Heteroscdasticidad y MCG

Walter Sosa-Escudero

May 24, 2009

Modelo lineal clasico:

1. Linealidad: $Y = X\beta + u$.
2. Exogeneidad: $E(u) = 0$
3. No Multicolinealidad: $\rho(X) = K$.
4. No heteroscedasticidad ni correlacion serial: $V(u) = \sigma^2 I_n$.

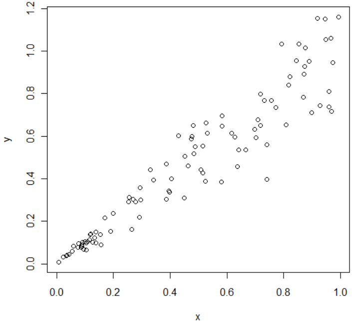
Teorema de Gauss/Markov $\hat{\beta} = (X'X)^{-1}X'Y$ es el mejor estimador lineal insesgado.

$\hat{V}(\hat{\beta}) = S^2(X'X)^{-1}$ es insesgado para $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$.

Que sucede si relajamos el supuesto de homocedasticidad?

- $\hat{\beta}$ (el estimador MCO) es lineal e insesgado (porque?).
- Pero no es mas el de minima varianza.
- $\hat{V}(\hat{\beta}) = S^2(X'X)^{-1}$ es sesgado. Invalida los estadisticos 't' y 'F'.

Intuition:



- La presencia de heterocedasticidad no altera la posicion 'central'de la linea de MCO (insesgadez).
- MCO pondera a todas las observaciones por igual. Pero en este caso parece mas razonable prestar mas atencion a aquellas en donde la varianza es mas chica.

Plan: que hacer con la heterocedasticidad?.

1. Antes de abandonar MCO: tests de heterocedasticidad.
2. *Estrategia 1:* Proponer otro estimador insesgado y eficiente para β (minimos cuadrados ponderados (WLS)) y un estimador apropiado para su varianza.
3. *Estrategia 2:* Mantener OLS (es insesgado, si bien ineficiente), pero reemplazar la varianza (la estandar es sesgada bajo heterocedasticidad).

Tests de Heterocedasticidad

a) Test de Breusch-Pagan/Godfrey/Koenker

Chequea si *ciertas* variables causan heterocedasticidad.
Consideremos el siguiente modelo heterocedastico:

$$Y = X\beta + u,\ u_i \sim \text{normal, con } E(u) = 0 \text{ y}$$

$$V(u_i) = h(\alpha_1 + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \ldots + \alpha_p Z_{pi})$$

donde $h(\ )$ es *cualquier* funcion positiva con dos derivadas.

Cuando $\alpha_2 = \ldots = \alpha_p = 0$, $V(u_i) = h(\alpha_1)$, una constante!!

Entonces, homocedasticidad $\Longleftrightarrow H_0 : \alpha_2 = \alpha_3 = \ldots = \alpha_p = 0$, y
$H_A : \alpha_2 \neq 0 \vee \alpha_3 \neq 0 \vee \ldots \vee \alpha_p \neq 0$.

Pasos para implementar el test:

1. Estimar por MCO y guardar los residuos al cuadrado, e_i^2.
2. Regresar e_i^2 en las variables Z_{ik}, $k = 2, \ldots, p$ and get (*ESS*). El estadistico es:

$$\frac{1}{2} ESS \sim \chi^2(p-1) \sim \chi^2(p)$$

bajo H_0, asintoticamente. Rechazamos si es demasiado grande.

Comentarios:

- Intuicion: El modelo auxiliar puede ser visto como un intento de modelar la varianza del termino de error. Si el R^2 de esta regresion es alto, entonces se puede explicar el comportamiento de los residuos al cuadrado, lo cual es compatible con que la varianza no es constante.
- Aceptar la hipotesis nula no implica que no haya heterocedasticidad. Porque?
- Test de muestras grandes.
- Koenker (1980) propone usar nR^2_A, que ademas es valido si los errores son no-normales.

b) Test de White

H_0 : no heteroscedasticidad, H_A : heterocedasticidad de alguna forma.

Consideremos un caso simple, con $K = 3$:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad 1, \ldots, n$$

Pasos para implementar el test:

1. Estimar por MCO, guardar los residuos al cuadrado e^2.
2. Regresar e^2 en todas las varaibles, sus cuadrados y todos los posibles productos cruzados no redundantes. En nuestro caso, regresar e^2 en $1, X_2, X_3, X_2^2, X_3^2, X_2X_3$, obtener el R^2 de esta regresion auxiliar.
3. Bajo H_0, $nR^2 \sim \chi^2(p)$. $p =$ numero de variables explicativas en el modelo auxiliar, menos una.
4. Rechzar H_0 si nR^2 es demasiado grande.

Comentarios:

- Intucion: hipotesis nula mas general que el de BPK.
- Informativo si no se rechaza la hipotesis nula (no heterocedasticidad).
- Si rechaza H_0 : hay heterocedasticidad. No tenemos informacion de que la causa.

Estimacion e inferencia bajo heterocedasticidad

Por simplicidad, consideremos el caso de dos variables

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

en donde ahora el termino de error es heteroscedastico, esto es

$$V(u_i) = \sigma_i^2, \quad i = 1, \ldots, n$$

Supondremos que todos los otros supuestos clasicos valen.

Dividamos cada observacion por σ_i:

$$\begin{aligned}\frac{Y_i}{\sigma_i} &= \beta_1 \frac{1}{\sigma_i} + \beta_2 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i} \\ Y_i^* &= \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \beta_k X_{ki}^* + u_i^*\end{aligned}$$

Notar que $V(u_i^*) = V(u_i/\sigma_i) = 1$, entonces, los residuos de este modelo transformado son *homoscedasticos.*

Entonces, si conocemos σ_i^2, el MELI es simplemente el estimador de MCO usando las variables transformadas.

En nuestro caso, el estimador MCO del modelo transformado es

$$\hat{\beta}_{2,wls} = \frac{\sum_{i=1}^{n} x_i^* y_i^*}{\sum_{i=1}^{n} x_i^{*2}}$$

$$\hat{\beta}_{1,wls} = \bar{Y}^* - \hat{\beta}_{2,wls}\bar{X}^*$$

con $Y_i^* = Y_i/\sigma_i$, $X_i^* = X_i/\sigma_i$, y las letras minuscuas indican diferencias con respecto a las medias muestrales.

Este es el estimador de minimos cuadrados ponderados.

El nombre minimos cuadrados ponderados viene del hecho de que este estimador se puede obtener resolviendo el siguiente problema de minimizacion:

$$\min \sum_{i=1}^{n} \frac{1}{\sigma_i^2} e_i^2,$$

esto es, los errores entran en la SRC ponderados por la inversa de la varianza de cada observaciono: estamos prestando mas atencion a las observaciones con menor varianza.

Problema: en la practica no conocemos σ_i^2. Esto conduce a dos estrategias.

1. Usas WLS: Pros: estimaciones insesgadas y eficientes. Contra: es necesario saber las varianzas de antemano (o realizar supuestos)
2. Mantener MCO pero cambiar la estimacion de la varianza: Pensemos en los efectos de la heterocedasticidad. MCO es insesgado pero ineficiente (not esta tan mal...). But, $S^2(X'X)^{-1}$ is biased, which invalidates inference (this is bad!). Then, a second strategy: keeping OLS for β and look for a valid estimator for its variance. Pros: no hacen falta supuestos, unbiased. Cons: perdermos eficiencia, pero con respecto a WLS (si esta disponible).

1) Varianza conocida: WLS

Consideremos el caso simple con dos variables:
$Y_i = \beta_1 + \beta_2 X_i + u_i$

Estrategia: suponer alguna forma particular de heterocedasticidad.

a) $V(u_i) = \sigma^2 X_i^2$

σ^2 es una constante desconocida. Dividamos todas las observaciones por X_i

$$\frac{Y_i}{X_i} = \beta_1 \frac{1}{X_i} + \beta_2 + \frac{u_i}{X_i}$$

$$Y_i^* = \beta_1 X_{0i}^* + \beta_2 + u_i^*$$

Notar que $E(u_i^*) = E(u_i/X_i)^2 = \sigma^2 \frac{X_i^2}{X_i^2} = \sigma^2$

Los errores del modelo transformado son homocedaticos: MCO en

- No hace falta conocer σ^2. Hemos dividido solo por la parte del error estandar que varia con las obervaciones, X_i.
- Esta estrategia prove una estimacion WLS.
- Cuidado con las interpretaciones. El intercepto de este modelo transformado es la *pendiente* del original y la pendiente del transformado es el intercepto del original.

b) $V(u) = \sigma^2 X_i$

$$\begin{aligned}\frac{Y_i}{\sqrt{X_i}} &= \beta_1 \frac{1}{\sqrt{X_i}} + \beta_2 \sqrt{X_i} + \frac{u_i}{\sqrt{X_i}} \\ Y_i^* &= \beta_1 X_{0i}^* + \beta_2 X_{1i}^* + u_i^*\end{aligned}$$

Con respecto a la implementacion e interpretacioon, notar que este modelo transformado no tiene intercepto, y que el coeficiente de la primera variable explicativa se corresponde con el intercepto y la segudna con la pendiente del modelo original.

Problema con estas estrategias: es muy dificil encontrar una forma exacta de heterocedasticidad.

2) Estructura de varianza desconocida

Estrategia alternativa: retener MCO (insesgado pero no eficiente) y buscar un estimador valido para la varianza. Bajo heterocedasticidad, la varianza de $\hat{\beta}_{OLS}$ es:

$$V(\hat{\beta}_{OLS}) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$$

$\Omega = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$.

White (1980): un estimador *consistentente* para $X'\Omega X$ es $X'DX$, $D = diag(e_1^2, e_2^2, \ldots, e_n^2)$, e_i's son los residuos de MCO.

Entonces, un estimador *consistente bajo heteroscedasticity* para la varianza es:

$$\hat{V}(\hat{\beta}_{OLS})_{HC} = (X'X)^{-1}X'DX(X'X)^{-1}$$

Estrategia: usar MCO pero reemplazar $S^2(X'X)^{-1}$ por el estimador de White. Esta estrategia no es eficiente, pero mo requiere supuestos acerca de la estructura de la heterocedasticidad.

Resumen

- La heterocedasicidad hace que MCO sea ineficiente e invalida el estimador estandar de la varianza, y por consiguiente, los procesos inferenciales estandar (t tests, F tests, etc.).
- El estimador WLS es eficiente e insesgado, pero depende de conocer la estructura de varianza. En la practica es muy raro que este disponible.
- En la practica es mas comun mantener el esimador MCO y reemplazar el estimador de la varianza pr el de White, que no requiere supuestos.

El model generalizado y MCG

Supongamos que todos los supuestos clasicos valen, pero ahora

- $V(u|X) = \sigma^2\Omega$ en donde Ω es cualquier matriz simetrica $n \times n$ y positiva definida.

Estamos permitiendo heterocedasticidad y/o correlacion serial, pero no estamos imponiendo ninguna estrucutra sobre Ω.

Plan

1. Consecuencias de relajar $V(u|X) = \sigma^2 I_n$.
2. Entontrar un estimador optimo (MCG).

Consecuencias de relajar $V(u|X) = \sigma^2 I_n$

- $\hat{\beta}$ lineal e insesgado, Gauss Markov Theorem no vale.
- $V(\hat{\beta}|X)$ es ahora $\sigma^2(X'X)^{-1}\Omega(X'X)^{-1}$ (probar este resultado).
- $S^2(X'X)^{-1}$ es sesgado para $V(\hat{\beta})$.
- Los tests t no tienen distribucion t. Los tests F tampoco son validos.

Entonces, ignorar heterocedasticidad o correlacion serial, esto es, usar $\hat{\beta}$ y $\hat{V}(\hat{\beta}|X) = S^2(X'X)^{-1}$, prserva la propiedad de insesadez, perdiendo eficiencia, pero la inferencia estandar deja de ser valida.

Minimos Cuadrados Generalizados

Necesitamos priemro un resultado simple: si Ω es $n \times n$ simetrica y pod, entonces existe una matriz $n \times n$, C, no singular tal que:

$$\Omega^{-1} = C'C$$

Que significa esto, intuitivamente?

Consideremos el siguiente modelo transformado

$$Y^* = X^* \beta + u^*$$

con $Y^* = CY$, $X^* = CX$ y $u^* = Cu$.

Es facil verificar:

1. $Y^* = X^*\beta + u^*$, so the transformed model is trivially linear.
2. $E(u^*|X) = CE(u|X) = 0$
3. $\rho(X^*) = \rho(CX) = K$, w.p.1. (CX is a rank preserving tranformation of X!).
4. $V(u^*|X) =$

$$\begin{aligned} V(Cu|X) = E(CuuC'|X) &= CE(uu'|X)C' \\ &= C\sigma^2\Omega C' \\ &= \sigma^2 C[\Omega^{-1}]^{-1}C' \\ &= \sigma^2 C\left[(C'C)^{-1}\right]^{-1} C \\ &= \sigma^2 I_n \end{aligned}$$

Entonces...?

Dado que el modelo transformado satisface todos los supuestsos clasicos, vale el Teorema de Gauss-Markov Theorem *para el modelo transformado*, entonces, el MELI es:

$$\hat{\beta}_{gls} = (X^{*\prime} X^*)^{-1} X^{*\prime} Y^*$$

Este es el estimador de minimos cuadrados generalizado

- Cuidado: el estimador de MCG es un estimador MCO para un modelo transformado.
- Provee un MELI bajo heterocedasticidad o correlacion serial.
- Ahoa es claro en que sentido $\hat{\beta}$ es ineficiente.
- Es importante ver que las propiedades estadisticas dependen de la estructura subyacente (no son propiedades del estimador *per-se*).

Notar que

$$\begin{aligned}
\hat{\beta}_{gls} &= (X^{**}X^{*})^{-1}X^{**}Y^{*} \\
&= (X'C'CX)^{-1}X'C'CY \\
&= (X'\Omega^{-1}X)X'\Omega^{-1}Y
\end{aligned}$$

- Cuando $\Omega = I_n$, $\hat{\beta}_{gls} = \hat{\beta}$.
- La implementacion practica de $\hat{\beta}_{gls}$ require conocer Ω (si bien no σ^2.)
- Es facil chequear que $V(\hat{\beta}_{gls}) = \sigma^2(X^{*\prime}X^{*})^{-1}$.

MGC Factible

Supongamos que existe un estimador para Ω, que llamaremos $\hat{\Omega}$. Entonces, reemplazando Ω por $\hat{\Omega}$:

$$\hat{\beta}_{fgls} = (X'\hat{\Omega}^{-1}X)X'\hat{\Omega}^{-1}Y$$

Este es el estimador de MCG factible.

Es lineal e insesgado? Eficiente?