تقدير معاملات الانحدار بعد إجراء اختبار عدم تجانس التباين

د.ايدن حسن الكناني*

المقدّمة:- Introduction

تعتبر مشاكل القياس الاقتصادي واحدة من المشاكل المهمة التي تواجه الباحثين عند وصفهم وبنائهم لنموذج انحدار خطي (Linear Regression Model) لمجموعة من الظواهر والمتغيرات التي تدعى بالمتغيرات التوضيحية (Explanatory Variables) والتي تؤثر وترتبط بمتغير أو ظاهرة معينة تدعى بمتغير الاستجابة (Response Variable) .

إنّ مشكلة عدم تجانس التباين (Heteroscedasticity) لحدود الأخطاء واحدة من أهم هذه المشاكل والتي ينبغي معالجتها ودراستها .

مشكلة عدم تجانس التباين :- Heteroscedasticity Problem

عند اعتمادنا على نموذج انحدار خطّي عام لتمثيل البيانات فإننا نستخدم طريقة المربعات الصغرى الاعتيادية (OLS) لتقدير معالم هذا النموذج . وتعتمد طريقة (OLS) عند تطبيقها على تحقق فرضيات وشروط أساسية معينة لكي تكون هناك دقة في تقدير معالم هذا النموذج .

وإحدى الفرضيات الأساسية التي نعتمدها عند تقدير معالم نموذج الانحدار الخطي بطريقة (OLS) هي فرضية ثبات تجانس التباين (Homoscedasticity) لحدود الأخطاء والتي يعبّر عنها رياضياً بالصيغة التالية:

$$Var(\underline{U}) = E(UU') = \sigma^2 In \qquad (1)$$

* استاذ مساعد/ الجامعة المستنصرية /كلية الادارة والاقتصاد / قسم الاحصاء

مقبول للنشر بتاريخ 2004/12/27

وفي الكثير من التطبيقات العملية فإنّ هذه الفرضية قد لا تتحقق وبالتالي فإنّ طريقة (OLS) لتقدير معالم نموذج الانحدار الخطي سوف لا تعطينا نتائج دقيقة وصحيحة مما يؤدي بنا إلى الوقوع في مشكلة عدم تجانس التباين والتي تصاغ رياضياً بالشكل التالي :

$$Var(\underline{U}) = E(UU') = \sigma^2 \ \Omega \qquad (2)$$

بالإضافة إلى إنّ المعالم المقدّرة في ظل وجود مشكلة عدم تجانس التباين سوف لا تعطينا خاصية أفضل تقدير خطّي غير متحيز (BLUE) وهذا يعني أنّ المعالم المقدّرة سوف لا تمتلك خاصية أقل تباين ممكن .

مقدّر Aitken ذو المرحلتين :- Two Stage Aitken Estimator

يعتبر مقدّر Aitken ذو المرحلتين والذي يرمز له اختصاراً (2 SAE) من أهم المقدرات التي تستخدم لتقدير معالم نموذج الانحدار في حالة وجود مشكلة عدم تجانس التباين بين حدود الأخطاء . فإذا فرضنا بأنّنا لدينا نموذج الانحدار الخطي العام (GLM) التالي :

$$\underline{y} = x\,\underline{\beta} + \underline{e} \qquad (3)$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \underline{\beta} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \qquad (4)$$

$$\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \sim N\left(\underline{0}, \begin{bmatrix} \sigma_1^2 In_1 & 0 \\ 0 & \sigma_2^2 In_2 \end{bmatrix}\right)$$

حيث أنّ :

y_i - متجه عشوائي ذو بعد n_i له متوسط $x_i\beta$.

x_i - مصفوفة ذات بعد $(n_i \times p)$ وإنّ $(n_i > p+2)$ ، $i = 1,2$.

$\underline{\beta}$ - متجه المعالم ذو بعد $p \times 1$.

e_i - متجه حدود الأخطاء ذو بعد n_i $(i = 1,2)$.

فإذا افترضنا بأننا لدينا معلومات مسبقة (Prior Knowledge) حول التباين حيث أن ($\sigma_1^2 \geq \sigma_2^2$) ، كذلك نفترض بأننا لدينا انحدار طبيعي ومتعامد (Orthonormality of the Regressors) في كل عينة جزئية بحيث أن :

$$x_1' x_1 = x_2' x_2 = Ip \qquad (5)$$

إنّ المعالم المقدرة بطريقة (2 SAE) للنموذج الخطي العام تكون بالصيغة التالية :

$$\underline{\hat{\beta}} = (x'\Phi x)^{-1} x'\Phi^{-1} y \qquad (6)$$

حيث أنّ $\Phi = \sigma_i^2$ وبالتالي يمكن كتابة المعادلة (6) بالشكل التالي :

$$(x'\Phi^{-1}x)^{-1} = \left(\frac{x_1^2}{\hat{\sigma}_1^2} + \frac{x_2^2}{\hat{\sigma}_2^2}\right)^{-1} = \left(\sum_{i=1}^{2} \hat{\sigma}_i^2 x_i^2\right)^{-1}$$

$$x'\Phi^{-1}y = \left(\frac{x_1 y_1}{\hat{\sigma}_1^2} + \frac{x_2 y_2}{\hat{\sigma}_2^2}\right) = \sum_{i=1}^{2} \hat{\sigma}_i^2 x_i y_i$$

$$\hat{\beta} = \left(\frac{x_1^2}{\hat{\sigma}_1^2} + \frac{x_2^2}{\hat{\sigma}_2^2}\right)^{-1} \left(\frac{x_1 y_1}{\hat{\sigma}_1^2} + \frac{x_2 y_2}{\hat{\sigma}_2^2}\right) = \left(\sum_{i=1}^{2} \hat{\sigma}_i^2 x_i^2\right)^{-1} \left(\sum_{i=1}^{2} \hat{\sigma}_i^2 x_i y_i\right) \qquad (7)$$

فإذا أردنا اختبار فرضية العدم $H_o : \sigma_1^2 = \sigma_2^2$ ضد الفرضية البديلة $H_1 : \sigma_1^2 > \sigma_2^2$ باستخدام احصاءة اختبار F عند درجة حرية $(n_1 - p \,,\, n_2 - p)$ ورفضنا فرضية العدم وقبلنا الفرضية البديلة H_1 فهذا يعني عدم تجانس تباين حدود الأخطاء وبالتالي فإننا نستخدم طريقة 2SAE لتقدير معالم نموذج الانحدار ويكون المقدر كالآتي :

$$\underline{\hat{\beta}} = w_1 \hat{\beta}_1 + w_2 \hat{\beta}_2 \qquad (8)$$

حيث أنّ :

$$w_1 = \frac{S_2^2}{S_1^2 + S_2^2} \,,\; w_2 = \frac{S_1^2}{S_1^2 + S_2^2}$$

$$\underline{\hat{\beta}}=\left[\frac{1}{S_2^2}+\frac{1}{S_1^2}\right]^{-1}\left[\frac{S_2^2\,\hat{\beta}_1+S_1^2\,\hat{\beta}_2}{S_1^2\;S_2^2}\right] \qquad (9)$$

$$\frac{S_1^2}{S_2^2}\sim F(n_1-p\;,n_2-p)$$

أما إذا قبلنا فرضية العدم $H_\circ$ ورفضنا الفرضية البديلة فهذا يعني تجانس تباين حدود الأخطاء وبالتالي فإننا نستخدم طريقة (OLS) لتقدير معالم نموذج الانحدار ويكون المقدر كالآتي :

$$b=\frac{b_1+b_2}{2} \qquad (10)$$

هدف البحث :- The Aim of Research

في هذا البحث سنهتم بمشكلة تقدير معالم نموذج الانحدار الخطي العام بعد إجراء اختبار تجانس التباين لحدود الأخطاء . لذا فإننا نهدف في هذا البحث على إيجاد مقدر جديد يدعى بمقدر الاختبار الأولي (Preliminary Test Estimator) والناتج من عملية ربط ودمج مقدر طريقة المربعات الصغرى الاعتيادية عند قبولنا لفرضية العدم بعد استخدام اختبار F لاختبار تجانس التباين ومقدر (2 SAE) عند رفضنا لفرضية العدم بعد استخدام اختبار F أيضاً .

إنّ مقدر الاختبار الأولي للمعلمة β يكون بالصيغة التالية :

$$\tilde{\beta}=\begin{cases}\hat{\beta}=\left(\dfrac{1}{S_1^2}+\dfrac{1}{S_2^2}\right)^{-1}\left(\dfrac{\hat{\beta}_1}{S_1^2}+\dfrac{\hat{\beta}_2}{S_2^2}\right) & if\quad \dfrac{S_1^2}{S_2^2}\geq\lambda\\[2ex] b=\dfrac{b_1+b_2}{2} & if\quad \dfrac{S_1^2}{S_2^2}<\lambda\end{cases} \qquad (11)$$

حيث أنّ λ تمثل القيمة الحرجة (Critical Value) لاختبار F . كما أننا سنشتق ونجد متوسط مربعات الخطأ (MSE) لمقدر الاختبار الأولي $\widetilde{\beta}$ بعد اختبار تجانس التباين .

متوسط مربعات الخطأ لمقدر الاختبار الأولي :-

يعتبر(MSE) من أهم المقاييس وأفضل خواص المقدّر الجيد لذا فإننا سنقوم باشتقاقه وإيجاد قيمته لمقدر الاختبار الأولي وحسب الخطوات التالية :

$$MSE(\widetilde{\beta}) = E\left[(\widetilde{\beta} - \beta)'(\widetilde{\beta} - \beta)\right] \qquad (12)$$

$$MSE(\widetilde{\beta}) = E\left[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)\right]\Pr\left(\frac{S_1^2}{S_2^2} \geq \lambda\right) + E\left[(b - \beta)'(b - \beta)\right]\Pr\left(\frac{S_1^2}{S_2^2} < \lambda\right) \qquad (13)$$

حيث أنّ :

$$\hat{\beta} = \frac{S_2^2}{S_1^2 + S_2^2}\hat{\beta}_1 + \frac{S_1^2}{S_1^2 + S_2^2}\hat{\beta}_2 \qquad (14)$$

وبضرب وبقسمة المعادلة (14) بالمقدار $\frac{\sigma_1^2}{\sigma_2^2}$ فإنّ مقدر $\hat{\beta}$ يصبح كالآتي :

$$\hat{\beta} = \frac{1}{1 + \theta u}\hat{\beta}_1 + \frac{\theta u}{1 + \theta u}\hat{\beta}_2 \qquad (15)$$

حيث أنّ :

$$\theta = \frac{\sigma_1^2}{\sigma_2^2} \qquad , \qquad u = \frac{S_1^2}{\sigma_1^2} \Big/ \frac{S_2^2}{\sigma_2^2}$$

$$\hat{\beta} - \beta = \frac{1}{1 + \theta u} x_1' e_1 + \frac{\theta u}{1 + \theta u} x_2' e_2 \qquad (16)$$

ويلاحظ بأنّ u , e_1 x_1' , e_2 x_2' مستقلة عن بعضها البعض وبالتالي فإنّ متوسط مربعات الخطــأ للمقدر $\widetilde{\beta}$ تكون كالآتي :

$$MSE(\widetilde{\beta}) = E\left[(\hat{\beta}-\beta)'(\hat{\beta}-\beta)\right]\Pr(u \geq \frac{\lambda}{\theta}) + E\left[(b-\beta)'(b-\beta)\right]\Pr(u < \frac{\lambda}{\theta}) \qquad (17)$$

ولإيجاد الحد الأول من متوسط مربعات الخطأ للمعادلة (17) فإنّ :

$$E\left[(\hat{\beta}-\beta)'(\hat{\beta}-\beta)\right]\Pr(u \geq \frac{\lambda}{\theta}) = P\int_{\frac{\lambda}{\theta}}^{\infty}\left[\frac{\sigma_1^2}{(1+\theta u)^2} + \frac{\theta^2 u^2 \sigma_2^2}{(1+\theta u)^2}\right] f(u)\,du \qquad (18)$$

$$f(u) = \begin{cases} \dfrac{m_1^{m_1}\ m_2^{m_2}\ u^{m_1-1}}{\beta(m_1,m_2)(m_1 u + m_2)^{m_1+m_2}} & 0 < u < \infty \\ 0 & 0/w \end{cases}$$

وبتبسيط المعادلة (18) نحصل على المقدار التالي :

$$\left[(\hat{\beta}-\beta)\right]\Pr(u \geq \frac{\lambda}{\theta}) = \frac{P m_1^{m_1} m_2^{m_2}}{\beta(m_1,m_2)}\left[\int_{\frac{\lambda}{\theta}}^{\infty}\frac{\sigma_1^2\, u^{m_1-1}}{(1+\theta u)^2 (m_1 u + m_2)^{m_1+m_2}}du + \int_{\frac{\lambda}{\theta}}^{\infty}\frac{\theta^2 u^2 \sigma_2^2\, u^{m_1-1}}{(1+\theta u)^2 (m_1 u + m_2)^{m_1+m_2}}du\right]$$

(19)

وبأستخدام أسلوب التحويلات (Transformation Technique) التالي فإن ناتج التكامل يكون:

$$u = \frac{m_2 z}{m_1(1-z)} \qquad , \qquad du = \frac{m_2}{m_1(1-z)^2}dz$$

$$E\left[(\hat{\beta}-\beta)'(\hat{\beta}-\beta)\right]\Pr(u\geq\frac{\lambda}{\theta})=\frac{P\sigma_1^2}{\beta(m_1,m_2)}\int_{z_\circ}^{1} z^{m_1-1}(1-z)^{m_2+1}\left[1-(1-\frac{\theta m_2}{m_1})z\right]^{-2}dz$$

$$+\frac{P\sigma_2^2}{\beta(m_1,m_2)}\int_{z_\circ}^{1} z^{m_1+1}(1-z)^{m_2-1}\left[1-(1-\frac{\theta m_2}{m_1})z\right]^{-2}dz \quad (20)$$

حيث أن :

$$z_\circ=\frac{m_1\lambda}{m_1\lambda+m_2\theta}$$

ولإيجاد الحد الثاني لمتوسط مربعات الخطأ للمقدر $\tilde{\beta}$ فإن :

$$E\left[(b-\beta)'(b-\beta)\right]\Pr(u<\frac{\lambda}{\theta})=E\left[(\frac{b_1+b_2}{2}-\beta)'(\frac{b_1+b_2}{2}-\beta)\right]\Pr(u<\frac{\lambda}{\theta}) \quad (21)$$

وبإجراء بعض التبسيطات والعمليات الحسابية نحصل على :

$$E\left[(b-\beta)'(b-\beta)\right]\Pr(u<\frac{\lambda}{\theta})=\frac{P}{4}(\sigma_1^2+\sigma_2^2)\Pr(u<\frac{\lambda}{\theta})=\int_0^{\frac{\lambda}{\theta}}\frac{P}{4}(\sigma_1^2+\sigma_2^2)$$

$$*\frac{m_1^{m_1}m_2^{m_2}u^{m_1-1}}{\beta(m_1,m_2)(m_1u+m_2)^{m_1+m_2}}du \quad (22)$$

وبأستخدام نفس أسلوب التحويلات التي استخدمت في الحد الأول نحصل على المقدار التالي :

$$E\left[(b-\beta)'(b-\beta)\right]\Pr(u<\frac{\lambda}{\theta})=P\sigma_2^2(\frac{1+\theta}{4})\frac{1}{\beta(m_1,m_2)}\int_0^{z_\circ}z^{m_1-1}(1-z)^{m_2-1}dz \quad (23)$$

حيث أنّ :

$$I_z(a,b)=\frac{1}{\beta(a,b)}\int_0^{z}t^{a-1}(1-t)^{b-1}dt$$

إذ أنّ $I_z(a,b)$ دالة بيتـا غيـر الكاملـة (Incomplete Beta Function) وبالتـالي فـإنّ $MSE(\widetilde{\beta})$ تكون صيغته النهائية بالشكل التالي :

$$MSE(\widetilde{\beta}) = P\sigma_2^2[\frac{1+\theta}{4} I_{t_\circ}(m_1,m_2) + \frac{\beta(m_1,m_2+2)\theta}{\beta(m_1,m_2)}\, {}_2F_1(m_1,2;m_1+m_2+2;1-\frac{m_2\theta}{m_1})$$

$$+\frac{\beta(m_1+2,m_2)m_2^2\theta^2}{\beta(m_1,m_2)m_2^2}\, {}_2F_1(m_1+2,2;m_1+m_2+2;1-\frac{m_2\theta}{m_1})] \qquad (24)$$

حيث أنّ ${}_2F_1(.,.;.;.)$ تمثل (Gauss Hyper geometric Function) والتي يمكن استخراج قيمها من خلال تطبيق المعادلات الموجودة في مصدر (1) والخاصة بمثل هذا النوع من الدوال .

التطبيق والنتائج :- Application and Results

بعد اشتقاق وإيجاد متوسط مربعات الخطأ لمقدر الاختبار الأولي $\widetilde{\beta}$ قام الباحث بكتابة برنـامج بلغة البيسك بطريقة سمبسون في التحليل العددي من أجل حساب تكامل دالة بيتا غير الكاملة فـي المعادلة رقم (23) وإيجاد قيم التكامل لها حسب الثوابت الموجودة في معادلـة متوسـط مربعـات الخطأ .

كما تمّ حساب دالة (Gauss Hyper Geometric) والتي يرمز لها ${}_2F_1(.,.;.;.)$ والتي عرفت في المعادلة رقم (24) من خلال العلاقات الموجودة في المصدر (1) وكالآتي :

$${}_2F_1(a,b;c;z) = \frac{1}{\beta(a,b)}\int_{t_\circ}^{1} t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\,dt \qquad (25)$$

حيث يمكن حساب المعادلة (25) من العلاقة التالية :

$$ {}_2F_1(a,b;c;z)=\sum_{n=0}^{\infty}\frac{\Gamma(a+n)\,\Gamma(b+n)\,z^n}{\Gamma(c+n)\;n!} \qquad (26) $$

حيث أنَ الثوابت الموجودة في المعادلة (26) تعرف بالشكل التالي :

$$ z=1-\frac{m_2\,\theta}{m_1}\ ,\quad c=m_1+m_2+2\ ,\quad b=2\ ,\quad a=m_1 $$

وبعد ذلك تمّ افتراض مجموعة من القيم إلى n_2, n_1 لحساب الثوابت $m_2=n_2-p\ ,\ m_1=n_1-p$ ، ومن ثم افتراض قيم إلى تباين العينتين الجزئيتين وهما $\sigma_2^2\ ,\ \sigma_1^2$ وتم اخذ حالة التساوي $\sigma_1^2=\sigma_2^2$ بالإضافة إلى حالة الأكبر $\sigma_1^2>\sigma_2^2$. وقد بُوبَتْ قيم متوسط مربعات الخطأ التي ظهرت من خلال حساب المعادلة رقم (24) في الجدول التالي :

جدول يوضَح قيم متوسط مربعات الخطأ لمقدر الاختبار الأولي $\tilde{\beta}$

n_1	n_2	$\sigma_1^2=\sigma_2^2=3$	$(\sigma_1^2>\sigma_2^2)=(5>3)$	$(\sigma_1^2>\sigma_2^2)=(10>4)$
10	10	3.3456	5.3270	5.9423
	15	3.2641	4.9624	3.7892
	20	2.9324	3.5435	4.1503
15	10	2.8527	6.3179	6.7324
	15	2.4688	4.6364	4.8249
	20	2.1373	3.3008	3.9427
20	10	4.0209	6.2683	6.5465
	15	3.1136	5.2118	5.9206
	20	1.7354	3.2387	2.7733

التعليق والاستنتاجات :-

يمكن توضيح الملاحظات التالية على قيم MSE المحسوبة في الجدول السابق وكالآتي :

1- عندما $\sigma_1^2 = \sigma_2^2$ يلاحظ بأن قيم متوسط مربعات الخطأ تكون قليلة ولكافة أحجام العينات المفترضة (n_2, n_1) ، حيث أن قيم $MSE(\tilde{\beta})$ تتناقص عند كافة أحجام العينات المتساوية والمختلفة .

2- عندما $\sigma_1^2 > \sigma_2^2$ يلاحظ بأن قيم متوسط مربعات الخطأ تكون صغيرة أيضا ولكافة أحجام العينات المفترضة (n_2, n_1) كما أن قيم $MSE(\tilde{\beta})$ تتناقص عند أحجام العينات المتساوية وغير المتساوية .

3- أما عند أحجام العينات المتساوية وغير المتساوية وبغض النظر عن تساوي أو عدم تساوي σ_2^2, σ_1^2 . فإننا نلاحظ تزايد قيم متوسط مربعات الخطأ لمقدر الاختبار الأولي $\tilde{\beta}$ ، ما عدا الحالتين التاليتين : $(n_1 = 10, n_2 = 20)$ ، $(n_1 = 20, n_2 = 20)$.

لذلك نستنتج بأن مقدر الاختبار الأولي الناتج من ربط مقدر (OLS) مع مقدر (SAE 2) يعطي نتائج كمتوسط مربعات الخطأ صغيرة وقليلة نسبياً .

وبالتالي فإننا نوصي إجراء مقارنة بين متوسط مربعات الخطأ لمقدر الاختبار الأولي وبين مقدر طريقة المربعات الصغرى ومقدر Aitken ذو المرحلتين، كما إننا نوصي بإيجاد المنطقة الحرجة المثلى (λ) (Optimal Critical Value) لمقدر الاختبار الأولي .

References :- المصـــادر

1- Abramowitz , M. and I.A. Stegun, (1972), "Handbook of Mathematical Functions"; Dover Publications, New York .

2- Mehta, J.S. and Gurland, (1969), "Combinations of Unbiased Estimators of the Mean which Consider Inequality of Unknown Variances", JASA, Vol.64, p.1042-1055 .

3- Toyoda, T. and T.D. Wallace, (1975), "Estimation of Variances After a Preliminary Test of Homogeneity and Optimal Level of Significance for the Pre-Test", Journal of Econometrics, Vol.3, p. 395-404 .

4- Toyoda, T. and K. Ohtani, (1980), "Estimation of Regression Coefficients After a Preliminary Test for Homoscedasticity" , Journal of Econometrics, Vol.12, p. 151-159 .

....

....

....