasso

يتم الحصول على هذا المقدر عن طريق تطوير أو تحسين مقدر Tukey – Lasso الذي يتم الحصول عليه من تعويض دالة خسارة توكي Tukey Loss ويتم التحسين عن طريق تعويض دالة الخسارة Pseudo Tukey بدلاً من خسارة توكي Tukey Loss في الصيغة (8) ، وتكتب دالة الخسارة Pseudo Tukey أو Scaled Tukey Loss كما يأتي [6] :

$$\underset{\text{Scaled Tukey Loss}}{L}(\cdot) = H_e(u_i) = \begin{cases} \frac{e^2}{6} [1 - (1 - (\frac{u_i}{e})^2)]^3 & \text{if } |u_i| \leq e \\ \frac{e^2}{6} & \text{if } |u_i| > e \end{cases} \quad (23)$$

إذ ان :

e: قيمة أو معامل الضبط tuning parameter في دالة خسارة Tukey وتستعمل e=4.685 ، لتكون مقدرات Tukey قريبة من مقدرات OLS بكفاءة 95% من التقارب للتوزيع الطبيعي عندما لا توجد هناك قيماً شاذة.

$$u_i = \frac{r_i}{\hat{\sigma}}$$

$$\hat{\sigma} = \frac{\text{MAD}}{0.6745} , \text{MAD} = \text{median}(r_i - \text{median}(r_i)) \quad (24)$$

وبتعويض  $H_e(u_i)$  يمكن الحصول على مقدر Pseudo Tukey-Lasso كما يأتي :

$$\underset{\text{Pseudo Tukey-Lasso}}{\hat{\beta}} = \text{argmin}_{\beta} \left( \frac{1}{n} \sum_{i=1}^n H_e(u_i) + \lambda \sum_{j=1}^{2k} |\beta_j| \right) \quad (25)$$

إذ ان :

$$H_e(u_i) = H_e\left(\frac{r_i}{\hat{\sigma}}\right) = H_e\left(\frac{Y_i - X_i' \beta_1 - X_i' \delta \cdot 1\{X_{ij} < \gamma\}}{\hat{\sigma}}\right) \quad (26)$$

ولكي نحل الصيغة (25) والحصول على مقدر Pseudo Tukey-Lasso باستعمال خوارزمية المربعات الصغرى الموزونة التكرارية Iterative Reweighted Least Squares ، ونحتاج الى ايجاد مشتقة دالة  $H_e(u_i)$  والتي عن طريقها تعرف دالة الوزن weight function التي تستعمل في الخوارزمية ولكل مشاهدة من مشاهدات العينة  $i = 1, 2, \dots, n$  ولكي نجدها يتم اشتقاق دالة الخسارة (23) لنحصل على الاتي:

$$\underset{\text{Scaled Tukey Loss}}{L'}(\cdot) = H'_e(u_i) = \begin{cases} u_i [(1 - (\frac{u_i}{e})^2)]^2 & \text{if } |u_i| \leq e \\ 0 & \text{if } |u_i| > e \end{cases}$$

$$\text{weight}(u_i) = w(u_i) = w_i = \frac{H'_e(u_i)}{u_i} = \begin{cases} [(1 - (\frac{u_i}{e})^2)]^2 & \text{if } |u_i| \leq e \\ 0 & \text{if } |u_i| > e \end{cases} \quad (27)$$

وعليه فان دالة الوزن weight function تعرف بشكل عام كما يأتي :

$$w(u) = \frac{H'_e(u)}{u} \quad (28)$$

واعتماداً على الصيغة (27) يمكن توضيح خطوات خوارزمية المربعات الصغرى الموزونة التكرارية IRLS للحصول على مقدر Pseudo Tukey – Lasso وكما يأتي :

1- اختر مقدرات ابتدائية  $\hat{\beta}^{(0)}$  لموجه المعلمات  $\beta$  كأن تكون مقدرات طريقة OLS أو أي مقدرات أخرى .

2- احسب  $u_i^{(t)}$  باستعمال الصيغة الآتية :

$$u_i^{(t)} = \frac{Y_i - X_i' \hat{\beta}_1^{(t)} - X_i' \hat{\delta}^{(t)} 1\{X_{ij} < \gamma\}}{\hat{\sigma}}, t = 0, 1, 2, \dots \quad (29)$$

إذ ان :

$$u_i^{(0)} = \frac{Y_i - X_i' \hat{\beta}_1^{(0)} - X_i' \hat{\delta}^{(0)} 1\{X_{ij} < \gamma\}}{\hat{\sigma}}$$

3- احسب  $w_i^{(t)}$  كما يأتي:

$$w_i^{(t)} = \begin{cases} \left[1 - \left(\frac{u_i^{(t)}}{4.685}\right)^2\right]^2 & \text{if } |u_i| \leq 4.685 \\ 0 & \text{if } |u_i| > 4.685 \end{cases}, t = 0, 1, 2, \dots \quad (30)$$

ويمكن افتراض ان :  $w_i^{(0)} = 1$  كما ورد في المصدر [13] .

4- احسب :

$$\hat{\beta}^{(t+1)} = \underset{\text{Pseudo Tukey-Lasso}}{\text{argmin}}_{\beta} \left( \frac{1}{n} \sum_{i=1}^n w_i^{(t)} \left( \frac{Y_i - X_i' \hat{\beta}_1^{(t)} - X_i' \hat{\delta}^{(t)} 1\{X_{ij} < \gamma\}}{\sigma} + \lambda \sum_{j=1}^{2k} |\hat{\beta}_j^{(t)}| \right) \right)$$

$$t = 0, 1, 2, \dots \quad (31)$$

5- تحقق من الشرط الآتي :

$$\left| \hat{\beta}^{(t+1)} - \hat{\beta}^{(t)} \right|_{\text{Pseudo Tukey-Lasso}} < 10^{-6} \quad (32)$$

6- اعد تطبيق الخطوات من 2- 5 في حالة عدم تحقق الشرط .

### 3- معايير المقارنة واختبار الفرضيات للجانبين التجريبي والتطبيقي

لاجراء المقارنة بين طرائق التقدير الحصينة الجزائية المستعملة في البحث تم استعمال عدة معايير في الجانبين التجريبي والتطبيقي فمعايير المقارنة في الجانب التجريبي هي كما يأتي [13] :

1-3 : معيار الاداء  $L_2$  -Oracal ( خطأ التقدير الميول الحدية مع الحد الثابت)

يمثل خطأ التقدير للمعلمة عند الافتراض معرفة المتغيرات التوضيحية المؤثرة فعلاً المهمة فقط في الانموذج ويستعمل في تجارب المحاكاة فقط بوصفها أداء نظرياً للطريقة ويرمز له بـ True وصيغته :

$$\text{True}^{(r)} = \|\hat{\beta}_{\text{Oracal}} - \beta_{(0)\text{true}}\|_2^{(r)} = \sqrt{\sum_{j=0}^k (\hat{\beta}_{\text{Oracal}(j)}^{(r)} - \beta_{\text{true}(j)}^{(r)})^2} \quad (33)$$

إذ ان:

r: يمثل عدد تكرارات تجربة المحاكاة.

اما الانحراف المعياري للمعيار في تجربة المحاكاة فيكون كما يأتي :

$$\text{SD}(\text{True}) = \sqrt{\frac{1}{R-1} \sum_{r=1}^R (\text{True}^{(r)} - \bar{L}_2)^2}, \quad \bar{\text{True}} = \frac{1}{R} \sum_{r=1}^R \text{True}^{(r)} \quad (34)$$

### 2-3 : معيار الاداء $L_2$ ( خطأ التقدير الكلي لمعلمت الانموذج بما فيها معلمت العتبة)

ويعتمد في احتسابه على معلمت الانموذج ككل ويستعمل لقياس المسافة التربيعية بين المقدرات والقيم الحقيقية، بذلك فهو يعكس الدقة الاجمالية للمقدرات، ويرمز له  $L_2$  وصيغته كما يأتي :

$$L_2^{(r)} = \|(\hat{\beta}, \hat{\gamma}) - (\beta, \gamma)\|_2^2 = \sum_{j=0}^k (\hat{\beta}_{\text{Oracal}(j)}^{(r)} - \beta_{(0j)\text{true}})^2 + (\hat{\gamma}^{(r)} - \gamma)^2 \quad (35)$$

$$L_2^{(r)} = \text{MSE}^{(r)} \quad \text{إذ ان :}$$

وان الوسط الحسابي والانحراف المعياري للمعيار هما :

$$\bar{L}_2 = \overline{\text{MSE}} = \frac{1}{R} \sum_{r=1}^R L_2^{(r)} = \frac{\sum_{r=1}^R \text{MSE}^{(r)}}{R} \quad (36)$$

$$\text{SD}(L_2) = \sqrt{\frac{1}{R-1} \sum_{r=1}^R (\text{MSE}^{(r)} - \overline{\text{MSE}})^2} \quad (37)$$

### 3-3 : معيار خطأ تقدير نقطة العتبة ( مقدار التحيز )

يمثل هذا المعيار الفرق المطلق بين مقدر المعلمة لنقطة العتبة والقيم الحقيقية لها، ويرمز له في المحاكاة  $\overline{\text{Tau Abs}}$ ، وتكتب صيغته كما يأتي :

$$\text{Tau Abs}^{(r)} = |\hat{\gamma}^{(r)} - \gamma| \quad (38)$$

$$\overline{\text{Tau Abs}} = \text{Bias}_{\gamma} = \frac{1}{R} \sum_{r=1}^R \text{Tau Abs}^{(r)} \quad (39)$$

والانحراف المعياري لمعيار خطأ تقدير نقطة العتبة في تجربة المحاكاة لـ R التكرارات يكتب كما يأتي :

$$SD(L_2) = \sqrt{\frac{1}{R-1} \sum_{r=1}^R (\text{Tau Abs}^{(r)} - \overline{\text{Tau Abs}})^2} \quad (40)$$

### Sensitivity of Selection

### 4-3 : معيار الاداء حساسية الاختيار

ويُسهّم هذا المعيار في التعرف على قدرة الانموذج على تحديد المتغيرات التوضيحية المهمة فعلاً عن طريق نسبة يمكن ايجادها وايجاد الوسط الحسابي والانحراف المعياري لها كما يأتي :

$$SEN^{(r)} = \frac{\text{عدد المتغيرات التوضيحية المؤثرة فعلاً المكتشفة}}{\text{عدد المتغيرات التوضيحية المؤثرة فعلاً الحقيقية}} \quad (41)$$

$$\overline{SEN} = \frac{1}{R} \sum_{r=1}^R SEN^{(r)} \quad (42)$$

$$SD(SEN) = \sqrt{\frac{1}{R-1} \sum_{r=1}^R (SEN^{(r)} - \overline{SEN})^2} \quad (43)$$

### Specificity

### 5-3 : معيار الاداء خصوصية الاختيار

$$SPE^{(r)} = \frac{\text{عدد المتغيرات التوضيحية غير المؤثرة المكتشفة بشكل صحيح}}{\text{عدد المتغيرات التوضيحية غير المؤثرة}} \quad (44)$$

$$\overline{SPE} = \frac{1}{R} \sum_{r=1}^R SPE^{(r)} \quad (45)$$

$$SD(SPE) = \sqrt{\frac{1}{R-1} \sum_{r=1}^R (SPE^{(r)} - \overline{SPE})^2} \quad (46)$$

اما الاختبارات المستعملة في الجانب التطبيقي حول انموذج انحدار العتبة TR فهي :

### 6-3 : اختبار نسبة الامكان الاعظم لوجود العتبة باستعمال البوتستراب

يُعد اجراء اختبار وجود العتبة في انموذج انحدار العتبة TR من الخطوات الضرورية للتحقق والتأكد من سلوك علاقة الانحدار بين المتغيرات (المعتمد وبعض أو كل المتغيرات التوضيحية في الانموذج) في التغير من عدمه عند نقطة عتبة (نقطة حرجة) بعبارة أخرى هل تتغير قيم معاملات انموذج الانحدار أم لا عبر جميع مشاهدات المتغيرات التوضيحية؟ وفي ضوء ذلك اقترح الباحث Hansen [9] طريقة اختبار تعتمد على نسبة الامكان الاعظم Likelihood Ratio Test المقترحة اصلاً من الباحث Fisher والمثبتة من الباحث Wills [18] [7] بانها تتوزع توزيعاً تقاربياً الى مربع كاي بدلاً من اختبار والد Wald بسبب ان الاخير يفترض ان تكون معاملات الانموذج جميعها معلومة بما فيها معلمة العتبة (نقطة العتبة) معرفة وقابلة للاشتقاق عند الفرضية الصفرية التي تتضمن تساوي الميول الحدية لان اصل الاختبار يعتمد على مصفوفة التباين والتباين المشترك لمقدّرات المعلمات التي في ضوئها نصل الى التوزيع التقاربي لصيغة الاختبار وهي مربع كاي ، وفي انموذج انحدار العتبة لا يتحقق التوزيع التقاربي المعتاد لإحصاء الاختبار (مربع كاي) بسبب عدم معرفة نقطة العتبة، ما يجعل الاعتماد على القيم الجدولية غير ممكن عند اختبار الفرضية الصفرية. وللتغلب على هذه المشكلة اقترح Bruce E. Hansen استعمال قيم

Bootstrap الحرجة التي يتم توليدها عددياً بالمحاكاة من عينات متعددة تحت فرضية العدم، ثم تُستعمل هذه القيم للمقارنة مع إحصاء الاختبار لاتخاذ قرار قبول أو رفض الفرضية الصفرية وخطواته كما يأتي:

1- كتابة فرضية الاختبار:

$$\left. \begin{array}{l} H_0: \beta_1 = \beta_2 \quad , \text{Traditional regression (لا يوجد عتبة)} \\ H_1: \beta_1 \neq \beta_2 \quad , \text{Threshold regression (يوجد عتبة)} \end{array} \right\} \quad (47)$$

2- تقدير أو اختيار العتبة عن طريق البحث الشبكي grid search

لكون معلمة أو نقطة العتبة  $\gamma$  غير معرفة يتم تقديرها أو اختيارها عن طريق البحث الشبكي grid search، يتكون مجموعة صغيرة بحجم  $m=10$  من القيم الأولية لنقاط العتبة المقترحة وكما يأتي [9]:

grid  $\gamma \in (Q 10 \% \text{ data}, Q 90\% \text{ data})$

$$\text{grid } \gamma = (\gamma_1, \gamma_2, \dots, \gamma_m) \rightarrow \gamma^0 \in (\gamma_1, \dots, \gamma_m)$$

وفي ضوء القيم الأولية لنقاط العتبة المقترحة تُحسب قيمة دالة الامكان لكل قيمة منها ونختار أو نُقدّر العتبة التي تقبل اعلى قيمة دالة امكان على وفق الآتي:

$$\hat{\gamma} = \arg \max_{\gamma_d \in \gamma^0} L(\hat{\beta}_1, \hat{\beta}_2 | \gamma_d) \quad (48)$$

3- احسب قيمة نسبة الامكان الاعظم LR على وفق الاتي :

$$LR = 2[\log L(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}) - \log L(\hat{\beta})] \quad (49)$$

4- احسب قيم البوتستراب bootstrap

عن طريق توليد B من العينات من البيانات الاصلية تحت شروط الفرضية الصفرية، يتم احتساب LR لكل عينة من البوتستراب، ومن ترتيبها تصاعدياً أي أن:

$$LR_{(1)} < LR_{(2)} < \dots < LR_{(B)}$$

وبتحديد قيمة مستوى المعنوية (مثلاً  $\alpha = 0.05$ )، يمكن استخراج قيمة البوتستراب الحرجة على وفق الاتي :

$$LR_{(1-\alpha)} = LR_{(B(1-\alpha))} \quad (50)$$

وعلى ضوء المعادلة (51) فان قيمة البوتستراب الحرجة  $LR_{(0.95)}$  هي القيمة التي تقع في الترتيب  $LR_{(950)}$  والقيمة الاخيرة هي التي تقارن مع قيمة  $LR_{(الاصلية)}$  اي على البيانات الاصلية فإذا كانت  $LR_{(الاصلية)}$  أكبر من  $LR_{(950)}$  فالقرار رفض الفرضية الصفرية وقبول الفرضية البديلة وهذا يعني انه توجد نقطة عتبة.

7-3 : اختبار وجود القيم الشاذة Cook's Distance

يُعد اختبار Cook's Distance احد اهم الاختبارات في تحليل الانحدار لتحديد المشاهدات ذات التأثير الكبير على تقديرات معاملات نموذج الانحدار بمختلف انواعه وتسمياته والتي تسمى بالقيم الشاذة outliers ، بسبب ان هذه القيم يمكن ان تؤثر على تقديرات المعلمات ويمكن تعريف صيغته كما يأتي [1] [3] :

$$D_i = \frac{(\hat{\theta}_{(i)} - \hat{\theta})' X' X (\hat{\theta}_{(i)} - \hat{\theta})}{(k+1)S_e^2} \quad i = 1, 2, \dots, n \quad (51)$$

$$\hat{\theta} = (\hat{\beta}_1 \hat{\beta}_2) \quad , \quad \hat{\theta}_{(i)} = (\hat{\beta}_{1(i)} \hat{\beta}_{2(i)}) \quad \text{إذ ان :}$$

إذ ان (i) :يرمز للمقدرات بعد حذف المشاهدة j.

وتقارن الصيغة (51) مع المقدار  $\frac{4}{n}$  ، ويكون القرار بان المشاهدة مؤثرة إذا كانت قيمة  $D_i$  أكبر من  $\frac{4}{n}$ .

## Simulation

## 4- المحاكاة

باستخدام برنامج Matlab، ستم مناقشة تحليل نتائج المحاكاة بناءً على ثلاثة اعتبارات. أولاً، وجود عدة أنواع من توزيعات الخطأ، وهي: التوزيع الطبيعي  $N(0,1)$ ، وتوزيع  $t$  بثلاث درجات حرية  $t(3)$ ، والتوزيع الطبيعي المختلط (مزيج)  $0.9 \cdot N(0,1) + 0.1N(0,10^2)$ . ثانياً، وجود نسبي تلوث، وهما 10% و 15%. ثالثاً، وجود التلوث أو عدمه في البيانات. وبناءً على ذلك، تكرر توليد البيانات 100 مرة. تم توليد البيانات وفقاً لنموذج الانحدار التالي مع تجربتين:  $(n=100, p=100)$ ،  $(n=100, p=300)$ ، وتم توليد القيم الحقيقية كقيم ابتدائية للمعاملات  $\beta$  مع أخذ أول 5 قيم فقط لا تساوي صفراً، و  $X \sim N(0, \rho^{|i-j|})$ ،  $\rho=0.1$ ، وتظهر النتائج كما يلي:

### 1-4: التجربة الافتراضية الأولى للحالة الأولى عندما $n=100, P=100$ (نسبة تلوث 10%)

من الجدول (1)، تُظهر النتائج أنه عندما يتساوى حجم العينة مع عدد المتغيرات التفسيرية، توجد اختلافات في أداء طريقتي التقدير الحصينة، فقد حققت طريقة Huber-L تفوقاً متوسطاً في الاداء والدقة ، مع استقرار أقل نوعاً ما عند الذبول الطويلة ، مع الحفاظ على توازن بين معايير الحساسية والجودة لاختيار المتغيرات التفسيرية المؤثرة في النموذج. في حين حققت طريقة Pseudo-Tukey-Lasso (PT-L) أخطاءً منخفضة عن تقدير المعلمات، ما يُشير إلى كفاءة الطريقة، ويمتد اداءها الجيد بعدم وجود التلوث، ما يعني عدم تأثرها بنوع التوزيع عند تقدير معاملات نموذج الانحدار العتبية مع البيانات عالية الأبعاد وهذا يدل على تفوقها على طريقة H-L ، ولأن اخطاء تقدير معلمة العتبة صغيرة جداً تُرك عمود Tau Abs فارغاً بعدم وجود التلوث بحيث لا وجود للاختلاف بينها وما تقدم يتوضح في الجدول الاتي :

جدول (1) نتائج المحاكاة لمقارنة طرائق التقدير الحصينة الجزائية لحالة  $(n=p=100)$  في HDTR

في ظل ظروف تلوث بنسبة 10% وظروف عدم التلوث

10%: نسبة التلوث ؛ بوجود التلوث ؛  $(n=100, p=100)$  :  $n = p$

التوزيع	الطريقة	L <sub>2</sub> mean ; SD	Tau Abs mean; SD	SEN mean; SD	SPE mean; SD
N (0,1)	H-L	0.458 (0.069)	3.241 (0.402)	0.034(0.014)	0.934 (0.086)
	P T-L	0.372(0.021)	3.604 (0.262)	0.072(0.028)	0.821 (0.073)
t (3)	H-L	5.812 (0.624)	0.162 (0.130)	0.806 (0.084)	0.727 (0.016)
	P T-L	4.423 (0.902)	0.075(0.031)	0.792 (0.042)	0.756(0.011)
0.9N (0,1) +0.1N (0,10 <sup>2</sup> )	H-L	16.924 (19.843)	0.261 (0.220)	0.694 (0.272)	0.722(0.007)
	P T-L	3.983(0.863)	0.059 (0.022)	0.824 (0.052)	0.769(0.008)
<b>n = p : (n=100 , p=100) ; بعدم وجود التلوث</b>					
N (0,1)	H-L	3.173(0.392)	-	0.936 (0.084)	0.762 (0.005)
	P T-L	3.694 (0.334)	-	0.772(0.052)	0.740 (0.015)
t (3)	H-L	5.286 (0.614)	-	0.824 (0.077)	0.724 (0.015)
	P T-L	5.183(0.431)	-	0.749(0.033)	0.732(0.017)
0.9N (0,1) +0.1N	H-L	4.983(1.138)	-	0.869(0.076)	0.722(0.013)

<b>(0,10<sup>2</sup></b>	<b>P T-L</b>	<b>5.286 (1.292)</b>	<b>-</b>	<b>0.813(0.0 84)</b>	<b>0.731 (0.011)</b>
--------------------------	--------------	--------------------------	----------	--------------------------	----------------------

2-4: التجربة الافتراضية الثانية للحالة الأولى عندما  $n=100, P=100$  (نسبة تلوث 15%)

جدول (2) نتائج المحاكاة لمقارنة طرائق التقدير الحصينة الجزائية لحالة ( $n=p=100$ ) في HDTR

في ظل ظروف تلوث بنسبة 15% وظروف عدم التلوث

<b>15%: نسبة التلوث ; بوجود التلوث ; <math>n = p : (n=100 , p=100)</math></b>					
<b>التوزيع</b>	<b>الطريقة</b>	<b>L<sub>2</sub> mean ; SD</b>	<b>Tau Abs mean; SD</b>	<b>SEN mean; SD</b>	<b>SPE mean; SD</b>
<b>N (0,1)</b>	<b>H-L</b>	<b>2.913(0.35 3)</b>	<b>0.026 (0.011)</b>	<b>0.946 (0.074)</b>	<b>0.773(0.00 4)</b>
	<b>P T-L</b>	<b>3.012 (0.219)</b>	<b>0.059(0.0 21)</b>	<b>0.836 (0.064)</b>	<b>0.782(0.01 0)</b>
<b>t (3)</b>	<b>H-L</b>	<b>5.084 (0.548)</b>	<b>0.142 (0.113)</b>	<b>0.824 (0.071)</b>	<b>0.748 (0.013)</b>
	<b>P T-L</b>	<b>4.104 (0.762)</b>	<b>0.067(0.0 26)</b>	<b>0.813(0.0 42)</b>	<b>0.781 (0.010)</b>
<b>0.9N (0,1) +0.1N (0,10<sup>2</sup></b>	<b>H-L</b>	<b>14.642 (16.804)</b>	<b>0.233(0.1 84)</b>	<b>0.724 (0.242)</b>	<b>0.741 (0.007)</b>
	<b>P T-L</b>	<b>3.641 (0.694)</b>	<b>0.052(0.0 19)</b>	<b>0.842 (0.046)</b>	<b>0.793(0.00 6)</b>
<b>15%: نسبة التلوث ; بعدم وجود التلوث ; <math>n = p : (n=100 , p=100)</math></b>					
<b>N (0,1)</b>	<b>H-L</b>	<b>3.041 (0.362)</b>	<b>-</b>	<b>0.945(0.0 78)</b>	<b>0.768 (0.005)</b>
	<b>P T-L</b>	<b>3.524</b>	<b>-</b>	<b>0.784</b>	<b>0.742</b>

		(0.301)		(0.046)	(0.014)
t (3)	H-L	5.124 (0.592)	-	0.836 (0.072)	0.729(0.01 4)
	P T-L	5.041 (0.418)	-	0.761 (0.031)	0.734 (0.016)
0.9N (0,1) +0.1N (0,10 <sup>2</sup> )	H-L	4.624 (1.010)	-	0.876 (0.071)	0.724 (0.012)
	P T-L	4.984 (1.214)	-	0.824 (0.081)	0.734 (0.010)

ومن الجدول (2) تُظهر النتائج أن طريقة PT-L تتفوق في تقدير معالم نموذج الانحدار العتبة بوجود البيانات عالية الأبعاد عند وجود تلوث بنسبة 15%، كما في الحالة السابقة. فهي تحافظ على قيم  $L_2$  منخفضة وتوازن جيد بين قيم الحساسية SEN والدقة SPE، ما يدل على مرونتها في مواجهة القيم الشاذة، وأظهرت طريقة H-L أداءً متوسطاً، في مواجهة الحالات الشاذة، ولكنها أقل استقراراً من PT-L في بعض الحالات. كما أظهرت التجارب أنه لا يوجد تأثير للتوزيعات ذات الذيل الثقيلة، مثل  $t(3)$ ، على كل من PT-L و H-L وهي أكثر مرونة. وبشكل عام، تُقدم PT-L أفضل توازن بين الدقة واختيار المتغيرات في ظل ظروف إحصائية مختلفة، حتى مع وجود القيم الشاذة، وعليه فإن P T-L هي الطريقة الفضلى على وفق ذلك

#### 3-4: التجربة الافتراضية الأولى للحالة الثانية عندما $n=100, P=200$ (نسبة تلوث 10%)

جدول (3) نتائج المحاكاة لمقارنة طرائق التقدير الحصينة الجزئية لحالة ( $n=100, p=200$ )

في HDTR في ظل ظروف تلوث بنسبة 10% وظروف عدم التلوث

10%: نسبة التلوث ; بوجود التلوث ; ( $n=100, p=200$ ) : $n > p$					
التوزيع	الطريقة	$L_2$ mean ; SD	Tau Abs mean; SD	SEN mean; SD	SPE mean; SD
N (0,1)	H-L	2.983(0.36 1)	0.031(0.0 12)	0.949(0.0 72)	0.783(0.00 4)
	P T-L	3.104 (0.220)	0.058 (0.0202)	0.842 (0.061)	0.792 (0.008)

<b>t (3)</b>	<b>H-L</b>	<b>5.084</b> <b>(0.548)</b>	<b>0.142</b> <b>(0.113)</b>	<b>0.824</b> <b>(0.071)</b>	<b>0.741(0.01</b> <b>284)</b>
	<b>P T-L</b>	<b>4.104</b> <b>(0.762)</b>	<b>0.067(0.0</b> <b>26)</b>	<b>0.813(0.0</b> <b>42)</b>	<b>0.781</b> <b>(0.010)</b>
<b>0.9N (0,1)</b> <b>+0.1N</b> <b>(0,10<sup>2</sup>)</b>	<b>H-L</b>	<b>14.642</b> <b>(16.804)</b>	<b>0.233(0.1</b> <b>84)</b>	<b>0.724</b> <b>(0.242)</b>	<b>0.741</b> <b>(0.007)</b>
	<b>P T-L</b>	<b>3.983(0.86</b> <b>3)</b>	<b>0.059</b> <b>(0.022)</b>	<b>0.824</b> <b>(0.052)</b>	<b>0.769(0.00</b> <b>8)</b>
<b>10% :نسبة التلوث ; بعدم وجود التلوث ; (n=100 , p=200) : n &gt; p</b>					
<b>N (0,1)</b>	<b>H-L</b>	<b>2.704</b> <b>(0.313)</b>	-	<b>0.941</b> <b>(0.052)</b>	<b>0.786</b> <b>(0.004)</b>
	<b>P T-L</b>	<b>2.504</b> <b>(0.238)</b>	-	<b>0.865(0.0</b> <b>41)</b>	<b>0.801</b> <b>(0.008)</b>
<b>t (3)</b>	<b>H-L</b>	<b>4.663(0.50</b> <b>4)</b>	-	<b>0.834</b> <b>(0.054)</b>	<b>0.749(0.01</b> <b>2)</b>
	<b>P T-L</b>	<b>4.062</b> <b>(0.322)</b>	-	<b>0.842</b> <b>(0.031)</b>	<b>0.783(0.00</b> <b>9)</b>
<b>0.9N (0,1)</b> <b>+0.1N</b> <b>(0,10<sup>2</sup>)</b>	<b>H-L</b>	<b>3.184</b> <b>(0.749)</b>	-	<b>0.872</b> <b>(0.051)</b>	<b>0.734</b> <b>(0.011)</b>
	<b>P T-L</b>	<b>2.905(0.68</b> <b>2)</b>	-	<b>0.874</b> <b>(0.046)</b>	<b>0.768</b> <b>(0.008)</b>

تُظهر النتائج في الجدول (3) ان طريقة Huber-Lasso تحقق أداءً متوازنًا بين الحساسية والخصوصية لاختيار المتغيرات، لكنها لا تزال تتأثر بشدة بالتلوث في القيم في الوقت نفسه، أظهرت النتائج قوة ومثانة طريقة PT-Lasso لمختلف التوزيعات وأخطاء التقدير، بالإضافة إلى التوازن بين الحساسية والخصوصية لاختيار المتغيرات في نموذج انحدار العتبة ذي الأبعاد العالية عندما يكون حجم العينة أصغر من عدد المتغيرات التفسيرية.

4-4 : التجربة الافتراضية الثانية للحالة الثانية عندما n=100, P=200 ( نسبة تلوث 15 % )

تتوافق نتائج الجدول (4) مع نتائج الجدول (3)، وهذا يُشير إلى أن هذه الطريقة PT-L تتفوق على طريقة H-L من حيث مقاومتها للتلوث، لا سيما في وجود توزيعات ذات ذيول طويلة وتوزيعات ملوثة مختلطة عندما يكون حجم العينة أصغر من عدد المتغيرات التفسيرية في نموذج انحدار العتبة، ومع البيانات ذات الأبعاد العالية، والجدول الآتي يوضح ما تقدم:

جدول (4) نتائج المحاكاة لمقارنة طرائق التقدير الحصينة الجزائئية لحالة (n= 100 , p=200)

في HDTR في ظل ظروف تلوث بنسبة 15% وظروف عدم التلوث

15%: نسبة التلوث ; بوجود التلوث ; (n=100 , p=200) : n > p					
التوزيع	الطريقة	L <sub>2</sub> mean ; SD	Tau Abs mean; SD	SEN mean; SD	SPE mean; SD
N (0,1)	H-L	0.371(0.0 91)	0.400(0.04 0)	2.837(1.1 92)	0.151(0.14 9)
	P T-L	0.205(0.0 41)	0.400(0.04 0)	2.876(1.1 91)	0.140(0.14 5)
t (3)	H-L	0.573(0.1 63)	5.485(2.75 2)	0.152(0.1 57)	0.517(0.24 5)
	P T-L	0.572(0.1 63)	5.489(2.75 8)	0.152(0.1 57)	0.517(0.24 5)
0.9N (0,1) +0.1N (0,10 <sup>2</sup> )	H-L	1.886(1.0 04)	53.411(44. 085)	0.300(0.2 13)	0.400(0.25 9)
	P T-L	1.883(1.0 03)	53.656(44. 773)	0.289(0.2 15)	0.400(0.25 9)
15%: نسبة التلوث ; بعدم وجود التلوث ; (n=100 , p=200) : n > p					
N (0,1)	H-L	0.406(0.0 61)	-	0.158(0.1 17)	0.508(0.24 1)
	P T-L	0.406(0.0 61)	-	0.158(0.1 17)	0.500(0.23 1)

		61)		17)	7)
t (3)	H-L	0.617(0.1 97)	-	0.194(0.2 19)	0.467(0.23 4)
	P T-L	0.617(0.1 96)	-	0.194(0.2 19)	0.467(0.26 0)
0.9N (0,1) +0.1N (0,10^2)	H-L	1.551(0.6 42)	-	0.330(0.1 88)	0.383(0.26 9)
	P T-L	1.549(0.6 41)	-	0.313(0.1 89)	0.375(0.27 7)

وخلص ما تقدم فقد حققت طريقة PT-L أداءً مستقرًا في التقدير واختيار المتغيرات التفسيرية في نموذج انحدار العتبة ذي الأبعاد العالية، وذلك بتحقيق توازن بين معياري الحساسية والنوعية في الاختيار، فضلًا عن عدم تأثرها بالقيم الشاذة ونوع التوزيعات المفترضة، بما في ذلك التوزيعات ذات الذيل الطويلة والتوزيعات المختلطة الملوثة، سواء كان حجم العينة مساويًا لعدد المتغيرات التفسيرية أو أقل منه. أما طريقة H-L، فقد حققت أداءً متوسطًا في موازنة الحساسية والنوعية عند التقدير واختيار المتغيرات في نموذج انحدار العتبة ذي الأبعاد العالية، لكنها تتأثر بشدة بوجود التوزيعات ذات الذيل السمكة، وبذلك تُعد طريقة PT-L هي الطريقة الفضلى في التقدير واختيار المتغيرات التوضيحية في نموذج HDTR .

## 5- الجانب التطبيقي

بالاعتماد على مجموعة بيانات رقمية معتمدة من البنك الدولي (World Development Indicators) ، وصندوق النقد الدولي (World Economic Outlook) ، وبرنامج الأمم المتحدة الإنمائي (UNDP) ، ومؤشرات الحوكمة العالمية (WGI) عن الاقتصاد العراقي والمتغيرات تم تعريفها في الملحق 2 ، وفي ضوء ذلك تم إجراء التحليل الاحصائي لها وكانت النتائج كما يأتي :

### 1-5 : اختبار وجود العتبة (Threshold Effect Test)

للتحقق من وجود عدم خطية في العلاقة بين المتغيرات المستقلة ونمو الناتج المحلي الإجمالي، تم تطبيق اختبار وجود العتبة المعروف بالصيغة (2-73) ، ويبين الجدول (5) نتائج اختبار العتبة والذي يشير إلى رفض الفرضية الصفرية وقبول الفرضية البديلة ، اي انه توجد نقطة عتبة ، إذ بلغت إحصاء اختبار (LR) 27.84، مع قيمة البوتستراب الحرجة (LR(95)) عند 1000 إعادة معاينة البوتستراب ولمستوى معنوية  $\alpha = 0.05$  ، وهذا القرار يسوغ استعمال نموذج انحدار العتبة (Threshold Regression) في تحليل العلاقة بين المتغيرات الاقتصادية ونمو

الناتج المحلي الإجمالي. بعبارة أخرى ان العلاقة بين محددات النمو الاقتصادي ونمو الناتج المحلي الإجمالي ليست خطية، وإنما تختلف باختلاف مستوى متغير العتبة، وهو ما يعني أن تأثير المتغيرات الاقتصادية يتغير عند الانتقال من نظام اقتصادي إلى آخر وكما يأتي :

جدول (5) نتائج اختبار الكشف عن نقطة العتبة

الاختبار	إحصاء الاختبار $LR_{(الاصلية)}$	قيمة البوتستراب الحرجة $LR_{(B(1-\alpha))}$	القرار
Threshold Test	27.84	$LR_{(950)}$	وجود عتبة

### 2-5 اختبار وجود القيم الشاذة (Cook's Distance)

تم فحص البيانات للكشف عن وجود القيم الشاذة باستخدام مؤشر Cook's Distance ، إذ يبين الجدول (6) وجود عدد من القيم المتطرفة والملاحظات ذات التأثير العالي. إذ تم تحديد عدد من المشاهدات ذات التأثير المرتفع على تقديرات النموذج، وذلك وفق المعيار  $D_i > 4/n$ . وتشير هذه النتائج إلى حساسية نموذج انحدار العتبة لعدد محدود من المشاهدات، مما يسوغ استعمال طرائق التقدير الحصينة (Robust Estimation) في تقدير معالمه ونتائج الاختبار توضح اختبار عينة من 100 مشاهدة موضحة كما في الجدول الاتي :

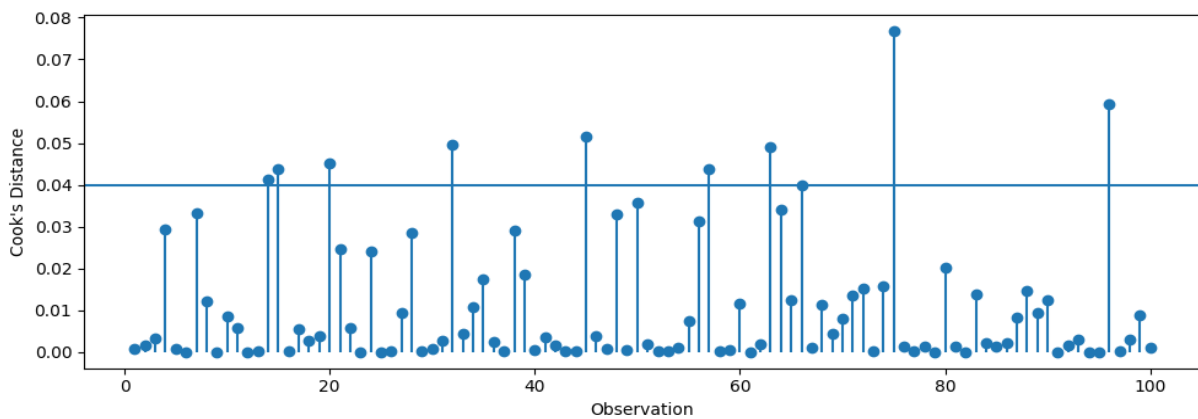
جدول (6) نتائج اختبار Cook's Distance للكشف عن القيم الشاذة في بيانات العينة 100 مشاهدة

i	$D_i$	Influentia ( $D_i > 4/n$ )	i	$D_i$	Influentia ( $D_i > 4/n$ )	i	$D_i$	Influentia ( $D_i > 4/n$ )	i	$D_i$	Influentia ( $D_i > 4/n$ )
1	0.00084	FALSE	2	0.00034	FALSE	5	0.00188	FALSE	76	0.00135	FALSE
2	0.00160	FALSE	7	0.00940	FALSE	2	0.00016	FALSE	77	0.00024	FALSE
3	0.00341	FALSE	8	0.00285	FALSE	5	0.0001	FALSE	78	0.00149	FALSE

				3			7				
<b>4</b>	0.029 44	FALS E	<b>2</b> <b>9</b>	0.0 001 8	FALS E	<b>5</b> <b>4</b>	0.0 011 8	FALS E	<b>79</b>	0.0000 2	FALSE
<b>5</b>	0.000 72	FALS E	<b>3</b> <b>0</b>	0.0 007 2	FALS E	<b>5</b> <b>5</b>	0.0 073 9	FALS E	<b>80</b>	0.0202 7	FALSE
<b>6</b>	0.000 06	FALS E	<b>3</b> <b>1</b>	0.0 026 7	FALS E	<b>5</b> <b>6</b>	0.0 313 2	FALS E	<b>81</b>	0.0014 6	FALSE
<b>7</b>	0.033 17	FALS E	<b>3</b> <b>2</b>	0.0 495 0	TRU E	<b>5</b> <b>7</b>	0.0 436 5	TRUE	<b>82</b>	0.0000 0	FALSE
<b>8</b>	0.012 19	FALS E	<b>3</b> <b>3</b>	0.0 044 3	FALS E	<b>5</b> <b>8</b>	0.0 002 6	FALS E	<b>83</b>	0.0137 4	FALSE
<b>9</b>	0.000 03	FALS E	<b>3</b> <b>4</b>	0.0 108 4	FALS E	<b>5</b> <b>9</b>	0.0 005 0	FALS E	<b>84</b>	0.0021 3	FALSE
<b>10</b>	0.008 50	FALS E	<b>3</b> <b>5</b>	0.0 174 7	FALS E	<b>6</b> <b>0</b>	0.0 115 3	FALS E	<b>85</b>	0.0014 7	FALSE
<b>11</b>	0.005 88	FALS E	<b>3</b> <b>6</b>	0.0 024 4	FALS E	<b>6</b> <b>1</b>	0.0 000 9	FALS E	<b>86</b>	0.0021 7	FALSE
<b>12</b>	0.000 05	FALS E	<b>3</b> <b>7</b>	0.0 002 3	FALS E	<b>6</b> <b>2</b>	0.0 018 6	FALS E	<b>87</b>	0.0081 7	FALSE

13	0.000 34	FALS E	3 8	0.0 289 9	FALS E	6 3	0.0 489 9	TRUE	88	0.0147 9	FALSE
14	0.041 27	TRU E	3 9	0.0 184 8	FALS E	6 4	0.0 340 3	FALS E	89	0.0092 6	FALSE
15	0.043 82	TRU E	4 0	0.0 003 9	FALS E	6 5	0.0 123 6	FALS E	90	0.0124 5	FALSE
16	0.000 25	FALS E	4 1	0.0 035 7	FALS E	6 6	0.0 398 5	FALS E	91	0.0000 0	FALSE
17	0.005 50	FALS E	4 2	0.0 017 0	FALS E	6 7	0.0 010 7	FALS E	92	0.0016 0	FALSE
18	0.002 71	FALS E	4 3	0.0 001 4	FALS E	6 8	0.0 113 2	FALS E	93	0.0030 1	FALSE
19	0.003 77	FALS E	4 4	0.0 001 3	FALS E	6 9	0.0 043 2	FALS E	94	0.0000 0	FALSE
20	0.045 28	TRU E	4 5	0.0 516 6	TRU E	7 0	0.0 081 4	FALS E	95	0.0000 0	FALSE
21	0.024 57	FALS E	4 6	0.0 038 9	FALS E	7 1	0.0 135 1	FALS E	96	0.0593 3	TRUE
22	0.005	FALS	4	0.0 008	FALS	7	0.0 152	FALS	97	0.0003	FALSE

	85	E	7	2	E	2	7	E		1	
<b>23</b>	0.000 00	FALS E	4 8	0.0 330 1	FALS E	7 3	0.0 001 9	FALS E	<b>98</b>	0.0030 0	FALSE
<b>24</b>	0.024 09	FALS E	4 9	0.0 003 9	FALS E	7 4	0.0 157 3	FALS E	<b>99</b>	0.0087 7	FALSE
<b>25</b>	0.000 00	FALS E	5 0	0.0 358 0	FALS E	7 5	0.0 767 3	<b>TRUE</b>	<b>100</b>	0.0010 1	FALSE



الشكل (1) رسم مخطط الساق والأوراق لقيم Cook's Distance

نلاحظ من الجدول (6) ان الحد الحرج للمقارنة  $4/n=0.04$  بسبب ان  $n=100$  ، وفي ضوء ذلك يتضح وجود 9 مشاهدات مؤثرة هي ( 14 ، 15 ، 20 ، 32 ، 45 ، 57 ، 63 ، 75 ، 96 )، ويمثل عدد هذه القيم ما نسبته 9 % من حجم العينة وهي ضمن النسبة المسموح بها وقرابية لنسبة 10% ، وتدل هذه النسبة على تأثير واضح على معاملات الانموذج الامر الذي يستوجب معه استعمال طرائق التقدير الحصينة عند تقدير معاملات انموذج انحدار التبة الممثل للبيانات .

ومن الشكل (1)، نلاحظ أن أغلب Cook's Distance تقع تحت الحد الحرج 0.04، وهذا يدل على عدم وجود تأثير واسع على تقديرات الانموذج وعلى الرغم من ذلك يوجد ما نسبته 9% من القيم المؤثرة التي تؤثر على تلك التقديرات الامر الذي يسوغ استعمال الطرائق الحصينة عند تقدير معاملات انموذج انحدار العتبة الممثل للبيانات بسبب ان لكل من هذه القيم تأثيره الفردي على تلك التقديرات.

### 3-5: تحليل البيانات:

جدول (7): مقارنة أداء طرائق Lasso الحصينة في تفسير نمو الناتج المحلي الإجمالي

الطريقة الحصينة الجزائية	Alpha	المعاملات غير الصفريّة	RMSE	R <sup>2</sup>	MAD _Residuals
H-L	0.1	14	0.68534	0.39010	0.45740
PT-L	0.1	17	0.51854	0.56460	0.43593

يبين جدول (7) مقارنة أداء نماذج لاسو الحصينة المختلفة في تفسير نمو الناتج المحلي الإجمالي، إذ أظهرت الطريقتان فروقاً واضحة في اختيار المتغيرات التوضيحية ودقة التقدير. إذ تفوقت طريقة PT-L على طريقة H-L، فقد حققت قيمة لجذر متوسط مربعات خطأ، إلى جانب أعلى معامل تحديد بلغ (0.56460)، ما يشير إلى قدرته الأعلى على تفسير التباين في نمو الناتج المحلي الإجمالي مع الاحتفاظ بعدد أكبر نسبياً من المتغيرات المؤثرة (17 متغيراً). بينما أظهرت طريقة H-L أداءً أضعف نسبياً من حيث القوة التفسيرية البالغة (0.39010)، على الرغم من اعتماده على عدد أقل من المتغيرات المختارة، وهو ما يعكس ميله إلى تقديرات أكثر تحفظاً في ظل وجود القيم الشاذة. ، ويمكن توضيح أداء الطريقتين كما يأتي :

جدول (8) نتائج تقييم أداء طرائق Robust LASSO

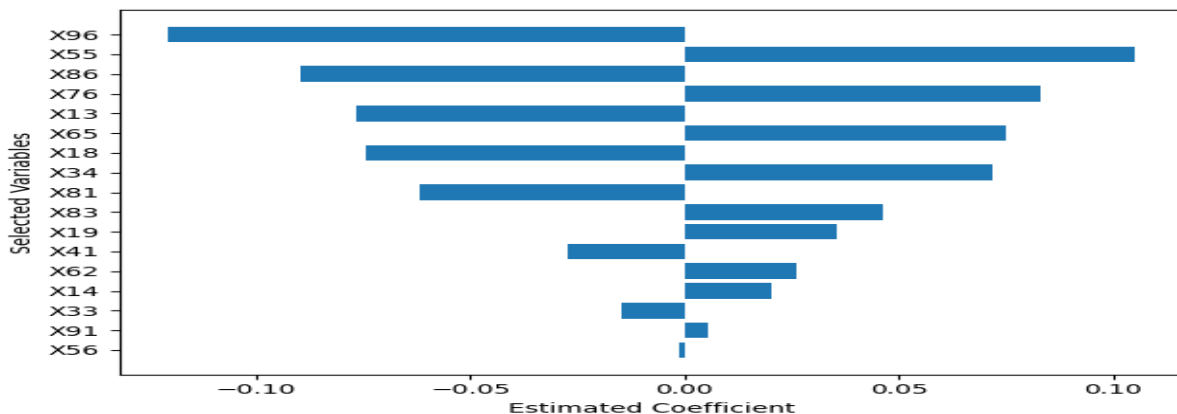
Method	True_Mean SD	L2_Mean SD	TauAbs_Mean SD	SEN_M eanSD	SPE_Mea nSD
H-L	0.00944	0.02697	0.00944	0.80000	0.82800
PT-L	0.00729	0.02287	0.00729	0.88000	0.85900

يبين جدول (8) القيم إلى فروق بين الطريقتين إذ أن PT-L قد حقق أداءً أفضل من طريقة H-L بسبب تسجيلها قيمة أقل لمقاييس الخطأ True\_MeanSD و L<sub>2</sub>\_MeanSD و TauAbs\_MeanSD، ما يدل على انخفاض الانحراف الكلي لمعاملات الانموذج واقتربها من القيم الحقيقية مقارنة بالطريقة H-L. كما حققت الطريقة أعلى قيمة لمؤشر SPE بلغ (0.859)، ما يعكس قدرتها أعلى استبعاد المتغيرات غير المؤثرة في المتغير المعتمد بدقة عالية، وهو أمر مهم في وجود مشكلة الأبعاد العالية. وفي ضوء النتائج السابقة يمكن عرض قيم تقديرات معاملات انحدار العتبة باستخدام طريقة PT-L كما يأتي:

جدول (9) تقديرات معاملات انحدار العتبة باستخدام طريقة PT-L

Variable	Coefficient	Variable	Coefficient
X96	-0.12066	X83	0.04632
X55	0.10506	X19	0.03552
X86	-0.08935	X41	-0.02719
X76	0.08295	X62	0.02614
X13	-0.07674	X14	0.02038
X65	0.07529	X33	-0.01463
X18	-0.07409	X91	0.00537
X34	0.07197	X56	-0.00103
X81	-0.06183		

يبين جدول (9) ان مجموعة محدودة من المتغيرات البالغ عددها (17) متغيراً ذات التأثير المعنوي في معدل نمو الناتج المحلي الإجمالي في العراق ، مع اختلاف في اتجاه وقوة هذا التأثير. وتظهر بعض المتغيرات بتقديرات سالبة ، مثل المتغير  $X_{96}$  و  $X_{86}$  و  $X_{13}$ ، مما يشير إلى أن ارتفاع هذه المتغيرات يرتبط بانخفاض معدل نمو الناتج المحلي الإجمالي، وهو ما قد يعكس آثاراً سلبية لعوامل اقتصادية أو مؤسسية معينة عند تجاوز مستوى العتبة. في المقابل، تسجل متغيرات أخرى مثل  $X_{55}$  و  $X_{76}$  و  $X_{65}$  تقديرات موجبة، ما يدل على دورها الإيجابي في دعم النمو الاقتصادي داخل النظام الاقتصادي محل الدراسة. ونلاحظ أن بعض المتغيرات جاءت بتقديرات صغيرة جداً قريبة من الصفر، مثل  $X_{91}$  و  $X_{56}$ ، وهو ما يعكس التأثير المحدود لها. وعليه تؤكد هذه النتائج قدرة طريقة PT-L على تحقيق توازن بين اختيار المتغيرات المؤثرة ، مما يعزز موثوقية التقديرات ما يسوغ استعمال هذا الأسلوب في تحليل بيانات اقتصادية تتسم بعدم الخطية ووجود قيم شاذة وبأبعاد عالية. وبتمثيل قيم التقديرات لمعاملات انموذج HDTR نحصل على الآتي :



## الشكل (2) تقدير معاملات المتغيرات التوضيحية المختارة انموذج انحدار العتبة باستعمال طريقة PT-L

توضح النتائج الاقتصادية أن نمو الناتج المحلي الإجمالي في العراق يتحدد على عدد محدود من المتغيرات التفسيرية، إذ اختارت طريقة PT-L عددًا قليلاً من المتغيرات 17 من أصل 100 متغير، شملت مؤشرات ترتبط بالإنتاجية والتحول الرقمي والانفتاح المالي والاستدامة المالية والحوكمة والثقة بالمؤسسات، فضلاً عن متغيرات القطاع الخارجي مثل شروط التبادل التجاري وتقلب أسعار النفط. كما اختارت الطريقة متغيرات رأس المال البشري والتحول الهيكلي، مثل مؤشر التنمية البشرية وجودة التعليم ونمو القطاع الزراعي، إضافة إلى متغيرات مالية ونقدية كالاستهلاك الحكومي وسعر الصرف الحقيقي، ما يؤكد أن النمو الاقتصادي يتحدد بعوامل أساسية تتفاعل بصورة غير خطية ضمن إطار انموذج انحدار العتبة. وأظهرت النتائج أن إنتاجية العمل والانفتاح المالي يؤثران إيجابياً في النمو، في حين كان لتقلب أسعار النفط ونمو القطاع الزراعي أثر سلبي، بما يعكس حساسية الاقتصاد العراقي للعوامل الخارجية والتحويلات الهيكلية.

### الاستنتاجات والتوصيات

أظهرت نتائج الجانبين التجريبي والتطبيقي أن طريقة PT-L حققت أفضل أداء وأكثره استقراراً في اختيار المتغيرات التفسيرية لانموذج انحدار العتبة ذي الأبعاد العالية، إذ إن تحقق توازن بين الحساسية والنوعية وعدم تأثرها بالقيم الشاذة أو بأنواع التوزيعات المختلفة، حتى مع العينات الصغيرة أو الملوثة. في المقابل، أظهرت طريقة LAD-L حساسية مرتفعة لكنها عانت من ضعف النوعية، بينما حققت طريقة H-L أداءً متوسطاً مع تأثير واضح بالتوزيعات ذات الذيل السمكة. كما بينت النتائج التطبيقية أن النمو الاقتصادي في العراق يتأثر بمزيج من الإنتاجية والانفتاح وجودة السياسات مع وجود آثار غير خطية لبعض المتغيرات. وبناءً على ذلك يوصي الباحثان باعتماد طريقة PT-L لتقدير معاملات انموذج انحدار العتبة في حالات الأبعاد العالية والقيم الشاذة، واعتماد انموذج العتبة المكون من 17 متغيراً لتفسير الناتج المحلي الإجمالي في العراق.

### المصادر

- 1- عبد المنعم ، زينة حكمت (2023). مقارنة بعض اساليب البوتستراب لتقدير معاملات انموذج الانحدار بوجود بيانات ذات الابعاد العالية . رسالة ماجستير، كلية الإدارة والاقتصاد، جامعة واسط.
- 2- يوسف ، علي حميد (2017) . استعمال المقدرات الحصينة الجزائية لنموذج الانحدار الخطي في ظل وجود مشكلتي الابعاد والقيم الشاذة مع تطبيق عملي) اطروحة دكتوراه في الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد.
- 3- Abdelmonim, Z. H. & Msallam, B. Sh. (2023). Weighted bootstrap probability - robust least angle regression method to selection variables in linear regression with high dimensional and outliers. Proc. SPIE 12936, International Conference on Mathematical and Statistical Physics, Computational Science, Education and Communication (ICMSCE 2023), 129361V.
- 4- Caner, M. & Hansen.B.(2001). Threshold Autoregression with A Unit Root. Econometrica, Vol. 69, No. 6, 1555 -1596.

- 5- Caner, M. (2002). A Note On Least Absolute Deviation Estimation of A Threshold Model. *Econometric Theory*, 18, 800-814.
- 6- Chang, L et al. (2017). Robust Lasso Regression Using Tukey's Biweight Criterion. *Technometrics*, ( ), 0–0. doi:10.1080/00401706.2017.13052.
- 7- Chong, T. T- L. (1999). Asymptotic distribution of the Sup-Wald statistic under specification errors. *Structural Change and Economic Dynamics*, 10(3-4), 421-430.
- 8- Gannaz, I. (2007). Robust estimation and wavelet thresholding in partially linear models. *Stat Comput* 17, 293–310.
- 9- Hansen, B.E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics* 93, 345-368.
- 10- Hansen, B.E. (2000). Sample Splitting and Threshold Estimation. *Econometrica*, Vol. 68, No. 3 ,575-603.
- 11- Hidalgo, J. et al (2019). Robust inference for threshold regression models. *Journal of Econometrics* 210, 291–309.
- 12- Lee, S. et al. (2016). The lasso for high dimensional regression with a possible change point. *J. R. Statist. Soc. B*, 78, Part 1, pp. 193–210.
- 13- Li, B.-w. et al. (2020). Robust Variable Selection and Estimation in Threshold Regression Model. *Acta Mathematicae Applicate Sinica, English Series*, Vol. 36, No. 2, 332-346.
- 14- Liu, B. et al. (2024). Robust High-Dimensional Regression with Coefficient Thresholding and its Application to Imaging Data Analysis. *Am Stat Assoc.*, 119(545): 715–729.
- 15- Su, p. et al. (2024). CR-Lasso: Robust cellwise regularized sparse regression. *Computational Statistics and Data Analysis* 197, 107971.
- 16- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *J. R. Statist. Soc. B* 58, No. 1, pp. 267-288.
- 17- Tong, H. (1978). On a Threshold Model in Pattern Recognition and Signal Processing. *Sijthoff & Noordhoff, Alphen aan den Rijn-The Netherlands*.
- 18- Tsay, R. S. (1998). Testing and Modeling Multivariate Threshold Models. *Journal of the American Statistical Association*, Vol. 93, No. 443.

الملحق (1) توليد البيانات

توليد المعلمات الحقيقية للانموذج  $\delta_0$  و  $\beta_0$

لغرض بناء نموذج انحدار العتبة TR تم تحديد عدد المعلمات غير الصفرية (Sparisty) كما يأتي:

$$sBeta = 5; \quad sDelta = 5$$

ومن ثم انشاء  $\beta_0$  و  $\delta_0$  على شكل متجهات sparse وكما يأتي:

$$beta0 = zeros(p,1);$$

$$delta0 = zeros(p,1);$$

ومن ثم تحديد مواقع القيم غير الصفرية وتعيينها وكما يأتي:

$$idxB = 1:sBeta;$$

$$idxD = (sBeta+1):(sBeta+sDelta);$$

$$beta0(idxB) = linspace(1.2, 0.6, sBeta)';$$

$$delta0(idxD) = linspace(-1.0, -0.5, sDelta)';$$

وهذا يعني ان أول  $sBeta$  معاملات في  $\beta_0$  غير صفرية وتكون مؤثرة قبل العتبة, وان معاملات  $\delta_0$  غير الصفرية في مواقع لاحقة (تغيير إضافي بعد العتبة).

## 2- توليد الضوضاء $\varepsilon$ ومكان دخول التلوث :

ويتم أولاً بتحديد معاملات الضوضاء/التلوث وكما يأتي:

$$sigma = 1.0;$$

$$t\_df = 2;$$

$$piOut = 0.15;$$

$$kappa = 20;$$

$sigma$ : حجم الضوضاء.

$piOut$ : نسبة التلوث (احتمال أن تكون المشاهدة شاذة ضمن الضوضاء)

$kappa$ : تضخيم قيمة الضوضاء عند المشاهدات الشاذة.

$t\_df$ : درجات حرية توزيع  $t$  (ذيل ثقيل).

ومن ثم توليد  $\varepsilon$  حسب نوع الضوضاء وكما يأتي:

```
eps = make_noise_scenario(n, sigma, Noise{kk}.mode, t_df, piOut, kappa);
```

وداخل كل حالة افتراضية يتم اختيار :

```
case "normal" -> eps = N(0, sigma^2)
```

```
case "t" -> eps ~ t_df
```

```
case "contam" ->
```

```
case "de" -> Double Exponential (Laplace)
```

وبالتحديد في حالة التلوث:

```
isOut = (rand(n,1) < piOut);
```

```
eps(isOut) = sigma*kappa. *randn(sum(isOut),1);
```

إذ ان التلوث يدخل عبر الضوضاء eps وبالتالي يظهر مباشرة في متغير الاستجابة y .

### 3- توليد المتغيرات التوضيحية X

لغرض توليد مصفوفة المتغيرات التوضيحية تم اتباع الخطوات الآتية:

- تحديد أبعاد مصفوفة المتغيرات التوضيحية : تم تحديد حجم العينة n وعدد المتغيرات التوضيحية p كما يأتي:

```
n = 100;
```

```
p = 100;
```

تم بعدها لتحكم في درجة الارتباط بين المتغيرات التوضيحية للتحكم في مشكلة التعدد الخطي، إذ تم افتراض أن المتغيرات التوضيحية تتبع توزيعاً طبيعياً متعدد المتغيرات بمصفوفة تغاير من نوع AR(1) وقد تم تحديد معامل الارتباط  $\rho$  وفق الآتي:

```
rhoList = [0.1 0.0];
```

حيث تمثل الحالة  $\rho=0.1$  وجود ارتباط موجب قليل بينها.

- بناء مصفوفة التغاير وتحليل Cholesky

تم إنشاء مصفوفة التغاير  $\Sigma$  ذات البنية AR(1) ثم تحليلها باستخدام Cholesky ، وذلك باستعمال الدالة:

```
C = ar1_chol (p, rho);
```

التي تعرف كما يأتي:

$$\text{Sigma} = \text{toeplitz}(\text{rho.} \wedge (0: p-1)) + 1e-10*\text{eye}(p);$$

$$C = \text{chol}(\text{Sigma}, 'lower');$$

حيث تضمن الإضافة الصغيرة للمصفوفة القطرية الاستقرار العددي.

وداخل الدالة: ar1\_chol:

$$v = \text{rho.} \wedge (0: p-1);$$

$$\text{Sigma} = \text{toeplitz}(v) + 1e-10*\text{eye}(p);$$

$$C = \text{chol}(\text{Sigma}, 'lower;')$$

إذ ان  $\text{toeplitz}(v)$  يبني مصفوفة AR(1)

+  $1e-10*\text{eye}(p)$  لضمان الاستقرار العددي وأن  $\Sigma$  مصفوفة pd

chol يعطي مصفوفة C بحيث  $\Sigma=CC'$

توليد مصفوفة عشوائية Z من التوزيع الطبيعي القياسي وكما يأتي:

$$\text{randn}(n,p)$$

$$X=ZCT \Rightarrow X \sim N(0, \Sigma)$$

ومن ثم تحويل Z إلى X بحيث تحمل الارتباط المطلوب وكما يأتي:

$$X = \text{randn}(n,p) * C';$$

وبهذا يكون كل صف من X متجه مشاهدات ، بينما تمثل الأعمدة المتغيرات التوضيحية التي قد تكون مستقلة أو مترابطة تبعاً لقيمة p .

### 1- حساب متغير الاستجابة y (انموذج انحدار العتبة)

$$y = X*\text{beta}0 + (X*\text{delta}0). *I0 + \text{eps};$$

ويتم وفق الانموذج الاتي:

### 2- تحديد قيمة العتبة (Threshold Value)

تم تحديد قيمة العتبة  $\tau$  في انموذج انحدار العتبة TR بوصفها متغيراً توضيحياً ، يتحكم في تقسيم البيانات إلى نظامين مختلفين من حيث معالم الانحدار ، ففي تجارب المحاكاة عرفت العتبة (Threshold) على أساس متغير العتبة  $q_i$  حيث يؤدي تجاوزه للقيمة  $\tau$  إلى انتقال النموذج من النظام الأول إلى النظام الثاني، وهو ما ينعكس في تفعيل معالم التغير  $\delta$  و عليه تم تثبيت قيمتين افتراضيتين للعتبة عند  $\tau_0=0.10, 0.15$  باستعمال الامر :  $\text{tau}0 = 0.15;$

### 3- تحديد متغير العتبة (Threshold Variable)

تم توليد متغير العتبة  $q_i$  في انموذج انحدار العتبة عشوائياً من توزيع منتظم مستقل عن المتغيرات التوضيحية وحد الخطأ، لضمان أن يكون متغير العتبة غير ملوث وغير مرتبط بالوضوءاء، مما يسمح بعزل تأثير العتبة عن تأثير التلوث في حد الخطأ وذلك وفق الأمر:

$$q = \text{rand}(n,1);$$

#### 4- ادراج قيمة العتبة في صيغة الانموذج

أدرج تأثير العتبة في صيغة متغير الاستجابة من خلال السماح بتغيير معاملات الانحدار عند تجاوز  $q_i$  لقيمة  $\tau_0$ ، كما في المعادلة الآتية:

$$y = X*\beta_0 + (X*\delta_0) * I_0 + \epsilon$$

ويمثل الحد  $(X*\delta_0)$  التغير أو الاختلاف في معاملات الانموذج الناتج عن وجود العتبة.

#### 5- تحديد مجموعة القيم المرشحة للعتبة (Grid)

لغرض تقدير قيمة العتبة، تم بناء شبكة من القيم المحتملة اعتماداً على كميات من متغير العتبة  $q$ ، مع استبعاد الأطراف لتفادي تأثير القيم المتطرفة، وذلك باستخدام الأمر الآتي:

$$qs = \text{quantile}(q, \text{linspace}(\text{trimQ}, 1-\text{trimQ}, \text{gridSize}));$$

إذ يحدد  $\text{trimQ}$  نسبة القص من الأطراف، بينما يحدد  $\text{gridSize}$  عدد نقاط الشبكة.

#### 6- تقدير العتبة باستخدام البحث الشبكي

تم تقدير قيمة العتبة من خلال البحث الشبكي عبر جميع القيم في  $qs$ ، واختيار القيمة التي تقلل دالة الهدف، وذلك من خلال الدالة الآتية:

$$[\text{bestB}, \text{bestD}, \text{bestTau}, \text{bestObj}] = \text{fit\_on\_grid}(X, y, q, qs, \dots);$$

#### 7- التمييز بين العتبة ونسبة التلوث

أن قيمة العتبة  $\tau$  ليست نسبة تلوث، ولا ترتبط بألية إدخال القيم الشاذة في البيانات. فالتلوث أُدخل بشكل مستقل عبر حد الخطأ العشوائي فقط، كما في:

$$\epsilon = \text{make\_noise\_scenario}(n, \text{sigma}, \text{Noise}\{kk\}.\text{mode}, t\_df, \text{piOut}, \text{kappa});$$

#### ملحق (2)

**Y** معدل نمو الناتج المحلي الإجمالي في الفترة المستقبلية (Future GDP Growth)

ويستخدم لقياس الأداء الاقتصادي المستقبلي، حيث يسعى انموذج الانحدار ذي العتبة الى تفسير العلاقة ، والمتغيرات التوضيحية كما يأتي :

أولاً: متغيرات الاقتصاد الكلي وتشمل المتغيرات التوضيحية (

الناتج المحلي الإجمالي الحقيقي (Real Gross Domestic Product – Real GDP) ، معدل نمو الناتج

المحلي الإجمالي (GDP Growth Rate) ، الناتج المحلي الإجمالي للفرد (GDP per Capita) ، معدل

التضخم السنوي (Inflation Rate) ، معدل البطالة (Unemployment Rate) ، عرض النقود الواسع (Broad Money Supply – M2) ، سعر الصرف الرسمي (Official Exchange Rate) ، الاحتياطيات الدولية (International Reserves) ، ميزان الحساب الجاري كنسبة من الناتج (Current Account Balance % of GDP) ، ميزان التجارة الخارجية (Trade Balance) ، نسبة الانفتاح التجاري (Trade Openness) ، معدل الادخار المحلي (Domestic Savings Rate) ، معدل الاستثمار الكلي (Gross Capital Formation) ، استهلاك القطاع الخاص (Private Consumption) ، استهلاك القطاع الحكومي (Government Consumption) ، معدل نمو الإنتاج الصناعي (Industrial Production Growth) ، معدل نمو القطاع الزراعي (Agricultural Growth Rate) ، معدل نمو قطاع الخدمات (Services Growth Rate) ، الناتج النفطي كنسبة من الناتج (Oil GDP Share) ، تقلبات الناتج المحلي (GDP Volatility)

ثانياً: المالية العامة والدين وتشمل المتغيرات التوضيحية ( )

الدين العام كنسبة من الناتج (Public Debt to GDP) ، الدين الخارجي (External Debt) ، عجز/فائض الموازنة العامة (Fiscal Balance) ، الإنفاق الحكومي الكلي (Total Government Expenditure) ، الإيرادات الحكومية (Government Revenue) ، الإيرادات النفطية (Oil Revenues) ، الإيرادات غير النفطية (Non-Oil Revenues) ، الإنفاق الاستثماري الحكومي (Public Investment Expenditure) ، الإنفاق الجاري الحكومي (Current Government Expenditure) ، خدمة الدين العام (Debt Service) ، نسبة الضرائب إلى الناتج (Tax Revenue % of GDP) ، الإنفاق على التعليم (Education Expenditure) ، الإنفاق على الصحة (Health Expenditure) ، الإنفاق العسكري (Military Expenditure) ، دعم الطاقة (Energy Subsidies) ، العجز الأولي (Primary Fiscal Balance) ، الدين المحلي (Domestic Debt) ، متوسط أجل استحقاق الدين (Average Debt Maturity) ، تكلفة الاقتراض الحكومي (Government Borrowing Cost) ، الاستدامة المالية (Fiscal Sustainability Index)

ثالثاً: القطاع الخارجي والنفط وتشمل المتغيرات التوضيحية ( )

صادرات النفط الخام (Crude Oil Exports) ، إنتاج النفط الخام (Crude Oil Production) ، أسعار النفط العالمية (Global Oil Price) ، واردات السلع والخدمات (Imports of Goods and Services) ، صادرات السلع والخدمات (Exports of Goods and Services) ، تدفقات الاستثمار الأجنبي المباشر (Foreign Direct Investment) ، صافي تدفقات رأس المال (Net Capital Flows) ، شروط التبادل التجاري (Terms of Trade) ، تحويلات العاملين (Workers' Remittances) ، الاحتياطيات النفطية المؤكدة (Proven Oil Reserves) ، إيرادات الصادرات غير النفطية (Non-Oil Export Revenues) ، درجة تنويع الصادرات (Export Diversification Index) ، الاعتماد على النفط (Oil Dependency Ratio) ، سعر الصرف الحقيقي (Real Exchange Rate) ، تقلب أسعار النفط (Oil Price Volatility) ، الانفتاح المالي (Financial Openness Index) ، تدفقات المساعدات الخارجية (Foreign Aid)

**Inflows** ، **الحساب الرأسمالي (Capital Account Balance)** ، **الميزان النفطي (Oil Trade Balance)** ، **نسبة الصادرات إلى الناتج (Exports to GDP)**

رابعاً: المتغيرات المؤسسية والحوكمة وتشمل المتغيرات التوضيحية (

**جودة المؤسسات (Institutional Quality Index)** ، **فعالية الحكومة (Government Effectiveness)** ، **سيادة القانون (Rule of Law)** ، **مكافحة الفساد (Control of Corruption)** ، **الاستقرار السياسي (Political Stability)** ، **جودة التنظيم الحكومي (Regulatory Quality)** ، **استقلالية البنك المركزي (Central Bank Independence)** ، **كفاءة الإدارة العامة (Public Administration Efficiency)** ، **الشفافية المالية (Fiscal Transparency)** ، **جودة السياسات الاقتصادية (Economic Policy Quality)** ، **حماية حقوق الملكية (Property Rights Protection)** ، **سهولة ممارسة الأعمال (Ease of Doing Business Index)** ، **الثقة في المؤسسات العامة (Institutional Trust Index)** ، **الحوكمة الاقتصادية (Economic Governance Index)** ، **الاستقرار المؤسسي (Institutional Stability)**

خامساً: رأس المال البشري والتنمية وتشمل المتغيرات التوضيحية (

**معدل نمو السكان (Population Growth Rate)** ، **نسبة السكان في سن العمل (Working Age Population Share)** ، **متوسط العمر المتوقع (Life Expectancy)** ، **معدل الأمية (Illiteracy Rate)** ، **معدل الالتحاق بالتعليم العالي (Tertiary Enrollment Rate)** ، **مؤشر رأس المال البشري (Human Capital Index)** ، **معدل الفقر (Poverty Rate)** ، **معامل جيني (Gini Coefficient)** ، **الإنفاق على البحث والتطوير (R&D Expenditure)** ، **إنتاجية العمل (Labor Productivity)** ، **مشاركة القوى العاملة (Labor Force Participation)** ، **معدل التحضر (Urbanization Rate)** ، **جودة التعليم (Education Quality Index)** ، **جودة الخدمات الصحية (Health Quality Index)** ، **التنمية البشرية (Human Development Index)**

سادساً: التكنولوجيا والبنية التحتية وتشمل المتغيرات التوضيحية (

**انتشار الإنترنت (Internet Penetration Rate)** ، **استخدام الهاتف المحمول (Mobile Phone Subscriptions)** ، **البنية التحتية للنقل (Transport Infrastructure Index)** ، **البنية التحتية للطاقة (Energy Infrastructure Index)** ، **إنتاج الكهرباء (Electricity Production)** ، **استهلاك الطاقة (Energy Consumption)** ، **كثافة الطاقة (Energy Intensity)** ، **الابتكار التكنولوجي (Technological Innovation Index)** ، **التحول الرقمي (Digital Transformation Index)** ، **الإنتاجية الكلية لعوامل الإنتاج (Total Factor Productivity)** .