Soluciones Hoja de Ejercicios 2

Econometría I

1. *Al preguntar el saldo Z (en miles de euros) de su cuenta de ahorro conjunta a un matrimonio madrileño tomado al azar, el marido responde $X = Z + U$ y la mujer responde $Y = Z + W$. Se sabe que en la población de interés las variables Z, U y W son independientes, con $E[Z] = 5$, $E[U] = 0$, $E[W] = 0$, $V[Z] = 30$, $V[U] = 6$, y $V[W] = 4$.*

   a) *Si el marido responde $X = 6$, prediga de la mejor forma posible dada la información disponible la cantidad que respondería la mujer.*

   Podemos usar un modelo lineal para hacer la predicción, y suponer que $E[Y|X = x] = \beta_0 + \beta_1 x$, donde

$$\begin{aligned} V[X] &= V[Z] + V[U] = 30 + 6 = 36 \\ C[Y,X] &= V[Z] = 30 \\ &\Rightarrow \beta_1 = \frac{C[Y,X]}{V[X]} = \frac{30}{36} = \frac{5}{6} \\ \beta_0 &= E[Y] - \beta_1 E[X] = 5 - \frac{5}{6}5 = \frac{5}{6} \end{aligned}$$

   y entonces $E[Y|X = 6] = \frac{5}{6} + \frac{5}{6}6 = 5 + \frac{5}{6} = 5.8333$.

   b) *Si la mujer responde $Y = 6$, prediga de la mejor forma posible dada la información disponible la cantidad que respondería el hombre.*

   Si proponemos que $E[X|Y = y] = \gamma_0 + \gamma_1 y$, donde

$$\begin{aligned} V[Y] &= V[Z] + V[W] = 30 + 4 = 34 \\ &\Rightarrow \gamma_1 = \frac{C[Y,X]}{V[Y]} = \frac{30}{34} \\ \gamma_0 &= E[X] - \gamma_1 E[Y] = 5 - \frac{30}{34}5 \end{aligned}$$

   y entonces $E[X|Y = 6] = 5 - \frac{30}{34}5 + \frac{30}{34}6 = 5.8824$.

2. *En el modelo lineal*

$$Y = \beta_0 + \beta_1 X_1 + U,$$

   supongamos que $E[U] \neq 0$. Escribe el modelo en términos de un error ε de media cero y una nueva constante.

   Podemos reescribir el modelo como

$$\begin{aligned} Y &= \{E[U] + \beta_0\} + \beta_1 X_1 + [U - E[U]] \\ &= \beta_0^* + \beta_1 X_1 + \varepsilon \end{aligned}$$

   donde la nueva constante es $\beta_0^* = \{E[U] + \beta_0\}$ y el nuevo error es $\varepsilon = U - E[U]$.

3. *La variable kids denota el número de hijos que ha tenido una mujer y educ los años de educación de la misma. Consideremos el siguiente modelo lineal que relaciona el número de niños con los años de educación,*

$$kids = \beta_0 + \beta_1 educ + u,$$

   donde u es un error no observado.

a) *¿Qué variables se espera que estén contenidas en u?*

Salario, renta, estado civil, salario y renta del marido, lugar de residencia, religión, raza, etc.

b) *¿Aceptaría una estimación de este modelo por MCO para estudiar la relación causal entre kids y educ?*

No, porque la estimación MCO no nos aproximaría el valor esperado de *kids* dado *educ*, ya que en principio $E[u|educ]$ no sería constante, ya que, por ejemplo, el salario cambia con el nivel de educación.

4. *En la función de consumo lineal*

$$\widehat{cons} = \hat{\beta}_0 + \beta_1 inc,$$

la propensión marginal al consumo (PMC) (estimada) de la renta es simplemente la pendiente, $\hat{\beta}_1$, *mientras que la propensión media al consumo (PMEC) es* $\widehat{cons}/inc = \hat{\beta}_0/inc + \hat{\beta}_1$. *Empleando observaciones de 100 familias sobre sus ingresos y consumo anuales (ambos medidos en dólares), obtenemos la siguiente ecuación,*

$$\begin{aligned} \widehat{cons} &= -124{,}84 + 0{,}853 inc \\ n &= 100, R^2 = 0{,}692. \end{aligned}$$

a) *Interpretar el término constante en esta ecuación y comentar su signo y magnitud.*

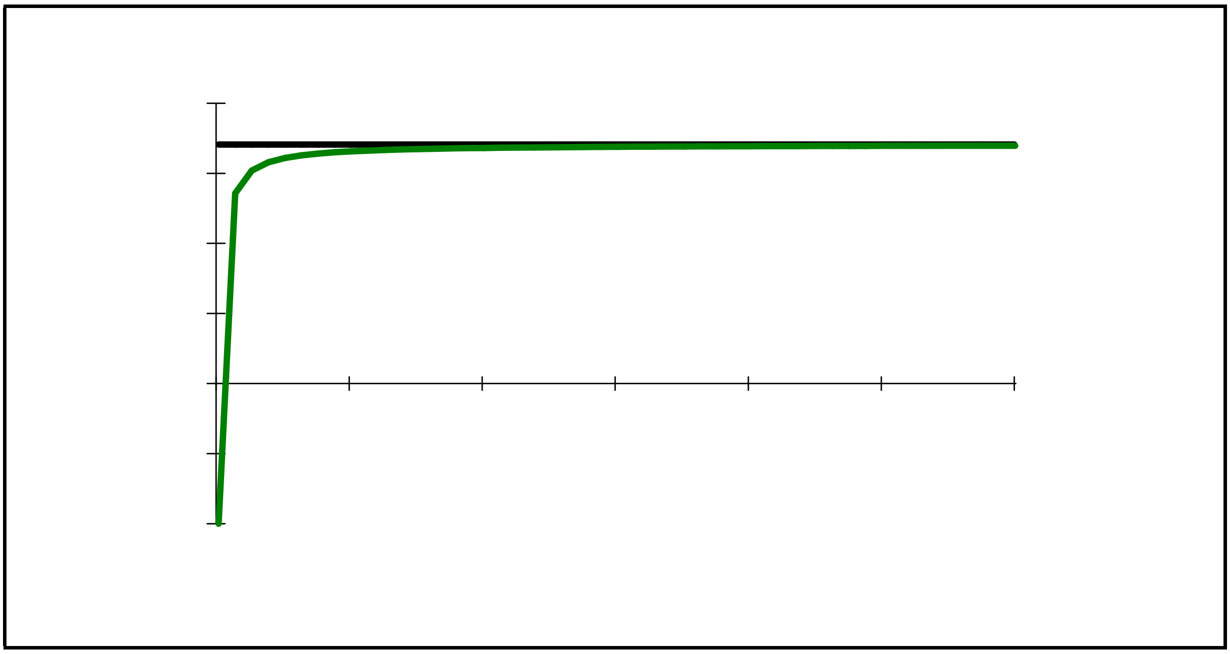
$-124{,}84$ es el nivel de consumo de alguien con $inc = 0$. No tiene mucho sentido un consumo negativo.

b) *¿Cuál es el consumo predicho cuando el ingreso de la familia es de 30000\$?*

$$\widehat{cons}(30000) = -124{,}84 + 0{,}853 * 30000 = 25465.$$

c) *Dibujar la curva de las PMC y PMEC estimadas, con inc en el eje de las x.*

0,853



5. *Consideremos la función de ahorro*

$$sav = \beta_0 + \beta_1 inc + u, \quad u = \sqrt{inc} \cdot e,$$

donde sav es ahorro, inc es renta, e tiene $E(e) = 0$ *y* $Var(e) = \sigma_e^2$ *; y e es independiente de inc.*

   a) *¿Se satisface el supuesto* $E(u|inc) = 0$*?*

   Sí,
$$E(u|inc) = E\left(\sqrt{inc} \cdot e|inc\right) = \sqrt{inc}E(e|inc) = \sqrt{inc}E(e) = \sqrt{inc} \cdot 0 = 0.$$

   b) *¿Se cumple el supuesto* $E(u^2|inc) = E(u^2)$ *(homoscedasticidad)?*

   No,

$$Var(u|inc) = Var\left(\sqrt{inc} \cdot e|inc\right) = inc \cdot Var(e|inc) = incVar(e) = inc \cdot \sigma_e^2.$$

   c) *¿Por qué puede aumentar la varianza del ahorro a medida que aumenta la renta de los individuos?*

   Porque el rango de ahorro aumenta con la renta, desde ahorrar casi toda la renta, a una gran posibilidad de endeudamiento.

6. *Consideremos el modelo estándar de regresión simple* $y = \beta_0 + \beta_1 x + u$ *con los supuestos RLS.1-4. Los estimadores MCO* $\hat{\beta}_0$ *y* $\hat{\beta}_1$ *habituales son insesgados para los parámetros poblacionales respectivos. Sea* $\tilde{\beta}_1$ *el estimador de* β_1 *obtenido con el supuesto de que el término constante,* β_0*, es cero.*

   a) *Obtener* $E\left(\tilde{\beta}_1\right)$ *en términos de las x y de* β_0, β_1*. Comprobar que* $\tilde{\beta}_1$ *no tiene sesgo cuando el término constante poblacional,* β_0*, es cero. ¿Hay otros casos en los que* $\tilde{\beta}_1$ *no tenga sesgo?*

   Tenemos que

$$\begin{aligned}
\tilde{\beta}_1 &= \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n (\beta_0 + \beta_1 x_i + u_i) x_i}{\sum_{i=1}^n x_i^2} \\
&= \frac{\sum_{i=1}^n \beta_0 x_i}{\sum_{i=1}^n x_i^2} + \frac{\sum_{i=1}^n \beta_1 x_i^2}{\sum_{i=1}^n x_i^2} + \frac{\sum_{i=1}^n u_i x_i}{\sum_{i=1}^n x_i^2} \\
&= \beta_0 \frac{n\bar{x}}{\sum_{i=1}^n x_i^2} + \beta_1 + \frac{\sum_{i=1}^n u_i x_i}{\sum_{i=1}^n x_i^2}.
\end{aligned}$$

   Condicional en las x_i,

$$\begin{aligned}
E\left(\tilde{\beta}_1\right) &= \beta_0 \frac{n\bar{x}}{\sum_{i=1}^n x_i^2} + \beta_1 + \frac{E\left(\sum_{i=1}^n u_i x_i\right)}{\sum_{i=1}^n x_i^2} \\
&= \beta_0 \frac{n\bar{x}}{\sum_{i=1}^n x_i^2} + \beta_1 + \frac{\sum_{i=1}^n E(u_i) x_i}{\sum_{i=1}^n x_i^2} \\
&= \beta_0 \frac{n\bar{x}}{\sum_{i=1}^n x_i^2} + \beta_1.
\end{aligned}$$

   Si $\beta_0 = 0$ entonces $E\left(\tilde{\beta}_1\right) = \beta_1$, pero también si $\bar{x} = 0$.

   b) *Obtener la varianza de* $\tilde{\beta}_1$.

Condicional en las x_i,

$$\begin{aligned} Var\left(\tilde{\beta}_1\right) &= Var\left(\frac{\sum_{i=1}^n u_i x_i}{\sum_{i=1}^n x_i^2}\right) \\ &= \frac{Var\left(\sum_{i=1}^n u_i x_i\right)}{\left(\sum_{i=1}^n x_i^2\right)^2} \\ &= \frac{\sum_{i=1}^n Var\left(u_i x_i\right)}{\left(\sum_{i=1}^n x_i^2\right)^2} \\ &= \frac{\sum_{i=1}^n \sigma^2 x_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}. \end{aligned}$$

c) *Demostrar que* $Var\left(\tilde{\beta}_1\right) \leq Var\left(\hat{\beta}_1\right)$. *Pista:* $\sum_{i=1}^n x_i^2 \geq \sum_{i=1}^n \left(x_i - \bar{x}\right)^2$.

Inmediato

$$\begin{aligned} Var\left(\tilde{\beta}_1\right) &= \frac{\sigma^2}{\sum_{i=1}^n x_i^2} \\ Var\left(\hat{\beta}_1\right) &= \frac{\sigma^2}{\sum_{i=1}^n \left(x_i - \bar{x}\right)^2} \end{aligned}$$

d) *Comentar la relación inversa entre sesgo y varianza al escoger entre* $\tilde{\beta}_1$ *y* $\hat{\beta}_1$.

$\tilde{\beta}_1$ tiene menos varianza pero puede tener sesgo, ya que usa un supuesto, β_0 que si es verdadero, ayuda en la estimación, pero si es falso, hacer estimar incorrectamente el parámetro de interés.

7. *Emplear los datos de WAGE2.RAW para estimar una regresión simple que explique el salario mensual (wage) en función del resultado QI (IQ).*

a) *Obtener el salario medio y el QI medio de la muestral. (Los resultados del QI están estandarizados de tal manera que la media poblacional es 100 con una desviación estándar de 15.)*

$$\begin{aligned} \overline{wage} &= 957{,}9455 \\ \overline{IQ} &= 101{,}2824. \end{aligned}$$

b) *Estimar un modelo de regresión simple en el que el aumento de un punto en IQ cambie wage en una cantidad constante en dólares. Utilizar este modelo para encontrar el aumento predicho de salario para un aumento de IQ en 15 puntos. ¿Explica IQ la mayor parte de la variación de wage?*

$$\begin{aligned} \widehat{wage} &= 116{,}9916 + 8{,}303064 IQ \\ n &= 935, R^2 = 0{,}095535. \end{aligned}$$

Predicción: si $\Delta IQ = 15$, entonces $\Delta\widehat{wage} = 15 * 8{,}303064 = 124{,}55\$$

No, porque R^2 es muy pequeño.

c) *Ahora, estimar un modelo en el que cada aumento de un punto de IQ tenga el mismo efecto porcentual sobre wage. Si IQ aumenta en 15 puntos, ¿cuál es el aumento predicho aproximado en porcentaje en wage?*

$$\begin{aligned} \log \widehat{wage} &= 5{,}886994 + 0{,}008807 * IQ \\ n &= 935, R^2 = 0{,}099091. \end{aligned}$$

Predicción: si $\Delta IQ = 15$, entonces $\Delta\,\%\widehat{wage} = 15 * 100 * 0{,}008807 = 13{,}211\,\%$.