ECONOMETRÍA

JOSÉ ALBERTO MAURICIO

2

REGRESIÓN LINEAL MÚLTIPLE I

2.1 ESPECIFICACIÓN I

Departamento de Análisis Económico y Economía Cuantitativa
Universidad Complutense de Madrid

EctrGr-JAM-2-1.pdf

ucm.randomshock.com/ectrgr

Versión 2.4 - Enero 2024

MATERIAL AUXILIAR Y COMPLEMENTARIO

El instrumental matemático que se utiliza en esta Sección 2.1 tiene que ver con variaciones (diferencias) absolutas y relativas, porcentajes, elasticidades, funciones lineales, cuadráticas, exponenciales, logarítmicas y recíprocas, y derivadas parciales. Se puede repasar todo ello en Hill, Griffiths, Lim (2018): Secciones A.1 - A.3, y en Wooldridge (2020): Secciones A-2 - A-5. En la PARTE 3 de esta Sección 2.1 se ofrece un breve resumen sobre variaciones y elasticidades.

Al final de esta Sección 2.1 se propone la realización de la parte teórica de las preguntas del Ejercicio 1 disponible en la página de la asignatura en Internet.

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 - INTRODUCCIÓN

La **especificación inicial** de un modelo RLM requiere considerar tres cuestiones:

① La elección de las **variables explicativas** (Tema 1).

② La elección de la **forma funcional** de la relación entre la variable dependiente y las variables explicativas (Tema 2: Sección 2.1).

③ El planteamiento de las **hipótesis** que garantizan unas buenas propiedades para los **métodos econométricos** que serán utilizados posteriormente (Tema 2: Sección 2.3).

Las cuestiones ①-② suelen resolverse mediante algún tipo de razonamiento lógico basado (quizás) en algún **modelo teórico** y (fundamentalmente) en un análisis inicial detallado de las **características muestrales** de los **datos** disponibles sobre todas las variables. También son importantes en ①-② la **intuición**, la **experiencia** y el **buen juicio** del analista.

La cuestión ③ suele resolverse planteando de manera tentativa (y, por lo tanto, revisable) sobre el resultado de ①-② un conjunto de hipótesis que justifican el empleo de métodos econométricos lo más sencillos posible y con buenas propiedades estadísticas.

PARTE 2 - FORMA FUNCIONAL

Un modelo RLM del tipo

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_K X_K + U \qquad [1]$$

es **lineal** porque los **parámetros** $\beta_1, \beta_2, \beta_3, ..., \beta_K$ figuran en su lado derecho de forma lineal, es decir, porque cada parámetro figura, a lo sumo, multiplicado por un término que no depende de ningún parámetro del modelo. No obstante, $Y, X_2, X_3, ..., X_K$ en [1] pueden representar variables **originales** de interés, o bien cualesquiera **transformaciones** lineales o no lineales de las mismas. Por ejemplo, si F y P representan variables originales, los modelos M1-M6 que figuran más abajo son todos casos particulares de [1] (definiendo adecuadamente Y, X_j y K en cada caso), de manera que M1-M6 son todos modelos lineales con respecto a sus parámetros. Al mismo tiempo, M2-M6 son modelos **no** lineales con respecto a las variables originales F y P, lo que permite considerar relaciones entre dichas variables (Figuras 1-6) que quizás resuman la información contenida en los datos disponibles mejor que una mera relación lineal como la que ofrece el modelo M1.

M1 $$F = \beta_1 + \beta_2 P + U.$$

M2 $$F = \beta_1 + \beta_2 P + \beta_3 P^2 + U.$$

M3 $$\ln F = \beta_1 + \beta_2 P + U, \text{ o bien } F = \exp[\beta_1 + \beta_2 P + U].$$

M4 $$F = \beta_1 + \beta_2 \ln P + U.$$

M5 $$\ln F = \beta_1 + \beta_2 \ln P + U, \text{ o bien } F = \exp[\beta_1 + \beta_2 \ln P + U].$$

M6 $$F = \beta_1 + \beta_2 \tfrac{1}{P} + U.$$

FIGURA 1

M1: $\beta_1+\beta_2 P$

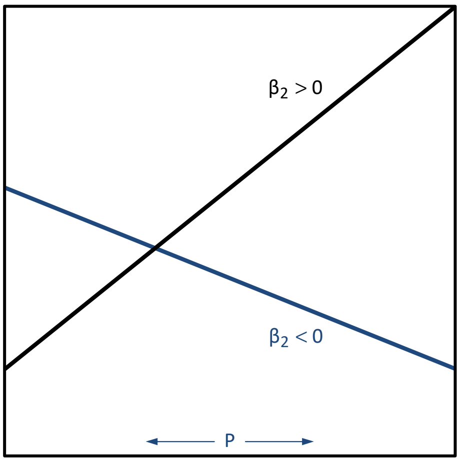


FIGURA 2

M2: $\beta_1+\beta_2 P+\beta_3 P^2$

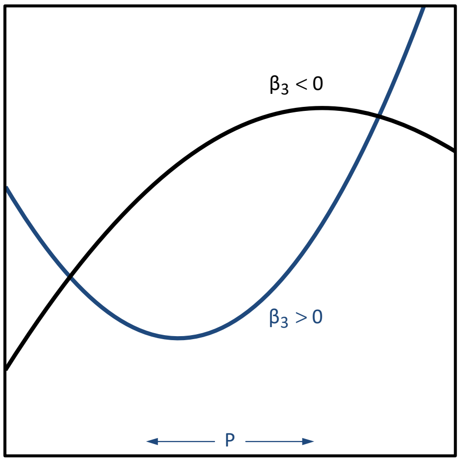


FIGURA 3

M3: $\exp[\beta_1+\beta_2 P]$

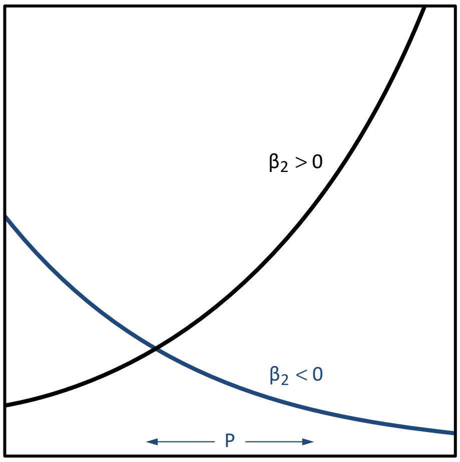


FIGURA 4

M4: $\beta_1+\beta_2 \ln P$

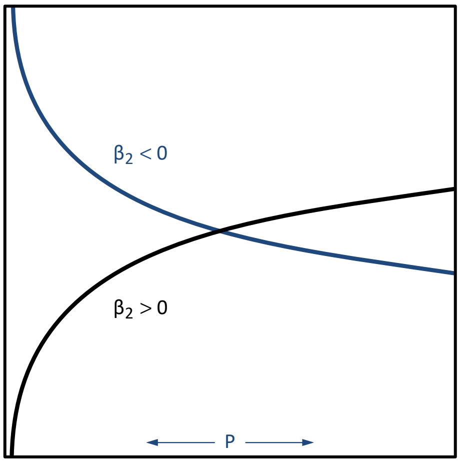


FIGURA 5

M5: $\exp[\beta_1+\beta_2 \ln P]$

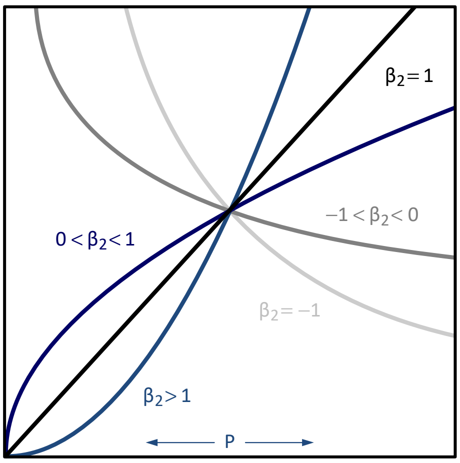
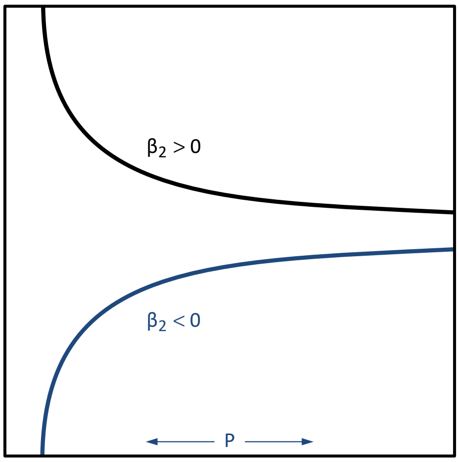


FIGURA 6

M6: $\beta_1+\beta_2 \frac{1}{P}$



Las posibilidades que ofrecen por separado modelos como M1-M6 se pueden incluir al mismo tiempo en un único modelo RLM (como en el ejemplo de la PARTE 6), lo que permite especificar la **forma funcional** de la relación entre la variable dependiente y la/s variable/s explicativa/s originales con mucha más flexibilidad de lo que sugiere la expresión [1] a primera vista.

PARTE 3 - VARIACIONES (DIFERENCIAS) Y ELASTICIDADES

Variación absoluta: $\Delta Y = Y_1 - Y_0$, donde Y_1, Y_0 son dos valores cualesquiera de Y.

Variación relativa o proporcional (tasa de variación):

Convencional: $\frac{\Delta Y}{Y_0} = \frac{Y_1 - Y_0}{Y_0}$.

Logarítmica: $\Delta \ln Y = \ln Y_1 - \ln Y_0 = \ln \frac{Y_1}{Y_0} = \ln\left[1 + \frac{\Delta Y}{Y_0}\right] \approx \frac{\Delta Y}{Y_0}$.

Observación 1: La relación $\Delta \ln Y = \ln[1 + \Delta Y / Y_0]$ equivale a que $\Delta Y / Y_0 = \exp[\Delta \ln Y] - 1$.

Observación 2: La relación $\Delta \ln Y \approx \Delta Y / Y_0$ se debe a que una aproximación de Taylor de primer orden (lineal) al valor de la función $\ln(1 + x)$ para valores de x próximos a x_* proporciona lo siguiente:

$$\ln(1+x) \approx \ln(1+x_*) + \frac{1}{1+x_*} \times [(1+x) - (1+x_*)] = \ln(1+x_*) + \frac{x - x_*}{1+x_*} = x \text{ cuando } x_* = 0,$$

de manera que $\Delta \ln Y = \ln[1 + \Delta Y / Y_0] \approx \Delta Y / Y_0$ cuando $\Delta Y / Y_0$ es una cantidad “pequeña”.

Variación porcentual: $\%\Delta Y = 100 \frac{\Delta Y}{Y_0} \approx 100 \Delta \ln Y$.

Elasticidad: $l_{YX} = \frac{\%\Delta Y}{\%\Delta X} \Rightarrow \%\Delta Y = l_{YX} \times \%\Delta X$, de manera que l_{YX} es la variación porcentual que tiene lugar en Y cuando X varía un 1%.

Semielasticidad: $l'_{YX} = \frac{\%\Delta Y}{\Delta X} \Rightarrow \%\Delta Y = l'_{YX} \times \Delta X$, de manera que l'_{YX} es la variación porcentual que tiene lugar en Y cuando X varía una unidad.

PARTE 4 - EFECTOS CAUSALES Y PARÁMETROS

La forma en la que las variables originales de interés figuran en un modelo de regresión lineal tiene consecuencias importantes sobre los tipos de efectos causales implícitos en el modelo y sobre el significado de sus parámetros, como se ilustra a continuación en relación con los modelos M1-M6 de la PARTE 2.

M1:

$$F = \beta_1 + \beta_2 P + U, \quad [1.1]$$

$$\frac{\partial F}{\partial P} = \beta_2 \text{ (constante)} \Rightarrow \frac{\Delta F}{\Delta P} = \beta_2 \Rightarrow \Delta F = \beta_2 \Delta P, \quad [1.2]$$

$$\beta_2 = \Delta F \text{ cuando } \Delta P = 1. \quad [1.3]$$

M2:

$$F = \beta_1 + \beta_2 P + \beta_3 P^2 + U, \quad [2.1]$$

$$\frac{\partial F}{\partial P} = \beta_2 + 2\beta_3 P \Rightarrow \frac{\Delta F}{\Delta P} \approx \beta_2 + 2\beta_3 P \Rightarrow \Delta F \approx (\beta_2 + 2\beta_3 P) \times \Delta P, \qquad [2.2]$$

$$\beta_2 + 2\beta_3 P \approx \Delta F \text{ cuando } \Delta P = 1. \qquad [2.3]$$

Adicionalmente,

$$(\partial F/\partial P) = 0 \Leftrightarrow P^* = -\frac{\beta_2}{2\beta_3},\ (\partial^2 F/\partial P^2) = 2\beta_3 \begin{cases} < 0\ (\Leftrightarrow \beta_3 < 0) \Rightarrow P^* \text{ máximo.} \\ > 0\ (\Leftrightarrow \beta_3 > 0) \Rightarrow P^* \text{ mínimo.} \end{cases}$$

Por otro lado, tanto F como $\ln F$ tienen su máximo o su mínimo con respecto a P en el mismo valor P^* (porque la función ln es una función monótona estrictamente creciente).

M3
$$\ln F = \beta_1 + \beta_2 P + U, \text{ o bien } F = \exp[\beta_1 + \beta_2 P + U]. \qquad [3.1]$$

$$\frac{\partial F}{\partial P} = \beta_2 F \Rightarrow \frac{\Delta F}{\Delta P} \approx \beta_2 F \Rightarrow \frac{\Delta F}{F} \approx \beta_2 \Delta P \Rightarrow \%\Delta F \approx (100\beta_2)\Delta P, \qquad [3.2]$$

$$100\beta_2 \approx \%\Delta F \text{ cuando } \Delta P = 1 \text{ (semielasticidad).} \qquad [3.3]$$

M4:
$$F = \beta_1 + \beta_2 \ln P + U, \qquad [4.1]$$

$$\frac{\partial F}{\partial P} = \beta_2 \frac{1}{P} \Rightarrow \frac{\Delta F}{\Delta P} \approx \beta_2 \frac{1}{P} \Rightarrow \Delta F \approx \beta_2 \frac{\Delta P}{P} \Rightarrow \Delta F \approx \left(\frac{\beta_2}{100}\right)\%\Delta P, \qquad [4.2]$$

$$\left(\frac{\beta_2}{100}\right) \approx \Delta F \text{ cuando } \%\Delta P = 1\%. \qquad [4.3]$$

M5
$$\ln F = \beta_1 + \beta_2 \ln P + U, \text{ o bien } F = \exp[\beta_1 + \beta_2 \ln P + U]. \qquad [5.1]$$

$$\frac{\partial F}{\partial P} = \beta_2 \frac{F}{P} \Rightarrow \frac{\Delta F}{\Delta P} \approx \beta_2 \frac{F}{P} \Rightarrow \frac{\Delta F}{F} \approx \beta_2 \frac{\Delta P}{P} \Rightarrow \%\Delta F \approx \beta_2 \%\Delta P, \qquad [5.2]$$

$$\beta_2 \approx \%\Delta F \text{ cuando } \%\Delta P = 1\% \text{ (elasticidad).} \qquad [5.3]$$

M6:
$$F = \beta_1 + \beta_2 \tfrac{1}{P} + U, \qquad [6.1]$$

$$\frac{\partial F}{\partial P} = \frac{-\beta_2}{P^2} \Rightarrow \frac{\Delta F}{\Delta P} \approx \frac{-\beta_2}{P^2} \Rightarrow \Delta F \approx -\beta_2 \frac{\Delta P}{P^2} \Rightarrow \Delta F \approx \left(-\frac{\beta_2}{100P}\right)\%\Delta P, \qquad [6.2]$$

$$\left(-\frac{\beta_2}{100P}\right) \approx \Delta F \text{ cuando } \%\Delta P = 1\%. \qquad [6.3]$$

PARTE 5 - REGRESIÓN NO LINEAL

Modelos como $Y = \beta_1 X_1 + \beta_1^2 X_2 + U$, $Q = \beta_0 L^{\beta_2} K^{\beta_3} + U$, son **no lineales** con respecto a sus parámetros (ninguno de ellos es un caso particular de [1]). Por otro lado, un modelo como $Q = \beta_0 L^{\beta_2} K^{\beta_3} \times \exp[U]$ es equivalente (tomando logaritmos en ambos lados) a $\ln Q = \beta_1 + \beta_2 \ln L + \beta_3 \ln K + U$ con $\beta_1 = \ln \beta_0$, que sí es lineal con respecto a los parámetros β_1, β_2 y β_3. Un modelo es no lineal cuando ni es lineal en su formulación original, ni se puede convertir en un modelo lineal mediante transformación alguna.

Observación 3: La manera (aditiva, multiplicativa, ...) en la que se incluye el término de error en un modelo tiene que ver, en general, con la cuestión ③ mencionada en la PARTE 1.

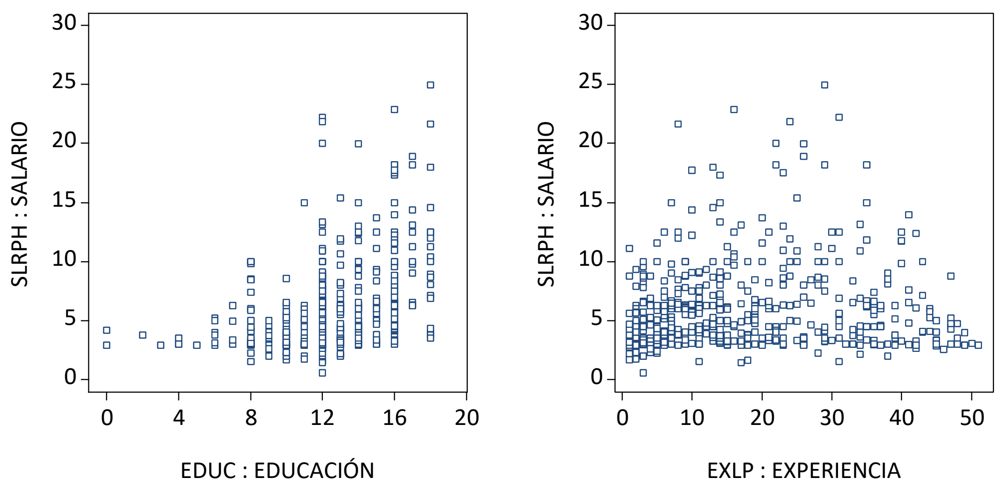
PARTE 6 - EJEMPLO

FIGURA 7

Datos de Sección Cruzada sobre Salarios, Educación y Experiencia

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*).

SC03-SALARIOS3.WF1



6.1 VARIABLES EXPLICATIVAS Y FORMA FUNCIONAL

El modelo más sencillo que puede considerarse para evaluar el efecto causal de la educación sobre el salario es un modelo RLS del tipo

$$SLRPH = \beta_1 + \beta_2 EDUC + U, \qquad [7]$$

aunque el panel izquierdo de la Figura 7 sugiere una posible relación parcial entre $SLRPH$ y $EDUC$ del tipo de la Figura 3 (con parámetro positivo):

RLS
$$\ln SLRPH = \beta_1 + \beta_2 EDUC + U. \qquad [8]$$

El problema fundamental de [8] reside en que si U contiene alguna influencia sobre el salario de un trabajador que esté relacionada con su educación (como, por ejemplo, su experiencia, o su habilidad o su capacidad para desempeñar adecuadamente su trabajo), entonces la posible independencia entre U y $EDUC$ no es justificable en [8]. En este sentido, un modelo RLM que incluye la experiencia como variable explicativa es

RLM.1
$$\ln SLRPH = \beta_1 + \beta_2 EDUC + \beta_3 EXLP + U, \qquad [9]$$

aunque el panel derecho de la Figura 7 sugiere una posible relación parcial entre $SLRPH$ y $EXLP$ del tipo de la Figura 2 (con parámetro negativo), lo que unido a [8] sugiere que un modelo RLM quizás más adecuado que [9] podría ser en este caso

RLM.2
$$\ln SLRPH = \beta_1 + \beta_2 EDUC + \beta_3 EXLP + \beta_4 EXLP^2 + U. \qquad [10]$$

La ventaja fundamental de [10] frente a [8] reside en que en [10] es seguro que U no contiene ni a $EXLP$ ni a $EXLP^2$, por lo que la posible relación entre U y $EDUC$ es seguramente notablemente menor en [10] que en [8]. En todo caso, para poder evaluar fiablemente tanto β_2 como β_3 y β_4 en el modelo RLM [10], U debe ser independiente no sólo de la educación, sino también de la experiencia. Aunque la presencia en U de factores no observables (como la habilidad para desempeñar un trabajo) siempre es problemática en este sentido, también es cierto que a [10] se le pueden añadir otras influencias observables relacionadas con la educación, con la experiencia o con ambas. En esta línea, una extensión razonable de [10] podría ser la siguiente:

RLM.3

$$\begin{aligned} \ln SLRPH &= \beta_1 + \beta_2 EDUC + \beta_3 EXLP + \beta_4 EXLP^2 \\ &\quad + \beta_5 (EDUC \times EXLP) + U. \end{aligned} \quad [11]$$

Si β_5 en [11] es distinto de cero, entonces U en [10] no puede ser independiente ni de la educación ni de la experiencia. Quizás U en [11] tampoco lo es, aunque seguro que el motivo no sería en ese caso la ausencia de un término asociado con el producto de la educación por la experiencia.

El modelo [11] se puede justificar dentro de un marco teórico como el llamado Modelo del Capital Humano [en inglés, *Human Capital Model*; ver, por ejemplo, el Capítulo 5 en Berndt, E.R. (1991), *The Practice of Econometrics*, Addison-Wesley]. No obstante, todo lo expuesto en este Apartado 6.1 sugiere que un modelo RLM como [11] puede surgir de manera bastante natural sin más que combinar algo de intuición y sentido común con lo que se puede observar acerca de las posibles relaciones (parciales) entre los datos sobre $SLRPH$ y $EDUC$, por un lado, y sobre $SLRPH$ y $EXLP$, por otro, en la Figura 7.

TABLA 1

Algunos Modelos de Regresión Lineal para los Datos de la Figura 7

RLS: $\overbrace{\ln \mathrm{SLRPH}}^{Y} = \beta_1 + \beta_2 \overbrace{\mathrm{EDUC}}^{X_2} + U.$

RLM.1: $\overbrace{\ln \mathrm{SLRPH}}^{Y} = \beta_1 + \beta_2 \overbrace{\mathrm{EDUC}}^{X_2} + \beta_3 \overbrace{\mathrm{EXLP}}^{X_3} + U.$

RLM.2: $\overbrace{\ln \mathrm{SLRPH}}^{Y} = \beta_1 + \beta_2 \overbrace{\mathrm{EDUC}}^{X_2} + \beta_3 \overbrace{\mathrm{EXLP}}^{X_3} + \beta_4 \overbrace{\mathrm{EXLP}^2}^{X_4} + U.$

RLM.3: $\overbrace{\ln \mathrm{SLRPH}}^{Y} = \beta_1 + \beta_2 \overbrace{\mathrm{EDUC}}^{X_2} + \beta_3 \overbrace{\mathrm{EXLP}}^{X_3} + \beta_4 \overbrace{\mathrm{EXLP}^2}^{X_4}$

$+ \beta_5 \overbrace{\mathrm{EDUC} \times \mathrm{EXLP}}^{X_5} + U.$

EJ1 - PARTE TEÓRICA

6.2 EFECTOS CAUSALES Y PARÁMETROS

El modelo RLS [8] implica que

$$\frac{\partial SLRPH}{\partial EDUC} = \beta_2 \times SLRPH, \qquad [12]$$

por lo que

$$\%\Delta SLRPH \approx 100\beta_2 \times \Delta EDUC, \qquad [13]$$

de manera que $100\beta_2$ es la **semielasticidad** (aproximada) del salario con respecto a los años de educación.

El modelo RLM.2 [10] implica, adicionalmente a [12]-[13], que

$$\frac{\partial SLRPH}{\partial EXLP} = (\beta_3 + 2\beta_4 EXLP) \times SLRPH, \qquad [14]$$

por lo que

$$\%\Delta SLRPH \approx 100(\beta_3 + 2\beta_4 EXLP) \times \Delta EXLP, \qquad [15]$$

de manera que $100(\beta_3 + \beta_4 EXLP)$ es la **semielasticidad** (aproximada) del salario con respecto a los años de experiencia. A diferencia de lo que ocurre con la semielasticidad con respecto a la educación (que es constante), la semielasticidad con respecto a la experiencia depende de los años de experiencia. Si $\beta_4 < 0$ (como sugiere el panel derecho de la Figura 7), entonces dicha semielasticidad disminuye con los años experiencia, que maximizan el salario (o su logaritmo) cuando $EXLP^* = -\beta_3/(2\beta_4)$.

El modelo RLM.3 [11] es una extensión del modelo RLM.2 [10] que implica, por un lado, que

$$\frac{\partial SLRPH}{\partial EDUC} = (\beta_2 + \beta_5 EXLP) \times SLRPH, \qquad [16]$$

por lo que

$$\%\Delta SLRPH \approx 100(\beta_2 + \beta_5 EXLP) \times \Delta EDUC, \qquad [17]$$

y, por otro lado, que

$$\frac{\partial SLRPH}{\partial EXLP} = (\beta_3 + 2\beta_4 EXLP + \beta_5 EDUC) \times SLRPH, \qquad [18]$$

por lo que

$$\%\Delta SLRPH \approx 100(\beta_3 + 2\beta_4 EXLP + \beta_5 EDUC) \times \Delta EXLP. \qquad [19]$$

En este caso, la semielasticidad con respecto a la educación en [17] depende de los años de experiencia, y la semielasticidad con respecto a la experiencia en [19] depende (además de los años de experiencia) de los años de educación, en ambos casos a través del parámetro β_5 (cuyo signo determina el sentido de dichas dependencias). Por este motivo, suele decirse que una variable explicativa como $EDUC \times EXLP$ (el producto de dos variables originales de interés) en el modelo [11] es una "variable de interacción".

Las operaciones realizadas y las conclusiones obtenidas en esta página son del mismo estilo (algunas son, de hecho, las mismas) que las necesarias para resolver el ejercicio que se indica al final de la Tabla 1 de la página anterior.