العنوان:	استخدام أسلوب البوتستراب الحصين لمعالجة مشكلة عدم تجانس التباين
المصدر:	المجلة العراقية للعلوم الإحصائية
الناشر:	جامعة الموصل - كلية علوم الحاسوب والرياضيات
المؤلف الرئيسي:	كاظم، أموري هادي
مؤلفين آخرين:	حسون، مريم، الصفاوي، صفاء يونس(م. مشارك)
المجلد/العدد:	ع4
محكمة:	نعم
التاريخ الميلادي:	2002
الشهر:	كانون الأول
الصفحات:	36 - 48
رقم MD:	866443
نوع المحتوى:	بحوث ومقالات
قواعد المعلومات:	EcoLink
مواضيع:	الإحصاء الرياضي
رابط:	http://search.mandumah.com/Record/866443

المجلة العراقية للعلوم الاحصائية (4) 2002
ص ص [36-48]

استخدام اسلوب البوتستراب الحصين لمعالجة مشكلة عدم تجانس التباين "Heteroscedasticity"

أموري هادي كاظم* صفاء يونس صفاوي** مريم حسون

الملخص:

في هذا البحث ستتم دراسة ما اذا كانت طرق البوتستراب Bootstrap حصينة او غير حصينة مقابل عدم تجانس التباين، حيث ان البوتستراب هو وجه اخر لاسلوب Jacknife وفكرته استخدام اسلوب المونت كارلو للمحاكاة المستندة الى التقدير اللامعلمي لتقدير التوزيع لدالة المشاهدات.

ومن خلال النتائج التي تم التوصل اليها فان مقدر التباين للبوتستراب مطابق لذلك في الحالة الاعتيادية، ويكون متحيزا وغير متسق في حالة الاخطاء غير المتجانسة. وبناء على ذلك لايمكن استخدام البوتستراب في حالة عدم تجانس التباين.

Using Bootstrap in Treating the Problem of Heteroscedasticity

ABSTRACT

This paper shows that the Robest Bootstrap method of estimation gives an unbiased and consistent estimators, and at the same time such a method treats the problem of Herteroscedasticity.

المقدمة وهدف البحث

ان الانحدار هو من اوسع الاساليب الاحصائية استخداما في مختلف العلوم وبخاصة الاقتصادية، وللحصول على نموذج جيد يتم تقدير معالمه بالطرائق الشائعة ومنها المربعات الصغرى التي تستلزم فروضا معينة لتطبيقها. وكما هو معروف في اغلب الظواهر الاقتصادية انها تعتمد على البيانات المقطعية Cross-section data غير المتجانسة وتوزيعها الاحتمالي غير معروف مما ينتج عنها مقدرات للتباين والتباين المشترك متحيزة تقاربيا تؤدي الى ابتعاد احصاءة الاختبار t و F عن قيمتها الحقيقية وتعطي قرارا خاطئا بشأن الفرضية.

* استاذ/ كلية الادارة والاقتصاد /جامعة بغداد.
** استاذ مساعد/ كلية علوم الحاسبات والرياضيات/ جامعة الموصل.

تمكن Eicker (1967، 1963) و White (1980) من تصحيح الاستدلال تقاربيا باستخدام HCCME (مقدر مصفوفة التباين والتباين المشترك المتسق بعدم التجانس heteroskedasticity consistent covariance matrix estimator) وتحت الشروط النظامية بين ان $T_n \to N(0,1)$ عندما $n \to \infty$. وفي (1985) درس Mackinno and White عدد من الانواع الممكنة لـ HCCME وبينا ان t و F لايزال متحيز خاصة في العينات الصغيرة. وبين Chesher and Jewitt (1987) علاقة التحيز بهيكلية النموذج. كما بين ان مقدر التباين متحيز للاسفل وعليه فاحتمال الرفض (RP) الصحيح للاختبار المعتمد على T_n يتجاوز (RP) الاسمي، وبين هذا الفرق بالنسبة الى العينات الصغيرة. وبعدها نفذ سلسلة من الدراسات باستخدام البوتستراب الا انه تبين ما ياتي:

1- لا يمكن تنفيذ البوتستراب باعادة معاينة بواقي OLS المستقلة في x.

2- لا يمكن تنفيذ البوتستراب بمعاينة المتغير العشوائي من نموذج معلمي لانه لم يحدد توزيع المتغير العشوائي ونوع التجانس غير معروف.

3- ان معاينة ازواج المشاهدات (x,y) لا تفترض القيد E(u/X: x=0) بعينة البوتستراب لذا استخدم الطريقة البديلة Wild Bootstrap التي قدمت من قبل Liu (1988). معززة بمقترح Wu(1986). وبين في عام (1993) قابليتها لتحسين التقارب لنموذج الانحدار بالاخطاء غير المتجانسة.

الا ان Fredman (1981) بين صحتها التقاربية (asymptotically valid) في حالة عدم تجانس التباين heteroskedasticity غير المعروف وذلك باسلوب اعادة معاينة ازواج المشاهدات للمبررات المبينة لاحقا. ومنه جاء هدف البحث الذي سيدرس ما اذا كانت طرائق البوتستراب حصينة ام غير حصينة مقابل عدم تجانس التباين وسيبين التحوير المناسب (كما في Wild Boots).

نشوء فكرة البوتستراب

ان البوتستراب هو وجه اخر لاسلوب Jackknife وفكرته هو استخدام محاكاة مونت كارلو المستندة الى التقدير اللامعلمي لتقدير التوزيع لدالة المشاهدات، هذا التوزيع يتم ايجاده بأبدال التوزيعات غير المعروفة بالتوزيع التجريبي للبيانات في تعريف الدالة الاحصائية، ثم اعادة معاينة البيانات لايجاد توزيع المتغيرات العشوائية الناتجة.

لقد برز اسلوب البوتستراب نتيجة للمشاكل التي تعرض لها اسلوب Jackknife الذي قدمه Quenouille-Tukey (1956-1958) لتقليل التحيز وتقدير الفترة الحصينة اثناء تطبيقه على نموذج الانحدار. اذ اعطى Miller (1974) اول تفصيل لاستخدام

Jakknife بنموذج الانحدار وذلك بشطب مشاهدة واحدة عشوائيا واعادة المعاينة من دون ارجاع الا ان هذه الطريقة تعرضت لمشكلتين هما:

1- اهمال الطبيعة غير المتزنة لبيانات الانحدار.

2- تقييد اختيار حجم اعادة العينة.

ونتيجة لهاتين المشكلتين فقد اشار Hinkely (1977) الى ثلاثة عيوب للمعالم المقدرة بهذه الطريقة واقترح لذلك تعديلا لها من خلال استخدام القيم الوهمية الموزونة التي امتازت بخاصية الاختيار المرن لحجم اعادة المعاينة وطبيعة الوزن. واستمر التفكير باسقاط عدد من المشاهدات في كل مرة الا انها ما زالت تتضمن عيوب اسقاط مشاهدة واحدة، لذلك تبنى Efron (1979) فكرة Jackkinfe باسلوب اكثر حصانة تمثل استخدام عينة عشوائية بسيطة مع الارجاع (عينة iid) لاعادة معاينة البيانات واعطى لها اسما مضللا Catchy name "Bootstrap" الذي يعود اساسه النظري الى المحاضرة التي القاها Efron (1977) في جامعة Stanford وناقش خلالها المشكلة الاتية: لعينة عشوائية ($X=x_1, \ldots\ldots x_n$) بحجم n مسحوبة من مجتمع له توزيع احتمالي غير معروف F، ثم بين Efron امكانية تطبيقه على نموذج الانحدار وبتطبيق الطرائق التقليدية ذاتها من خلال اسلوبي اعادة المعاينة مع الارجاع (المتمثل باعادة معاينة البواقي واعادة معاينة ازواج المشاهدات) وبين بعض مشاكلهما التي درست من قبل Singh (1981) وبين Freedman and Bikel (1981) شرعيتها للتوزيع التقاربي.

واقترح Wu (1986) اسلوبا بتحوير اعادة معاينة البواقي البوتسترابية لمعالجة الاخطاء العشوائية ذات التوزيع غير المعلوم وغير المتجانس وقدم هذا المقترح بعد اجراء تعديل عليه من قبل Liu (1988) واطلق عليه Wild Bootstrap. وبين Mammen (1993) امكانيته في تحسين التقارب لنموذج الانحدار وتم تطبيقه في حالة الانحدار غير المعلمي من قبل Hardleand (1993). كما بين Davidon (1999) قابلية الاسلوبين (الاصلي والمحور) على تحسين التقارب asymptotic refirement.

اسلوب البوتستراب في الانحدار

ان اسلوب البوتستراب في نموذج الانحدار يعطي نتائج الطرائق التقليدية حيث تتم عملية اعادة المعاينة باسلوبين هما:

الاول: اعادة معاينة البواقي، اذ تتم اعادة سحب عينة من تقديرات البواقي $\hat{e}$ من خلال العلاقة:

$$\hat{e} = y - X\hat{B}$$

وهذه التقديرات للاخطاء لها توزيع احتمالي غير معروف وليكن F، ويتم ايجاد $\hat{F}$ التوزيع التجريبي للاخطاء القياسية $\overset{*}{e}$ الممركزة حول الوسط:

$$\overset{*}{e} = \frac{1}{n}(\hat{e} - \hat{\mu})$$

حيث $\hat{\mu}$: الوسط الحسابي للاخطاء التقديرية $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\hat{e}_i$

وبفرض $\overset{*}{e}_1, \ldots\ldots, \overset{*}{e}_n$ مستقلة بصورة شرطية بتوزيع $\hat{F}$، يصبح نموذج الانحدار بالصيغة الاتية:

$$\bar{y} = X\hat{B} + \overset{*}{e}$$

حيث $\bar{y}$: تولدت من البيانات باستخدام نموذج الانحدار $\hat{B}$

$\hat{B}$ قيمة المعالم

$\hat{F}$ توزيع حدود الاضطراب

$\overset{*}{e}$: استخرجت من خلال اعادة المعاينة مع الارجاع للاخطاء المختزلة

فتصبح البيانات(x,y) التي تستخدم لاستخراج تقدير اخر لمتجه المعالم والذي هو تقدير المربعات الصغرى

$(x'x)^{-1}x'\overset{*}{y}$. $\overset{**}{B}$ وتكرر عملية اعادة المعاينة بـــ N من تكرار مونت كارلو

اساس اسلوب البوتستراب هو التوزيع $\sqrt{n}(\overset{**}{B} - \hat{B})$ الذي يمكن حسابه بصورة مباشرة من البيانات والذي يقترب من التوزيع $\sqrt{n}(\hat{B} - B)$ وهذا التقريب جيد في حالة (n) كبيرة. ولا يمكن تطبيق اسلوب $\sqrt{n}(\overset{**}{B} - \hat{B})$ في حالة عدم الحصول على البواقي المركزة قبل اعادة المعاينة وذلك: افرض ان متجهات الثوابت هي اما ان تضمن او تكون عمودية على فراغ space المتجه x فالتوزيع لـــ $\sqrt{n}(\overset{**}{B} - \hat{B})$ الذي يعتمد على $E_1, \ldots\ldots, E_n$ العشوائي يكون متحيزا ويكون طبيعيا محدودا غير منحل not dagenarte وهذا مخالف للحقيقة لان التوزيع التجريبي للبواقي غير المركزية يقترب الى F.

كما اثبت Freedman فشل عينة البوتستراب بالاحتمال والتوزيع التقاربي الشرطي المحوري أي ان*:

$$\sqrt{n}(\hat{B}^{*}-\hat{B})\xrightarrow[\text{weakly}]{\text{converg}}N(0,\sigma^{2}U^{-1})$$

$$\hat{\sigma}^{*}\xrightarrow{\text{con.}}\sigma$$

$$(x'x)^{\frac{1}{2}}(\hat{B}^{*}-\hat{B})/\hat{\sigma}^{*}\xrightarrow{\text{con.}}N(0,1)$$

الثاني: اعادة معاينة ازواج المشاهدات (x,y) باستخدام اسلوب البوتستراب تتم اعادة معاينة ازواج الشاهدات (x,y) عشوائيا مع الارجاع ومن دون افتراض القيد (E(u/X:x=0)** ، حيث يتم سحب عينة $(\dot{x},\dot{y})$iid من الازواج (x,y) فيصبح نموذج الانحدار

$$\dot{y}=\dot{x}B+e$$

فيكون التقدير بالصيغة

$$B\hat{}^{*}=(\dot{x}'\dot{x})^{-1}\dot{x}'\dot{y}$$

النتائج تحمل صفات مطابقة لسابقتها من حيث شرعيتها بالاحتمال والقانون الشرطي الذي اثبتها Freedman

$$\sqrt{m}(\dot{x}'\dot{x})\xrightarrow{\text{con.}}\sum$$

$$\sqrt{m}(\hat{B}^{*}-\dot{B})\rightarrow N(0,\sum^{-1}M\sum^{-1})$$

وان فكرة البوتستراب تعتمد على نوعين مختلفين من النماذج "الانحدار" و "الارتباط"

في نموذج الانحدار y=xB+e حيث x يكون محدودا fixed والخطأ يكون متجانسا homoscedastic

لكن في نموذج الارتباط y=xB+e يكون x عشوائيا random والخطأ غير متجانس heteroscedastic

حيث يقدر النموذج مستوى الانحدار لمجتمع معين بالاعتماد على عينة عشوائية بسيطة ويقيس اخطاء المعاينة، بفرض:x عشوائي و E ليا علاقة بـ x ، المتجهات (x_i,y_i) مستقلة ليا توزيع مشترك غير معروف (μ) بالفراغ R^{P+1} وان $E//x_iy_i//^4<\infty$ حيث //.// الفراغ الاقليدي، فحسب الطريقة الاقليدية xi المتجه الصفي $\sum=E(x_i'x_i)$ فتعرف B بانها متجه

* انظر (7) برهان النظرية في صفحة (1198)

** على اساسه يستخدم البديل Wild Bootstrap (4) (10) والذي انتقده Efron (12)

المعالم الذي يجعل $\min \; E//y_i-x_iB//^2$ أي المتجه $y_i-x_{ij}B$ عمودي على المتجـــه x_i أي ان $Ex_{ij}E_i=0$ وان $E_i=y_i-x_{ij}B$ فان

$$\hat{B} = \sum^{-1} E(x_i' y_i)$$

وان المصفوفة M محددة غير سالبة $M_{ij} = E(x_{ij} - x_{ik}\varepsilon_i^2)$ فالتقارب بالنموذج يختصر بالاتي:

$$\sqrt{n}(\hat{B}-B) = (\frac{1}{n}x'_{(n)}x_{(n)})^{-1}\frac{1}{\sqrt{n}}x'_{(n)}E(n)$$

وان

$$\frac{1}{n}x'x \underset{a.e}{\overset{\text{almost everywhere}}{\longrightarrow\!\!\!\rightarrow}} \sum$$

لذا فان $\sqrt{n}(\hat{B}-B)$ محاذٍ طبيعي بوسط صفر ومصفوفة تباين وتباين مشترك $\sum^{-1} M \sum^{-1}$، اذ ان Σ بنموذج الارتباط لها دور V في نموذج الانحدار المحاذي asymptotic اذ يكون مخالفا تماما ما لم يكن نموذج الارتباط متجانس والذي يفسر a.e $E(E_i^2/x_i) = \sigma^2$ حيث $\hat{\sigma}^2 = Eei^2$ وان $M = \sigma^2 \sum$

مقدر التباين Bootstrap الحصين (المكافئ لمقدر التباين Jackknife الحصين)

بين Beran (1984) ان تقدير Jackknife للتحيز والتباين والالتواء يقتـــرب الـــى نظيره البوتستراب وكل من هذه التقديرات متسقة وان تقديرهما للتبـــاين تقاربيـــا طبيعـــي و minimax . افرض نموذج الانحدار $y_i = x_i'B + ei$ حيث xi متجـــه محـــدد (k x 1)، B متجه المعالم (k x 1)، ei الاخطاء غير المرتبطة بوسط صفر وتبـــاين σ_i^2، $e=(e_1,\dots e_n)$ وان $y=(y_1,\dots y_n)$ و $x=(x_1,\dots x_n)$ أي النموذج $y=x'B+e$ وان $dig(\cdot{}^2_1,\dots,\sigma_n^2)$. , . V(e)، فمقدر OLS هو

$$\hat{B} = (x'x)^{-1}x'y$$

وان Jackknife الاعتيادي للنموذج معطى بالاتي:

افرض ان $\hat{B}_{(i)}$ تقدير لـــ B يحصل من اعادة تقدير $\hat{B}$ بعد حذف (i) من ازواج المشاهدات من العينة، لتقدير الدالة غير الخطية لـــ B و $\theta=g(B)$ و $\hat{\theta} = g(\hat{B})$ والقيمة الوهمية

$$P_i = n\hat{\theta} - (n-1)\hat{\theta}_{(i)}$$

التقدير النقطي الجاكي لـــ θ هو

$$\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}p_i$$

ومقدر التباين هو

$$\upsilon_J=\frac{1}{n(n-1)}\sum^{n}(p_i-\tilde{\theta})(p_i-\tilde{\theta})'$$

لكن اشار Hinkely الى ثلاثة عيوب للمعالم الخطية θ=B و $\sigma_i^2=\sigma_i^2$ المقدرة بهذه الطريقة هي:

1- $\tilde{B}$ غير متحيزة لكن لها تباين اكبر من مقدر $LSE(\hat{B})$

2- . متحيز لتقدير $V(\hat{B})$ و $V(\tilde{B})$

3- رتبة التحيز لـ $\tilde{\theta}$ بين (n^{-2}, n^{-1}) يعتمد على اتزان مصفوفة x.

هذه الانتقادات ناتجة من اهمال الطبيعة غير المتزنة لبيانات الانحدار لذلك اقترح Hinkely استخدام القيم الوهمية الموزونة المعرفة بالاتي:

$$\theta_i=\hat{\theta}+n(1-w_i)(\hat{\theta}-\hat{\theta}_{(i)})$$

وان $w_i=x_i'(x'x)^{-1}x_i$

وان مقدر الجاكي

$$\tilde{\theta}_w=\frac{1}{n}\sum_{i=1}^{n}\theta_i$$

والتباين

$$\upsilon_{H(1)}=\frac{1}{n(n-k)}\sum_{i=1}^{n}(\theta_i-\tilde{\theta}_w)(\theta_i-\tilde{\theta}_w)'$$

وعندما B=θ ، $\tilde{B}_w$ تطابق لـ $\hat{B}$ و $\upsilon_{H(1)}$ يكون:

$$\upsilon_{H(1)}=\sum\frac{(1-w_i)^2}{1-n^{-1}k}(\hat{B}_{(i)}-\hat{B})(\hat{B}_{(i)}-\hat{B})'$$
$$=(x'x)^{-1}\sum\frac{r_i^2}{1-n^{-1}k}x_ix_i'(x'x)^{-1}$$

حيث $r_i=y_i-x_i'\hat{B}$

وان $\upsilon_{H(1)}$ متحيز تقاربيا في حالة تجانس التباين لكنه حصين مقابل عدم تجانس التباين كما مبين لاحقا لذلك فان $\hat{\theta}_w$ تحتفظ بخصائص تقليل التحيز.

في عام (1984) اقترح Wu طريقة Jackknife العامة الموزونة والتي لا تمتلك العيوب المذكورة سابقا والمستندة الى:

$$\hat{B} = \frac{\sum_r |x'_s x_s| \hat{B}_s}{\sum_r |x'_s x_s|}$$

$$\Rightarrow \sum_r |x'_s x_s| (\hat{B}_s - \hat{B}) = 0$$

فان التباين Jackknife

$$\upsilon_{J,r}(\hat{\theta}) = \frac{r-k+1}{n-r} \sum_r w_s (\hat{\theta}_s - \hat{\theta})(\hat{\theta}_s - \hat{\theta})'$$

وان

$$w_s \varsigma |x'_s x_s| \qquad , \qquad \sum_r w_s = 1$$

وحسب قواعد المصفوفات فان مقدر Jackknife للتباين:

$$\upsilon_{J,r}(\hat{\theta}) = \begin{pmatrix} n-k \\ r-k+1 \end{pmatrix}^{-1} |x'x|^{-1} \sum_r |x'_s x_s| (\hat{\theta}_s - \hat{\theta})(\hat{\theta}_s - \hat{\theta})'$$

افرض ان $\hat{\theta} = g(\hat{B})$

حيث g دالة تمهيدية لــ B و $\hat{\theta}_s = g(\hat{B}_s)$ و $\sum_r$ المجموع العام لجميع المجاميع الجزئية ذات الحجم r وان $x'_s x_s$ غير مفردة nonsingular.

العامل العددي $\frac{r-k+1}{n-r}$ هو مطبق ظاهريا (externally) لــ $\hat{\theta}_s - \hat{\theta}$ بعد التحويل g يحصل مقدر اخر بتطبيق العامل نفسه داخليا لــ $\hat{B}_s - \hat{B}$ (internally) لذلك يصبح التباين:

$$\tilde{\upsilon}_{J,r}(\hat{\theta}) = \sum_r w_s (\tilde{\theta}_s - \hat{\theta})(\tilde{\theta}_s - \hat{\theta})' \qquad \ldots\ldots\ldots\ldots\ldots (1)$$

وان

$$\tilde{\theta} = g(\tilde{B}), \tilde{B}_s = \hat{B} + \left(\frac{r-k+1}{n-r}\right)^{1/2} (\hat{B}_s - \hat{B})$$

وتحت الشروط التمهيدية المعقولة under reasonable smoothness cond. لكلا التباينين $\upsilon_{J,r}(\hat{\theta})$ و $\tilde{\upsilon}_{J,r}(\hat{\theta})$ يقتربان الى مقدر التباين Jakknife الخطي:

$$\upsilon_{Lin} = g'(\hat{B})' \upsilon_{J,r}(\hat{B}) g'(\hat{B})$$

حيث $g'(\hat{B})$ مشتقة g بالنسبة الى $\hat{B}$

اذا $r = \frac{n+k-1}{2}$ فان العامل العددي $\left(\frac{r-k+1}{n-r}\right)$ يساوي واحدا وان $\upsilon_{J,r} = \tilde{\upsilon}_{J,r}$ وللمعلمة الخطية θ=B فان $\upsilon_{J,r}(\hat{B}) \approx \tilde{\upsilon}_{J,r}(\hat{B})$ وله الخصائص التالية:

1- غير متحيز لــ $V(\hat{B})$ عندما $V(e) = \sigma^2 I$ (حالة التجانس حسب نظرية 3 (11))

2- التعريف المناسب لــ $\upsilon_{J,k}$ يكافئ مقدر التباين الاعتيادي $\hat{\sigma}^2 (x'x)^{-1}$ (حسب نظرية 4 المصدر السابق نفسه).

3- $\upsilon_{J,n-1}$ هو حصين مقابل عدم تجانس تباين الخطأ كما مبين ادناه:

(حيث $\upsilon_{J,n-1}$ تعني تباين Jackknife بعد حذف مشاهدة واحدة ويكتب ايضا $\upsilon_{J(1)}$).

بما ان التباين للجاكنايف الموزنة العامة (حذف مشاهدة واحدة) هو

$$\upsilon_{J(1)} = (x'x)^{-1}\sum_{1}^{n}\frac{r_i^2}{1-w_i}x_i x_i'(x'x)^{-1} \qquad \ldots\ldots\ldots\ldots\ldots (2)$$

حيث $\hat{B}_{(i)}$ هي LSE بــ i من المشاهدات المحذوفة وان $r_i = y_i - x_i'\hat{B}$

$$w_i = x_i'(x'x)^{-1}x_i$$

وتحت افتراض عدم التجانس $V(e) = \mathrm{dig}\sigma_i^2$ فان تباين $\hat{B}$ هو

$$V(\hat{B}) = (x'x)^{-1}\sum_{1}^{n}\sigma_2^2 x_i'(x'x)^{-1} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

وبمقارنة (2) و (3) يقترح ان $\upsilon_{J(1)}$ حصين لتقدير $v(\hat{B})$ تحت افتراض عدم التجانس ويكون التقارب تحت الافتراضات الاتية:

1- x_n تشير الى مصفوفة x بالمعادلة y=xB+e لــ n مشاهدة، وان

$$\max_{1\le i\le n} x_i'(x_n'x_n)^{-1}x_i \le c/n$$

c مستقلة عن n

2- $\max \sigma_i^2 < \infty$

3- اكبر واقل قيمة ذاتية (eign value) لــ $n^{-1}x_n'x_n$ هي uniformly bounded محددة بين (0) والــ (∞)

4- عناصر محددة بانتظام uniformly bounded

يتوقف عدم التحيز لــ $\upsilon_{J(1)}$ لتقدير $v(\hat{B})$ على العلاقة

$$Er_i^2 = (1-w_i)\sigma_i^2$$

فان شرط التقارب الشرطي معطى بالقاعدة الاتية:

اذا $\sigma_i \neq \sigma_j$ لاي i,j $w_{ij} = x_i'(x'x)^{-1}x_j = 0$ ......(4) فان

$$Er_i^2 = (1-w_i)\sigma_i^2 \qquad \ldots\ldots\ldots\ldots (5)$$

وان

$$w_i = x_i'(x'x)^{-1}x_i$$

وحسب الافتراض (1) و (2) فان الشكل العام هو:

$$Er_i^2 = (1-w_i)\sigma_i^2 + O(n^{-1}) \qquad \ldots\ldots\ldots\ldots (6)$$

ونبرهن بالاتي:

$$\because r_i = y_i - x_i'\hat{B}$$
$$= e_i - x_i(x'x)^{-1}x'e$$
$$Er_i^2 = \sigma_i^2 - 2w_i\sigma_i^2 + \sum_{j=1}^{n} w_{ij}^2\sigma_j^2$$
$$= (1-w_i)\sigma_i^2 + \sum_{j=1}^{n} w_{ij}^2(\sigma_j^2 - \sigma_i^2) \qquad (7)$$

حيث $w_{ij} = x_i'(x'x)^{-1}x_j$ وان الحد الثاني من (7) ياتي من $w_i = \sum_{j=1}^{n} w_{ij}^2$ وواضح الان (5) تنتج من (4) ، افتراض (1) و (2) فان

$$\left|\sum_j w_{ij}^2(\sigma_j^2 - \sigma_i^2)\right| \le 2(\max_i \sigma_i^2)\sum w_{ij}^2 = 2(\max_i \sigma_i^2)w_i$$

وهذه برتبة $O(n^{-1})$، لذلك المعادلة (6) تنتج من (7)

وبمقارنة المعادلة (2) و (3) وحسب القاعدة المذكورة نحصل على النتيجة الاتية:

تحت شرط المعادلة (4) $\quad E\upsilon_{J(1)} = VAR(\hat{B})$ i-

تحت الافتراضين (1) و (2) $\quad E\upsilon_{J(1)} = VAR(\hat{B})(1+o(n^{-1}))$ ii-

تم تعميم هذه النتيجة لــ $\upsilon_{J(1)}$ من قبل Wu و Shao(1985) وان r عدد المشاهدات المحذوفة لاسلوب Jackknife.

وبما ان $\upsilon_{J(1)}$ يكافئ تقاربيا لــ $\upsilon_{H(1)}$ لذلك $\upsilon_{H(1)}$ حصين بمعقولية النتيجة اعلاه (ii) وعند تعطيل الافتراض $V(e) = \text{diag}\sigma_i^2$ فانه يكون:

بافتراض $v(e) = \sigma^2 I$ وتحت الشرط للافتراض (1)

$$\text{i)}\ E\upsilon_{H(1)} = VAR(\hat{B})(1+O(n^{-1}))$$

تحت افتراض $v(e) = \text{diag}\,\varepsilon_i^2$ وتحقق الافتراضين (1) و (2)

$$\text{ii)}\ E\upsilon_{H(1)} = VAR(\hat{B})(1+O(n^{-1}))$$

وبالنسبة الى مقدر التباين البوتسترابي (v) فانه يقترب الى $\upsilon_{H(1)}$ (1) (12) كما مبين:

$$\bar{\ } = \sum_{i=1}^{n} \; i \; i^{-1} \sum_{i=1}^{n} \; i \; i$$

وان

$$\overset{*}{B} - \hat{B} = (\sum_{1}^{n} \overset{*}{p_i}\, x_i x_i')^{-1} \sum_{1}^{n} \overset{*}{p_i}\, x_i r_i$$

حيث

$$r_i = y_i - x_i' \hat{B}$$

حيث $\left.\overset{*}{p_i}\right)_1^n$ لها توزيع ذو الحدين المتعدد multinomial $Mult_n(n,(1/n)1)$ واذا اهمل الاختلاف بـ $\sum_{1}^{n} \overset{*}{p_i}\, x_i x_i'$ ، وان $\sum_{1}^{n} x_i r_i = 0, E_*(\overset{*}{p_i}) = 1$

فان $\upsilon^* = E_*(\overset{*}{B} - \hat{B})(\overset{*}{B} - \hat{B})'$

يمكن ان يقترب بـ:

$$\approx (\sum_{1}^{n} (x_i x_i')^{-1} \operatorname{var}_* (\sum_{1}^{n} \overset{*}{p_i}\, x_i r_i)(\sum_{1}^{n} x_i x_i')^{-1}$$

أي

$$\upsilon^* = (x_i x_i')^{-1} \sum_{1}^{n} r_i^2 x_i x_i')(x'x)^{-1}$$

ولكي يقترب لـ $\upsilon_{H(1)}$ فانه يعدل بـ $\frac{n}{n-k}$ لذا

$$\upsilon^{*'} = \frac{n}{n-k} \upsilon^* \approx V_{H(1)} \qquad \ldots\ldots\ldots (8)$$

$\therefore \upsilon^{*'}$ حصين مقابل عدم تجانس التباين

اما اعادة معاينة البواقي البوتسترابية فان مقدر التباين يكون غير حصين مقابل عدم تجانس التباين كما مبين بالاتي:

ان اعادة معاينة البواقي تستند على سحب عينة iid $\{e_i^*\}_1^n$ من البواقي $\left.(\frac{r_i}{1-n^{-1}k})^{1/2}\right\}_1^n$ ومشاهدة البوتستراب تعرف

$$y_i^* = x_i' \hat{B} + e_i^*$$

حيث تعامل $\hat{B}$ كمعلمة حقيقية و $(\frac{r_i}{1-n^{-1}k})^{1/2}$ كمجتمع للاخطاء وتقدير البوتستراب $\hat{B} = (x'x)^{-1} x'y^*$ للمقدر غير الخطي $\hat{\theta} = g(\hat{B})$ ، مقدر التباين البوتستراب

$$V_b = E_*(\overset{*}{\theta}-\hat{\theta})(\overset{*}{\theta}-\hat{\theta})'$$

* تشير الى عينة البوتستراب (iid) من مجتمع البواقي، وان $\overset{*}{\theta}=\hat{\theta}$ غير الموزونة لـــ

$\theta = B$ وان $E_*\overset{*}{B} = \hat{B}, V_b = \hat{V} = \hat{\sigma}^2(x'x)^{-1}$

وان $\hat{\sigma}^2 = \frac{1}{n-k}\sum r_i^2$

وهذا يعني ان مقدر التباين البوتسترابي مطابق لمقدر التباين الاعتيادي $\hat{V}$ للاخطاء المتجانسة لذلك V_b غير متحيز ولكن يكون متحيز وغير متسق للاخطاء غير المتجانسة حيث التحيز في مقدر البوتستراب هو $\hat{B}_b = E_*\overset{*}{\theta}-\hat{\theta}$

ومن مفكوك تايلر

$$\overset{*}{\theta} = \hat{\theta} + g'(\hat{B})'(\overset{*}{B}-\hat{B}) + \frac{1}{2}(\overset{*}{B}-\hat{B})'g''(\hat{B})(\overset{*}{B}-\hat{B}) + \zeta_*$$

لذا فان

$$\hat{B}_b = \frac{1}{2}tr(g''(\hat{B})\upsilon_b)E_*(\zeta_*)$$

وبأخذ التوقع فان

$$E(\hat{B}_b) = B(\hat{\theta}) + (1 + O(n^{-1})) \qquad \text{..........} (9)$$

لان $E(E_*(\zeta_*)) = O(n^{-2})$

لذلك لايمكن تطبيق (9) في حالة عدم تجانس التباين بسبب υ_b متحيز كما ان سحب العينة يعتمد على افتراض البواقي متقلبة باعتدال، لذلك مزج البواقي لا يسمح لعدم التجانس لـــ ri و σ_i^2 بالتاثير بتقدير التباين كما في حالة Jackknife، لذلك فانه غير حصين.

كما بين Wu (1986) اسلوب لمعالجة عدم التجانس بطريقة مشابهة لطريقة Jackknife (تم عرضه من قبل Liu (1988) باسم Wild Bootstrap والذي يكافئ $\upsilon_{J(1)}$، لــذلك يتمتــع بخصائص التحيز الحصين مقابل عدم تجانس التباين وذلك :

افرض:

$$y_i^* = x_i'\hat{B} + \frac{r_i}{\sqrt{1-w_i}}t_i^* \qquad \text{...........} (10)$$

وان $w_i = x_i'(x'x)^{-1}x_i$ $\qquad r_i = y_i - x_i\hat{B}$

وان $t^* = t_i^*\}_1^n$ تحصل بموجب اعادة المعاينة * الافتراضات الاتية

$E_*t^*=0 \qquad , \qquad Var_*(t^*)=I$

وان $\overset{*}{B} = (x'x)^{-1}x'y^*$

حيث $y^*=(y^*_1,\ldots\ldots y^*_n)$

الاستدلال حول B او $\theta=g(B)$ من خلال تقلبية $\overset{*}{B}$ فيقدر $Var(\hat{\theta})$ بالاتي

$$\ldots E_*(\hat{.}^*\ \hat{.})(\hat{.}^*\ \hat{\theta})' \qquad \hat{\theta}^* = g(\hat{B})^*, \hat{\theta} = g(\hat{B})$$

وللمعالم الخطية $\theta=B$ فان

$$\upsilon_* = (x'x)^{-1}\sum\frac{r_i^2}{1-w_i}x_i x_i'(x'x)^{-1}$$

$$\approx \upsilon_{J(1)}$$

الا ان Eforn (10) انتقد اسلوب Wild Bootstrap.

المصادر

1. Beran, R. (1984). "Jackknife Approximation to bootstrap estimates", Annals of Stat., 12, 101-118.
2. Beran, R. (1986). Discussion of "Jackknife, bootstrap and other resampling methods regression analysis" by Wu., Annal of Stat. 14, 1295-1298.
3. Chesher A., and Jewitt (1987). "The bais of a heteroskedasticy consistent covariance matrix estimator", Econom. etrica, 55, 1217-1222.
4. Davidson and Department Economics Queen's university (2001). "The wild Bootstrap, Tamed at least", Discussion paper No. DARP 58.
5. Efron, (1979). "Bootstrap methods: another look at the Jackknife", Annals of Stat. 7, 1-26.
6. Freedman, D. A. and Bickel, P. J. (1981). "some a symptotic theory for the bootstrap", the Annal of Stat.. 9, 1196-1917.
7. Freedman, D. A. (1981). "Bootstrapping regression models", Annals of stat. 9, 1218-1228.
8. Hahn, J. (1996). "A note on bootstrapping generalized method of moments estimator", Econometric Theory, 12, 187-197.
9. Hoprowitz, J. L. (1998). "Bootstrap method for median regression models", Econometrica, 66, 1327-1351.
10. Mammen, E. (1993). "Bootstrap and Wild bootstrap for high dimensional linear models", Annals of stat. 21, 255-285.
11. Wu, C. F. J. (1986). "Jackknife, bootstrap and other resampling methods in regression analysis", Annals of Stat., 14, 1261-1295.