CASO PRÁCTICO DEL MODELO LINEAL DE ECONOMETRÍA

R. Amor Pulido, E. Álvarez-Verdejo, P.J. Moya-Fernández

Departamento de Métodos Cuantitativos para la Economía y la Empresa. Universidad de Granada

RESUMEN:

Uno de los principales problemas que hemos detectado para la comprensión por parte de los alumnos del modelo lineal de Econometría es saber interpretar los conocimientos teóricos en un problema numérico real. Para solucionar este problema, hemos seleccionado un problema, que vamos a ir desarrollando junto con la teoría. Por tanto, partiremos de los datos propiamente dichos, construiremos las matrices y vectores iniciales del modelo e iremos aplicándole todas las técnicas estudiadas en la teoría, desde la obtención de los parámetros del modelo, hasta el estudio de la multicolinealidad, heterocedasticidad y autocorrelación. De esta manera pueden ver que todo está conectado y que la Econometría no son procedimientos aislados unos de otros.

Con este método de trabajo, hemos conseguido una mejor comprensión por parte del alumno de qué es la Econometría, para qué sirve y cómo se trabaja en un caso práctico. Además, con solo este ejemplo, se abarca todo el contenido de la asignatura y ya deben saber hacer cualquier otro problema parcial.

Queremos explicar el beneficio de las empresas en cientos de miles de € (Y) en función del número de cientos de empleados (X₂) y de la inversión inicial en millones de € (X₃), a partir de los siguientes datos muestrales:

Observación	Y	X ₂	X ₃
1	10	1	1
2	13	2	2
3	25	2	2
4	22	3	3
5	30	4	5

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

Agrupamos los datos en matrices y realizamos los primeros cálculos:

$$Y = \begin{pmatrix} 10 \\ 13 \\ 25 \\ 22 \\ 30 \end{pmatrix} \quad X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \\ 1 & 3 & 3 \\ 1 & 4 & 5 \end{pmatrix} \quad X'X = \begin{pmatrix} 5 & 12 & 13 \\ 12 & 34 & 38 \\ 13 & 38 & 43 \end{pmatrix} \quad X'Y = \begin{pmatrix} 100 \\ 272 \\ 302 \end{pmatrix} \quad (X'X)^{-1} = \begin{pmatrix} 2.25 & -2.75 & 1.75 \\ -2.75 & 5.75 & -4.25 \\ 1.75 & -4.25 & 3.25 \end{pmatrix}$$

Obtenemos los estimadores de MCO y el modelo estimado queda:

$$\hat{\beta} = (X'X)^{-1} X'Y = \begin{pmatrix} 2.25 & -2.75 & 1.75 \\ -2.75 & 5.75 & -4.25 \\ 1.75 & -4.25 & 3.25 \end{pmatrix} \begin{pmatrix} 100 \\ 272 \\ 302 \end{pmatrix} = \begin{pmatrix} 5.5 \\ 5.5 \\ 0.5 \end{pmatrix} \quad \hat{Y}_t = 5.5 + 5.5X_{2t} + 0.5X_{3t}$$

• Por cada 100 empleados de más, los beneficios aumentan en 550000 €

• Por cada millón de € de inversión inicial, los beneficios aumentan en 500000€

Los residuos se obtienen como: $Y - \hat{Y} = Y - X\hat{\beta} = \begin{pmatrix} 10 \\ 13 \\ 25 \\ 22 \\ 30 \end{pmatrix} - \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \\ 1 & 3 & 3 \\ 1 & 4 & 5 \end{pmatrix} \begin{pmatrix} 5.5 \\ 5.5 \\ 0.5 \end{pmatrix} = \begin{pmatrix} -1.5 \\ -4.5 \\ 7.5 \\ -1.5 \\ 0 \end{pmatrix}$

Calculamos las distintas sumas de cuadrados para posteriores cálculos:

$$\begin{cases} SCT = Y'Y - n\bar{Y}^2 = \sum_{i=1}^n Y_i^2 - n \left(\frac{\sum_{i=1}^n Y_i}{n} \right)^2 = 2278 - 5 \cdot \left(\frac{100}{5} \right)^2 = 278 \\ SCR = Y'Y - \hat{\beta}'X'Y = 2278 - (5.5 \quad 5.5 \quad 0.5) \begin{pmatrix} 100 \\ 272 \\ 302 \end{pmatrix} = 2278 - 2197 = 81 \\ SCE = \hat{\beta}'X'Y - n\bar{Y}^2 = 2197 - 2000 = 197 \end{cases}$$

Calculamos el coeficiente de determinación y el coeficiente de determinación corregido:

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = \frac{197}{278} = 0.7086$$

• La variación de los beneficios se puede explicar en un 0.86% a partir de la variación del número de empleados y la inversión inicial.

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k} = 1 - (1 - 0.7086) \frac{5-1}{5-3} = 0.4173$$

• Al incluir variables en el modelo, el coeficiente de determinación aumenta aunque las variables incluidas no sean significativas. Este problema se resuelve con el coeficiente de determinación corregido, que penaliza la inclusión de variables.

Calculamos la estimación de la varianza de las perturbaciones y la matriz de varianzas-covarianzas de los parámetros:

$$\hat{\sigma}^2 = \frac{SCR}{n-k} = \frac{81}{5-3} = 40.5$$

$$\widehat{Var}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} = \begin{pmatrix} 91.125 & -111.375 & 70.875 \\ -111.375 & 232.875 & -172.125 \\ 70.875 & -172.125 & 131.625 \end{pmatrix}$$

Realizamos los contrastes de significación individual:

$$\begin{cases} H_0: \beta_i = 0 \\ H_1: \beta_i \neq 0 \end{cases} t_{exp} = \frac{\hat{\beta}_i}{\sqrt{Var(\hat{\beta}_i)}} = t_{n-k, 1-\alpha/2}$$

- $H_0: \beta_1 = 0 \Rightarrow t_{exp} = \frac{5.5}{\sqrt{91.125}} = 0.576 < t_{2, 0.975} = 4.303 \Rightarrow$ Término independiente no significativo
- $H_0: \beta_3 = 0 \Rightarrow t_{exp} = \frac{0.5}{\sqrt{131.625}} = 0.043 < t_{2, 0.975} = 4.303 \Rightarrow$ Variable X₃ no significativa
- $H_0: \beta_2 = 0 \Rightarrow t_{exp} = \frac{5.5}{\sqrt{232.875}} = 0.360 < t_{2, 0.975} = 4.303 \Rightarrow$ Variable X₂ no significativa

Realizamos el contraste de significación conjunta: $\begin{cases} H_0: \beta_2 = \beta_3 = 0 \\ H_1: \text{Alguna} \neq 0 \end{cases}$

FV	GL	SC	CM = SC/GL	F _{exp}	F _{2,2,0.95}
Regresión	k-1 = 2	SCE = 197	197/2=98.5	98.5/40.5=2.43209877<	19.00
Residuos	n-k=2	SCR = 81	81/2=40.5	Acepto H ₀	
Total	n-1 = 4	SCT = 278		Modelo no significativo	

Vemos la significación del coeficiente de determinación:

R² es significativo si

$$R^2 > R_{sig}^2 = \frac{\frac{k-1}{n-k} \cdot F_{k-1, n-k, 1-\alpha}}{1 + \frac{k-1}{n-k} \cdot F_{k-1, n-k, 1-\alpha}} \quad R^2 = 0.7086 < R_{sig}^2 = \frac{\frac{2}{2} \cdot 19}{1 + \frac{2}{2} \cdot 19} = 0.95 \Rightarrow R^2 \text{ no es significativo}$$

Calculamos los intervalos de confianza para los parámetros al 95%:

$$\beta_i \in \left(\hat{\beta}_i \pm t_{n-k, 1-\alpha/2} \sqrt{Var(\hat{\beta}_i)} \right) \Rightarrow \begin{cases} \beta_1 \in (5.5 \pm 4.303\sqrt{91.125}) = (-35.573, 46.573) \\ \beta_2 \in (5.5 \pm 4.303\sqrt{232.875}) = (-60.1595, 71.1595) \\ \beta_3 \in (0.5 \pm 4.303\sqrt{131.625}) = (-48.8634, 49.8634) \end{cases}$$

Una empresa tiene 500 trabajadores y realiza una inversión inicial de 3 millones de € $\Rightarrow x_0' = (1 \quad 5 \quad 3)$

• **Predicción puntual:** $\hat{Y}_t = 5.5 + 5.5 \cdot 5 + 0.5 \cdot 3 = 23.5$ miles de €

• **Predicción por intervalo:**

$$\hat{Y} \in \left(x_0' \pm t_{n-k, 1-\alpha/2} \hat{\sigma} \sqrt{1 + x_0'(X'X)^{-1}x_0} \right) = (23.5 \pm 4.303\sqrt{40.5\sqrt{30.75}}) = (-128.3523, 175.3523)$$

con $x_0'(X'X)^{-1}x_0 = (1 \quad 5 \quad 3) \begin{pmatrix} 2.25 & -2.75 & 1.75 \\ -2.75 & 5.75 & -4.25 \\ 1.75 & -4.25 & 3.25 \end{pmatrix} \begin{pmatrix} 1 \\ 5 \\ 3 \end{pmatrix} = 30.75$

• **Valor esperado de Y:**

$$\hat{Y} \in \left(x_0' \pm t_{n-k, 1-\alpha/2} \hat{\sigma} \sqrt{1 + x_0'(X'X)^{-1}x_0} \right) = (23.5 \pm 4.303\sqrt{40.5\sqrt{1+30.75}}) = (-130.8017, 177.8017)$$

• Si una empresa en las condiciones anteriores tiene unos beneficios de 800 mil €, ¿están estos datos conformes al modelo de trabajo?

Como 8 está en el intervalo del valor esperado, concluimos que el modelo sigue siendo válido para el dato considerado.

Queremos ver si podemos aceptar la restricción de que $\beta_2 + \beta_3 = 7$ y, en su caso, obtener los estimadores que la cumplan.

a) Se rechaza la restricción si $F_{exp} = (R\hat{\beta} - r)' \frac{[R(X'X)^{-1}R']^{-1}}{q \cdot \hat{\sigma}^2} (R\hat{\beta} - r) > F_{q, n-k-1, \alpha}$

$$\begin{cases} H_0: \beta_2 + 2\beta_3 = 7 \\ H_1: \beta_2 + 2\beta_3 \neq 7 \end{cases} \Rightarrow \begin{cases} q = 1 \text{ (n}^{\circ} \text{ restricciones)} \\ R = (0 \quad 1 \quad 2) \\ r = 7 \end{cases}$$

$$F_{exp} = (R\hat{\beta} - r)' \frac{[R(X'X)^{-1}R']^{-1}}{q \cdot \hat{\sigma}^2} (R\hat{\beta} - r) = (-0.5) \frac{1.75^{-1}}{1 \cdot 40.5} (-0.5) = 0.003527 < F_{1, 2, 0.95} = 18.51, \text{ con}$$

$$(R\hat{\beta} - r) = (0 \quad 1 \quad 2) \begin{pmatrix} 5.5 \\ 5.5 \\ 0.5 \end{pmatrix} - 7 = -0.5$$

$$R(X'X)^{-1}R' = (0 \quad 1 \quad 2) \begin{pmatrix} 2.25 & -2.75 & 1.75 \\ -2.75 & 5.75 & -4.25 \\ 1.75 & -4.25 & 3.25 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} = (0 \quad 1 \quad 2) \begin{pmatrix} 0.75 \\ -2.75 \\ 2.25 \end{pmatrix} = 1.75$$

entonces, puedo aceptar la restricción como cierta.

b) El estimador que cumple la restricción viene dado por:

$$\hat{\beta}_R = \hat{\beta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) = \begin{pmatrix} 5.5 \\ 5.5 \\ 0.5 \end{pmatrix} - \begin{pmatrix} 0.75 \\ -2.75 \\ 2.25 \end{pmatrix} 1.75^{-1}(-0.5) = \begin{pmatrix} 5.7142 \\ 4.7142 \\ 1.1429 \end{pmatrix}$$

que cumple la restricción $\beta_2 + 2\beta_3 = 4.7142 + 2 \cdot 1.1429 = 7$

MULTICOLINEALIDAD:

• **Exacta:** No existe Multicolinealidad exacta porque la matriz X tiene rango máximo al ser invertible la matriz X'X.

• **Aproximada:** Hacemos la regresión lineal de la variable X₂ respecto de la variable X₃: $\hat{X}_{2t} = 0.4783 + 0.7391X_{3t}$ con $R^2 = 0.9665$

Entonces, el factor de agrandamiento de la varianza de la variable X₂ viene dado por: $FAV(X_2) = \frac{1}{1 - R^2} = \frac{1}{1 - 0.9665} = 29.85$

Por tanto, el modelo tiene multicolinealidad aproximada y deberíamos eliminar una de las 2 variables independientes. Esta es la razón por la que las varianzas son tan elevadas.

El modelo debería quedar: $\hat{Y}_t = 5.2308 + 6.1538X_{2t}$ con $R^2 = 0.7083$

HETEROCEDASTICIDAD:

Comprobamos si el modelo original presenta o no heterocedasticidad, utilizando el test de Breusch-Pagan. Tenemos que realizar la regresión auxiliar con variable dependiente el cuadrado de los residuos y variables independientes todas las independientes del modelo inicial, es decir,

$$e_t^2 = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

$$\hat{e}_t^2 = 20.25 + 20.25X_{2t} - 20.25X_{3t} \text{ con } R_{aux}^2 = 0.1443$$

$$\chi_{exp}^2 = nR_{aux}^2 = 5 \cdot 0.1443 = 0.7215 < \chi_{q, 1-\alpha}^2 = \chi_{2, 0.95}^2 = 5.991$$

por lo que acepto la hipótesis nula de que no existe heterocedasticidad.

AUTOCORRELACIÓN:

Comprobamos si el modelo original presenta o no autocorrelación, utilizando el test de Durbin-Watson.

e _t	e _t ²	e _{t-1}	e _t - e _{t-1}	(e _t - e _{t-1}) ²
-1,5	2,25			
-4,5	20,25	-1,5	-3	9
7,5	56,25	-4,5	12	144
-1,5	2,25	7,5	-9	81
0	0	-1,5	1,5	2,25
	81			236,25

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} = \frac{236.25}{81} = 2.9167$$

-Si está en (0, d_L) existe autocorrelación positiva.

- Si está en (d_L, d_U) o en (4-d_U, 4-d_L) existe indeterminación.

-Si está en (d_U, 4-d_U) no existe autocorrelación. d_L y d_U dependen de n y k' = k-1

- Si está en (4-d_L, 4) existe autocorrelación negativa.

Este valor tenemos que ver en qué intervalo se encuentra, pero en nuestro ejemplo no podemos obtener estos valores, pues necesitaríamos un tamaño muestral de, al menos, 7 unidades.