# Repaso de MCO


Gustavo A. García

ggarci24@eafit.edu.co



Econometría II

Programa de Economía

Universidad EAFIT

Link slides en formato html

Link slides en formato PDF

# En este tema

- El modelo
- Estimación MCO
- Propiedades de los estimadores MCO
- Ejercicio aplicado en R: tasas de retorno a la educación
- Ejercicio aplicado en Python: tasas de retorno a la educación

# Lecturas

- Wooldridge, J. (2013). *Introducción a la econometría*. 5a edición, Cengage Learning. Caps 3, 4, 5

- Gujarati, D. y Porter, D. (2010). *Econometría*. 5a edición, Mc Graw Hill. Caps 7, 8

- Judge, G., Hill, R., Griffiths, W., Lütkepohl, H. y Lee, T. (1988). *Introduction to the theory and practice of econometrics*. 2a edición, John Wiley & Sons. Cap. 5

# El modelo

Un modelo de regresión lineal múltiple tiene la siguiente estructura:

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_k X_{ik} + u_i$$

El modelo entonces contiene $k$ parámetros poblacionales (desconocidos). Se trabaja, entonces, con $n - k$ gdl

Recordando los supuestos iniciales:

- Los coeficientes $\beta_j$ con $j = 1, 2, 3, \ldots, k$ son fijos y desconocidos

- Las $X_{ij}$ son fijas o no estocásticas (no aleatorias) para $j = 2, 3, 4, \ldots, k$. Este es un supuesto de partida propio del laboratorio. Un supuesto más real en economía y que lleva a resultados similares es: las variables explicatorias son exógenas. Esto implica que:

$$Cov(X_{i2}, u_i) = 0$$

$$Cov(X_{i3}, u_i) = 0$$

$$\vdots$$

$$Cov(X_{ik}, u_i) = 0$$

- El modelo esta completo: $E(u_i) = 0, \forall i = 1 \ldots n$
- Homoscedasticidad: $Var(u_i) = E(u_i - E(u_i))^2 = E(u_i^2) = \sigma_u^2$
- No autocorrelación: $Cov(u_i, u_j) = E[(u_i - E(u_i))(u_j - E(u_j))] = E(u_i u_j) = 0, \forall i \neq j$
- Normalidad: $u_i \sim NID(0, \sigma_u^2)$

# El modelo

El modelo general es un polinomio de regresión de la forma:

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_k X_{ik} + u_i$$

Lo que el modelo dice es:

$$Y_1 = \beta_1 + \beta_2 X_{12} + \beta_3 X_{13} + \ldots + \beta_k X_{1k} + u_1$$

$$Y_2 = \beta_1 + \beta_1 X_{22} + \beta_3 X_{23} + \ldots + \beta_k X_{2k} + u_2$$

$$\vdots$$

$$Y_n = \beta_1 + \beta_2 X_{n2} + \beta_3 X_{n3} + \ldots + \beta_k X_{nk} + u_n$$

En matrices se tiene

$$\mathbf{Y}_{n\text{x}1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \mathbf{u}_{n\text{x}1} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad \mathbf{B}_{k\text{x}1} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \quad \mathbf{X}_{n\text{x}k} = \begin{bmatrix} 1 & X_{12} & X_{23} & X_{1k} \\ 1 & X_{22} & X_{23} & X_{2k} \\ \vdots & & & \\ 1 & X_{n2} & X_{n3} & X_{nk} \end{bmatrix}$$

En esta notación $X_{ij}$ indica: fila $i$ (observación), columna $j$ (variable explicatoria). El polinomio de regresión en álgebra matricial se puede escribir como:

$$\mathbf{Y}_{n\text{x}1} = \mathbf{X}_{n\text{x}k}\mathbf{B}_{k\text{x}1} + \mathbf{u}_{n\text{x}1}$$

# El modelo

La especificación del modelo de Regresión Lineal Múltiple (RLM) haciendo uso del álgebra matricial se tiene

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{u}$$

- Modelo completo: $E(\mathbf{u}) = \mathbf{0}$
- Exogeneidad: $E(\mathbf{X}'\mathbf{u}) = \mathbf{0}$
- Perturbaciones esféricas: $Cov(\mathbf{u}) = E(\mathbf{u}\mathbf{u}') = \sigma_u^2\mathbf{I}_n$ (homocedasticidad y no autocorrelación)
- No multicolinealidad perfecta: $\rho(\mathbf{X}_{n\text{x}k}) = k < n$
- Normalidad: $\mathbf{u}_{n\text{x}1} \sim \mathbf{N}(\mathbf{0}_{n\text{x}1}, \sigma_u^2\mathbf{I}_n)$

# Estimación MCO

Se construye matricialmente la sumatoria de cuadrados de los residuales $(SCR)$ y se deriva respecto a $\hat{\mathbf{B}}$

El modelo estimado es $\mathbf{Y} = \mathbf{X}\hat{\mathbf{B}} + \hat{\mathbf{u}}$ y sin necesidad de supuestos se sabe que

$$\hat{\mathbf{u}} = \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix}$$

y por ende la $SCR = \sum \hat{u}_i^2 = \hat{u}_1^2 + \hat{u}_2^2 + \ldots + \hat{u}_n^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}$

Si se transpone un vector y se premultiplica por el vector original se obtiene la suma de cuadrados de los elementos del vector

$$\begin{aligned} SCR = \hat{\mathbf{u}}'\hat{\mathbf{u}} &= (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) \\ &= (\mathbf{Y}' - \hat{\mathbf{B}}'\mathbf{X}')(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\mathbf{B}} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} + \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} \end{aligned}$$

Se tiene que $(\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y})' = \mathbf{Y}'\mathbf{X}\hat{\mathbf{B}}$ y que $\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}$ y $\mathbf{Y}'\mathbf{X}\hat{\mathbf{B}}$ son dos escalares iguales, entonces

$$SCR = \hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{Y}'\mathbf{Y} - \underbrace{2\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}}_{\text{forma lineal en } \hat{\mathbf{B}}} + \underbrace{\hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}}_{\text{forma cuadrática en } \hat{\mathbf{B}}}$$

# Estimación MCO

Derivando respecto a $\hat{\mathbf{B}}$ e igualando a cero

$$\frac{\partial \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\partial \hat{\mathbf{B}}} \implies \frac{\partial \mathbf{Y}'\mathbf{Y}}{\partial \hat{\mathbf{B}}} = \mathbf{0}$$
$$\implies \frac{\partial 2\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}}{\partial \hat{\mathbf{B}}} = 2\mathbf{X}'\mathbf{Y}$$
$$\implies \frac{\partial \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}}{\partial \hat{\mathbf{B}}} = 2(\mathbf{X}'\mathbf{X})\hat{\mathbf{B}}$$

$$\frac{\partial \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\partial \hat{\mathbf{B}}} = -2\mathbf{X}'\mathbf{Y} + 2(\mathbf{X}'\mathbf{X})\hat{\mathbf{B}} = \mathbf{0}$$

$$(\mathbf{X}'\mathbf{X})\hat{\mathbf{B}} = \mathbf{X}'\mathbf{Y} : \text{Ecuaciones normales}$$

$$\hat{\mathbf{B}}_{MCO} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

# Propiedades de los estimadores MCO

## Linealidad

Es demostrar que $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ es lineal en $Y$ y en $u$. Observando el vector $\hat{\mathbf{B}}$ se puede hacer lo siguiente

$$\hat{\mathbf{B}} = ((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')_{k\mathrm{x}n}\mathbf{Y}_{n\mathrm{x}1}$$

Se puede definir una matriz de $k\mathrm{x}n$

$$\mathbf{C}'_{k\mathrm{x}n} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Entonces

$$\hat{\mathbf{B}} = \mathbf{C}'\mathbf{Y}$$

Con

$$\mathbf{C}' = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ C_{k1} & C_{k2} & \cdots & C_{kn} \end{bmatrix}$$

Cada estimador $\hat{\beta}_j, j = 1, 2, \ldots, k$, puede escribirse como

$$\hat{\beta}_j = \sum C_{ij} Y_i$$

combinación lineal en $Y_i$

# Propiedades de los estimadores MCO

**Linealidad**

De la misma forma

$$\hat{\mathbf{B}} = \mathbf{C}'\mathbf{Y} = \mathbf{C}'(\mathbf{X}\mathbf{B} + \mathbf{u}) = \mathbf{C}'\mathbf{X}\mathbf{B} + \mathbf{C}'\mathbf{u}$$

Se observa que

$$\mathbf{C}'\mathbf{X} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}_k$$

Entonces

$$\hat{\mathbf{B}} = \mathbf{B} + \mathbf{C}'\mathbf{u}$$

Expresión aleatoria de $\hat{\mathbf{B}}$. Donde cada $\hat{\beta}_j$ es

$$\hat{\beta}_j = \beta_j + \sum C_{ij} u_i$$

Una combinación lineal en $u_i$

# Propiedades de los estimadores MCO

**Insesgadez**

Partiendo de la expresión aleatoria

$$E(\hat{\mathbf{B}}) = E(\mathbf{B} + \mathbf{C}'\mathbf{u}) = \mathbf{B} + \mathbf{C}'E(\mathbf{u}) = \mathbf{B}$$

**Mínima varianza**

Primero calculamos la matriz de varianzas-covarianzas de $\hat{\mathbf{B}}$ y luego corroboramos el Teorema de Gauss-Markov para demostrar que la varianza calculada es mínima

$$Cov(\hat{\mathbf{B}}) = E[(\hat{\mathbf{B}} - E(\hat{\mathbf{B}}))(\hat{\mathbf{B}} - E(\hat{\mathbf{B}}))'] = E[(\hat{\mathbf{B}} - \mathbf{B})(\hat{\mathbf{B}} - \mathbf{B})']$$

De la expresión aleatoria tenemos que

$$\hat{\mathbf{B}} = \mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$
$$\hat{\mathbf{B}} - \mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$
$$(\hat{\mathbf{B}} - \mathbf{B})' = \mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Entonces

$$Cov(\hat{\mathbf{B}}) = E[(\hat{\mathbf{B}} - \mathbf{B})(\hat{\mathbf{B}} - \mathbf{B})'] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Por el supuesto de perturbaciones esféricas $E(\mathbf{u}\mathbf{u}') = \sigma_u^2\mathbf{I}_n$

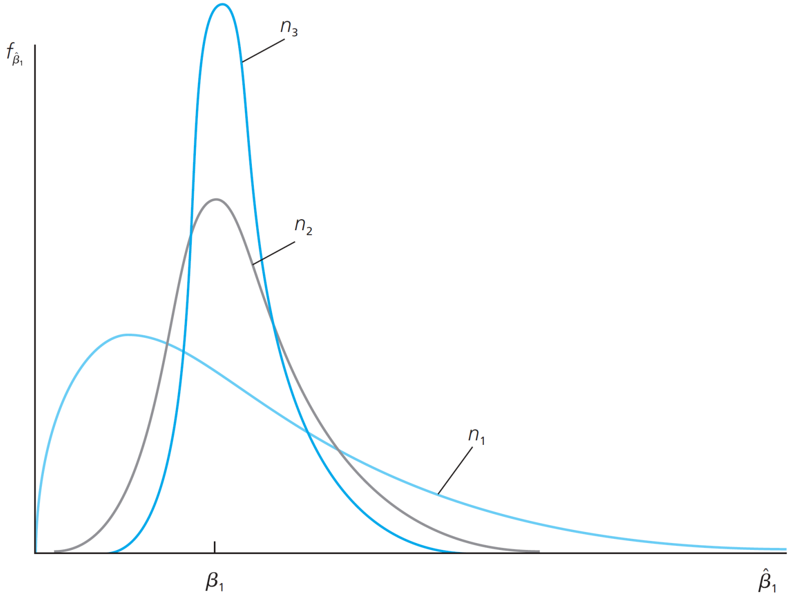
$$Cov(\hat{\mathbf{B}}) = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\underbrace{\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}}_{\mathbf{I}_k}$$

$$Cov(\hat{\mathbf{B}}) = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$$

# Propiedades de los estimadores MCO

## Consistencia

- Aunque la insesgadez de los estimadores es importante, no siempre puede lograrse: por ejemplo $\widehat{\sigma}_u^2$ es sesgado
- Aunque no todos los estimadores útiles son insesgados, casi todos los economistas están de acuerdo en que la consistencia es un requisito mínimo para un estimador
- Si el estimador de un determinado parámetro poblacional no es consistente, entonces se está perdiendo el tiempo
- $\widehat{\beta}_j$ es un estimador de $\beta_j$ y como es insesgado la distribución de probabilidad de $\widehat{\beta}_j$ tiene una media de $\beta_j$. Como estimador consistente, entonces a medida que el tamaño de la muestra aumenta, la distribución de $\widehat{\beta}_j$ se estrechará cada vez más entorno a $\beta_j$
- Esto significa que si es posible recolectar tantos datos como se desee, entonces puede hacerse que el estimador esté arbitrariamente cerca de $\beta_j$



# Propiedades de los estimadores MCO

**Normalidad asintótica**

- Se sabe que la normalidad no juega ningún papel en la insesgadez de los estimadores MCO y tampoco afecta las conclusiones de MCO es el mejor estimador lineal insesgado bajo los supuestos estándar
- Pero la inferencia exacta basada en los estadísticos $t$ y F requiere el supuesto que los errores se distribuyen normal
- Aunque las $Y_i$ no provienen de una distribución normal, puede emplearse el teorema central del límite para concluir que los estimadores de MCO satisfacen la normalidad asintótica, lo cual significa que están distribuidos de manera aproximadamente norma cuando se tienen muestras de tamaño suficientemente grande
- Este teorema indica es que, sin importar la distribución de la población de $u$, los estimadores de los MCO, cuando se estandarizan de manera apropiada, tienen distribuciones normales estándar aproximadas

# Ejercicio aplicado en R: tasas de retorno a la educación

```r
library(wooldridge); library(tidyverse); library(summarytools); library(modelsummary); library(gt)

data('wage1')
```

## Estimación por OLS

```r
ols1 <- lm(lwage ~ educ, data = wage1)
summary(ols1)
```

```
Call:
lm(formula = lwage ~ educ, data = wage1)

Residuals:
     Min       1Q   Median       3Q      Max
-2.21158 -0.36393 -0.07263  0.29712  1.52339

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.583773   0.097336   5.998 3.74e-09 ***
educ        0.082744   0.007567  10.935  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4801 on 524 degrees of freedom
Multiple R-squared:  0.1858,    Adjusted R-squared:  0.1843
F-statistic: 119.6 on 1 and 524 DF,  p-value: < 2.2e-16
```

```r
ols2 <- lm(lwage ~ educ + female, data = wage1)
summary(ols2)
```

```
Call:
lm(formula = lwage ~ educ + female, data = wage1)

Residuals:
     Min       1Q   Median       3Q      Max
-2.02672 -0.27470 -0.03731  0.26219  1.34738

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.826269   0.094054   8.785   <2e-16 ***
educ         0.077203   0.007047  10.955   <2e-16 ***
female      -0.360865   0.039024  -9.247   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4455 on 523 degrees of freedom
Multiple R-squared:  0.3002,    Adjusted R-squared:  0.2975
F-statistic: 112.2 on 2 and 523 DF,  p-value: < 2.2e-16
```

# Ejercicio aplicado en R: tasas de retorno a la educación

Utilizamos el paquete `modelsummary` para generar tablas editadas (Word, tex, text, png, html...)

```
modelos <- list("OLS1" = lm(lwage ~ educ, data = wage1),
                "OLS2"  = lm(lwage ~ educ + female, data = wage1))

modelsummary(modelos, output = 'gt', coef_map = c('educ' = 'Educación', 'female' = 'Mujer (=1)'), stars = c('*'=.1, '**'=.05, '***'=.01), statistic =
tab_style(style = cell_text(size = "x-small"), locations = cells_source_notes()) |>  tab_source_note(source_note = "Nota: Errores estándar en paréntes
```

Tabla 1. Determinantes de los salarios (Y = log(salario))

| | OLS1 | OLS2 |
|---|---|---|
| Educación | 0.083*** | 0.077*** |
| | (0.008) | (0.007) |
| Mujer (=1) | | -0.361*** |
| | | (0.039) |
| Num.Obs. | 526 | 526 |
| R2 | 0.186 | 0.300 |
| R2 Adj. | 0.184 | 0.298 |
| F | 119.582 | 112.189 |

* p < 0.1, ** p < 0.05, *** p < 0.01

Nota: Errores estándar en paréntesis

# Ejercicio aplicado en Python: tasas de retorno a la educación

En el siguiente link se puede descagar el código en Python: MCO.py