مشكلة عدم ثبات التباين (عدم التجانس)

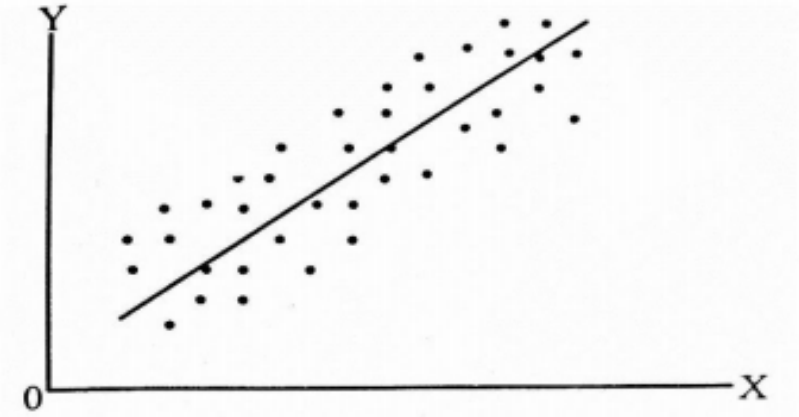
محاضرة رقم 3

مشكلة عدم ثبات التباين (عدم التجانس) Heteroskedasticity

1.بيان المشكلة

ان استعمال النماذج الانحدارية المتعددة خاصة يتوقف على مدى احترام الفرضيات الأساسية لطريقة التقدير المستعملة، ومن بين هذه الفرضيات فرضية تجانس تباين الخطأ أي أن تباين المتغير العشوائي أو البواقي يكون متجانسا وثابتا، غير أن هذا التجانس لا يكون دائما صحيحا عند دراسة حالات تطبيقية واقعية وباستعمال بيانات مقطعية جزئية يظر تشتت كبير للبواقي. وبالتالي لا يظهر التجانس في هذا التباين حيث يكون لكل مشاهدة تباين للخطأ يختلف عن المشاهدة الأخرى. كما هو موضح من خلال الشكل التالي:

الشكل رقم(3.5): حالة ثبات التباين



وعليه فان عدم تحقق فرضية ثبات تباين الخطأ لكل قيم المتغيرات المستقلة في الانحدار فإننا نواجه مشكلة عدم ثبات التباين، مما يؤدي الى تقديرات غير كفؤة، وفي هذه الحالة تكون الأخطاء حول خط الانحدار المقدرة ليس لها نفس التباين، أي أن[4]:

$$V(u_i) = \sigma_{u_i}^2 \neq \sigma_u^2$$

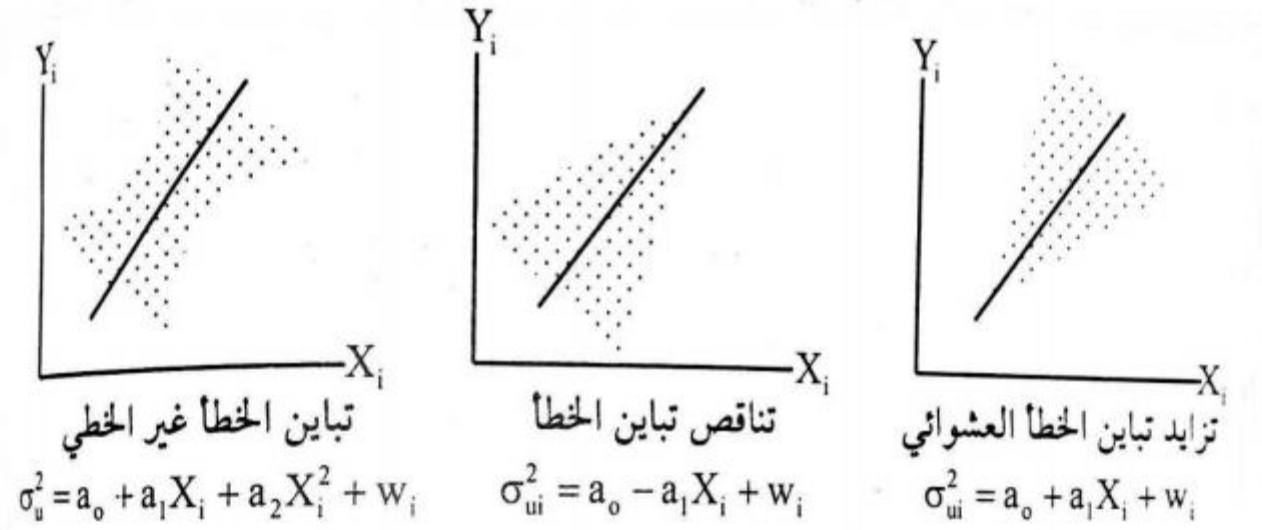
لاحظ انه لايمكن التعبير عن تباين الأخطاء بمقدار ثابت (أي σ_u^2) بسبب ان تباين الخطأ يختلف من مشاهدة إلى أخرى، مما يترتب عليه أيضا عدم القدرة على التعبير على مصفوفة التباين والتباين المشترك بالصيغة ($\sigma_u^2 I_n$).

$$cov(u) = \begin{pmatrix} \sigma_{u_1}^2 & 0 \dots & 0 \\ 0 & \sigma_{u_2}^2 \dots & 0 \\ 0 & 0 \dots & \sigma_{u_n}^2 \end{pmatrix}$$

لاحظ هنا ان: $\sigma_{u_1}^2 \neq \sigma_{u_2}^2 \neq \cdots \neq \sigma_{u_n}^2$

ويمكن تفسير عدم ثبات التباين بيانيا من خلال تقلب كل من الأخطاء العشوائية حول وسطها دالة في المتغير المستقل X_i، وفي مثل هذه الحالة يعتمد شكل عدم تجانس الخطأ على شكل العلاقة بين تباين الأخطاء والمتغير المستقل، وفيما يلي بعض من هذه الاشكال:

الشكل رقم (5.5): أشكال حالات عدم تباين الأخطاء العشوائية



المصدر: حسام علي داود وخالد محمد السواعي، مرجع سبق ذكره، ص275.

وهكذا يتضح لنا أن اتجاه تباين الخطأ نحو التزايد مع زيادة قيمة X_i ، أو اتجاهه نحو التناقص مع زيادة قيمة X_i ، أو أي شكل آخر غير الثبات عند كل مستوى للمتغير المستقل X_i. وهذا ما يعرف بمشكلة عدم ثبات تباين الأخطاء العشوائية.

2.، أثر وجود المشكلة

يترتب على وجود مشكلة عدم ثبات التباين عدة آثار، أهمها:

1- عند وجود مشكلة عدم ثبات التباين تظل تقديرات المعالم باستخدام طريقة المربعات الصغرى العادية غير متحيزة ومتسقة، لكنها تفقد خاصية الكفاءة، وذلك لأن تباين التقديرات في ظل مشكلة عدم ثبات التباين لا يكون أقل ما يمكن سواء في العينات الصغيرة أو الكبيرة، وهذا يني أنها أقل مصداقية من تنبؤات أخرى تبنى على طرق تخلو من مشكلة عدم ثبات التباين مثل طريقة المربعات الصغرى المعممة (GLS).

2- لا يمكن صيغة تباين المعلمات من أجل الوصول إلى اختبارات المعنوية أو تكوين مجالات الثقة، والسبب في ذلك يرجع إلى أن الحصول على تباينات التقديرات تم في ظل افتراض تجانس تباين أخطاء جميع المشاهدات أي (σ_u^2)، مما ادى إلى أخذ σ_u^2 كعامل مشترك.

$$V(\hat{a}) = \sigma_u^2 \frac{\sum x^2}{n\sum x^2} \, , V(\hat{b}) = \sigma_u^2 \frac{1}{n\sum x^2}$$

بيد أنه في ظل افتراض عدم ثبات التباين من مشاهدة إلى أخرى نتيجة لتغييره مع تغير قيمة المتغير المستقل X_i، فإنه لا يمكن أخذ $(\sigma_{u_i}^2)$ كعامل مشترك ومن ثم لا يمكن أن تتحقق الصيغ السابقة بل نكون بصدد صيغ مختلفة تماما.

3- مشكلة عدم ثبات التباين تؤثر على دقة التقديرات المتحصل عليها وينعكس في أن القيم المتنبأ بها للدالة تصبح غير دقيقة (أي عدم كفاءة التنبؤات).

3. طرق الكشف عن المشكلة

هناك عدة طرق للكشف عن مشكلة عدم ثبات تباين الأخطاء، ومن أهمها الطريقة البيانية واختبار جولد فيلد وكوانت. وفيما يلي عرض لمحتوى هذه الاختبارات:

أ- الطريقة البيانية

يتم اجراء انحدار بطريقة المربعات الصغرى لبيانات العينة والحصول على البواقي $\hat{e}_i = y_i - \hat{y}_i$ ثم نحسب قيم مربع الأخطاء $\hat{e}_i^2$.

ثم تتم دراسة الرسوم البيانية للعلاقة بين مربع الأخطاء والمتغير المستقل X_i ومن ثم يمكن استنتاج فيما إذا كان هناك ثبات للتباين أو عدم ثبات التباين.

ب- اختبار جولد فيلد وكوانت Goldfeld and Quandt Test

يستخدم هذا الاختبار في العينات الكبرى، كما يشترط أن يكون حجم العينة مساويا على الأقل ضعف عدد المعالم المراد تقديرها. كما يفترض في هذا الاختبار أن المتغير العشوائي يتبع التوزيع الطبيعي، وعدم وجود الارتباط الذاتي للأخطاء:

$$E(u_i, u_j) = 0 \quad \forall i \neq j$$

يتم تقسيم العينة الى ثلاثة اقسام ويتم استبعاد القسم الذي في المنتصف. ثم يحسب تباين الأخطاء بالنسبة للجزء لأول والجزء الثالث، ثم اختبار مدى تساويهما باستخدام اختبار F، وذلك باتباع الخطوات التالية:

- ترتيب المشاهدات تصاعديا طبقا لقيم المتغير المستقل X_i؛
- اختيار عدد من المشاهدات التي تقع في الوسط ونقوم بحذفها، على الاغلب يكون عددها يمثل ربع العدد الكلي للمشاهدات أو ثلثه؛
- بقية البيانات والتي قسمت الى مجموعتين تمثل الأولى القيم الصغيرة للمتغير X_i وليكن حجمها n_1، والثانية تضم القيم الكبيرة للمتغير X_i وليكن حجمها n_2؛
- يتم تقدير معادلة انحدار لكل مجموعة بطريقة المربعات الصغرى العادية بهدف الوصول إلى مجموع مربع البواقي لكل منهما $(\sum \hat{e}_1^2 , \sum \hat{e}_2^2)$:
- للحصول على تباين الأخطاء للمجموعتين الأولى والثانية نقوم بتقسيم مجموع مربعات الأخطاء لكل مجموعة على درجات الحرية df، حيث k تمثل عدد المعالم المقدرة في النموذج.
- حساب الإحصائية F المحسوبة (F_{cal})، حيث تتبع توزيع F برجتي الحرية n_2-k و n_1-k، والجدول الموالي يلخص خطوات تطبيق الاختبار:

١. خطوات تطبيق اختبار جولد فيلد وكوانت Goldfeld and Quandt Test

الخطوة	المجموعة الأولى (القيم الصغيرة للمتغير المستقل)	المجموعة الثانية (القيم الكبيرة للمتغير المستقل)
1	$Y_{1i} = a_1 + b_1X_{1i} + e_{1i}$ $\hat{Y}_1 = \hat{a}_1 + \hat{b}_1X_{1i}$	$Y_{2i} = a_2 + b_2X_{2i} + e_{2i}$ $\hat{Y}_2 = \hat{a}_2 + \hat{b}_2X_{2i}$
2	$\hat{e}_{1i} = y_{1i} - \hat{y}_{1i} \Rightarrow \sum \hat{e}^2_{1i}$	$\hat{e}_{2i} = y_{2i} - \hat{y}_{2i} \Rightarrow \sum \hat{e}^2_{2i}$
3	$df_1 = n_1 - k$	$df_2 = n_2 - k$
4	تباين الأخطاء: $\hat{\sigma}^2_{e_{1i}} = \frac{\sum \hat{e}^2_{1i}}{df_1}$	تباين الأخطاء: $\hat{\sigma}^2_{e_{2i}} = \frac{\sum \hat{e}^2_{2i}}{df_2}$
5	$F_{cal} = \frac{\hat{\sigma}^2_{e_{2i}}}{\hat{\sigma}^2_{e_{1i}}} = \frac{\sum \hat{e}^2_{2i}/n_2 - k}{\sum \hat{e}^2_{1i}/n_1 - k}$	
6	$F_{tab} = F_{n_2-k,n_1-k}$	

- وبمقارنة حساب الإحصائية F المحسوبة (F_{cal})، بقيمة F الجدولية(F_{tab}) **نقبل فرضية العدم بثبات التباين عندما يكون $F_{tab} > F_{cal}$** ، والعكس صحيح.

مثال(4): فيما يلي بيانات عن الدخل والاستهلاك في دولة معينة من 1983-2013:

قبل ترتيب الدخل X_i تصاعدياً

السنة	الدخل	استهلاك
1983	3838	3600
1984	2770	2782
1985	3580	3426
1986	3766	3550
1987	4238	3722
1988	4780	3646
1989	4930	3700
1990	5038	3758
1991	4904	3942
1992	4778	4328
1993	4314	4314
1994	4652	4328
1995	4586	4350
1996	4856	4518
1997	4902	4498
1998	4954	4530
1999	5074	4524
2000	5758	4626
2001	5236	4838
2002	5378	4890
2003	5392	4902
2004	5362	4870
2005	5490	5044
2006	5492	5072
2007	5556	5092
2008	5698	5236
2009	5802	5358
2010	6124	5572
2001	7494	5806
2012	6652	6028
2013	6850	6134

بعد ترتيب الدخل X_i تصاعدياً

السنة	دخل	استهلاك
1984	2770	2782
1985	3580	3426
1986	3766	3550
1983	3838	3600
1987	4238	3722
1993	4314	4314
1995	4586	4350
1994	4652	4328
1992	4778	4328
1988	4780	3646
1996	4856	4518
1997	**4902**	**4498**
1991	**4904**	**3942**
1989	**4930**	**3700**
1998	**4954**	**4530**
1990	**5038**	**3758**
1999	**5074**	**4524**
2001	**5236**	**4838**
2004	**5362**	**4870**
2002	**5378**	**4890**
2003	**5392**	**4902**
2005	5490	5044
2006	5492	5072
2007	5556	5092
2008	5698	5236
2000	5758	4626
2009	5802	5358
2010	6124	5572
2012	6652	6028
2013	6850	6134
2011	7494	5806

المطلوب: اختبر وجود مشكل عدم ثبات تباين الأخطاء العشوائية عند مستوى معنوية 5%.

الحل:

بداية يجب ترتيب المشاهدات تصاعديا كما هو في الجدول المعطى وتقسيم البيانات الى ثلاث مجموعات، يتم حذف ربع البيانات الوسطى وعددها 9 مشاهدات، وبالتالي نحصل على مجموعتين من البيانات وعدد مشاهداتها 11. وبعد القيام بعملية الانحدار لكل مجموعة حصلنا نتائج كل مجموعة كما يلي:

انحدار المجموعة 1:

$$Y_{1i} = 819.76 + 0.727X_{1i}$$

$$\bar{R}^2 = 0.744$$

$$\sum \hat{e}_{1i}^2 = 667723.6 \qquad df_1 = 11-2=9$$

انحدار المجموعة 2:

$$Y_{2i} = 1847.856 + 0.581X_{2i}$$

$$\bar{R}^2 = 0.647$$

$$\sum \hat{e}_{2i}^2 = 732280.73 \qquad df_2 = 11-2=9$$

وبعد الحصول على مجموع مربعات الأخطاء يمكن اجراء الاختبار وذلك بحساب الإحصائية F المحسوبة (F_{cal}) ومقارنتها بقيمة F الجدولية(F_{tab}). ونجري الاختبار التالي: $H_0 : \hat{\sigma}_{ui}^2 = \hat{\sigma}_u^2$

$$F_{cal} = \frac{\sum \hat{e}_{2i}^2 / n_2 - k}{\sum \hat{e}_{1i}^2 / n_1 - k} = \frac{732280.73/11-2}{667723.6/11-2} = 1.097$$

$$F_{tab} = F_{(9;9;0.05)} = 3.18$$

بما ان قيمة F المحسوبة (F_{cal}) أقل من قيمة F الجدولية(F_{tab}) فإننا نقبل فرضية العدم H_0 أي هناك ثبات لتباين الأخطاء.

4. طرق تصحيح المشكلة

غالبا يتم حل هذه المشكلة بقسمة كل حدود النموذج على X ثم إعادة تقدير الانحدار باستخدام المتغيرات المحولة.

مثال اسفرت نتائج انحدار متوسط الأجور Y على عدد العمال X في 30 شركة في احدى الصناعات على ما يلي:

$$Y=7.5+0.009X \quad R^2=0.90$$

(40.27) (16.10)

نتائج انحدار متوسط الأجور Y على عدد العمال X في الاثني عشرة مشاهدة الأولى على الترتيب

$$Y=7.5+0.009X \quad R^2=0.66 \quad \sum \hat{e}_{1i}^2=0.507$$

(40.27) (16.10)

$$Y=6.1+0.013X \quad R^2=0.60 \quad \sum \hat{e}_{1i}^2=3.095$$

(4.16) (3.89)

حيث أن ESS1/ESS2= 3.09095 تجاوز F10.10=2.97 عند مستوى معنوية 5% ومنه نقبل فرضية وجود اختلاف التباين.

وباعادة تقدير النموذج المحول لتصحيح اختلاف التباين نحصل على:

$$Y/X=0.008+7.8(1/X) \quad R^2=0.99 \quad \sum \hat{e}_{1i}^2=0.507$$

(14.43) (75.58)

تمرين 01: يوضح الجدول بيانات عن الدخل الشخصى الممكن إنفاقه بعد الضريبة ونفقات الاستهلاك الشخصى فى الولايات المتحدة بالألف دولار عن الفترة من 1962 إلى 1979 . تفترض النظرية الاقتصادية أن النفقات الحقيقية للاستهلاك الشخصى فى عام ما (C_t) هو فى الأصل دالة خطية للدخل الحقيقى الصافى لنفس العام (Y_{d_t}) .

(أ) وضح لماذا تشير معادلة انحدار C_t على Y_{d_t} إلى وجود مشكلة ارتباط ذاتى بينما لا يشير انحدار ΔC_t على ΔY_{d_t} إلى وجود تلك المشكلة ؟

(ب) اشرح كيفية إجراء معادلة انحدار ΔC_t على ΔY_{d_t} .

C: $\hat{C}_t = -11.40 + 0.93 Y_{d_t}$ $\quad R^2 = 0.999$

$= 0.75 \quad d \quad (144.33)$

D: $\widehat{\Delta C_t} = 0.94 \Delta Y_{d_t}$ $\quad R^2 = 0.960$

$= 2.44 \quad d \quad (37.79)$

الحـــل

يشير الانحدار الأول ، **C** ، إلى وجود مشكلة ارتباط ذاتى عند مستوى الأهمية 5% ، وذلك لأن

$d = 0.75$ أى أقل من القيمة الحرجة $d_L = 0.95$ عند $k' = 1$ (عدد المتغيرات المستقلة) و $n = 18$ (عدد المشاهدات) بالجدول C.4 فى الملحق C . أما الانحدار الثانى ، **D** ، فلا يشير إلى وجود هذه المشكلة عند مستوى 5% لأن $d = 2.24$ ، وهو أكبر من القيمة الحرجة $d_U = 1.12$ عند $k' = 1$ و $n = 18$ بالجدول C.4 .

(ب)لإجراء انحدار للتغير فى الاستهلاك على التغير فى الدخل نقوم أولاً بحساب معدلات التغير لكل عنصر اقتصادى ، ثم نقوم بإجراء الانحدار باستخدام المتغيرات الناتجة . وبالتحديد فإننا نقوم بطرح القيمة الأولى من القيمة الثانية بكل عامود بيانات للحصول على عمود جديد به البيانات الجديدة لكل متغير . وبهذا لا نفقد سوى مشاهدة واحدة ، لكن بما أنه يتم إجراء الانحدار فى التغيرات الأولى بدون حدود ثابتة (كما هو الحال هنا) ، فإن عدد درجات الحرية لا يتغير

السنة	الدخل	الاستهلاك
1962	383.9	355.2
1963	402.8	374.6
1964	437.0	400.4
1965	472.2	430.2
1966	510.4	464.8
1967	544.5	490.4
1968	588.1	535.9
1969	630.4	579.7
1970	685.9	618.8
1971	742.8	668.2
1972	801.3	733.0
1973	901.7	809.9
1974	984.6	889.6
1975	1,086.7	979.1
1976	1,184.5	1,089.9
1977	1,305.1	1,210.0
1978	1,458.4	1,350.8
1979	1,623.2	1,509.8

التمرين 02:

وضح :

(أ) ما هو الارتباط الذاتى ولماذا يعد مشكلة ؟

(ب) كيف يمكن التحقق من وجود هذه المشكلة ؟

(ج) كيف يمكن التغلب عليها ؟

الحـل

(أ) تنشأ مشكلة الارتباط الذاتى فى بيانات السلاسل الزمنية عندما يكون للانحرافات المعيارية المتتابعة نفس الإشارة أو تغيير الإشارة يتم كثيراً مما يؤدى إلى قيم مبالغ فيها لإحصاء t مما يؤثر على قيم R^2 و F .

(ب) يمكن التحقق من وجود مشكلة الارتباط الذاتى من خلال دراسة شكل يحتوى على الانحرافات أو النتائج مع مرور الوقت . والأكثر انتشاراً ودقة هو استخدام اختبار Durbin-Watson

مصدر المحاضرات
الدكتورة اشواق قدور
والدكتور عادل عبد الغني