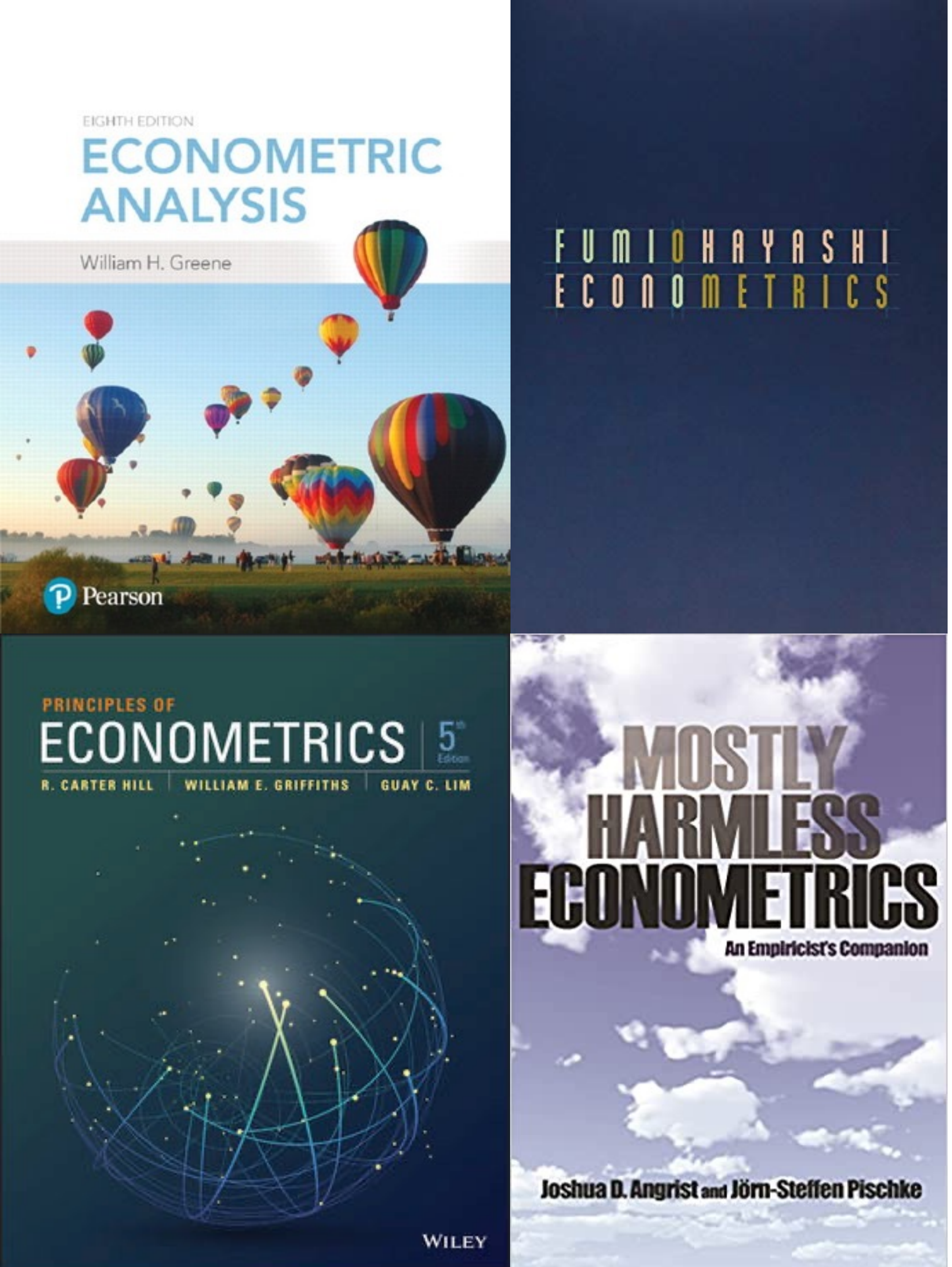
PUCP



MAESTRÍA EN ECONOMÍA
ECONOMETRÍA INTERMEDIA
ECO743 – MÓDULO 2

Sesión 1
Perturbaciones No Esféricas - Heterocedasticidad

Docente: Juan Palomino

PUCP

Índice

1. Supuestos MCO

Supuestos MCO

Supuesto 1: Linealidad

La relación entre la variable dependiente y las variables independientes es lineal. El modelo poblacional es:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_K x_{iK} + \varepsilon_i \qquad i = 1, \ldots, n$$

Donde β's son parámetros desconocidos a ser estimados, y ε_i son términos de errores observados.

Nota: La propiedad de linealidad es una propiedad de los parámetros, no de las variables.

Supuestos MCO

Supuesto 2: Exogeneidad Estricta

El valor esperado de la perturbación aleatoria debe ser cero para cualquier observación:

$$E[\varepsilon_i|X] = 0 \qquad i = 1,2,\dots,n$$

Para cualquier valor de $x_1, x_2, \dots, x_K$ en la población, el promedio no observable es igual a cero.

Supuestos MCO

Consecuencias de Exogeneidad Estricta:

Definición: Ley de Esperanzas Iteradas

Si $E|y| < \infty$, es decir, la esperanza poblacional de y existe, entonces para cualquier vector x aleatorio:

$$E_X[E(y|x)] = E(y)$$

1. Media Incondicional

La media incondicional del término de error es cero:

$$E(\varepsilon_i) = 0$$

Consecuencias de Exogeneidad Estricta:

2. Regresores son ortogonales al término de error

Si $E(xy)$ de dos variables aleatorias x e y es cero, entonces decimos que x es ortogonal a y. Bajo exogeneidad estricta, los regresores son ortogonales al término de error para todas las observaciones, es decir, no comparten información.

$$E\left(x_{jk}\varepsilon_i\right) = 0 \qquad i, j = 1,2, \dots, n \qquad k = 1, \dots, K$$

o

$$E(x \cdot \varepsilon_i) = \begin{pmatrix} E(x_{j1}\varepsilon_i) \\ E(x_{j2}\varepsilon_i) \\ \vdots \\ E(x_{jK}\varepsilon_i) \end{pmatrix} = \underbrace{0}_{K\times 1} \qquad \forall i, j$$

Supuestos MCO

Consecuencias de Exogeneidad Estricta:

3. Condición de cero correlación

El regresos no está correlacionado con el término de error:

$$\text{Cov}\left(\varepsilon_i, x_{jk}\right) = 0$$

4. Media condicional de $\boldsymbol{y_i}$

La media condicional de la variable dependiente es una función lineal de los regresores.

Bajo el supuesto de linealidad y exogeneidad estrictra:

$$E(y_i|X) = x_i'\beta \qquad\qquad i = 1,2,\ldots,n$$

Supuestos MCO

Supuesto 3: Rango Completo

Asume que ninguna de las variables en el modelo es una combinación lineal exacta de las otras.

$$X \text{ } es \text{ } una \text{ } matriz \text{ } n \times K \text{ } con \text{ } rango \text{ } K$$

Significa que X tiene rango de columna completo.

Nota: El rango de una matriz es igual al número de columnas linealmente independientes de la matriz. También se le conoce como **condición de identificación.**

Supuesto 4: Perturbaciones Esféricas

❖ Homocedasticidad

Asume que todas las unidades tienen el mismo error de varianza

$$E\left[\varepsilon_i^2 \middle| X\right] = \sigma_0^2 > 0 \qquad i = 1,2,\dots,n$$

❖ No Correlación

Asume también que no hay correlación entre las observaciones

$$E\left[\varepsilon_i \varepsilon_j \middle| X\right] = 0 \qquad i,j = 1,2,\dots,n \qquad i \neq j$$

Supuestos MCO

Supuesto 5: Regresores no estocásticos

Las observaciones de x_i son fijos en muestras repetidas.

Asumir que los X son fijos quiere decir que, en repetidas muestras de X, los valores obtenidos $x_1, x_2, \ldots, x_K$ van a ser siempre los mismos; es decir, dejan de ser aleatorios.

Bajo este supuesto, ya no es necesario hablar de esperanzas condicionales:

$$E(\varepsilon_i) = 0$$
$$\mathrm{Var}(\varepsilon_i) = \sigma^2$$
$$\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$$

Supuesto 6: Normalidad de los errores

ε_i distribuye normal con media cero y varianza σ^2 condicional a X:

$$\varepsilon_i | X \sim N(0, \sigma^2), \qquad i = 1,2, \ldots, n$$

2. Propiedades Estadísticas de MCO

Insesgadez del Estimador de MCO

Estimador Insesgado

Un estimador $\hat{\theta}$ para θ_0 es insesgado si $E(\hat{\theta}) = \theta_0$

Teorema 1: Insesgadez del Estimador MCO

Bajo el supuesto de Linealidad, Exogeneidad Estricta y Rango Completo:

$$E(\hat{\beta}|X) = \beta_0$$

Entonces, $\hat{\beta}$ es un estimador insesgado de β_0

Insesgadez del Estimador de MCO

Ejemplo:

Tenemos el siguiente modelo:

$$bmi_i = \alpha_0 + \beta_0 income_i + \varepsilon_i$$

Donde β_0 es el efecto del ingreso sobre obesidad y es igual a 1.

El siguiente paso es conseguir una muestra de la población y obtener un estimado de β_0, llamado $\hat{\beta}$.

Imaginen lo siguiente: $B = 50$ muestras de 1000 individuos y para cada muestra estimamos β.

1era muestra → $\hat{\beta}_1 = 0.9$

2da muestra → $\hat{\beta}_1 = 0.94$

3era muestra → $\hat{\beta}_1 = 0.7$

50ava muestra → $\hat{\beta}_1 = 0.5$

En promedio de las 50 muestras, conseguir: $(\frac{1}{B})\sum_{b=1}^{B} \hat{\beta}_b \approx 1$

Insesgadez del Estimador de MCO

Demostración

Tenemos:

$$\hat{\beta} = (X'X)^{-1}X'y$$

$$\hat{\beta} = (X'X)^{-1}X'(X\beta_0 + \varepsilon)$$

$$\hat{\beta} = (X'X)^{-1}X'X\beta_0 + (X'X)^{-1}X'\varepsilon$$

$$\hat{\beta} = \beta_0 + (X'X)^{-1}X'\varepsilon$$

$$\hat{\beta} - \beta_0 = A\varepsilon$$

Donde $A = (X'X)^{-1}X'$. Se le conoce como **sampling error**.

Varianza del Estimador MCO

Teorema 2: Varianza del estimador MCO

Bajo el supuesto de Linealidad, Exogeneidad Estricta, Rango Completo y Errores Esféricos:

$$\mathrm{Var}(\hat{\beta}_{OLS}) = \sigma_0^2 (X'X)^{-1}$$

Entonces, $\hat{\beta}$ es un estimador insesgado de β_0

Teorema de Gauss Markov

Teorema 3: Gauss-Markov

Bajo todos los supuestos, el estimador MCO es eficiente en la clase de estimadores lineales insesgados. Es decir, para cualquier estimador insesgado $\tilde{\beta}$ que es lineal en y,

$$\mathrm{Var}(\tilde{\beta}) \geq \mathrm{Var}(\hat{\beta}_{OLS}|X)$$

Este teorema señala que el estimador MCO es eficiente en el sentido que su matriz de varianza-covarianza $Var(\hat{\beta}_{OLS}|X)$ es el más pequeño entre todos los estimadores insesgados. Por esta razón, al estimador MCO se le conoce como Mejor Estimador Lineal Insesgado (MELI) [Best Lineal Unbiased Estimator (BLUE)].

Covarianza entre $\widehat{\beta}$ y $\hat{\varepsilon}$

Teorema 4: Covarianza entre $\widehat{\beta}$ y $\hat{\varepsilon}$

Bajo todos los supuestos de Linealidad, Exogeneidad Estricta, Rango Completo, y Errores Esféricos:

$$Cov(\hat{\beta}, \hat{\varepsilon} | X) = 0$$

Donde $\hat{\varepsilon} \equiv y - X\widehat{\beta}$

Covarianza entre $\widehat{\beta}$ y $\hat{\varepsilon}$

Demostración

$$
\begin{aligned}
Cov(\hat{\beta},\hat{\varepsilon}|X) &= E\{[(\hat{\beta}-E(\hat{\beta}|X))][\hat{\varepsilon}-E(\hat{\varepsilon}|X)]'|X\} \\
&= E\{[(\hat{\beta}-\beta_0|X)][M\varepsilon-E(M\varepsilon|X)]'|X\} \\
&= E[A\varepsilon\varepsilon'M|X] \\
&= AE[\varepsilon\varepsilon'|X]M' \\
&= AE[\sigma_0^2 I_n]M \\
&= \sigma_0^2(X'X)^{-1}X'M \\
&= \sigma_0^2(X'X)^{-1}\underbrace{(MX)'}_{0} \\
&= 0
\end{aligned}
$$

Teorema 5: Insesgadez de $\hat{\sigma}^2$

Bajo todos los supuestos de Linealidad, Exogeneidad Estricta, Rango Completo, y Errores Esféricos:

$$\mathrm{E}(\hat{\sigma}^2|X) = \sigma_0^2$$

Siempre que $n > K$, de modo que $\hat{\sigma}^2$ está bien definido.

Insesgadez de $\hat{\sigma}^2$

Demostración

Ya que $\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(n-K)}$, la prueba equivale a mostrar que $E(\hat{\varepsilon}'\hat{\varepsilon}|X) = (n-K)\sigma_0^2$. Sabemos que $\hat{\varepsilon}'\hat{\varepsilon} = \varepsilon M \varepsilon$, donde M es la matriz aniquiladora. La demostración consiste en probar dos propiedades: (1) $E(\varepsilon M \varepsilon|X) = \sigma_0^2 \cdot traza(M)$, y (2) $traza(M) = n - K$.

1. Probar que$E(\varepsilon M \varepsilon|X) = \sigma_0^2 \cdot traza(M)$: ya que $\varepsilon' M \varepsilon = \sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij}\, \epsilon_i \epsilon_j$, tenemos:

$$E(\varepsilon M \varepsilon|X) = \sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij} E\left(\epsilon_i \epsilon_j \middle| X\right)$$

Ya que m_{ij}'s son funciones de X

$$E(\varepsilon M \varepsilon|X) = \sum_{j=1}^{n} m_{ii}\, \sigma^2$$

Ya que $E\left(\varepsilon_i \varepsilon_j \middle| \mathrm{X}\right) = 0$ para $i \neq j$ por supuesto 4

$$E(\varepsilon M \varepsilon|X) = \sigma^2 \sum_{j=1}^{n} m_{ii} = \sigma^2 \cdot \mathrm{traza(M)}$$

Insesgadez de $\hat{\sigma}^2$

Demostración

2. Probar que $traza(M) = n - K$

$$traza(M) = traza(I_n - P) \qquad \text{Ya que } P = X(X'X)^{-1}X'$$

$$traza(M) = traza(I_n) - traza(P) \qquad \text{El operador de la traza es lineal}$$

$$traza(M) = n - traza(P)$$

y

$$traza(P) = traza[X(X'X)^{-1}X'] \qquad \text{Ya que } P = X(X'X)^{-1}X'$$

$$traza(P) = traza[(X'X)^{-1}X'X] \qquad \text{Ya que } traza(AB) = traza(BA)$$

$$traza(P) = traza[I_K]$$

$$traza(P) = K$$

Por lo tanto, $traza(M) = n - K$

3. Perturbaciones Esféricas y No Esféricas

3.1 Perturbaciones Esféricas

Supuesto: Perturbaciones Esféricas

❖ Homocedasticidad

Asume que todas las unidades tienen el mismo error de varianza

$$E\left[\varepsilon_i^2 \middle| X\right] = \sigma_0^2 > 0 \qquad i = 1,2,\ldots,n$$

❖ No Correlación

Asume también que no hay correlación entre las observaciones

$$E\left[\varepsilon_i \varepsilon_j \middle| X\right] = 0 \qquad i,j = 1,2,\ldots,n \qquad i \neq j$$

Perturbaciones Esféricas

Este supuesto señala que la varianza de todos los términos de errores es constante.

Para ver esto:

$$Var(\varepsilon_i|X) = E\left(\varepsilon_i^2|X\right) - [E(\varepsilon_i|X)]^2 \qquad \textbf{(Por varianza condicional)}$$

$$Var(\varepsilon_i|X) = E\left(\varepsilon_i^2|X\right) \qquad \textbf{(Por exogeneidad estricta)}$$

Similarmente, la covarianza de la distribución conjunta de $(\varepsilon_i\varepsilon_j)$ condicional a X es cero.

$$Cov\left(\varepsilon_i\varepsilon_j\middle|X\right) = 0 \qquad (i, j = 1,2, \dots, n \qquad i \neq j)$$

$$E\left(\varepsilon_i\varepsilon_j\middle|X\right) - E(\varepsilon_i|X)E(\varepsilon_j|X) = 0$$

$$E\left(\varepsilon_i\varepsilon_j\middle|X\right) = 0$$

Perturbaciones Esféricas

En notación matricial:

$$E(\varepsilon\varepsilon'|X) = \begin{pmatrix} E(\varepsilon_1\varepsilon_1|X) & E(\varepsilon_1\varepsilon_2|X) & \cdots & E(\varepsilon_1\varepsilon_n|X) \\ E(\varepsilon_2\varepsilon_1|X) & E(\varepsilon_2\varepsilon_2|X) & \cdots & E(\varepsilon_2\varepsilon_n|X) \\ \vdots & \vdots & \ddots & \vdots \\ E(\varepsilon_n\varepsilon_1|X) & E(\varepsilon_n\varepsilon_2|X) & \cdots & E(\varepsilon_n\varepsilon_n|X) \end{pmatrix}$$

$$E(\varepsilon\varepsilon'|X) = \begin{pmatrix} \sigma_0^2 & 0 & \cdots & 0 \\ 0 & \sigma_0^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_0^2 \end{pmatrix} = \sigma_0^2 I_n$$

Los términos de errores que cumplen los supuestos de homocedasticidad y no autocorrelación son llamados "**perturbaciones o errores esféricas**".

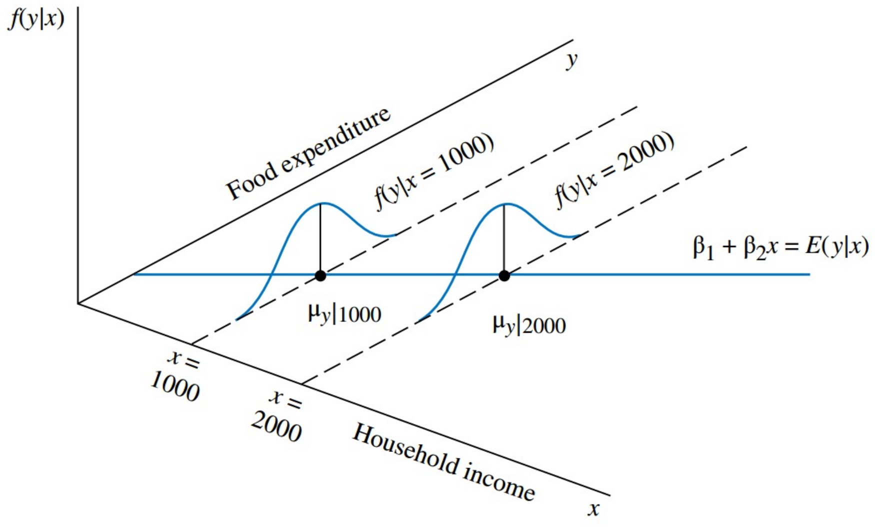
Perturbaciones Esféricas

- ¿Qué consecuencias tiene el no cumplimiento de este supuesto?
- ¿Cómo afecta las propiedades estadísticas del estimador de MCO?

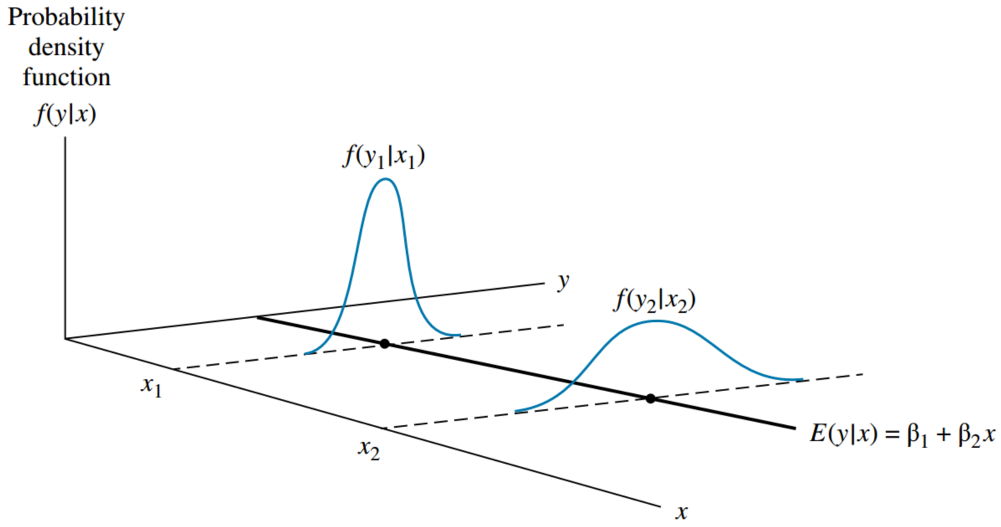
Motivación

Supuesto de Homocedasticidad

Varianza del error no observable ε condicionado sobre las variables explicativas es constante.



Homocedasticidad



Heterocedasticidad

3. Perturbaciones No Esféricas

3.2 Perturbaciones No Esféricas

Perturbaciones No Esféricas - Definición

Recordemos que hemos asumido que:

$$Var(\varepsilon|X) = \sigma_0^2 I_n = \begin{pmatrix} \sigma_0^2 & 0 & \cdots & 0 \\ 0 & \sigma_0^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_0^2 \end{pmatrix}$$

Eliminando el supuesto de homocedasticidad:

$$Var(\varepsilon|X) = E(\varepsilon\varepsilon'|X) = \sigma_0^2 \Omega(X)$$

Perturbaciones No Esféricas - Definición

La matriz $\Omega(X)$ es no singular y conocida. Esta matriz puede contener elementos distintos en su diagonal principal y elementos diferentes de cero fuera de su diagonal:

$$\Omega(X) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{12} & \sigma_{2n} & \dots & \sigma_n^2 \end{bmatrix}$$

Es decir, la varianza no es necesariamente constante y la covarianza entre diferentes perturbaciones podría ser diferente de cero.

4. Propiedades del Estimador $\widehat{\beta}$ en presencia de Perturbaciones No Esféricas

Consecuencias de relajar el supuesto: Insesgadez

Evaluar insesgadez:

$$E\left(\hat{\beta}_{MCO}\middle|X\right) =?$$

Demostración?

Consecuencias de relajar el supuesto: Insesgadez

Demostración

Tenemos:

$$\hat{\beta} = (X'X)^{-1}X'y$$

$$\hat{\beta} = (X'X)^{-1}X'(X\beta_0 + \varepsilon)$$

$$\hat{\beta} = (X'X)^{-1}X'X\beta_0 + (X'X)^{-1}X'\varepsilon$$

$$\hat{\beta} = \beta_0 + (X'X)^{-1}X'\varepsilon$$

Luego:

$$E(\hat{\beta}_{MCO}|X) = \beta + (X'X)^{-1}X'E(\varepsilon|X)$$

$$E(\hat{\beta}_{MCO}|X) = \beta$$

Sigue siendo **insesgado**.

Consecuencias de relajar el supuesto: Varianza

Recordemos que la matriz de varianza-covarianza homocedástica es:

$$Var(\hat{\beta}_{MCO}) = \sigma_0^2(X'X)^{-1}$$

Demostración?

Consecuencias de relajar el supuesto: Varianza

Demostración: Varianza Homocedástica

Tenemos:

$$Var(\hat{\beta}|X) = E(\hat{\beta} - E[\hat{\beta}]|X)^2 \quad \textbf{(Por varianza condicional)}$$

$$Var(\hat{\beta}|X) = E(\hat{\beta} - \beta_0|X)^2 \quad \textbf{(Por insesgadez)}$$

$$Var(\hat{\beta}|X) = E(A\varepsilon|X)^2 \quad \textbf{(Por sampling error)}$$

$$Var(\hat{\beta}|X) = E(A\varepsilon\varepsilon'A'|X) \quad \textbf{(Forma cuadrática)}$$

$$Var(\hat{\beta}|X) = AE(\varepsilon\varepsilon'|X)A' \quad \textbf{(A es una función de X)}$$

$$Var(\hat{\beta}|X) = A(\sigma_0^2 I_n)A' \quad \textbf{(Por homocedasticidad)}$$

$$Var(\hat{\beta}|X) = \sigma_0^2 AA'$$

$$\boldsymbol{Var(\hat{\beta}|X) = \sigma_0^2(X'X)^{-1}}$$

Consecuencias de relajar el supuesto: Varianza

Evaluar matriz de varianza-covarianza no homocedástica:

$$Var(\hat{\beta}_{MCO}) = ?$$

Demostración?

Consecuencias de relajar el supuesto: Varianza

Demostración: Varianza No Homocedástica

Tenemos:

$$Var(\hat{\beta}_{MCO}|X) = E[(\hat{\beta} - E(\hat{\beta}))\left(\hat{\beta} - E(\hat{\beta})\right)'|X]$$

$$Var(\hat{\beta}_{MCO}|X) = E(\hat{\beta} - E(\hat{\beta})|X)^2$$

$$Var(\hat{\beta}_{MCO}|X) = E(\hat{\beta} - \beta_0|X)^2$$

$$Var(\hat{\beta}|X) = E(A\varepsilon|X)^2$$ **(Por sampling error)**

$$Var(\hat{\beta}_{MCO}|X) = E(A\varepsilon\varepsilon'A'|X)$$ **(Forma cuadrática)**

$$Var(\hat{\beta}_{MCO}|X) = AE(\varepsilon\varepsilon'|X)A'$$ **(A es una función de X)**

$$Var(\hat{\beta}_{MCO}|X) = (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1}$$

$$\boldsymbol{Var(\hat{\beta}_{MCO}|X) = (X'X)^{-1}X'\Omega X(X'X)^{-1}}$$

Esta difiere de la matriz de varianza-covarianza homocedástica

Consecuencias de relajar el supuesto

El estimador de σ^2, $\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-K}$, ya no es insesgado. Tarea: **Demostrarlo**.

El Teorema de Gauss Markov no se mantiene, por lo tanto el estimador MCO no son MELI. ¿Por qué?

El estadístico t y F no son válidas. Los estadísticos utilizados en los intervalos de confianza y en los contrastes de hipótesis ya no siguen las distribución señaladas, de modo que carecen de validez.

5. Mínimos Cuadrados Generalizados

5.1 MCG

Mínimos Cuadrados Generalizados

- Dado los fallos que ocurren en las propiedades de los estimadores MCO, surge la conveniencia de buscar estimadores alternativos que verifiquen mejores propiedades que los de MCO.

- Este es el caso de los estimadores de Mínimos Cuadrados Generalizados (MCG)

- Para construirlos basta observar una propiedad del nuevo modelo, que depende de la descomposición de la matriz de varianza-covarianza de las perturbaciones.

Modelo Transformado

- Dado que Ω por construcción es una matriz simétrica y definida positiva, podemos encontrar una matriz C no singular (cuadrada) y de orden n tal que:

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

- La inversa de Ω también debe ser una matriz positiva definida y simétrica, y C es una matriz $n \times n$ no singular. Es posible reescribir:

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

$$C\Omega C' = CC^{-1}(C')^{-1}\,C' = I$$

Modelo Transformado

- Crear un nuevo modelo de regresión transformando (y, X, ε) por C como:

$$\tilde{y} \equiv Cy \qquad \tilde{X} \equiv CX \qquad \tilde{\varepsilon} \equiv C\varepsilon$$

- Por supuesto de linealidad para (y, X, ε) implica que $(\tilde{y}, \tilde{X}, \tilde{\varepsilon})$ también satisface linealidad:

$$\tilde{y} = \tilde{X}\beta_0 + \tilde{\varepsilon}$$

- Supongamos que la matriz C, que aún no sabemos como es, cumple su objetivo, teniendo entonces que $E(\tilde{\varepsilon}|\tilde{X}) = 0$ y $Var(\tilde{\varepsilon}|\tilde{X}) = \sigma_0^2 I_n$.

Modelo Transformado

Este modelo transformado cumple supuesto de exogeneidad estricta:

$$E\left(\tilde{\varepsilon}\middle|\tilde{X}\right) =?$$

Demostración: Exogeneidad Estricta

Tenemos:

$$\begin{aligned} E\left(\tilde{\varepsilon}\middle|\tilde{X}\right) &= E(\tilde{\varepsilon}|X) \\ &= E(C\varepsilon|X) \\ &= CE(\varepsilon|X) \\ &= 0 \end{aligned}$$

Modelo Transformado

Supuesto de homocedasticidad se cumple para el modelo transformado:

$$E\left(\tilde{\varepsilon}\tilde{\varepsilon}'\middle|\tilde{X}\right) = ?$$

Demostración: Homocedasticidad

Tenemos:

$$\begin{aligned} E\left(\tilde{\varepsilon}\tilde{\varepsilon}'\middle|\tilde{X}\right) &= E(\tilde{\varepsilon}\tilde{\varepsilon}'|X) \\ &= CE(\varepsilon\varepsilon'|X)C' \\ &= C\sigma_0^2\Omega C' \\ &= \sigma_0^2 C\Omega C' \\ &= \sigma_0^2 I_n \end{aligned}$$

Finalmente, $\tilde{\varepsilon}|\tilde{X}$ es normal porque la distribución de $\tilde{\varepsilon}|\tilde{X}$ es la misma que $\tilde{\varepsilon}|X$ y $\tilde{\varepsilon}$ es una transformación lineal de ε

Estimador MCG

- Estimador de mínimo cuadrados generalizados

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

$$\hat{\beta}_{MCG} = [(CX)'(CX)]^{-1}(CX)'Cy$$

$$\hat{\beta}_{MCG} = (X'C'CX)^{-1}(X'C'Cy)$$

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

Propiedades de MCG

Demostración: Insesgadez de $\widehat{\boldsymbol{\beta}}_{\boldsymbol{MCG}}$

Tomando el valor esperado de $\hat{\beta}_{MCG}$:

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

$$E(\hat{\beta}_{MCG}|X) = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(X\beta_0 + \varepsilon|X)$$

$$E(\hat{\beta}_{MCG}|X) = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X\beta_0$$

$$E(\hat{\beta}_{MCG}|X) = \beta_0$$

El estimador sigue siendo insesgado.

Propiedades de MCG

Su varianza condicional es:

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

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

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

El estimador $\hat{\beta}_{MCG}$ es eficiente (mínima varianza).

Propiedades de MCG

❖ Teorema de Aitken

El estimador $\hat{\beta}_{MCG}$ es lineal, insesgado y su varianza es incluso menor que la varianza de MCO.

$$Var(\hat{\beta}_{MCG}|X) < Var(\hat{\beta}_{MCO}|X)$$

$$\sigma^2(X'\Omega^{-1}X)^{-1} < \sigma_0^2(X'X)^{-1}$$

Para el modelo de perturbaciones no esféricas, el estimador MCG es el de menor varianza dentro de la clase de estimadores lineales e insesgados.

Propiedades de MCG

- El estimador insesgado de σ^2 es:

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta}_{MCG})'\Omega^{-1}(y - X\hat{\beta}_{MCG})'}{n-k}$$

- Se verifica que el estadístico:

$$\frac{\hat{\beta}_{MCG} - \beta}{\sqrt{\widehat{Var}(\hat{\beta}_{MCG})}} \sim t(n-k)$$

Inferencia

- Dicho estadístico se puede utilizar para obtener el intervalo de confianza para β a un nivel de confianza $1-\alpha$:

$$\hat{\beta}_{MCG} \pm t_{n-k;1-\frac{\alpha}{2}}\sqrt{\widehat{Var}(\hat{\beta}_{MCG})}$$

- Así como resolver contrastes de hipótesis sobre β
- Si se quiere contrastar:

$$H_0: R\beta = r$$

$$H_1: R\beta \neq r$$

- El estadístico de contraste es:

$$F = \frac{\left(R\hat{\beta}_{MCG} - r\right)'[R(X'\Omega^{-1}X)R']^{-1}\left(R\hat{\beta}_{MCG} - r\right)/\#r}{q\hat{\sigma}^2_{MCG}} \sim F(q, n-k)$$

De forma que se rechaza H_0 si $F_{exp} > F_{q,n-k;1-\alpha}$.

5. Mínimos Cuadrados Generalizados

5.2 Mínimos Cuadrados Generalizados Factibles

¿Y si Ω no es conocida?

- En la práctica, la matriz Ω raramente es conocida, por lo que el estimador MCG no se puede calcular.
- La alternativa es sustituir esta matriz por un estimador suyo $\widehat{\Omega}$, obteniendo así el estimador de mínimos cuadrados generalizados factibles (MCGF)

$$\widehat{\boldsymbol{\beta}}_{MCGF} = (\boldsymbol{X}'\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}'\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{y})$$

- Como consecuencia, las propiedades que satisface el vector de estimadores MCG ya no se mantienen. De esta forma, el estimador MCGF, en general, no es insesgado ni óptimo.

6. Heterocedasticidad

Naturaleza de la Heterocedasticidad

❖ Hipótesis de homocedasticidad

$$var(\varepsilon_i) = \sigma^2, \forall i = 1, \dots, n$$

❖ Hipótesis de heterocedasticidad

$$var(\varepsilon_i) = \sigma_i^2, \forall i = 1, \dots, n$$

Caso especial de perturbaciones esféricas que se presenta cuando la varianza de los términos de perturbación no es la misma para cada individuo o unidad de análisis:

$$Var(\varepsilon|X) = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

Los términos de pertubación siguen siendo independientes.

Naturaleza de la Heterocedasticidad

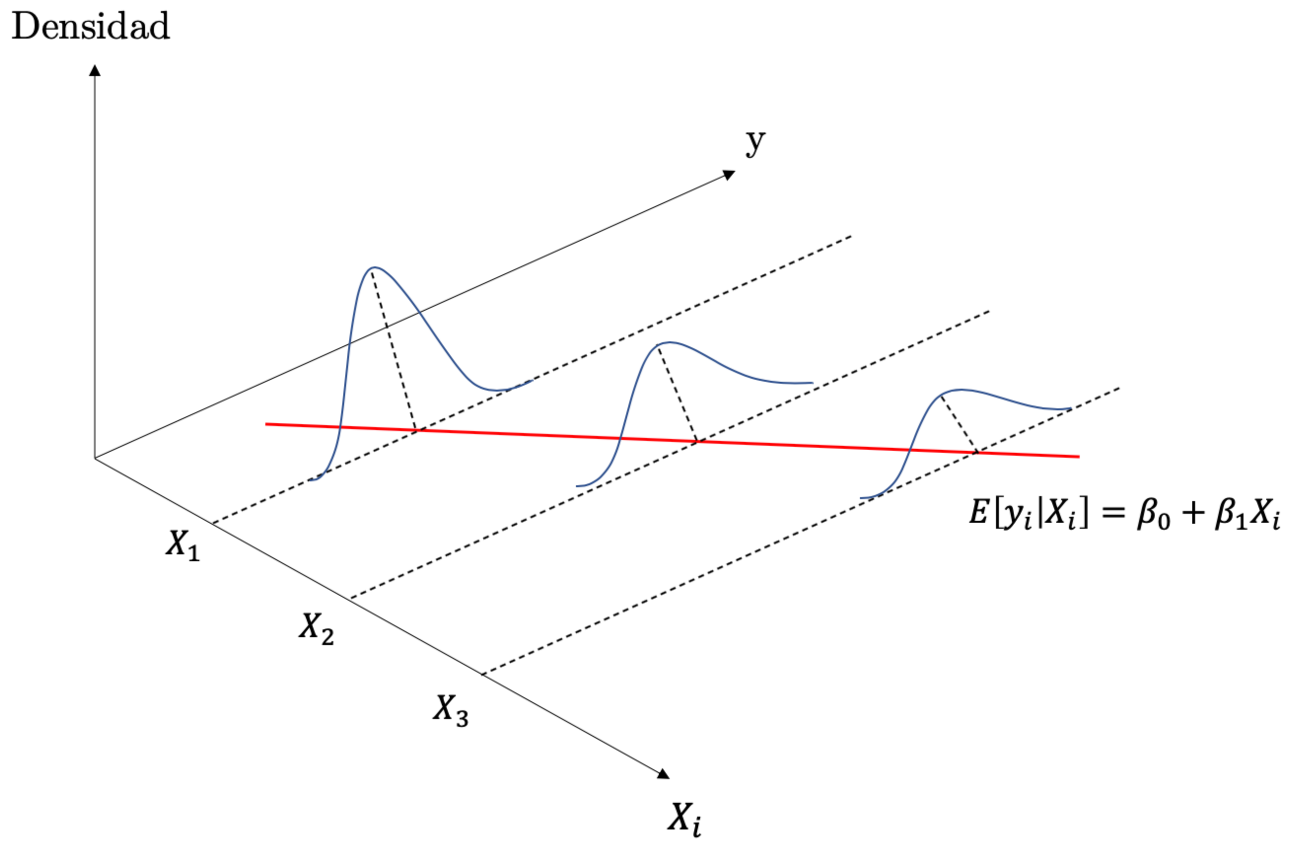
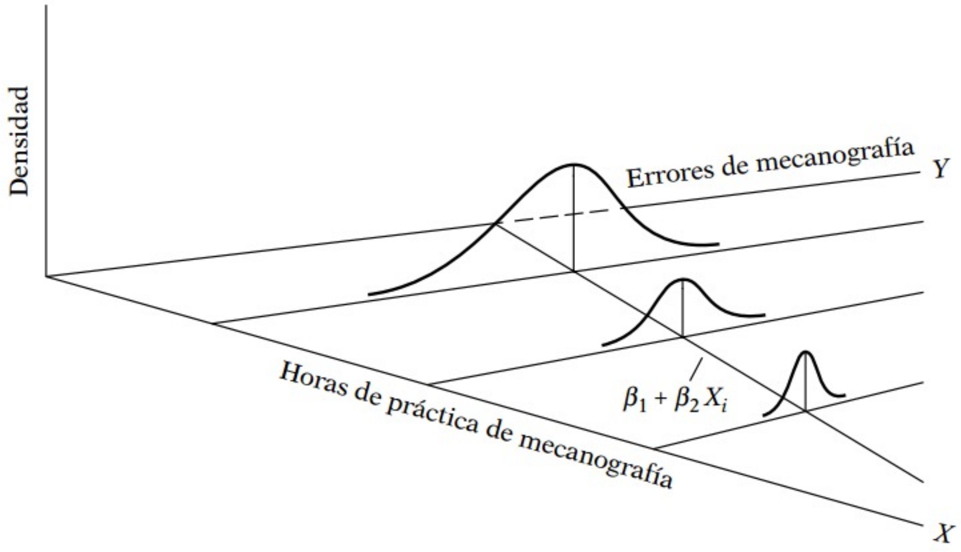


Figura 1. Heterocedasticidad

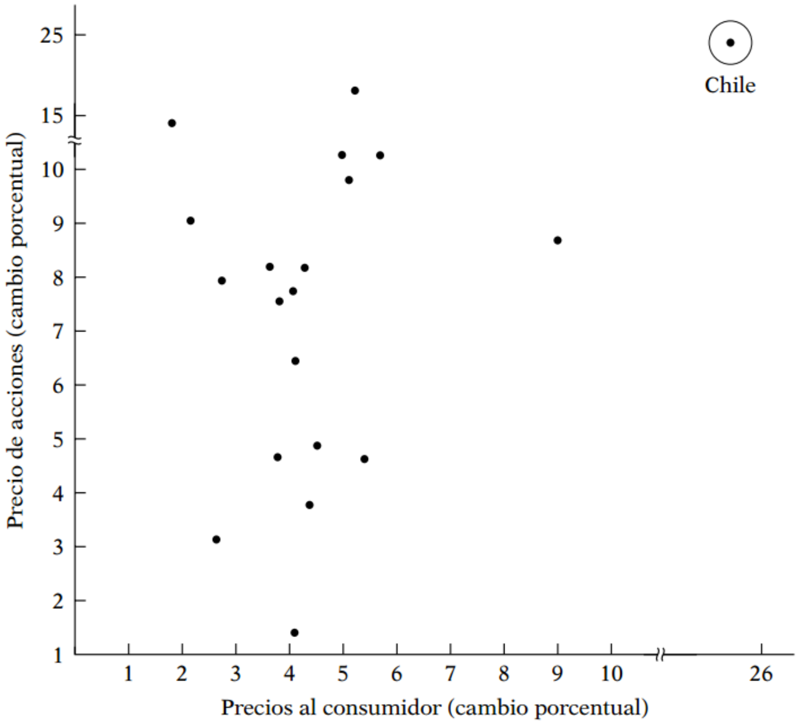
Causas de la Heterocedasticidad

1. Con base en los modelos de aprendizaje de los errores, a medida que las personas aprenden, disminuyen sus errores de comportamiento con el tiempo.



Ejemplo. Prueba de mecanografía

2. La presencia de datos atípicos provoca una mayor variabilidad del error.



Ejemplo. Relación entre precios de acciones y precios al consumidor

Causas de la Heterocedasticidad

3. A medida que mejoran las técnicas de recolección de datos, es probable que σ_i^2 se reduzca. **Ejemplo:** bancos con equipos complejos de procesamiento de información cometan menos errores en los informes mensuales.

4. Otra fuente de la heterocedasticidad es la asimetría en la distribución de uno o más regresores incluidos en el modelo. **Ejemplo:** variables económicas como riqueza, ingreso y escolaridad. La riqueza le corresponde a uno o cuantos individuos pertenecientes a los estratos superiores.

5. Surge debido a i) la incorrecta transformación de los datos (transformaciónes de razón o de primeras diferencias) y ii) una forma funcional incorrecta (modelos lineales frente a modelos log-lineales)

6. Omisión de variables importantes en el modelo.

Nota: La heterocedasticidad es un problema típico de modelos que utilizan datos de corte transversal.

Consecuencias de la Heterocedasticidad

- Estimación sesgada de la varianza, σ^2
- Los estimadores MCO son insesgados, pero menos eficientes que los MCG (es decir, no tienen varianza mínima).
- Los intervalos de confianza y los estadísticos utilizados para resolver contrastes de hipótesis ya no son fiables. Por tanto, existe la posibilidad de extraer conclusiones erróneas.

7. Detección de Heterocedasticidad

Detección de heterocedasticidad

- No es fácil detectar la heterocedasticidad en una situación concreta, ya que σ_i^2 solo se puede conocer si tenemos toda la población de Y correspondiente al valor X_i.
- En algunas ocasiones la propia naturaleza del problema sugiere si es probable que exista heterocedasticidad.
- Disponemos de varias herramientas de diagnóstico que nos pueden ayudar a detectar la heterocedasticidad.
- Se clasifican en:
  - Procedimientos informales o subjetivos
  - Procedimientos formales u objetivos

8. Detección de Heterocedasticidad

8.1 Procedimientos Informales

Procedimientos Informales

Son subjetivos y se basan en la representación gráfica de los residuos:

- Se calculan los residuos estimados del modelo por MCO.
- Se representan los $\hat{\varepsilon}_i^2$ o los $\hat{\varepsilon}_i$ frente a los $\hat{Y}_i$, o bien frente a la variable que se cree que provoca la heterocedasticidad.
- Si los puntos se disponen en el gráfico de forma aleatoria, no hay evidencia de que se incumpla la hipótesis de homocedasticidad.
- Si el gráfico muestra un patrón de comportamiento (lineal, cuadrático, exponencial, etc.) es probable que haya heterocedasticidad.

Procedimientos Informales

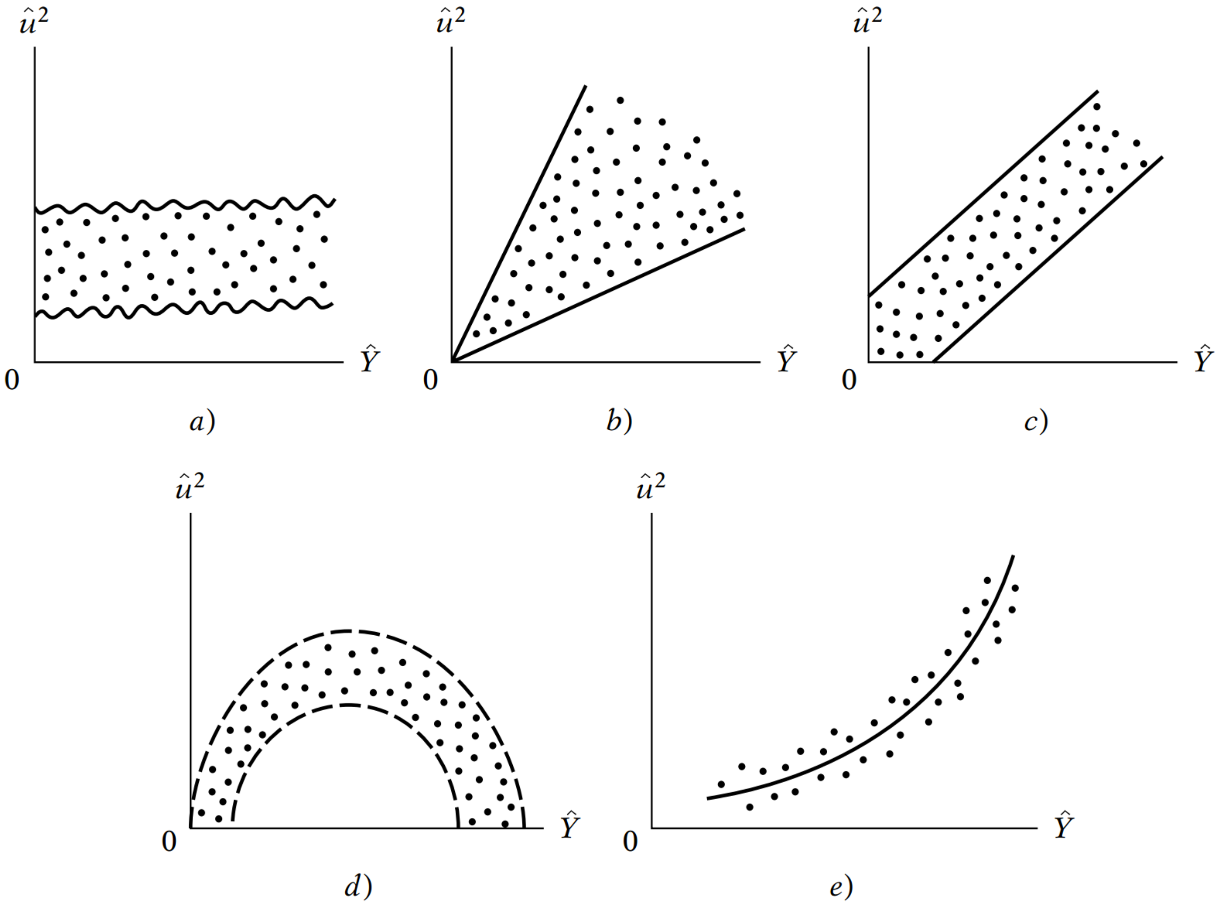


Figura 2. Procedimiento Informal

8. Detección de Heterocedasticidad

8.2 Procedimientos Formales

Procedimientos Formales

Son objetivos ya que se trata de contrastes de hipótesis.

- Existe diversos test de heterocedasticidad, cada uno de los cuáles parte de un determinado supuesto acerca del posible patrón de heterocedasticidad:
  - Test de Goldfeld-Quant
  - Test de Breusch-Pagan (1979)
  - Test de Glesjer
  - Test de White (1980)
- Cada uno de estos test plantea como hipótesis nula la ausencia de heterocedasticidad pero se diferencian entre sí en la forma de plantear la hipótesis alternativa.

Test de Goldfeld-Quant

- Este test supone que la varianza de las perturbaciones (heterocedasticidad) está relacionada, normalmente, con una de las variables explicativas del modelo.
- Bajo el supuesto de que dicha dependencia sea, por ejemplo, positiva (los menores/mayores valores de la varianza se producen cuando los valores de la variable son menores/mayores).

Test de Goldfeld-Quant

❖ **Pasos:**

1. Ordenar de menor a mayor los valores respecto a X_i, donde se supone que ésta es la variable explicativa causante de la heterocedasticidad.
2. Se omiten las c observaciones centrales.
3. Realizar dos estimaciones por MCO utilizando las $\frac{N-c}{2}$ primeras observaciones y las $\frac{N-c}{2}$ últimas observaciones.
4. Calcular la suma de los cuadrados de los residuos asociados a las dos estimaciones (SCR_1, SCR_2)
5. Bajo el supuesto de homocedasticidad se verifica que:

$$\frac{SCR_2}{SCR_1} \sim F_{\frac{N-c}{2}-k}$$

Con lo cual, fijado un nivel de significancia α, rechazaremos la homocedasticidad si:

$$\frac{SCR_2}{SCR_1} > F_{\frac{N-c}{2}-k;1-\alpha}$$

Test de Goldfeld-Quant

❖ **Incovenientes:**

- ¿Cómo elegir c? → $c \approx \frac{N}{3}$
- ¿Y si desconocemos cuál es la variable X_i causante de la heterocedasticidad o se sospecha de varias variables a la vez?

Test de Breusch Pagan/Godfrey

- **Supuesto**: La heterocedasticidad está provocada por una o más variables, no necesariamente presentes en el modelo lineal inicial.
- Se supone que la varianza del término de error depende de un vector de p variables Z, del siguiente modo:

$$\sigma_i^2 = f(\alpha_0 + \alpha_1 Z_{1i} + \cdots + \alpha_p Z_{pi})$$

- Donde f es una forma funcional cualquiera, α_j, $j = 0, \ldots, p$, parámetros y $Z_1, \ldots, Z_p$ las variables presumiblemente causantes de la heterocedasticidad.
- En consecuencia, contrastar la heterocedasticidad equivale a probar la hipótesis:

$$H_0\colon\ \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$$

$$H_0\colon Algunos\ de\ ellos\ es\ no\ nulo$$

Test Breusch Pagan/Godfrey

❖ Pasos:

1. Estimar la ecuación principal $y = X\beta + \varepsilon$ por MCO y calcular los residuos y los residuos al cuadrado.
2. Calcular el estimador de σ^2: $\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n}$
3. Generar un nueva variable $g_i = \frac{\hat{\varepsilon}_i^2}{\hat{\sigma}^2}$, $i = 1, \dots, n$
4. Regresionar g_i contra $Z_1, Z_2, \dots, Z_p$:

$$g_i = \alpha_0 + \alpha_1 Z_{1i} + \cdots + \alpha_p Z_{pi} + v_i$$

5. Estimar $\alpha_0, \dots, \alpha_p$ por MCO y calcular la SCE de la regresión auxiliar.
6. Bajo la hipótesis nula de homocedasticidad y normalidad del término de error; se sabe que:

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

7. Por tanto, se rechaza la hipótesis nula de homocedasticidad cuando:

$$Q_{exp} > \chi^2_{(p-1);1-\alpha}$$

Test de White

- **Supuesto**: La heterocedasticidad está provocada por una o más variables del modelo:

$$\sigma_i^2 = f(X_2, X_3, \dots, X_k)$$

Donde $f(\cdot)$ es una función polinómica.

Test de White

❖ **Pasos:**

1. Estimar por MCO el modelo $y = X\beta + \varepsilon$ y obtener los residuos $\hat{\varepsilon}_i$ y $\hat{\varepsilon}_i^2$.

2. Hacer la regresión de los residuos al cuadrado con respecto a todas las variables explicativas, sus cuadrados y todos sus productos cruzados. Por ejemplo, para un modelo con dos variables explicativas:

$$\hat{\varepsilon}_i^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{1i}^2 + \alpha_4 X_{2i}^2 + \alpha_5 X_{1i} X_{2i} + \upsilon_i$$

3. La hipótesis nula de homocedasticidad es: H_0: $\alpha_i = 0$

4. Bajo esta hipótesis, el estadístico:

$$n \times R_{aux}^2 \sim \chi_p^2$$

   Donde p es el número de variables explicativas en dicha regresión.

5. Así, se rechaza la hipótesis de homocedasticidad si:

$$n \times R_{aux}^2 > \chi_{p,1-\alpha}^2$$

8. Estimación bajo Heterocedasticidad

8.1 Matriz Ω conocida

1. Mínimos Cuadrados Ponderados (MCP)

- En presencia de heterocedasticidad, MCO ya no es el mejor estimador lineal insesgado.
- Si se conoce la forma de la heterocedasticidad puede usarse la estimación por Mínimos Cuadrados Ponderados (MCP) para obtener estimadores más eficientes que los de MCO.
- Los estimadores MCP nos conducen a nuevos estadísticos t y F que tienen distribuciones t y F, respectivamente.
- La idea básica del procedimiento de Mínimos Cuadrados Ponderados se basa en transformar el modelo verdadero para que el término de error sea homocedástico.

1. Mínimos Cuadrados Ponderados (MCP)

Siendo el modelo a estimar:

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

No sabemos a ciencia cierta de qué depende la varianza del error ε_i, y, por lo tanto, no sabemos cómo construir la matriz $\Omega(X)$.

Empezamos haciendo una conjetura acerca de esta estructura de heterocedasticidad, asumiendo que:

$$Var(\varepsilon_i | x_1, \dots, x_n) = \sigma_i^2 = \sigma_0^2 \exp(x_i' \alpha)$$

Donde $\exp(\cdot)$ asegura que $Var(\varepsilon_i | x_1, \dots, x_n) > 0$. Asumimos que Ω es diagonal. Sea $h_i(X)$ los elementos diagonal de $\Omega(X)$. Por lo tanto,

$$Var(\varepsilon_i | X) = \sigma_i^2 = \sigma_0^2 \mathrm{h_i}(X)$$

Donde $h_i(X)$ son conocidos y positivos. Asimismo, $h_i(X)$ determina la heterocedasticidad.

1. Mínimos Cuadrados Ponderados (MCP)

- ¿Qué modelo estimamos? El modelo transformado es:

$$Cy = CX\beta_0 + C\varepsilon$$

Recordar que el estimador MCG es:

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

Donde:

$$\Omega^{-1} = \begin{pmatrix} 1/h_i & 0 & \cdots & 0 \\ 0 & 1/h_i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/h_i \end{pmatrix} = C'C$$

- Entonces, en este caso particular, Ω^{-1} es matriz diagonal cuyo elemento es $1/h_i$. Es decir:

$$C = \begin{pmatrix} \sqrt{h_1} & 0 & \cdots & 0 \\ 0 & \sqrt{h_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{h_n} \end{pmatrix} \qquad \Omega = \begin{pmatrix} h_1 & 0 & \cdots & 0 \\ 0 & h_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_n \end{pmatrix}$$

1. Mínimos Cuadrados Ponderados (MCP)

Si dividimos el modelo por $\sqrt{h_i}$ obtenemos un modelo transformado de la forma:

$$\tilde{y}_i = \beta_0 + \beta_1 \tilde{X}_{1i} + \beta_2 \tilde{X}_{2i} + \cdots + \beta_k \tilde{X}_{ki} + \tilde{\varepsilon}_i$$

Donde:

$$\tilde{y}_i = \frac{y_i}{\sqrt{h_i}} \qquad \tilde{X}_{ji} = \frac{X_{ji}}{\sqrt{h_i}} \qquad \tilde{\varepsilon}_i = \frac{\varepsilon_i}{\sqrt{h_i}}$$

El estimador MCG se obtiene regresionando:

$$Cy = \begin{pmatrix} \frac{y_1}{\sqrt{h_1}} \\ \frac{y_2}{\sqrt{h_2}} \\ \vdots \\ \frac{y_n}{\sqrt{h_n}} \end{pmatrix} \qquad Cx = \begin{pmatrix} \frac{x_1}{\sqrt{h_1}} \\ \frac{x_2}{\sqrt{h_2}} \\ \vdots \\ \frac{x_n}{\sqrt{h_n}} \end{pmatrix}$$

1. Mínimos Cuadrados Ponderados (MCP)

Aplicando MCO al modelo transformado, se obtiene el estimador WLS (mínimos cuadrados ponderados):

$$\hat{\beta}_{WLS} = [\sum_{i=1}^{n} w_i x_i x_i']^{-1}[\sum_{i=1}^{n} w_i x_i y_i]$$

Donde $w_i = 1/h_i$.

Los errores de este modelo son homocedásticos dado que:

$$E(\tilde{\varepsilon}_i|X) = E\left(\frac{\varepsilon_i}{\sqrt{h_i}}\middle|X\right) = \frac{E(\varepsilon_i|X)}{\sqrt{h_i}} = 0$$

$$Var(\tilde{\varepsilon}_i|X) = E\left(\frac{E(\varepsilon_i^2|X)}{h_i}\middle|X\right) = \frac{\sigma^2 \mathrm{h_i}}{h_i} = \sigma^2$$

8. Estimación bajo Heterocedasticidad

8.2 Matriz Ω desconocida

Usando varianzas corregidas de White

El problema con el enfoque anterior es que depende de la forma inicial de la estructura de la varianza del error

- Asumir que β de MCO es insesgado y consistente.
- Si n es muy grande, el estimador de la matriz de $var - cov$ de MCO consistentes con heterocedasticidad de White es:

$$Var(\hat{\beta}_{MCO}) = (X'X)^{-1}X'\widehat{\Omega}X(X'X)^{-1}$$

- Estimar la matriz Ω mediante los cuadrados de los residuos, como si fueran estimadores de las varianzas:

$$\widehat{\Omega} = \begin{bmatrix} \hat{\varepsilon}_1^2 & 0 & \cdots & 0 \\ 0 & \hat{\varepsilon}_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\varepsilon}_n^2 \end{bmatrix}$$

Referencias

Capítulo 11 - Gujarati, D., & Porter, D. (2010). Econometría (Quinta edición ed.). *& P. Carril Villareal, Trad.) México: Mc Graw Hill educación*.

Capítulo 12 - Greene, W. H. (2000). *Análisis econométrico*. Tercera edición. Mardrid.

PUCP