Hoja de Ejercicios 5
Errores de especificación

Nota: En aquellos ejercicios en los que se incluyen estimaciones y referencia al archivo de datos utilizado, el estudiante debería comprobar los resultados obtenidos en Gretl.

1. Supongamos que el modelo
$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$
verifica que $E(\varepsilon|X_1) = 0$ para todo X_1, pero que estimamos por MCO la regresión sin constante de Y sobre X_1, esto es, $\widetilde{Y} = \delta_1 X_1$. ¿Es $\widehat{\delta_1}$ un estimador consistente de β_1? Justifique la respuesta. ¿Qué concluiría entonces acerca de la posibilidad de excluir el término constante de la regresión?

2. Suponga que el verdadero modelo de regresión es de la forma
$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i2}^2 + u_i$$
pero se estiman los coeficientes de la regresión
$$Y_i = \beta_1^* + \beta_2^* X_{i2} + \varepsilon_i$$
   a) ¿Es consistente el estimador MCO de β_2^*?.
   b) ¿Cuándo el sesgo es positivo y cuándo negativo?
   c) ¿Se hubieran presentado los mismos problemas si la regresión inicial hubiera sido
$$Y_i = \beta_1 + \beta_2 \ln X_{i2} + \beta_3 \ln X_{i2}^2 + v_i?$$

3. [Este ejercico se corresponde con el ejercicio 9.4 del Wooldridge] La siguiente ecuación caracteriza las horas semanales de televisión que ve un niño en función de su edad, educación de la madre, educación del padre y número de hermanos:
$$tvhours^* = \beta_0 + \beta_1 age + \beta_2 age^2 + \beta_3 motheduc + \beta_4 fatheduc + \beta_5 sibs + u.$$
Cabe sospechar que $tvhours^*$ esté medida con error en nuestra encuesta. Si $tvhours$ es el número de horas de televisión declaradas por semana:
   a) ¿Qué requiere el supuesto clásico de errores en variables en esta aplicación?
   b) ¿Es verosímil que se cumpla el supuesto clásico de errores en variables en este caso? Justifique su respuesta.

4. El fichero `WAGE2.GDT` contiene datos usados en el artículo de Blackburn y Neumark (QJE n^o 107, pags. 1421-1436, 1992). Incluye información sobre salarios mensuales, educación y ciertas variables demográficas y cocientes IQ de 935 hombres en 1980. Como forma de evitar el sesgo por omisión de habilidad, añadimos IQ a la ecuación habitual de $\log(wage)$:
$$\begin{aligned} \log(wage) &= \beta_0 + \beta_1 educ + \beta_2\ exper + \beta_3 tenure + \beta_4 married + \beta_5 south \\ &\quad + \beta_6 urban + \beta_7 black + u \end{aligned}$$
$$\begin{aligned} \log(wage) &= \beta_0 + \beta_1 educ + \beta_2\ exper + \beta_3 tenure + \beta_4 married + \beta_5 south \\ &\quad + \beta_6 urban + \beta_7 black + \beta_8 IQ + u \end{aligned}$$

a) Estima ambas ecuaciones y compara el retorno estimado a la educación. ¿Cuál es el signo de la correlación entre educación y la habilidad omitida y el del sesgo por omisión correspondiente en la primera ecuación?

b) Interpreta los coeficientes de *educ* e IQ. Compara el incremento salarial producido por 15 puntos más en el IQ con el de un año adicional de educación. Si la desviación típica poblacional de IQ es de 15, ¿qué puedes concluir sobre la importancia relativa de la habilidad y de la educación en los ingresos?

c) Compara los R^2 de ambos modelos.

d) ¿Qué puedes decir sobre las diferencias entre hombres negros y no negros a la luz de la estimación de ambos modelos?

e) Ahora incluimos la interacción entre *educ* y IQ en la ecuación para comprobar si el retorno de la educación es mayor para personas con mayor IQ :

$$\begin{aligned}\log(wage) &= \beta_0 + \beta_1 educ + \beta_2\ exper + \beta_3 tenure + \beta_4 married + \beta_5 south \\ &\quad + \beta_6 urban + \beta_7 black + \beta_8 IQ + \beta_9 educ \cdot IQ + u.\end{aligned}$$

¿Apoyan los datos esa conjetura? ¿Qué ocurre con la significatividad individual de *educ* y IQ en la nueva ecuación? ¿Qué modelo es más apropiado?

f) Usa la variable KWW (resultados del test "knowledge of the world of work") como proxy para habilidad en lugar de IQ. ¿Cuál es el retorno a la educación en este caso?

g) Ahora usa IQ y KWW conjuntamente como variables proxy. ¿Qué ocurre ahora con el retorno a la educación?

h) En el apartado anterior, ¿son IQ y KWW individualmente significativos? ¿Son conjuntamente significativos?

5. El gerente de un centro comercial esta interesado en conocer cómo influye la renta de sus clientes (R, expresada en euros) y su sexo ($S = 1$ si el cliente es un hombre y 0 si se trata de una mujer) en las compras (en cientos de euros) que realizan sus clientes (V). Para investigar el comportamiento de sus clientes utiliza el siguiente modelo:

$$E(V|\, R, S) = \beta_0 + \beta_1 \ln(R) + \beta_2 S$$

Para estimar el modelo anterior, el empresario pregunta a 528 clientes (seleccionados de manera aleatoria) su renta cada vez que realizaban una compra.

OLS, using observations 1–528
Dependent variable: V

	Coefficient	Std. Error	t-ratio	p-value
const	2,5210	0,0175	144,31	0,000
$\ln R$	0,0160	0,0014	11,25	0,000
S	−0,0640	0,0060	−10,61	0,000

Sum squared resid	14,6110	S.E. of regression	0,1670
R^2	0,2944	Adjusted R^2	0,2917

a) Si los clientes mienten al indicar cual es su renta anual, y denotando $\widehat{\beta}_{\cdot 1}$, $\widehat{\beta}_{\cdot 2}$ como los estimadores MCO de β_1, β_2, indique cuál de las siguientes afirmaciones es CIERTA:

i) $\widehat{\beta}_1$ será inconsistente y $\widehat{\beta}_2$ será consistente.
ii) $\widehat{\beta}_1$ será consistente y $\widehat{\beta}_2$ será inconsistente.
iii) En general, tanto $\widehat{\beta}_1$ como $\widehat{\beta}_2$ serán consistentes.
iv) En general, tanto $\widehat{\beta}_1$ como $\widehat{\beta}_2$ serán inconsistentes.

b) Indique cuál de las siguientes afirmaciones es cierta:

i) Cuanto mayor sea la varianza del error de medida respecto a la varianza de la renta, mayor será el sesgo de inconsistencia de $\widehat{\beta}_1$.
ii) Cuanto mayor sea la varianza del error de medida respecto a la varianza de la renta, menor será el sesgo de inconsistencia de $\widehat{\beta}_1$.
iii) El sesgo de inconsistencia de $\widehat{\beta}_1$ no depende de la relación entre las varianzas del error de medida y de la renta.
iv) $\widehat{\beta}_1$ es consistente siempre y cuando el modelo sea estimado por MCO.

c) Suponga que el error de medida en la renta es una proporción fija de la renta, de manera que si la verdadera renta es R^*, su relación con la renta observada es $R = \delta R^*$, con $0 < \delta < 1$.Entonces:

i) $\widehat{\beta}_1$ será inconsistente y $\widehat{\beta}_2$ será consistente.
ii) $\widehat{\beta}_1$ será consistente y $\widehat{\beta}_2$ será inconsistente.
iii) En general, tanto $\widehat{\beta}_1$ como $\widehat{\beta}_2$ serán consistentes.
iv) En general, tanto $\widehat{\beta}_1$ como $\widehat{\beta}_2$ serán inconsistentes.

d) Suponga que la renta es medida sin error, pero existe un error en la contabilización de las compras (V). En este caso, si la covarianza entre el error de medida y la renta es nula y además la esperanza del error de medida también es nula:

i) $\widehat{\beta}_1$ será inconsistente y $\widehat{\beta}_2$ será consistente.
ii) $\widehat{\beta}_1$ será consistente y $\widehat{\beta}_2$ será inconsistente.
iii) En general, tanto $\widehat{\beta}_1$ como $\widehat{\beta}_2$ serán consistentes.
iv) En general, tanto $\widehat{\beta}_1$ como $\widehat{\beta}_2$ serán inconsistentes.

e) Si no tenemos error de medida ni en la renta, ni en las ventas contabilizadas, ¿cuál es la diferencia en media entre las compras de los hombres y de las mujeres para un mismo nivel de renta?

i) 252 euros.
ii) 6,4 euros menos para los hombres.
iii) No hay diferencia significativa entre las compras de los hombres y el de las mujeres.
iv) 6,4 euros más para los hombres.