A Comparison of Parameters Estimation Methods for the Negative Binomial Regression Model under Multicollinearity Problem by Using Simulation

مقارنة طرائق تقدير معلمات أنموذج انحدار ثنائي الحدين السالب في ظل وجود مشكلة التعدد الخطي باستعمال المحاكاة

أ.م. سهيل نجم عبود/ جامعة بغداد/ كلية الادارة والاقتصاد/ قسم الاحصاء
الباحث/ ايناس صلاح خورشيد

OPEN ACCESS

P - ISSN 2518 - 5764
E - ISSN 2227 - 703X

Received:10/7/2018
Accepted: 24/6/2018

الخلاصة

ناقش هذا البحث مقدر متحيز لأنموذج انحدار ثنائي الحدين السالب (Negative Binomial Regression Model) ومعرف بالمقدر ليو(Liu Estimator)، اذ استعمل هذا المقدر لتقليل التباين والتغلب على مشكلة التعدد الخطي بين المتغيرات التوضيحية، كما تم استخدام بعض التقديرات منها مقدر انحدار الحرف (Ridge Regression) ومقدر الامكان الاعظم (Maximum Likelihood)، اذ يهدف هذا البحث الى المقارنات النظرية بين مقدر (Liu Estimator) ومقدرات الامكان الاعظم (Maximum Likelihood) وانحدار الحرف (Ridge Regression) باستخدام معيار متوسط مربعات الخطأ (MSE)، اذ يكون تباين مقدر الامكان الاعظم (MLE) متضخم في ظل وجود مشكلة التعدد الخطي بين المتغيرات التوضيحية، وتم في هذا البحث تصميم المحاكاة (مونت كارلوا) لتقييم اداء المقدرات باستخدام معيار مقارنة متوسط مربعات الخطأ (MSE)، حيث اظهرت نتائج المحاكاة اهمية مقدر ليو وتفوقها على مقدري انحدار الحرف (RR) والامكان الاعظم (MLE) عندما يكون عدد المتغيرات التوضيحية (p=5) ولحجم العينة (n=100)، اما عندما يكون عدد المتغيرات التوضيحية (p=3) ولكافة الحجوم، وكذلك عندما (p=5) ولكافة الحجوم ماعدا حجم العينة (n=100) طريقة انحدار الحرف RR هي الافضل.

المصطلحات الرئيسة للبحث/ مقدر الامكان الاعظم، مقدر انحدار حرف، مقدر ليو، أنموذج انحدار ثنائي الحدين السالب، التعدد الخطي.

*البحث مستل من رسالة ماجستير.

Journal of Economics and Administrative Sciences
2019; Vol. 25, No.110
Pages: 466- 488

المبحث الاول/ المقدمة العامة

1-1 المقدمة introduction

تكمن فلسفة الإحصاء من حيث آلية التطبيق في محاولة نمذجة الظواهر المختلفة بنماذج اقرب مايكون الى الواقع الفعلي، وان هذه النماذج تقاس درجة قوتها بحسب درجة تقاربها مع الخواص الاحصائية وهي على اشكال وانواع مختلفة ومنها النماذج السببية التي تعتمد في صياغة نماذجها على ما يعرف بالسبب ونتيجة السبب وتاتي في مقدمة هذه النماذج ما تسمى بنماذج الانحدار اذ تقوم نماذج الانحدار باكتشاف العلاقة مابين السبب والذي يعرف احصائيا بالمتغيرات التوضيحية (التفسيرية)، وبين ماهو نتيجة السبب او مايعرف بالمتغير الاستجابة (المعتمد). وظهرت انواع مختلفة من نماذج الانحدار منها أنموذج انحدار بواسون وأنموذج انحدار ثنائي الحدين السالب (Negative Binomial Regression Model). اذ سنقوم في هذا البحث دراسة المقارنة بين طرائق تقدير معلمات أنموذج انحدار ثنائي الحدين السالب التي غالبا ما تستخدم في دراسة الظواهر الصحية والاجتماعية والاقتصادية وكذلك في علوم فيزيائية . وأنموذج الانحدار ثنائي الحدين السالب هو نادر الاستخدام في البحوث التطبيقية عندما ياتي المتغير التابع y_i في شكل اعداد صحيحة غير السالبة وهذا أنموذج هو اكثر فائدة من أنموذج انحدار بواسون لانها تتمكن من التعامل مع البيانات التي تكون فوق التشتت (over dispersion) لانه يسمح للاختلاف عشوائي في متوسط بواسون مشروط.

2-1 مشكلة البحث

تؤثر مشكلة التعدد الخطي بين المتغيرات التوضيحية على تباين مقدرات معلمات أنموذج انحدار ثنائي الحدين السالب بشكل سلبي، تحصل مشكلة التعدد الخطي عندما يرتبط اثنان او اكثر من المتغيرات التوضيحية بعلاقة خطية قوية جدا، اذ يتعذر تقدير معالم الأنموذج عندما تكون هناك علاقة خطية شبه تامة بين المتغيرات التوضيحية لذلك تم استخدام طرائق التقدير في عملية معالجة مشكلة التعدد الخطي وتقدير معلمات أنموذج انحدار ثنائي الحدين السالب في ظل وجود مشكلة التعدد الخطي.

3-1 هدف البحث

ان الهدف من البحث هو مقارنة طرائق تقدير معلمات أنموذج انحدار ثنائي الحدين السالب في ظل وجود مشكلة التعدد الخطي عبر طرائق التقدير ممكنة وهي طريقة مقدر الامكان الاعظم (MLE) وطريقة انحدار الحرف (RR) فضلا عن مقدر ليو (LE).

المبحث الثاني/ الجانب النظري

1-2 المقدمة

في هذا المبحث سيتم عرض ودراسة أنموذج انحدار ثنائي الحدين السالب وصيغتها العامة. والتطرق الى دراسة مشكلة التعدد الخطي وكيفية الكشف عنها و دراسة طرائق تقدير معلمات أنموذج الأنحدار ثنائي الحدين السالب في ظل وجود مشكلة التعدد الخطي.

2-2 أنموذج انحدار ثنائي الحدين السالب *Negative Binomial Regression Model*

يعرف أنموذج انحدار ثنائي الحدين السالب بانه احد انواع نماذج الانحدار المعممة التي تنضوي تحت مظلة نماذج الانحدار الخطية – اللوغارتمية وهو أنموذج نادر الاستخدام والذي تستخدم كبديل لنموذج انحدار بواسون (Poisson Regression Model) عندما تكون الشروط الاساسية للنموذج غير متوفرة ، وان هذا الأنموذج يمكن استخدامه لتحليل البيانات الكمية فوق التشتت (Over dispersion) اي انه يحل افتراضات مقيدة للمتوسط والتباين للمتغير المعتمد [(ix:pp179)]، اذ ان التباين يساوي المتوسط في أنموذج بواسون اي عندما تكون معلمة التشتت ($\theta=0$) فان تباين المتغير الاستجابة يساوي وسطه الحسابي وبالتالي نحصل على توزيع بواسون.

اذ تعد نمذجة المتغيرات الكمية من المهام المهمة في العديد من المجالات منها الاقتصادية والعلوم الاجتماعية والطبية وغالبا مايستخدم أنموذج انحدار بواسون لنمذجة ويكون هذا أنموذج غير ملائم عندما يكون هناك مشكلة فوق التشتت (Over dispersion) اي عندما تكون قيمة معلمة التشتت ($\theta>0$) فان تباين متغير الاستجابة سوف يتجاوز المتوسط الحسابي وعليه يستخدم أنموذج الانحدار ثنائي الحدين السالب (Negative Binomial Regression Model)[iii:pp19]. ويمكن اعتبار أنموذج (ثنائي الحدين السالب) تعميم لأنموذج بواسون لانه يحتوي على نفس هيكل أنموذج بواسون و يحتوي على معلمة اضافية لأنموذج فوق التشتت .يتم الحصول على نماذج انحدار ثنائي الحدين السالب بنفس طريقة نماذج انحدار بواسون عن طريق ربط المتوسط μ بالمتجه من المتغيرات التوضيحية x يمكن كتابتها بشكل الاتي [v:pp5]:

$$\mu = \exp^{(x_i'\beta)} \qquad (2-1)$$

اذ ان:

x_i': تمثل الصف (i) من المصفوفة X.

β : موجه معلمات ذو الدرجة ((p+1)x1).

كما يمكن اشتقاق أنموذج انحدار ثنائي الحدين السالب كخليط اثنين من توزيعات (بواسون – كاما) خليط الاوزان باستخدام معلمة اضافية (معلمة التشتت)، وهذا خليط تم اشتقاقه اولا من قبل الباحثان (Greenwood & yul) عام 1920 وهذا التوزيع استخدم لمعالجة حالات فوق التشتت (Over Dispersion) والملحوظة بشكل شائع في بيانات متقطعة (منفصلة) او في بيانات العد [viii:pp D-1]. اذ ان المتوسط المشروط والتباين المشروط للتوزيع تعطى كالاتي:

$$E(Y_i/X_i) = \mu_i \qquad (2-2)$$

$$Cov(Y_i/X_i) = \mu_i(1+\theta\mu_i) \qquad (2-3)$$

اذ ان دالة كثافة الاحتمالية لتوزيع ثنائي الحدين السالب بمعلمتين (θ, μ_i) بافتراض أن $\theta = \frac{1}{\delta}$ كالاتي:

$$f(y_i,\mu_i,\theta) = \frac{\Gamma(y_i+\theta^{-1})}{\Gamma(y_i+1)\Gamma\theta^{-1}} \left(\frac{\theta^{-1}}{\mu_i+\theta^{-1}}\right)^{\theta^{-1}} \left(\frac{\mu_i}{\mu_i+\theta^{-1}}\right)^{y_i} \qquad (2-4)$$

ويمكن ان تعرف دالة كثافة الاحتمالية لتوزيع ثنائي الحدين السالب ايضا بشكل الاتي ايضا:

$$f(y_i,\mu_i,\theta) = C_{\theta-1}^{y_i+\theta-1} \left(\frac{\theta}{\mu_i+\theta}\right)^{\theta} \left(\frac{\mu_i}{\mu_i+\theta}\right)^{y_i} \qquad (2-5)$$

استخدم أنموذج ذي الحدين السالب مسمى NB2 في الآونة الاخيرة طريقة في تحليل نماذج استجابة العد، اي في نمذجة المتغير المعتمد كونه متغير استجابة عندما تكون قيم ذلك المتغير بهيئة قيم معدودة (Count Data) ولكن عدد قليل نسبيا من الباحثين او الاحصائيين طبقوا على انواع من نماذج ذات الحدين السالب المتاحة.

3-2 الصيغة العامة لأنموذج انحدار ثنائي الحدين السالب

General Form of Negative Binomial Regression Model

يمكن التعبير عن أنموذج انحدار ثنائي الحدين السالب بالصيغة التالية [iv:pp2]:

$$\underline{Y} = e^{X\underline{\beta}+\underline{U}} \qquad (2-6)$$

إذ أن:

Y: موجه متغير الاستجابة ذي درجة (nx1).

X: مصفوفة المتغيرات التوضيحية ذات الدرجة ((n x (p+1)).

β: موجه المعلمات ذو الدرجة (P+1) x 1)).

U: موجه الاخطاء العشوائية (nx1).

n: حجم العينة. P: عدد المتغيرات التوضيحية.

2- 4 **مشكلة التعدد الخطي** *Multicollinearity Problem*

اصبحت مشكلة الارتباط الخطي المتعدد (Multicollinearity) معروفة منذ ان اكتشفها العالم Fisher عام (1934) عند دراسته لسلسلة زمنية تشمل عدة متغيرات وقد تبين بعد ذلك مدى خطورتها ومقدار دقتها. ان مشكلة التعدد الخطي تظهر فقط عندما تكون هناك علاقة خطية بين بعض أو كل المتغيرات التوضيحية وان الارتباطات بين هذه المتغيرات يعرف بالتعدد الخطي، بحيث يصبح من الصعب فصل اثر كل متغير عن متغير الاستجابة في الواقع التطبيقي، كما تحدث مشكلة التعدد الخطي حينما تكون قيمة احد متغيرات التوضيحية متساوية لكافة المشاهدات، او عندما تعتمد قيمة احد المتغيرات التوضيحية على واحد او اكثر من المتغيرات التوضيحية الاخرى في الأنموذج المدروس[ii:143] . بشكل عام مشكلة التعدد الخطي تكون على نوعين هما:

2-4-1 التعدد الخطي التام (Perfect Multicollinearity)

في مثل هذه الحالة محدد مصفوفة المعلومات او مصفوفة فيشر (Fisher Matrix) مساوية للصفر $|X'X| = 0$ فانه لايمكن إيجاد معكوس مصفوفة المعلومات وبالتالي لاتوجد إمكانية لتقدير معلمات الأنموذج ($\widehat{B}$) ولغرض معالجة هذه المشكلة لابد من حذف المتغيرات التوضيحية المسببة للتعدد الخطي ومن ثم تقدير الأنموذج وابرز طرق المعالجة في هذه الحالة هو استخدام المركبات الرئيسية (Principle Components)[i:pp190].

2-4-2 التعدد الخطي شبه التام (Semi Multicollinearity)

في هذه الحالة قيمة محدد مصفوفة المعلومات لا يساوي الصفر وإنما قريب منه اي انها صغيرة جدا وتظهر هذه الحالة عندما تميل المتغيرات للتحرك سوية بالزيادة أو النقصان أو في حالة استخدام المتغيرات المرتدة زمنيا (*Lagged variables*) ففي هذه الحالة يمكن تقدير معلمات الأنموذج ولكن هذه التقديرات سوف تكون غير دقيقة وغير ممثلة لواقع المشكلة المدروسة بسبب تباين المعلمات المقدرة ستكون كبيرة جدا ، وبالتالي سيظهر اختبار أن عدم معنوية t للمعلمات في حين أنها في الواقع معنوية احصائياً، ولكن بناء الأنموذج يعجز عن إظهار اثر هذه المتغيرات بشكل منفصل نظرا لارتباط هذه المتغيرات بعضها ببعض وابرز طرق المعالجة في هذه الحالة طريقة انحدار الحرف Ridge Regression Method[i:pp 190].

2-5 **تقدير معلمات أنموذج انحدار ثنائي الحدين السالب**

Estimation of the Parameters for Negative Binomial Regression Model

هناك عدة طرق لتقدير معلمات أنموذج انحدار ثنائي الحدين السالب ومنها :

2-5-1 طريقة تقدير الامكان الاعظم Maximum likelihood Estimation Method (MLE)

ناقش تقديرات الإمكان الأعظم من قبل العالم Fisher عام (1941) لتوزيع ثنائي الحدين السالب على اساس الصيغة (2-4)[xiii:pp864] ووصف خصائص تقدير (MLE) لمعاملات أنموذج الانحدار ثنائي الحدين السالب قدرت عن طريق اخذ شرط من الدرجة الاولى وجعلها مساوية للصفر. اذ ان المتغير المعتمد y_i تتبع توزيع ثنائي الحدين السالب بمعلمتين (μ,θ) اذ (μ>0) ويمثل متوسط المتغير Y، θ هي معلمة التشتت، فتكون دالة الكثافة الاحتمالية لنموذج ثنائي الحدين السالب كالاتي والمذكورة سلفا في الصيغة (2-4)[xiii:pp1]:

$$p(y_i) = \frac{\Gamma(\theta^{-1} + y_i)}{\Gamma(1 + y_i)\Gamma\theta^{-1}} \left(\frac{\theta^{-1}}{\theta^{-1} + \mu_i}\right)^{\theta^{-1}} \left(\frac{\mu_i}{\theta^{-1} + \mu_i}\right)^{y_i} \qquad y_i = 0,1,2, \ldots$$

ومن خلال تعظيم المشاهدات لتوزيع المتغير (y_i) الواردة في الصيغة اعلاه تكون دالة الامكان بشكل الاتي (vi:pp3)(viii :pp D-4):

$$L(y_1, y_2, \dots y_n, \mu_i, \theta) = \prod_{i=1}^{n} p(y_i) \qquad (2-7)$$

$$= \prod_{i=1}^{n} \left[\frac{\Gamma(\theta^{-1} + y_i)}{\Gamma(1 + y_i)\Gamma\theta^{-1}} \left(\frac{\theta^{-1}}{\theta^{-1} + \mu_i}\right)^{\theta^{-1}} \left(\frac{\mu_i}{\theta^{-1} + \mu_i}\right)^{y_i} \right] \qquad (2-8)$$

وبأخذ اللوغاريتم الطبيعي لدالة الامكان للمشاهدات اعلاه نحصل على:

$$\log L(y_i / x_i, \beta, \theta) = \sum_{i=1}^{n} \left[(\log\Gamma(\theta^{-1} + y_i) - \log\Gamma(1 + y_i) - \log\Gamma\theta^{-1}) + \theta^{-1}\log\left(\frac{\theta^{-1}}{\theta^{-1} + \mu_i}\right) + y_i \log\left(\frac{\mu_i}{\theta^{-1} + \mu_i}\right) \right] \qquad (2-9)$$

ويمكن التعبير عن الحد الاول في الصيغة (2-9) بالشكل الاتي:

$$\log\left[\frac{\Gamma(\theta^{-1} + y_i)}{\Gamma(\theta^{-1})}\right] = \sum_{j=0}^{yi-1} \log(j + \theta^{-1}) \qquad (2-10)$$

وبالاعتماد على الفروض الاساسية لنموذج انحدار ثنائي الحدين السالب $\mu_i = e\{x_i'\beta\}$ يتم التعويض هذا الافتراض بالصيغة رقم (2-9) اعلاه كما يلي (ix:pp179):-

$$\log L(y_i/x_i, \beta, \theta) = \sum_{i=1}^{n} \left[\left\{ \sum_{j=0}^{yi-1} \log(j + \theta^{-1}) \right\} - \log(y_i\ !) + \theta^{-1}\log\left(\frac{1}{1+\theta e^{x_i'\beta}}\right) + y_i \log\left(\frac{\theta e^{x_i'\beta}}{1+\theta e^{x_i'\beta}}\right) \right] \quad (2-11)$$

وباشتقاق الصيغة (2-11) بالنسبة لمواجه المعلمات (ß) نحصل على :-

$$S(\beta) = \frac{\partial Log L(\theta, \beta)}{d\beta} = \sum_{i=1}^{n} \frac{(yi - e^{x_i'\beta})}{1 + \theta\ e^{x_i'\beta}} X_i \qquad (2-12)$$

تقديرات موجه معلمات ß لنموذج ثنائي الحدين السالب يمكن حصول عليها من خلال مساواة مشتقة دالة لوغارتيم الامكان الاعظم للصفر في معادلة رقم (2-12):-

$$\sum_{i=1}^{n} \frac{(y_i - e^{x_i'\beta})}{1 + \theta e^{x_i'\beta}} X_i = 0 \rightarrow (2-13)$$

في الخطوة الاخيرة يتبين التكرار الطبيعي للخوارزمية وهذه الطريقة تعرف بخوارزمية المربعات الصغرى التكرارية الموزونة **(Iteratively Re–Weighted Least Squares Algorithm) (IRLS)** إذ تعطي مقدرات المعلمات (β) لأنموذج انحدار ثنائي الحدين السالب حسب الصيغة التالية (x:pp2):

$$\hat{\beta}_{MLE} = (X'\widehat{W}X)^{-1}(X'\widehat{W}\underline{\hat{Z}}) \qquad (2-14)$$

$\hat{\beta}_{MLE}$: موجه معلمات انموذج انحدار ثنائي الحدين السالب المقدرة وفق طريقة الامكان الاعظم.

$\hat{w}$: مصفوفة قطرية عناصر القطر فيها تساوي القيم المقدرة لمعلمة توزيع ثنائي الحدين السالب ($\hat{\mu}_i$) (x:pp4)

$$\hat{w} = \begin{bmatrix} e^{\hat{\mu}} & \cdots & 0 \\ \vdots & e^{\hat{\mu}} & \vdots \\ 0 & \cdots & e^{\hat{\mu}} \end{bmatrix}$$

$$\underline{\hat{z}} = \log(\hat{\mu}_i) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}$$

ومصفوفة التباين المشترك لمقدرات الإمكان الأعظم لأنموذج انحدار ثنائي الحدين السالب تكون كالاتي:

$$Cov\left(\underline{\hat{\beta}}_{ML}\right) = [E\{\frac{\partial^2 L(X;\beta)}{\partial\beta\partial\beta'}\}]^{-1} = \sigma_u^2 (X'\widehat{W}X)^{-1} \quad (2-15)$$

اذ ان:

σ_u^2 : تباين الخطأ العشوائي للمجتمع.

وبذلك يكون متوسط مربعات الخطأ لمعلمات أنموذج انحدار ثنائي الحدين السالب المقدرة وفق طريقة الإمكان الأعظم كما يلي (iv:pp7)(ix:pp179)**:**

$$MSE\left(\hat{\beta}_{ML}\right) = E(\underline{\hat{\beta}}_{ML} - \underline{\beta})'\left(\underline{\hat{\beta}}_{ML} - \underline{\beta}\right) = tr(X'\widehat{W}X)^{-1} = \sum_{i=1}^{P} \frac{1}{\lambda_j} \quad (2-16)$$

اذ ان :

λ_j : هي القيمة المميزة للعنصر (j_{th}) للمصفوفة $(X'\widehat{W}X)$.

عند النظر في متوسط مربعات الخطأ لطريقة الامكان الاعظم فانه يمكن بسهولة ملاحظة أنها جاءت متضخمة فالقيم المميزة ستكون صغيرة عندما مصفوفة ($X\widehat{W}X$) تكون (ill-conditional) اي بمعنى اخر الارتباط بين المتغيرات التوضيحية تكون عالية تقود الى عدم ثبات واستقرار المقدرات بالتالي تاتي التباين المتضخم لمعلمات المقدرة وفق طريقة الامكان الاعظم عند وجود مشكلة التعدد الخطي في هذا الوضع يصبح من الصعب جدا تفسير المعلمات.

2-5-2 طريقة مقدرات إنحدار الحرف Ridge Regression Estimators Method

تم التطرق الى طريقة انحدار حرف لأول مرة من قبل (Hoerl & Kennard) عام (1970a) اذا ثبت في الوقت الحاضر بانها الطريقة الاكثر كفاءة لبيان كيفية التعامل مع المشاكل العامة التي تسببها مشكلة التعدد الخطي، اذ تتلخص هذه طريقة بإضافة كمية صغيرة موجبة الى عناصر قطر مصفوفة المعلومات ($X'X$) حسب اقتراح الباحثان (Hoerl & Kennard)، ويعتبر اسلوب انحدار حرف (RR) احد بدائل طرق التقدير عندما يكون هناك تعدد خطي بين المتغيرات التوضيحية، الميزة الرئيسية لطريقة انحدار حرف (RR) هو ان متوسط مربعات الخطأ (MSE) يصغر، لهذا تم اختيار مقدر ليكن ($\underline{B}_{RR}$) فيمكن كتابة مجموع مربعات الاخطاء الموزون على النحو التالي (ix:pp179)**:**

$$u'u = (\underline{y} - \widehat{\beta})'\left(\underline{y} - \widehat{\beta}\right) = (\underline{y} - X\underline{\beta}_{ML})'\left(\underline{y} - X\underline{\beta}_{ML}\right) + (\underline{\beta}_{RR} - \underline{\beta}_{ML})'\left(X'\widetilde{W}\left(\underline{\beta}_{ML}\right)X\right)\left(\underline{\beta}_{RR} - \underline{\beta}_{ML}\right)$$
$$= \emptyset_{min} + \emptyset\left(\underline{\beta}_{RR}\right) \quad (2-17)$$

اذ ان:

$\emptyset_{min}$: تمثل الزيادة في متوسط مربعات الأخطاء الموزون في حال استبدال المعلمات المقدرة بطريقة الإمكان الأعظم (β_{ML}) بالمعلمات المزمع ايجادها (β_{RR})، ويلاحظ زيادة متوسط مربعات الخطأ (MSE) أي تضخم التباين لمجموع مربعات الاخطاء الموزونة عندما استخدم طريقة الامكان الاعظم (ML) لتقدير معلمات الأنموذج β_{ML} لهذا تم استبداله بمقدر انحدار الحرف (RR). وان مقدر انحدار الحرف لأنموذج ثنائي الحدين السالب يمكن ايجاده عن طريق تقليل طول $\hat{\beta}'\hat{\beta}$ من خلال تطبيق الشرط التالي:

$$\emptyset\left(\underline{\beta}_{RR}\right) = \emptyset_0 \qquad (2-18)$$

ويمكن القول بان استعمال مضاعف لاكرانج فإن موجه المعلمات المزمع احتسابه β_{RR} سيصغر مجموع مربعات الاخطاء المؤزون وفقا للقيد التالي (ix:pp180):

$$Minimize\ F = \underline{\beta}'_{RR}\underline{\beta}_{RR} + (\frac{1}{K})(\underline{\beta}_{RR} - \underline{\beta}_{ML})'(X'\widehat{W}(\underline{\beta}_{ML})X)\left(\underline{\beta}_{RR} - \underline{\beta}_{ML} - \emptyset_0\right) \quad (2-19)$$

اذ ان $(1/k)$ هو مضاعف لاكرانج، وبذالك سيكون مجموع مربعات الأخطاء المؤزون:

$$u'u = (\underline{y} - X\underline{\beta}_{ML})'\left(\underline{y} - X\underline{\beta}_{ML}\right) + \underline{\beta}'_{RR}\underline{\beta}_{RR} + (\frac{1}{K})(\underline{\beta}_{RR} - \underline{\beta}_{ML})'\left(X'\widehat{W}\left(\underline{\beta}_{ML}\right)X\right)\left(\underline{\beta}_{RR} - \underline{\beta}_{ML} - \emptyset_0\right)$$

(2-20)

وباشتقاق الصيغة اعلاه بالنسبة لموجه المعلمات β_{RR} وبمساواة المشتقة للصفر، يمكن ايجاد مقدرات معلمات أنموذج انحدار الثنائي الحدين السالب وكما يلي:

$$\frac{\partial u'u}{\partial \beta_{RR}} = 2\underline{\beta}_{RR} + \left(\frac{1}{K}\right)\left(2X'\widehat{W}\left(\underline{\beta}_{ML}\right)X\underline{\beta}_{RR} - 2X'\widehat{W}(\underline{\beta}_{ML})X\underline{\beta}_{ML}\right) = 0 \qquad (2-21)$$

واخيراً نحصل على مقدر انحدار حرف لأنموذج ثنائي الحدين السالب بحل معادلة اعلاه (xiv:pp180):

$$\underline{\hat{\beta}}_{RR} = (X'\widehat{W}\left(\underline{\beta}_{ML}\right)X + KI)^{-1}\left(X'\widehat{W}\left(\underline{\beta}_{ML}\right)X\underline{\beta}_{ML}\right) \qquad (2-22)$$

$$\underline{\hat{\beta}}_{RR} = (X'\widehat{W}X + KI)^{-1}(X'\widehat{W}X)\underline{\hat{\beta}}_{ML} = Z\underline{\hat{\beta}}_{ML} \qquad (2-23)$$

اذ ان :

$$Z = (X'\widehat{W}X + KI)^{-1}(X'\widehat{W}X) \qquad (2-24)$$

I: هو مصفوفة احادية ذي الدرجة (p+1) x (p+1).

K : هو ثابت غير سالب يسمى المتحيز او معلمة حرف، مع ملاحظة عندما K=0 نحصل على تقديرات المربعات الصغرى الاعتيادية (OLS)، اذ أن مقدرات انحدار الحرف متحيزة بتزايد قيمة K، اذ انها تعطي تقديرات اكثر دقة من مقدرات المربعات الصغرى الاعتيادية (OLS) لموجه المعلمات.

وبعد ان تم ايجاد مقدرات حرف تبين ان تلك المقدرات متحيزة لكنها اكفأ من نظيرتها المقدرة باستعمال طريقة الامكان الاعظم (ML) اذ ان تناقص معاملات القيم الموجبة الصغيرة من K تحسن تكييف المشكلة وتقلل تباين المقدرات على الرغم من التحيز. كما أن مصفوفة التباين والتباين المشترك لمقدرات انحدار الحرف تكون كما يلي:

$$Var - Cov\left(\underline{\hat{\beta}}_{RR}\right) = Z\ \ Var - Cov\left(\underline{\hat{\beta}}_{ML}\right)\ \ Z'$$
$$= Z\ \sigma_u^2\ \ (X'\widehat{W}X)^{-1}\ \ Z' = \sigma_u^2\ \ Z\ \ (X'\widehat{W}X)^{-1}\ \ Z' \qquad (2-25)$$

اذ بين (Mansson & Shukur 2011) ان متوسط مربعات الخطأ (MSE) لمعلمات أنموذج الثنائي الحدين السالب المقدرة يساوي (5:pp2)(ix:PP180):

$$MSE(\hat{\beta}_{RR}) = \sum_{j=1}^{J} \frac{\lambda_j}{(\lambda_j + K)^2} + k^2 \sum_{j=1}^{J} \frac{\alpha_j^2}{(\lambda_j + k)^2} = \gamma_1(k) + \gamma_2(k) \qquad (2-26)$$

اذ α_j تعرف كنعصر j_{th} من معلمات أنموذج انحدار ثنائي الحدين السالب عند اخذ اللوغارتيم للأنموذج المبين في الصيغة (2-25)، وتساوي المقدار $\gamma\underline{\hat{\beta}}_{ML}$ ، اي ان:

$$\alpha_j = \gamma\underline{\hat{\beta}}_{ML} \qquad (2-27)$$

γ : تمثل المتجه المميز (*Eigen Vector*) للمصفوفة $X'\widehat{W}X$.

2- 6 مقدرات معلمة حرف Ridge Parameter Estimators

اقترح (Hoerl & Kennard) مقدر انحدار حرف كبديل لمقدر مربعات الصغرى الاعتيادية (OLS) في وجود مشكلة تعدد خطي. واحد من العقبات الرئيسية في استخدام الانحدار حرف في اختيار قيمة مناسبة لمعلمة التحيز K وقد تم تطوير العديد من التقنيات لتقدير معلمة حرف K وتم استعراضها وتصنيفها الى اشكال مختلفة واقترح اساليب جديدة لتقدير معلمة انحدار حرف (معلمة التحيز)(ix:pp180):

1- صيغة Hoerl & Kennard عام 1970:

تعتبر هذه الصيغة اول واقدم المقترحات لتقدير معلمة حرف K اذ اقترحوا تقدير معلمة حرف والتي ينتج مع اقل متوسط مربعات الخطأ عن طريق تعظيم الثابت (α_i^2) باستخدام متوسط مربعات الخطأ اقترحت معلمة الحرف الامثل وكالاتي:

$$\hat{k}_{HK}^{FM} = \frac{\hat{\sigma}^2}{Max(\hat{\alpha}_i^2)} \qquad i = 1,2,\dots,p \qquad (2-28)$$

اذ ان:

$\hat{\sigma}^2$: متوسط مربعات الخطأ والذي يحسب من أنموذج انحدار ثنائي السالب وفق الصيغة ادناه:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-p-1} \qquad (2-29)$$

$Max(\hat{\alpha}_i^2)$: تمثل اكبر عنصر في الصيغة (2-27).

$\sum_{i=1}^{n} e_i^2$: تمثل مجموع مربعات الاخطاء العشوائية.

n : تمثل حجم العينة.

p : تمثل عدد المتغيرات مستقلة .

2- صيغة Hoerl واخرون عام (1975) .

اقترح Hoerl وآخرون تقدير مختلف لمعلمة التحيز k من خلال اخذ المتوسط التوافقي لمعلمة حرف k_{HK} وحسب الصيغة التالية:

$$\hat{k}_{HK}^{HM} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{n} \alpha_i^2} \qquad (2-30)$$

3- صيغة kibria عام 2003:

اقترح kibria بعض المقدرات الجديدة لمعلمة التحيز K من خلال اخذ المتوسط الهندسي، المتوسط الحسابي، والوسيط $(P \geq 3)$ *لمعلمة* k_{HK}، هذه المقدرات على التوالي تعطى وفق الصيغ التالية:

$$\hat{k}_{HK}^{GM} = \frac{\hat{\sigma}^2}{(\prod_{i=1}^{p} \hat{\alpha}_i^2)^{\frac{1}{p}}} \qquad (2-31)$$

$$\hat{k}_{HK}^{AM} = \frac{1}{p}\sum_{i=1}^{p}(\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}) \qquad (2-32)$$

$$\hat{k}_{HK}^{M} = \text{Median}\left(\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}\right) \qquad (2-33)$$

4- صيغة Muniz & Kibria عام 2009:

اقترح Muniz & Kibria بعض المقدرات للمعلمة k، والصيغ هي الجذر التربيعي للوسط الهندسي لمعلمة k_{HK}، ومعكوس الوسط الهندسي لمعلمة k_{HK}، والجذر التربيعي للوسيط لمقدر (Hoerl & Kennard)، ومعكوس الجذر التربيعي للوسيط لمقدر (Hoerl & Kennard)،ووفق الصيغ التالية:

$$\hat{k}_{HK}^{GMSR} = \sqrt{\frac{\hat{\sigma}^2}{(\prod_{i=1}^{P} \hat{\alpha}_i^2)^{\frac{1}{P}}}} \qquad (2-34)$$

$$\widehat{K}_{HK}^{GMRSR} = \frac{1}{\sqrt{\frac{\hat{\sigma}^2}{(\prod_{i=1}^{P} \hat{\alpha}_i^2)^{\frac{1}{P}}}}} \qquad (2-35)$$

$$\widehat{K}_{HK}^{MSR} = Median\left(\sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}}\right) \qquad (2-36)$$

$$\hat{k}_{HK}^{MRSR} = Median\left(\frac{1}{\sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}}}\right) \qquad (2-37)$$

5- صيغة مقدرات مستندة على مقترح Lawless& Wang عام 1976:

في تحسين للمقترح الذي اقترحه وقدمه الباحثون (Hoerl واخرون عام (1975))، اقترح الباحثان Lawless & Wang مقدر مختلف للمعلمة K ناتجة من اخذ الوسط التوافقي لمعلمة الحرف ومعرفة كالاتي:

$$\widehat{K}_{LW}^{HM} = \frac{p\hat{\sigma}^2}{\sum_{j=1}^{p} \lambda_j \hat{\alpha}_j^2} \qquad (2-38)$$

اذ أن:

λ_j : تمثل القيم المميزة (Eigen Values) للمصفوفة $X'\widehat{W}X$.

6- **صيغة AL Khamisi واخرون عام 2006:**

ربط هذا المقدر بين القيم المميزة وتباين الاخطاء العشوائية فضلاً عن الأخذ بالاعتبار أثر الموجهات المميزة عبر احتساب اكبر قيمة، وإن القيمة المثلى لمعلمة التحيز عند تحليل أنموذج انحدار ثنائي الحدين السالب هي:

$$\widehat{K}_{AKS_i}^{FM} = Max\{s_j\} \qquad (2-39)$$

إذ أن:

$$S_j = \frac{\lambda_j \hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + \lambda_j \hat{\alpha}_j^2} \qquad (2-40)$$

7- **صيغة Muniz & Kibria عام 2009:**

اقترح Muniz & Kibria مقدر معلمة حرف K كمتوسط هندسي لمعلمة حرف K_{AKS_i} التي قدمها (AL Khamisi واخرون) ومعرفة كالاتي:

$$\widehat{K}_{AKS}^{GM} = \left(\prod_{j=1}^{p} \frac{\lambda_j \hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + \lambda_j \hat{\alpha}_j^2}\right)^{\frac{1}{p}} \qquad (2-41)$$

3-5-2 طريقة مقدرات ليو Liu Estimators Method

هذه الطريقة تتضمن مقدر ليو الذي استخدم لمعالجة المشاكل التي تسببها مشكلة تعدد الخطي، ومن اجل حل مشكلة تضخم متوسط مربعات الخطأ للطريقة الكلاسيكية الامكان الاعظم (ML)، اقترح هذا المقدر من قبل ليو (1993) و ليو (2003) وتم تطبيق هذه الاسلوب في حالة النماذج الخطية المتعددة، اذ تحدث مشكلة التعدد الخطي عندما المتغيرات التوضيحية لنموذج الانحدار ترتبط بشكل كبير اذ تؤدي الى عدم استقرار المعايير التي تقدرها طريقة الامكان الاعظم ML [v:pp7] وتضخم التباين. اذ افترض ان اداء مقدر ليو $\hat{\beta}_d$ افضل من مقدر الامكان الاعظم $\hat{\beta}_{ML}$ عندما وجود التعدد الخطي بين المتغيرات التوضيحية، وقد تظهر هذه من خلال دراسة خصائص متوسط مربعات الخطأ، اذ ادعى ان هذا المقدر يقلل من متوسط مربعات الخطأ (MSE)، ويعرف مقدر ليو لأنموذج انحدار ثنائي الحدين السالب كالاتي[x :pp1)(xi:pp2]:

$$\underline{\hat{\beta}}_{Liu} = (X'\widehat{W}X + I)^{-1}\left((X'\widehat{W}X)\underline{\beta}_{ML} + d\underline{\hat{\beta}}_{ML}\right)$$

$$\underline{\hat{\beta}}_{Liu} = (X'\widehat{W}X + I)^{-1}(X'\widehat{W}X + dI)\underline{\hat{\beta}}_{ML} = Z\,\underline{\hat{\beta}}_{ML} \qquad (2-42)$$

إن مقدرات (Liu) تعد متحيزة عندما $(d > 0)$ ومقدار التحيز يكون[iv:PP8]:

$$Bias\left(\underline{\hat{\beta}}_{Liu}\right) = E\left(\underline{\hat{\beta}}_{Liu}\right) - \underline{\beta} = Z\underline{\beta} - \underline{\beta} = (Z-I)\underline{\beta} \qquad (2-43)$$

اذ أن :

$$Z = (X'\widehat{W}X + I)^{-1}(X'\widehat{W}X + dI) \qquad (2-44)$$

ومصفوفة التباين والتباين المشترك لمقدرات (Liu) تعرف كالاتي :

$$Var - Cov\left(\hat{\underline{\beta}}_{Liu}\right) = Z\ Var - Cov\left(\hat{\underline{\beta}}_{ML}\right)\ Z'$$
$$= \sigma_u^2\ Z\,(X'\widehat{W}X)^{-1}Z' \qquad (2-45)$$

أما متوسط مربعات الخطأ (MSE) لمعلمات أنموذج انحدار ثنائي الحدين السالب المقدرة وفق طريقة مقدرات (Liu) كالاتي [(xi:pp3)(x:pp3)]**:**

$$MSE\left(\hat{\underline{\beta}}_{LE}\right) = \mathrm{E}(\hat{\underline{\beta}}_{Liu} - \underline{\beta})'(\hat{\underline{\beta}}_{Liu} - \underline{\beta})$$
$$= \sum_{j=1}^{p}\frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j+1)^2} + (d-1)^2\sum_{j=1}^{p}\frac{\alpha_j^2}{(\lambda_j+1)^2} \qquad (2-46)$$

اذ أن:

λ_j**: تمثل القيم المميزة (Eigen Values) للمصفوفة** $X'\widehat{W}X$.

مقدرات معلمة ليو المتحيزة Liu Biased Parameter Estimators

مقدرات ليو متحيزة، اذ ان سبب التحيز هو وجود قيمة مضافة (d)، وان تلك القيمة تتراوح بين (1،0)، ولايجاد المقدرات المقترحة لمعلمة التحيز (d) يجب ايجاد القيمة المثلى لها وذلك من خلال مساواة قيمة المشتقة لمتوسط مربعات الخطأ المبينة في الصيغة (2-42) بالنسبة الى (d) ومساواتها بالصفر[(vii:PP3)]**:**

$$\frac{\partial MSE\left(\hat{\underline{\beta}}_{Liu}\right)}{\partial d} = 0$$

$$2\frac{\lambda_j + \hat{d}}{\lambda_j(\lambda_j+1)^2} + 2\left(\hat{d} - 1\right)\frac{\alpha_j^2}{(\lambda_j+1)^2} = 0 \qquad (2-47)$$

وبإجراء العمليات الجبرية نحصل على:

$$\hat{d} = \frac{\dfrac{\alpha_j^2 - 1}{(\lambda_j+1)^2}}{\dfrac{\dfrac{1}{\lambda_j} + \alpha_j^2}{(\lambda_j+1)^2}} = \frac{\alpha_j^2 - 1}{\dfrac{1}{\lambda_j} + \alpha_j^2} \qquad (2-48)$$

ويتم عرض القيمة المثلى للمعلمة التحيز (d) للقيم المختلفة من المعلمة α_j^2 **والقيم المميزة** λ_j **يمكن ملاحظة بوضوح بان القيمة المثلى سالبة عندما تكون قيمة المعلمة** α_j^2 **اقل من الواحد، وموجبة عندما تكون قيمته اكبر من الواحد، وكلما زادت قيمة معلمة** α_j^2 **تكون القيمة المثلى كبيرة ايضاً، وكما اشار (Liu) بأن تلك القيمة لمعلمة التحيز تقع ضمن المجال (0,1)، وبذلك لاتوجد قاعدة محددة لتقديرها، وهناك عدة طرق مختلفة لتقدير معلمة التحيز (d) عن طريق ايجاد قيمة واحدة لتلك المعلمة وكما يلي**[(x:PP4)(xi:PP4)]**:**

1- الصيغة الاولى: وتعطى كالآتي:

$$d_1 = Max\left[0, \frac{\hat{\alpha}_{max}^2 - 1}{\frac{1}{\lambda_j} + \hat{\alpha}_{max}^2}\right] \qquad (2-49)$$

اذ $\hat{\alpha}_{max}^2$ تعرف كنعصر تعظيم ل $\hat{\alpha}_j^2$ و λ_{max} تمثل اكبر قيمة مميزة في المصفوفة $X'\hat{W}X$.
وباستبدال قيم المعلمات المجهولة مع هذه القيمة لتعظيم المقدرات غير المتحيزة هي فكرة ماخوذة من (Hoerl & Kennard) وتتضمن إضافة قيمة موجبة لعناصر قطر المصفوفة $(X'\hat{W}X)$ لأنموذج الانحدار الخطي.

2- الصيغة الثانية:

يعتمد هذا المقترح على إيجاد قيمة الوسيط باعتباره احد مقاييس النزعة المركزية بالنسبة لمقدر معلمة التحيز لأنه يتعامل مع القيم الشاذة في البيانات التي قد تسبب التحيز، وتعطى كالآتي:

$$d_2 = Max\left[0, median\left[\frac{\alpha_j^2 - 1}{\frac{1}{\lambda_j} + \hat{\alpha}_j^2}\right]\right] \qquad (2-50)$$

3- الصيغة الثالثة

فكرة هذا المقترح تأتي من ذات التطبيق لأنموذج الانحدار الخطي المتعدد عبر التعامل مع الوسط الحسابي كأبرز مقاييس النزعة المركزية ،فضلا عن أخذه بنظر اعتبار عدد المتغيرات التوضيحية في الأنموذج.

$$d_3 = Max\left[0, \frac{1}{p}\left[\sum_{j=}^{p} \frac{\hat{\alpha}_j^2 - 1}{\frac{1}{\lambda_j} + \hat{\alpha}_j^2}\right]\right] \qquad (2-51)$$

4- الصيغة الرابعة والخامسة:

أن مقدرات d_2 و d_3 لها نظائر في ((kibria(2003))، اذ أن المقترحان (4) و(5) يعتمدان على احتساب اكبر واصغر قيمة لمقدر معلمة التحيز على التوالي، واستخدم كذلك الكميات الاخرى مثل القيمة العظمى تم تطبيقها بنجاح من قبل ((Khalaf and Shukur (2005)) والفكرة الكامنة وراء المقدرين d_4 و d_5 ماخوذة منهم [xi:pp5].

$$d_4 = Max\left[0, max\left[\frac{\hat{\alpha}_j^2 - 1}{\frac{1}{\lambda_j} + \hat{\alpha}_j^2}\right]\right] \qquad (2-52)$$

$$d_5 = Max\left[0, min\left[\frac{\hat{\alpha}_j^2 - 1}{\frac{1}{\lambda_j} + \hat{\alpha}_j^2}\right]\right] \qquad (2-53)$$

المبحث الثالث/ الجانب التجريبي

3-1 المقدمة

في المبحث الثاني قدمت بعض المقارنات النظرية وهي واسعة النطاق، اما في هذا المبحث تم تصميم دراسة اسلوب المحاكاة مونت – كارلو لتقييم اداء المقدرات ولمقارنة طرائق تقدير معلمات أنموذج ثنائي الحدين السالب (NBM) وذلك بحساب المؤشر المقارنة متوسط مربعات الخطأ (MSE)، ويتم وصف موجز لكيفية إنشاء البيانات جنباً الى جنب مع مناقشة النتائج التي توصلنا اليها في المحاكاة.

3-2 المفهـــوم العــــــام للمحاكــــاة General Understanding of Simulation

المحاكاة هى عملية تقليد (محاكاة) لشيء أو ظرف ما حقيقى أو عملية واقعية، وهى تتضمن بصفة عامة بعض الخصائص الاساسية لسلوكيات النظام المادى أو المجرد المعنى به، وهى تستخدم فى العديد من السياقات المهمة التى تخدم الإنسان اذ تتضمن محاكاة نماذج النظم الطبيعية والإنسانية لتوسيع المدارك و الفهم حول وظائف تلك النظم، محاكاة التكنولوجيا من أجل التوصل إلى تحسين وملائمة الأداء على سبيل المثال لكل من الأمن الصناعى والإنشاءات والعمليات الهندسية، إختبار المواد، التدريب والتعليم والبحث العلمي ....الخ، كما تعمل على إظهار الأثار الحقيقية للمواقف والأوضاع البديلة أو المختلفة ومسارات العمل والأفعال وردود الأفعال المتوقعة .

3-3 وصف تجربة المحاكاة

تضمنت تجارب المحاكاة كتابة عدد من البرامج بلغة(MATLAB 2010)، اذ ان الأنموذج الذي اعتمد يكون وفق الصيغة (2-6) الواردة في الجانب النظري، وسيتم شرح المراحل الخاصة بتجربة المحاكاة بالكامل، اذ في المرحلة الاولى تم اخذ بنظر الاعتبار العامل الاول وهو حجم العينة كونه عاملاً مؤثراً متغيراً، اذ تم اختيار اربعة أحجام للعينات المفترضة وهي:

(n =25, 50, 100,250)

اما العامل (المتغير) الآخر الذي تم اخذه بنظر الاعتبار، هو معامل الارتباط البسيط اذ سيتم اخذ ثلاث قيم مختلفة لمدى قوة العلاقة بين المتغيرات المستقلة وهي (0.90,0.95,0.99) ومبينة في الجدول رقم (1).

بعد ذلك تم توليد متغير الخطأ العشوائي في أنموذج انحدار ثنائي الحدين السالب تبعا لتوزيع ثنائي الحدين السالب بمعلمة (μ_i , θ).

اما في المرحلة الثانية يتم انشاء او توليد المتغيرات التوضيحية التفسيرية واخضاعها الى علاقات ارتباط قوية فيما بينها بغية خلق مشكلة التعدد الخطي وفقا للصيغة التالية:

$$X_{ij} = (1-\rho^2)^{\frac{1}{2}} Z_{ij} + \rho Z_{ip} \qquad (3-1)$$

إذ أن :

n: تمثل عدد المشاهدات i=1,2,……………,n

P : تمثل عدد المتغيرات التوضيحية j=1,2,……….,p

ρ^2 : يمثل معامل الارتباط البسيط بين كل زوج من المتغيرات التوضيحية.

Z_{ij} : تمثل الارقام العشوائية المولدة وفق التوزيع الطبيعي القياسي.

سوف ندرس تأثير وجود ثلاث متغيرات مستقلة وخمس متغيرات مستقلة في عملية المقارنة بين طرائق التقدير معلمات أنموذج ثنائي الحدين السالب والجدول (1) يوضح العوامل التي تم اخذها بنظر الاعتبار.

في المرحلة الثالثة من تجربة المحاكاة نولد متغير معتمد لأنموذج انحدار ثنائي الحدين السالب باستخدام الارقام عشوائية التالية لتوزيع ثنائي الحدين السالب $NB(\mu_i, \mu_i + \theta\mu_i^2)$ وحسب أنموذج ثنائي الحدين السالب يتم حساب قيم متغير الأستجابة (Y_i) المذكور في الصيغة (6-2):

$$\underline{Y} = \mu_i = e^{X\underline{\beta}+\underline{U}}$$

إذ عرفنا سلفا قيمة (μ_i) وفق الصيغة (6-2) وعبر تعميمها اكثر لتكون كما يلي:

$$Y_i = \mu_i = \text{EXP}\{\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots.. + \beta_P X_{iP}\} \qquad (3-2)$$

في المرحلة الرابعة، يتم اختيار قيم معلمات أنموذج في الصيغة اعلاه كالاتي:

$$\sum_{j=1}^{p} \beta_j^2 = 1 \qquad (3-3)$$

والقيم الافتراضية التي تم اخذها لمعلمات الأنموذج هي:

$$\beta_1 = 0.67, \quad \beta_2 = 0.2\,, \quad \beta_3 = 0.3, \quad \beta_4 = 0.4, \quad \beta_5 = 0.5$$

هذا القيد شائع الاستخدام في دراسات المحاكاة كما في الدراسات السابقة لبعض الباحثين منهم Kibria(2003) [(v:pp182)].

جدول (1): العوامل الثلاث الموثرة وقيمها في تجربة المحاكاة

العامل	القيم
حجم العينة N	25, 50, 100, 250
معامل الارتباط البسيط ρ	0.90, 0.95, 0.99
عدد المتغيرات المستقلة p	3, 5

في المرحلة الخامسة تقدر معالم أنموذج ثنائي الحدين السالب على وفق طرائق التقدير المدروسة في الجانب النظري وهي:

1- طريقة مقدر الامكان الاعظم (Maximum Likelihood Estimator).
2- طريقة مقدر انحدار حرف (Ridge Regression Estimator).
3- طريقة مقدر ليو (Liu Estimator).

ولاجل اجراء المقارنة بين الطرائق التقدير المختلفة، سنعتمد على متوسط مربعات الخطأ(MSE) كمعيار لمقارنة متوسطات تقدير المعلمات (β_j) وذلك تبعا للصيغة ادناه:

$$MSE = \frac{\sum_{r=1}^{R}(\hat{\beta}-\beta)'_r(\hat{\beta}-\beta)_r}{R} \qquad r= 1, 2,...., 1000 \qquad (3-4)$$

$\hat{\beta}$: قيمة المعلمة المقدرة وفق طرائق التقدير المختلفة.

β : قيمة المعلمة في الصيغة (3-3).

R : عدد مرات تكرار التجربة والتي ستؤخذ لتكون مساوية 1000.

4-3 نتائج المحاكاة

يظهر الجدول (2) والجدول (3) نتائج تجربة المحاكاة والتي تم الحصول عليها عبر تنفيذ البرنامج لايجاد متوسط مربعات الخطأ (MSE) لكافة الطرائق من خلال المعادلة (2-14) لطريقة الامكان الاعظم وعبر التعويض المباشر للصيغ (2-28)، (2-29)، (2-30)، (2-31)، (2-32)، (2-33)، (2-34)، (2-35)، (2-36)، (2-37)، (2-38)، (2-39)، (2-40)، (2-41) في المعادلة (2-23) لطريقة انحدار الحرف، والتعويض المباشر للصيغ (2-49)، (2-50)، (2-51)، (2-52)، (2-53) في المعادلة (2-42) لطريقة مقدرات ليو، اذ تعكس النتائج قيم متوسط مربعات الخطأ (MSE) لكافة طرائق التقدير السابقة.

<u>الحالة الاولى: عدد المتغيرات (P=3):</u>

من خلال الجدول (2) نلاحظ ما يلي:

1- عند حجم العينة (25) ولقيمة الارتباط (ρ=0.90, 0.95) نلاحظ ان طريقة صيغة تقدير انحدار الحرف (K_6) تمتلك اقل متوسط مربعات الخطأ مقارنة مع الطرائق الاخرى، بينما لقيمة ($\rho = 0.99$) ظهر ان تقدير انحدار الحرف (K_3) بصيغتها الاولى تمتلك اقل متوسط مربعات الخطأ (MSE) مقارنة مع الطرائق الاخرى.

2- عند حجم العينة (50) ولقيمة ($\rho = 0.90$) نلاحظ ان صيغة تقدير انحدار الحرف (K_6) تمتلك اقل متوسط مربعات الخطأ (MSE) مقارنة مع الطرائق الاخرى، ولقيمة ($\rho = 0.95$) ظهر ان مقدر انحدار الحرف (K_3) بصيغته الثانية يمتلك اقل متوسط مربعات الخطأ (MSE)، بينما لقيمة ($\rho = 0.99$) ظهر ان مقدر انحدار الحرف (K_4) بصيغته الثانية يمتلك اقل متوسط مربعات الخطأ (MSE) مقارنة مع الطرائق الاخرى.

3- عند حجم العينة (100) ولقيم (ρ=0.90, 0.95) نلاحظ صيغة تقدير انحدار الحرف (K_4) بصيغته الثانية يمتلك اقل متوسط مربعات الخطأ (MSE) مقارنة مع الطرائق الاخرى، ومقدر انحدار الحرف (K_6) تمتلك اقل متوسط مربعات الخطأ (MSE) عند قيمة الارتباط ($\rho = 0.99$).

4- عند حجم العينة (250) ولجميع قيم ρ نلاحظ مقدر انحدار الحرف (K_4) بصيغتها الثانية تمتلك اقل متوسط مربعات الخطأ (MSE) مقارنة مع الطرائق الاخرى.

الحالة الثانية: عدد المتغيرات (P=5):

من خلال الجدول (3) نلاحظ ما يلي:

1- عند حجم العينة (25) نلاحظ لقيمتي الارتباط ($\rho = 0.90, 0.99$) ان صيغة مقدر انحدار الحرف (K_6) تمتلك اقل متوسط مربعات الخطأ مقارنة مع طرائق التقدير الاخرى، بينما لقيمة الارتباط ($\rho = 0.95$) ان الصيغة الاولى لمقدر انحدار الحرف (K_3) تمتلك اقل متوسط مربعات الخطأ مقارنة مع طرائق التقدير الاخرى.

2- عند حجم العينة (50) نلاحظ لقيمتي الارتباط ($\rho = 0.90,0.95$) ان مقدر انحدار الحرف (K_4) بصيغتها الثانية تمتلك اقل متوسط مربعات الخطأ (MSE) مقارنة ببقية صيغ تقدير والطرائق الاخرى. بينما لقيمة الارتباط ($\rho = 0.99$) ظهرت صيغة مقدر انحدار الحرف (K_6) تمتلك اقل متوسط مربعات الخطأ مقارنة ببقية صيغ تقدير وطرائق الاخرى.

3- عند حجم العينة (100) ولقيمة ($\rho = 0.90$) نلاحظ ان صيغة تقدير انحدار الحرف (K_4) الثانية تمتلك اقل متوسط مربعات الخطأ (MSE) مقارنة مع الطرائق الاخرى، ولقيمة الارتباط ($\rho = 0.95$) ظهر ان صيغ طريقة مقدر ليو (Liu-Estimator) تمتلك اقل متوسط مربعات الخطأ (MSE) مقارنة مع الطرائق الاخرى، بينما لقيمة ($\rho = 0.99$) ظهر ان صيغة تقدير انحدار الحرف تقدير انحدار الحرف (K_6) تمتلك اقل متوسط مربعات الخطأ (MSE).

4- عند حجم العينة (250) نلاحظ لقيمتي ($\rho = 0.90,0.95$) ان صيغة التقدير الثانية لـ (K_4) من مقدرات انحدار الحرف تمتلك اقل متوسط مربعات الخطأ (MSE)، ولقيمة الارتباط ($\rho = 0.99$) تمتلك (K_6) اقل متوسط مربعات الخطأ (MSE) مقارنة ببقية صيغ تقدير لمقدر انحدار الحرف وطرائق الاخرى.

جدول (2) قيم متوسط مربعات الخطأ لكافة طرائق التقدير عندما يكون عدد المتغيرات المستقلة (P=3)

n	ρ	EML	Ridge Estimator							Liu Estimator				
			K_1	K_2	K_3	K_4	K_5	K_6	K_7	d_1	d_2	d_3	d_4	d_5
25	0.90	0.132667	0.131080	0.129104	0.0981971 0.1316350 0.1326660	0.111137 0.125175 0.132666 0.1326666	0.132462	0.094232	0.121176	0.119790	0.119790	0.119790	0.119790	0.119790
	0.95	0.142574	0.140239	0.138667	0.123921 0.140523 0.142574	0.126371 0.130431 0.142574 0.142575	0.142307	0.121994	0.132731	0.128533	0.128533	0.128533	0.128533	0.128533
	0.99	0.165738	0.162898	0.1607674	0.130775 0.163355 0.165737	0.13982397 0.15375174 0.16573755 0.16573700	0.165334	0.140701	0.153537	0.147972	0.1479721	0.1479721	0.1479721	0.1479721
50	0.90	0.052909	0.053638	0.0527306	0.05038125 0.05285750 0.05290800	0.05073988 0.05131605 0.05290878 0.05290800	0.052904	0.050011	0.051861	0.051049	0.0510493	0.0510493	0.0510493	0.0510493
	0.95	0.056665	0.057455	0.056469	0.05381299 0.05660685 0.05666500	0.05425988 0.05496041 0.05666497 0.056665	0.056659	0.054925	0.055580	0.054639	0.054639	0.054639	0.054639	0.054639
	0.99	0.065383	0.066224	0.065137	0.062956 0.065294 0.065383	0.062928 0.062872 0.065383 0.0653834	0.065373	0.063041	0.064283	0.062900	0.062900	0.062900	0.062900	0.062900

تابع جدول (2)

n	ρ	MLE	Ridge Estimator											Liu Estimator
			K_1	K_2	K_3	K_4	K_5	K_6	K_7	d_1	d_2	d_3	d_4	d_5
100	0.90	0.026989	0.026978	0.026963	0.026650 0.026982 0.026989	0.026576 0.026376 0.026989 0.026987	0.026989	0.026741	0.026773	0.026486	0.026486	0.026486	0.026486	0.026486
	0.95	0.029078	0.029064	0.029049	0.0286835 0.0290684 0.029078	0.028610 0.028422 0.029078 0.029076	0.029077	0.028719	0.028846	0.028524	0.028524	0.028524	0.028524	0.028524
	0.99	0.033909	0.0338905	0.033872	0.0333855 0.0338953 0.0339080	0.0333070 0.0331130 0.0339087 0.0339090	0.033907	0.032913	0.033659	0.033217	0.033217	0.033217	0.033217	0.033217
250	0.90	0.0105694	0.0105681	0.010568	0.0105115 0.0105687 0.0105695	0.010503 0.0104825 0.010569 0.010568	0.010569	0.0105658	0.0105531	0.010494	0.010494	0.010494	0.010494	0.010494
	0.95	0.011338	0.011337	0.011336	0.0113276 0.0113373 0.011339	0.0113087 0.0111167 0.0113384 0.011338	0.0113383	0.011299	0.0113285	0.011257	0.011257	0.011257	0.011257	0.011257
	0.99	0.0131231	0.0131214	0.0131206	0.0131124 0.0131214 0.0131231	0.0130903 0.0128206 0.0131231 0.0131123	0.0131231	0.0131077	0.0131132	0.013023	0.013023	0.013023	0.013023	0.013023

جدول (3) قيم متوسط مربعات الخطأ لكافة طرائق التقدير عندما يكون عدد المتغيرات المستقلة (P=5)

n	ρ	MLE	Ridge Estimator							Liu Estimator				
			K_1	K_2	K_3	K_4	K_5	K_6	K_7	D_1	D_2	D_3	D_4	D_5
25	0.90	0.161425	0.160264	0.160222	0.137361 0.161101 0.161425	0.143415 0.151632 0.161425 0.161427	0.1612993	0.122772	0.1510586	0.1480902	0.1480902	0.1480902	0.1480902	0.1480902
	0.95	0.1752521	0.173935	0.173905	0.150899 0.174880 0.175252	0.156129 0.163693 0.175252 0.175252	0.1750963	0.1553452	0.1645855	0.1603422	0.1603422	0.1603422	0.1603422	0.1603422
	0.99	0.2070599	0.205122	0.205395	0.186378 0.206412 0.207059	0.1874797 0.1895252 0.2070599 0.2070599	0.2068606	0.1761900	0.1970058	0.1885281	0.1885281	0.1885281	0.1885281	0.1885281
50	0.90	0.0683267	0.068129	0.068235	0.067319 0.068298 0.068327	0.0666774 0.0640110 0.0683267 0.0683267	0.0683234	0.0671345	0.0675124	0.0656449	0.0656449	0.0656449	0.0656449	0.0656449
	0.95	0.0746748	0.074461	0.074574	0.073645 0.074641 0.074675	0.0729275 0.0697699 0.0746748 0.0746748	0.0746704	0.0701529	0.0738488	0.0717300	0.0717300	0.0717300	0.0717300	0.0717300
	0.99	0.0894418	0.089327	0.089308	0.083577 0.089413 0.089442	0.0847579 0.0864786 0.0894418 0.089441	0.0894328	0.0840142	0.0877192	0.0857121	0.0857121	0.0857121	0.0857121	0.0857121

تابع لجدول (3)

n	ρ	MLE	Ridge Estimator							Liu estimator				
			K_1	K_2	K_3	K_4	K_5	K_6	K_7	D_1	D_2	D_3	D_4	D_5
100	0.90	0.0307735	0.0307548	0.030762	0.030598 0.030770 0.030774	0.0304757 0.0300294 0.0307736 0.0307736	0.030773	0.030153	0.030629	0.0302709	0.0302709	0.0302709	0.0302709	0.0302709
	0.95	0.0338859	0.0338646	0.033873	0.033692 0.033882 0.033886	0.03355577 0.03343491 0.03388594 0.03388594	0.033886	0.033591	0.033731	0.0333242	0.0333242	0.0333242	0.0333242	0.0333242
	0.99	0.0411423	0.0410688	0.041125	0.040483 0.041138 0.041142	0.04043766 0.04033878 0.04114239 0.04114238	0.041142	0.039939	0.040867	0.0403898	0.0403898	0.0403898	0.0403898	0.0403898
250	0.90	0.0132918	0.0132909	0.013292	0.013287 0.013291 0.013292	0.01327214 0.01312654 0.01329176 0.0132917	0.013291	0.013263	0.013287	0.0132064	0.0132064	0.0132064	0.0132064	0.0132064
	0.95	0.0146314	0.0146305	0.014631	0.014626 0.014631 0.014631	0.01460809 0.01425363 0.01463140 0.01463140	0.014631	0.014587	0.014625	0.0145368	0.0145368	0.0145368	0.0145368	0.0145368
	0.99	0.0177563	0.0177550	0.017756	0.017736 0.017756 0.017757	0.01770761 0.01767147 0.01775625 0.01775625	0.017756	0.017597	0.017744	0.0176382	0.0176382	0.0176382	0.0176382	0.0176382

المبحث الرابع/ الاستنتاجات والتوصيات

يتضمن هذا المبحث أهم الاستنتاجات التي تم التوصل اليها من خلال المباحث الثلاثة السابقة في هذا البحث، وكذلك التوصيات التي يوصى بها الباحث.

1-4 الاستنتاجات

1- في ظل وجود مشكلة التعدد الخطي في أنموذج انحدار ثنائي الحدين السالب تكون مقدرات طريقة الامكان الاعظم (MLE) غير دقيقة لعدم استقرار المعلمات المقدرة على الرغم من عدم تحيز المقدرات وتبين انها تمتلك اكبر متوسط مربعات للخطأ (MSE) من بين طرائق التقدير الاخرى.

2- اثبتت طريقة مقدر انحدار حرف (Ridge Regression Estimator) لأنموذج انحدار ثنائي الحدين السالب انها تفوق طريقتي تقدير مقدرات الامكان الاعظم (MLE) وطريقة مقدرات ليو (Liu-Estimators) من خلال معيار المقارنة متوسط مربعات الخطأ (MSE) في صيغها (K_6, K_4, K_3).

3- ظهر تفوق طريقة مقدرات ليو (Liu-Estimators) في تقدير معلمات أنموذج انحدار ثنائي الحدين السالب في حالة وجود مشكلة التعدد الخطي عندما يكون عدد المتغيرات التوضيحية (P=5) وحجم العينة (n=100) وعند الارتباط ($\rho = 0.95$).

2-4 التوصيات

1- ضرورة استخدام أنموذج انحدار ثنائي الحدين السالب في دراسة وتحليل الظواهر والاحداث نادرة الوقوع فضلا عن كونها ذات قيم موجبة وبهيئة بيانات معدودة.

2- دراسة وايجاد طرائق لتقدير معلمات أنموذج انحدار ثنائي الحدين السالب عند وجود مشكلة التعدد الخطي التام.

3- اعتماد طريقة مقدرات انحدار الحرف في تقدير معلمات أنموذج انحدار ثنائي الحدين السالب في حالة وجود مشكلة التعدد الخطي.

4- اعتماد طريقة مقدرات ليو (Liu Estimators) في تقدير معلمات أنموذج انحدار ثنائي الحدين السالب في حالة وجود مشكلة التعدد الخطي عندما يكون عدد المتغيرات التوضيحية (P=5) وحجم العينة (n=100) وعند الارتباط ($\rho = 0.95$).

المصادر References

1.كاظم ،اموري هادي ، ومسلم ، باسم شليبة (٢٠٠٢) م. " القياس الاقتصادي المتقدم النظرية والتطبيق"، مطبعة دنيا الأمل، العراق ، بغداد.

2. يحيى ، م. م. مزاحم محمد (٢٠٠٥) م. "استخدام المكونات الرئيسة وانحدار الحرف في تقدير معادلة السعر العالمي للقمح للفترة من (١٩٦١- ٢٠٠٢)، كلية الادارة والاقتصاد ،جامعة تكريت.

1. Algamal , Z. Y & Rasheed, K.B (2012). " Re-Sampling Techniques in Count Data Regression Models " , Iraqi Journal Of Statistical Sciences ,pp [15-25], Mosul University, Iraq.

2. Asar ,Y (2016). " Liu – type Negative Binomial Regression : A comparison of Recent Estimators And Applications", Department Of Mathematics Necmettin Erbakan University Konya, Turkey.

3. Asar,Y (2016). "Liu Type Logistic Estimators With Optimal Shrinkage Parameter", Necmettin Erbakan University, Journal Of Modern Applied Statistical Methods ,Konya,Turkey.

4. Garaya, A.M & Hashimotob, E. M & Ortega, E.M.M & Lachos ,V.H(2011). "On estimation and influence Diagnostics For Zero –Inflated Negative Binomial Regression Models ",a Departmento de Estatistica,bDepartmento de Ciências Exatas, Universidad de São Paulo, Brazil.

5. Lukman , A. F & Ayinde , K (2015). "Review And Glassifications Of The Ridge Parameter Estimation Techniques", Department of Statistics, Ladoke Akintola University of Technology, Ogbomoso, Nigeria

6.- Lord,D & Park,B.J (2012). "Negative Binomial Regression Models And Estimation Models ", Texas A&M University ,Korea Transport Institute.

7. Månsson, K (2012). "On Ridge Estimators For The Negative Binomial Regression Model", Department Of Economics And Statistics, Jònkòping University, Sweden.

8. Månssona, K (2012). "Developing A Liu Estimator For The Negative Binomial Regression Model: Method And Application", Department Of Economics ,Finance And Statistics ,Jònkòping University,Journal Of Statistical Computation And Simulation,Sweden.

9. Mansson, K. & Kebria, B. M & Sjolander, P & Shukur, G (2012). "Improved Liu Estimators for The Poisson Regression Mode ", International Jornal of Statistics and Probability, Vol. 1, No. 1, pp. 2-6.

10. Mansson, K & Kebria, B . M & Sjolander, P & Shukur, G. (2012). "New Liu Estimators for the Poisson Regression Model : Methods and Application" Jönköping, Sweden.

11. Piegorsch , Walter W (1990). " Maximum Likelihood Estimation For The Negative Binomial Dispersion Parameter", Statistics and Biomathematics Branch, National Institute Of Environmental Health Sciences Research Triangle Park, North Carolina. U.S.A.

12. Zwilling , Michael L (2013). "Negative Binomial Regression ", the Mathematica R Journal.

A Comparison of Parameters Estimation Methods for the Negative Binomial Regression Model under Multicollinearity Problem by Using Simulation

Abstract

This study discussed a biased estimator of the Negative Binomial Regression model known as) Liu Estimator ،(This estimate was used to reduce variance and overcome the problem Multicollinearity between explanatory variables ،Some estimates were used such as Ridge Regression and Maximum Likelihood Estimators ،This research aims at the theoretical comparisons between the new estimator (Liu Estimator) and the estimators of Maximum Likelihood) ML (and Ridge Regression (RR (by using the mean square error (MSE) criterion ،where the variance of the Maximum Likelihood) ML) comes in the presence of the problem Multicollinearity between the explanatory variables. In this study, the Monte Carlo simulation was designed to evaluate the performance of estimations using the criterion for comparison ،the mean square error) MSE). The simulation results showed important an estimated Liu and superior to the RR and MLE estimator Where the number of explanatory variables is) p=5 (and the sample size is) n=100 ،(where the number of explanatory variables is) p=3 (and for all sizes, and also when) p=5 (for all sizes except size) n=100 ،(the RR regression method is the best.

Key word/ Maximum Likelihood Estimator, Ridge Regression Estimator, Liu Estimator, Negative Binomial Regression Model, Multicollinearity Problem.