79. The Heteroscedasticity Problem مشكلة عدم تجانس التباين

من الافتراضات الاساسية التي يقوم عليها النموذج الخطي (البسيط والعام) هي ثبات التباين لحدود الخطأ (تجانس تباين الخطأ) لجميع المشاهدات والتي يعبر عنها في النموذج الخطي البسيط بالشكل الآتي

$$E(u_i^2) = \sigma_u^2$$

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

وبموجب هذه الفرضية فان

$$\sigma_{u_1}^2 = \sigma_{u_2}^2 = \cdots = \sigma_{u_n}^2$$

وفي واقع الامر فاننا كثيراً ما نواجه حالات يتعذر فيها استيفاء الشرط اعلاه، ومن ثم فان التباين لا يكون ثابتاً بل يختلف لكل مشاهدة من مشاهدات العينة وتصبح لدينا قيم مختلفة وغير ثابتة لتباينات حدود الخطأ العشوائية، وعليه فان القطر الرئيسي لمصفوفة التباينات والتباينات المشتركة الخاصة بحدود الخطأ تحتوي على قيم مختلفة وغير ثابتة أي ان

$$E(u_i^2) \neq \sigma_u^2$$

$$\sigma_{u_1}^2 \neq \sigma_{u_2}^2 \neq \cdots \neq \sigma_{u_n}^2$$

تحدث هذه أساساً في الدراسات القياسية وخاصة الدراسات التي تعتمد على البيانات الاحصائية التي تأخذ شكل البيانات المقطعية كما هو الحال في بيانات بحوث ميزانية الأسرة التي تشتمل على أسر متباينة بشكل كبير في مستويات دخولها، فتشتت مشاهدات البيانات المقطعية الخاصة بالمتغير المعتمد قد يختلف اختلافاً شديداً من مستوى الى اخر في مستويات المتغير المستقل، مثال ذلك دراسة دالة الاستهلاك التي تعتمد على دخل وانفاقات العوائل على مختلف السلع والخدمات، فالعوائل ذات الدخول المرتفعة تتمتع بمرونة كبيرة في الانفاقات، اما انفاقات العوائل ذات الدخول الواطئة (المنخفضة) فانه يقع عادة ضمن حدود

ضيقة، وكذلك فإن التباين عند فئة الدخول المرتفعة تكون أكبر من التباين عند فئة الدخول المنخفضة .. وهكذا نجد أن فرضية تجانس تباين الخطأ تصبح تحكمية الحدوث في مثل هذه الحالات وإن افتراضنا هذا يؤدي الى حدوث مشكلة عدم تجانس التباين

Heteroscedasticity problem

وفي ظل وجود مشكلة عدم تجانس التباين تكون استخدام طريقة (OLS) لتقدير معالم النموذج غير مجدية، حيث أن المعالم المقدرة بمثل هذا الأسلوب سوف لن تكون أفضل تقدير خطي غير متحيز (BLUE) أي لا تمتلك المعالم المقدرة بهذا الأسلوب فاعلية أقل تباين ممكن.

أسباب عدم تجانس تباين حد الخطأ

هناك عدة أسباب لظهور هذه المشكلة ~~[illegible]~~ ومن أهمها الآتي:

1- سلوكية وتصرف الأفراد والتي تقل الأخطاء فيها بمرور الزمن وكذلك فإن تباين حد الخطأ (σ^2) يتناقص هو الآخر خلال الفترة نفسها.

2- يزداد تباين حد الخطأ (σ^2) مع زيادة الدخل وذلك لتباين وتعدد اختيارات الناس في سلوكهم. مثال ذلك تباين الإنفاق على المواد الغذائية بين الأسر يمكن أن يزيد بزيادة دخل الأسرة.

3- مع تحسن أساليب جمع البيانات يقل تباين حد الخطأ (σ^2) فجمع البيانات الدقيقة والواقعية تقلل من الأخطاء. مثال ذلك أن الأخطاء التي ترد في بيانات المؤسسات المصرفية والتي تستخدم الحاسبة في تحليل البيانات تكون أقل من مثيلاتها في المؤسسات التي لا تستخدم الحاسبة الآلية.

(82)

③ - اختبارات الكشف عن وجود مشكلة عدم ثبات تجانس التباين

هناك عدة اختبارات يمكن بواسطتها الكشف عن وجود مشكلة عدم ثبات تجانس التباين منها

1- اختبار كولد فيلد - وكوانت Goldfeld and Quandt test

2- اختبار معامل ارتباط الرتب لسبيرمان Spearman's Rand Correlation Coefficient test

3- اختبار بارك - كليجسر Park - Glejser test

4- اختبار بارتليت Bartlett test

1- اختبار كولد فيلد - وكوانت

يعد هذا الاختبار من الاختبارات المهمة لغرض الكشف عن مشكلة عدم تجانس تباين الخطأ ويتم استخدامه مع العينات الكبيرة. وتتلخص خطوات إجراء هذا الاختبار كما يلي:

1- ترتيب قيم المتغير المستقل X_i تصاعدياً

2- حذف المشاهدات الوسيطة (المركزية) من بيانات العينة ويفضل حذف $(\frac{1}{5})$ من (حجم العينة) المشاهدات الكلية، بهدف جعل الاختبار الإحصائي أكثر حساسية، ويفضل الحذف (6 - 8) مشاهدات عندما تكون العينة تتكون من (30) مشاهدة

3- نقسم المشاهدات الباقية على عينتين جزئيتين متساويتين احدهما تتضمن العينة الاولى من قيم X_i الصغيرة وتتضمن العينة الثانية من قيم X_i الكبيرة

4- يتم احتساب انحدار لكل عينة جزئية على حدة ثم يتم احتساب تباين حد الخطأ العشوائي لكل عينة جزئية وفقاً للصيغة التالية

تباين الخطأ للعينة الجزئية الاولى.

$$\sigma_{u_1}^2 = \frac{\sum e_1^2}{n_1 - K}$$

تباين الخطأ للعينة الجزئية الثانية.

$$\sigma_{u_2}^2 = \frac{\sum e_2^2}{n_2 - K}$$

حيث أن: مجموع مربعات البواقي للعينة الجزئية الأولى

$ESS_1 = \sum e_1^2 = \sum y_1^2 - \sum \hat{y}_1^2$

$ESS_2 = \sum e_2^2 = \sum y_2^2 - \sum \hat{y}_2^2$

" " " الثانية

n_1, n_2 : حجم العينة الجزئية الأولى والثانية على التوالي

K : عدد معلمات النموذج المقدر

5- ثم احتساب إحصاءة F وفقاً للصيغة الآتية

$$F = \frac{\sigma_{e_1}^2}{\sigma_{e_2}^2}$$

ثم تقارن قيمة F المحسوبة مع قيمة F الجدولية عند مستوى معنوية معين (1% أو 5% أو 10%) ودرجات حرية قدرها (n_1-k, n_2-k) للبسط والمقام

فإذا كانت قيمة F المحسوبة أكبر من الجدولية

F المحسوبة $>$ F الجدولية

نرفض فرض العدم (H_0) ونقبل بالفرض البديل (H_1) والذي ينص على وجود مشكلة عدم ثبات تباين حد الخطأ العشوائي

$$H_1: \sigma_1^2 \neq \sigma_2^2 \neq \cdots \neq \sigma_n^2$$

وبالعكس إذا كانت قيمة F المحسوبة أصغر من قيمة F الجدولية

F المحسوبة $<$ F الجدولية

نقبل بفرض العدم (H_0) ونرفض الفرض البديل (H_1) مما يعني عدم وجود المشكلة أي ثبات تباين الخطأ

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2$$

٢- اختبار معامل ارتباط الرتب لسبيرمان

يعد هذا الاختبار من ابسط واسهل الاختبارات المستخدمة في اكتشاف مشكلة عدم تجانس التباين، ويمكن تطبيقه على العينات الكبيرة والصغيرة، ويعتمد هذا الاختبار على القيم المطلقة للأخطاء (e_i) (اي اهمال الاشارة) وقيم المتغير المستقل (X_i) موضوع الدراسة ويمكن تلخيص الخطوات الحسابية لهذا المعامل بالآتي:

1- تقدير النموذج باستخدام (OLS)

2- احتساب قيم البواقي (e_i) ، حيث ان: $e_i = Y_i - \hat{Y}_i$

3- ترتيب قيم المتغير المستقل (X_i) والانحرافات المطلقة (e_i) ترتيباً تصاعدياً او تنازلياً واعطاء كل منها رتباً معينة (مثل سلسلة اعداد طبيعية ($1, 2, 3 ---- n$)) وفق تسلسل القيم ثم تحسب فروقات الرتب وتربيعها، ومن ثم يستخرج معامل ارتباط الرتب ما بين القيم المطلقة للانحرافات (e_i) وقيم المتغير المستقل المعني (X_i) وفق قانون سبيرمان لارتباط الرتب الآتي:

$$r_s = 1 - \frac{6\sum D_i^2}{n(n^2-1)}$$

حيث ان:

D_i: يمثل الفرق بين رتب القيم المطلقة للأخطاء (e_i) ورتب المتغير المستقل (X_i) القيم المناظرة (اي الفرق بين رتب الازواج المتناظرة لكل من (e_i, X_i)) و n تمثل عدد المشاهدات.

n : حجم العينة.

وكلما اقتربت قيمة r_s من الواحد الصحيح دل ذلك على وجود علاقة قوية بين e_i و X_i ومن ثم وجود مشكلة عدم تجانس التباين.

ويمكن التأكد من وجود المشكلة من خلال اختبار (t) مع وفق الصيغة الآتية:

$$t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}}$$

وتقارن قيمة (t) المحسوبة مع نظيرتها الجدولية (t^*) عند درجة حرية ($n-k$) ومستوى معنوية 1% او 5% فاذا كانت

84

t المحسوبة $>$ (أكبر) t الجدولية

نرفض فرضية العدم ونقبل الفرض البديل (H_1) مما يشير على وجود مشكلة عدم تجانس التباين حالات

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2 = \cdots \neq \sigma_n^2$$

اما اذا كانت t المحسوبة $<$ (أقل) t الجدولية

نقبل فرضية العدم ونرفض الفرض البديل مما يعني عدم وجود مشكلة عدم تجانس التباين (تجانس تباين الخطأ).

③ - اختبار بارك - كليجسر Park - Glejser

من الاختبارات الاخرى المستخدمة في الكشف عن عدم تجانس التباين هو اختبار بارك - كليجسر ويعتمد هذا الاختبار على شكل العلاقة بين الاخطاء العشوائية (e_i) والمتغير المستقل الذي نعتقد انه سبب الاخطاء

① يشمل الخطوة الاولى في هذا الاختبار الاتي
تقدير معالم العلاقة الخطية وايجاد قيم الاخطاء العشوائية الناتجة من الفرق بين القيم الحقيقية والتقديرية للمتغير المعتمد.

② اما الخطوة الثانية فيتم فيها تقدير صيغ مختلفة وحسب نوع العلاقة التي نعتقد بانها موجودة ما بين القيم المختلفة للاخطاء العشوائية وقيم المتغير المستقل ومن امثلة على هذه الصيغ المختلفة الاتي

$$|e_i| = a + bX_i$$

$$|e_i| = a + b_1\sqrt{X_i}$$

$$|e_i| = a + \frac{b_1}{X_i}$$

③ - الخطوة الاخيرة من هذا الاختبار تتمثل في اجراء اختبار معنوية العلاقة المقدرة حيث يستدل منها فيما اذا كانت هناك مشكلة عدم تجانس التباين او لا وتحديد الصيغة العلاقة بين الاخطاء العشوائية والمتغير المستقل شكلها
تجدر الاشارة هنا الى ان هذا الاختبار له اهمية في تحديد الصيغة الملائمة من التجربة.

٢-٢) اختبار بارتليت Bartlett

وتتلخص الفكرة الأساسية لهذا الاختبار بتجزئة العينة تحت البحث الى (m) من العينات الجزئية، ومن ثم احتساب تباين الخطأ لكل عينة جزئية (σ_i^2) بدرجة حرية (n_i-1)، ومن ثم البحث عن احتمال انتساب هذه العينات الجزئية من مجتمع معين فاذا قبلت فرضية العدم التالية

$$H_0: \sigma_1^2=\sigma_2^2=\dots=\sigma_n^2$$

$$H_0: \sigma_i^2=\sigma^2$$

أو

دل ذلك على ان العينات الجزئية مسحوبة من مجتمع متجانس

مقابل الفرضية البديلة

$$H_1: \sigma_1^2\neq\sigma_2^2\neq\dots\neq\sigma_n^2$$

$$H_1: \sigma_i^2\neq\sigma^2$$

أو

اي ان تباين الخطأ المحتسب من العينات الجزئية غير متجانس

ويجدر الاشارة هنا الى ان هذا النوع من الاختبار غالباً ما يطبق على العينات التي تتوفر فيها تكرارات عديدة لكل قيمة من قيم المتغير المستقل، لذا فان هذا الاختبار يستوجب تقسيم المتغير المستقل الى عدة مستويات ولنفرض ان هناك (n_i) من المشاهدات مقابل كل مستوى حيث ان ($i=1,2,\dots,m$)، وهنا يكون المجموع الكلي للمشاهدات

$$\sum_{i=1}^{m} n_i = N$$

اضافة الى ذلك، تفترض ان المتغير المعتمد يرتبط بالمتغير المستقل بواسطة الصيغة التالية

$$Y_{ij} = a + bX_{ij} + U_{ij}$$ حيث

$$j=1,2,\dots,n_i$$