Econometria II
Part 3: Heteroscedastisitat

Riste Gjorgjiev
pareto.uab.es/∼rgjorgjiev/econ_cat

Lorenzo Burlon
idea.uab.es/lburlon/teaching_econometria2.html

Universitat Autònoma de Barcelona

Heteroscedastisitat

L'objectiu d'aquest tema és poder contestar les següents preguntes

- Què és l'Heteroscedastisitat
- Quins són els seus efectes al MQO
- Com la podem detectar
- Què podem fer per evitar els efectes negatius

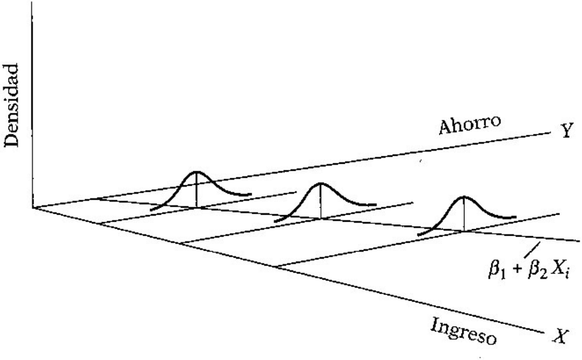
El supòsit 4 del MLG és:

- La variància de les pertorbacions és constant, $var(u_i) = \sigma^2$

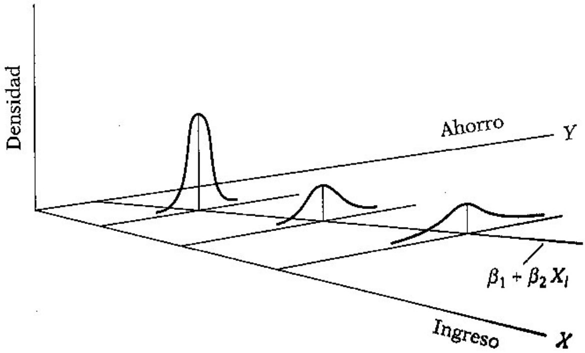
Què implica aquest supòsit sobre la variancia de Y?

- $Var(Y_i) = Var(u_i) =$const

Exemple: Regressió de l'estalvi Y sobre l'ingrés X, $(Y = \beta_1 + \beta_2 X + u)$



Quina diferencia podem notar en el gràfic:



- Quina és la relació entre l'ingrés i l'estalvi en els dos casos? - positiva
- Es canvia la variància del Y?
- Què vol dir aquest canvi? - les families amb ingressos més alts estalvien amb més variabilitat

En el exemple anterior podem notar què la variància de les perturbacions no és constant o les perturbacions són heteroscedastiques.
Formal, tenim la definició

Definició

El model $Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$ té les pertorbacions heteroscedastiques si

$$var(u_i) = \sigma_i^2 = \sigma^2 \omega_i, \; i \; \exists i \neq j, \; \sigma_i \neq \sigma_j.$$

Indicació per presència d'heteroscedastisitat

Utilitzant les dades de l'arxiu *dades1.xls* sobre

- la despesa en alimentació - Y
- la despesa total - X_2
- el número d'adults - X_3
- el número de nens y - X_4
- l'edat del marit - X_5

d'una família, fem una regressió de la despesa en alimentació sobre les altres variables

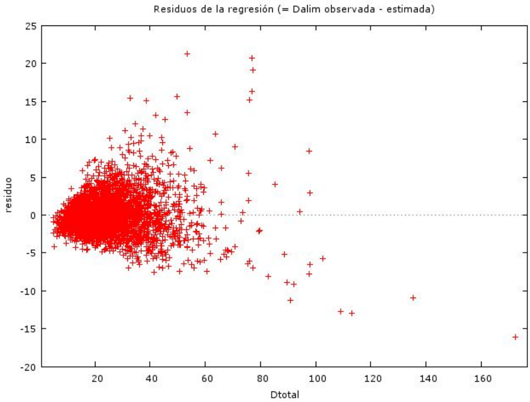
$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u.$$

Els resultats són

$$\widehat{\text{Dalim}} = \underset{(0.28168)}{0.612480} + \underset{(0.0039081)}{0.132891}\,\text{Dtotal} + \underset{(0.052484)}{0.554692}\,\text{Adults} + \underset{(0.047104)}{0.588184}\,\text{Nens}$$
$$+ \underset{(0.0058684)}{0.0204881}\,\text{Edat}$$
$$T = 3371 \quad \bar{R}^2 = 0.3572 \quad F(4, 3366) = 469.26 \quad \hat{\sigma} = 2.6616$$

Per poder fer comentari sobre la variància del les pertorbacions mirem al gràfic dels residus (les pertorbacions estimades) sobre Dtotal.
En Gretl

- gràfics
- gràfics dels residus
- contra la variable desitjada (en aquest cas Dtotal)



Què podem dir d'aquest gràfic

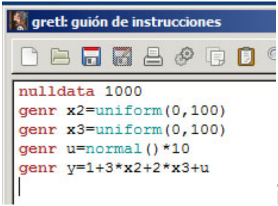
- sembla que la variància dels residus augmenta amb Dtotal
- concloem què els pertorbacions són heteroscedastiques - No!
- hi ha indicació de pertorbacions heteroscedastiques - Si

Unes raons per presència d'heteroscedastisitat

- aprenentatge durant del temps, un experiment en laboratori (número d'errors fets Y sobre l'hores de practica X)
- ingres més alt $\Rightarrow$ més posibilitats per estalvar, l'exemple anterior
- millorar a les tecniques de recollir les dades provoca disminució dels errors
- errors d'espificació del model - exclucióo d'una variable significativa pot augmentar la variació de la variable dependent (un exemple amb gretl)

Excluciío d'una variable significativa del MQO

Generem un conjunt de dades



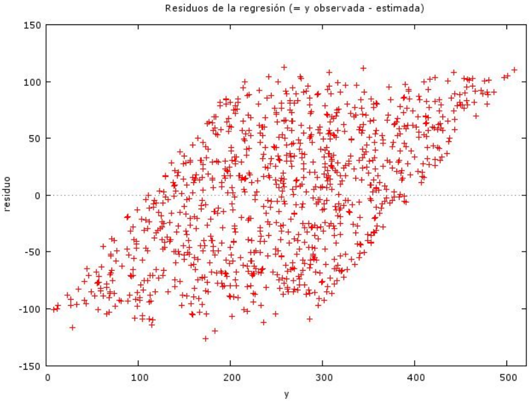
i fem una estimació amb X_2 com l'única variable explicativa

$$\widehat{y} = \underset{(3.7172)}{107.140} + \underset{(0.063467)}{2.91456}\, x2$$

$$T = 1000 \quad \bar{R}^2 = 0.6785 \quad F(1,998) = 2108.8 \quad \hat{\sigma} = 57.998$$

(Desviaciones tpicas entre parntesis)

Mirem al gràfic dels residus sobre la Y



- es canvia variància de u amb Y?
- hi ha indicació d'heteroscedastisitat?

Propietats del MQO sota heteroscedastisitat

Si $var(u_i) = \sigma^2$, sabem què

- $E(\hat{\beta}_{MQO}) = \beta$
- $var(\hat{\beta}_{MQO}) = \sigma^2(X'X)^{-1}$

però en el cas d'heteroscedastisitat

- $E(\hat{\beta}_{MQO}) = \beta$
- $var(\hat{\beta}_{MQO}) = (X'X)^{-1}\left(\sigma^2 \sum_{i=1}^{T} \omega_i X_i X_i'\right)(X'X)^{-1}$

Aleshores, en un model amb pertorbacions heteroscedastiques

- $var(\hat{\beta}_{MQO})$ no és pot estimar amb $\hat{\sigma}^2(X'X)^{-1}$
- $\hat{\beta}_{MQO}$ no és el millor estimador lineal sense biaix
- els estadístics t què surten en Gretl no són valides

Mínims Quadrats Generalitzats (MQG)

Abans de fer una regressió a un model heteroscedastic,

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u, \; X_1 = \text{const}$$

on $var(u_i) = \sigma_i^2 = \sigma^2 \omega_i$, fem la següent transformació

$$Y_i^* = Y_i/\sigma_i, \; X_{ji}^* = X_{ji}/\sigma_i, \; u_i^* = u_i/\sigma_i$$

Ara tenim el model transformat

$$Y^* = \beta_1^* X_1^* + \beta_2^* X_2^* + \ldots + \beta_k^* X_k^* + u^*$$

Notem què $\beta_i^* = \beta_i$

Quins supòsits compleix el model transformat

1. Les variables independents són fixes, $E(X^*) = X^*$
2. Les variables independents no estan correlacionades, $(X^{*\prime}X^*)^{-1}$ existeix
3. El valor mitjà de les pertorbacions és zero, $E(u_i^*) = 0, \forall i$
4. No hi ha correlació entre les pertorbacions, $cov(u_i^*, u_s^*) = 0$
5. La variància de les pertorbacions és constant, $var(u_i^*) = 1$
6. Les pertorbacions segueixen la distribució normal, $u_i^* \sim N(0, 1)$

Doncs,

- el model transformat compleix tots els supòsits del MLG
- podem aplicar el MQO al model transformat

Definició

El mètode d'estimació quan s'aplica MQO a les variables transformades es diu Mínims Quadrats Generalitzats (MQG), $\hat{\beta}_{MQG}$

Què fa el MQG? - Busca els valors de β^*, què minimitzan la suma quadratica residual

$$\sum \hat{u}_i^{*\,2} = \sum (Y_i^* - X_i^{*\prime} \beta)^2, \tag{1}$$

què també es pot escriure com

$$\sum \left(\frac{\hat{u}_i}{\sigma_i} \right)^2 = \sum (\frac{Y_i}{\sigma_i} - \frac{X_i'}{\sigma_i} \beta)^2 = \sum \frac{1}{\sigma_i} (Y_i - X_i' \beta)^2. \tag{2}$$

Conclusió:

- Els valors dels paràmetres β què minimitzan (1), també minimitzan (2)
- l'estimador $\hat{\beta}_{MQG}$ és el millor estimador lineal sense biaix de β

Què passaria amb un model heteroscedastic ($\text{var}(u_i) = \sigma_i^2 = \sigma^2 \omega_i$) si féssim la transformació

$$Y_i^* = Y_i / \sqrt{\omega_i},\ X_{ji}^* = X_{ji} / \sqrt{\omega_i},\ u_i^* = u_i / \sqrt{\omega_i}$$

En aquest cas, la variància de les pertorbecions transformats serà σ^2 (model homoscedastic)

Per fer utilitzar el MQG, necessitem saber el σ_i^2. En pràctica, és molt sovint que les variàncies de u_i siguin desconegudes.
En aquesIls casos, hem de trobar manera d'estimar-les

$$\hat{\sigma}_i^2 \text{ és el valor estimat de } \sigma_i^2$$

Després fem la transformació

$$Y_i^* = Y_i/\hat{\sigma}_i,\ X_{ji}^* = X_{ji}/\hat{\sigma}_i,\ u_i^* = u_i/\hat{\sigma}_i$$

$$Y^* = \beta_1^* X_1^* + \beta_2^* X_2^* + \ldots + \beta_k^* X_k^* + u^*$$

Definició

L'estimador dels parametres de l'última equació amb MQO es diu Mínims Quadrats Generalitzats Factibles (MQGF), $\hat{\beta}_{MQGF}$

Exercici 1

El model amb una variable explicativa

$$Y_i = \beta X_i + u_i,\ i = 1, \ldots, 60$$

cumpleix tots els supòsits del MLG excepte l'homoscedastisitat.
Es sap què

$$var(u_i) = 2,\ i = 1, \ldots, 20 \text{ i}$$
$$var(u_i) = \sigma^2,\ i = 21, \ldots, 60$$

on σ^2 no és conegut.

Escriure les expressions de $\hat{\beta}_{MQG}$ i $\hat{\beta}_{MQGF}$
Pel segon, trobar una manera d'estimar σ^2

Exercici 2

El model amb una variable explicativa

$$Y_i = \beta X_i + u_i, \; i = 1, \ldots, 60$$

cumpleix tots els supòsits del MLG excepte l'homoscedastisitat $(var(u_i) = \sigma_i^2)$.

- trobar la variància de l'estimador MQO, $var(\beta_{MQO})$
    1. trobar una manera d'estimar-la, $\hat{var}(\beta_{MQO}) =?$
- trobar la variància de l'estimador MQG, $var(\beta_{MQG})$

Definició

La variància estimada en el punt 1 es la diu variància de White

$$\hat{var}(\hat{\beta}_{MQO}) = (X'X)^{-1} \left(\sum_{i=1}^{T} \hat{u}_i^2 X_i X_i' \right) (X'X)^{-1}$$

Supòsits sobre la forma de σ_i^2

- per poder utilitzar l'estimador de MQG, hem dit què necessitem trobar manera d'estimar σ_i^2
- per això, és necessari un supòsit addicional sobre la seva forma
- aquí presentarem uns supòsit més tipics
- i mirarem com serà l'estimador MQG
- en tots els casos, treballarem amb un model amb una constanta i una variable explicativa

Supòsit 1

En el model

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{3}$$

suposem què

$$\sigma_i^2 = \sigma^2 X_i^2$$

Aquí hem de dividir totes les variables amb X_i o transformar l'equació (3)

$$\frac{Y_i}{X_i} = \frac{\beta_1}{X_i} + \beta_2 + \frac{u_i}{X_i}, \left(Y_i^* = \beta_2 + \beta_1 \frac{1}{X_i} + u_i^*\right)$$

$\hat{\beta}_{MQG}$ serà $\hat{\beta}_{MQO}$ a Y^* sobre *const* i $1/X$

Atenció: Hi ha canvi al ordre dels coeficients β

Supòsit 2

En el model

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{4}$$

suposem què

$$\sigma_i^2 = \sigma^2 X_i$$

Aquí hem de dividir totes les variables amb $\sqrt{X_i}$ o transformar l'equació (4)

$$\frac{Y_i}{\sqrt{X_i}} = \frac{\beta_1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + \frac{u_i}{\sqrt{X_i}} \quad \left(Y_i^* = \beta_1\frac{1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + u_i^*\right)$$

$\hat{\beta}_{MQG}$ serà $\hat{\beta}_{MQO}$ a Y^* sobre $1/\sqrt{X}$ i $\sqrt{X}$

Atenció: No hi ha constanta

Supòsit 3

En el model

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{5}$$

suposem què $\sigma_i^2 = \sigma^2 [E(Y_i)]^2$
Aquí hem de dividir totes les variables amb $E(Y_i)$ o transformar l'equació (5)

$$\frac{Y_i}{E(Y_i)} = \frac{\beta_1}{E(Y_i)} + \beta_2 \frac{X_i}{E(Y_i)} + \frac{u_i}{E(Y_i)}$$

Nosaltres no sabem el valor de $E(Y_i) = \beta_1 + \beta_2 X_i$, però podem utilitzar el seu valor estimat $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$. Llavors l'equació per estimar és

$$\frac{Y_i}{\hat{Y}_i} = \frac{\beta_1}{\hat{Y}_i} + \beta_2 \frac{X_i}{\hat{Y}_i} + \frac{u_i}{\hat{Y}_i} \; (Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + u_i^*)$$

$\hat{\beta}_{MQGF}$ serà $\hat{\beta}_{MQO}$ a Y^* sobre X_1^* i X_2^*

Prova de White

Model amb k variables explicatives

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

La prova de White s'utilitza per poder decidir de rebutjar o no, l'hipòtesi seguënt:

$$H_0 : \sigma_t^2 = \sigma^2$$
$$H_1 : \sigma_t^2 \text{ és una funció de les variables independents}$$

L'algoritme constitueix dels següents pasos:

1. fer la regressió $Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$
2. guardar els residus estimats $\hat{u}$
3. fer la regressio auxiliar

$$\hat{u}^2 = \alpha_1 + \sum_{i=2}^{k} \alpha_i X_i + \sum_{i=2}^{k} \sum_{j=i}^{k} \alpha_{ij} X_i X_j + \nu$$

4. guardar el R^2 de l'última regressió

$$\text{sota } H_0 \;\; W = TR^2 \sim \chi^2_l, \; l = \frac{(k+1)k}{2} - 1$$

5. si TR^2 és major què el valor de χ^2_l crític, rebutjem l'H_0 (hi ha heteroscedastisitat)

Avantatges:

- no cal especificar la relacó entre u^2 i les variables
- no necessita el supòsit de pertorbacions normals

Desavantatges:

- no podem fer la transformació per utilitzar MQG

Exemple

Fem la prova de White per l'exemple d'abans. Tenim les dades *dades1.xls* sobre

- la despesa en alimentació - Y
- la despesa total - X_2
- el número d'adults - X_3
- el número de nens y - X_4
- l'edat del marit - X_5

1. fem una regressió de la despesa en alimentació sobre les altres variables
$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u.$$
2. guardem els residus quadrats (Guardar->residuos al cuadrado), $\hat{u}^2$
3. generem les variables necessaris per la regressió auxiliar[29] (quantes hi ha?)
4. fem la regressió
$$\hat{u}^2 = \alpha_1 + \sum_{i=2}^{5} \alpha_i X_i + \sum_{i=2}^{5} \sum_{j=i}^{5} \alpha_{ij} X_i X_j + \nu$$

Modelo 2: MCO, usando las observaciones 1–3371
Variable dependiente: usq1

	Coeficiente	Desv. Tpica	Estadstico t	Valor p
const	−12.4597	8.92600	−1.3959	0.1628
Dtotal	0.141710	0.185590	0.7636	0.4452
Adults	−1.34758	2.97947	−0.4523	0.6511
Nens	−1.78054	2.28070	−0.7807	0.4350
Edat	0.710747	0.370613	1.9178	0.0552
Dtotal2	0.00998597	0.000742135	13.4557	0.0000
Adults2	0.360182	0.227870	1.5806	0.1141
Nens2	0.334575	0.216884	1.5426	0.1230
Edat2	−0.00704638	0.00436625	−1.6138	0.1067
DtotAdul	−0.0751237	0.0293681	−2.5580	0.0106
DtotNens	0.0531168	0.0297663	1.7845	0.0744
DtotEdat	−0.00190837	0.00382035	−0.4995	0.6174
AdulNens	0.210808	0.404052	0.5217	0.6019
AdulEdat	0.00107992	0.0514184	0.0210	0.9832
NensEdat	−0.0359792	0.0471233	−0.7635	0.4452

Media de la vble. dep.	7.073432	D.T. de la vble. dep.	20.80356
Suma de cuad. residuos	1150583	D.T. de la regresin	18.51604
R^2	0.211117	R^2 corregido	0.207826

$W = TR^2 = 3371 * 0,211117 = 711.675 > \chi^2_{14}$ per qualsevol nivell de significació

$\Rightarrow$Rebutjem l'H_0

Una altra manera per calcular W

Després d'estimar el model inicial, utilitzant la pantalla de l'estimació en Gretl fem: contrastes $->$ heteroscedastisitat $->$ contraste de White

Tabla Contraste de heterocedasticidad de White
MCO, usando las observaciones 1-3371
Variable dependiente: $uhat^2$

	Coeficiente	Desv. Tpica	Estadstico t	Valor p
const	-12.4597	8.92600	-1.396	0.1628
Gtotal	0.141710	0.185590	0.7636	0.4452
Hijos	-1.78054	2.28070	-0.7807	0.4350
Adultos	-1.34758	2.97947	-0.4523	0.6511
Hedad	0.710747	0.370613	1.918	0.0552
sq-Gtotal	0.00998597	0.000742135	13.46	3.09e-040
X2-X3	0.0531168	0.0297663	1.784	0.0744
X2-X4	-0.0751237	0.0293681	-2.558	0.0106
X2-X5	-0.00190837	0.00382035	-0.4995	0.6174
sq-Hijos	0.334575	0.216884	1.543	0.1230
X3-X4	0.210808	0.404052	0.5217	0.6019
X3-X5	-0.0359792	0.0471233	-0.7635	0.4452
sq-Adultos	0.360182	0.227870	1.581	0.1141
X4-X5	0.00107992	0.0514184	0.02100	0.9832
sq-Hedad	-0.00704638	0.00436625	-1.614	0.1067

R-cuadrado = 0.211117

Estadístico de contraste: $TR^2 = 711.674784$,

con valor p = P(Chi-cuadrado(14) ¿ 711.674784) = 0.0000

El codi de l'exemple

```
genr Dtotal2=Dtotal*Dtotal
genr Adults2=Adults*Adults
genr Nens2=Nens*Nens
genr Edat2=Edat*Edat
genr DtotAdul=Dtotal*Adults
genr DtotNens=Dtotal*Nens
genr DtotEdat=Dtotal*Edat
genr AdulNens=Adults*Nens
genr AdulEdat=Adults*Edat
genr NensEdat=Nens*Edat
```

Prova de Goldfeld-Quandt

Model amb k variables explicatives

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

Suposem que la variància de les pertorbacions és

$$\sigma_t^2 = \sigma^2 X_{jt}^2$$

La prova de Goldfeld-Quandt s'utilitza per poder decidir de rebutjar o no, l'hipòtesi seguënt:

$$H_0 : \text{model homoscedastic}$$

L'algoritme constitueix dels següents pasos:

1. ordenar totes les observacions segons la variable X_j
2. omet c observacions centrals i parteix la mostra en dues parts amb grandàries T_I i T_{II}
   - c pot ser zero
3. fer dues regressions, utilitzant les mostres partides
4. guardar les sumes quadratiques residuals de les dues regressions, e_I i e_{II}
5. sota el supòsit de normalitat de les pertorbacions i sota l'hipòtesi zero
   - $GQ = \frac{e_{II}/(T_{II}-k)}{e_I/(T_I-k)} \sim F_{T_{II}-k,T_I-k}$
6. si el valor de GQ és major que el valor crític de $F_{T_{II}-k,T_I-k}$ rebutjem l'hipòtesi zero

En el cas de rebutjar, podem concloure que no hi ha heteroscedastisitat?

Exemple

Fem la prova de Goldfeld-Quandt pel model què explica el consum utilitzant l'ingrés

$$Y = \beta_1 + \beta_2 X + u, \ Y\text{-consum}, \ X\text{-ingrés}$$

L'arxiu *GQ.xls* conté les dades de les dues variables. Les dades en les coulumnes RY i RX estan ordernades per X.
Traiem les 4 observacions del mig i fem les dues regressions amb les primers i les últims 13 observacions

Per restringir la mostra, al Gretl escollim

- muestra$-$ $>$establecer rango

Modelo 1: MCO, usando las observaciones 1–13
Variable dependiente: RY

	Coeficiente	Desv. Tpica	Estadstico t	Valor p
const	3.40943	8.70492	0.3917	0.7028
RX	0.696774	0.0743660	9.3695	0.0000

Suma de cuad. residuos	377.1663	D.T. de la regresin	5.855582

Modelo 2: MCO, usando las observaciones 18–30 ($n = 13$)
Variable dependiente: RY

	Coeficiente	Desv. Tpica	Estadstico t	Valor p
const	−28.0272	30.6421	−0.9147	0.3800
RX	0.794137	0.131582	6.0353	0.0001

Suma de cuad. residuos	1536.800	D.T. de la regresin	11.81986

- $GQ = \frac{1536/11}{377/11} = 4.07$
- el valor crític de $F_{11,11}$ al nivell de 5% és 2.82
- $GQ > 2.82$
- $\Rightarrow$ rebutjem la hipòtesi de homoscedasticitat

Prova de Breusch-Pagan-Godfrey

Model amb k variables explicatives

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

Suposem que la variància de les pertorbacions és

$$\sigma_t^2 = f(\alpha_1 + \alpha_2 Z_2 + \ldots + \alpha_m Z_m),$$

on les variables Z_i són variables no estocàstiques.

La prova de Breusch-Pagan-Godfrey s'utilitza per poder decidir de rebutjar o no, l'hipòtesi següënt:

$$H_0 : \alpha_2 = \ldots = \alpha_m = 0$$

L'algoritme constitueix dels següents pasos:

1. fer MQO al model
$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$
2. guardar els residus $\hat{u}_i$ (o els quadrats dels residus $\hat{u}_i^2$)
3. obtenir la constanta $\tilde{\sigma}^2 = \sum \hat{u}_i^2 / T$
4. definir la nova variable $p = \hat{u}^2 / \tilde{\sigma}^2$
5. regressar p sobre const, $Z_2, \ldots Z_m$
6. obtenir la suma quadràtica explicada, SQE
7. sota el supòsit de normalitat dels pertorbacions i sota l'hipòtesi zero
$$\Theta = \frac{1}{2} SQE \sim \chi^2_{m-1}$$
8. si el valor de Θ és major que el valor crític de χ^2_{m-1}, rebutjem l'H_0

Exemple

Fem la prova de Breusch-Pagan-Godfrey pel mateix exemple d'abans

1. MQO de Y sobre const, X
2. $\hat{u}^2$ - residus
3. $\tilde{\sigma}^2 = 2361/30 = 78$
4. definir $p = \hat{u}^2/78$
5. regressar p sobre const, X
6. calcular $SQE = \frac{R^2}{1-R^2} SQR = 10.42$
7. $\Theta = SQE/2 = 5.21$

El valor crític de χ^2_1 de 5% és 3.8141, $\Rightarrow$ rebutjem l'H_0

Prova de Park

Sota el supòsit

$$\sigma_i^2 = \sigma^2 X_{mi}^{\beta} e^{\nu_i}$$

La prova de Park construeix un estadístic per rebutjar o no la següent hipòtesi

$$H_0 : \beta = 0$$

Per poder construir-lo, utilitzem transformació logarítmica del supòsit anterior

$$\log \sigma_i^2 = \log \sigma^2 + \beta \log X_{mi} + \nu_i$$

Per construir l'estadístic fem els pasos:

1. fer la regressió de Y sobre les variables del model
2. guardar els residus quadrats, $\hat{u}^2$
3. aplicar una transformació logarítmica a $\hat{u}^2$ i X
4. regressar $\log \hat{u}^2$ sobre const i $\log X$
5. l'estadístic de Park és el valor t del β
6. si el valor de t és major que el valor crític de t_{T-k}, rebutjem l'H_0

Exemple

Utilitzem l'exemple amb les dades sobre la mortalitat infantil (MI), PNB per càpita (PNBC) i la taxa d'analfabetisme de les dones (TAD).
Volem investigar l'influència de PNBC i TAD a la MI, utilitzant l'equació

$$MI = \beta_1 + \beta_2 PNBC + \beta_3 TAD + u_i$$

Com que diferents països tenen diferents experiències amb la mortalitat, es sospita heteroscedastisitat. Fem la prova de Park

1. regressió de MI sobre const i $PNBC$ I TAD
2. guardar els residus quadrats, $\hat{u}^2$
3. transformar les variables $\hat{u}^2$ i TAD (o l'altra)
4. els resultats de la regressió són

$$\widehat{\log\text{RES}} = \underset{(4.126)}{7.57744} - \underset{(-0.915)}{0.439433}\,\log\text{TAD}$$

$$T = 64 \quad \bar{R}^2 = -0.0026 \quad F(1,62) = 0.83657 \quad \hat{\sigma} = 2.3811$$

(entre parntesis, los estadsticos t)

- l'estadístic t és inferior al valor crític de 5% de la distribució $t_{62} = 2$
- no rebutjem l'H_0, o no detectem heteroscedastisitat

Proves d'heteroscedasticitat - resum

	White	Goldfeld-Quandt	Breush-P-G
supòsits	cap	$\sigma_t^2 = \sigma^2 X_{jt}^2$	$\sigma_t^2 = f(\alpha_1 + \alpha_2 Z_2 + \ldots + \alpha_m Z_m)$
necessita	mostra gran	$u \sim N$, ordenar l'X_j, c	$u \sim N$
per heter. global	si	no	no sempre
info per fer MQG	no	si	si

	Park
supòsits	$\sigma_t^2 = \sigma^2 X_{jt}^{\beta} e^{\nu_t}$
necessita	$\nu_i \sim N(0, const)$
per heter. global	no
info per fer MQG	si

Contrastes sota heteroscedasticitat 1

Al model amb k variables explicatives

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

es sap que $\sigma_t^2 = \sigma^2 X_{2t}^{\alpha}$, on $\sigma_t^2 = var(u_i)$ i α no és coneguda.

Com podem fer el següent contraste

$$H_0 : \beta_2 = \beta_0?$$

MQG és aplicable?

- sota tots els supòsits del model lineal general, utilitzàvem l'estadístic

$$t^* = \frac{\hat{\beta}_2 - \beta_0}{ee(\hat{\beta}_2)}$$

- t^* ara no és aplicable. Per què?
- per poder contestar a l'H_0 hem de utilitzar el MQG o MQGF
- dividir totes les variables amb $\sqrt{X_{2t}^{\alpha}}$ o $(X_{2t}^{\alpha/2})$
- és MQG aplicable?
- com podem estimar l'α
  1. fer transformació log al supòsit $\sigma_t^2 = \sigma^2 X_{2t}^{\alpha}$
  2. $\sigma_t^2 = \sigma^2 X_{2t}^{\alpha}$
  3. regressar $\log \hat{u}^2$ sobre const i $\log X_2$ - ($\hat{u}^2$ són els residus del model inicial)
  4. el coeficient de $\log X_2$ representa el valor estimat d'α

- ara tenim el valor $\hat{\alpha}$
- i podem transformar les variables del model

$$X_{ji}^* = \frac{X_{ji}}{X_{2i}^{\alpha^*/2}},\ Y_i^* = \frac{Y_i}{X_{2i}^{\alpha^*/2}},\ u_i^* = \frac{u_i}{X_{2i}^{\alpha^*/2}}$$

- Ara el model és

$$Y^* = \beta_1 \frac{X_1}{X_{2i}^{\alpha^*/2}} + \beta_2 \frac{X_2}{X_{2i}^{\alpha^*/2}} + \ldots + \beta_k X_k^* + u^*$$

- l'últim model cumpleix totes els supòsits de MLG
- l'estadístic t serà

$$t = \frac{\hat{\beta}_2^* - \beta_0}{ee(\hat{\beta}_2^*)}$$

- la conclució de rebutjar o no depèn de t i el valor crític de la distribució t_{T-k}
- què pasaria si el valor estimat d'α fos 2?

Contrastes sota heteroscedasticitat 2

Al model amb k variables explicatives

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

- es sap que $\sigma_t^2 = \gamma_1 + \gamma_2 X_{2t} + \gamma_3 X_{3t}$, on
- $\sigma_t^2 = var(u_i)$
- $\gamma_1, \gamma_2, \gamma_3$ no són conegudes

Com podem fer el següent contraste

$$H_0 : \beta_2 = \beta_0?$$

MQG és aplicable?

- al model inicial les pertorbecions són heteroscedastiques
- la variància σ_t^2 depend d'uns paràmetres desconegudes
- $\Rightarrow$ MQG no és aplicable
- doncs, hem de trobar una manera d'estimar els valors de γ_i, $i = 1, 2, 3$
- obtenir $\hat{\sigma}_t^2$ segons

$$\hat{\sigma}_t^2 = \hat{\gamma}_1 + \hat{\gamma}_2 X_{2t} + \hat{\gamma}_3 X_{3t}$$

- transformar totes les variables del model inicial

$$X_{ji}^* = \frac{X_{ji}}{\hat{\sigma}_i},\ Y_i^* = \frac{Y_i}{\hat{\sigma}_i},\ u_i^* = \frac{u_i}{\hat{\sigma}_i}$$

- fer MQGF (o MQO al model)

$$Y^* = \beta_1 X_1^* + \beta_2 X_2^* + \ldots + \beta_k X_k^* + u^*$$

- calcular l'estadístic $t = \frac{\hat{\beta}_2 - \beta_0}{ee(\hat{\beta}_2^*)}$
- rebutjar si $t > t_{T-k}$

- ara l'únic problema és com obtenir $\hat{\sigma}_t^2$
- si sabem que $\sigma_t^2 = \gamma_1 + \gamma_2 X_{2t} + \gamma_3 X_{3t}$,
- podem regressar els residus quadrats del model inicial sobre una *cons*, X_2 i X_3

$$\hat{u}^2 = \gamma_1 + \gamma_2 X_2 + \gamma_3 X_3 + \nu$$

- els valors estimats dels paràmetres de la última regression, són els valors $\hat{\gamma}_1$, $\hat{\gamma}_2$ i $\hat{\gamma}_3$ que necessitem
- l'últim pas serà calcular $\hat{\sigma}_t^2$, segons

$$\hat{\sigma}_t^2 = \hat{\gamma}_1 + \hat{\gamma}_2 X_{2t} + \hat{\gamma}_3 X_{3t}$$

Contrastes sota heteroscedasticitat 3

Tenim un model amb k variables explicatives

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

Imaginem que la prova de White detecta heteroscedasticitat. Com podem fer el següent contraste

$$H_0 : \beta_2 = \beta_0?$$

- en aquest cas $t^* = \frac{\hat{\beta}_2 - \beta_0}{ee(\hat{\beta}_2)}$ tampoc és aplicable
- aquesta prova no ens dona informació per poder fer el MQG
- però hi ha manera per poder constestar a la pregunata a sobre
- recordem a la variància estimada de White
$$\hat{var}_w(\hat{\beta}_{MQO}) = (X'X)^{-1} \left(\sum_{i=1}^{T} \hat{u}_i^2 X_i X_i' \right) (X'X)^{-1}$$

- $\hat{var}_w(\hat{\beta})^\star$ és la variància estimada del $\hat{\beta}_{MQO}$ sota heteroscedasticitat
- $\hat{var}_w(\hat{\beta})^{\star\star}$ és una matriu que el seu valor a segona fila i a la segona columna $\hat{var}_w(\hat{\beta})_{2,2}$, representa la variància estimada del $\hat{\beta}_2$
- doncs,

$$t = \frac{\hat{\beta}_2 - \beta_0}{ee_w(\hat{\beta}_2)}$$

  serà l'estadístic que utilitzarem per decidir sobre l'H_0
- si $t > t_{T-k} \Rightarrow$ rebutjem

$\star$ ometo la notació MQO

$\star\star$ no confonguem $\hat{var}_w(\hat{\beta})$ i $var_w(\hat{\beta}) = (X'X)^{-1}\left(\sum_{i=1}^{T} \sigma_i^2 X_i X_i'\right)(X'X)^{-1}$

Contrastes sota heteroscedasticitat 4

més que una restricció

Model amb k variables explicatives

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

La prova de White detecta heteroscedasticitat. Com podem contestar a l'hipòtesi següent

$$H_0 : R\beta = r$$

- sota <u>tots els supòsits del MLG</u> utilitzàvem
$$F = (R\hat{\beta} - r)'[R \cdot \tfrac{SQR}{T-k}(X'X)^{-1} \cdot R']^{-1}(R\hat{\beta} - r)/q \sim F_{q,T-k}$$
- on $\frac{SQR}{T-k}(X'X)^{-1}$ és l'estimador de $\sigma^2(X'X) = var(\hat{\beta})$

- però en el nostre cas, no el podem aplicar perquè

$$var(\hat{\beta}) = (X'X)^{-1}\left(\sum_{i=1}^{T}\sigma_i^2 X_i X_i'\right)(X'X)^{-1}$$

- llavors, al seu lloc podem ficar la variància estimada de White
- i utilitzar l'estadístic

$$F = (R\hat{\beta} - r)'[R \cdot \hat{var}_w(\hat{\beta}) \cdot R']^{-1}(R\hat{\beta} - r)/q \sim F_{q,T-k}$$

- però, sota el supòsit de pertorbacions normals
- en el cas contrari, hem de utilitzar la distribució χ_q^2
- si $F > F_{q,T-k}$, rebutjem

Exercici 1

Una empresa vol estimar la demanda de bitllets Y en funció de

- una constanta
- el preu X_2
- qualitat del servei X_3

Utilitzant 50 observacions ordenades segons la X_2, s'han fet els seguents estimacions

1. $Y_t = 0.68 - 0.3X_{2t} + 0.33X_{3t} + \hat{u}_t \qquad \hat{\sigma} = 184$
2. $\hat{u}_t^2 = \hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \hat{\beta}_3 X_{3t} + \hat{\beta}_4 X_{2t}^2 + \hat{\beta}_5 X_{3t}^2 + \hat{\beta}_6 X_{2t}X_{3t} + \hat{\nu}_t, \quad R^2 = 0.053$
3. $\hat{u}_t^2 = \hat{\gamma}_1 + \hat{\gamma}_2 \hat{u}_{t-1}^2 + \hat{\omega}_t \quad R^2 = 0.025$
4. $\hat{Y}_t = -0.10 - 0.93X_{2t} + 1.48X_{3t}, \qquad \hat{\sigma} = 204$
5. $\hat{Y}_t = 0.42 - 1.85X_{2t} + 1.66X_{3t}, \qquad \hat{\sigma} = 150$

on les primeres 3 estimacions són fetes utilitzant totes les observacions, la (4) amb les primers 20 i la (5) amb les últims 20 observacions

(a) fer el contraste de White per contrastar al supòsit d'homoscedasticitat

(b) contrastar el supòsit d'homoscedasticitat amb la prova de Goldfeld-Quandt

Exercici 2

Al model amb només una variable dependent X i sense constanta

$$Y_t = \beta X_t + u_t, \; t = 1 \dots T$$

es sap que $var(u_t) = \sigma^2 X_t$. A més, es disposa la següent informació

- $\sum X_t = 1977, \quad \sum Y_t = 1578, \quad \sum X_t^4 = 41630$
- $X_t^2 = 4897, \quad \sum X_t^2 e_t^2 = 6121341, \quad \sum \hat{u}_t^2 = 260125$
- $\sum X_t Y_t = 3851, \quad \sum e_t^2 = 380560, \quad T = 1000$

1. contrastar utilitzant el MQO, $H_0 : \beta = 0$, v.s. $H_1 : \beta \neq 0$
2. obtenir l'estimador MQG de β i calcular la seva variància