

Soluciones Examen de Econometría I, 24 de Enero del 2000

Problema 1.

$$\pi_i = \beta \ln R_i + u_i$$

donde π_i = beneficio por empleado en millones de pesetas y R_i = gasto propio en I+D. Se dispone de una muestra aleatoria de 5000 empresas manufactureras de distintos sectores. Datos:

$$\begin{array}{lll} \sum \pi_i \ln R_i = 6.13 & \sum \pi_i \ln S_i = 74.05 & \sum \ln R_i \ln S_i = 128.17 \\ \sum \pi_i^2 = 128.5 & \sum (\ln R_i)^2 = 4.59 & \sum (\ln S_i)^2 = 28931.2 \end{array}$$

1.a. **Estimar β por MCO. Interpretar el resultado.**

Por definición, el estimador mínimo cuadrático de β viene dado por:

$$\hat{\beta}_{MCO} = \frac{\sum \pi_i \ln R_i}{\sum (\ln R_i)^2} = \frac{6.13}{4.59} = 1.3353$$

Su interpretación es la siguiente:

$\hat{\beta}_{MCO}$ = incremento medio en el beneficio por empleado ante un incremento de un 1% en el gasto propio en I+D.

1.b.i. **¿Cómo afectaría la endogeneidad del gasto propio en I+D a las propiedades del estimador de MCO?**

No sería consistente.

1.b.ii. **Proponga y calcule un estimador consistente alternativo al de MCO.**

Para obtener un estimador consistente tenemos que usar variables instrumentales. Como variable instrumental vamos a usar S_i , donde S_i mide el gasto total en I+D en el sector al que pertenece la empresa i -ésima (excluido el gasto en dicha empresa). Este estimador está definido por:

$$\hat{\beta}_{VI} = \frac{\sum \pi_i \ln S_i}{\sum \ln S_i \ln R_i} = \frac{74.05}{128.17} = 0.5777$$

- 1.b.iii. **En base a dicho estimador, y utilizando $\hat{\sigma}^2 = 0.024$ como una estimación consistente a la varianza de la perturbación, contraste la hipótesis de que el gasto propio en I+D no afecta al nivel de los beneficios de la empresa.**

Tenemos que contrastar

$$H_0 : \beta = 0 \quad vs. \quad H_0 : \beta \neq 0$$

Para ello utilizamos el contraste dado por el estadístico

$$t = \frac{\hat{\beta}_{VI}}{\sqrt{Var(\hat{\beta}_{VI})}} \stackrel{H_0}{\sim} N(0, 1), \quad n > 30$$

Para poder calcular este estadístico necesitamos conocer el valor estimado de su varianza, que viene dada por:

$$\begin{aligned} Var(\hat{\beta}_{VI}) &= \hat{\sigma}^2 \frac{\sum (\ln S_i)^2}{[\sum \ln S_i \ln R_i]^2} = \\ &= 0.024 \frac{28931.2}{128.17^2} = 0.04226 \end{aligned}$$

El valor que toma el estadístico es:

$$t = \frac{0.5777}{\sqrt{0.04226}} = 2.81$$

con

$$Z_{0.025} = 1.96$$

por lo que rechazamos H_0 al 5%, es decir, hay suficiente evidencia para afirmar que el gasto propio en I+D afecta al nivel de beneficios de la empresa.

- 1.c. **Comparar los anteriores estimadores y contrastar si el gasto propio en I+D es exógeno. Qué estimador elegiría?**

Utilizamos el estadístico propuesto por Hausman-Wu para contrastar la exogeneidad del gasto propio en I+D. El estadístico viene dado por:

$$H = \frac{(\hat{\beta}_{VI} - \hat{\beta}_{MCO})^2}{Var(\hat{\beta}_{VI}) - Var(\hat{\beta}_{MCO})} \stackrel{H_0}{\sim} \chi_1^2$$

La varianza estimada del estimador MCO viene dada por

$$\begin{aligned} Var(\hat{\beta}_{MCO}) &= \hat{\sigma}^2 \frac{1}{\sum (\ln R_i)^2} = \\ &= 0.024 \frac{1}{4.59} = 0.005228 \end{aligned}$$

Por tanto:

$$H = \frac{(0.5777 - 1.3353)^2}{0.04226 - 0.005228} = \frac{0.5739}{0.03703} = 15.49$$

con

$$\chi_{1,0.05}^2 = 3.84$$

Hay suficiente evidencia en contra de la exogeneidad del gasto propio en I+D, por lo que debemos usar el estimador de variables instrumentales.

Problema 2.

$$y_t = \alpha_1 + \alpha_2 d_{2t} + \alpha_3 d_{3t} + \alpha_4 d_{4t} + u_t \quad (1)$$

donde y_t = ventas trimestrales de una empresa y $d_{jt} = \{ 1 \text{ si } t \text{ pertenece a trimestre } j\text{-ésimo y } 0 \text{ es otro caso} \}$.

2.a. Interprete los coeficientes del modelo (1).

Las ventas esperadas en cada trimestre vienen dadas por

$$\begin{aligned} \alpha_1 &\rightarrow \text{valor medio esperado de } y_t \text{ en el } 1^{er} \text{ trimestre.} \\ \alpha_1 + \alpha_2 &\rightarrow \text{valor medio esperado de } y_t \text{ en el } 2^{o} \text{ trimestre.} \\ \alpha_1 + \alpha_3 &\rightarrow \text{valor medio esperado de } y_t \text{ en el } 3^{er} \text{ trimestre.} \\ \alpha_1 + \alpha_4 &\rightarrow \text{valor medio esperado de } y_t \text{ en el } 4^{o} \text{ trimestre.} \end{aligned}$$

luego, $\alpha_j, j \neq 1$, mide la diferencia en las ventas esperadas del trimestre j -ésimo con respecto a las del 1^{er} trimestre, en millones de pesetas.

2.b. Modelo alternativo

$$y_t = \beta_0 + \beta_1 d_{1t} + \beta_2 d_{2t} + \beta_3 d_{3t} + \beta_4 d_{4t} + \varepsilon_t$$

¿Existe algún problema para estimar dicho modelo?

Existe un problema de multicolinealidad perfecta, ya que

$$d_{1t} + d_{2t} + d_{3t} + d_{4t} = 1 \quad \forall t$$

2.c. **Dado el siguiente modelo:**

$$y_t = \gamma_1 d_{1t} + \gamma_2 d_{2t} + \gamma_3 d_{3t} + \gamma_4 d_{4t} + v_t \quad (2)$$

comente qué relación existe entre los coeficientes del modelo (1) y los del modelo (2).

Es claro que

$$d_{1t} = 1 - d_{2t} - d_{3t} - d_{4t}$$

por lo que

$$\begin{aligned} y_t &= \gamma_1 + (\gamma_2 - \gamma_1) d_{2t} + (\gamma_3 - \gamma_1) d_{3t} + (\gamma_4 - \gamma_1) d_{4t} + v_t \\ \Rightarrow \alpha_1 &= \gamma_1; \alpha_2 = \gamma_2 - \gamma_1; \alpha_3 = \gamma_3 - \gamma_1; \alpha_4 = \gamma_4 - \gamma_1. \end{aligned}$$

2.d. **Estime por MCO los parámetros del modelo (2).**

Sabemos que

$$\hat{\gamma}_{MCO} = (\mathbf{d}'\mathbf{d})^{-1} (\mathbf{d}'y)$$

Debemos observar que

$$(\mathbf{d}'\mathbf{d})^{-1} = \begin{pmatrix} (\sum d_1^2)^{-1} & 0 & 0 & 0 \\ 0 & (\sum d_2^2)^{-1} & 0 & 0 \\ 0 & 0 & (\sum d_3^2)^{-1} & 0 \\ 0 & 0 & 0 & (\sum d_4^2)^{-1} \end{pmatrix}$$

con $\sum d_j^2 = \sum d_j$, por lo que llegamos a que cada $\hat{\gamma}_j$ viene dado por

$$\hat{\gamma}_j = \frac{\sum d_{ij} y_i}{\sum d_j}$$

por tanto

$$\hat{\gamma}_1 = \frac{20}{10} = 2; \quad \hat{\gamma}_2 = \frac{30}{10} = 3; \quad \hat{\gamma}_3 = \frac{70}{10} = 7; \quad \hat{\gamma}_4 = \frac{30}{10} = 3.$$

Para los del examen bilingüe, que tenían en el enunciado $\sum d_j = 40$, los resultados son:

$$\hat{\gamma}_1 = \frac{20}{40} = 0.5; \quad \hat{\gamma}_2 = \frac{30}{40} = 0.75; \quad \hat{\gamma}_3 = \frac{70}{40} = 1.75; \quad \hat{\gamma}_4 = \frac{30}{40} = 0.75.$$

2.e. **Sabiendo que $\sum \hat{v}_t^2 = 720$, contraste que las ventas durante el primer trimestre son iguales en media a las del segundo trimestre.**

Lo que tenemos que contrastar es:

$$H_0 : \gamma_1 = \gamma_2 \quad vs. \quad H_1 : \gamma_1 \neq \gamma_2$$

donde el estadístico del contraste viene dado por

$$t = \frac{|\hat{\gamma}_1 - \hat{\gamma}_2|}{sq(\hat{\gamma}_1 - \hat{\gamma}_2)} \stackrel{H_0}{\sim} N(0, 1), \quad n > 30$$

con

$$\begin{aligned} sq(\hat{\gamma}_1 - \hat{\gamma}_2) &= \sqrt{Var(\hat{\gamma}_1) + Var(\hat{\gamma}_2) - 2Cov(\hat{\gamma}_1, \hat{\gamma}_2)} \\ &= \sqrt{s^2 \left(\frac{1}{\sum d_{i1}} + \frac{1}{\sum d_{i2}} \right)} = \\ &= \sqrt{\frac{720}{40 - 4} \left(\frac{1}{10} + \frac{1}{10} \right)} = \sqrt{4} = 2 \end{aligned}$$

Por tanto

$$t = \frac{|\hat{\gamma}_1 - \hat{\gamma}_2|}{sq(\hat{\gamma}_1 - \hat{\gamma}_2)} = \frac{|2 - 3|}{2} = \frac{1}{2} = 0.5$$

Para los del examen en Inglés:

$$\begin{aligned} sq(\hat{\gamma}_1 - \hat{\gamma}_2) &= \sqrt{Var(\hat{\gamma}_1) + Var(\hat{\gamma}_2) - 2Cov(\hat{\gamma}_1, \hat{\gamma}_2)} \\ &= \sqrt{s^2 \left(\frac{1}{\sum d_{i1}} + \frac{1}{\sum d_{i2}} \right)} = \\ &= \sqrt{\frac{720}{40 - 4} \left(\frac{1}{40} + \frac{1}{40} \right)} = \sqrt{1} = 1 \end{aligned}$$

Por tanto

$$t = \frac{|\hat{\gamma}_1 - \hat{\gamma}_2|}{sq(\hat{\gamma}_1 - \hat{\gamma}_2)} = \frac{|0.5 - 0.75|}{1} = \frac{0.25}{1} = 0.25$$

mientras que

$$Z_{0.025} = 1.96$$

No hay suficiente evidencia estadística, al 5%, para rechazar que las ventas durante el primer trimestre son iguales en media a las del segundo trimestre, en ambos exámenes, el de castellano y el de inglés.

Problema 3.

Hay dos modelos alternativos

$$S_i = b_1 F_i + u_i \quad (3)$$

$$S_i = \beta_1 F_i + \beta_2 Z_i + v_i \quad (4)$$

donde S_i =salarios de empleados jóvenes, F_i =años de formación y Z_i =otra variable explicativa. Todas las variables están expresadas respecto a la media.

3.a. En qué modelo el coeficiente de determinación será mayor?

En el modelo (4) ya que el coeficiente de determinación nunca decrece cuando se añade otra variable a la ecuación de regresión.

3.b. Qué inconveniente presenta el modelo (4) si queremos contrastar la significación de F_i en el caso de que el verdadero valor de β_2 sea 0?

Incluir Z puede aumentar la varianza de $\hat{\beta}_1$, ya que si denotamos por $\mathbf{X}' = (F, Z)$, tenemos que:

$$\begin{aligned} Var(\hat{\beta}_1) &= \sigma^2 \{(\mathbf{X}'\mathbf{X})\}_{11}^{-1} = \\ &= \sigma^2 (F'F - F'Z(Z'Z)^{-1}Z'F)^{-1} \end{aligned}$$

Mientras que

$$Var(\hat{b}_1) = \sigma^2 (F'F)^{-1}$$

3.c. Expresión del estimador de MCO del parámetro asociado a F_i en cada modelo.

Tienen la siguiente expresión:

$$\begin{aligned} \hat{b}_1 &= (F'F)^{-1}(F'S) = \frac{\sum F_i S_i}{\sum F_i^2} \\ \hat{\beta}_1 &= \frac{\sum F_i S_i \sum Z_i^2 - \sum Z_i S_i \sum F_i Z_i}{\sum F_i^2 \sum Z_i^2 - (\sum F_i Z_i)^2} \end{aligned}$$

equivalentemente, usando fórmulas de la regresión particionada:

$$\begin{aligned}\hat{\beta}_1 &= (\mathbf{F}'M_Z\mathbf{F})^{-1}(\mathbf{F}'M_ZS) \quad \text{con} \\ M_Z &= I - Z(Z'Z)^{-1}Z'\end{aligned}$$

3.d. **Bajo qué condiciones $\hat{b}_1 = \hat{\beta}_1$?**

De la fórmulas anteriores es claro que la condición es

$$\sum F_i Z_i = 0 \equiv F'Z = 0$$

a) 0.5 puntos

En el modelo (5). **0.2 P**

Como $R^2 = \frac{\text{var.explicada}}{\text{var.total}}$, en donde *var.total* es igual para (4) y (5). Pero la varianza explicada no puede ser menor en (5) como eso es la varianza de S que se explica por F y Z . Aunque Z no explicará nada, como todo que hemos incluido en (4) también se tiene en (5), (4) no puede explicar más. **0.3 P**

b) 1 punto

Incluyendo variables no relevantes en un modelo puede aumentar la varianza de todos los parámetros estimados. **0.4.**

(En consecuencia, el estadístico para contrastar la significancia de F sería menor, el contraste perdería potencia, **0.2**)

así que rechazaríamos más pronto. **0.4**

c) 0.5 punto

Para $\hat{b}_1$, $(F'F)^{-1}F'S \underline{\text{O}} \frac{\sum F_i S_i}{\sum F_i^2}$ o $\frac{\sum (F_i - \bar{F}) S_i}{\sum (F_i - \bar{F})^2}$. **0.2**

Para $\hat{\beta}_1$ se puede escribir $\hat{\beta} = (X'X)^{-1}X'S$ con $X' = (F, Z) \underline{\text{O}} \hat{\beta}_1 = (F'M_ZF)^{-1}F'M_ZS$ con $M_Z = I - Z(Z'Z)^{-1}Z'$. **0.3**

d) 1 punto

¡ Sólo cuando $Z'F = 0$ (son ortogonales) (Independencia no es suficiente) **1 P.**

Si dicen algo de independencia, damos **0.2 P.**