CAPíTULO 11

Heterocedasticidad

11.1. Introducción

En el modelo lineal general

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, \ldots, n$$

hemos supuesto que $E(u_i) = 0$, $E(u_i^2) = \sigma_u^2$ y $E(u_i u_j) = 0\,\forall i \neq j$. El supuesto de homocedasticidad o varianza de los errores constante, $E(u_i^2) = \sigma_u^2$, es difícil de justificar en algunas situaciones, especialmente cuando utilizamos datos de sección cruzada. Veamos algunos ejemplos.

Ejemplo 18 (Datos de sección cruzada). *Si disponemos de observaciones sobre gasto en consumo Y_i y renta X_i para distintas familias, la ecuación de regresión simple*

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (i = 1, \ldots, n)$$

postula, con $\beta_2 > 0$, que el gasto en consumo esperado para la i-ésima familia es una función creciente de la renta

$$E(Y_i) = \beta_1 + \beta_2 X_i$$

Además, la ecuación postula que el gasto en consumo esperado será el mismo para todas las familias con la misma renta X_i. Sin embargo, sabemos que no todas las familias que tienen la misma renta realizan el mismo consumo, algunas gastan más y otras gastan menos. El error aleatorio u_i recoge las discrepancias entre el consumo relizado y el consumo esperado, $u_i = Y_i - E(Y_i)$. Cuanto mayores sean estas discrepancias, tanto mayor será la varianza del error. Ahora bien, las familias con una renta muy baja gastan prácticamente toda su renta, por lo que presentan un gasto en consumo muy similar y la varianza del error es pequeña. En cambio, las familias con una renta alta tendrán un gasto en consumo muy dispar, por lo que la varianza del error será mayor. En consecuencia, el término de error en este modelo es heterocedástico porque su varianza varía con la renta. El gráfico 18 representa esta forma de heterocedasticidad. Un argumento similar se aplica en el análisis de la relación entre dividendos Y_i y beneficios X_i empresariales. La varianza de la variable dividendos será mayor en el grupo de empresas grandes que en el grupo de empresas pequeñas.

Ejemplo 19 (Datos agregados). *En la ecuación de regresión*

$$Y_{ij} = \beta_1 + \beta_2 X_{ij} + u_{ij} \quad i = 1, \ldots, N;\ j = 1, \ldots, n_i$$

en donde $E(u_{ij}) = 0$, $E(u_{ij}^2) = \sigma_u^2$ y $E(u_{ij} u_{kl}) = 0\,\forall i \neq k, j \neq l$, podemos demostrar que la ecuación de regresión

$$\bar{Y}_{i.} = \beta_1 + \beta_2 \bar{X}_{i.} + \bar{u}_{i.} \quad i = 1, \ldots, N$$

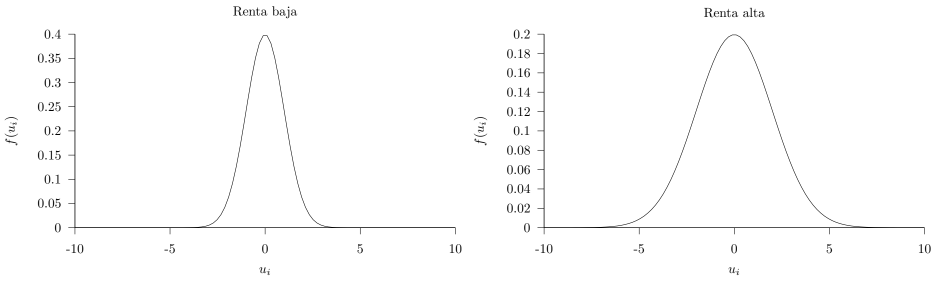

Figura 1: Errores heterocedásticos

presenta heterocedasticidad, siendo

$$\bar{Y}_{i.} = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i} \quad \bar{X}_{i.} = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i} \quad \bar{u}_{i.} = \frac{\sum_{j=1}^{n_i} u_{ij}}{n_i}$$

El doble subíndice de la ecuación podría representar a una persona j que reside en una provincia i, o bien a una empresa j de una industria i. El número de personas de la provincia i (o de empresas de la industria i) en la muestra es n_i.

Vamos a obtener las propiedades estadísticas del término de error $\bar{u}_{i.}$:

- *Media*

$$E(\bar{u}_{i.}) = E\left(\frac{\sum_{j=1}^{n_i} u_{ij}}{n_i}\right) = \frac{\sum_{j=1}^{n_i} E(u_{ij})}{n_i} = 0$$

- *Varianza*

$$E(\bar{u}_{i.})^2 = E\left(\frac{\sum_{j=1}^{n_i} u_{ij}}{n_i}\right)^2 = E\left(\frac{\sum_{j=1}^{n_i} u_{ij}^2 + \sum_{j\neq l}^{n_i} u_{ij}u_{il}}{n_i^2}\right) = \frac{\sum_{j=1}^{n_i} E(u_{ij}^2) + \sum_{j\neq l}^{n_i} E(u_{ij}u_{il})}{n_i^2} = \frac{\sigma_u^2}{n_i}$$

- *Covarianza*

$$E(\bar{u}_{i.}\bar{u}_{k.}) = E\left(\frac{\sum_{j=1}^{n_i} u_{ij}\sum_{l=1}^{n_k} u_{kl}}{n_i^2}\right) = E\left(\frac{\sum_{j=1}^{n_i}\sum_{l=1}^{n_k} u_{ij}u_{kl}}{n_i^2}\right) = \frac{\sum_{j=1}^{n_i}\sum_{l=1}^{n_k} E(u_{ij}u_{kl})}{n_i^2} = 0$$

Vemos que, en la ecuación de regresión con datos per cápita, el término de error es heterocedástico: su varianza varía inversamente con el número de observaciones usadas para calcular la media $E(\bar{u}_{i.})^2 = \sigma_u^2/n_i$.

Estos dos ejemplos ilustran las dos situaciones más comunes en que podemos encontrar peturbaciones heterocedásticas con datos de sección cruzada: (1) cuando las observaciones de la variable dependiente pueden dividirse en grupos dependientes de la varianza local y (2) cuando trabajamos con datos agrupados o datos *per cápita*. Muchas series temporales también presentan heterocedasticidad, pero pueden transformarse en homocedásticas con la transformación logarítmica: si la varianza local de Y_t aumenta linealmente con la media local de Y_t, entonces la serie $\log(Y_t)$ suele ser homocedástica.

DEFINICIÓN 87. *El modelo lineal general correctamente especificado presenta* **heterocedasticidad pura** *cuando los errores tienen distinta varianza, $E(u_i^2) = \sigma_i^2$ (el subíndice de σ_i^2 indica que la varianza cambia con la observación i).*

La **heterocedasticidad impura** viene causada por un error de especificación debido a la omisión de variable dependiente. Por ejemplo, si la variable Y depende de X_2

y X_3

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i, \quad u_i \sim N(0, \sigma_u^2)$$

y erróneamente especificamos la relación

$$Y_i = \beta_1 + \beta_2 X_{2i} + e_i$$

entonces el error $e_i = \beta_3 X_{3i} + u_i$ presentará heterocedasticidad

$$E(e_i)^2 = \sigma_u^2 + \beta_3^2 X_{3i}^2$$

Observación 65. *La distinción entre heterocedasticidad pura e impura es importante para determinar cómo solucionar el problema. En el segundo caso, el tratamiento consiste en incluir la variable omitida.*

En un modelo de regresión con heterocedasticidad pura, la matriz de varianzas y covarianzas de los errores tiene la forma

$$E(\mathbf{u}\mathbf{u}') = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix} \tag{11.1}$$

que contiene n parámetros desconocidos. De aquí, el número total de parámetros del modelo, $n + k$, es mayor que el número de observaciones, n, por lo que no es posible la estimación conjunta de los mismos.

Tradicionalmente, para resolver este problema se ha supuesto que la varianza del error depende de alguna variable conocida. Por ejemplo, en la relación entre consumo y renta parece razonable pensar que la varianza del error es función de la renta, de modo que podemos especificar $E(u_i^2) = \sigma_u^2 X_i$; y en el modelo con datos medios, $E(\bar{u}_{i.})^2 = \sigma_u^2 / n_i$. De esta forma, conseguimos reducir el número de parámetros desconocidos de $n + k$ a $k + 1$, siendo ahora posible la estimación del modelo.

Es conveniente escribir la matriz de varianzas y covarianzas (11.1) como

$$E(\mathbf{u}\mathbf{u}') = \sigma_u^2 \begin{pmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_n \end{pmatrix} = \sigma_u^2 \mathbf{\Omega} \tag{11.2}$$

en donde podemos definir $\sigma_u^2 = \sigma_1^2$ y $\omega_i = \sigma_i^2 / \sigma_1^2$. De este modo, para $\mathbf{\Omega} = \mathbf{I}_n$ obtenemos como caso especial el modelo de regresión con las hipótesis básicas.

Observación 66. *Las formulaciones (11.1) y (11.2) de la matriz de varianzas y covarianzas son idénticas y contienen el mismo número de parámetros.*

El modelo de regresión con heterocedasticidad es un caso especial del modelo de regresión con perturbaciones no esféricas. Por tanto, en presencia de heterocedasticidad, el estimador de mínimos cuadrados ordinarios es lineal, insesgado y consistente, pero deja de ser eficiente. Para aplicar mínimos cuadrados generalizados necesitamos proponer una forma para la heterocedasticidad de las perturbaciones. Por tanto, resulta interesante disponer de métodos de detección que nos indiquen si los residuos de un modelo son homocedásticos o heterocedásticos, y, en su caso, la forma de la heterocedasticidad. Es este tema estudiamos, primero, métodos gráficos y contrastes estadísticos para detectar heterocedasticidad en el análisis de datos transversales y, después, una variante del método de estimación MCG: mínimos cuadrados ponderados.

11.2. Detección de heterocedasticidad

Si, como se ilustra en el ejemplo (18), pensamos que la varianza del error σ_i^2 puede depender de la renta, entonces podemos especificar la ecuación de regresión

$$u_i^2 = \alpha_1 + \alpha_2 X_i + e_i$$

Suponiendo que el error e_i tiene media cero, tenemos

$$E(u_i^2) \equiv \sigma_i^2 = \alpha_1 + \alpha_2 X_i$$

Esta ecuación nos sugiere dos ideas interesantes. En primer lugar, en la detección de heterocedasticidad es conveniente analizar la relación que u_i^2 mantiene con las variables explicativas del modelo. En segundo lugar, como los datos de sección cruzada no presentan una ordenación natural, podemos ordenarlos de acuerdo a los valores crecientes o decrecientes de X_i. De este modo, podemos evaluar si valores grandes de u_i^2 se corresponden con valores grandes ($\alpha_2 > 0$) o pequeños ($\alpha_2 < 0$) de X_i. Ahora bien, como los errores u_i no se observan, podemos utilizar los residuos del modelo estimado por mínimos cuadrados ordinarios. Los contrastes de heterocedasticidad tratan de encontrar relaciones entre los cuadrados de los residuos $\hat{u}_i^2$ y las variables explicativas.

Como en el caso de detección de autocorrelación, los métodos gráficos nos ayudan a entender la intuición subyacente a los contrastes estadísticos más formales.

11.2.1. Métodos gráficos.

11.2.1.1. Gráfico de residuos. El gráfico de los residuos, o de sus cuadrados, es un diagrama de dispersión de $\hat{u}_i$ frente a i, siendo una herramienta muy útil en la detección de heterocedasticidad. Al examinar un gráfico de residuos debemos formar grupos de observaciones (por ejemplo, la primera mitad de observaciones y la segunda) y comprobar si la varianza local permanece aproximadamente constante en cada grupo. En el gráfico (2) podemos ver que la varianza local para el grupo de observaciones 30-40 es mayor que la varianza local para el grupo observaciones 70-80.

La forma de un gráfico de residuos depende del orden de los datos. Por tanto, hay que asegurarse que los datos de sección cruzada están ordenados de acuerdo a los valores de una variable dependiente.

11.2.2. Gráficos de dispersión. Si sospechamos que una variable explicativa puede causar heterocedasticidad, entonces el diagrama de dispersión de los residuos frente a esta variable es una herramienta más informativa que el gráfico de residuos. En el primer gráfico de la figura (3) podemos ver que la varianza local de los residuos aumenta con el valor de x_i. De aquí, concluimos que la varianza del error depende de x_i. Otros diagramas de dispersión que podemos examinar son los residuos al cuadrado frente a las variables x_i, x_i^2, $\hat{y}_i$ o $\hat{y}_i^2$. En el segundo gráfico de la figura (3) podemos ver que el cuadrado del residuo aumenta con el cuadrado de x_i.

11.2.3. Métodos formales.

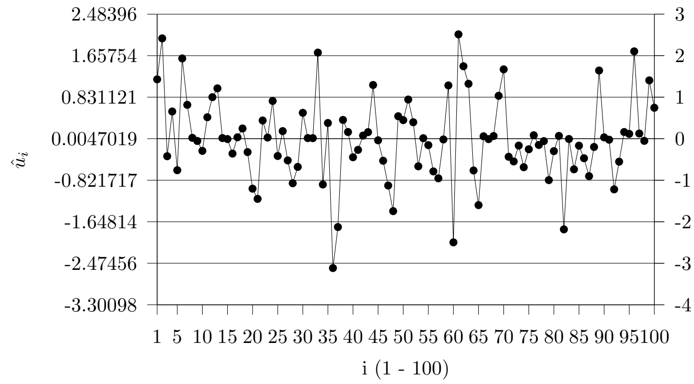

Figura 2: Gráfico de residuos

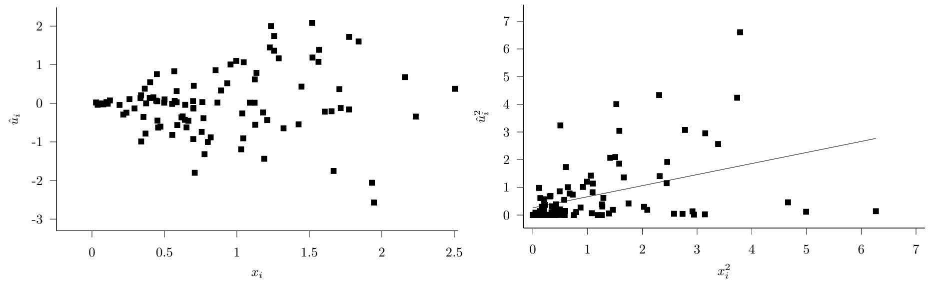

Figura 3: Diagramas de dispersión de $\hat{u}_i$ versus x_i y de $\hat{u}_i^2$ versus x_i^2

11.2.3.1. El contraste de White. Para simplificar la exposición, vamos a describir el contraste en una ecuación de regresión con término constante y dos variables explicativas. La extensión del contraste al modelo lineal general es trivial.

Consideramos pues el modelo

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

y deseamos contrastar las hipótesis

$$H_0 : E(u_i^2) = \sigma_u^2 \quad (Homocedasticidad)$$
$$H_1 : E(u_i^2) = \sigma_i^2 \quad (Heterocedasticidad)$$

Los pasos para realizar el contraste son los siguientes

1. Estimar por mínimos cuadrados ordinarios la ecuación de regresión de interés

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

   y obtener los residuos $\hat{u}_i$.

2. Estimar por mínimos cuadrados ordinarios la ecuación de regresión auxiliar

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2_i} X_{3i} + e_i$$

   y calcular el coeficiente de determinación R^2.

3. Calcular el estadístico de contraste nR^2, que sigue asintóticamente una distribución Chi-cuadrado con $p-1$ grados de libertad, donde p es el número de parámetros de la regresión auxiliar.

4. La hipótesis H_0 se rechaza al nivel de significación α, si $nR^2 > c$, donde c es el valor crítico para el cual $Prob(\chi^2_{p-1} > c) = \alpha$.

Suele decirse que el test de White es general porque no necesitamos conocer las variables que causan la heterocedasticidad. Sin embargo, de la ecuación de regresión auxiliar vemos que la forma de heterocedasticidad implícita es

$$\sigma_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2_i} X_{3i}$$

y las hipótesis a contrastar son

$$\begin{aligned} H_0 :&\alpha_2 = \cdots = \alpha_p = 0 \\ H_1 :&\alpha_i \neq 0 \text{ para algún } i = 2, \ldots, p \end{aligned}$$

Bajo H_0, la varianza del error es constante $\sigma_i^2 = \alpha_1$; bajo H_1, hay heterocedasticidad del tipo

$$\hat{\sigma}_i^2 = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2i} + \hat{\alpha}_3 X_{3i} + \hat{\alpha}_4 X_{2i}^2 + \hat{\alpha}_5 X_{3i}^2 + \hat{\alpha}_6 X_{2_i} X_{3i}$$

La regresión auxiliar incluye como variables explicativas todas las variables explicativas del modelo de interés, sus cuadrados y los productos cruzados, siempre que tales variables no sean redundantes, es decir, no aparezcan ya en la ecuación de regresión. Por ejemplo, si X_{2i} es una variable ficticia tomando los valores 0 y 1, entonces $X_{2i}^2 = X_{2i}$ y se dice que X_{2i}^2 es redundante.

En resumen, algunas ventajas del test de White son:

1. es un test general,
2. es un test constructivo, nos sugiere una forma de heterocedasticidad si se rechaza H_0,
3. es muy simple de aplicar.

y como inconvenientes

1. la ecuación de regresión auxiliar puede incluir muchas variables explicativas, $k(k+1)/2$.
2. la ecuación de regresión auxiliar no está exenta de los errores de especificación de cualquier regresión,
3. es un contraste asintótico, válido para muestras muy grandes.

11.2.3.2. El contraste de Breush-Pagan/Godfrey, BPG. Godfrey (1978) considera el modelo de regresión múltiple

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, \ldots, n$$

con heterocedasticidad del tipo

$$\sigma_i^2 = e^{(\alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_p Z_{pi})}$$

Esta ecuación indica que la varianza σ_i^2 es función exponencial de una combinación lineal de variables conocidas. En la práctica, las variables Z_j $(j = 2, \ldots, p)$ coinciden con las variables explicativas X_j $(j = 1, \ldots, k)$.

Breusch y Pagan (1979) consideran una forma más general de heterocedasticidad

$$\sigma_i^2 = h(\alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_p Z_{pi})$$

donde h es una función no especificada.

Las hipótesis a contrastar son

$$H_0 : \alpha_2 = \cdots = \alpha_p = 0 \quad \text{(homocedasticidad)}$$
$$H_1 : \alpha_i \neq 0 \text{ para algún } i = 2, \ldots, p \quad \text{(heterocedasticidad)}$$

Los pasos para realizar el contraste son los siguientes:

1. Estimar por mínimos cuadrados la ecuación de regresión de interés obtener los residuos $\hat{u}_i$, y calcular $\tilde{\sigma}_u^2 = \sum_{i=1}^n \hat{u}_i^2/n$
2. Estimar por mínimos cuadrados la ecuación de regresión auxiliar
$$\left(\frac{\hat{u}_i}{\tilde{\sigma}_u}\right)^2 = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_p Z_{pi} + e_i$$
y calcular la suma de cuadros explicada, SCE.
3. Calcular el estadístico de contraste $SCE/2$, que sigue bajo H_0 (y suponiendo normalidad) una distribución Chi-cuadrado con $p-1$ grados de libertad.
4. La hipótesis H_0 se rechaza al nivel de significación α, si $SCE/2 > c$, donde c es el valor crítico para el cual $Prob(\chi^2_{p-1} > c) = \alpha$

Un procedimiento equivalente es el siguiente

1. Estimar por mínimos cuadrados la ecuación de regresión de interés obtener los residuos $\hat{u}_i$.
2. Estimar por mínimos cuadrados la ecuación de regresión auxiliar
$$\hat{u}_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_p Z_{pi} + e_i$$
y calcular el coeficiente de determinación, R^2.
3. Calcular el estadístico de contraste nR^2, que sigue bajo H_0 (y suponiendo normalidad) una distribución Chi-cuadrado con $p-1$ grados de libertad.
4. La hipótesis H_0 se rechaza al nivel de significación α, si $nR^2 > c$, donde c es el valor crítico para el cual $Prob(\chi^2_{p-1} > c) = \alpha$

Veamos la lógica del contraste. Si H_0 es cierta, entonces la varianza σ_i^2 no depende de las variables Z_j $(j = 2, \ldots, p)$ y el R^2 en la regresión auxiliar será bajo. Por el contrario, si H_1 es cierta, entonces la varianza σ_i^2 depende de las variables Z_j $(j = 2, \ldots, p)$ y el R^2 será alto. El problema está en decidir cuando el R^2 se considera bajo y alto. Pues bien, será bajo cuando $nR^2 < c$ y alto cuando $nR^2 > c$.

Observe que cuando las variables que causan la heterocedasticidad, Z_j $(j = 2, \ldots, p)$ coinciden con las variables explicativas X_j $(j = 1, \ldots, k)$, la ecuación de regresión auxiliar de este contraste está contenida en la regresión auxiliar del test de White.

En resumen, algunas ventajas del test de BPG son:

1. es muy simple de aplicar,
2. no requiere conocer la forma funcional de la heterocedasticidad,
3. es un test constructivo, nos sugiere una forma de heterocedasticidad si se rechaza H_0,

y como inconvenientes

1. descansa en el supuesto de normalidad de los errores,
2. la ecuación de regresión auxiliar no está exenta de los errores de especificación de cualquier regresión,

Observación 67. En los exámenes, algunos estudiantes confunden los tests de autocorrelación y heterocedasticidad de Godfrey. Recordemos que en el primer caso nos interesa la dependencia del residuos $\hat{u}_t$ de su pasado $\hat{u}_{t-k}$.

11.2.4. El contraste de Goldfeld-Quandt. El contraste de Goldfeld-Quandt (1965) se aplica cuando sospechamos que la varianza del error aumenta con los valores de una variable conocida Z

$$\sigma_i^2 = \sigma_u^2 Z_i$$

El problema a contrastar lo podemos formular como

$$\begin{aligned} H_0 :& \sigma_i^2 = \sigma_u^2 \quad \text{(homocedasticidad)} \\ H_1 :& \sigma_i^2 = \sigma_u^2 Z_i \text{ para algún } i = 2, \ldots, p \quad \text{(heterocedasticidad)} \end{aligned}$$

Los pasos del contraste de Goldfeld-Quandt son los siguientes:

1. Identificar la variable que causa la heterocedasticidad, digamos Z.
2. Ordenar la tabla de datos según los valores crecientes de Z.
3. Dividir la tabla de datos en tres submuestras. La submuestra central con m observaciones, y las otras dos submuestras con $(n-m)/2$ observaciones.
4. Omitir las observaciones centrales, y estimar por mínimos cuadrados ordinarios la ecuación de regresión en cada submuestra.
5. Calcular la suma de cuadrados de los residuos en cada submuestra: SCR_1 y SCR_2. Si H_1 es cierta, entonces $SCR_2 > SCR_1$.
6. Calcular el estadístico de contraste
$$F = \frac{SCR_2}{SCR_1}$$
que sigue una distribución F con $(n-m)/2$ grados de libertad en el numerador y denominador.
7. La hipótesis H_0 se rechaza al nivel de significación α, si $F > c$, donde c es el valor crítico para el cual $Prob(F_{(n-m)/2,(n-m)/2} > c) = \alpha$

La elección del número de observaciones a omitir, m, juega un papel clave en el contraste. Cuanto mayor sea m, tanto mayor será la diferencia entre las sumas de cuadrados de los residuos SCR_1 y SCR_2 y más probable será rechazar H_0 si es falsa. El problema es que disminuyen los grados de libertad de cada regresión estimada, con lo cual el valor crítico c aumenta y menos probable es rechazar H_0 si es falsa. Por ejemplo, el valor crítico para $Prob(F_{4,4} > c) = 0{,}95$ es $c = 6{,}388$ y $Prob(F_{8,8} > c) = 0{,}95$ es $c = 3{,}438$. En la práctica, m suele elegirse igual a $n/3$.

Observación 68. Si σ_i^2 depende negativamente de los valores de Z_i, entonces el cociente SCR_1/SCR_2 debería ser mayor que uno, y deberíamos basar el contraste en esta ratio.

11.2.4.1. El contraste de razón de verosimilitudes. Cuando disponemos de una muestra grande, podemos formar g grupos de residuos y estimar en cada grupo la varianza residual $\hat{\sigma}_i^2 = \sum_{j \in i} \hat{u}_i/n_i$, en donde n_i es el número de residuos en el grupo i-ésimo. Se demuestra que bajo la hipótesis nula de homocedasticidad

$$[n_1 \log(\hat{\sigma}_1^2) + \cdots + n_g \log(\hat{\sigma}_g^2)] - n \log(\hat{\sigma}_u^2) \sim \chi^2_{g-1}$$

11.3. Tratamiento de la heterocedasticidad

Los remedios que vamos a ver para este problema dependen de si la heterocedasticidad es conocida o desconocida. En el primer caso, el método de estimación preferido será el de mínimos cuadrados generalizados; en el segundo, podemos optar por mínimos cuadrados generalizados fáctibles o mínimos cuadrados ordinarios.

11.3.1. Heterocedasticidad conocida. El modelo lineal general con heterocedasticidad

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, \ldots, n \tag{11.3}$$

en donde $E(u_i) = 0$, $E(u_i^2) = \sigma_u^2 \omega_i$ y $E(u_i u_j) = 0 \,\forall i \neq j$, es un caso especial del modelo con perturbaciones no esféricas. En este marco, el estimador lineal, insesgado, consistente y óptimo es el estimador de mínimos cuadrados generalizados

$$\hat{\boldsymbol{\beta}}_{MCG} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$$

cuya matriz de varianzas y covarianzas es

$$V(\hat{\boldsymbol{\beta}}_{MCG}) = \sigma_u^2(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}$$

en donde $\boldsymbol{\Omega} = diag(\omega_1, \ldots, \omega_n)$ es una matriz diagonal, cuya inversa es simplemente $\boldsymbol{\Omega}^{-1} = diag(1/\omega_1, \ldots, 1/\omega_n)$. El cálculo directo de $\hat{\boldsymbol{\beta}}_{MCG}$ y $V(\hat{\boldsymbol{\beta}}_{MCG})$ requiere crear la matriz $\boldsymbol{\Omega}^{-1}$ de orden $n \times n$. En la práctica, podemos evitar la creación de esta matriz usando el método de mínimos cuadrados ponderados.

11.3.1.1. Mínimos cuadrados ponderados. En el modelo (11.3), los errores cuasi-tipificados

$$u_i^* = \frac{u_i}{\sqrt{\omega_i}}$$

cumplen las hipótesis básicas:

1. Media cero
$$E(u_i^*) = E\left(\frac{u_i}{\sqrt{\omega_i}}\right) = \frac{E(u_i)}{\sqrt{\omega_i}} = 0$$
2. Varianza constante
$$E(u_i^{*2}) = E(\frac{u_i^2}{\omega_i}) = \frac{E(u_i^2)}{\omega_i} = \frac{\sigma_u^2\omega_i}{\omega_i} = \sigma_u^2$$
3. Covarianza cero
$$E(u_i^* u_j^*) = E\left(\frac{u_i u_j}{\sqrt{\omega_i}\sqrt{\omega_j}}\right) = \frac{E(u_i u_j)}{\omega_i \omega_j} = 0 \quad \forall i \neq j$$

De aquí, si transformamos el modelo de interés (11.3) dividiendo por la cuasi-desviación típica de los errores $\sqrt{w_i}$ obtenemos

$$\frac{Y_i}{\sqrt{\omega_i}} = \beta_1 \frac{1}{\sqrt{\omega_i}} + \beta_2 \frac{X_{2i}}{\sqrt{\omega_i}} + \cdots + \beta_k \frac{X_{ki}}{\sqrt{\omega_i}} + \frac{u_i}{\sqrt{\omega_i}}, \quad i = 1, \ldots, n$$

o bien

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \cdots + \beta_k X_{ki}^* + u_i^*, \quad i = 1, \ldots, n \tag{11.4}$$

El modelo transformado (11.4) contiene los mismos parámetros $\beta_1, \ldots, \beta_k$ que el modelo de interés (11.3), pero su error cumple las hipótesis básicas. Por tanto, el estimador de mínimos cuadrados ordinarios en (11.4) será lineal, insesgado y eficiente, que se denomina estimador de mínimos cuadrados ponderados. Podemos interpretar este método de estimación como un método en dos estapas: en una primera etapa, se transforman las variables dividiendo los datos por la cuasi-desviación típica de los errores y, en una segunda etapa, se estiman los parámetros por mínimos cuadrados ordinarios.

El método de mínimos cuadrados ponderados es una forma conveniente de obtener el estimador de mínimos cuadrados generalizados bajo heterocedasticidad. No debemos considerarlo como un método de estimación alternativo. Podemos comprobar la equivalencia entre los dos métodos usando la descomposición $\mathbf{\Omega}^{-1} = \mathbf{P}'\mathbf{P}$ y expresando el estimador MCG como

$$\hat{\boldsymbol{\beta}}_{MCG} = (\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{y}$$

en donde

$$\mathbf{P} = \begin{pmatrix} 1/\sqrt{\omega_1} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\omega_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \dots & 1/\sqrt{\omega_n} \end{pmatrix}$$

Ahora es claro que los datos transformados

$$\mathbf{Py} = \begin{pmatrix} y_1/\sqrt{\omega_1} \\ y_2/\sqrt{\omega_2} \\ \vdots \\ y_n/\sqrt{\omega_n} \end{pmatrix} \quad \text{y} \quad \mathbf{PX} = \begin{pmatrix} 1/\sqrt{\omega_1} & X_{21}/\sqrt{\omega_1} & \dots & X_{k1}/\sqrt{\omega_1} \\ 1/\sqrt{\omega_2} & X_{22}/\sqrt{\omega_2} & \dots & X_{k2}/\sqrt{\omega_2} \\ \vdots & \vdots & \ddots & \\ 1/\sqrt{\omega_n} & X_{2n}/\sqrt{\omega_n} & \dots & X_{kn}/\sqrt{\omega_n} \end{pmatrix}$$

se corresponden con los datos de las variables divididos por las cuasi-desviaciones típicas de los errores correspondientes.

Ejemplo 20. *Sea la ecuación de regresión simple con heterocedasticidad*

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

en donde $E(u_i) = 0$, $E(u_i^2) = \sigma_u^2 X_i^2$ *y* $E(u_i u_j) = 0 \, \forall i \neq j$. *El modelo transformado con errores homocedásticos se obtiene dividiendo la ecuación por* X_i

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

en donde $u_i^* = u_i/X_i$ *cumple las propiedades* $E(u_i^*) = 0$, $E(u_i^{*^2}) = \sigma_u^2$ *y* $E(u_i^* u_j^*) = 0 \, \forall i \neq j$. *Vemos que el modelo transformado contiene un término constante. Por tanto, los residuos* $\hat{u}_i^*$ *de la estimación de mínimos cuadrados tendrán media cero y el* R^2 *estará acotado entre 0 y 1.*

Ejemplo 21. *Sea la ecuación de regresión simple con heterocedasticidad*

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

en donde $E(u_i) = 0$, $E(u_i^2) = \sigma_u^2 X_i$ *y* $E(u_i u_j) = 0 \, \forall i \neq j$. *El modelo transformado con errores homocedásticos se obtiene dividiendo la ecuación por* $\sqrt{X_i}$

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

en donde $u_i^* = u_i/\sqrt{X_i}$ *cumple las propiedades* $E(u_i^*) = 0$, $E(u_i^{*^2}) = \sigma_u^2$ *y* $E(u_i^* u_j^*) = 0\,\forall i \neq j$.

Observamos que el modelo transformado no contiene un término constante. Por tanto, los residuos $\hat{u}_i^*$ *de la estimación de mínimos cuadrados no tendrán necesariamente media cero y el* R^2 *no estará necesariamente acotado entre 0 y 1.*

Ejemplo 22. *Sea la ecuación de regresión simple con heterocedasticidad*

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

en donde $E(u_i) = 0$, $E(u_i^2) = \sigma_u^2 e^{X_i}$ *y* $E(u_i u_j) = 0\,\forall i \neq j$. *El modelo transformado con errores homocedásticos se obtiene dividiendo la ecuación por* $\sqrt{e^{X_i}}$

$$\frac{Y_i}{\sqrt{e^{X_i}}} = \beta_1 \frac{1}{\sqrt{e^{X_i}}} + \beta_2 \frac{X_i}{\sqrt{e^{X_i}}} + u_i^*$$

en donde $u_i^* = u_i/\sqrt{e^{X_i}}$ *cumple las propiedades* $E(u_i^*) = 0$, $E(u_i^{*^2}) = \sigma_u^2$ *y* $E(u_i^* u_j^*) = 0\,\forall i \neq j$.

Para comprender porqué el método de estimación descrito se denomina método de mínimos cuadrados ponderados, hay que notar que la suma de cuadrados en el modelo transformado

$$\sum_{i=1}^n \hat{u}_i^{*^2} = \sum_{i=1}^n (Y_i^* - \hat{\beta}_1 X_{1i}^* + \hat{\beta}_2 X_{2i}^* + \cdots + \hat{\beta}_k X_{ki}^*)^2$$

puede escribirse como una suma ponderada de los cuadrados de los residuos

$$\sum_{i=1}^n \hat{u}_i^{*^2} = \sum_{i=1}^n \frac{1}{\omega_i} (Y_i - \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki})^2$$

en donde la ponderación asignada a cada residuo es inversamente proporcional a la varianza del error. De modo que, cuanto mayor sea la varianza tanto menor será el peso que se asigna a cada residuo, y viceversa.

11.3.2. Heterocedasticidad desconocida.

11.3.2.1. Mínimos cuadrados generalizados factibles. En la práctica cuando deseamos estimar una ecuación de regresión múltiple

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, \dots, n$$

desconocemos la forma de heterocedasticidad. Si sospechamos que

$$\sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_p Z_{pi}$$

podríamos aplicar el test de Breusch-Pagan-Godfrey. Si rechazamos la hipótesis nula de homocedasticidad, entonces debemos estimar por mínimos cuadrados generalizados. Ahora bien, como no conocemos las varianzas σ_i^2, tendremos estimarlas utilizando los valores ajustados obtenidos en la estimación de la regresión auxiliar

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_p Z_{pi} + e_i$$

esto es

$$\hat{\sigma}_i^2 = \hat{\alpha}_1 + \hat{\alpha}_2 Z_{2i} + \cdots + \hat{\alpha}_p Z_{pi}$$

Tipificando los datos, obtenemos el modelo transformado

$$\frac{Y_i}{\hat{\sigma}_i} = \beta_1 \frac{1}{\hat{\sigma}_i} + \beta_2 \frac{X_{2i}}{\hat{\sigma}_i} + \cdots + \beta_k \frac{X_{ki}}{\hat{\sigma}_i} + \frac{u_i}{\hat{\sigma}_i}, \quad i = 1, \ldots, n$$

en donde el estimador de mínimos cuadrados ordinarios se denomina mínimos cuadrados generalizados factibles (MCGF). El término factible enfatiza que la estimación se ha realizado reemplazando σ_i^2 por una estimación $\hat{\sigma}_i^2$.

Las propiedades estadísticas de este estimador en muestras finitas son desconocidas; mientras que en muestras grandes, dependen de las propiedades de $\hat{\sigma}_i^2$: si $\hat{\sigma}_i^2$ es un estimador consistente de σ_i^2, entonces el estimador MCGF también lo será.

11.3.2.2. Mínimos cuadrados ordinarios. Cuando la heterocedasticidad es desconocida, el estimador MCO cumple algunas propiedades deseables: es lineal, insesgado y consistente. El problema surge porque el estimador es ineficiente y, en consecuencia, los estadísticos t y F estarán sesgados. White demostró que la matriz de varianzas y covarianzas

$$V(\hat{\boldsymbol{\beta}}_{MCO}) = \sigma_u^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} \left(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}\right) \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

puede estimarse consistentemente reemplazando la matriz desconocida $\boldsymbol{\Omega}$ por una matriz diagonal que contiene los residuos al cuadrado de la estimación MCO, $\hat{\boldsymbol{\Omega}} = diag(\hat{u}_1^2, \ldots, \hat{u}_n^2)$. Algunos autores prefieren el estimador MCO al MCGF, y calculan los estadísticos t y F usando la matriz de varianzas y covarianzas consistente en presencia de heterocedasticidad desconocida.

11.4. Resumen

1. El supuesto de homocedasticidad en las perturbaciones debe cuestionarse cuando trabajamos con datos de sección cruzada.
2. La forma de heterocedasticidad más común es $\sigma_i^2 = \sigma_u^2 \omega_i$, en donde la varianza del error es proporcional a los valores de una variable ω_i.
3. El análisis gráfico de los residuos de un modelo estimado por mínimos cuadrados es útil para detectar la presencia y forma de heterocedasticidad en los errores.
4. Los contrastes de White y Godfrey son los métodos de deterción de heterocedasticidad más populares.
5. El procedimiento de míninimos cuadrados ponderados es un remedio simple al problema de heterocedasticidad conocida.
6. El estimador de MCO sigue siendo útil cuando los constrastes de hipótesis se basan en la matriz de varianzas y covarianzas consistente a la heterocedasticidad.

Palabras clave

Datos de sección cruzada
Homocedasticidad
Heterocedasticidad pura
Heterocedasticidad impura
Varianza local
Gráfico de dispersión
Mínimos cuadrados ponderados
Errores estándar consistentes ante heterocedasticidad

11.5. Ejercicios

1. Explique la diferencia entre heterocedasticidad pura e impura.

2. Describa el contraste de Goldfeld y Quandt.
3. Glejser propuso un test de heterocedasticidad basado en la regresión simple de $|\hat{u}_i|$ sobre Z_i^δ para $\delta = -1$, $-0{,}5$, $0{,}5$, 1. ¿Cómo debería ser la pendiente de Z_i^δ bajo homocedasticidad? ¿Y el R-cuadrado? Dibuje los diagramas de dispersión de $|\hat{u}_i|$ versus Z_i^δ para $\delta = -1$, $-0{,}5$, $0{,}5$, 1 en presencia de heterocedasticidad.
4. Considere la regresión de $Y_i \, (i = 1, \ldots, n)$ sobre un término constante. Demuestre que, en presencia de heterocedasticidad $E(u_i^2) = \sigma_u^2 \omega_i$, la varianza de la media $\bar{Y}$ es
$$Var(\bar{Y}) = \frac{\sigma_u^2}{n^2} \sum_{i=1}^n \omega_i$$
5. En el modelo de regresion simple con heterocedasticidad
$$Y_i = \beta_1 + \beta_2 X_i + u_i \, (i = 1, \ldots, n), \quad u_i \sim N(0, \sigma_i^2)$$
demuestre que
$$Var(\hat{\beta}_{2,MCO}) = \frac{\sum_{i=1}^n \omega_i x_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2}$$

11.6. Ejercicios resueltos

Ejercicio 5. *Diga si las siguientes afirmaciones son verdaderas o falsas, justificando sus respuestas:*

(a) *En presencia de heterocedasticidad, los estimadores MCO son sesgados e ineficientes.*
(b) *Si hay heterocedasticidad, las pruebas t y F no son válidas.*
(c) *En presencia de heterocedasticidad, el método MCO sobreestima siempre los errores estándar de los estimadores.*
(d) *Si los residuos de una regresión MCO muestran un patrón sistemático, esto significa que hay presencia de heterocedasticidad en los datos.*
(e) *No hay una prueba general de heterocedasticidad que esté libre de supuesto alguno sobre cuál de las variables está correlacionada con el término de error.*
(f) *Si el modelo de regresión está mal especificado (por ejemplo, se ha omitido una variable importante), los residuos MCO mostrarán un patrón claramente distinguible.*
(g) *Si un regresor que tiene varianza no constante se omite (incorrectamente) de un modelo, los residuos MCO serán heterocedásticos.*

Solución

a. *Falso, la presencia de heterocedasticidad en el término perturbación no afecta a la propiedad de insesgadez de las estimadores MCO, lo que si es cierto es que son ineficientes.*
b. *Son inválidas en el caso en el que el modelo ha sido estimado por MCO, pero si el modelo es estimado por Mínimos Cuadrados Ponderados las pruebas son válidas.*

c. *Correcto, el método MCO en presencia de heterocedasticidad es menos eficiente que MCG, ya que:*

$$V(\hat{\beta}_{MCO}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$$
$$V(\hat{\beta}_{MCG}) = \sigma^2(X'\Omega^{-1}X)^{-1}$$

y aplicando el teorema de Aitken tenemos que la estimación MCG es más eficiente que MCO.

d. *Correcto, un patrón sistemático en los residuos indica heterocedasticidad en los datos provocada por la omisión de alguna variable relevante en la explicación de la variable dependiente Y.*

e. *Falso, en el contraste de WHITE no hay ningún supuesto sobre la forma de la heterocedasticidad en la especificación de la hipótesis alternativa.*

f. *Correcto, si se realiza un gráfico entre los residuos MCO al cuadrado y la variable que hemos omitido observaremos un patrón muy claro de relación entre ambos, lo que nos dará una idea de la transformación a realizar en las variables del modelo.*

g. *Correcto, supongamos que el modelo correcto es, $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$, pero incorrectamente se estima: $Y_i = \beta_1 + \beta_2 X_{2i} + V_i$, el término perturbación $V_i = \beta_3 X_{3i} + U_i$ recogería el comportamiento de la variable omitida y los residuos MCO recogerán el comportamiento obviado en la especificación del modelo.*

Ejercicio 6. *Suponga el siguiente modelo de regresión:*

$$y_i = \mu + u_i \qquad donde \qquad E(u_i) = 0 \quad y \quad Var(u_i) = \sigma^2 x_i^2$$

(a) *Hallar el estimador MCO de μ y calcular su varianza.*

(b) *¿Existe un estimador más eficiente de μ? En caso de que exista, halle su expresión y la de su varianza.*

Solución

a. *La estimación por MCO será la siguiente:*

$$SCR_{MCO} = \sum_{i=1}^{N} \widehat{u}_i^2 = \sum_{i=1}^{N} (y_i - \widehat{\mu}_{MCO})^2$$

Derivando con respecto a $\widehat{\mu}_{MCO}$ obtenemos su estimador MCO:

$$\frac{\partial SCR_{MCO}}{\partial \widehat{\mu}_{MCO}} = -2\sum_{i=1}^{N}(y_i - \widehat{\mu}_{MCO}) = 0 \Rightarrow \widehat{\mu}_{MCO} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

La esperanza y la varianza de este estimador serán las siguientes:

$$\widehat{\mu}_{MCO} = \frac{1}{N}\sum_{i=1}^{N}(\mu + u_i) = \mu + \frac{1}{N}\sum_{i=1}^{N} u_i$$

$$E(\widehat{\mu}_{MCO}) = E(\mu + \frac{1}{N}\sum_{i=1}^{N} u_i) = \mu + \frac{1}{N}\sum_{i=1}^{N} E(u_i) = \mu$$

$$V(\widehat{\mu}_{MCO}) = E[\widehat{\mu}_{MCO} - E(\widehat{\mu}_{MCO})]^2 = E[\mu + \frac{1}{N}\sum_{i=1}^{N} u_i - \mu]^2 = \frac{1}{N^2}\sum_{i=1}^{N} E(u_i^2)$$

$$V(\widehat{\mu}_{MCO})\frac{\sigma^2}{N^2}\sum_{i=1}^{N} x_i^2$$

b. *El estimador del apartado anterior no es el más eficiente dada la presencia de heterocedasticidad, por lo que habrá que estimar el modelo por MCG, para lo que debemos transformarlo de la siguiente manera:*

$$\frac{y_i}{x_i} = \mu\frac{1}{x_i} + \frac{u_i}{x_i}$$

Al igual que en el apartado anterior, obtenemos el estimador de μ de la siguiente manera:

$$SCR_{MCG} = \sum_{i=1}^{N}(\frac{\widehat{u}_i}{x_i})^2 = \sum_{i=1}^{N}(\frac{y_i}{x_i} - \widehat{\mu}_{MCG}\frac{1}{x_i})^2$$

$$\frac{\partial SCR_{MCG}}{\partial \widehat{\mu}_{MCG}} = -2[\sum_{i=1}^{N}(\frac{y_i}{x_i} - \widehat{\mu}_{MCG}\frac{1}{x_i})]\frac{1}{x_i} = 0 \Rightarrow \widehat{\mu}_{MCG} = \frac{\sum_{i=1}^{N}(y_i/x_i^2)}{\sum_{i=1}^{N}(1/x_i^2)}$$

Calculamos ahora la esperanza y la varianza de $\widehat{\mu}_{MCG}$:

$$\widehat{\mu}_{MCG} = \frac{\sum_{i=1}^{N}(\frac{\mu+u_i}{x_i^2})}{\sum_{i=1}^{N}(\frac{1}{x_i^2})} = \mu + \frac{\sum_{i=1}^{N}(u_i/x_i^2)}{\sum_{i=1}^{N}(1/x_i^2)}$$

$$E(\widehat{\mu}_{MCG}) = \mu + \frac{\sum_{i=1}^{N}(E(u_i)/x_i^2)}{\sum_{i=1}^{N}(1/x_i^2)} = \mu$$

$$V(\widehat{\mu}_{MCG}) = E[\widehat{\mu}_{MCG} - E(\widehat{\mu}_{MCG})]^2 = E[\mu + \frac{\sum_{i=1}^{N}(u_i/x_i^2)}{\sum_{i=1}^{N}(1/x_i^2)} - \mu]^2 =$$

$$= \frac{\sum_{i=1}^{N}(E(u_i^2)/x_i^4)}{(\sum_{i=1}^{N}(1/x_i^2))^2} = \frac{\sigma^2\sum_{i=1}^{N}(1/x_i^2)}{(\sum_{i=1}^{N}(1/x_i^2))^2}$$

$$V(\widehat{\mu}_{MCG}) = \frac{\sigma^2}{\sum_{i=1}^{N}(1/x_i^2)}$$

Comparando ambas varianzas:

$$V(\widehat{\mu}_{MCO}) = \frac{\sigma^2}{\sum(1/x_i^2)}\sum(\frac{1}{x_i^2})\frac{\sum x_i^2}{N^2} > \frac{\sigma^2}{\sum(1/x_i^2)} = V(\widehat{\mu}_{MCG})$$

EJERCICIO 7. *Dado el modelo $y_t = \alpha + \beta x_t + u_t$ donde $E(u_t) = 0$, $E(u_t^2) = \sigma_t^2$ y $E(u_t, u_s) = 0 \quad t \neq s$ Se propone la siguiente estimación para término constante:*

$$\widehat{\alpha}_{MCG} = \frac{1}{T}(\sum_{t=1}^{T}\frac{y_t}{\sigma_t} - \widehat{\beta}_{MCG}\sum_{i=1}^{T}\frac{x_t}{\sigma_t})$$

¿Es correcta esta expresión? Justifique su respuesta.

Solución

Bajo la presencia de heterocedasticidad, el estimador insesgado más eficiente es el de MCG. Para estimarlo, debemos transformar el modelo de la siguiente manera:

$$\frac{y_t}{\sigma_t} = \alpha\frac{1}{\sigma_t} + \beta\frac{x_t}{\sigma_t} + \frac{u_t}{\sigma_t} \Rightarrow y_T^* = \alpha_{MCG}x_{1t}^* + \beta_{MCG}x_{2t}^* + v_t$$

Obtenemos el estimador MCG de α de la siguiente manera:

$$SCR_{MCG} = \sum_{t=1}^{T}(\widehat{v}_t)^2 = \sum_{t=1}^{T}(y_t^* - \widehat{\alpha}_{MCG}x_{1t}^* - \widehat{\beta}_{MCG}x_{2t}^*)^2$$

$$\frac{\partial SCR_{MCG}}{\partial\widehat{\alpha}_{MCG}} = -2\sum_{t=1}^{T}(y_t^* - \widehat{\alpha}_{MCG}x_{1t}^* - \widehat{\beta}_{MCG}x_{2t}^*)x_{1t}^* = 0$$

$$\sum_{t=1}^{T}y_t^*x_{1t}^* - \widehat{\alpha}_{MCG}\sum_{t=1}^{T}x_{1t}^{*2} - \widehat{\beta}_{MCG}\sum_{t=1}^{T}x_{2t}^*x_{1t}^* = 0$$

$$\widehat{\alpha}_{MCG} = \frac{1}{\sum_{t=1}^{T}x_{1t}^{*2}}\sum_{t=1}^{T}(y_t^*x_{1t}^* - \widehat{\beta}_{MCG}x_{2t}^*x_{1t}^*) = \frac{1}{\sum_{t=1}^{T}(1/\sigma_t)^2}\sum_{t=1}^{T}(\frac{y_t}{\sigma_t^2} - \widehat{\beta}_{MCG}\frac{x_t}{\sigma_t^2})$$

Por lo tanto, el estimador MCO propuesto no está correctamente calculado.

EJERCICIO 8. *Considere el modelo $y_t = \alpha + \beta x_t + u_t$, donde $E(u_t) = 0$ y $Var(u_t) = \sigma_t^2 = \sigma^2 x_t^2$, del que se tienen las siguientes observaciones:*

t	y_t	x_t
1	4	2
2	2	1
3	1	4
4	5	2
5	3	1

(a) *Obtenga los estimadores MCO de α y β, así como su matriz de covarianzas. ¿Existe algún problema en la definición de dichos estimadores?*

(b) *Obtenga los estimadores MCG de α y β, así como su matriz de covarianzas. ¿Existe algún problema en la definición de dichos estimadores? Compare su matriz de covarianzas con la del estimador MCO.*

Solución

a. $\hat{\beta}_{MCO} = (X'X)^{-1}X'y = \begin{pmatrix} 5 & 10 \\ 10 & 26 \end{pmatrix}^{-1}\begin{pmatrix} 15 \\ 27 \end{pmatrix} = \begin{pmatrix} 4 \\ -0,5 \end{pmatrix}$

$V(\hat{\beta}_{MCO}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} =$

$\sigma^2\begin{pmatrix} 5 & 10 \\ 10 & 26 \end{pmatrix}^{-1}\begin{pmatrix} 26 & 82 \\ 82 & 290 \end{pmatrix}\begin{pmatrix} 5 & 10 \\ 10 & 26 \end{pmatrix}^{-1} =$

$\sigma^2\begin{pmatrix} 4,3733 & -2,6666 \\ -2,6666 & 1,8333 \end{pmatrix}$

El único problema es que no se verifican todas las hipótesis que garantizan que el estimador es ELIO dentro de los estimadores lineales e insesgados.

b. $\hat{\beta}_{MCG} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = \begin{pmatrix} 2,5625 & 3,25 \\ 3,25 & 5 \end{pmatrix}^{-1}\begin{pmatrix} 7,3125 \\ 9,75 \end{pmatrix} = \begin{pmatrix} 2,1666 \\ 0,541667 \end{pmatrix}$

$$V(\hat{\beta}_{MCG}) = \sigma^2(X'\Omega^{-1}X)^{-1} = \sigma^2 \begin{pmatrix} 2,5625 & 3,25 \\ 3,25 & 5 \end{pmatrix}^{-1} = \sigma^2 \begin{pmatrix} 2,2222 & -1,4444 \\ -1,4444 & 1,1388 \end{pmatrix}$$

La estimación por MCG proporciona estimaciones más eficientes que la estimación por MCO, ya que tenemos que las varianzas de los estimadores son más pequeñas.

EJERCICIO 9. *Encuentre una expresión analítica general para la matriz de varianzas y covarianzas de los estimadores de MCO y MCG en el modelo:*

$$y_t = \beta_1 + \beta_2 x_t + u_t \qquad E(u_t) = 0 \qquad Var(u_t) = kx_t$$

donde k es una constante. ¿Qué peculiaridad presentarían ambas matrices en el caso en que se tuviese que $\sum_{i=1}^{T} x_t = 0$

Solución

Expresión analítica general para la matriz de varianzas y covarianzas del estimador MCO en el modelo $y_t = \beta_1 + \beta_2 x_{2t} + u_t \quad Var(u_t) = kx_t$

$V(\hat{\beta}_{MCO}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$ *vamos a ver la expresión que tiene cada uno de los componentes de la matriz de varianzas y covarianzas del estimador MCO,*
$(X'X)^{-1} = \frac{1}{(T\sum_{t=1}^{T} x_t^2 - (\sum_{t=1}^{T} x_t)^2)} \begin{pmatrix} \sum_{t=1}^{T} x_t^2 & -\sum_{t=1}^{T} x_t \\ -\sum_{t=1}^{T} x_t & T \end{pmatrix}$,

$X'\Omega X = \begin{pmatrix} \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} x_t^2 \\ \sum_{t=1}^{T} x_t^2 & \sum_{t=1}^{T} x_t^3 \end{pmatrix}$,

Si hacemos el producto entre esas matrices obtenemos la siguiente expresión para $V(\hat{\beta}_{MCO})$: $V(\hat{\beta}_{MCO}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} =$

$$\left(\frac{k}{(T\sum_t x_t^2 - (\sum_t x_t)^2)^2}\right) *$$
$$\begin{pmatrix} -\sum_t x_t[(\sum_t x_t^2)^2 - \sum_t x_t \sum_t x_t^3] & T[(\sum_t x_t^2)^2 - \sum_t x_t \sum_t x_t^3] \\ T[(\sum_t x_t^2)^2 - \sum_t x_t \sum_t x_t^3] & \sum_t x_t(\sum_t x_t)^2 - 2T\sum_t x_t^2] + T^2 \sum_t x_t^3] \end{pmatrix}$$

Matriz de varianzas y covarianzas del estimador MCG:

$$V(\hat{\beta}_{MCG}) = k(X'\Omega^{-1}X)^{-1} = \frac{k}{\sum_t \frac{1}{x_t} \sum_t x_t - T^2} \begin{pmatrix} \sum_t x_t & -T \\ -T & \sum_t \frac{1}{x_t} \end{pmatrix}$$

EJERCICIO 10. *Suponga que el modelo de regresión lineal simple:*

$$y_{ig} = \beta x_{ig} + u_{ig} \qquad E(u_{ig}) = 0 \qquad Var(u_{ig}) = \sigma_u^2$$

es válido para cada observación i de cada uno de los cinco subgrupos en que se ha dividido la muestra $g = 1, 2, 3, 4, 5$, con $n_1 + n_2 + n_3 + n_4 + n_5 = 100$.

Suponga asimismo que se observan tan sólo las medias muestrales de cada grupo:

Grupo	*1*	*2*	*3*	*4*	*5*
n_g	*10*	*20*	*15*	*30*	*25*
$\bar{y}_g$	*8*	*16*	*12*	*4*	*20*
$\bar{x}_g$	*2*	*8*	*4*	*2*	*12*

Se proponen tres estimadores de β:

$$\widehat{\beta}_1 = \frac{\bar{y}}{\bar{x}} = \frac{\sum_g n_g \bar{y}_g / \sum_g n_g}{\sum_g n_g \bar{x}_g / \sum_g n_g}$$

$$\widehat{\beta}_2 = \frac{\sum_g \bar{y}_g \bar{x}_g}{\sum_g \bar{x}_g^2}$$

$$\widehat{\beta}_3 = \frac{\sum_g n_g \bar{x}_g \bar{y}_g}{\sum_g n_g \bar{x}_g^2}$$

(a) *Pruebe que los tres estimadores son insesgados.*

(b) *Obtenga la varianza de cada estimador.*

(c) *Calcule numéricamente cada estimador y su varianza suponiendo que $\sigma_u^2 = 1$. ¿Cuáles son los estimadores más eficientes?*

(d) *Demuestre que la varianza del tercer estimador es siempre menor que la del segundo.*

Solución

Antes de calcular las esperanzas y varianzas de los tres estimadores propuestos, los expresamos en función de la perturbación aleatoria:

$$\widehat{\beta}_1 = \frac{\sum_g n_g \bar{y}_g}{\sum_g n_g \bar{x}_g} = \frac{\sum_g n_g(\beta \bar{x}_g + \bar{u}_g)}{\sum_g n_g \bar{x}_g} = \beta + \frac{\sum_g n_g \bar{u}_g}{\sum_g n_g \bar{x}_g}$$

$$\widehat{\beta}_2 = \frac{\sum_g \bar{y}_g \bar{x}_g}{\sum_g \bar{x}_g^2} = \frac{\sum_g (\beta \bar{x}_g + \bar{u}_g)\bar{x}_g}{\sum_g \bar{x}_g^2} = \beta + \frac{\sum_g \bar{u}_g \bar{x}_g}{\sum_g \bar{x}_g^2}$$

$$\widehat{\beta}_3 = \frac{\sum_g n_g \bar{x}_g \bar{y}_g}{\sum_g n_g \bar{x}_g^2} = \frac{\sum_g n_g \bar{x}_g(\beta \bar{x}_g + \bar{u}_g)}{\sum_g n_g \bar{x}_g^2} = \beta + \frac{\sum_g n_g \bar{x}_g \bar{u}_g}{\sum_g n_g \bar{x}_g^2}$$

a. *Las esperanzas de los estimadores son las siguientes:*

$$E(\widehat{\beta}_1) = E[\beta + \frac{\sum_g n_g \frac{1}{n_g} \sum_{i=1}^{n_g} u_{ig}}{\sum_g n_g \bar{x}_g}] = \beta + \frac{\sum_g \sum_{i=1}^{n_g} E(u_{ig})}{\sum_g n_g \bar{x}_g} = \beta$$

$$E(\widehat{\beta}_2) = E[\beta + \frac{\sum_g \bar{x}_g \frac{1}{n_g} \sum_{i=1}^{n_g} u_{ig}}{\sum_g \bar{x}_g^2}] = \beta + \frac{\sum_g \bar{x}_g \frac{1}{n_g} \sum_{i=1}^{n_g} E(u_{ig})}{\sum_g \bar{x}_g^2} = \beta$$

$$E(\widehat{\beta}_3) = E[\beta + \frac{\sum_g n_g \bar{x}_g \frac{1}{n_g} \sum_{i=1}^{n_g} u_{ig}}{\sum_g n_g \bar{x}_g^2}] = \beta + \frac{\sum_g \bar{x}_g \sum_{i=1}^{n_g} E(u_{ig})}{\sum_g n_g \bar{x}_g^2} = \beta$$

b. *Las varianzas de los estimadores serán:*

$$V(\widehat{\beta}_1) = E[\widehat{\beta}_1 - E(\widehat{\beta}_1)]^2 = E[\beta + \frac{\sum_g \sum_{i=1}^{n_g} u_{ig}}{\sum_g n_g \bar{x}_g} - \beta]^2 = \frac{N\sigma_u^2}{(\sum_g n_g \bar{x}_g)^2}$$

$$V(\widehat{\beta}_2) = E[\widehat{\beta}_2 - E(\widehat{\beta}_2)]^2 = E[\beta + \frac{\sum_g \bar{x}_g \frac{1}{n_g} \sum_{i=1}^{n_g} u_{ig}}{\sum_g \bar{x}_g^2} - \beta]^2 = \frac{\sigma_u^2 \sum_g \frac{1}{n_g} \bar{x}_g^2}{(\sum_g \bar{x}_g^2)^2}$$

$$V(\widehat{\beta}_3) = E[\widehat{\beta}_3 - E(\widehat{\beta}_3)]^2 = E[\beta + \frac{\sum_g n_g \bar{x}_g \frac{1}{n_g} \sum_{i=1}^{n_g} u_{ig}}{\sum_g n_g \bar{x}_g^2} - \beta]^2 = \frac{\sigma_u^2}{\sum_g n_g \bar{x}_g^2}$$

c. *Los valores numéricos de los estimadores y sus varianzas son:*

$$\widehat{\beta}_1 = 2 \qquad \widehat{\beta}_2 = 1,90 \qquad \widehat{\beta}_3 = 1,83$$

$$V(\widehat{\beta}_1) = 0,00028 \qquad V(\widehat{\beta}_2) = 0,00020 \qquad V(\widehat{\beta}_3) = 0,00019$$

Los estimadores más eficientes (con menor varianza) son $\widehat{\beta}_2$ y $\widehat{\beta}_3$.

d. *Vamos a obtener el estimador más eficiente entre los lineales e insesgados que, dado el problema de heterocedasticidad por grupos que presenta el modelo propuesto, corresponde con el de MCG.*

En primer lugar, derivamos las propiedades del término de error del modelo:

$$E(\bar{u}_g) = E(\frac{1}{n_g}\sum_{i=1}^{n_g} u_{ig}) = \frac{1}{n_g}\sum_{i=1}^{n_g} E(u_{ig}) = 0$$

$$V(\bar{u}_g) = E[\bar{u}_g]^2 = E(\frac{1}{n_g^2}\sum_{i=1}^{n_g} u_{ig}^2) = \frac{\sigma_u^2}{n_g}$$

Por lo tanto, hemos demostrado cómo al considerar los datos en medias, estamos introduciendo heterocedasticidad en el modelo. Para obtener la estimación MCG de β transformamos el modelo de la siguiente manera:

$$\sqrt{n_g}\bar{y}_g = \beta\sqrt{n_g}\bar{x}_g + \sqrt{n_g}\bar{u}_g \Rightarrow y_g^* = \beta x_g^* + v_g$$

Estimamos ahora el modelo transformado por Mínimos Cuadrados:

$$SCR_{MCG} = \sum_g (\widehat{v}_g)^2 = \sum_g (y_g^* - \widehat{\beta}_{MCG} x_g^*)^2$$

$$\frac{\partial SCR_{MCG}}{\partial \widehat{\beta}_{MCG}} = -2\sum_g (y_g^* - \widehat{\beta}_{MCG} x_g^*) x_g^* = 0$$

$$\widehat{\beta}_{MCG} = \frac{\sum_g y_g^* x_g^*}{\sum_g x_g^{*2}} = \frac{\sum_g n_g \bar{y}_g \bar{x}_g}{\sum_g n_g \bar{x}_g^2}$$

Como vemos, el estimador MCG de β coincide con el estimador $\widehat{\beta}_3$ propuesto en el ejercicio. Por lo tanto, éste último siempre tendrá menor varianza que cualquier otro, por ser, por definición el estimador insesgado más eficiente.