Econometría II. 3^a LADE
Tema 5. Heteroscedasticidad

1 Introducción

En los temas anteriores hemos analizado el caso en que la perturbación del modelo tiene un matriz de varianzas y covarianzas no escalar. Este supuesto se viola tanto en aquellos casos en los que los valores fuera de la diagonal principal son distintos de 0, caso de autocorrelación, como en aquellos en los que los valores de la diagonal principal son distintos entre sí. Este es el caso de heteroscedasticidad, que vamos a estudiar en este tema.

Una de las hipótesis del modelo lineal general es que $Var(u_i) = \sigma^2$ $\forall\, i = 1, 2, ..., N$. No obstante, existen situaciones en las que este supuesto no es asumible. Al margen de aquellas en las que se produce un error en la especi cación del modelo, caso estudiado en los capítulo anteriores, existen otras circunstancias en las que, en ausencia de errores en la especi cación, es posible observar problemas de heteroscedasticidad. Un ejemplo de ello lo tenemos en el uso de variables que miden comportamientos medios de grupos de agentes económicos. El uso de estos datos agregados puede producir la presencia de problemas de heteroscedasticidad, por cuanto las propias observaciones de la variable endógena no tienen por qué tener una misma varianza. En estas condiciones, el uso de mínimos cuadrados generalizados puede ser de gran utilidad.

El resto del tema se organiza de forma similar al anterior. En primer lugar estudiaremos los problemas que conlleva la presencia de heteroscedasticidad sobre la estimación mco. A continuación analizaremos cómo podemos determinar si en un modelo existen o no problemas de heteroscedasticidad. El tema termina discutiendo las soluciones posibles para el problema de heteroscedasticidad.

2 Consecuencias sobre la estimación mco

Supongamos queremos estudiar el modelo:

$$y = X\beta + u$$

disponiendo, para ello, de una muestra de tamaño N. Suponemos que el todas las hipótesis del modelo lineal general se cumple, con excepción de la que hace referencia a la matriz de varianzas y covarianzas del vector de perturbaciones que es igual a:

$$Var\left(u\right) = \quad = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{pmatrix}$$

En el caso que $\sigma_1^2 = \sigma_2^2 = ... = \sigma_N^2 = \sigma^2$, la matriz es diagonal y nos encontramos en el caso del modelo lineal general. Pero, si esta restricción no se cumple, entonces las varianzas son distintas entre sí, por lo que tenemos problemas de heteroscedasticidad. Por tanto, esto es un caso particular del que analizamos en el tema de mínimos cuadrados generalizados, por lo que conocemos qué efectos tiene sobre la estimación mco.

Las causas de por qué se producen estos problemas de heteroscedasticidad son diversas. En cualquier caso, todas ellas están relacionadas con el uso de datos de corte transversal, careciendo prácticamente de interés el estudio de problemas de heteroscedasticidad en presencia de datos de series temporales, al menos el tipo de heteroscedasticidad que estamos considerando en este tema. Para entender por qué se producen los problemas de heteroscedasticidad pensemos, por ejemplo, en un modelo que quiera determinar el consumo de un bien en función de la renta de un grupo de familias. Debemos tener en cuenta que el consumo de este bien dependerá de, al margen de factores del precio o de la renta, de otros menos tangibles como son los gustos. Por tanto, dos familias que tengan la misma renta, pueden consumir distintas cantidades del bien que estamos estudiando, simplemente por que a una de ellas le gusta el bien y a la otra no. No obstante, es también cierto que para aquellas familias de menor poder adquisitivo, la mayor parte de la renta estará destinada a cubrir los aspectos básicos como son comida, alimentación, etc. Por tanto, la diferencia de consumo en el bien no será diferente, dado que no pueden destinar más que una mínima parte de la renta a él. Esto supone que la varianza en el consumo será pequeña. Por el contrario, las familias que dispogan de mayor renta, y que deseen consumir este bien, podrán destinar buena parte de su renta a su consumo. Pero aquellos que no les guste el bien, podrán no consumir nada. En consecuencia, la varianza del consumo para aquellas familias de mayor renta será elevada. En de nitiva, que la varianza del consumo aumenta con la renta de las familias, lo que conlleva la existencia de los problemas de heteroscedasticidad.

Otra fuente de problemas de heteroscedasticidad viene dada por el uso de datos agregados. El uso de estos datos es relativamente frecuente en las aplicaciones empíricas, bien por que las propias fuentes nos ofrecen valores medios de empresas o familias, bien por que sea el propio investigador se construya este tipo de datos.

Por último, también es posible que los problemas de heteroscedasticidad pueden aparecer como consecuencia de una mala especi ación. Tal y como hicimos en el tema de autocorrelación, supongamos que el proceso generador de los datos viene dado por la siguiente relación:

$$PGD: \qquad y_i = \beta_1 + \beta_1 x_{2i} + \beta_3 x_{3i} + v_i$$

donde,por ejemplo, y es el consumo de un bien por parte de una conjunto de familias, x_2 es la renta de estas familias, x_3 es el precio de este bien y v_t es una perturbación aleatoria que sigue una distribución $niid(0, \sigma_v^2)$.

Supongamos ahora que el investigador encargado del análisis de la variable plantea una especi cación que no coincide con el proceso generador de los datos, por cuanto omite la variable x_2, por ejemplo. El modelo empírico, por tanto, queda de nido en los siguientes términos:

$$ME: \qquad y_i = \beta_1 + \beta_3 x_{3i} + e_i \tag{1}$$

donde e es la perturbación del modelo empírico. Es evidente, que dado que la especi cación empírica no coincide con el proceso generador de los datos, la perturbación del modelo empírico contiene no solo un componente aleatorio, sino que además incluye el componente determinístico omitido. Así, resulta que $e_i = \beta_2 x_{2i} + v_i$. Por tanto, el vector de perturbaciones tiene esperanza distinta de 0, lo que hace que el vector de estimadores sea sesgado. Pero este no es el único problema, ya que su varianza es igual a:

$$\begin{aligned} Var(e_i) &= Var(\beta_2 x_{2i} + v_i) = Var(\beta_2 x_{2i}) + Var(v_i) + 2\,Cov(\beta_2 x_{2t}, v_i) \\ &= \beta_2^2 Var(x_{2i}) + \sigma_v^2 \end{aligned}$$

donde hemos incluido el supuesto de no relación entre las variables explicativas y la perturbación del proceso generador de los datos, lo que implica que $Cov(\beta_4 x_{4t}, v_t) = 0$. En el caso de que $Var(x_{2i})$ no sea constante, entonces una incorrecta especi cación lleva acarreado la presencia de problemas de heteroscedasticidad. Desde este de punto de vista, la presencia de heteroscedasticidad se debe asociar a problemas de especi cación.

3 Contrastes de heterocedasticidad

En este apartado vamos a estudiar diversos métodos que nos permitan determinar la existencia o no de problemas de heterocedasticidad. El planteamiento que vamos a seguir es similar para todos ellos. Suponemos que estamos interesados en estimar el siguiente modelo:

$$y = X\beta + u \tag{2}$$

donde, bajo la hipótesis nula, resulta que $Var(u_i) = \sigma^2$. Disponemos de una muestra de tamaño N para cada una de las variables.

3.1 Razón de varianzas

Bajo este supuesto de heteroscedasticidad, la varianza de la perturbación es no es idéntica para todas las observaciones. Sin embargo, esto no implica que sea necesario que todas las varianzas sean distintas entre sí. Este último lleva consigo que si queremos estimar cada una de las distintas varianzas sea necesario determinar la existencia de N varianzas distintas entre sí. Dado que disponemos de N observaciones, no va a ser posible estimar todos los parámetros del modelo. Por tanto, resulta aconsejable introducir algún tipo de restricción en el comportamiento de las varianzas. Por ejemplo, podemos considerar que la muestra total se puede subdividir en m submuestras, tales que todas ellas presentan una varianza constante dentro de cada submuestra, aunque no todas ellas coinciden entre sí. De acuerdo a este supuesto, el modelo anterior se puede particionar de la siguiente manera:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix} \beta + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix}$$

donde, $\forall\, i = 1, 2, ..., m$, Y_i es un vector columna de orden $N_i \times 1$, X_i es un matriz de orden $N_i \times k$, , β es un vector de parámetros de orden $k \times 1$ y, nalmente, u_i es un vector de perturbaciones aleatorias de orden $N_i \times 1$. Implícitamente estamos suponiendo que $\sum_{i=1}^{m} N_i = N$ y que $N_i > k$, $\forall\, i = 1, 2$. La matriz de varianzas y covarianzas del vector de perturbaciones es igual a:

$$
\begin{aligned}
Var\ (u) &= E\left(u\,u'\right) = E\left[\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix}\begin{pmatrix} u_1' & u_2' & \cdots & u_m' \end{pmatrix}\right] = \\
&= \begin{pmatrix} \sigma_1^2 I_{N_1} & 0 & \cdots & 0 \\ 0 & \sigma_2^2 I_{N_2} & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & \sigma_m^2 I_{N_m} \end{pmatrix}
\end{aligned}
$$

Si se cumple que $\sigma_1^2 = \sigma_2^2 = ... = \sigma_m^2$, la hipótesis de homoscedasticidad es cierta. Por el contrario, si al menos dos submuestras presentan valores de la varianza distintos, tenemos problemas de heteroscedasticidad. Para estudiar esta última circunstancia podemos desarrollar un estadístico basado en el uso de la razón de verosimilitud. Esto supone que es necesario conocer el valor de la función de verosimilitud tanto bajo la hipótesis nula como bajo la alternativa.

La función log- verosimilitud procedente de las estimación del modelo (2), bajo el supuesto del cumplimiento de la hipótesis nula, es igual a:

$$
\mathcal{L}\left(\beta, \sigma^2; y\right) = -\frac{N}{2}\ln\, 2\pi - \frac{N}{2}\ln\, \sigma^2 - \frac{u'u}{2\,\sigma^2}
$$

mientras que si estimamos este modelo para cada una de las submuestras, la función log-verosímil adopta esta forma:

$$
\mathcal{L}_i\left(\beta, \sigma_i^2; y\right) = -\frac{N_i}{2}\ln\, 2\pi - \frac{N_i}{2}\ln\, \sigma_i^2 - \frac{u_i'u_i}{2\,\sigma_i^2}\,,\ \ i = 1, 2, ..., m
$$

Bajo la hipótesis alternativa de que las submuestras no tienen una única varianza, la función de verosimiliud del conjunto es igual a:

$$
\begin{aligned}
\mathcal{L}_A &= \sum_{i=1}^{m} \mathcal{L}_i = \sum_{i=1}^{m}\left(-\frac{N_i}{2}\ln\, 2\pi - \frac{N_i}{2}\ln\, \sigma_i^2 - \frac{u_i'u_i}{2\,\sigma_i^2}\right) \\
&= -\frac{N}{2}\ln\, 2\pi - \frac{1}{2}\sum_{i=1}^{m} n_i\, \ln\, \sigma_i^2 - \frac{1}{2}\sum_{i=1}^{m}\frac{u_i'u_i}{\sigma_i^2}
\end{aligned}
$$

Los estimadores máximo verosímiles de cada una de las submuestras son $\tilde{\beta}_i = (X'X)^{-1}X'y_i$, $\tilde{\sigma}_i^2 = \frac{\hat{u}_i'\hat{u}_i}{N_i}$, donde $\hat{u}_i = y_i - X\tilde{\beta}_i$, siendo $i = 1, 2, ..., m$, mientras que los obtenidos para el total de la muestra son $\tilde{\beta} = (X'X)^{-1}X'y$, $\tilde{\sigma}^2 = \frac{\hat{u}'\hat{u}}{N}$, donde $\hat{u} = y - X\tilde{\beta}$. Si concentramos las diversas funciones de verosimilitud, tenemos que:

$$
\begin{aligned}
\mathcal{L}^* &= -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln \tilde{\sigma}^2 - \frac{\left(y - X\tilde{\beta}\right)'\left(y - X\tilde{\beta}\right)}{2\,\tilde{\sigma}^2} \\
&= -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln \tilde{\sigma}^2 - \frac{N}{2} \qquad (3)
\end{aligned}
$$

y para cada una de las submuestras:

$$
\begin{aligned}
\mathcal{L}_i^* &= -\frac{N_i}{2}\ln 2\pi - \frac{N_i}{2}\ln \tilde{\sigma}_i^2 - \frac{\left(y_i - X\tilde{\beta}_i\right)'\left(y_i - X\tilde{\beta}_i\right)}{2\,\tilde{\sigma}_i^2} \\
&= -\frac{N_i}{2}\ln 2\pi - \frac{N_i}{2}\ln \tilde{\sigma}_i^2 - \frac{N_i}{2},\; i = 1, 2, ..., m
\end{aligned}
$$

por lo que la función de verosimilitud concentrada bajo la hipótesis alternativa es:

$$
\begin{aligned}
\mathcal{L}_A^* &= -\sum_{i=1}^{m}\frac{N_i}{2}\ln 2\pi - \sum_{i=1}^{m}\frac{N_i}{2}\ln \tilde{\sigma}_i^2 - \sum_{i=1}^{m}\frac{N_i}{2} \\
&= -\frac{N}{2}\ln 2\pi - \sum_{i=1}^{m}\frac{N_i}{2}\ln \tilde{\sigma}_i^2 - \frac{N}{2} \qquad (4)
\end{aligned}
$$

El contraste de la hipótesis nula $\mathrm{H}_o : \sigma_1^2 = \sigma_2^2 = ... = \sigma_m^2$ es muy sencillo ahora ya que podemos construir un estadístico basado en el principio de la razón de verosimilitud. Este estadístico se de ne de la siguiente forma:

$$
-2\,(\mathcal{L} - \mathcal{L}_A) = -N\ln \tilde{\sigma}^2 - \sum_{i=1}^{m} N_i \ln \tilde{\sigma}_i^2 \overset{AS}{\sim} \chi_{m-1}^2
$$

donde hemos hecho uso de los resultados expresados en (3) y en (4). Este estadístico sigue asintóticamente una distribución χ_{m-1}^2.

3.2 Goldfeld-Quandt

En el caso anterior hemos supuesto la existencia de m posibles submuestras. En el caso de que $m = 2$, podemos utilizar un estadístico alternativo para contrastar la presencia de problemas de heteroscedasticidad. Si estimamos el modelo (2) por mco para cada una de las 2 submuestras disponibles, conocemos que es cierto el siguiente resultado:

$$\frac{\hat{u}_i'\hat{u}_i}{\sigma_i^2} \sim \chi^2_{N_i-k}, \;\; i = 1, 2$$

Entonces, asumiendo sin pérdida de generalidad que $\sigma_2^2 > \sigma_1^2$, se puede construir el siguiente estadístico:

$$\frac{\frac{\hat{u}_2'\hat{u}_2}{(N_2-k)\,\sigma_2^2}}{\frac{\hat{u}_2'\hat{u}_2}{(N_1-k)\,\sigma_1^2}} = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \sim F_{(N_2-k),\,(N_1-k)} \tag{5}$$

donde $\hat{\sigma}_i^2 = \frac{\hat{u}_i'\hat{u}_i}{(N_i-k)}$, $i = 1, 2$ El ratio anterior es un cociente de distribuciones χ^2. Dado que ambas son independientes entre sí, ya que hemos mantenido el supuesto de incorrelación de las perturbaciones, entonces bajo la hipótesis nula H_o: $\sigma_1^2 = \sigma_2^2$ este cociente se distribuye según una distriución F de $(N_2 - k)$ y $(N_1 - k)$ grados de libertad en el numerador y en el denominador, respectivamente. Cuanto más similares sean las varianzas de las dos submuestras, más próximo a la unidad se encontrará el valor del estadístico anterior. Por contra, cuanto más diferentes sean, mayor será el valor del estadístico, ya que $\hat{\sigma}_2^2 > \hat{\sigma}_1^2$, y, por tanto, se tenderá hacia el rechazo de la hipótesis nula.

Basándose en este resultado, Goldfeld y Quandt (1965) proponen un estadístico que mejora al anterior en un aspecto fundamental: la selección de las submuestras. En efecto, tanto en el estadístico de nido en (5) como en el caso general estudiado en la sección anterior, hemos supuesto que se conocen tanto el número de segmentos en los que hay de dividir la muestra, como en el punto exacto en el que se producen los saltos en las varianzas. En la práctica esto no es asumible, ya que el investigador va a carecer de esta información. Para solucionar el problema, podemos utilizar el estadístico diseñado en Goldfeld y Quandt (1965). Estos autores suponen la existencia de 2 submuestras que presentan varianzas distintas entre sí. Además, consideran que, bajo la hipótesis alternativa, las varianzas de las dos submuestras dependen de una variable explicativa z, que puede estar incluída en la especi cación del modelo o no. De esta forma, bajo la hipótesis alternativa tenemos que $\sigma_1^2 = \sigma^2 f(z)$ y $\sigma_2^2 = \sigma^2 h(z)$, con $\sigma_1^2 \neq \sigma_2^2$. Por facilidad en la exposición, vamos a considerar que $\sigma_1^2 < \sigma_2^2$. Con estas características, estos autores proponen obtener un estadístico similar a (5), pero obtenido de la siguiente manera:

1. Ordenamos todas las observaciones de la muestra en función de la variable z

2. Dividimos el total de la muestra en dos partes iguales

3. Estimamos por mco el modelo que estamos estudiando para cada una de las dos submuestras.

4. Calculamos la suma residual asociada a cada una de las estimaciones del apartado anterior, a las que denominamos $\hat{u}_1'\hat{u}_1$ y $\hat{u}_2'\hat{u}_2$

5. El estadístico de contraste se de ne como:

$$GQ = \frac{\hat{u}_2'\hat{u}_2}{\hat{u}_1'\hat{u}_1} \sim F_{\frac{N}{2}-k,\ \frac{N}{2}-k} \tag{6}$$

bajo la hipótesis nula $H_o: \sigma_1^2 = \sigma_2^2$, este estadístico sigue una distribución $F_{\frac{N}{2}-k,\ \frac{N}{2}-k}$. Cabe destacar que este estadístico se puede obtener a partir de (5), sin más que imponer que las dos submuestras tengan los mismos grados de libertad. Como La intuición de este estadístico nos dice que cuanto más similares sean las sumas residuales de ambos modelos, más próximo estará el valor del estadístico GQ a la unidad y, por tanto, con mayor facilidad se aceptará la hipótesis nula. Por contra, cuanto más diferentes sean estas sumas residuales, mayor será el valor del estadístico, lo que nos conducirá a rechazar la hipótesis nula. Por otro lado, hay que señalar que un valor del estadístico GQ muy próximo a 0 no implica la aceptación de la hipótesis nula. Supongamos, por ejemplo, que para una muestra de tamaño $N = 100$ con $k = 2$, las sumas residuales de las dos submuestras son $\hat{u}_1'\hat{u}_1 = 1.000$ y $\hat{u}_2'\hat{u}_2 = 1$. Tal y como está de nido, el valor del estadístico GQ es 0.001. Dado un nivel de signi cación del 5%, el valor crítico de una disstribución F de 48 grados de libertad, tanto en el numerador como en el denominador, es 1.60. Si comparamos ambos valores, la conclusión a la que llegamos es que la hipótesis nula es cierta. Sin embargo, esta conclusión se contrapone con la propia comparación de las sumas residuales, claramente diferentes entre sí. Para resolver este punto, debemos tener en cuenta que en (6) hemos considerado que $\hat{\sigma}_1^2 < \hat{\sigma}_2^2$, supuesto que no se cumple en nuestro ejemplo. Si de nimos el estadístico GQ de forma inversa todos los resultados concuerda, ya que ahora $GQ = 1.000$, lo que supone el rechazo de la hipótesis nula. para el nivel de signi cación elegido.

Otro problema que se nos puede presentar a la hora de calcular el estadístico de GQ es que las últimas observaciones de la primera submuestra serán muy similares a las primeras de la segunda submuestra. Por tanto, esto puede hacer que si los patrones de heteroscedastiticidad no están nítidamente diferenciados, la potencia del contraste disminuya. Para mejorar el funcionamiento de este contraste, los autores proponen eliminar las c observaciones centrales. De esta forma, los comportamientos medios quedan excluidos del cálculo, lo que hace que las diferencias entre las submuestras aumenten. La única repercusión que tiene la eliminación de estas observaciones centrales sobre el estadístico de Goldfeld-Quandt es que ahora las observaciones efectivas son $N - c$, por lo que los grados de libertad de cada submuestra son $\frac{N-c}{2} - k.$ Esto supone que a la hora de realizar el contraste haya que comparar el valor empírico del estadístico con los con los valores críticos de una dsitribución F de $\frac{N-c}{2} - k$ grados de libertad en el numerador y el denominador. Más di cultades plantea el cálculo apropiado del valor de c. Siguiendo a Goldfeld-Quandt, este valor no debe nunca exceder de $\frac{N}{3}$.

3.3 Glejser

Los estadísticos anteriores son capaces de darnos información sobre el grado de igualdad de las varianzas de una serie de submuestras. Sin embargo, no hacen ningún intento de modelizar el patrón heterocedástico que sigue la varianza de la perturbación. En ese sentido, los estadísticos anteriores son no constructivos. Debemos tener en cuenta que para aplicar mínimos cuadrados generalizados resulta indispensable conocer qué tipo de heterocedasticidad sigue la varianza de la perturbación. Desde este punto de vista, parece conveniente utilizar métodos que nos den alguna ayuda en este sentido.

Un primer intento es el contraste propuesto en Glejser (1969). La motivación de este estadístico es que bajo la hipótesis alternativa, la varianza de la perturbación del modelo (2) depende de una variable z. Entonces, este autor propone estimar la relación que existe que vincula a estas dos variables. Para ello, dado que desconocemos el valor de la varianza, utiliza como variable proxy de la desviación típica de la perturbación a $|\hat{u}|$, donde $\hat{u}$ son los residuos mco procedentes de la estimación de (2). Es posible también utilizar $\hat{u}^2$ como variable proxy de la varianza de la perturbación. Una vez determinada la variable a explicar, el siguiente paso es realizar una regresión entre ésta y la variable que creemos causa los problemas de heteroscedasticidad, en nuestro caso z. Como la forma funcional no tiene que ser lineal, se pueden utilizar formas funcionales alternativas. Por tanto, el estadístico de Glejser está basado en la estimación de la siguiente relación:

$$|\hat{u}_i| = \alpha_o + \alpha_1 \, z_i^h + e_i$$

donde h puede tomar valores $\{-1, 0, 1, 2\}$, por ejemplo. El siguiente paso es determinar qué forma funcional es la que mejor se adecúa a nuestro caso. Esto se puede realizar comparando, por ejemplo, los coe cientes de determinación del modelo. Una vez estimada la mejor de las relaciones, debemos contrastar la hipótesis nula $H_o : \alpha_1 = 0$, mediante el uso del estadístico t_{α_1} Si aceptamos esta hipótesis nula, es tanto como decir que no existe relación entre la desviación típica del modelo (2) y la variable z, lo que equivale a aceptar la hipótesis de homoscedasticidad. Si, por el contrario, encontramos que $\hat{\alpha}_1$ es signi cativamente distinto de 0, entonces concluímos que σ no es constante, por lo que debemos rechazar la hipótesis nula de homoscedasticidad.

3.4 Breush-Pagan LM test (Econometrica, 1979)

El planteamiento del contraste es el siguiente. Supongamos que queremos estudiar la existencia de heterocedasticidad en este modelo (2) donde sospechamos que la perturbación del modelo tiene un comportamiento heterocedástico. En concreto, estos autores suponen que:

$$u_i \sim N(0, \sigma_i^2), \ i = 1, 2, ..., N$$

Bajo la hipótesis nula, suponemos que la especi cación anterior es correcta, y que no existen problemas de heterocedasticidad, lo que se puede expresar así:

$$H_0 : \sigma_t^2 = \sigma^2 \quad \forall \ t = 1, 2, ..., N$$

Mientras que, bajo la hipótesis alternativa, la varianza de la perturbación sigue una función que depende de algunas variables. Esto se modeliza de la siguiente manera:

$$H_1 : \sigma_t^2 = h(z_t'\alpha)$$

donde z_t es un vector de p variables. Por simplicidad, asumimos linelalidad de forma que la relación anterior queda en los siguiente términos:

$$h(z_t'\alpha) = \alpha_o + \alpha_1 z_{1t} + ... + \alpha_p z_{pt}$$

De aquí que, la hipótesis nula de homoscedasticidad sea equivalente a formular la siguiente hipótesis

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \ldots = \alpha_q = 0.$$

En genereal, debemos considerar que la función h is basically unspeci ed but we need to specify a set of possible z variables (may include some or all of the X's).

El contrastes de Breusch-Pagan es un contrastes de multiplicadores de Lagrange. Para obtener esta distribución partimos de analizar la función de verosimilitud bajo la hipótesis nula. Si denotamos la función log-versoímil con $\mathcal{L}$, tenemos que:

$$\mathcal{L}(\beta, \alpha; y) = -\frac{T}{2}\ell n(2\pi) - \frac{1}{2}\Sigma_{t=1}^T \ell n\left[h\left(z_t'\alpha\right)\right] - \frac{1}{2}\Sigma_{t=1}^T \underbrace{\left(y_t - x_t'\beta\right)}_{u_t}^2 / h\left(z_t'\alpha\right).$$

Veámos primero que la matriz hesiana es diagonal por bloques:

- i) $\dfrac{\partial \mathcal{L}}{\partial \beta} = \left[\Sigma_{t=1}^T \left(y_t - x_t'\beta\right) x_t'\right] / h\left(z_t'\alpha\right);$

- ii) $\dfrac{\partial^2 \mathcal{L}}{\partial \beta_j \partial \alpha_i} = -\dfrac{\left[\Sigma_{t=1}^T \overbrace{\left(y_t - x_t'\beta\right)}^{u_t} x_{it}\right] h'\left(z_t'\alpha\right) z_{it}}{\left[h\left(z_t'\alpha\right)\right]^2}.$

Dado que asumismo que x_{it} y z_{it} son deterministas,

$$E\left[\frac{\partial^2 \mathcal{L}}{\partial \beta_j \partial \alpha_i}\right] = 0.$$

Por tanto, el Hessiano es diagonal por bloques y podemos olvidarnos de los elementos relacionados con β en el contraste. Entonces, el contraste de multiplicadores de Lagrange se de ne de la siguiente manera:

$$LM = \left(\overset{1\times q}{\frac{\partial \mathcal{L}(\tilde{\theta})}{\partial \alpha}} \right)' \overset{q\times q}{V_\alpha(\tilde{\theta})} \left(\overset{q\times 1}{\frac{\partial \mathcal{L}(\tilde{\theta})}{\partial \alpha}} \right)$$

donde

$$V_\alpha\left(\tilde{\theta}\right) = -\left[E\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial\alpha\partial\alpha'}\right]\right]^{-1}_{|\tilde{\theta}}$$

la matriz de información evaluada bajo la estimación restringida de $\tilde{\theta}$.

Tomando derivadas y evaluandalas en $\alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$, la estimación restringida, tenemos que

$$\frac{\partial \mathcal{L}(\theta)}{\partial \alpha_i} = -\frac{1}{2}h'(\cdot)\,\Sigma_{t=1}^T z_{it}/h_t + \frac{1}{2}h'(\cdot)\,\Sigma_{t=1}^T \frac{z_{it}u_t^2}{h_t^2},$$

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\theta)}{\partial\alpha_i\partial\alpha_j} &= \frac{1}{2}\left[h'(\cdot)\right]^2 \Sigma_{t=1}^T \frac{z_{it}z_{jt}}{h_t^2} - \left[h'(\cdot)\right]^2 \Sigma_{t=1}^T \frac{z_{it}z_{jt}u_t^2}{h_t^3} \\ &\quad + h''(\cdot)\, z_{jt}\{-\frac{1}{2}\Sigma_{t=1}^T z_{it}/h_t + \frac{1}{2}\Sigma_{t=1}^T \frac{z_{it}u_t^2}{h_t^2}\} \end{aligned}$$

y

$$\begin{aligned} E\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial\alpha_i\partial\alpha_j}\right] &= +\frac{1}{2}\left[h'(\cdot)\right]^2 \Sigma_{t=1}^T \frac{z_{it}z_{jt}}{h_t^2} - \left[h'(\cdot)\right]^2 \Sigma_{t=1}^T \frac{z_{it}z_{jt}h_t}{h_t^3} \\ &= -\frac{1}{2}\left[h'(\cdot)\right]^2 \Sigma_{t=1}^T \frac{z_{it}z_{jt}}{h_t^2} \end{aligned}$$

ya que $E\left[u_t^2\right] = h_t$. Evaluaando en $\tilde{\theta}$; tenemos que $\alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$, y $h(z_t'\tilde{\theta}) = \hat{\sigma}^2$, la estimación mco de la varianza de la perturbación. Por tanto,

$$\frac{\partial \mathcal{L}(\tilde{\theta})}{\partial \alpha_i} = \left(\frac{1}{2\hat{\sigma}^2}\right) h' \Sigma_{t=1}^T z_{it} f_t$$

donde

$$f_t = \frac{\hat{u}_t^2}{\hat{\sigma}^2} - 1$$

y

$$E\left[\frac{\partial^2\mathcal{L}(\theta)}{\partial\alpha_i\partial\alpha_j}\right]\tilde{\theta} = -\frac{1}{2\hat{\sigma}^4}(h')^2\,\Sigma_{t=1}^T z_{it}z_{jt}$$

El contrates LM es igual a: [notar que h' es un escalar]

$$\begin{aligned} LM &= \left[\left(\frac{1}{2\hat{\sigma}^2}\right)h'\Sigma_{t=1}^T z_t' f_t\right]'\left[\frac{1}{2\hat{\sigma}^4}(h')^2\,\Sigma_{t=1}^T z_t z_t'\right]^{-1}\left[\left(\frac{1}{2\hat{\sigma}^2}\right)h'\Sigma_{t=1}^T z_t' f_t\right] \\ &= \frac{1}{2}\left(\Sigma_{t=1}^T z_t' f_t\right)'\left[\Sigma_{t=1}^T z_t z_t'\right]^{-1}\Sigma_{t=1}^T z_t' f_t \\ &= \frac{1}{2}f'Z\,(Z'Z)^{-1}\,Z'f \end{aligned}$$

donde

$$\begin{aligned} f' &= [f_1,\ldots,f_T] \\ Z' &= [z_1,\ldots,z_T] \end{aligned}$$

Si suponemos que:

$$f_t = \frac{\hat{u}^2}{\hat{\sigma}^2} - 1 \equiv g_t - 1,$$

o, en forma de vector

$$f = g - i$$

con i siendo un vector de columna de unos. Por tanto, el contraste LM se puede escribir como:

$$LM = (\frac{1}{2})\,(g-i)'\,P_Z\,(g-i)$$

y, dado que $P_Z i = i$ (Z contiene una column de 1),

$$LM = \frac{1}{2}g'P_Z g + \frac{i'i}{2} - \frac{1}{2}g'i - \frac{1}{2}i'g.$$

Pero

$$g'i = \frac{1}{\hat{\sigma}^2}\Sigma_{t=1}^T\hat{u}_t^2 = \left(\frac{\Sigma_{t=1}^T\hat{u}_t^2}{T}\right)^{-1}\Sigma_{t=1}^T\hat{u}_t^2 = T = i'i.$$

Entonces,

$$LM = \frac{1}{2}g'P_Z g - \frac{1}{2}T.$$

Si recordamos que en una regresión de y sobre X, la suma explicada es igual a:

$$SE = \hat{y}'\hat{y} - T\bar{y}^2 = y'P_X y - T\overline{y}^2.$$

Entonces,

$$\begin{aligned} LM &= \frac{1}{2}g'P_Z g - \frac{T}{2} \\ &= \frac{1}{2}\left[g'P_Z g - T \underbrace{\overline{g}^2}_{1}\right]. \end{aligned}$$

since $\overline{g}^2 = 1$. Si consideramos una regression de g sobre Z; we have $\overline{g} = 1$ since $g'i = T$. Hence,

$$\boxed{LM = \frac{1}{2}SE \text{ de una regresión de } g \text{ sobre } Z.}$$

The procedure to construct the test can be summarized as follows:

1. Estimate the regression $y = X\beta + u$ by OLS and obtain the residuals $\hat{u}$;
2. Compute $g_t = \hat{u}_t^2/\hat{\sigma}^2$ where $\hat{\sigma}^2 = T^{-1}\Sigma_{t=1}^T \hat{u}_t^2$;
3. Run OLS of $g = (g_1, ..., g_T)$ with $g_t = \hat{u}^2/\hat{\sigma}^2$on Z (including a constant) and compute
$$LM = \frac{1}{2}ESS.$$
4. The asymptotic distribution under the null hypothesis is χ_p^2.

3.5 White test (*Econometrica* 1980)

Este estadístico se puede interpretar como un contraste basado en los multiplicadores de Lagrange. Tiene la ventaja, importante, de que no requiere que se determirme a priori el patrón de heteroscedasticidad que suponemos que sigue la perturbación aleatoria. El proceso que debemos seguir para su obtención es el siguiente:

1. Estimar el modelo $y = X\beta + u$ mediante mco y obtener el vector de residuos $\hat{u}$

2. Estimar por mco la regresión auxiliar

$$\hat{u}_t^2 = \alpha_0 + \Sigma_{j=1}^k \Sigma_{i=j}^k \alpha_{ij} X_{ti} X_{tj} + w_t;$$

3. El estadístico de White es igual a TR^2 donde T es el tamaño muestral dispònible y R^2 es el coe ciente de determinación de la regresión auxiliar

4. El estadístico de White sigue asintóticamente una distribución $\chi^2_{k(k-1)/2}$, donde los grados de libertad coinciden con el número de variables explicativas utilizadas en la regresión auxiliar.

Remark 1 *The test is based on comparing the lower triangular elements of*

$$\frac{T}{T-k} \underbrace{(X'X)^{-1} \left[\Sigma_{t=1}^T \hat{u}_t^2 x_t x_t'\right] (X'X)^{-1}}_{\substack{\text{heteroskedasticity-consistent} \\ \text{covariance matrix}}}$$

and

$$\hat{\sigma}^2 (X'X)^{-1}.$$

Remark 2 *This is a test of general misspeci cation as well as heteroskedasticity.*

4 Soluciones

En el caso de que los problemas de heteroscedasticidad vengan causados por una mala especi cación, la solución de los mismos tiene que venir de la mano de la corrección de estos problemas. Por ejemplo, la omisión de una variable relevante, la incorrecta selección de la forma funcional o no costancia de los parámetros del modelo pueden generar estos problemas. Otra posibilidad es que los problemas de heteroscedasticidad no tengan como origen una mala especi cación. En estos casos, la solución que debemos adoptar es el uso de estimadores e cientes, como es el caso de la estimación mínimo cuadrática generalizada. La estimación máximo verosímil también se puede utilizar, siempre que realicemos algún supuesto sobre la forma que adopta el patrón de heteroscedasticidad. En lo que sigue discutimos estos dos últimos métodos.

4.1 Mínimos cuadrados ponderados

La estimación por mínimos cuadrados generalizados ha sido objeto de estudio en temas precedentes, aunque desde un punto de vista muy general. En el caso de que la matriz de varianzas y covarianzas sea escalar, pero no diagonal, este método tiene las siguientes características. En primer lugar hay que tener en cuenta que la matriz de varianzas y covarianzas adopta la siguiente forma:

$$Var\left(u\right) = \quad = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_N \end{pmatrix}$$

donde hemos realizado el supuesto que $\sigma_i^2 = \sigma^2 \, w2_i$, $i = 1, 2, ..., N$, donde w_i puede representar una varaible que esté presente en la especi cación del modelo. La estimación mínimo cuadrática generalizada se de ne como:

$$\hat{\beta}_{GLS} = \left(X' \quad ^{-1}X\right)^{-1} X' \quad ^{-1}Y$$

donde $P'P = \quad ^{-1}$. En este caso, tenemos que:

$$^{-1} = \frac{1}{\sigma^2} \begin{pmatrix} w_1^{-1} & 0 & \cdots & 0 \\ 0 & w_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_N^{-1} \end{pmatrix}$$

por lo que ahora es muy sencillo obtener la matriz P:

$$P = \begin{bmatrix} w_1^{-1/2} & & \\ & \ddots & \\ & & w_N^{-1/2} \end{bmatrix}.$$

Por tanto, la transformación que es necesaria para realizar la estimación mínimo cuadrática generalizada es ahora muy sencilla. Si recordamos que este modelo transormado se de ne como:

$$y^* = X^* \beta + u^* \tag{7}$$

donde

$$y^* = P\,y = \begin{pmatrix} \frac{y_1}{w_1^{-1/2}} \\ \frac{y_2}{w_2^{-1/2}} \\ \vdots \\ \frac{y_N}{w_N^{-1/2}} \end{pmatrix}$$

$$X^* = P\,X = \begin{pmatrix} \frac{1}{w_1^{-1/2}} & \frac{x_{21}}{w_1^{-1/2}} & \cdots & \frac{x_{k1}}{w_1^{-1/2}} \\ \frac{1}{w_2^{-1/2}} & \frac{x_{21}}{w_2^{-1/2}} & \cdots & \frac{x_{k2}}{w_2^{-1/2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{w_N^{-1/2}} & \frac{x_{21}}{w_N^{-1/2}} & \cdots & \frac{x_{kN}}{w_N^{-1/2}} \end{pmatrix}$$

$$u^* = P\,u = \begin{pmatrix} \frac{u_1}{w_1^{-1/2}} \\ \frac{u_2}{w_2^{-1/2}} \\ \vdots \\ \frac{u_N}{w_N^{-1/2}} \end{pmatrix}$$

Como vemos, los mínimos cuadrados generalizados se obtienen simplemente de actando las observaciones de cada una de las variables utilizando $w_i^{-1/2}$ $(i = 1, 2, ..., N)$ como factor de de ación o ponderación. De ahí que a este caso particular se le conoce como mínimos cuadrados

ponderados. En lo que sigue vamos a estudiar este método de estimación para diversos patrones de heteroscedasticidad.
Case 1 *:* $Var(u_t) \propto x_{jt}^2$, one of the regressors ($\propto$ denotes proportional to). This case is useful if we have an obvious choice available. Then,

$$w_t = x_{jt}^2$$

and

$$P = \begin{bmatrix} x_{j1}^{-1} & & \\ & \ddots & \\ & & x_{jT}^{-1} \end{bmatrix}$$

The transformation is

$$\left(\frac{y_t}{x_{jt}}\right) = \beta_1 \left(\frac{1}{x_{jt}}\right) + \beta_2 \left(\frac{x_{2t}}{x_{jt}}\right) + \ldots + \beta_k \left(\frac{x_{kt}}{x_{jt}}\right) + e_t$$

Remark 3 *β_1 becomes a slope coefficient and β_j the intercept.*

Case 2 *:* $Var(u_t) \propto [E(y_t)]^2 = [x_t'\beta]^2$.

1) we run OLS of y on X to get $\hat{\beta}_{OLS}$,
2) then compute $\hat{w}_t = \left(x_t'\hat{\beta}\right)^2$,
3) transform the model by dividing through by $x_t'\hat{\beta}$ and run OLS.

Remark 4 *If the $X's$ are exogenous, $\hat{\beta}_{FA}$ is consistent and asymptotically efficient.*

Case 3 : $\ell n\,(\sigma_t^2) = z_t'\alpha$.

The model is

$$y = X\beta + u \tag{8}$$

with

$$\begin{aligned} V(u) &= diag\left(\sigma_1^2, \ldots, \sigma_T^2\right) \\ \ell n\left(\sigma_t^2\right) &= \underbrace{\alpha_1 + \alpha_2 z_{2t} + \ldots + \alpha_p z_{pt}}_{\text{p regressors including a constant}} . \end{aligned}$$

To obtain the feasible GLS estimates:

a) Run OLS on regression (8) to get $\hat{u}$.
b) Run the OLS regression

$$\ell n(\hat{u}_t^2) = z_t'\alpha + v_t$$

to obtain

$$\hat{\alpha} = (z'z)^{-1} z'h$$

where

$$h' = \left[\ell n\left(\hat{u}_1^2\right), \ldots, \ell n\left(\hat{u}_T^2\right)\right]$$

c) Transform the model with the matrix

$$P = diag\left[\hat{\sigma}_1^{-1}, \ldots, \hat{\sigma}_T^{-1}\right]$$

where

$$\hat{\sigma}_t^2 = \exp\left(z_t'\hat{\alpha}\right)$$

and run OLS.

Remark 5 *This requires the speci cation of the z_t variables (need not be the regressors x_t).*

4.2 Maximum likelihood for the case with $\ell n\left(\sigma_t^2\right) = z_t'\alpha$.

The likelihood function of the errors is:

$$L(\alpha, \beta | e) = \frac{1}{(2\pi)^{T/2}} \frac{1}{(\sigma^2)^{T/2}} \exp\left[-\frac{1}{2\sigma^2} e'e\right]$$

where

$$\begin{aligned} e &= Qu \\ Q &= diag\left(\sigma_1^{-1}, \ldots, \sigma_T^{-1}\right) \\ \sigma_t^2 &= \exp\left(z_t\alpha\right). \end{aligned}$$

The transformed likelihood in terms of the data y is:

$$L(\alpha, \beta; y) = \frac{1}{(2\pi)^{T/2}} \frac{1}{(\sigma^2)^{T/2}} \exp\left[-\frac{1}{2\sigma^2}(y^* - X^*\beta)'(y^* - X^*\beta)\right] \cdot |J|$$

where

$$\underset{T\times T}{\underset{\uparrow}{|J|}} = \left|\frac{\partial e_t}{\partial y_t}\right| = \left|D\left(e^{1/2z_t'\alpha}\right)\right|,$$

is the Jacobian of the transformation, and where $D(a_t)$ is a diagonal matrix with elements a_t. Taking the logarithm:

$$\ell nL(\alpha, \beta, y) = const - \frac{1}{2}\Sigma_{t=1}^{T} z_t'\alpha - \frac{1}{2}\Sigma_{t=1}^{T} \exp(-z_t'\alpha)(y_t - x_t'\beta)^2.$$

Note that here we have subsumed σ under α.

The rst-order conditions are:

$$\begin{aligned}
\frac{\partial \ell nL}{\partial \beta} &= \Sigma_{t=1}^{T} x_t \exp(-z_t'\alpha)(y_t - x_t'\beta) = 0 \quad \text{(k equations)} \\
\frac{\partial \ell nL}{\partial \alpha} &= \Sigma_{t=1}^{T} Z_t \left[\exp(-z_t'\alpha)(y_t - x_t'\beta)^2 - 1\right] = 0 \quad \text{(p equations)}
\end{aligned}$$

The $(p+k)$ equations are highly nonlinear in α and β. We may use the Newton-Raphson method to optimize numerically.

Otros textos recomendados

1. Greene, W.H., 1999, Análisis Econométrico, Ed. Prentice Hall Iberia, Madrid. Cap. 12.

2. Johnston, J., 1987, Métodos de Econometría, Ed. Vicens Vives, Barcelona. Cap. 8.

3. Judge, G.G., R. C. Hill, W. E. Griffiths, H. Lütkepohl y T.C. Lee, 1988, Introduction to the Theory and Practice of Econometrics, Ed. John Wiley & Sons. Cap. 9.

4. Kmenta, J., 1985, Elementos de Econometría, Ed. Vicens Vives, Barcelona. Cap. 8.

5. Maddala, G.S. (1992) Introduction to Econometrics. Ed: MacMillan. Cap. 5.

6. Novales, A. (1993). Econometría. Ed. McGraw-Hill. Cap. 6

7. Pindyck, R.S. y D. L. Rubinfeld, 2000, Econometría: Modelos y Pronósticos, Ed. Mc Graw-Hill. Cap. 6

Referencias

[1] Breusch, T. S. y A. R. Pagan (1979). A simple test for heteroscedasticity and random coefficient variation , Econometrica, 43, 1287-1294.

[2] Breusch, T. S. y A. R. Pagan (1980). The Lagrange multiplier test and is applications to model speci cation in econometrics , Review of Economic Studies, 47, 239-254.

[3] Glejser, H. (1969). A new test for heteroscedasticity , Journal of the American Statistical Association, 64, 316-323.

[4] Goldfeld, S. M. y R. E. Quandt (1965). Some tests for homoscedasticity , Journal of the American Statistical Association, 60, 539-547

[5] White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity , Econometrica, 48, 817-838.