# ECONOMETRÍA

JOSÉ ALBERTO MAURICIO

2

## REGRESIÓN LINEAL MÚLTIPLE I

### 2.4 ESPECIFICACIÓN II

Departamento de Análisis Económico y Economía Cuantitativa
Universidad Complutense de Madrid

EctrGr-JAM-2-4.pdf



ucm.randomshock.com/ectrgr

Versión 2.4 - Enero 2024

## MATERIAL AUXILIAR Y COMPLEMENTARIO

En esta Sección 2.4 se supone conocido todo el instrumental matemático y estadístico utilizado en las Secciones 2.1 - 2.3. También se suponen bien conocidos todos los aspectos teóricos y prácticos del Análisis de Regresión Lineal cubiertos en el Tema 1 y en las Secciones 2.1 - 2.3.

Adicionalmente, en esta Sección 2.4 se indica la obligatoriedad de estudiar con atención fuera de clase tres apartados del manual de Wooldridge (2020). También se mencionan algunas secciones del manual de Hill, Griffiths, Lim (2018) como material complementario opcional para esta Sección 2.4.

## BIBLIOGRAFÍA PARA TODO EL TEMA 2

Hayashi (2000): Capítulo 1.
Hill, Griffiths, Lim (2018): Capítulos 2 - 6.
Wooldridge (2020): Capítulos 2 - 6 y Apéndice E.

## PARTE 1 - OMISIÓN DE VARIABLES EXPLICATIVAS RELEVANTES

BE.1 $$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + U\,,\ \text{con}\ \beta_3 \neq 0\,. \qquad [1]$$

ME.1 $$Y = \beta_1 + \beta_2 X_2 + U^*,\ \text{con}\ U^* = \beta_3 X_3 + U\ \ (\beta_3 \neq 0). \qquad [2]$$

BE.1 $$\mathrm{E}[\hat{\beta}_{2\mathbf{W}}] = \beta_2\,,\ \mathrm{Var}[\hat{\beta}_{2\mathbf{W}}] = \frac{\sigma^2}{\mathrm{SCT}_2 \times (1 - R_2^2)}\,. \qquad [3]$$

ME.1 $$\mathrm{E}[\hat{\beta}^*_{2\mathbf{W}}] = \beta_2 + \beta_3\,\frac{\widehat{\mathrm{cov}}[\mathbf{X}_2, \mathbf{X}_3]}{\widehat{\mathrm{var}}[\mathbf{X}_2]} \neq \beta_2\,,\ \mathrm{Var}[\hat{\beta}^*_{2\mathbf{W}}] = \frac{\sigma^2}{\mathrm{SCT}_2} \leq \mathrm{Var}[\hat{\beta}_{2\mathbf{W}}]. \qquad [4]$$

SESGO POR OMISIÓN DE UNA VARIABLE EXPLICATIVA RELEVANTE

$$\hat{\beta}^*_{2\mathbf{W}} = \frac{\sum_{i=1}^N (X_{i2}-\bar{X}_2)(Y_i-\bar{Y})}{\sum_{i=1}^N (X_{i2}-\bar{X}_2)^2} = \frac{\sum_{i=1}^N (X_{i2}-\bar{X}_2)Y_i}{\sum_{i=1}^N (X_{i2}-\bar{X}_2)^2} = \frac{\sum_{i=1}^N (X_{i2}-\bar{X}_2)(\beta_1+\beta_2 X_{i2}+\beta_3 X_{i3}+U_i)}{\sum_{i=1}^N (X_{i2}-\bar{X}_2)^2}$$

$$= \frac{1}{\sum_{i=1}^N (X_{i2}-\bar{X}_2)^2}\left[\beta_1 \overbrace{\sum_{i=1}^N (X_{i2}-\bar{X}_2)}^{0} + \beta_2 \overbrace{\sum_{i=1}^N (X_{i2}-\bar{X}_2)X_{i2}}^{\sum_{i=1}^N (X_{i2}-\bar{X}_2)^2} + \right.$$

$$\left. + \beta_3 \underbrace{\sum_{i=1}^N (X_{i2}-\bar{X}_2)X_{i3}}_{\sum_{i=1}^N (X_{i2}-\bar{X}_2)(X_{i3}-\bar{X}_3)} + \sum_{i=1}^N (X_{i2}-\bar{X}_2)U_i \right]$$

$$= \beta_2 + \beta_3\,\frac{\widehat{\mathrm{cov}}[\mathbf{X}_2, \mathbf{X}_3]}{\widehat{\mathrm{var}}[\mathbf{X}_2]} + \frac{\sum_{i=1}^N (X_{i2}-\bar{X}_2)U_i}{\sum_{i=1}^N (X_{i2}-\bar{X}_2)^2} \Rightarrow \mathrm{E}[\hat{\beta}^*_{2\mathbf{W}}] - \beta_2 = \beta_3\,\frac{\widehat{\mathrm{cov}}[\mathbf{X}_2, \mathbf{X}_3]}{\widehat{\mathrm{var}}[\mathbf{X}_2]}\,.$$

**Observación:** En el Apéndice se consideran los efectos de la omisión del término constante.

## PARTE 2 - INCLUSIÓN DE VARIABLES EXPLICATIVAS IRRELEVANTES

ME.2 $$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + U\,,\ \text{con}\ \beta_3 = 0\,. \qquad [5]$$

BE.2 $$Y = \beta_1 + \beta_2 X_2 + U\,. \qquad [6]$$

ME.2 $$\mathrm{E}[\hat{\beta}_{2\mathbf{W}}] = \beta_2\,,\ \mathrm{E}[\hat{\beta}_{3\mathbf{W}}] = \beta_3 = 0,\ \mathrm{Var}[\hat{\beta}_{2\mathbf{W}}] = \frac{\sigma^2}{\mathrm{SCT}_2 \times (1 - R_2^2)}\,. \qquad [7]$$

BE.2 $$\mathrm{E}[\hat{\beta}^*_{2\mathbf{W}}] = \beta_2\,,\ \mathrm{Var}[\hat{\beta}^*_{2\mathbf{W}}] = \frac{\sigma^2}{\mathrm{SCT}_2} \leq \mathrm{Var}[\hat{\beta}_{2\mathbf{W}}]. \qquad [8]$$

RELACIÓN CON MÍNIMOS CUADRADOS RESTRINGIDOS (SECCIÓN 2.3 - APÉNDICE 1)

[2] es [1] restringido por $\beta_3 = 0$ FALSA, mientras que [6] es [5] restringido por $\beta_3 = 0$ CIERTA.

**EJEMPLO PARTE 1 - SC00-NOTAS.WF1**

BE.1

$$\begin{array}{ccccccc} \widehat{NF} & = & 2.270 & + & 0.140\,HA & + & 1.413\,HF, \\ & & (0.169) & & (0.112) & & (0.091) \\ & & [0.000] & & [0.213] & & [0.000] \end{array}$$

$$N = 200,\ \hat{\sigma}^2 = 1.267,\ R^2_{HA} = 0.602.$$

ME.1

$$\begin{array}{ccccc} \widehat{NF} & = & 2.469 & + & 1.491\,HA, \\ & & (0.250) & & (0.105) \\ & & [0.000] & & [0.000] \end{array}$$

$$N = 200,\ \hat{\sigma}^2 = 2.795.$$

**EJEMPLO PARTE 2 - SC00-NOTAS.WF1**

ME.2

$$\begin{array}{ccccccc} \widehat{NF} & = & 2.270 & + & 1.413\,HF & + & 0.140\,HA, \\ & & (0.169) & & (0.091) & & (0.112) \\ & & [0.000] & & [0.000] & & [0.213] \end{array}$$

$$N = 200,\ \hat{\sigma}^2 = 1.267,\ R^2_{HF} = 0.602.$$

BE.2

$$\begin{array}{ccccc} \widehat{NF} & = & 2.375 & + & 1.502\,HF, \\ & & (0.147) & & (0.058) \\ & & [0.000] & & [0.000] \end{array}$$

$$N = 200,\ \hat{\sigma}^2 = 1.271.$$

## PARTE 3 - REVISIÓN DEL CONTENIDO DEL CONJUNTO DE VARIABLES EXPLICATIVAS

El sesgo asociado con la omisión de variables explicativas relevantes y la pérdida de precisión asociada con la inclusión de variables explicativas irrelevantes, ponen de relieve la importancia de considerar y, en su caso, revisar cuidadosamente el contenido inicial del conjunto de variables explicativas en un modelo estimado.

En la práctica, la señal más evidente del sesgo asociado con la omisión de alguna/s variable/s explicativa/s relevante/s suele ser la obtención de estimaciones con valores y/o signos chocantes (distintos de lo razonablemente esperado).

Por otro lado, algunos contrastes de significación pueden ayudar a decidir si una o varias variables explicativas son relevantes o no en un modelo, aunque la posible escasa

**significación estadística** de uno o varios parámetros no debe llevar directamente a la eliminación de sus variables respectivas. En primer lugar, porque la significación estadística depende del tamaño muestral y de la posible presencia de **multicolinealidad** [Wooldridge (2020): Apartado 3-4A] [$*$]. Y, sobre todo, porque la **significación práctica** de una variable explicativa puede ser tan relevante o más que su significación estadística [Wooldridge (2020): Apartado 4-2F] [$*$].

Sin olvidar las consideraciones anteriores, también se puede recurrir a ciertos **criterios de especificación** como ayuda para seleccionar un conjunto adecuado de variables explicativas.

## PARTE 4 - CRITERIOS DE INFORMACIÓN - EVALUACIÓN DE PREVISIONES - RESET

*Akaike Information Criterion:*

$$\text{AIC} = -\tfrac{2}{N}\, l_* + \tfrac{2K}{N} \approx \ln\left(\tfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N}\right) + \tfrac{2K}{N}\,. \qquad [9]$$

*Schwarz (Bayesian) Information Criterion:*

$$\text{BIC} = -\tfrac{2}{N}\, l_* + \tfrac{\ln(N)K}{N} \approx \ln\left(\tfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N}\right) + \tfrac{\ln(N)K}{N}\,. \qquad [10]$$

En AIC y BIC, $l_*$ es el ln de la "función de verosimilitud" bajo HC1-HC5 [lo que significa, simplemente, que $-(2/N)l_*$ y $\ln(\hat{\mathbf{u}}'\hat{\mathbf{u}}/N)$ difieren tan sólo en la constante $1+\ln(2\pi)$].

$$\begin{array}{c} \text{Muestra} \\ \text{disponible} \\ \text{completa} \end{array} \left[ \begin{array}{l} \begin{array}{c} \text{Muestra} \\ \text{de} \\ \text{modelización} \end{array} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \longrightarrow \text{Modelo} \longrightarrow \text{Previsiones: } \hat{y}_{N+1}, \hat{y}_{N+2}, \ldots, \hat{y}_{N+P}\,. \\ \begin{array}{c} \text{Muestra} \\ \text{de} \\ \text{previsión} \end{array} \begin{bmatrix} y_{N+1} & - & \hat{y}_{N+1} & = & \hat{e}_1 \\ y_{N+2} & - & \hat{y}_{N+2} & = & \hat{e}_2 \\ \vdots & & \vdots & & \vdots \\ y_{N+P} & - & \hat{y}_{N+P} & = & \hat{e}_P \end{bmatrix} \longrightarrow \text{Errores: } \hat{e}_1, \hat{e}_2, \ldots, \hat{e}_P\,. \end{array} \right.$$

*Root Mean Squared Error:*

$$\text{RMSE} = \sqrt{\tfrac{1}{P}\textstyle\sum_{f=1}^{P} \hat{e}_f^2}\,. \qquad [11]$$

*Mean Absolute Error:*

$$\text{MAE} = \tfrac{1}{P}\textstyle\sum_{f=1}^{P} \left|\hat{e}_f\right|. \qquad [12]$$

*Mean Absolute Percentage Error:*

$$\text{MAPE} = 100 \times \left(\tfrac{1}{P}\textstyle\sum_{f=1}^{P} \left|\tfrac{\hat{e}_f}{y_{N+f}}\right|\right). \qquad [13]$$

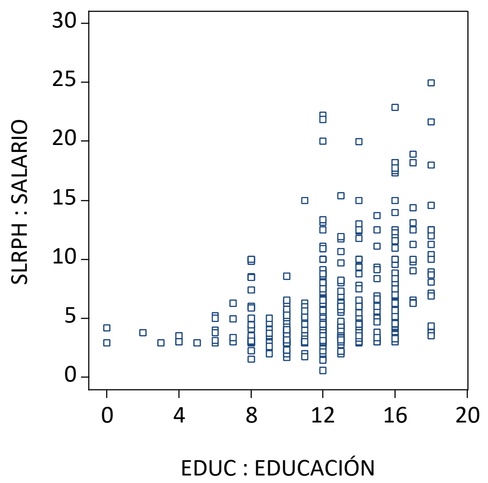
*RESET (REgression Specification Error Test):*

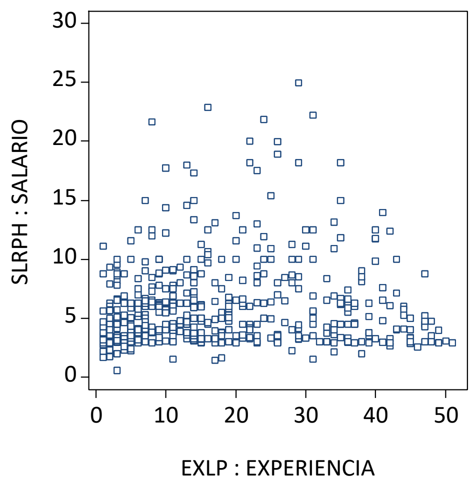
Wooldridge (2020): Apartado 9-1A [$*$].

## EJEMPLO PARTE 4 - SC03-SALARIOS3.WF1

### DATOS

Salario medio anual (dólares por hora), educación (años) y experiencia laboral (años) de 526 personas trabajadoras entrevistadas en 1976 (*Current Population Survey - U.S. Census Bureau*).




### ALGUNOS MODELOS CONSIDERADOS EN LA SECCIÓN 2.1: PARTE 6

**RLS** $\ln SLRPH = \beta_1 + \beta_2 EDUC + U$

**RLM.1** $\ln SLRPH = \beta_1 + \beta_2 EDUC + \beta_3 EXLP + U$

**RLM.2** $\ln SLRPH = \beta_1 + \beta_2 EDUC + \beta_3 EXLP + \beta_4 EXLP^2 + U$

**RLM.3** $\ln SLRPH = \beta_1 + \beta_2 EDUC + \beta_3 EXLP + \beta_4 EXLP^2 + \beta_5 (EDUC \times EXLP) + U$

### CRITERIOS DE ESPECIFICACIÓN

| | $\bar{R}^2$ | AIC | BIC | RMSE | MAE | MAPE | RESET |
|---|---|---|---|---|---|---|---|
| **RLS** | 0.1248 | 1.3474 | 1.3644 | 0.59 | 0.50 | 21.90 % | 0.002 |
| **RLM.1** | 0.1853 | 1.2777 | 1.3032 | 0.55 | 0.46 | 20.26 % | 0.038 |
| **RLM.2** | 0.2361 | 1.2154 | **1.2494** | 0.52 | 0.44 | 19.31 % | **0.103** |
| **RLM.3** | **0.2435** | **1.2076** | 1.2501 | **0.50** | **0.41** | **18.73 %** | **0.206** |

Modelos estimados con las N = 495 primeras observaciones.

RMSE, MAE y MAPE calculados a partir de previsiones de Y = ln SLRPH para las observaciones 496-526 (P = 31 observaciones). P-valores RESET calculados con $\hat{Y}^2$ e $\hat{Y}^3$.

Operaciones con EViews en pp. 48-50 de la guía *Introducción al Uso de EViews 4.1*.

[∗] Material adicional **obligatorio** para esta Sección 2.4 en Wooldridge (2020): Apartados 3-4A (multicolinealidad), 4-2F (significación práctica vs. significación estadística) y 9-1A (RESET). Material complementario **opcional** en Hill, Griffiths, Lim (2018): Secciones 6.3 - 6.6.

## APÉNDICE - EFECTOS DE LA OMISIÓN DEL TÉRMINO CONSTANTE

BE.3 $$Y = \beta_1 + \beta_2 X_2 + U\,,\ \text{con}\ \beta_1 \neq 0\,. \qquad \text{[A1]}$$

ME.3 $$Y = \beta_2 X_2 + U^*,\ \text{con}\ U^* = \beta_1 + U\ \ (\beta_1 \neq 0). \qquad \text{[A2]}$$

BE.3 $$\mathrm{E}[\hat{\beta}_{2\mathbf{W}}] = \beta_2\,, \mathrm{Var}[\hat{\beta}_{2\mathbf{W}}] = \frac{\sigma^2}{\mathrm{SCT}_2} = \frac{\sigma^2}{\mathbf{X}_2'\mathbf{X}_2 - N\bar{X}_2^2}\,. \qquad \text{[A3]}$$

ME.3 $$\mathrm{E}[\hat{\beta}^*_{2\mathbf{W}}] = \beta_2 + \beta_1 \frac{N\bar{X}_2}{\mathbf{X}_2'\mathbf{X}_2} \neq \beta_2\,, \mathrm{Var}[\hat{\beta}^*_{2\mathbf{W}}] = \frac{\sigma^2}{\mathbf{X}_2'\mathbf{X}_2} \leq \mathrm{Var}[\hat{\beta}_{2\mathbf{W}}]. \qquad \text{[A4]}$$

### EJEMPLO - SC00-NOTAS.WF1

BE.3

$$\begin{array}{ccccc} \widehat{NF} & = & 2.375 & + & 1.502\,HF, \\ & & (0.147) & & (0.058) \\ & & [0.000] & & [0.000] \end{array}$$

$$N = 200,\ \hat{\sigma}^2 = 1.271,\ \overline{HF} = 2.137\ [0.000].$$

ME.3

$$\begin{array}{ccc} \widehat{NF} & = & 2.286\,HF, \\ & & (0.048) \\ & & [0.000] \end{array}$$

$$N = 200,\ \hat{\sigma}^2 = 2.935.$$