

# Econometría II

---

## *La heteroscedasticidad*

Alejandro Mosiño - *Universidad de Guanajuato*

v. 2025.02.10

**¿Qué es la heteroscedasticidad?**

# ¿En qué consiste la heteroscedasticidad?

Uno de los problemas que podemos encontrar en datos de *sección cruzada* es el de **heteroscedasticidad**. Formalmente, existe heteroscedasticidad cuando:

$$\begin{aligned}\text{var}(u_i) &= \mathbb{E}(u_i^2) \\ &= \sigma_i^2,\end{aligned}$$

para  $i = 1, 2, \dots, n$ .

Esto implica que la matriz de covarianzas de los errores se vea como:

$$\mathbf{\Omega} = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}$$

# ¿Cuáles son las consecuencias de la heteroscedasticidad?

En presencia de heteroscedasticidad:

- Los estimadores de MCO *siguen siendo* insesgados y consistentes.
- Los estimadores de MCO *dejan de ser* eficientes.
- Los resultados de las pruebas  $t$  y  $F$  *dejan de ser* confiables.

# Detección de heteroscedasticidad

---

# Detección de heteroscedasticidad (1/2)

Consideremos el modelo de regresión simple:

$$\text{Gasto en alimentos}_i = \beta_1 + \beta_2 \text{Ingreso}_i + u_i$$

Donde el gasto en alimentos se mide en dólares y el ingreso se mide en cientos de dólares por semana.

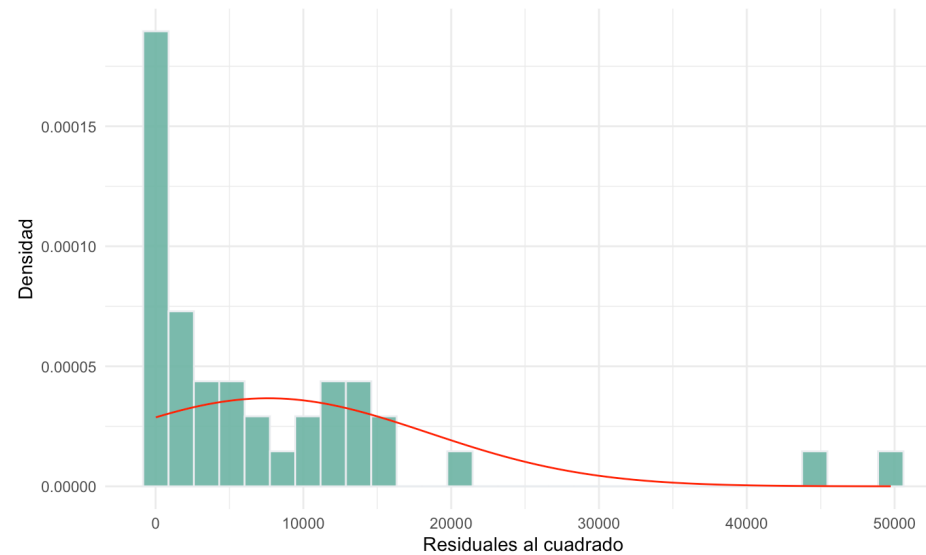
# Detección de heteroscedasticidad (2/2)

## Ejercicio

Utiliza los datos del archivo [E201-food.csv](#). Estima el modelo utilizando el método de los MCO. Explica tus resultados.

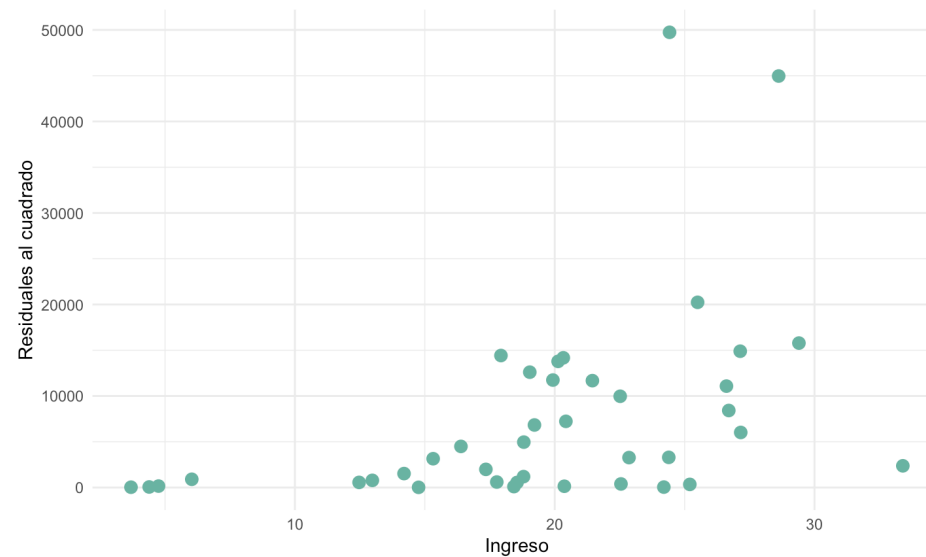
# Método gráfico (1/2)

Para detectar la presencia de heteroscedasticidad, primero calculamos el cuadrado de los residuales del modelo. Un histograma del **cuadrado de los residuales** muestra si estos no son homoscedásticos.



## Método gráfico (2/2)

También podemos graficar el cuadrado de los residuales estimados contra la variable explicativa para ver si existe algún patrón identificable.



**Nota:** Si tenemos muchas variables explicativas, también podríamos graficar el cuadrado de los residuales contra el valor estimado del gasto en alimentación,  $\hat{y}_i$ .

# Prueba de Breusch-Pagan (1/3)

En la prueba de **Breusch-Pagan** se plantea la siguiente hipótesis:

$$H_0 : \sigma_i^2 = \sigma^2$$

$$H_1 : \sigma_i^2 = h(\alpha_1 + \alpha_2 z_{2i} + \cdots + \alpha_s z_{si})$$

donde la función  $h(\cdot) > 0$  y las variables explicativas  $z_{2i}, z_{3i}, \dots, z_{si}$  pueden ser diferentes a las variables  $x_{2i}, x_{3i}, \dots, x_{ki}$  del modelo original.

# Prueba de Breusch-Pagan (2/3)

Los pasos para realizar la prueba de Breusch-Pagan son los siguientes:

1. Estimamos el modelo por MCO y guardamos el cuadrado de los residuales.
2. Regresamos el modelo:

$$e_i^2 = \alpha_1 + \alpha_2 z_{2i} + \cdots + \alpha_s z_{si} + v_i$$

3. Calculamos el estadístico  $LM = n \times R^2$  y lo comparamos contra una distribución  $\chi^2$  con  $s - 1$  grados de libertad, para algún nivel de significancia,  $\alpha$ .

# Prueba de Breusch-Pagan (3/3)

En nuestro ejemplo tenemos:

```
##  
## Breusch-Pagan test  
##  
## data: food_exp ~ income  
## BP = 7.3439, df = 1, p-value = 0.006729
```

Nota que, dado que no tenemos variables adicionales disponibles, en la prueba solo se ha usado la variable explicativa del modelo original.

# Prueba de White (1/3)

Un problema con la prueba de Breusch-Pagan es que supone que conocemos quiénes son las variables  $Z_{2i}, Z_{3i}, \dots, Z_{si}$ . Si no conocemos estas, aplicamos la **prueba de White**. En esta:

$$H_0 : \sigma_i^2 = \sigma^2 \forall i$$

$$H_1 : \sigma_i^2 \neq \sigma_j^2 \text{ para al menos un } i \neq j$$

# Prueba de White (2/3)

Los pasos para realizar la prueba de White son idénticos a los de la prueba de Breusch-Pagan, pero definimos las variables  $z_{2i}, z_{3i}, \dots, z_{si}$  a partir de las variables del modelo original. Por ejemplo, si:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

Tenemos que:

$$z_{2i} = x_{2i}; \quad z_{3i} = x_{3i}; \quad z_{4i} = x_{2i}^2; \quad z_{5i} = x_{3i}^2; \quad z_{6i} = x_{2i}x_{3i}$$

# Prueba de White (3/3)

En nuestro ejemplo:

```
## White's test results
##
## Null hypothesis: Homoskedasticity of the residuals
## Alternative hypothesis: Heteroskedasticity of the residuals
## Test Statistic: 7.56
## P-value: 0.022879
```

# Prueba de Goldfeld-Quandt (1/3)

La prueba de **Goldfeld-Quandt** está diseñada para analizar dos grupos de datos, **1** y **2**, con varianzas (posiblemente) diferentes,  $\sigma_1^2$  y  $\sigma_2^2$ , respectivamente. La hipótesis es:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

# Prueba de Goldfeld-Quandt (2/3)

Para realizar la prueba, se estima por MCO cada subgrupo por separado y se calcula el estadístico:

$$GQ = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{df_1, df_2},$$

donde  $df_1 = n_1 - k_1$  y  $df_2 = n_2 - k_2$ . Nota que, cuando la hipótesis nula es verdadera:

$$GQ = \frac{s_1^2}{s_2^2} \sim F_{df_1, df_2}$$

# Prueba de Goldfeld-Quandt (3/3)

En nuestro ejemplo, podemos definir como:

- Grupo 1: las 20 familias con los ingresos más bajos
- Grupo 2: las 20 familias con los ingresos más altos.

En este caso:

```
##  
## Goldfeld-Quandt test  
##  
## data: food_exp ~ income  
## GQ = 3.6148, df1 = 18, df2 = 18, p-value = 0.004596  
## alternative hypothesis: variance increases from segment 1 to 2
```

# Medidas correctivas de la heteroscedasticidad

---

# Medidas correctivas de la heteroscedasticidad (1/2)

Si logramos mostrar que nuestro modelo presenta problemas de heteroscedasticidad, el siguiente paso es aplicar alguna medida correctiva. Si conocemos la matriz de varianzas y covarianzas de los errores, aplicamos el método de los MCG.

Sin embargo, dado que el término de error es desconocido, la aplicación de este método es inviable.

# Medidas correctivas de la heteroscedasticidad (2/2)

Para corregir el problema de la heteroscedasticidad:

- Si no conocemos la matriz de varianzas y covarianzas de los errores, pero podemos “adivinar” su forma, aplicamos el método de los **MCG Factibles** (MCGF).
- Si no podemos adivinar la forma de la matriz de varianzas y covarianzas de los errores, estimamos el modelo por MCO, pero utilizando el **procedimiento de White** para calcular **varianzas robustas** (*varianzas asintóticamente válidas*).

# Mínimos Cuadrados Generalizados Factibles

Si no conocemos la varianza de los errores, también podemos estimar los coeficientes aplicando el método de los mínimos cuadrados generalizados factibles. En este, deseamos encontrar un estimador de  $\mathbf{\Omega}$  de tal forma que:

$$\begin{aligned} p \lim \frac{\mathbf{x}' \hat{\mathbf{\Omega}}^{-1} \mathbf{x}}{\sqrt{n}} &= p \lim \frac{\mathbf{x}' \mathbf{\Omega}^{-1} \mathbf{x}}{\sqrt{n}} \\ p \lim \frac{\mathbf{x}' \hat{\mathbf{\Omega}}^{-1} \mathbf{u}}{\sqrt{n}} &= p \lim \frac{\mathbf{x}' \mathbf{\Omega}^{-1} \mathbf{u}}{\sqrt{n}} \end{aligned}$$

Entonces, los estimadores de MCG y MCGF son *asintóticamente* equivalentes.

# MCGF: Heteroscedasticidad multiplicativa

Supongamos que:

$$\begin{aligned}\sigma_i^2 &= \exp(\alpha_1 + \alpha_2 z_{2i} + \cdots + \alpha_s z_{si}) \\ &= \exp(\alpha_1) \exp(\alpha_2 z_{2i} + \cdots + \alpha_s z_{si}) \\ &= \sigma^2 h(\alpha_2 z_{2i}, \dots, \alpha_s z_{si})\end{aligned}$$

Este es un modelo particularmente flexible y se conoce como **modelo de heteroscedasticidad multiplicativa**.

# Heteroscedasticidad multiplicativa: caso 1

Un caso especial del modelo de heteroscedasticidad multiplicativa es consiste en suponer que  $s = 2$  y  $z_{i2} = \ln(x_i)$ . De tal forma que:

$$\begin{aligned}\sigma_i^2 &= \exp(\alpha_1 + \alpha_2 z_{i2}) \\ &= \exp(\alpha_1) \exp(\alpha_2 \ln[x_i]) \\ &= \sigma^2 x_i^{\alpha_2},\end{aligned}$$

donde  $\alpha_2$  es un parámetro desconocido. Naturalmente, si este es igual a cero, no tendríamos problemas de heteroscedasticidad.

# Heteroscedasticidad multiplicativa: caso 2

Consideremos ahora dos grupos, 1 y 2. Definimos una variable dummy tal que  $D_i = 1$  si  $i$  pertenece al primer grupo y  $D_i = 0$  si  $i$  pertenece al segundo. En este caso, si  $s = 2$ :

$$\begin{aligned} \text{var}(u_i | x_i) &= \exp(\alpha_1 + \alpha_2 D_i) \\ &= \begin{cases} \exp(\alpha_1) = \sigma^2; & \text{si } D_i = 0 \\ \exp(\alpha_1 + \alpha_2) = \sigma^2 \exp(\alpha_2); & \text{si } D_i = 1 \end{cases} \end{aligned}$$

Naturalmente, si  $\alpha_2 = 0$  la varianza entre ambos grupos es la misma y no habría problemas de heteroscedasticidad.

# Estimación del modelo multiplicativo (1/3)

Para estimar un modelo con heteroscedasticidad multiplicativa seguimos los siguientes pasos:

1. Estimamos el modelo para  $y_i$  utilizando MCO. Guardamos el cuadrado de los residuales,  $e_i^2$ .
2. Una simple manipulación del modelo multiplicativo revela que  $\alpha_1, \alpha_2, \dots, \alpha_s$  pueden estimarse a partir de:

$$\ln(e_i^2) = \alpha_1 + \alpha_2 z_{2i} + \dots + \alpha_s z_{si} + v_i$$

**Nota.** Puede demostrarse que los estimadores de MCO para  $\alpha_1, \alpha_2, \dots, \alpha_s$  son consistentes.

# Estimación del modelo multiplicativo (2/3)

Para estimar un modelo con heteroscedasticidad multiplicativa seguimos los siguientes pasos:

3. Calculamos el valor estimado de  $h(\mathbf{z}), \hat{h}(z_{2i}, z_{3i}, \dots, z_{si})$ .
4. Multiplicamos cada variable del modelo por el factor de ponderación:

$$\frac{1}{\sqrt{w_i}} = \frac{1}{\sqrt{\hat{h}(z_{2i}, z_{3i}, \dots, z_{si})}}$$

5. Estimamos el modelo resultante por MCO.

# Estimación del modelo multiplicativo (3/3)

## Ejercicio

Muestra que los coeficientes del modelo multiplicativo,  $\alpha_1, \alpha_2, \dots, \alpha_s$ , pueden estimarse a partir del modelo:

$$\ln(e_i^2) = \alpha_1 + \alpha_2 z_{2i} + \dots + \alpha_s z_{si} + u_i$$

# MCGF vs MCP (1/2)

## Nota

En el contexto de la heteroscedasticidad, el método de los mínimos cuadrados generalizados factibles también se conoce como el **método de los mínimos cuadrados ponderados**.

## MCGF

*Cada observación* (incluido el error) se divide por el factor:

$$\sqrt{w_i} = \frac{1}{\sqrt{\hat{h}(z_{i2}, \dots, z_{is})}}$$

## MCP

*La varianza del error* se pondera por el factor:

$$w_i = \frac{1}{\hat{\sigma}_i^2}$$

## MCGF vs MCP (2/2)

Los métodos de MCGF y de MCP son equivalentes debido a las propiedades de la varianza. En específico:

$$\text{var}(\sqrt{w_i}u_i) = w_i \text{var}(u_i)$$

Donde el lado izquierdo de la igualdad corresponde al procedimiento de los MCGF y el lado derecho al procedimiento de los MCP.

# El método de los MCP: ejemplos (1/3)

Imaginemos que la varianza del error es proporcional a uno de los regresores. En este caso, estimamos (por MCO) el modelo usando el *inverso* de ese regresor como factor de ponderación.

```
##  
## Call:  
## lm(formula = food_exp ~ income, data = data, weights = w1)  
##  
## Weighted Residuals:  
##      Min      1Q  Median      3Q      Max  
## -45.37 -11.13  -1.33   15.02   39.23  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept)   78.684     23.789   3.308  0.00206 **  
## income        10.451      1.386   7.541 4.61e-09 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 18.75 on 38 degrees of freedom  
## Multiple R-squared:  0.5994, Adjusted R-squared:  0.5889  
## F-statistic: 56.87 on 1 and 38 DF,  p-value: 4.614e-09
```

## El método de los MCP: ejemplos (2/3)

Imaginemos ahora que la varianza del error es proporcional *al cuadrado* de uno de los regresores. En este caso, estimamos (por MCO) el modelo usando el *inverso del cuadrado* ese regresor como factor de ponderación.

```
##
## Call:
## lm(formula = food_exp ~ income, data = data, weights = w2)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -9.2423 -2.5664 -0.2214  3.4069  7.2566
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   76.270     12.004   6.354 1.86e-07 ***
## income        10.612       1.024  10.367 1.24e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.048 on 38 degrees of freedom
## Multiple R-squared:  0.7388, Adjusted R-squared:  0.7319
## F-statistic: 107.5 on 1 and 38 DF,  p-value: 1.24e-12
```

# El método de los MCP: ejemplos (3/3)

## Ejercicio

Como hemos dicho, cuando aplicamos el método de los MCP utilizando alguna ponderación,  $w$ , en realidad lo que estamos haciendo es dividir las variables del modelo por la raíz cuadrada de  $w$  y luego estimar los coeficientes utilizando el método de los MCO.

Averigua si los dos modelos estimados antes siguen presentando problemas de heteroscedasticidad.

# Errores estándar robustos de White

White sugirió un procedimiento que nos permite estimar el modelo por el método de los MCO, pero modificando los errores estándar para que estos tomen cuenta la heteroscedasticidad.

```
##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  83.4160    27.4637   3.0373  0.004299 **
## income       10.2096     1.8091   5.6436 1.755e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Como resultado los estimadores son insesgados, y además las pruebas  $t$  y  $F$  son válidas.

# Práctica

---

# Práctica

Utiliza los datos: [E202-abortion](#). El objetivo es determinar la tasa de abortos en los 50 estados de los Estados Unidos.

Usando la variable ABR (Abortion Rate) como dependiente y el resto de las variables como explicativas, realiza un análisis completo del problema de la heteroscedasticidad.