مجلة أبحاث كلية التربية الأساسية ، المجلد ١٩، العدد (٢)، لسنة ٢٠٢٣

College of Basic Education Researchers Journal. ISSN: 7452-1992 *Vol. (19), No.(2), (2023)*

المقدرات المقلصة الخطية في نماذج بيانات العد: مراجعة مقال

نوال محمود حمد **ندوى خزعل رشاد**

قسم نظم المعلومات الادارية/ كلية الادارة والاقتصاد/ جامعة الموصل

(قدم للنشر في ٩/ ١٠/ ٢٠٢٢،قبل للنشر في ٦/ ١٢/ ٢٠٢٢)

المستخلص

تعد مشكلة التعدد الخطي التي تحدث نتيجة للإرتباط العالي والتلازم الخطي بين متغيرين أو أكثر من المتغيرات التوضيحية الداخلة في بنائها واحدة من أهم مشاكل بناء النماذج ، مما يؤثر بشكل سلبي على عملية تقدير معلمات انموذج الانحدار. تهدف هذه الدراسة الى استعراض المقدرات المقلصة التي تستخدم لمعالجة مشكلة التعدد الخطي التي تظهر في نماذج بيانات العد وخصوصا انموذج انحدار بواسون. يتعتبر هذا النموذج من اكثر النماذج. المستخدمة في حالة كون بيانات متغير الاستجابة ذات قيم قابلة للعد ولا تخضع للتوزيع الطبيعي. من خلال تجارب محاكاة مونت-كارلو تبين بان المقدر المقلص ذو المعلمتين افضل المقدرات المقترحة بسبب تقليله متوسط مربعات الخطأ للنموذج.

الكلمات المفتاحية: التعدد الخطي، مقدر الحرف، مقدر ليو، انموذج انحدار بواسون، نماذج بيانات العد

Linear Shrinkage Estimators in Count Data Models: an Article Review

Nawal Mahmoud Hamad **Nadwa Khazal Rashad**

Department of Management Information Systems/ College of Administration and Economics/ University of Mosul

Abstract

The problem of multilinearity that occurs as a result of the high correlation and linear correlation between two or more explanatory variables included in its construction is one of the most important model building problems, which negatively affects the process of estimating the regression model parameters. This study aims to review the shrinking estimators that are used to address the problem of multicollinearity that appears in counting data models, especially the Poisson regression model. This model is considered one of the most used models in the case that the data of the response variable have countable values and are not subject to a normal distribution. Through Monte-Carlo simulation experiments, it was found that the two-parameter shrink estimator is the best proposed estimator because it reduces the mean square error of the model.

١– المقدمة

إحدى الغايات الرئيسية من دراسة أية مشكلة أو ظاهرة سواء كانت اقتصادية أَم اجتماعية أَم طبيية أَم غيرها من المجالات هو إيجاد المعادلة الرئيسية التي تُمثلها بشكل دقيق والتي ستكون المدخل الرئيسي لفهم هذه الظاهرة وتحديد معالمها الرئيسية وهو ما يعرف في علم الإحصاء بنمذجة الظواهر (طراد, ٢٠١٣).

كما أنّ فلسفة علم الإحصاء تكمن من حيث آلية التطبيق ومحاولة نمذجة الحالات المختلفة بنماذج تكون أقرب ما يمكن إلى الواقع الحقيقي، إذ إنّ جودة هذه النماذج تُقاس بحسب درجة تقاربها مع الخواص الاستدلالية الإحصائية، وهذه النماذج على أشكال مختلفة، فمنها: الاحتمالي التى تعتمد في صياغتها على الاحتمالات الصِرفة، ومنها التي تُصاغ نماذجها على ما يعرف بالسببِ ونتيجته وهي النماذج السببية، وتأتي في مقدمة هذه النماذج ما يُسمى بنماذج الانحدار (Regression Models). التي تُعد أحد أكثر أساليب الإحصاء التطبيقي استعمالاً وأهميةً، إذ تقوم بتفسير العلاقة بين المتغيرات ووصفها على شكل نموذج يتم من خلالهِ تفسيرُ العلاقة بين واحد أو أكثر من المتغيرات السببية الذي تُعرف إحصائياً بالمتغيرات التوضيحية(Explanatory Variables) وبين نتيجتهُ أو ما يُعرف بمتغير الاستجابة (Response Variable) تستعمل لغرض التقدير والتنبوء بما ينفع في التخطيط واتخاذ القرارت الصائبة.

يعد نموذج انحدار بواسون أحد أهم نماذج الانحدار اللوغاريتمية الخطية, وهو الأداة التي يتم من خلالها نمذجة المتغير المعتمد عندما تكون قيم ذلك المتغير على شكل قيم قابلة للعد. وكغيره من سائر نماذج الانحدار, قد يحتوي النموذج على متغيرات مستقلة كثيرة ما يؤثر سلباً على دقة النموذج وبساطته في تفسير النتائج.

إن غالبية البيانات في الواقع التطبيقي الحقيقي لا تخلو من وجود مشاكل كمشكلة التعدد الخطي شبه التام التي تحصل في هذه النماذج، وهي من المشاكل القائمة الوجود التي باتت معروفة لدى الباحثين الإحصائيين وكذلك آثارها السلبية في عمليةِ التقدير، إذ إنّ هذه المشكلة في أبسط حالتها تؤدي بنا إلى استنتاج أنّ بعض المتغيرات التوضيحية غير

معنوية ولكن في الواقع هي مهمة ومعنوية؛ ولكن بناء النموذج يَعجزُ عن إظهارِ أثر كل منها بشكل منفصل بسبب ارتباط (تلازم) هذه المتغيرات مع بعضها، وكذلك تأثيرها على المقدرات والمعلمات وتباينات المعلمات المقدرة المحسوبة . لذا وجب تفادي هذه المشكلة ووضع الحلول المناسبة لها ،إذ أُستَعِملت طرائق عدة للتخلص منها ومعالجتها، ومن هذه الطرائق ماتمَّ تقديمهُ في عام (١٩٧٠) من قبل (Hoerl & Kennard)، والمُسماة بطريقة انحدار الحرف (Ridge Regression) التي تُعدّ من أكثر الطرائق المتحيزة شيوعاً، التي تَتَمثلُ بإضافة كمية موجبة صغيرة إلى العناصر القطرية لمصفوفة المعلومات التي تعمل على فك الارتباطات بين هذه المتغيرات (النعيمي, ٢٠٠٥; حسن, ٢٠١٧ ; حمود و صالح, ٢٠١٦).

٢ انموذج انحدار بواسون

يعُد تحليل الانحدار أحد أهم الوسائل المُستخدمة لتحليل البيانات التي تحوي على متغيرين فأكثر. إذ يرتبطُ مفهوم الانحدار بدراسة الاعتمادية لأحد المتغيرات الذي يسمى بمتغير الاستجابة (y) (Response variable)، المعتمد على متغير واحد أو أكثر من المتغيرات المستقلة التي تُسمى بالمتغيرات التوضيحية، من خلال إنتاج معادلة إحصائية تُدعى بنموذج الانحدار (التميمي, السعدون, و الثعلبي, ٢٠١٤).

إذ إننا في تحليل الانحدار نُجري توافقا بين النموذج التنبؤي والبيانات المتوفرة، إذ يتم استخدام البيانات لتكوين نموذج يمكنهُ أن يصف الظاهرة المدروسة بشكل جيد، ونستخدم هذا النموذج لنتوقع قيماً لمتغير الاستجابة اعتماداً على متغير أو أكثر من المتغيرات التوضيحية، وهذا يعني أنّه يمكننا الحصول على النتيجة باستعمال نموذج ملائم للبيانات مع إضافة نوع من الخطأ(حسن و رنة, ٢٠١٩).

يعد نموذج انحدار بواسون أحد أهم نماذج الانحدار اللوغاريتمية الخطية, وهو الأداة التي يتم من خلالها نمذجة المتغير المعتمد عندما تكون قيم ذلك المتغير على شكل قيم قابلة للعد. وكغيره من سائر نماذج الانحدار, قد يحتوي النموذج على متغيرات مستقلة كثيرة ما يؤثر سلباً على دقة النموذج وبساطته في تفسير النتائج. يفترض هذا الأنموذج أن المتغير المعتمد y_i هو متغير استجابة يتبع توزيع بواسون وبمعلمة قدرها (μ)، كما تتبع الأخطاء العشوائية في الأنموذج توزيع بواسون بمعلمة قدرها (μ) Mansson and Kubria ((2012) Hossain And Ahmed,) . (2012) ويعرف وفق الدالة الاحتمالية المعرفة بالصيغة الاتية.

$$y_i = e^{XB+U} \qquad \dots (1)$$

ويمكن التعبير عنه ايضا بصيغة المصفوفات وكالتالي:

$$\mathbf{y} = Exp(\boldsymbol{X}\boldsymbol{\beta} + \mathbf{U}) \qquad \dots (2)$$

إذ أن:

y: موجه المتغير التابع ذي درجة, $(n\times 1)$ **X** : مصفوفة المتغيرات المستقلة (التوضيحية) ذات الدرجة $(n\times(p+1))$

β: موجه المعلمات ذو الدرجة $((p+1)\times 1)$ **U**: موجه الأخطاء العشوائية ذي الدرجة $(n\times 1)$ n : حجم العينة P : عدد المتغيرات المستقلة (التوضيحية).

لأجل تقدير معلمات أنموذج انحدار بواسون باستخدام طرائق الإمكان الجزائية سيتم اللجوء الى تعظيم المشاهدات لتوزيع المتغير المعتمد (y_i)، إذا كان المتغير المعتمد (y_i) يتبع توزيع بواسون بمعلمة قدرها (μ_i) فتكون دالة التوزيع كما في الصيغة (1) والمعرفة سلفا بالشكل التالي:

$$f(y_i/\mu_i)=\frac{e^{-\mu i}\,\mu_i^{y_i}}{y_i!}$$

ومن خلال تعظيم المشاهدات لتوزيع المتغير المعتمد (y_i) الوارد في الصيغة أعلاه تكون دالة الإمكان الأعظم بالشكل الآتي:

$$L(y_1,y_2,\dots,y_n;\mu_i)=\frac{Exp\{-\sum_{i=1}^{n}\mu_i\}\,\mu_i^{\sum_{i=1}^{n}y_i}}{\prod_{i=1}^{n}y_i\,!}\qquad \dots(3)$$

وبأخذ اللوغاريتم الطبيعي لدالة الإمكان الأعظم للمشاهدات أعلاه نحصل على:

$$LogL(\mathbf{y}_{\mathrm{i}}|x_i, \boldsymbol{\beta}\ ) = -\sum_{i=1}^{n} \mu_i + \sum_{i=1}^{n} y_i\ (\mathrm{Log}\{\mu_i\}) - \mathrm{Log}\left\{\prod_{i=1}^{n} y_i\ !\right\} \quad ...(4)$$

وبالاعتماد على الافتراض الثاني من الفروض الأساسية لأنموذج انحدار بواسون $\mu_i = \mathrm{Exp}\{x_i^T \boldsymbol{\beta}\}$ ، يتم تعويض هذا الافتراض بالدالة (4) أعلاه وكما يلي:

$$LogL(\mathbf{y}_{\mathrm{i}}|x_i, \boldsymbol{\beta}\ ) = -\sum_{i=1}^{n}\left(\mathrm{Exp}\{x_i^T \boldsymbol{\beta}\}\right) + \sum_{i=1}^{n} y_i \left(\mathrm{Log}\left\{\mathrm{Exp}\{x_i^T \boldsymbol{\beta}\}\right\}\right) - \mathrm{Log}\left\{\prod_{i=1}^{n} y_i\ !\right\}$$

$$= \sum_{i=1}^{n}\left(y_i x_i^T \boldsymbol{\beta} - EXP\left(x_i^T \boldsymbol{\beta}\right) - \log y_i!\right) \qquad ...(5)$$

ومقدر الامكان الاعظم لمعلمات انموذج بواسون هي

$$\hat{\boldsymbol{\beta}}_{PRM} = (\mathbf{X}^T\, \hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^T\, \hat{\mathbf{W}}\hat{\mathbf{u}}, \qquad (6)$$

حيث ان $\hat{\mathbf{W}} = \mathrm{diag}(\hat{\theta}_i)$ و $\hat{v}_i = \ln(\hat{\theta}_i) + ((y_i - \hat{\theta}_i)/\hat{\theta}_i)$.

كما تعرف مصفوفة التباين لمقدر الامكان الاعظم بالشكل الاتي:

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}_{PRM}) = \left[-E\left(\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial\beta_i\ \partial\beta_k}\right)\right]^{-1} = (\mathbf{X}^T\, \hat{\mathbf{W}}\mathbf{X})^{-1}. \qquad (7)$$

اما معيار MSE لمقدر الامكان الاعظم يعرف رياضيا بالشكل الاتي

$$
\begin{aligned}
\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{PRM}) &= E(\hat{\boldsymbol{\beta}}_{PRM} - \hat{\boldsymbol{\beta}})^T (\hat{\boldsymbol{\beta}}_{PRM} - \hat{\boldsymbol{\beta}}) \\
&= tr[(\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1}] \qquad (8) \\
&= \sum_{j=1}^{p} \frac{1}{\lambda_j},
\end{aligned}
$$

٣ مشكلة التعدد الخطي

إنّ مصطلح الارتباط الخطي المتعدد (Multicollinearity) أو التداخل الخطي المتعدد هو مصطلح إنكليزي مكون ثلاثة مقاطع: (Multi) متعدد، و (co) وتعني مرتبط أو متداخل، و(linearity) ومعناها خطي. إذ إنّ هذا المصطلح يعود الى العالم النرويجي Frisch عام ١٩٣٤م، من خلال تسليطه الضوء على هذه الظاهرة في معرض تحليله لبيانات السلاسل الزمنية. فوجد أنّ المتغيرات الاقتصادية التوضيحية متناغمة عبر الزمن وهذا ما يجعل بعضها يفقد المقدرة التفسيرية في النموذج (التميمي وآخرون, 2014).

في تحليل الانحدار غالباً مايواجه الباحثون مشكلة تعدد العلاقة الخطية (Multicollinearity)، وهذا ما جعل هذه المشكلة محط اهتمام الكثير من البحوث والدراسات لما تسببه من تشويه لمقدرات المربعات الصغرى الاعتيادية (OLS)؛ إذ تكون التقديرات لمعاملات الانحدار كبيرة في القيمة المطلقة في ظل وجود التعدد الخطي(Akdeniz,F et al., 2015). إذ تحصل هذه المشكلة عندما تكون هناك علاقة خطية قوية بين اثنين أو أكثر من المتغيرات التوضيحية في نموذج الانحدار مما يؤدي إلى خرق في أحد فروض التحليل (فرض استقلالية المتغيرات التوضيحية)، بحيث يكون من الصعب فصل أثر كل متغير عن متغير الاستجابة، وإنّ الارتباط بين هذه المتغيرات يعرف بالتعدد الخطي، وكذلك تظهر مشكلة التعدد الخطي عند اعتماد قيمة أحد المتغيرات التوضيحية على قيمة واحدة

أو أكثر من المتغيرات التوضيحية الأخرى في النموذج، علما بأن هذه المشكلة قد تواجه الباحث في حالة تجانس أو عدم تجانس التباين.

وهناك العديد من الطرائق الإحصائية المختلفة التي يتم بواسطتها الكشف عن مشكلة التعدد الخطي بين المتغيرات التوضيحية، منها فحص مصفوفة الارتباط، عامل تضخم التباين، ومقياس العدد الشرطي.

للتغلب على مشكلة التعدد الخطي توجد عدة إجراءات وطرائق مقترحة، فمن هذه الإجراءات ماهو منصب على معالجة بعض الاختلالات في البيانات من دون التعديل على بنية معادلة طريقة (OLS)، كالتحويل المعياري لقيم المتغيرات، الذي من شأنه أضعاف التداخل الخطي وبالنتيجة خفض قيمة معامل الارتباط بين أزواج المتغيرات التوضيحية، أو عن طريق حذف(إسقاط) أحد المتغيرات المتسببة في مشكلة التعدد الخطي أو استبدالها، ومما يعاب على هذا الإجراء حصول مشكلة بتوصيف البيانات، وعدم استقرارية المعلمات المقدرة. وكذلك يمكن التخفيف من وطأة التعدد الخطي بإضافة بيانات جديدة إلى البيانات الأصلية وهذا الإجراء مشابه لأسلوب زيادة حجم العينة، إذ ترتفع القيم الذاتية لمحدد مصفوفة المعلومات($|Z^T Z|$)، ولهذا الإجراء بعض السلبيات منها ما يعود إلى أسباب اقتصادية أو تغيرات في المجتمع المبحوث، كما أنّ تضخم البيانات قد يؤدي إلى تطرفها(السواعي, ٢٠١٨; دبدوب و النعيمي, ٢٠٠٦).

كما ويمكن استخدم الطرائق الإحصائية البديلة التي تعمل على تخليص النموذج من مشكلة التعدد الخطي، إذ تعد مقدرات التقليص(المتحيّزة)(shrinkage estimator)، من أكثر الطرائق الإحصائية شيوعاً في هذا المجال، إذ إنّ أُسلوب التقليص يشير إلى نهج مقيد للتقدير أو التنبوء عند ارتباط اثنين أو أكثر من المتغيرات التوضيحية في نموذج الانحدار الخطي المتعدد(Duran et al., 2011)، إذ إنّ الهدف من التقليص هو تقليل الارتباطات بين المتغيرات

التوضيحية وبالتالي تقليل التباين للمعلمات المقدرة. وبالرغم من أنّ هذه الطريقة تُعطي مقدرات متحيزة إلا أنّها تكون أكثر دقة من المقدرات غير المتحيزة، وإنّ مقدراتها تتميز بقدر كبير من الاستقرارية مما جعلها محل اهتمام الباحثين.

في الانحدار الخطي عادة ما يتم تقدير المعلمات بطريقة المربعات الصغرى الاعتيادية (OLS)، التي تعطي تقديرات غير متحيزة، إلا أنّ هذه التقديرات ستكون ضعيفة وعادة ما تكون غير مستقرة وتعطي معلومات مضللة عند وجود تعدد العلاقة الخطية (Duran & Akdeniz, 2013; Wu, 2018)؛ لذا يتم استخدام مقدرات متحيزة مثل مقدر الحرف، ومقدر LIU ،ومقدر LIU-type التي تميل إلى تحسين دقة التنبوء بتقليص بعض المعلمات، لكن جميع هذه المقدرات لا تقلص قيم المعلمات وتجعلها مساوية إلى الصفر.

٤ معالجة مشكلة التعدد الخطي في انموذج انحدار بواسون المتضخم صفريا

من أجل التغلب على مشكلة التعدد الخطي شبه التام في نموذج انحدار بواسون المتضخم صفريا لغرض تخليص النموذج من هذه المشكلة، وسنتناول في هذه الدراسة بعضاً من هذه المقدرات المتحيزة: كمقدر الحرف (Ridge)، مقدر LIU، مقدر LIU type، فضلاً عن المُقدر ذو المعلمتين (Tow Parameters).

٤-١ مقدر الحرف Ridge Estimator

اقترح الباحثان (Hoerl & Kennard) في عام ١٩٧٠م طريقة بديلة لطريقة (OLS) لتقدير معلمات نموذج الانحدار الخطي المتعدد في ظل وجود مشكلة التعدد الخطي، وذلك لأن طريقة (OLS) وبالرغم ما تتصف به من مميزات جيدة: كسهولة حساب تقدير المعلمات، ومنطقية نتائجها، وسهولة فهم آلية عملها، إلا أنها تعاني من عدم

استقرارية المعلمات المقدرة، لذا تم اللجوء إلى طريقة بديلة وهي طريقة انحدار الحرف الاعتيادية، التي تُعد واحدة من أكثر الطرائق المتحيّزة شيوعا في معالجة مشكلة التعدد الخطي شبه التام؛ إذ تتمثل هذه الطريقة بإضافة قيمة صغيرة موجبة(k) تدعى بمعلمة التحيّز(التقليص) (Bias parameter) إلى العناصر القطرية لمصفوفة المعلومات $\mathbf{X}^T\mathbf{X}$ إذ تعمل هذه القيمة المضافة على تقليص قيمة المعلمة المقدرة عن طريق فك الارتباطات بين المتغيرات التوضيحية مما يؤدي إلى ثبات قيمة المعلمة المقدرة من حيث الإشارة والقيمة (أي تكون أكثر استقراراً)، وكذلك تعمل على خفض التباين للمعلمات المقدرة عن طريق تقليل قيم العناصر القطرية لمعكوس مصفوفة المعلومات $\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}$ وعلى الرغم من أنّ هذه الطريقة تعطي تقديرات متحيزة، إلا أنّها أكثر دقة من المقدرات غير المتحيزة لطريقة (OLS) (دبدوب و النعيمي, ٢٠٠٦).

وقد حَصَل (Hoerl & Kennard) على مُقدر الحرف عن طريق تقليل ($\hat{\beta}^T\hat{\beta}$) ليخضع للشرط الآتي:

$$\left(\hat{\beta}-\hat{\beta}_{OLS}\right)^T Z^T Z\left(\hat{\beta}-\hat{\beta}_{OLS}\right)=C \qquad (9)$$

إذ إنّ C ثابت، ونتيجة لمعادلة لاكرانج يمكن الحصول على المعادلة الآتية:

$$F=\hat{\beta}^T\hat{\beta}+\frac{1}{k}\left[\left(\hat{\beta}-\hat{\beta}_{OLS}\right)^T Z^T Z\left(\hat{\beta}-\hat{\beta}_{OLS}\right)-C\right] \qquad (10)$$

إذ إنّ ($\frac{1}{k}$) هو مضاعف لاكرانج، وباشتقاق المعادلة الأخيرة بالنسبة لـ($\hat{\beta}$) نحصل على الصيغة الأتية:

$$\frac{\partial F}{\partial\hat{\beta}}=2\hat{\beta}+\frac{1}{k}\left[2Z^T Z\hat{\beta}-2Z^T Z\hat{\beta}_{OLS}\right]=0 \qquad (11)$$

يتم الحصول على مقدر الحرف في انموذج بواسون بالصيغة الآتية

$$\hat{\boldsymbol{\beta}}_{Ridge} = (\mathbf{X}^T \hat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^T \hat{\mathbf{W}}\mathbf{X}\hat{\boldsymbol{\beta}}_{PRM} = (\mathbf{X}^T \hat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^T \hat{\mathbf{W}}\hat{\mathbf{u}}, \quad (12)$$

إذ إنّ k: تمثل معلمة التحيز $(k > 0)$، وإنّ I: تمثل مصفوفة أُحادية. ان مصفوفة التباين والتباين المشترك للمقدر

Variance -Covariance matrix of an estimator

تكون وفق الصيغة التالية

$$\text{MSE}(\hat{\boldsymbol{\beta}}_{Ridge}) = \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \sum_{j=1}^{p} \frac{\alpha_j}{(\lambda_j + k)^2}, \quad (13)$$

٢-٤ مقدر LIU **LIU Estimator**

اقترح الباحثLIU عام ١٩٩٣ م مقدراً متحيزاً لتقدير معلمات نموذج الانحدار عندما يعاني من مشكلة التعدد الخطي شبه التام، عن طريق الجمعِ بينَ مميزات كلٍ من مقدر انحدار الحرف (Ridge) ومقدر(Stein,1956) بمقدرٍ جديد أُطلقَ عليه مُقدر(LIU Estimator) (E. Akdeniz et al., 2018). إذ إنّ هناك مُميزات وعيوب لكل من مقدر انحدار الحرف ومقدر Stein ، فمن مميزات مقدر انحدار الحرف أنّهُ فعال في التطبيق العملي، لكن من عيوبهِ أنّهُ دالة معقدة في (k)، أما مقدر Stein فمن مميزاتهِ أنّهُ دالة خطية في (c) لكن يعاب عليهِ أنّ التقليص(shrinkage) هو نفسه لكل عنصر من عناصر $\left(\hat{\beta}_{Stein}\right)$، وهذا ما يجعلهُ لا يسلكُ سلوكاً جيداً في التطبيق العملي.

$$\hat{\boldsymbol{\beta}}_{stein} = \mathbf{c}\hat{\boldsymbol{\beta}}_{OLS} \quad , \quad \mathbf{0} < \mathbf{c} < \mathbf{1}.$$

وإنّ الصيغة الرياضية لمقدر LIU تكون بالشكل الآتي:

$$\hat{\boldsymbol{\beta}}_{LIU} = (\mathbf{X}^T \hat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}^T \hat{\mathbf{W}}\mathbf{X} + d\,\mathbf{I})\hat{\boldsymbol{\beta}}_{PRM}\,, \quad 0 < d < 1 \qquad (14)$$

إذ يُفضل مقدر LIU على مقدر انحدار الحرف لأمرين: الأول هو أنّ مقدر LIU دالة خطية في المعلمة (d) وهذا ما يجعل حساب قيمة معلمة (d) أسهل من حساب معلمة انحدار الحرف (k)، أما الثاني فهو أنّ مقدر LIU دالة متزايدة في المعلمة (d)، بينما مقدر انحدار الحرف هو دالة متناقصة في المعلمة (k) (السراي, ٢٠١٥).

ان مصفوفة التباين والتباين المشترك للمقدر Variance -Covariance matrix of an estimator تكون وفق الصيغة التالية

$$\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{LIU}) = \sum_{j=1}^{p} \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2} + (d-1)^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + 1)^2}. \qquad (15)$$

٤-٣ مقدر LIU-type　LIU-type Estimator

في عام 2003م قدم الباحث LIU مقدراً جديدًا أطلق عليه (LIU-type Estimator) وصيغته:

$$\hat{\boldsymbol{\beta}}_{PLT} = (\mathbf{X}^T \hat{\mathbf{W}}\mathbf{X} + k\,\mathbf{I})^{-1}(\mathbf{X}^T \hat{\mathbf{W}}\mathbf{X} - d\,\mathbf{I})\hat{\boldsymbol{\beta}}_{PRM}\,, k > 0\,,\ 0 < d < 1. \qquad (١٦)$$

ان مصفوفة التباين والتباين المشترك للمقدر Variance -Covariance matrix of an estimator تكون وفق الصيغة التالية

$$\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{PLT}) = \sum_{j=1}^{p} \frac{(\lambda_j - d)^2}{\lambda_j(\lambda_j + k)^2} + (d+k)^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + k)^2}. \qquad (17)$$

٤-٤ مقدر ذو المعلمتين Two-parameters Estimator

اقترح الباحثان(Asar & Genç) في عام 2018م مقدرًا متحيزاً جديداً ثنائي المعلمات (Two-parameters) ، إذ إنّ فكرة المقدر الجديد تقوم على أساس الجمع بين مقدرين مختلفين: مقدر الحرف (Ridge)، ومقدر LIU ، على أمل أنّ الجمع بين المقدرين قد يَكسبُ المقدر الجديد مزايا كل منهما، وبالتالي الحصول على مقدر بكفاءة أفضل. إذ تمّ تسمية المقدر الجديد بمقدر المعلمتين (Two-parameters Estimator)، وهو بالصيغة الآتية:

$$\hat{\boldsymbol{\beta}}_{PTP} = (\mathbf{X}^T\mathbf{X} + k\,\mathbf{I})^{-1}(\mathbf{X}^T\mathbf{X} + k\,d\,\mathbf{I})\hat{\boldsymbol{\beta}}_{PRM}\,,\ k > 0,\ 0 < d < 1. \qquad (18)$$

ان مصفوفة التباين والتباين المشترك للمقدر Variance -Covariance matrix of an estimator تكون وفق الصيغة التالية

$$\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{PTP}) = \sum_{j=1}^{p+1}\left[\frac{(\lambda_j + kd)^2}{\lambda_j(\lambda_j + k)^2} + k^2(d-1)^2\frac{\alpha_j^2}{(\lambda_j + k)^2}\right]. \qquad (19)$$

٥- نتائج المحاكاة

يُعدّ استعمال أُسلوب المحاكاة وسيلة لحل ودراسة المشاكل المعقدة، والمفهوم العام للمحاكاة أنّها عملية تصميم نموذج مشابه بمعطياته وفرضياته للنظام الحقيقي من أجل التعرف على سلوكية النظام الحقيقي، وبمعنى آخر هو تمثيل أو تقليد للواقع الحقيقي للنظام باستخدام نماذج معينة مكتوبة وفق أسلوب برمجي معين، وبتكرارات محددة؛

لغرض الحصول على نتائج تجريبية هدفها هو التأكد من النتائج التي تم الحصول عليها من الجانب النظري من أجل الحصول على الأُسلوب الأمثل أو اختيار الطريقة الأفضل للتحليل.

ويعد أسلوب (Mont Carlo) أحد أساليب المحاكاة الأكثر شيوعاً واستعمالاً في التحليل الذي تم استخدامه في توليد البيانات إذ يقوم بتوليد البيانات لمعظم التوزيعات الاحتمالية المعروفة من خلال أنواع العلاقات الرياضية والمنطقية.

٥-١ وصف مراحل تجربة المحاكاة Description Stages Simulation Experiment

تم تصميم تجربة ومحاكاتها باستعمال لغة البرمجة (R) وكما في الخطوات الاتية:

١- تعيين قيم عدد المتغيرات التوضيحية (p) واحجام العينات (n) ومعاملات الارتباط (r)، اذ ستؤخذ ثلاثة اعداد من المتغيرات التوضيحية (4,8,12) وخمسة احجام من العينات (30,50,100,150,200) وثلاثة قيم من معاملات الارتباط (0.90,0.95,0.99) كما مبينة في الجدول (3-1).

الجدول (1): يبين قيم العوامل p,n,r

القيم	العامل
4 8 12	p
30 50 100 150 200	n
0.90 0.95 0.99	ρ

٢- توليد قيم المتغيرات التوضيحية $(\mathbf{x}_j)$ من خلال اسلوب محاكاة مونت-كارلو حيث تم توليد مصفوفة ذات ابعاد $(n \times p)$ وتعاني بذات الوقت من مشكلة التعدد الخطي بين المتغيرات التوضيحية من خلال المعادلة التالية (McDonald and Galarneau, 1975)

$$x_{ij} = (1-\rho^2)^{1/2} w_{ij} + \rho w_{ip}, \; i = 1,2,\ldots,n, \quad j = 1,2,\ldots,p \qquad (20)$$

حيث ان:

ρ : هو معامل الارتباط البسيط بين المتغيرات التوضيحية.

$\mathbf{W}$: هي مصفوفة ذات ابعاد $(n \times p)$ تم توليدها من خلال اسلوب مونت-كارلو وفق التوزيع الطبيعي بمتوسط (0) وتباين (1).

حيث أن:

٣- تم ايجاد قيم المعلمات الاولية $(\boldsymbol{\beta})$ والتي تساوي متجه القيم المميزة الاكبر والتي يتم استخراجها من المصفوفة $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ بغض النظر عن قيمة B_0.

٤- يتم انشاء متغير الاستجابة بواسطة المعادلة رقم ٢

٥- لغرض تقليل التحيز في تجارب مونت-كارلو سوف يتم تكرار التوليد (١٠٠) مرة.

٦- لأجل المقارنة بين طرائق التقدير المستخدمة وبيان الطريقة الأفضل من خلال الاعتماد على متوسط مربعات الخطأ:

$$MSE = \frac{\sum_{r=1}^{R} \left(\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}\right)^T \left(\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}\right)}{R} \qquad (٢١)$$

اذ أن:

$\hat{\boldsymbol{\beta}}_{\mathbf{r}}$: قيمة المعلمة المقدرة وفق طرائق التقدير المختلفة.

β: قيمة المعلمة الحقيقية.

R : هي عدد مرات تكرار التجربة .

٣-٥ نتائج المحاكاة The results of simulation

تمَّ تطبيق الأفكار النظرية الواردة في الجزء النظري باستخدام أسلوب المحاكاة الذي أُجري على انموذج انحدار بواسون المتضخم صفريا، وبعد إيجاد متوسط مربعات الخطأ (MSE) لكل مقدر من المقدرات، وباعتبارهِ كمعيار للمقارنة تمت المفاضلة بين المقدرات التي تم توضيحها في الجانب النظري. الجداول ٢ الى ٦ توضح المقارنة بين هذه المقدرات.

من خلال النتائج التي تم الحصول عليها من الجداول ٢-٦ الخاصة بقيم (MSE) لكل مقدر من المقدرات ولجميع أحجام العينات وقيم الارتباط المستعملة تمَّ تلخيص النتائج بالنقاط الآتية:

١-أظهرت النتائج تفوق المقدر ذو المعلمتين على جميع المقدرات المستخدمة لتقدير المعلمات في انموذج انحدار بواسون ولجميع حجوم العينات وقيم الارتباط المستعملة لكونه يمتلك أقل قيم (MSE)، ويليه بالمرتبة الثانية مُقدر LIU-typeبكونه أفضل مقدر بعد مقدر ذو المعلمتين، إذ أظهرت قيم الـ(MSE) الخاصة بهِ تقارباً في النتائج مع المقدر ذو المعلمتين.

٢- مُقدر LIU جاء بالمرتبة الثالثة، بعد مُقدر LIU-type لكونه يمتلك قيماً (MSE) أقل بقليل من مقدر الحرف (Ridge) الذي بدورهِ جاء بالمرتبة الرابعة ولجميع قيم الأرتباط وأحجام العينات.

٣- كما أشارت النتائج أنّ أقل المقدرات كفاءةً وتمثيلاً للنموذج المفترض هو مقدر (MLE) ولجميع قيم الارتباط وأحجام العينات بسبب امتلاكهِ أعلى قيم (MSE)، وإنّ جميع المقدرات المتحيزة الأخرى تعد أفضل منه بسبب امتلاكها قيم (MSE) أقل.

٤-انخفاض قيم (MSE) عند ازدياد أحجام العينات لجميع قيم الارتباط ولجميع المقدرات المستخدمة ، وهذا يَدلُ على أن العلاقة بين حجم العينة وقيم الـ(MSE)علاقة عكسية.

٥- بشكل عام فإن قيم (MSE) تزداد بزيادة قيم معاملات الارتباط في جميع طرائق التقدير ولجميع حجوم العينات ، أي إنّ العلاقة بينهما علاقة طردية.

جدول ٢: نتائج ال MSE عندما n=30

p	r	PTP	PLT	LIU	Ridge	PRM
4	0.90	**3.323**	3.655	4.914	3.909	7.587
	0.95	**4.462**	5.069	9.27	9.005	10.805
	0.99	**6.055**	6.414	9.705	9.467	13.702
8	0.90	**4.001**	4.333	5.592	4.587	8.265
	0.95	**5.14**	5.747	9.948	9.683	11.483
	0.99	**6.733**	7.092	10.383	10.145	14.38
12	0.90	**4.128**	4.46	5.719	4.714	8.392

	0.95	**5.267**	5.874	10.075	9.81	11.61
	0.99	**6.86**	7.219	10.51	10.272	14.507

جدول 3: نتائج ال MSE عندما n=50

p	r	PTP	PLT	LIU	Ridge	PRM
4	0.90	**3.32**	3.652	4.911	3.906	7.584
	0.95	**4.459**	5.066	9.267	9.002	10.802
	0.99	**6.052**	6.411	9.702	9.464	13.699
8	0.90	**3.998**	4.33	5.589	4.584	8.262
	0.95	**5.137**	5.744	9.945	9.68	11.48
	0.99	**6.73**	7.089	10.38	10.142	14.377
12	0.90	**4.125**	4.457	5.716	4.711	8.389
	0.95	**5.264**	5.871	10.072	9.807	11.607
	0.99	**6.857**	7.216	10.507	10.269	14.504

جدول 4: نتائج ال MSE عندما n=100

p	r	PTP	PLT	LIU	Ridge	PRM
4	0.90	**2.262**	2.594	3.853	2.848	6.526

	0.95	**3.401**	4.008	8.209	7.944	9.744
	0.99	**4.994**	5.353	8.644	8.406	12.641
8	0.90	**2.94**	3.272	4.531	3.526	7.204
	0.95	**4.079**	4.686	8.887	8.622	10.422
	0.99	**5.672**	6.031	9.322	9.084	13.319
12	0.90	**3.067**	3.399	4.658	3.653	7.331
	0.95	**4.206**	4.813	9.014	8.749	10.549
	0.99	**5.799**	6.158	9.449	9.211	13.446

جدول 5: نتائج ال MSE عندما n=150

p	r	PTP	PLT	LIU	Ridge	PRM
4	0.90	**0.768**	1.1	2.359	1.354	5.032
	0.95	**1.907**	2.514	6.715	6.45	8.25
	0.99	**3.5**	3.859	7.15	6.912	11.147
8	0.90	**1.446**	1.778	3.037	2.032	5.71
	0.95	**2.585**	3.192	7.393	7.128	8.928
	0.99	**4.178**	4.537	7.828	7.59	11.825
12	0.90	**1.573**	1.905	3.164	2.159	5.837
	0.95	**2.712**	3.319	7.52	7.255	9.055

	0.99	**4.305**	4.664	7.955	7.717	11.952

جدول 6: نتائج ال MSE عندما n=200

p	r	PTP	PLT	LIU	Ridge	PRM
4	0.90	**0.34**	0.672	1.931	0.926	4.604
	0.95	**1.479**	2.086	6.287	6.022	7.822
	0.99	**3.072**	3.431	6.722	6.484	10.719
8	0.90	**1.018**	1.35	2.609	1.604	5.282
	0.95	**2.157**	2.764	6.965	6.7	8.5
	0.99	**3.75**	4.109	7.4	7.162	11.397
12	0.90	**1.145**	1.477	2.736	1.731	5.409
	0.95	**2.284**	2.891	7.092	6.827	8.627
	0.99	**3.877**	4.236	7.527	7.289	11.524

٦- الاستنتاجات

من خلال ما تمَّ التوصل إليه من نتائج في الجانب التجريبي ، أظهرت النتائج وفي ظل وجود مشكلة التعدد الخطي شبه التام الأفضلية للمقدر ذو المعلمتين في تقدير متجه معلمات انموذج انحدار بواسون ولجميع حجوم العينات، ومعاملات الارتباط، وذلك بامتلاكه أقل متوسط مربعات خطأ. كذلك جاءَ مُقدر LIU-type كأفضل مُقدر

لتقدير متجه معلمات معلمات انموذج انحدار بواسون. كما تبين بان أقل المقدرات كفاءةً في تقدير متجه معلمات معلمات انموذج انحدار بواسون هو مقدر MLE ولجميع قيم معاملات الارتباط وحجوم العينات ,وعدد المتغيرات المستقلة، بسبب أعطائهِ أعلى القيم لمتوسط مربعات الخطأ مقارنة بالمقدرات الأخرى.

المصادر

1-المصادر العربية

١- التميمي، زهرة عباس والسعدون، فوزية غالب و الثعلبي, ساهرة حسين ،(٢٠١٤)، "تحليل الانحدار". مديرية دار الكتب للطباعة والنشر ، جامعة البصرة، العراق.

٢- الراوي، خاشع محمود ،(١٩٨٧)، "المدخل إلى تحليل الانحدار" ، مديرية دار الكتب للطباعة والنشر، جامعة الموصل، العراق .

٣- السراي، حامد حران، (٢٠١٥)، "مقارنة مقدرات الإنتروبي العظمى ومقدرات ليوفين مع بعض المقدرات الأخرى لأنموذج الانحدار الخطي بوجود مشكلة التعدد الخطي شبه التام مع تطبيق عملي"، رسالة ماجستير في الإحصاء، كلية الإدارة والاقتصاد، جامعة بغداد.

٤- السواعي، خالد محمد، (٢٠١٨)، "الاقتصاد القياسي"، دار الكتاب الثقافي، أربد – الأردن.

٥- النعيمي، أسوان محمد، (٢٠٠٥)، "اختيار المتغيرات في انحدار الحرف"، رسالة ماجستير في الإحصاء، كلية علوم الحاسوب والرياضيات، جامعة الموصل.

٦- دبدوب، مروان عبد العزيز و النعيمي، أسوان محمد، (٢٠٠٦)، "طرائق مقترحة في انحدار الحرف"، المجلة العراقية للعلوم الإحصائية، المجلد(١٠)، الصفحات [٨٥-١٠٦].

٧- عبد المنعم، ثروت محمد، (٢٠٠٥)، "الانحدار"، مكتبة الإنجلو المصرية، القاهرة، مصر.

٨- كاظم، أموري هادي و مسلم، باسم شليبة، (٢٠٠٢)، القياس الاقتصادي المتقدم النظرية والتطبيق، مطبعة دنيا الأمل، بغداد، العراق.

2-المصادر الأجنبية

9- Akdeniz Duran, E., Härdle, W. K., & Osipenko, M., (2012), "Difference based ridge and Liu type estimators in semiparametric regression models" ,Journal of Multivariate Analysis, Vol. 105(1), pp[164-175].

10- Akdeniz, E., Akdeniz, F., & Roozbeh, M. (2018), "A new difference-based weighted mixed Liu estimator in partially linear models", Statistics, Vol. 52(6), pp[1309-1327].

11- Akdeniz, F., Akdeniz Duran, E., Roozbeh, M., & Arashi, M., (2015), "Efficiency of the generalized difference-based Liu estimators in semiparametric regression models with correlated errors", Journal of Statistical Computation and Simulation, Vol. 85(1), pp[147-165].

12- Akdeniz, F., & Duran, E. A., (2010), "Liu-type estimator in semiparametric regression models", Journal of Statistical Computation and Simulation, Vol. 80(8), pp[853-871].

13- Akdeniz, F., & Roozbeh, M., (2019), "Generalized difference-based weighted mixed almost unbiased ridge estimator in partially linear models", Statistical Papers, Vol. 60(5),pp[1717-1739].

14- Akdeniz, F., Roozbeh, M., Akdeniz, E., & Khan, N. M., (2020), "Generalized difference-based weighted mixed almost unbiased liu estimator in semiparametric regression models", Communications in Statistics-Theory and Methods, pp[1-22].

15- Akdenz, F., & Tabakan, G., (2009), "Restricted Ridge Estimators of the Parameters in Semiparametric Regression Model", Communications in Statistics - Theory and Methods, Vol. 38(11), pp[1852-1869].7

16- Akram, M. N., Amin, M., & Amanullah, M., (2020), "Two-parameter estimator for the inverse Gaussian regression model" Communications in Statistics - Simulation and Computation, pp[1-19].8

17- Algamal, Z. Y., & Asar, Y., (2020), "Liu-type estimator for the gamma regression model", Communications in Statistics-Simulation Computation, Vol. 49(8), pp[2035-2048].

18- Amini, M. & Roozbeh, M., (2015), "Optimal partial ridge estimation in restricted semiparametric regression models", Journal of Multivariate Analysis, Vol.(136), pp[26-40].

19- Asar, Y., & Genç, A., (2018), "A New Two-Parameter Estimator for the Poisson Regression Model", Iranian Journal of Science and Technology, Transactions A: Science, Vol. 42(2), pp[793-803].

20- Aydın, D., (2011), "Partially linear models based on smoothing spline estimated by different selection methods: a simulation study", Department of Statistics, Faculty of Arts and Sciences, Mugla University.

21- Duran, E. A., Akdeniz, F., & Hu, H., (2011), "Efficiency of a Liu-type estimator in semiparametric regression models", Journal of Computational and Applied Mathematics, Vol. 235(5), pp[1418-1428].

22- Emami, H., (2016), "Local influence in ridge semiparametric models", Journal of Statistical Computation and Simulation, Vol. 86(17), pp[3357-3370].

23- Engle, R. F., Granger, C. W., Rice, J., & Weiss, A., (1986), "Semiparametric estimates of the relation between weather and electricity sales", Journal of the American statistical Association, Vol. 81(394), pp[310-320].

24- Eubank, R., Kambour, E., Kim, J., Klipple, K., Reese, C., Schimek, M., & Analysis, D., (1998), "Estimation in partially linear models", Computational Statistics, Vol. 29(1), pp[27-34].

25- Fan, J., & Wu, Y., (2008), "Semiparametric estimation of covariance matrixes for longitudinal data", Journal of the American statistical Association, Vol. 103(484), pp[1520-1533].

26- Farrar, D. E., & Glauber, R. R., (1967), "Multicollinearity in regression analysis: the problem revisited", The Review of Economic and statistics, Vol.(49), pp[92-107].

27- Gujarati, D. N., (2004), "Basic econometrics", In: The Mc-Graw Hill.

28- Härdle, W. K., Müller, M., Sperlich, S., & Werwatz, A., (2004), "Nonparametric and semiparametric models An Introduction", Springer Science & Business Media.

29- Hoerl, A. E., & Kennard, R. W., (1970), "Ridge Regression, Biased Estimation for Nonorthogonal Problems", Technometrics., Vol. (12), pp[55-67].

30- Hu, H., (2005), "Ridge estimation of a semiparametric regression model", Journal of Computational Applied Mathematics, Vol. 176(1), pp[215-222].