المجلة العراقية للعلوم الاحصائية (9) 2006
ص ص [150-140]

تقويم طريقة (Liu) المقيدة في معالجة مشكلة التعدد الخطي

خيري بدل رشيد الرشيداني*

الملخص :

في هذا البحث تمت مقارنة مقدرات المربعات الصغرى المقيــدة $(\hat{\beta}_{RLS})$ مع مقدرات الطريقة المقترحة من قبل الباحــث ((1993) *Liu, K.J.*)[5] والتــي أطلق عليها اسم مقدرات *Liu* المقيدة $(\hat{\beta}_{RLIU})$،وذلك في حالــة وجــود مشــكلة التعددية الخطية (*Multicollinearity*) ، ولقد استخدم المعيار الإحصائي (*MSE*) لاختيار الطريقة الأكفأ لمعالجة هذه مشكلة. إذ تــم التوصــل إلــى أن مقــدرات طريقة(*Liu*) المقيدة تمتلك متوسط مربعات خطأ اقــل مــن مقــدرات المربعــات الصغرى المقيدة.

Assessment Restricted Liu Estimator to treating Multicollinearity Problem

ABSTRACT

In this research, we compared restricted least squares $(\hat{\beta}_{RLIU})$ with restricted Liu estimator $(\hat{\beta}_{RLS})$. by using (*MSE*) criterion in the existence of multicollinearity. We found that restricted Liu estimator is the best in comparison.

المقدمة :

إن مشكلة التعددية الخطية (تعدد العلاقة الخطية) أصبحت معروفــة لــدى العديد من الباحثين الإحصائيين وكذلك معرفة تبعاتها الإحصــائية علــى معلمــات الانحدار الخطي المتعدد إذ تؤدي هذه المشكلة في ابسط حالاتها إلى ابتعاد معلمات النموذج المقدرة عن خصائصها العلمية المرجوة منها في تفسير الظاهرة العلميــة

* مدرس مساعد / كلية علوم الحاسبات والرياضيات- قسم الإحصاء

تاريخ التسلم : 4/ 9 /2005 ــــــــــــــــ تاريخ القبول : 26/ 9 /2005

بالأسلوب العلمي الصحيح، لذا وجب تفادي هذه المشكلة وذلك بوضع الحلول المناسبة لها[4].

من المعروف أن طريقة المربعات الصغرى الاعتيادية (OLS) من أكفأ طرائق التقدير المعروفة غير المتحيزة ويشترط فيها توفر فروض التحليل التي من أهمها استقلالية المتغيرات التوضيحية.

في بعض الأحيان تزودنا النظرية الاقتصادية ببعض المعلومات النظرية بهيئة قيود(قيد واحد متطابق أو أكثر حول المعالم المراد تقديرها) من خارج نطاق عينة البحث، في هذه الحالة فان طريقة المربعات الصغرى الاعتيادية تؤول إلى صيغة أخرى وهي صيغة المربعات الصغرى المقيدة(RLS)، بحيث من الممكن (في بعض الدوال السلوكية) توظيف هذه القيود لمعالجة مشكلة التعدد الخطي[1] .

في عام 1993 قام الباحث (*Liu*) بطرح نموذج جديد لمعالجة هذه المشكلة إذ قام بدمج مقدرات (*Stein*) ومقدرات انحدار الحرف الاعتيادية ($\hat{\beta}_{ORR}$) بمقدر خاص أطلق عليه اسم مقدرات ($\hat{\beta}_{OLiu}$) وعند وضع القيود على المقدرات الأخيرة تصبح مقدرات ($\hat{\beta}_{RLiu}$) المقيدة[2] .

قسمت هذه الدراسة إلى جزأين إذ تضمن الجزء الأول الجانب النظري والذي تم فيه شرح أسلوب إيجاد مقدرات المربعات الصغرى ومقدرات (*Liu*) المقيدتين ، أما الجزء الثاني فتضمن الجانب التجريبي والذي استخدم فيه البرنامج التطبيقي الجاهز (*Minitab under Windows*) في توليد البيانات باستخدام أسلوب مونتيكارلو للمحاكاة ووضع صيغ الارتباطات وإيجاد متوسط مربعات الخطأ لكلا الطريقتين[3] .

الجانب النظري:

1- مقدرات *OLS* و *RLS* :

تعد طريقة المربعات الصغرى الاعتيادية من الطرائق المهمة والأكثر شيوعا لتقدير معلمات نموذج الانحدار الخطي وتمتاز بصفات أكثر فعالية من غيرها من الطرائق وبسهولة حساب تقدير معلماتها ومنطقية النتائج المستحصل عليها وسهولة فهم ميكانيكية عملها وتمتاز بان مقدراتها تمتلك اقل التباينات من بين كل الطرائق غير المتحيزة الاخرى في حال توافر فروض التحليل للنموذج الخطي والمتمثلة بكون الخطأ هو متغير عشوائي موزع توزيعا طبيعيا بمعدل قدره (0) وتباين قدره(σ_u^2) مما يعني أن الأخطاء غير مرتبطة ذاتيا، أن متغير الاستجابة(Y) هو متغير عشوائي يتوزع توزيعا طبيعيا بمعدل قدره (μ) وتباين قدره ($\sigma_u^2 \ I$) ، فضلا عن أن أعمدة مصفوفة المتغيرات التوضيحية (X) مستقلة خطيا" (*Linearly Independent*) وان (Rank(X)= P < n)[4]. في هذا البحث تم اخذ النموذج الخطي الاتي:

$$\underline{Y} = X\underline{\beta} + \underline{U} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

حيث أن :

$\underline{Y}_{n*1}$: متجه متغير الاستجابة .

X_{n*p+1} : مصفوفة المتغيرات التوضيحية .

$\underline{\beta}_{p+1*1}$: متجه معلمات الانحدار .

$\underline{U}_{n*1}$: متجه الأخطاء العشوائية حيث أن $E(U)=0$ ، وان $V(U) = \sigma_u^2 I_n$

(P) : تمثل عدد المتغيرات التوضيحية ، (n) تمثل حجم العينة .

I_{n*n} : مصفوفة أحادية .

من المعروف أن مقدرات المربعات الصغرى يتم الحصول عليها من تطبيق الصيغة الاتية :

$$\hat{\beta}_{OLS} = (X'X)^{-1} X'Y \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

وان مصفوفة متوسط مربعات الخطأ للمقدر $\hat{\beta}_{OLS}$ من β_{OLS} هي:

$$MSE(\hat{\beta}_{OLS}) = E(\hat{\beta}_{OLS} - \beta_{OLS})(\hat{\beta}_{OLS} - \beta_{OLS})' = V(\hat{\beta}_{OLS}) + \left[(bias(\hat{\beta}_{OLS}))(bias(\hat{\beta}_{OLS}))'\right] \quad \ldots. (3)$$

حيث أن:

$$bias(\hat{\beta}_{OLS}) = E(\hat{\beta}_{OLS}) - \beta_{OLS}$$

وعند افتراض وجود (q)من القيود المفروضة على النموذج الخطي والتي من الممكن أن تكون بالشكل الاتي[2]:

$$R\beta_{OLS}=r \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

حيث أن :

R_{q*p} مصفوفة، صفوفها ذات رتبة كاملة وان $p \succ q$ ، وهي مصفوفة معلومة .

r_{q*1}: متجه . وهو معلوم أيضا" .

لذا فأن مقدرات ($\hat{\beta}_{RLS}$) وطبقا" لقيود المعادلة (4) أعلاه تكون بالشكل التالي:

$$\hat{\beta}_{RLS} = \hat{\beta}_{OLS} + S^{-1} R' (R S^{-1} R')^{-1} (r - R \hat{\beta}_{OLS}) \ldots\ldots\ldots\ldots (5)$$

حيث أن:

$$\hat{\beta}_{OLS} = S^{-1} X' Y \qquad and \qquad S = X'X$$

وإذا افترضنا أن :

$$A = S^{-1} - S^{-1} R' (R\ S^{-1}\ R')^{-1}\ R\ S^{-1} \qquad and \quad \delta = r - R\ \hat{\beta}_{OLS} \quad \ldots\ldots\ldots.. (6)$$

نحصل على :

$$V(\hat{\beta}_{RLS}) = \sigma^2\ A \quad \ldots\ldots\ldots\ldots\ldots\ldots (7)$$

نفرض أن مقدر (Liu) هو ($\hat{\beta}_{LIU}$) للمعلمة $(0 \prec d \prec 1)$ بالشكل التالي[2]:

$$\hat{\beta}_{LIU} = (S + I)^{-1} (X'Y + d\ \hat{\beta}_{OLS}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

وذا افترضنا أن :

$$F_d = (S + I)^{-1} (S + d\ I) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (9)$$

إذا" مقدر $\hat{\beta}_{LIU}$ يمكن أن يكتب بالشكل الاتي:

$$\hat{\beta}_{LIU} = F_d \; \hat{\beta}_{OLS} \qquad \ldots\ldots\ldots\ldots(10)$$

بحيث أن:

$$E(\hat{\beta}_{LIU}) = F_d \; \beta_{OLS} \qquad \ldots\ldots\ldots\ldots(11)$$

وان :

$$V(\hat{\beta}_{LIU}) = \sigma^2 \; F_d \; S^{-1} F_d' \qquad \ldots\ldots\ldots\ldots(12)$$

من التعريف أعلاه وطبقا" للقيود في المعادلة (4) نجد أن مقــدر (Liu) المقيــد يكون بالصيغة الاتية:

$$\hat{\beta}_{RLIU} = F_d \; \hat{\beta}_{RLS} \qquad \ldots\ldots\ldots\ldots(13)$$

بحيث أن :

$$E(\hat{\beta}_{RLIU}) = F_d \; \beta_{OLS} + F_d \, S^{-1} R' (R \;\; S^{-1} R')^{-1} \delta \qquad \ldots\ldots(14)$$

وان:

$$V(\hat{\beta}_{RLIU}) = \sigma^2 \; F_d \; A \; F_d' \qquad \ldots\ldots\ldots\ldots(15)$$

2-مقارنة بين مقدرات $\hat{\beta}_{RLS}$ ومقدرات $\hat{\beta}_{RLIU}$:

عند اخذ الفرق بين تبايني كلا المقدرين نجد أن [2] :

$$V(\hat{\beta}_{RLS}) - V(\hat{\beta}_{RLIU}) = \sigma^2 \, (A - F_d \; A F_d')$$
$$= \sigma^2 \, (1-d) \; (S+I)^{-1} \left[AS + SA + (1+d)A \right] (S+I)^{-1}$$
$$= \sigma^2 \, (1-d) \; (S+I)^{-1} \left[G - H \right] (S+I)^{-1} \qquad \ldots\ldots\ldots\ldots(16)$$

حيث أن : $G = 2I + (1+d)\, S^{-1}$ مصفوفة محــددة وموجبــة *((Positive Definite (p.d)* ، وأن :

$$H = S^{-1} R' (R \;\; S^{-1} R')^{-1} R + R' \, (R \, S^{-1} \, R')^{-1} \, R \, S^{-1} + (1+d) \, S^{-1} \, R' (R \, S^{-1} \, R')^{-1} R \, S^{-1}$$

وهي مصفوفة متماثلة *(Symmetric matrix)* . وبما أن G مصفوفة محددة وموجبة فأن هنالك مصفوفة أحادية مثل Q بحيث أن :

$Q' G Q = I$ ، وان $Q' H Q = \Lambda$ ، حيث أن Λ مصفوفة قطريـــة عناصـــر القطر فيها عبارة عن الجذور المميزة لـــ $|H - \lambda G| = 0$

فأذا كانت $(G-H)$ مصفوفة محـــددة غيــر ســالبة ($(n.n.d)$ *Non Negative Definite*) .

وان : $Q'GQ-Q'HQ=I-\Lambda$ هي ايضا مصفوفة محددة غير سالبة إذا"[5] :

$$1-\lambda_i \geq 0$$

وايضا نجد أن :

$$\lambda_{\max}(G^{-1}H) \leq 1$$

مما سبق نجد أن :

$$\lambda_p \leq \frac{x'Hx}{x'Gx} \leq \lambda_1$$

حيث أن :

$$\lambda_1(G^{-1}H) \geq \lambda_2(G^{-1}H) \geq .... \geq \lambda_p(G^{-1}H)$$

من هنا نجد أن: $x'Hx \leq x'Gx$ ، وان : $(G-H)$ هي ايضا $(n.n.d)$.

من هذه المعلومات نجد انه من الواضح أن :

$$V(\hat{\beta}_{RLS})-V(\hat{\beta}_{RLIU}) \qquad is\ (n.n.d) \qquad \forall \quad 0 \prec d \prec 1 \ \ldots. \ (17)$$

أي أن تباين مقدرات(Liu) المقيدة اقل من تباين مقدرات المربعات الصغرى المقيدة إذا وفقط إذا

$$\lambda_{\max}(G^{-1}H) \leq 1 \ .$$

الجانب التجريبي :

1–وصف النموذج ومحاكاته :

في هذا البحث تم اســتخدام أسـلوب مونتيكــارلو المحاكــاة (*Monte Carlo simulation*)[3] في توليد مشاهدات نموذج الانحدار في المعادلة(1) لخمس عينات كل عينة بحجم (100) مشاهدة وكل مشاهدة مكررة مائة مرة ثم تم اخذ معدل هــذه المشاهدات واخيرا قسمت هذه المشاهدات على شكل خمسة أعمدة كل عمود يحوي (100) مشـــاهدة وبهـــذا سيتشـــكل لـــدينا خمســـة أعمـــدة (Z_{ik} , $i=1,2,3,...,100$, $k=1,2,...,5$) والتي سيتم اســتخدامها فــي توليد قيم (X_{ij} , $i=1,2,3,....,100$, $j=1,2,....,4$) ، وان المتغير (Z_{i5})

سيمثل عمود الخطأ ولكن بعد طرح وسطه الحسابي منه أي أن $(U_i = Z_{i5} - \bar{Z}_5)$
ومن ثم سيتم تقدير قيم Y_i باستخدام المعادلة الاتية :

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + U_i$ ، حيث انه يتم اخـــذ قيم $(\beta_1, \beta_2, \beta_3, \beta_4)$ بالتطابق مع الجذور المميزة لمصفوفة $(X'X)$ القياســية مع افتراض أن قيمة $\beta_0 = 0$ ، ومن ثم سيتم وضع (افتراض) علاقــة خطيــة اة عدمه بين المتغيرات التوضيحية X_{ij} المولدة ، وذلك باستخدام العلاقة الاتية :

$$\begin{aligned} X_{ij} &= \sqrt{(1-\alpha^2)}\ Z_{ij} + \alpha\ Z_{i5} \qquad , j = 1,2 \quad , i = 1,2,\ldots\ldots,100 \\ X_{ij} &= \sqrt{(1-\alpha^{*2})}\ Z_{ij} + \alpha^*\ Z_{i5} \qquad , j = 3,4 \quad , i = 1,2,\ldots\ldots,100 \end{aligned} \quad \ldots(18)$$

حيث أن قيمة (α) تتمثل بدرجة الارتبـــاط بـــين المتغيـــرين (X_1, X_2) ، وان (α^*) تتمثل بدرجة الارتباط بين المتغيرين (X_3, X_4) . وفـــي هـــذا البحـــث افترضت ثلاث حالات ارتباط وهي :

- ارتباط عال" للمجموعتين متمثلا بالقيم $(\alpha = 0.9 \ , \ \alpha^* = 0.9)$.
- ارتباط واطئ للمجموعتين متمثلا بالقيم $(\alpha = 0.3 \ , \ \alpha^* = 0.3)$.
- ارتباط عال للمجموعة الأولى وواطـــئ للمجموعـــة الثانيـــة متمـــثلا بـــالقيم $(\alpha = 0.9 \ , \ \alpha^* = 0.1)$.

وأخيرا" يتم تحويل المتغيرات التوضيحية إلى شكلها القياسي باستخدام صيغة الثابت بطول واحد و كالاتي :

$$W_i^* = \frac{W_i - \bar{W}}{\sqrt{\sum_{i=1}^{n} (W_i - \bar{W})^2}} \ldots\ldots\ldots\ldots\ldots\ldots(19)$$

حيث أن :

W_i هو أي متغير عشوائي ، وان W^* يمثل الشكل القياسي لذلك المتغير.

2-مرحلة تقدير المعلمات :

في هذا البحث تم افتراض وجود القيدين المتطابقين الاتيين :

1-إن المجموع النهائي لكل المعالم المقدرة مساويا" للواحد الصـــحيح أي أن:

$$\sum_{j=1}^{p} \beta_j = 1$$

2-إن الميل الحدي الأول مساوٍ للميل الحدي الثاني أي أن :

$$\beta_1 = \beta_2 \qquad or \quad \beta_1 - \beta_2 = 0$$

إن توظيف مثل القيود السابقة يستوجب تطبيق أسلوب المربعات الصغرى المقيــدة والذي يتطلب إعادة صياغة القيود أعلاه باستخدام المصفوفات والمتجهات وكالأتي:

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} \quad ، \qquad r = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad ،$$

$$R = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix}$$

تم في هذه المرحلة تقدير المعلمات بطريقة المربعات الصغرى المقيــدة وطبقــا" للمعادلة (5) ولكل حالة ارتباط ، كما تم أيضا" تقدير معلمات طريقة (Liu) المقيدة طبقا للمعادلة(13) ولكل حالة ارتباط أيضا" مع اختيار أربع قيم ل (d) وهي (0.1, 0.4، 0.6، 0.9)، وتم الحصول على النتائج كما في الجدول (A) الاتي :

بعد أن تم أيجاد مقدرات كلتا الطريقتين يتم ألان إيجاد متوسط مربعات الخطأ لكلتا الطريقتين ولجميع حالات الارتباط وذلك عن طريق المعادلتين الاتيين :

فبالنسبة إلى مقدرات ($\hat{\beta}_{RLS}$) فأن :

$$MSE(\hat{\beta}_{RLS}) = \frac{\sum_{j=1}^{4}(\hat{\beta}_{RLS} - \beta_{eign(S)})^2}{n}$$

أما بالنسبة إلى مقدرات($\hat{\beta}_{RLIU}$) فان :

$$MSE(\hat{\beta}_{RLIU}) = \frac{\sum_{j=1}^{4}(\hat{\beta}_{RLIU} - \beta_{eign(S)})^2}{n}$$

حيث أن : $\beta_{eign(S)}$ تمثل الجذور المميزة لمصفوفة S القياسية .

والجدول (B) أدناه يوضح نتائج (MSE) لكلتا الطريقتين :

	$\alpha = 0.9$, $\alpha^* = 0.9$	$\alpha = 0.1$, $\alpha^* = 0.9$	$\alpha = 0.3$, $\alpha^* = 0.3$
$MSE\,\hat{\beta}_{RLS}$	**0.085**	**1.202**	**0.121**
$MSE\,\hat{\beta}_{RLIU}$ $d = 0.1$	**0.074**	**0.209**	**0.042**
$MSE\,\hat{\beta}_{RLIU}$ $d = 0.4$	**0.073**	**0.433**	**0.088**
$MSE\,\hat{\beta}_{RLIU}$ $d = 0.6$	**0.084**	**0.641**	**0.098**
$MSE\,\hat{\beta}_{RLIU}$ $d = 0.9$	**0.084**	**1.044**	**0.063**

الاستنتاجات :

1–من ملاحظة جدول (B) نجد أن طريقة (*Liu*) هي اكثر دقة من طريقة المربعات الصغرى المقيدة .

2– يمكن استخدام هذه الطريقة حتى في حالة عدم وجود التعددية الخطية بين المتغيرات التوضيحية .

المصادر :

1 –كاظم ، أموري هادي و مسلم ، باسم شليبة ، (2002) " القياس الاقتصادي المتقدم النظرية والتطبيق "، دار الكتب والوثائق ،جامعة بغداد .

2- Akdeniz, F and Kaciranlar, S., (2001)" More on the new biased estimator in linear regression ", Sankhya, Vol.63, Ser. B, Pt.3, pp.321-325.

3- Ghrishtopher, Z .M., (1997)"Monte Carlo Simulation", SAGE Publications, Delhi.

4-Gunst, R. F. and Mason, R. L. and Webster, J. F., (1975) "Regression Analysis and problems of Multicollinerity "Communication in statistics, Vol. 22, No.3. "

5-Liu, Communication in statistics , Ser. A , Vol . 22, pp. 393-402.