مجلة التجارة والتمويل

/https://caf.journals.ekb.eg

كلية التجارة – جامعة طنطا

العدد : الأول

مارس ٢٠٢٣

دراسة مقارنة لطرق علاج مشكلة الازدواج الخطي
بالتطبيق علي الهجرة الداخلية في مصر

الدكتور

عبد الرحيم عوض عبد الخالق بسيوني

المستخلص:

استهدف البحث المفاضلة بين طريقة المربعات الصغرى (OLS) وطريقة انحدار الحرف Ridge Regression (RR) وطريقة المكونات الرئيسية (PCR) وطريقه انحدار لاسو (Lasso) وذلك للوصول للطريقة الأفضل لعلاج مشكلة الازدواج الخطي في نموذج الانحدار المتعدد وتوصلت الدراسة إلى أفضلية طريقة انحدار لاسو (Lasso) عن كُلاً من طريقة المُربعات الصغرى (OLS) وطريقه المكونات الرئيسية (PCR) وطريقه انحدار الحرف (RR) وَذلك لأنها صاحبة أقل متوسط مُربعات أخطاء (MSE) وأقل قيم لمُعامل تضخم التباين VIF.

Abstract:

The research aimed to compare between ordinary the least squares method (OLS), the Ridge regression (RR) ,the principal components regression (PCR) and the Lasso Regression in order to reach the optimal method for treating the problem of Multicollinearity The study concluded that the Lasso Regression method is superior to the least squares method and the principal components method, and Ridge Regression as it has the least mean squares errors (MSE) and the lowest values of the variance inflation factor (VIF).

مقدمة: -

يُعد تحليل الانحدار أحد أهم الأساليب الإحصائية والأكثر استخدامًا في تحليل العلاقات بين المتغيرات ووضعها في شكل نموذج يتكون من متغير تابع وواحد أو أكثر من مُتغير مُستقل وفي حالة أن احتواء نموذج الانحدار على متغير مستقل واحد يسمى بالانحدار البسيط وفي حالة تعدد المتغيرات المستقلة التي تدخل في تفسير الظاهره محل الدراسه والمُتداخلة فيما بينها بحيث يصعب قياس أثر كل منها على حدى على الظاهرة ومن أكثر الطرق استخدامًا في تقدير معلمات الانحدار هي طريقة المربعات الصغرى العاديةس Ordinary least squares (OLS) والتي تتميز بفاعليتها عن باقي الطرق وتتطلب طريقة المربعات الصغرى مجموعة من الافتراضات التي قد يصعب توافرها في الواقع وفي حالة توافرها تصبح طريقة المربعات الصغرى هي الأكثر مُلائمة لتقدير معلمات الانحدار أما في حالة عدم توافرها تكون طريقة المربعات الصغرى غير صالحة وبالتالي يتعين البحث عن طرق تقدير بديلة لها كما ينتج من عدم توافر هذه الافتراضات ظهور المشاكل القياسية التي تجعل تقديرات طريقة المربعات الصغرى متحيزة وغير كفء ومن المشاكل القياسية الناتجة هي مشكلة الازدواج الخطي Multicollinearity، ولقد تناول هذه المُشكلة العديد من الدراسات السابقة حيث قارنت دراسة Fuwenjiang(1998) بين طريقة انحدار الحرف (RR) وطريقة انحدار لاسو (LASSO) مستخدما اسلوب المحاكاه وتبين أفضلية انحدار لاسو كما استعرض الشيخ (2011) مستخدما أسلوب انحدار الحرف (Ridge) المُناسب في حالة وجود مُشكلة الازدواج الخطي وذلك دون اللجوء لحذف المُتغيرات ذات الارتباط العالي وتم تطبيق هذة الدراسة علي دالة الاستهلاك بالجزائر خلال الفترة من (1910 - 2011) كما استخدمت دراسة Lukman et al (2014) مقدرات طريقة انحدار الحرف ومقدرات طريقة المُربعات الصغرى في حالة وجود مشكلة الازدواج الخطي والقيم المتطرفة كما سعت دراسة طه (2014) بمعالجة مشكلة الازدواج الخطي

باستخدام أسلوب انحدار الحرف وذلك على بيانات في شركة النيل الأزرق للطباعة والتغليف خلال الفترة من (1986 - 2010). أما دراسه Fonti (2017) فقد تناولت شرح طريقه لاسو في خاصيه الاختيار لعدد أقل من المتغيرات المستقله وذلك لوصف المتغير التابع ولجعل النموذج أسهل في التفسير. كما تعرضت دراسة سالم (2018) لمعالجة مشكلة الازدواج الخطي بين المُتغيرات المستقلة عن طريق استخدام كلاً من طريقة المكونات الرئيسية وأسلوب انحدار الحرف وطريقة حذف المتغيرات واختيار أفضل طريقة لعلاج مُشكلة الازدواج الخطي كما استعانت دراسة القصاب، آمين (2018) بأثر الحرف Ridge Trace في تحديد قيمة معلمة الحرف (k) لطريقة انحدار الحرف وذلك للوصول للقيمة المثلى للثابت k للتخلص من مشكلة الازدواج الخطي كما قامت دراسة العبيدي (2019) بمقارنة بين طرق تقدير معلمة انحدار الحرف المعممه مع التطبيق على بيانات مرضى الفشل الكلوي المزمن واقتراح طريقة لمعلمة التحيز وتبين أفضلية الطرق المقترحة كما تناولت دراسه الكفيشي (2019) تقدير معلمات نموذجي انحدار لاسو وانحدار الحرف العادية وانحدار الحرف البيزية ،كما قارنت دراسة حسن واخرون (2020) بين طريقة انحدار الحرف وطريقة المركبات الرئيسية من خلال معيار متوسط مربعات الأخطاء ومنها تبين افضلية انحدار الحرف في علاج مشكلة الازدواج الخطي واخيرًا سعت دراسة الحلواني (2022) إلى استخدام أسلوب انحدار الحرف (RR) وذلك بالمقارنة مع طريقة المربعات الصغرى العادية (OLS) حيث تبين أفضلية طريقة انحدار الحرف. من خلال الدراسات التي تم عرضها تبين كثره استخدام انحدار الحرف وندره استخدام انحدار لاسو Lasso حيث يعد من الموضوعات الحديثه التي لم يتطرق اليها دراسات كثيره.

مُشكلة البحث:

تعتبر مُشكلة الازدواج الخطي إحدى المشاكل القياسية الناجمة عن اسقاط أحد الافتراضات التي يبنى عليها نموذج الانحدار المتعدد وتقوم عليها طريقة المربعات

الصغرى (OLS) وترجع هذه المشكلة لوجود ارتباط خطي تام أو شبه تام بين المتغيرات المستقلة في نموذج الانحدار المُتعدد وتكون عند حدها الأقصى إذا كان هناك ارتباط تام بين المتغيرات المستقلة بحيث يصعب فصل أو عزل تأثير أحدهما عن الآخر علي المتغير التابع وتنعدم في حالة عدم وجود ارتباط بين المتغيرات المستقلة أي تكون المتغيرات مُتعامدة orthongal ووجود مثل هذه المشكلة يضعف من كفاءة مقدرات طريقة المُربعات الصغرى كما ان الابقاء على نموذج انحدار يعاني من مشكلة الازدواج الخطي لا يمكن الاعتماد عليه لان نتائجه غير دقيقة ومضللة.

أهمية البحث:

يَستَمد البحث أهميته من خطورة مُشكلة الازدواج الخطي وتأثيرها على نموذج الانحدار وعلى مقدرات طريقة المُربعات الصغرى كما يستمد أهميته من خطورة الطرق التقليدية في علاج مُشكلة الازدواج الخطي ومنها حذف المُتغيرات الأكثر ارتباطًا وقد يكون منها ما هو ذات أهمية تنبؤية عالية في تفسير الظاهرة محل الدراسة وحذفها يُضعف من قدرة النموذج التفسيرية بالإضافة للوقوع في خطأ توصيف النموذج ومن الطرق التقليدية أيضًا أخذ الفروق الأولى للمتغيرات مما يفقد البيانات أهميتها.

كما يستمد البحث أهميته من خلال أهمية الطُرق المُتبعة لعلاج مُشكلة الازدواج الخطي مثل طريقة انحدار الحرف Ridge Regression (RR) وطريقة المكونات الرئيسية Principal Component Regression (PCR) وطريقه انحدار لاسو Lasso Regression كأحد الطرق الحديثه في علاج مشكله الازدواج الخطي وذلك بالتطبيق علي بيانات خاصة بالهجرة الداخلية في مصر والتوصل الي العديد من النتائج التي تساعد متخذي القرارات في رسم السياسات الخاصة بالسكان.

أهداف البحث: -

يتمثل الهدف الأساسي للبحث هوَ المقارنة بين طرق علاج مشكلة الازدواج الخطي وهي طريقة المُربعات الصُغرى (OLS) وطريقة انحدار الحرف (RR) وطريقة المكونات الرئيسية (PCR) وطريقة انحدار لاسو (Lasso) للوصول الي أفضلها لعلاج مُشكلة الازدواج الخطي ويتم ذلك بالاعتماد علي عده معايير أهمها متوسط مربعات الاخطاء MSE ومعامل التحديد R^2

مُتغيرات وحدود البحث: -

يَتكون نموذج الدراسة من عدة مُتغيرات هيَ: - المُتغير التابع (Y) حجم الهجرة الداخلية في مصر.

المُتغيرات المُستقلة: -

X1 :عدد السكان X2: عدد المواليد X3: عدد الوفيات

X4:الهجره الخارجيه X5:متوسط الدخل السنوي X6: حجم البطاله

بيانات سنوية خلال الفترة من (1996-2022).

مُشكلة الازدواج الخطي

Multicollinearity

تُعد مُشكلة الازدواج الخطي إحدى مشاكل الانحدار الخطي المُتعدد وهي راجعة الي اسقاط أحد الافتراضات التي تقوم عليها طريقة المُربعات الصغرى وهي عدم وجود علاقة خطية بين المُتغيرات المُستقلة وبالتالي عدم صلاحية طريقة المُربعات الصغرى في تقدير معلمات نموذج الانحدار الخطي المُتعدد حيث تنشأ هذه المُشكلة نتيجة وجود

ارتباط تام أو شبة تام بين المتغيرات المستقلة المؤثرة في المتغير التابع في نموذج الانحدار الخطي المتعدد الذي يأخذ الشكل الآتي: -

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + u_i \qquad (1)$$

أي تكون

$$Y = X\beta + U$$

حيث ان: -

Y: متجه المتغير التابع (P×1).

X: مصفوفة المتغيرات المُستقلة (p+1) ×p.

β: متجه معلمات نموذج الانحدار 1 × (p+1).

U: مُتجه الأخطاء العشوائية 1 × p.

وَبتقدير معلمات نموذج الانحدار المُتعدد باستخدام طريقة المُربعات الصغرى كالتالي:

$$\hat{\beta} = (\grave{x}x)^{-1}\,\grave{x}y \qquad (2)$$

حيث ان $\hat{\beta}$ مُتجه مُقدرات نموذج الانحدار باستخدام طريقة المربعات الصغرى.

أولاً: طُرق اكتشاف مُشكلة الازدواج الخطي: -

هُناك طُرق عديدة لاكتشاف مُشكلة الازدواج الخطي أهمها: -

١- اختبار (Farrar& Glauber; 1967):-

يعتمد هذا الاختبار على توزيع χ^2 حيث تنص الفروض الإحصائية لهُ على عدم وجود مشكلة الازدواج الخطي كفرض عدمي ووجود مشكلة الازدواج الخطي كفرض بديل كما انَ إحصائية الاختبار الخاصة بهِ هيَ: -

$$\chi^2 = -\left[(n-1) - \frac{1}{6}\,(2k+5)\right]\ln|R| \qquad (3)$$

حيث ان: -

n: حجم العينة أو عدد المُشاهدات.

k: عدد المُتغيرات المُستقلة.

$\ln|R|$: اللوغاريتم الطبيعي لقيمة مُحدد مصفوفة مٌعاملات الارتباط بين المُتغيرات المُستقلة.

القيمة الجدولية للاختبار: -

$$\chi^2 \left(\frac{1}{2}k(k-1),\propto\right)$$

فإذا كانت إحصائية الاختبار أكبر من القيمة الجدولية يتم رفض الفرض العدم وبالتالي هُناك مُشكلة الازدواج الخطي والعكس صحيح.

٢- مُعامل تضخم التباين Variance inflation Factor (VIF)

يستخدم معامل تضخم التباين في الكشف عن وجود مشكلة الازدواج الخطي وكذلك تحديد المتغير المستقل المُتسبب في وجود هذه المُشكلة ويتم حساب معامل تضخم التباين كالتالي:

$$VIF = \frac{1}{1-R_j^2} \qquad j = 1, 2, \quad \dots\dots k \qquad (4)$$

حيث ان R_j^2 هوَ مُعامل التحديد الناتج عن انحدار المُتغير المُستقل (x_j) على باقي المُتغيرات المُستقلة في الانحدار $x_j = f\ (x_1, x_2, \ \dots\dots\ x_k)$

وإذا كانتَ قيمة VIF > 10 فذلك دليل على وجود مُشكلة الازدواج الخطي بين المتغير x_j وباقي المُتغيرات وبالتالي ينصح بحذفه لأنه هو السبب في وجود مُشكلة الازدواج الخطي أما إذا كان VIF < 10 لا يوجد مُشكلة ازدواج خطي.

٣- المؤشر الشرطي Condition index (CI)

ويتم حساب المؤشر الشرطي (CI) بالاعتماد على الجذور المميزة (eigen values) لمصفوفة($\grave{x}x$)

كالتالي:

$$C\,I = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \qquad (5)$$

حيث ان:-

λ_{max} : أكبرجذر مميز للمصفوفة $\grave{x}x$

λ_{min} : أصغر جذر مميز للمصفوفة $\grave{x}x$

فإذا كانت:

$CI > 100$ دل ذلك على وجود ازدواج خطي حاد.

$30 < CI < 100$ دل ذلك على وجود ازدواج خطي عالي.

$CI < 30$ لا يوجد ازدواج خطي.

ثانيًا: طرق معالجة مُشكلة الازدواج الخطي:

تعددت طرق معالجة مشكلة الازدواج الخطي ومنها طريقة انحدار الحرف التي قدمها (Horel& Kennard; 1970;2000) وهي طريقة بديلة لطريقة المُربعات الصغرى العادية *(OLS)* لتقدير معلمات نموذج الانحدار المتعدد في ظل وجود مُشكلة الازدواج الخطي وذلك لما تعانيه طريقة المربعات الصغرى من تضخم في الأخطاء المعيارية للمقدرات وتعتبر من أكثر الطرق شيوعًا في حالة ظهور مشكلة الازدواج الخطي التام وشبه التام وطريقة المكونات الرئيسية وطريقة انحدار لاسو.

١ـ انحدار الحرف Ridge Regression (RR)

احدي الطرق المتخصصه في تحليل الانحدار المتعدد عندما يعاني من مشكلة الازدواج الخطي Multicollinearity وقد أظهرت هذه الطريقة نتائج فعاله في التخلص من مشكلة الازدواج الخطي حيث ان وجود هذه المشكله يؤدي لكبر تباين المعلمات والحصول علي نتائج مضلله عند استخدام طريقة المربعات الصغري وتقوم فكرة انحدار الحرف علي ايجاد قيمة الثابت "k" والمسمي بمعلمه التحيز وهي كميه موجبه تضاف الي عناصر القطر في المصفوفة $(\acute{x}x)$ تؤدي لتخفيض تباين المعلمات المقدره حيث ان أضافة الثابت (k) بقيم صغيره تعمل علي حدوث تغير سريع في قيم المعلمات المقدره ومع زياده قيمة k تبدأ تلك القيم في الاستقرار تدريجيا الي ان تصل الي حد يكون التغيير ضئيل جدا وتقوم طريقه انحدار الحرف بتقدير المعلمات بحيث تقلل مجموع مربعات الخطأ (SSE) كالتالي :

$$\widehat{\mathcal{B}} = \min \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{i\,j})^2 \qquad (6)$$

وتكون معادله التقلص اوالانكماش Shrinkage كالتالـــــي:

$$\widehat{\mathcal{B}} = \min \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{i\,j})^2 + k \sum_{j=1}^{p} \beta_j^2 \qquad (7)$$

وتتكون المعادلة من جزئين الاول مجموع مربعات الاخطاء (SSE) والثاني دالـه الجزاء كالتالي :

$$\widehat{\mathcal{B}} = SSE + k \sum_{j=1}^{p} \beta_j^2$$

وبتدنيه مجموع مربعات الاخطاء نحصل علي مقدرات انحدار الحرف كالتالـــــــــي:

$$\widehat{\mathcal{B}}_{RR} = \left[\grave{x}x + \grave{}K\, I_p\ \right]^{-1}(x'y) \qquad (8)$$

$\mathcal{B}_{RR}$: متجه المعلمات المقدره من انحدار الحرف.

K: معلمه التحيز bias parameter

ولايجاد العلاقة بين مقدرات طريقة انحدار الحرف والمربعات الصغري (OLS) كالتالي:

$$\hat{\beta}_{RR} = [(\grave{x}x) + K\,(\grave{x}x)(\acute{x}x)^{-1}]^{-1}(x'y)$$

$$= (\acute{x}x)^{-1}[\mathrm{I_P} + K\,(\acute{x}x)^{-1}]^{-1}(x'y)$$

$$= [\mathrm{I_P} + K\,\mathrm{I_P}\,(\acute{x}x)^{-1}]^{-1}(\acute{x}x)^{-1}(x'y)$$

$$\widehat{\mathcal{B}}_{RR} = Z_{RR}\hat{\beta}_{ols} \qquad (9)$$

حيث ان :ـــ

$$Z_{RR} = [\mathrm{I_P} + K\,\mathrm{I_P}\,(\acute{x}x)^{-1}]^{-1}$$

وبالتالي فإن مقدرات طريقة انحدار الحرف هي تحويل خطي لمقدرات المربعات الصغري حيث ان:ـــ

$$MSE_R = Variance\,(\hat{\beta}_R) + \,(bias\ in\ \hat{\beta}_R)^2$$

كلما زادت قيمة K فان مقدار التحيز يزيد والتباين يقل وبالتالي يجب اختيار k بحيث يكون الانخفاض في قيم التباين أكثر من الارتفاع في مقدار مربع التحيز عند ذلك فإن متوسط مربعات الخطأ لانحدار الحرف يكون أقل من تباين مقدرات المربعات الصغري.

كما ان زيادة k تقلل من قيمة معامل التحديد (R^2) وبالتالي فان مقدرات انحدار الحرف ليس من الضروري ان تعطي أفضل نموذج ملائم للبيانات بل تبحث عن أفضل معادلة ذات مقدرات ثابتة (غير متحيزه مع زيادة k) ALKhamisi (2007).

٢ـ طريقة المكونات الرئيسية Principal component Regression (PCR)

قدمَ طريقة المكونات الرئيسية في شكلها النهائى Hotellingعام١٩٣٣ وتعد طريقة مكونات الرئيسية واحدة من النماذج الخطية المتحيزة الواسعة الاستخدام لتخطي مشكلة الازدواج الخطي التي يعاني منها نموذج الانحدار الخطي المتعدد وتقوم طريقة المكونات الرئيسية على فكرة تحويل المتغيرات التفسيرية الاصلية المرتبطة دون حذف أي منها إلى متغيرات جديدة متعامدة ومستقلة تسمى مكونات رئيسية وكل مكون أو مركب رئيسي عبارة عن توليفة خطية في المتغيرات المستقلة الاصلية ويتم ترتيب المكونات حسب أهميتها من خلال مساهمتها في مقدار التباين الكلي فيها تهدف إلى استنتاج مجموعة من المغيرات الجديدة تكون أقل عدًا من المتغيرات الأصلية.

حيث تستخدم هذه المتغيرات الجديدة في دراسة وتحليل هيكل العلاقة بين المتغيرات الأصلية في الدراسة مع العلم ان فكرة المكونات الرئيسية هي الحصول على متغيرات جديدة ومستقلة عن بعضها البعض.

نموذج المكونات الأساسية: (Johanson et al; 1998)

ويتمثل الأساس الرياضي لطريقة المكونات الرئيسية في إذا كان هُناك عدد p من المتغيرات المستقلة ويراد تخفيض عددها فحينئذ تسمى مكونات وحتى تكون هذه المكونات معنوية وذات فائدة أي تحتوي على نفس القدر من المعلومات التي يتم الحصول عليها فيما لو استخدمنا كل المتغيرات الأصلية فلابد ان تحتوي هذه المكونات على قدر كافي (n) من البيانات وذلك للحصول على الجزء الأكبر من التباين الكلي: - فإذا كان المتجه العشوائي للمتغيرات الأصلية

$\acute{x} = [x_1, x_2, \dots x_p]$ ومصفوفة التباين والتغاير $\sum$ وجذورها المميزة eigenvalues هي

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$$

فإذا كانت المجموعات أو المكونات الخطية Linear Combinations كالتالي:

$$y_1 = \grave{L}_1\, x = L_{11}\, x_1 + L_{21}\, x_2 + \cdots + L_{p1}\, x_p$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$y_p = \grave{L}_p\, x = L_{1p}\, x_1 + L_{2p}\, x_2 + \cdots + L_{pp}\, x_p$$

حيث إن L تمثل أوزان أو مُعاملات يتم اختيارها لتعظيم نسبة التباين التي يشرحها المكون (y) من إجمالي التباين كما يتم اختيارها أيضًا لتحقيق استقلال المكونات مع بعضها على ان يكون مجموع مربعاتها مساويًا للواحد الصحيح.

هذا ومن المعلوم إنه في حالة وجود q من التوليفات الخطية في p في المتغيرات الأصلية وتم ضرب المتغيرات في ثابت C.

أي كان:

$$Z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_q \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ & & & \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = CX$$

$$Z = CX$$

وبالتالي متجه المتوسطات μ_2 ومصفوفة التباين والتغاير $\sum_z$ للتوليفة Z

$$E\,(Z) = C\,E\,(x) = C\mu_x$$

$$V\,(Z) = C \textstyle\sum_x \grave{C}$$

حيث ان μ_x، $\sum_x$ متجه المتوسطات ومصفوفة التباينات والتغايرات للمتجه العشوائي x.

وعليه فإن تباين المكون (i) يكون

$$Var\,(y_i) = \grave{l}_i \sum l_i \qquad i = 1, 2 \quad \ldots p$$

وَتغاير المكونين i، k يكون

$$Cov\ (y_i, y_k) = \grave{l}_i \sum l_k$$

وبالتالي فالمكونات الرئيسية ما هي إلا مجموعات خطية من المُتغيرات المستقلة الأصلية ويشترط استقلالها عن بعضها البعض وإن يكون تباين كل مكون منها أكبر من الذي يليه قدر الإمكان *(Manly; 1994)*.

$$var\ (y_1) \geq var\ (y_2)\ \geq var\ (y_3) \ldots\ \geq var\ (y_p)$$

ومن ثم يمكن تعريف المكونات الرئيسية كالتالي:

المكون الأساسي الأول:

هو المجموعة الخطية $\grave{l}_1 x$ صاحب أكبر تباين لـ $var\ (\grave{l}_x)$ بشرط $\grave{l} = i$

المكون الأساسي الثاني:

صاحب المجموعة الخطية $\grave{l}_2\ x$ صاحب أكبر تباين بعد المكون الأول

$$\grave{l}_2\ l_2 = i \qquad , var\ (\grave{l}_1 x) > var\ \ (\grave{l}_2 x)$$

$$Cov\ (\grave{l}_1 x, \grave{l}_2 x) = 0$$

وهكذا حتى الحصول على المكون الأخير.

وبناءً على ما سبق إذا كان متجه المتغيرات الأصلية $\grave{x} = [x_1, x_2 \ldots x_p]$ له مصفوفة التغاير $\sum$ وجذورها ومتجهاتها المميزة هي:

$$(\lambda_1, e_1), (\lambda_2, e_2) \ldots (\lambda_p, e_p)$$

حيث ان:

$$\lambda_1\ \geq\ \lambda_2\ \geq \cdots \geq\ \lambda_p\ \geq 0$$

ومع افتراض ان المكونات الرئيسية: -

$$y_1 = \grave{e}_1 x \qquad , y_2 = \grave{e}_2 x, \ldots, y_p = \grave{e}_p x$$

$$\therefore\ t_r\ (\sum) = \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} = \sum_{i=1}^{p} var(x_i)$$

$$= \lambda_1 + \lambda_2 + \cdots \lambda_p$$

$$= \sum_{i=1}^{p} var\ (y_i)$$

وبناءً عليه فإنه يمكن حساب نسبة التباين لكل مكون أساسي (K) كما يلي:

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 \dots + \lambda_p} = \frac{\lambda_k}{tr(\Sigma)} \quad k = 1, 2, \dots p$$

فإذا كانت مصفوفة الارتباط هي الخاضعة للدراسة فإن المقام عدد المتغيرات المستقلة (p)

$$\frac{\lambda_k}{tr\,(R)} = \frac{\lambda_k}{p} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots \lambda_3}$$

٣- انحدار لاسو LASSO Regression

يعد انحدار لاسو Lasso اختصار لـــ Least Absolute shrinkage and selection operator والمقترح من قبل العالم (Tibshirani,1996) وهو احد أنواع الانحدار الخطي وصمم اساسا لنماذج المربعات الصغري Least squares Models ويقوم علي فكرة الانكماش "التقلص" shrinkage لمعلمات نموذج الانحدار ويعد من أفضل النماذج ملائمة في حالة تعرض نموذج الانحدار الخطي الي درجة عالية من الازدواج الخطي Multicollineanty حيث انه يعمل علي تصغير مجموع مربعات الاخطاء وذلك بناء علي قيد او شرط معين يتم فرضه علي المعلمات والذي يمثل المجموع المطلق للمعلمات ومن خلال هذا القيد فإن مقدر Lasso يعمل علي جعل عدد من المعلمات لنموذج الانحدار مساوية للصفر وتقليص او انكماش باقي المعلمات بمقدار معين حيث انه بالاضافة الي تقدير معلمات النموذج يقوم باختيار وتنظيم المتغيرات الداخله في النموذج وذلك لزيادة القدرة التفسيرية لنموذج الانحدار المستخدم في تحليل الظاهره محل الدراسة وتعتبر طريقة انحدار لاسو امتداد لكلا من طريقة الاختبار المتدرج Stepwise selection والتي تعمل علي تحسين دقة النموذج في حالات معينة وخاصة عندما يكون هناك ارتباط قوي بين المتغير التابع والمتغيرات المستقلة والذي يجعل التنبؤ أقل دقة. بالاضافة الي طريقة انحدار الحرف Ridge Regression والذي يشيع استخدامها لتحسين دقة التنبؤ لنموذج الانحدار وبالتالي تقليص وانكماش معاملات الانحدار الكبيرة ولكنه لا يقوم بمهمة اختيار المتغيرات الملائمة للنموذج وبالتالي فإن مقدر لاسو يقوم بالجمع بين كلا من مهام الاختيار المشترك والانكماش لمعلمات نموذج الانحدار وذلك من خلال جعل مجموع القيم المطلقة لمعلمات الانحدار أقل من قيمة ثابتة مما يجبر بعض معاملات الانحدار لتساوي الصفر مع اختيار نموذج بسيط يستبعد فيه هذه المتغيرات التي انكمشت معاملاتها للصفر وبالتالي فإن مقدر لاسو يجمع بين

مهام مقدر انحدار الحرف ومقدر اختيار مجموعه من المتغيرات. (Wang et al ; (2007

ويهدف انحدار Lasso الحصول علي أقل مجموع مربعات الخطأ وذلك من خلال وضع قيد أو شرط معين وهو ان المجموع المطلق لمعلمات الانحدار تكون أقل من ثابت معين وبذلك تنكمش او تتقلص بعض معاملات نموذج الانحدار للصفر وبالتالي تنقسم المتغيرات في النموذج الي نوعين احدهما انكمش معامله للصفر وتم استبعاده من النموذج وآخر انكمش معامله بقدر ما ويتم ادراجه ضمن النموذج وبالتالي تحقيق الهدف وهو تقليل خطأ التنبؤ وفي طريقة Lasso هناك معلمه تسمي معلمه ضبط وهي التي تتحكم في معلمات نموذج الانحدار فعندما تكون المعلمة كبيره تضطر بعض المعاملات الي ان تكون مساوية للصفر وعندما تكون معلمه الضبط مساوية للصفر سنحصل علي مقدرات طريقة المربعات الصغري العادية (OLS) ويتميز انحدار لاسو في حالة وجود عدد من المشاهدات أقل وعدد كبير من المتغيرات حيث ان تقليص او ازاله المتغيرات تخفض التباين دون الزيادة الكبيرة في درجة التحيز كما أنه يسهل عملية تفسير النموذج من خلال استبعاد المتغيرات غير المرتبطة بالمتغير التابع.

نموذج انحدار لاســو LASSO Regression

تقدر معلمات انحدار لاسو طبقا للمربعات الصغري (OLS) كالتالي:

اذا كان هناك عينة حجمها n من الحالات كل حاله تحتوي علي "k" من المتغيرات المستقله ومتغير تابع واحد "y_i" وليكن "X" متجه المتغيرات المستقله للحالــة j^{th} فيكون هدف انحدار لاسو هو الوصول لحل المعادله الاتية:

$$min\left[\frac{1}{n}\sum_{i=1}^{n}(y_i-\beta_0-\grave{x}_i\beta)^2\right] \qquad (10)$$

$$Subject\ to\ \sum_{j=1}^{p}\left|\beta_j\right| \le t$$

حيث ان:

t: معلمه تحدد مسبقا وتمثل مقدار الانكماش (التقلص).

$\grave{x}$: مصفوفة المتغيرات المستقلة.

X_i: هي الصف j^{th} من المصفوفه X.

ويمكن صياغه لاسو كالتالي :

$$\min_{\beta_0,\beta}\left[\frac{1}{n}\|y-\ \beta 0\ I_n-x\ \beta\|\ \frac{2}{2}\right] \qquad (11)$$

$$Subject\ to\ \|\beta\|\leq t$$

حيث ان

$\|\boldsymbol{\beta}\|_p=(\sum_{i=1}^{n}|\boldsymbol{\beta}_i|^p)^{\frac{1}{p}}$ وعندما يصبح $\mathbf{p}=\mathbf{1}$ يصبح $\|\boldsymbol{\beta}\|_1$ الطول القياسي L^p

وبتحويل المتغيرات الي قيم معيارية نحصل علي الشكل الاتـــــــي :

$$\min_{\beta 0,\beta}\left[\frac{1}{n}\|y-\ x\ \beta\|2^2\right]$$

$$Subject\ to\ \|\beta\|\leq t$$

ويكون بصيغه مضاعف لاجرانج كالتالـــــي:

$$min_{B\in R^p}\left[\frac{1}{n}\ \|y-x\beta\|\ \frac{2}{2}+\ \lambda\|\beta\|_1\right] \qquad (12)$$

حيث ان λ معلمه الضبط والتحكم والانكماش لمعلمات نموذج الانحدار .

الجانب التطبيقي

استهدف البحث المفاضلة بين طريقة المربعات الصغرى (OLS) وطريقة انحدار الحرف (RR) وطريقة المكونات الرئيسية (PCR) وطريقه انحدار لاسو (Lasso) وذلك للوصول للطريقة الأفضل في علاج مشكلة الازدواج الخطي وذلك بالتطبيق على بيانات الهجرة الداخلية في مصر خلال الفترة من (1996-2022) بيانات سنوية وحيث ان المتغير التابع (y) حجم الهجرة الداخلية في مصر ومجموعة من المتغيرات المستقلة هي (x_1) عدد السكان، (x_2) عدد المواليد، (x_3) عدد الوفيات، (x_4) الهجرة الخارجية، (x_5) متوسط الدخل السنوي للفرد، (x_6) معدل البطالة

أولاً: تقدير نموذج الانحدار باستخدام طريقة المربعات الصغرى: -

وبتقدير نموذج الانحدار المتعدد باستخدام طريقة المربعات الصغرى تم الحصول على جدول (١) التالي: -

جدول (١) المعلمات والأخطاء المعيارية المقدرة باستخدام (OLS)

Model	B	St- Error	Sig	R^2	R
Constant	-78954.782	319345.125	0.807		
x_1	-0.0000138	000	0.185		
x_2	944.091	177.979	000	0.982	0.99
x_3	-11304.064	2858.001	-001		
x_4	-1357.539	597.977	0.0.034		
x_5	10488.321	3014.952	0.002		
x_6	1374.782	630.518	0.041		

وَمِن المُلاحظ من نتائج تقدير مَعلمات نموذج الانحدار باستخدام طريقة المربعات الصغرى ظهور بعض المتغيرات بإشارات تخالف الواقع وأخطاء معيارية عالية مع معنوية غالبية المتغيرات حيث ان Sig أقل من 0.05 لمعظم المتغيرات وظهور مُعامل التحديد بقيمة عالية $R^2 = 98.2\%$ وكل ذلك مؤشرات هامة وخطيرة عن وجود مشكلة الازدواج الخطي بين المتغيرات المستقلة لنموذج الانحدار المتعدد.

جدول (2) جدول تحليل التباين

Model	Sum of Squares	df	Mean Square	F	p-value
Regression	4101841837386.554	6	683640306231.092	89.691	.000b
Residual	152443149008.113	20	7622157450.406		
Total	4254284986394.667	26			

من جدول (2) نلاحظ ان النموذج ذو دلالة إحصائية حيث ان قيم p-value أقل من 0.05 وبالتالي رفض الفرض العدمي وقبول الفرض البديل بمعنوية نموذج الانحدار المُتعدد

ويمكن صياغة نموذج الانحدار المتعدد باستخدام طريقة المربعات الصغرى(OLS) كالتالي:

$$y = -78954.782 - 0.0000138\, x_1 + 944.091\, x_2$$

$$- 11304.064\, x_3 - 1357.539 x_4 + 10488.32 x_5 + 1374.782\, x_6$$

$$R^2 = 98.2\% \qquad , \qquad MSE = 7622157450$$

ثانيًا الكشف عن مُشكلة الازدواج الخطي Multicollinearity: -

هناك عدة اختبارات لاكتشاف وجود مشكلة الازدواج الخطي أهمها:

١- اختبار Farrar& Glauber : -

وينص الفرض العدمي له على عدم وجود مشكلة الازدواج الخطي والفرض البديل على وجود مشكلة الازدواج الخطي أما إحصائية الاختبار: -

$$\chi^2 = -\left[(n-1) - \frac{1}{6}(2k+5)\right]\ln|R|$$

$$n = 27, k = 6 \ , \ln|R| = -12.518$$

$$\chi^2 = -\left[(0.27-1) - \frac{1}{6}(2\times 6+5)\right] * -12.518$$

$$= 290$$

أما القيمة الجدولية:

$$x^2\left(\frac{1}{2}k\ (k-1), \propto\right)$$

$$x^2\ (15, 0.05) = 24.996$$

وحيث ان إحصائية الاختبار أكبر من القيمة الجدولية وبالتالي يتم رفض الفرض العدمي وقبول الفرض البديل بوجود مشكلة الازدواج الخطي بين المتغيرات المستقلة.

١. حساب قيمة مُعامل تضخم التباين (VIF) للمتغيرات المستقلة كانت كالتالي: -

جدول (3) يتم قيم معامل تضخم التباين VIF

	VIF	Torelence
x_1	1.386	0.721
x_2	32.683	0.031
x_3	179.221	0.006
x_4	54.763	0.018
x_5	167.418	0.006
x_6	59.823	0.017

من خلال جدول (3) نلاحظ ان هناك متغيرات مستقلة قيمة VIF لها أكبر من 10 وبالتالي هناك ازدواج خطي بين المتغيرات المستقلة.

٢. حساب المؤشر القياسي (CI):

يتم حساب قيمة CI بالاعتماد على قيم الجذور المميزة eigen values لمصفوفة $(\grave{x}x)$ وتم الحصول عليها كالتالي:

جدول (4) قيم المؤشر القياسي

	Eigen values	Condition index (CI)
λ_1	5.841	1
λ_2	0.921	2.520
λ_3	0.2180	5.174
λ_4	0.020	17.198
λ_5	0.001	63.822
λ_6	0.001	96.555

نلاحظ من خلال قيم CI في جدول (4) ان هناك قيم بين 30 و 100 وذلك مؤشر قوي على وجود مشكلة الازدواج الخطي بين المتغيرات المستقلة في النموذج.

ثالثًا: طرق علاج مشكلة الازدواج الخطي:

تعددت طرق علاج مشكلة الازدواج الخطي التي استخدمها كثير من الباحثين لإيجاد الحلول المناسبة لعلاج المشكلة سواء كان بحذف المتغيرات ذات الارتباط العالي والمعنوي أو بزيادة حجم العينة وذلك على أساس ان مشكلة الازدواج الخطي هي مشكلة عينة وليس مجتمع (عطية ، 1998) أو استخدام طرق أخرى ولكنها متحيزه ومنها:

١ـ تقدير معلمات النموذج باستخدام طريقة انحدار الحرف

Ridge Regression (RR)

بناء علي ما توصلنا اليه من نتائج تقدير معلمات النموذج باستخدام طريقة المربعات الصغري (OLS) والتي تعتبرها مضلله من الناحيه الاحصائيه وذلك لتضخم الاخطاء المعياريه للمعلمات المقدره وظهور بعض المتغيرات باشارات مخالفه للواقع سوف نستخدم طريقه انحدار الحرف لتقدير معلمات النموذج وبالاعتماد علي أسلوب أثر الحرف ridge Trace لتحديد قيمه معلمه الحرف او التحيز (K) المثلي والتي ينحصر قيمتها بين الصفر والواحد وذلك بمقدار زياده قدره 0.05 للوصول للنموذج الاكثر استقرارا وصاحب اقل متوسط مربعات للخطأ وذلك باستخدام برنامج Stat graphics 18 , MAT Lab تم التوصل للقيمه المثلي لـــــــ K عند 0.30 وبالتالي فالنموذج المقدر كالتالي:

جدول (5) تقديرات معلمات نموذج الانحدار المقدر بطريقة انحدار الحرف عند K= 0.30

Variable	β	S.E	P.value	VIF
X_1	1.8E-10	0.2131	0.1530	0.6178
X_2	0.0369	0.0460	0.000	0.258
X_3	- 0.2773	0.6192	000	0.146
X_4	0.01073	0.0765	0.008	0.259
X_5	0.3198	0.0787	0.016	0.164
X_6	0.7236	0.4465	0.0461	0.2525

ومن جدول (5) نلاحظ ان المتغيرات (X_2) عدد المواليد، (X_3) عدد الوفيات، (X_4) حجم الهجره الخارجيه، (X_5) متوسط دخل الفرد، (X_6) حجم البطاله. متغيرات ذات دلاله احصائيه عاليه حيث ان P-Value اقل من مستوي المعنوية 0.05 وان المتغير (X_1) عدد السكان غير دال احصائيا وبالتالي فان معادله الانحدار المقدره بطريقه انحدار الحرف وعند معلمه (K) المثلي وهي 0.30 فالنموذج كالتالي :

$$\hat{y} = 1.8E - 10x_1 + 0.0369x_2 - 0.2773x_3 + 0.01073x_4 + 0.3198x_5 + 0.07236x_6$$

ولاختبار معنويه نموذج الانحدار المقدر باستخدام طريقة انحدار الحرف تصاغ الفروض الاحصائيه كالتالي:

النموذج غير معنوي $H_o : B_i = 0$

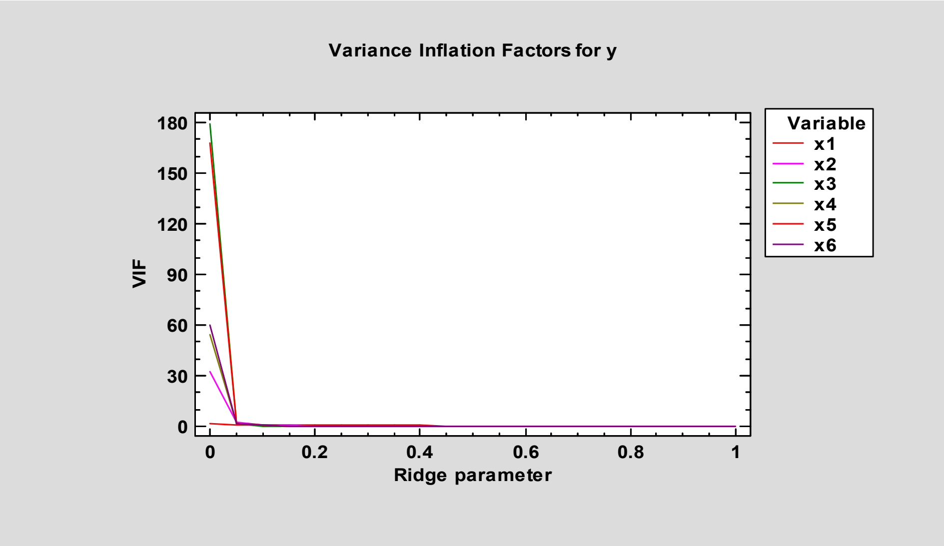
النموذج معنوي $H_1 : B_i \neq 0$

ومن خلال جدول تحليل التباين ANOVA:

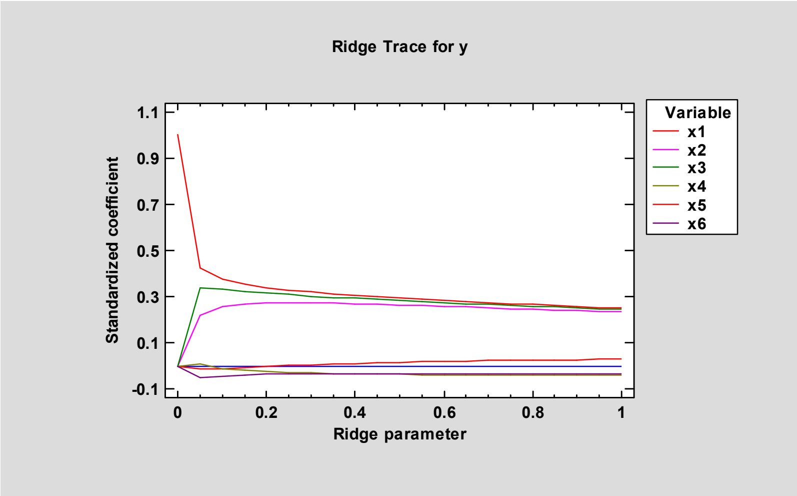
Model	SS	df	MS	F	p-value
Regression	21648.96	6	3608.16	34.72	000
Error	2078.3	20	103.915		
Total	23727.26	26			

حيث ان p-value اقل من 0.05 وبالتالي فالنموج المقدر بطريقة انحدار الحرف ذو دلاله احصائيه وذلك بمعامل تحديد $R^2 = 90.4\%$ ومتوسط مربعات اخطاء MSE = 103.915 وقد تم الوصول للقيمة المثلي لمعلمه التحيز في نموذج انحدار الحرف (K) باستخدام برنامج Stat graphics 18 ومن خلال شكل Ridge Trace والذي يوضح العلاقه بين قيم المعلمات المعياريه للمتغيرات المتوقع تأثيرها علي المتغير التابع والقيم المختلفة لمعلمه الحرف (K=0.30) وذلك لانها اقل قيمة تبدأ بعدها المعلمات المعيارية في الاستقرار وقد تم اختيار هذه القيمة بعد اجراء عده مقارنات بينها وبين القيم السابقه واللاحقه لها وكذلك في شكل (2) تظهر العلاقة بين القيم المختلفه لمعلمه الحرف ومعامل تضخم التباين (VIF) ويتضح من الشكل ان القيمة المثلي لمعلمه الحرف K=0.30 وذلك لان بعدها تبدأ عوامل تضخم التباين في الانخفاض ببطء شديد حيث ان قيم VIF اقل من (10) وبالتالي ظهور بعض المتغيرات باشارات مخالفه ومعلمات عاليه وغير معنويه دليل علي ان طريقه انحدار الحرف لم تتخلص من مشكله الازدواج الخطي بشكل كامل.

شكل (1)



شكل (2)



2ـ طريقة المكونات الرئيسية:

للتأكد من وجود مشكلة الازدواج الخطي بين المتغيرات المستقلة تم فحص مصفوفة الارتباط البسيط بين المتغيرات المستقلة التالية:

$$\begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{array} \begin{array}{c} \begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \end{array} \\ \begin{bmatrix} 1 & 0.341 & 0.322 & 0.06 & 0.28 & 0.136 \\ 0.341 & 1 & 0.981 & -0.16 & 0.972 & -0.095 \\ 0.322 & 0.981 & 1 & -0.239 & 0.994 & -0.0177 \\ 0.06 & -0.16 & -0.239 & 1 & -0.28 & 0.985 \\ 0.28 & 0.972 & 0.994 & -0.28 & 1 & -0.228 \\ 0.136 & -0.095 & 0.0177 & 0.985 & -0.228 & 1 \end{bmatrix} \end{array}$$

وَمِن خلال مصفوفة الارتباط نجد إن هناك ارتباط قوي ومعنوي بين (x_2, x_3) وبين (x_2, x_5) وبين (x_3, x_5)و(x_4, x_6) وذلك مؤشر قوي لوجود مشكلة الازدواج الخطي بين المُتغيرات المُستقلة في النموذج كما يمكن استخدام اختبار Kaiser وBartlett's لاختبار كفاية العينة كالتالي:

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.599
Bartlett's Test of Sphericity	Approx. Chi-Square	290.029
	df	15
	Sig.	.000

نلاحظ إن إحصائية Kaiser تجاوزت0.5 وهو الحد الأدنى لكفاية العينة ولاختبار ان مصفوفة الارتباط هي مصفوفة الوحدة أم لا نجد إن Sig أقل من ٠,٠٥ وبالتالي نرفض الفرض العدمي وبالتالي فان مصفوفة الارتباط لا تمثل مصفوفة الوحدة كما يمكن استخدام الجذور المميزة y اكتشاف وجود مشكلة الازدواج الخطي حيث ان:

$$\lambda_1 = 3.257 \quad \lambda_3 = 0.802 \quad \lambda_5 = 0.12$$
$$\lambda_2 = 1.903 \quad \lambda_4 = 0.023 \quad \lambda_6 = 0.003$$

نُلاحظ إن هُناك جذور مميزة قيمتها قريبة جدًا من الصفر وذلك مؤشر على وجود مُشكلة الازدواج الخطي.

كما يمكن حساب المؤشر القياسي (CI) كالتالي :

$$CI = \sqrt{\frac{\lambda_{max}}{\lambda_{main}}} = \sqrt{\frac{3.257}{0.003}} = 32.94$$

حيث ان $CI > 30$ مؤشر يدل على ان هناك احتمالية عالية لوجود مُشكلة الازدواج الخطي

المكونات الرئيسية الناتجة:

حيث ان هناك جذران مميزان قيمتهما اكبر من الواحد الصحيح وبالتالي فإن هُناك مكونان رئيسيان هُما:

المكونات الرئيسية

	Component	
	1	2
x2	.980	-.078
x3	.975	-.160
x5	.960	-.213
x1	.490	.257
x6	-.037	.989
x4	-.107	.977

مصفوفة الارتباط للمكونات الرئيسية

Component	1	2
1	.934	-.357
2	.357	.934

يُمكن كتابة المكونات الرئيسية كالتالي:

$$P_{c1} = 0.490\, x_1 + 0.980\, x_2 + 0.975\, x_3 - 0.107 x_4 + 0.960\, x_5 - 0.037 x_6$$

$$P_{c2} = 0.257\, x_1 - 0.078\, x_2 - 0.160\, x_3 - 0.977 x_4 + 0.213\, x_5 + 0.989 x_6$$

ويتم تقدير نموذج انحدار المكونات الرئيسية بطريقة المُربعات الصغرى كالتالي:

Model	Unstandardized Coefficients		Collinearity Statistics	
	B	Std. Error	Tolerance	VIF
1 (Constant)	1735704.111	29748.097		
P_{c1}	371236.224	30314.779	1.000	1.000
P_{c2}	-61272.773	30314.779	1.000	1.000

$$\hat{y} = 1733704.111 + 371236.22\, P_{c1} - 61272.773\, P_{c2}$$

ونلاحظ خلو النموذج من مشكلة الازدواج الخطي حيث ان قيمة VIF = 1 وهي أقل من ١٠ وبالتالي قامت طريقة المكونات الرئيسية بالمُهمة المنوطة لها وهي التخلص من مُشكلة الازدواج الخطي.

3- تقدير معلمات النموذج بطريقه انحدار لاسو Lasso Regression

تقوم طريقه انحدار لاسو علي نفس فكره طريقه انحدار الحرف تعمل علي تقليص او انكماش المعلمات للنموذج ولا توصلها للصفر ليأتي مقدر انحدار لاسو بانكماش اعلي لتصل الي الصفر ويتم استبعادها من نموذج الانحدار المقدر وهي المتغيرات الغير معنويه احصائيا مع المتغير التابع ليبقي في النموذج المتغيرات الاكثر معنويه والتي تزيد القدره التفسيريه للنموذج المقدر.

جدول (6) تقديرات معلمات النموذج بطريقه انحدار لاسو $\lambda = 0.30$

Variable	β	S.E	P.value
X_1	0	—	انكماش (تقلص)
X_2	0.0326	0.042	0.000
X_3	-0.2637	0.5961	0.002
X_4	0.0106	0.0732	0.0162
X_5	0.2986	0.0695	0.0156
X_6	0	—	انكماش (تقلص)

ومن خلال جدول (6) تبين ان كلا من المتغيرات (X_2) عدد المواليد، (X_3) عدد الوفيات، (X_4) حجم الهجره الخارجيه، (X_5) متوسط دخل الفرد. متغيرات ذات دلاله احصائيه حيث ان P-Value اقل من مستوي المعنوية 0.05. وبالتالي تدرج ضمن النموذج المقدر اما كلا من (X_1) عدد السكان ، (X_6) حجم البطاله قد تقلصت معلماتهم للصفر وبالتالي يتم استبعادها من النموذج المقدر حيث انها ليست ذات دلاله احصائيه مع المتغير التابع. وبالتالي فالنموذج الانحدار المقدر بطريقة انحدار لاسو كالتالي:

$$\hat{y} = 0.0326x_2 - 0.2637x_3 + 0.0106x_4 + 0.2773x_3 + 0.2986x_5$$

ولاختبار معنويه النموذج المقدر بطريقه انحدار لاسو تصاغ الفروض الاحصائيه كالتالي:

$H_o : B_i = 0$

$H_1 : B_i \neq 0$

ومن خلال جدول تحليل التباين المقدر بطريقه انحدار لاسو

Model	SS	df	MS	F	p-value
Regression	54300.34133	4	13575.085	177.88	000
Error	1679.392	22	76.336		
Total	55979.733	26			

ومن جدول تحليل التباين نلاحظ ان قيمة p-value اقل من 0.05. وبالتالي فالنموذج المقدر بطريقه انحدار لاسو ذو دلاله احصائيه وذلك بمعامل تحديد قدره $R^2 = 97\%$ ومتوسط مربعات الاخطاء MSE=76.3365.

وبذلك سجلت طريقه انحدار لاسو معامل تحديد اعلي من معامل تحديد طريقه انحدار الحرف واقل منها من متوسط مربعات الاخطاء.

كما لاحظنا ان نموذج انحدار لاسو ادي الي تقلص كلا من X_1،X_6 والتي ظهرت من خلال طريقه الانحدار الحرف غير معنويه احصائيا مع المتغير التابع وبمقارنه الطرق السابقه كالتالي:ـ

جدول (7) مقارنه طرق التقدير للنموذج

Method	MSE	R^2	P.value
OLS	7622157450	98.2%	0.00
RR	2.389 E10	86.5%	0.00
PCR	103.95	90.5%	0.00
Lasso	76.336	97%	0.00

ومن خلال جدول (7) نلاحظ ان أفضل طريقه لتقدير نموذج الانحدار في حاله تعرضه لمشكله الازدواج الخطي من الدرجه العاليه هي طريقه انحدار لاسو حيث سجل اقل متوسط مربعات اخطاء MSE مقارنه بباقي الطرق يليها طريقه انحدار الحرف اما كلا من طريقه المكونات الرئيسيه (PCR) وطريقه المربعات الصغري (OLS) ظهر متوسط مربعات الاخطاء بقيم عاليه جدا وبالتالي فهناك افضليه لطريقه الانحدار بطريقه لاسو.

النتائج والتوصيات

استهدف البحث المفاضلة بين طريقة المربعات الصغرى وطريقة انحدار الحرف وطريقة المكونات الرئيسية وطريقه انحدار لاسو وذلك للوصول للطريقة المثلى لعلاج مشكلة الازدواج الخطي وتوصلت الدراسة إلى:

بفحص مصفوفة معاملات الارتباط تبين وجود ارتباط متوسط وعالي وذات دلالة إحصائية بين المتغيرات المستقلة وذلك مؤشر على وجود مشكلة الازدواج الخطي وبتقدير معلمات نموذج الانحدار باستخدام طريقة المربعات الصغرى وجدنا ظهور متغيرات بإشارات تخالف الواقع وذات أخطاء معيارية عالية مع معنوية المتغيرات ومعامل تحديد ومعامل ارتباط مرتفع وكل ذلك دليل علي وجود مشكله الازدواج الخطي وباستخدام معامل تضخم التباين VIF وجدنا إن هناك متغيرات مستقلة قيمة ($10 <$ VIF) وكذلك المؤشر الشرطي ($CI > 30$) وكل ذلك يدل على ان هناك احتمالية عالية لوجود ازدواج خطي، وتم استخدام طريقة انحدار الحرف (RR) للتخلص من مشكلة الازدواج الخطي وتم التوصل للقيمة المثلى لمعلمة الحرف ($k = 0.30$) والحصول على نموذج انحدار أكثر استقرارًا بمعامل تحديد $R^2 = 90.4\%$ ومتوسط مربعات أخطاء 103.95 وهي أقل بكثير من طريقة المربعات الصغرى.كما تم استخدام طريقة المكونات الرئيسية (PCR) لعلاج مشكلة الازدواج الخطي حيث اثبتت فاعليتها في الحصول على نموذج مستقر بمعامل تحديد $R^2 = 86.5\%$ ومتوسط مربعات خطأ 2.389E10 وتم استخدام طريقه انحدار لاسو(Lasso) وتم الحصول علي نموذج بمعامل تحديد $R^2 = 97\ \%$ ومتوسط مربعات اخطاء MSE=76.336 وقد تم التخلص من مشكله الازدواج الخطي وبالمفاضلة تبين أفضلية انحدار لاسو (Lasso) عن كلا من طريقة انحدار الحرف (RR) وطريقه المكونات الرئيسية (PCR) وطريقة المربعات الصغرى (OLS) وذلك لأنها صاحبة أقل متوسط مربعات أخطاء MSE وأقل قيمة معامل تضخم التباين VIF.

وبناءً على ما سبق يوصي الباحث:

١- استخدام طرق معالجة أخرى لمشكلة الازدواج الخطي ومقارنتها بالطرق محل الدراسة مثل Bayesian Ridge Regression ، Robust Ridge.

٢- التأكد من كفاية حجم العينة قبل استخدام طريقة المكونات الرئيسية.

٣- التوسع في دراسه طريقة انحدار لاسو (Lasso) حيث انها تعد من الطرق الحديثه في تقدير النماذج الاحصائيه التي تعاني من مشكله الازدواج الخطي.

٤- محاولة الوصول إلى معايير أخرى للمفاضلة بين طرق علاج مُشكلة الازدواج الخطي غير MSE، VIF، CI، R^2.

المراجع العربية:

١- الحلواني، ماجي خليل (٢٠٢٢) "استخدام أسلوب انحدار الحرف لتقدير حجم الهجرة الداخلية في مصر" جامعة عين شمس، كُلية التجارة، قسم الإحصاء والرياضة، المجلة المصرية للتنمية والتخطيط.

٢- الشيخ، ساوس (2011) "معالجة مشكلة الازدواج الخطي باستخدام انحدار الحرف دراسة تطبيقية على الانفاق الاستهلاكي في الجزائر خلال الفترة من 1970-2011"، مجلة الحقيقة، جامعة ادرار، الجزائر، العدد 29.

٣- الكفيشي،ساره ماجد (2019)"تقدير معلمات نموذجي انحدار لاسو وانحدار الحرف مع تطبيق عملي" جامعه كربلاء، كليه الاداره والاقتصاد،قسم الاحصاء،رساله دكتوراه.

٤- القصاب، موفق محمد – آمين، مُعاذ عبد الرحيم (2018) "استخدام أسلوب انحدار الحرف لمعالجة مشكلة التعدد الخطي مع التطبيق" المجلة الأكاديمية لجامعة لوزوز، المُجلد السابع، العدد الثاني.

٥- العُبيدي، ندى نزار (2019) "مقارنة بين طرق تقدير معلمة انحدار الحرف المُعممة مع التطبيق على بيانات مرض العجز الكلوي المُزمن" مجلة جامعة الأنبار للعلوم الاقتصادية والإدارية، المجلد (11)، العدد (25).

٦- عطية، عبد القادر محمد (1998) "الاقتصاد القياسي بين النظرية والتطبيق" جامعة الإسكندرية، كُلية التجارة، الطبعة الثانية.

٧- طه، امتثال ميرغني (2014) "معالجة مشكلة التداخل الخطي المُتعدد لشركة النيل الأزرق للتغليف والطباعة باستخدام انحدار الحرف (1986-2010)" جامعة السودان للعلوم والتكنولوجيا، كُلية الدراسات العُليا، قسم الإحصاء التطبيقي.

٨- سالم، محمد عبد الوهاب (2018) "طُرق معالجة التداخل الخطي المتعدد لبيانات الاستثمار الكُلي في السودان (1980-2014)" جامعة السودان للعلوم والتكنولوجيا، كُلية الدراسات العُليا، قسم الإحصاء التطبيقي، رسالة دكتوراة.

٩- حسن،محمود- شاكر،هديل- حبي،ناصر (2020) "مقارنة بين طريقتي انحدار الحرف وانحدار المركبات الرئيسية باستعمال محاكاة مونت كارلو من خلال متوسط مربعات الخطأ MSE" مجلة كُلية الرافدين الجامعة للعلوم، العدد (46).

المراجع الأجنبية:

1- Duzan, H., and N.S.B M. Shariff (2015) “Ridge regression for solving the multicollinearity problem review of methods and models”, Journal of Applied science.
2- Farrar, D. E, &Glaubar, R. R, (1967) “Multicollinearity in regression analysis the problem revisited” The Review of Economic and Statistics, 92-107.
3- Fonti,.V,(2017),"Feature Selection using LASSO" Research paper in Business Analytics,VU Amesterdam.
4- Fuwenjiang, J. (1998)” penalized regressions: The Ridge versus The Lasso” Journal of Computational and graphical statistics , VOL.7 (3), pp. 397 – 416 .
5- Hoerl, A. E and Kennard (1970) “Ridge Regression Application to non orthogonal problem”.
6- Hoerl.E., and Robert w. Kennard (2000) “Ridge Regression: Biased Estimation for Non orthogonal problems “Technometric, Vol, 42, No. 1, PP 80-86.
7- Johnson, R. A., and Wicheren, D. W (1998) “Applied Multivariate statistical Analysis” Prentice – Hall, Inc. N. Y.
8- Lukman, A.,Arowolo O., and Ayide, k., (2014) “Some robust ridge regression for handing multicollinearity and outlier” International Journal of Science: Basic Applied Research 6 (2): 192-2022.

9- Manly, B. F., J (1994) “Multivariate statistical Methods” Chapman and Hall, London.

10- Tibshirani, R.,(1996),"Regression Shrinkage and selection Via the Lasso", J.R.Statist.SOC.B58,NO.1,pp.267-288.

11- Wang, H., Li, G. and tasi ,C., (2007)” regression Coefficient and auto regressive order shrinkage and selection via the Lasso” J.R.statist. SOC. B, VOL.(69), pp. 63 – 78 .