

Examen de Econometría I
Universidad Carlos III de Madrid
2aa Convocatoria (11 de septiembre del 2003)
Curso 2002/2003

Conteste las preguntas siguientes en 2 horas y media

Valores críticos de diferentes distribuciones pueden encontrarse al final del examen.

1. Suponga que:

$$E(Y|X) = \beta_0 + \beta_1 X + \beta_2 X^2,$$

donde todos los coeficientes del modelo son distintos de cero.

El predictor lineal óptimo (*PLO*) de Y dado X es:

$$PLO(Y|X) = \delta_0 + \delta_1 X.$$

¿Bajo que circunstancias β_1 es igual a δ_1 ? Justifique su respuesta. (1 punto).

SOLUCIÓN: Si $E(X) = E(X^3) = 0$, entonces $C(X, X^2) = E(X^3) = 0$ y,

$$\delta_1 = \beta_1 + \beta_2 \frac{C(X^2, X)}{V(X)} = \beta_1.$$

2. Interprete los coeficientes β_1 y β_2 en el modelo:

$$\ln Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 X_2 + \varepsilon,$$

donde ε es un error de media cero e independiente de X_1 y de X_2 . (1 punto).

SOLUCIÓN:

$$\beta_1 = \frac{\partial}{\partial \ln X_1} E(\ln Y | X_1, X_2)$$

$\Rightarrow \beta_1$ es la elasticidad esperada de Y respecto a X_1

$\Rightarrow$ Cuando X_1 varía en un 1%, manteniéndose X_2 constante, se espera que Y varíe en un $\beta_1\%$

$$\beta_2 = \frac{\partial}{\partial X_2} E(\ln Y | X_1, X_2)$$

$\Rightarrow \beta_2$ es la semi-elasticidad esperada de Y respecto a X_2

$\Rightarrow$ Cuando X_2 varía en una unidad, manteniéndose X_1 constante, se espera que Y varíe en un $100 \cdot \beta_2\%$

3. Sea el modelo de regresión lineal:

$$E(\log Y | X_1, X_2, X_3, X_4, X_5) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

donde:

$$\begin{aligned} Y &= \text{salario-hora de un individuo en pesetas} \\ X_1 &= \begin{cases} 1 & \text{si el individuo reside en la comunidad de Madrid} \\ 0 & \text{en caso contrario} \end{cases} \\ X_2 &= \begin{cases} 1 & \text{si el individuo trabaja en el sector agrícola} \\ 0 & \text{en caso contrario} \end{cases} \\ X_3 &= \begin{cases} 1 & \text{si el individuo trabaja en el sector servicios} \\ 0 & \text{en caso contrario} \end{cases} \\ X_4 &= \text{años de educación del individuo} \\ X_5 &= \text{años de experiencia laboral del individuo} \end{aligned}$$

Existen tres sectores productivos: agricultura, servicios e industria.

Empleando datos de 3360 individuos asalariados se han realizado las siguientes estimaciones:

SALIDA 1

Dependent Variable: LOG(Y)

Method: Least Squares

Sample: 1 3360

Included observations: 3360

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.678950	0.025663	221.2936	0.0000
X ₁	0.077972	0.023293	3.347419	0.0008
X ₂	-0.376339	0.036398	-10.33950	0.0000
X ₃	-0.032378	0.015776	-2.052339	0.0402
X ₄	0.066317	0.001974	33.59876	0.0000
X ₅	0.014673	0.000601	24.42582	0.0000
R-squared	0.330751	Mean dependent var	6.544462	
Adjusted R-squared	0.329753	S.D. dependent var	0.520908	
S.E. of regression	0.426460	Akaike info criterion	1.135189	
Sum squared resid	609.9863	Schwarz criterion	1.146117	
Log likelihood	-1901.117	F-statistic	331.5171	
Durbin-Watson stat	1.862043	Prob(F-statistic)	0.000000	

SALIDA 2

Dependent Variable: LOG(Y)

Method: Least Squares

Sample: 1 3360

Included observations: 3360

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.628126	0.025348	222.0386	0.0000
X ₁	0.087191	0.023590	3.696049	0.0002
X ₄	0.068040	0.001923	35.37594	0.0000
X ₅	0.014675	0.000610	24.05864	0.0000
R-squared	0.309419	Mean dependent var	6.544462	
Adjusted R-squared	0.308802	S.D. dependent var	0.520908	
S.E. of regression	0.433074	Akaike info criterion	1.165375	
Sum squared resid	629.4290	Schwarz criterion	1.172660	
Log likelihood	-1953.830	F-statistic	501.2254	
Durbin-Watson stat	1.842341	Prob(F-statistic)	0.000000	

SALIDA 3

Dependent Variable: LOG(Y)

Method: Least Squares

Sample: 1 3360

Included observations: 3360

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.677029	0.025695	220.9383	0.0000
X ₂	-0.380988	0.036427	-10.45897	0.0000
X ₃	-0.029525	0.015777	-1.871368	0.0614
X ₄	0.067127	0.001962	34.21496	0.0000
X ₅	0.014795	0.000601	24.63570	0.0000
R-squared	0.328515	Mean dependent var	6.544462	
Adjusted R-squared	0.327714	S.D. dependent var	0.520908	
S.E. of regression	0.427108	Akaike info criterion	1.137929	
Sum squared resid	612.0242	Schwarz criterion	1.147035	
Log likelihood	-1906.720	F-statistic	410.3469	
Durbin-Watson stat	1.856654	Prob(F-statistic)	0.000000	

- a. ¿Cuál es la diferencia porcentual media estimada entre el salario de un individuo que trabaja en el sector industrial en la comunidad de Madrid y un individuo que trabaja en el sector agrícola en Andalucía, que tengan igual educación y experiencia laboral? (1 punto)
- b. Contraste al 5% de significación que la ecuación salarial es independiente del sector productivo en que trabaje un individuo. (1 punto).

SOLUCIÓN:

- a. $E(\log Y | X_1 = 1, X_2 = 0, X_3 = 0, X_4, X_5) - E(\log Y | X_1 = 0, X_2 = 1, X_3 = 0, X_4, X_5) = \beta_1 - \beta_2$. La diferencia porcentual media (aproximada) entre el salario de un trabajador del sector industrial de Madrid y un trabajador del sector agrícola de Andalucía es del $(\beta_1 - \beta_2) \times 100\%$. Por tanto, la diferencia estimada por MCO es del $(\hat{\beta}_1 - \hat{\beta}_2) \times 100\% = 0,077972 - (-0,376339) \times 100\% = 45,43\%$.
- b. La hipótesis nula a contrastar es $H_0 : \beta_2 = \beta_3 = 0$. Nótese que la Salida 2 es la estimación del modelo de la Salida 1 tras imponer esta hipótesis. El estadístico de contraste es:

$$W^0 = \frac{(R_1^2 - R_2^2)}{(1 - R_1^2)} \times (n - k - 1)$$

donde R_1^2 es el coeficiente de determinación de la Salida 1 (Modelo sin restringir), R_2^2 el coeficiente de determinación de la Salida 2 (Modelo restringido) y k el número de regresores del modelo sin restringir, k=5. Por tanto:

$$W^0 = \frac{(0,330751 - 0,309419)}{(1 - 0,330751)} \times (3360 - 5 - 1) = 106,91$$

Alternativamente:

$$W^0 = \frac{(SR_2 - SR_1)}{SR_1} \times (n - k - 1)$$

donde SR_1 es la suma de cuadrados de los residuos de la Salida 1 (Modelo sin restringir), SR_2 la suma de cuadrados de los residuos de la Salida 2 (Modelo restringido) y k el número de regresores del modelo sin restringir, k=5. Por tanto:

$$W^0 = \frac{(629,429 - 609,9863)}{609,9863} \times (3360 - 5 - 1) = 106,91$$

Sabemos que:

$$W^0 \underset{aprox}{\sim} \chi_2^2 \text{ bajo } H_0$$

El valor crítico al 5% de significación es 5,99 y por tanto, rechazamos la hipótesis nula.

4. Dos investigadores proponen el siguiente modelo para analizar la relación entre producción (Y) y factores productivos (L=trabajo y K=stock de capital):

$$Y = \beta_0 + \beta_1 L + \beta_2 K + \varepsilon$$

donde $E(\varepsilon|X) = 0$ y $V(\varepsilon|X) = \sigma^2$.

Aunque, el stock de capital, K, no se observa, como están interesados en estimar β_1 , piensan que esto no será un problema muy grave. Sin embargo, no se ponen de acuerdo en cómo obtener estimaciones consistentes de β_1 .

El primer investigador, propone estimar por MCO el modelo:

$$Y = \alpha + \beta L + u \quad (1)$$

El segundo investigador propone estimar por MCO el modelo:

$$Y = \gamma_0 + \gamma_1 L + \gamma_2 K^* + v \quad (2)$$

donde $K^* = K + w$, siendo w un error de medida que cumple $E(w) = 0$, $E(w^2) = \sigma_w^2$, $E(w\varepsilon) = E(wK) = 0$.

¿Con cuál de las dos estrategias se obtendrá un estimador consistente de β_1 ? Justifique su respuesta (1,5 punto).

SOLUCIÓN:

Los estimadores *MCO* de los parámetros de la ecuación (1) son inconsistentes, al omitir una variable significativa (K) que está correlacionada con la otra variable explicativa (el factor trabajo L).

Los estimadores *MCO* de los parámetros de la ecuación (2) son inconsistentes, al incluir en el modelo la variable significativa (K) medida con error.

5. Considere el modelo de regresión

$$E(Y|X) = \beta \cdot X$$

donde β es un parámetro y

Y = total de pescado capturado en miles de kilos en un día

X = número de barcos que salieron a pescar en ese día.

Esto es, el verdadero modelo de regresión no tiene constante.

El coeficiente β se estima por *MCO* en un modelo con constante. ¿Es el estimador resultante consistente? ¿Es el estimador resultante eficiente? Justifique sus respuestas. (1 punto)

SOLUCIÓN:

Al ser la constante igual a cero, los estimadores *MCO* del modelo con constante son consistentes (el estimador de la constante estima consistentemente 0 y el de la pendiente estima consistentemente β). No obstante, no se está utilizando la información de que la constante es cero y, consecuentemente, el estimador de MCO que incorpora la restricción (estimador de mínimos cuadrados restringidos) será más eficiente el de MCO que no utiliza la restricción.

6. Utilizando datos de sección cruzada sobre 59 viviendas familiares vendidas en La Jolla (el nombre americano es con "ll") y en su Ciudad Universitaria, se obtuvieron datos de las siguientes variables:

<i>PRICE</i>	=	precio de la vivienda en miles de dólares
<i>SQFT</i>	=	pies cuadrados útiles
<i>YARD</i>	=	tamaño de la parcela en pies cuadrados
<i>POOL</i>	=	1 si la vivienda tiene piscina y 0 en otro caso
<i>VIEW</i>	=	1 si la vivienda tiene vistas y 0 en otro caso.
<i>LAJOLLA</i>	=	1 si la vivienda está en La Jolla y 0 en otro caso.

Utilizando las salidas de E-Views que se presentan abajo:

- Utilizando la salida 1. ¿Cual es el precio esperado de una casa en La Jolla con vistas, sin piscina, con 600 pies cuadrados útiles y 5000 pies cuadrados de parcela. (1 punto)
- Contraste al 1% de significación si las variables explicativas afectan de distinta forma a los precios de las viviendas en La Jolla y en la Ciudad Universitaria. (1 punto).

SALIDA 1

Dependent Variable: PRICE

Method: Least Squares

Date: 08/25/03 Time: 18:32

Sample: 1 59

Included observations: 59

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	33.13433	56.29804	0.588552	0.5589
SQFT	0.055836	0.026629	2.096812	0.0412
YARD	0.001319	0.005069	0.260250	0.7958
VIEW	15.88825	24.91806	0.637620	0.5267
POOL	12.16879	29.98681	0.405805	0.6867
LAJOLLA	65.55319	68.08578	0.962803	0.3404
LAJOLLA*SQFT	0.020636	0.033166	0.622200	0.5367
LAJOLLA*YARD	0.001085	0.005431	0.199778	0.8425
LAJOLLA*POOL	57.20759	43.61570	1.311628	0.1958
LAJOLLA*VIEW	-4.325599	38.22952	-0.113148	0.9104
R-squared	0.726726	Mean dependent var	228.6294	
Adjusted R-squared	0.676532	S.D. dependent var	112.4359	
S.E. of regression	63.94701	Akaike info criterion	11.30725	
Sum squared resid	200371.8	Schwarz criterion	11.65938	
Log likelihood	-323.5639	F-statistic	14.47855	
Durbin-Watson stat	2.185908	Prob(F-statistic)	0.000000	

SALIDA 2

Dependent Variable: PRICE

Method: Least Squares

Date: 08/25/03 Time: 18:36

Sample: 1 59

Included observations: 59

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.603636	38.30177	-0.094085	0.9254
SQFT	0.085252	0.019974	4.268191	0.0001
YARD	0.004637	0.002305	2.011589	0.0493
VIEW	43.21288	23.83722	1.812832	0.0754
POOL	23.12639	27.75670	0.833182	0.4084

R-squared	0.465839	Mean dependent var	228.6294
Adjusted R-squared	0.426272	S.D. dependent var	112.4359
S.E. of regression	85.16439	Akaike info criterion	11.80798
Sum squared resid	391660.5	Schwarz criterion	11.98404
Log likelihood	-343.3355	F-statistic	11.77328
Durbin-Watson stat	1.355576	Prob(F-statistic)	0.000001

SOLUCIÓN

a.

$$\widehat{PRECIO} = (33,134 + 65,553) + (15,888 - 4,325) + 600 \times (0,0558 + 0,021) + 5000 \times (0,001 + 0,001) = 166,33$$

El precio predicho es de 166.330 dólares.

b.

$$W^0 = \frac{0,726 - 0,465}{1 - 0,726} \times (59 - 9 - 1) = 46,67$$

Bajo H_0 , $W^0 \sim_{approx} \chi_5^2$. El valor crítico al 1% de significación es 15,1. Por tanto, se rechaza la hipótesis nula de que los efectos de las diferentes variables explicativas son idénticos en las dos localizaciones.

- 7 Considérese el siguiente modelo lineal, para los diferentes estados de Estados Unidos, que relaciona número de delitos violentos (*VIOLNTCRIME*), gasto en fuerzas de seguridad (*POLICE*), edad mediana en la población (*MEDAGE*), y una variable ficticia que toma el valor 1 si en el estado considerado existe pena capital y 0 en otro caso (*DEATH*).

$$VIOLNTCRIME = \beta_0 + \beta_1 POLICE + \beta_2 MEDAGE + \beta_3 DEATH + U,$$

donde U es un término de error.

- a. ¿Puede estar correlacionada la variable *POLICE* con el término de error U ? Ilústrelolo con un ejemplo. Comente las consecuencias sobre los estimadores de los coeficientes por MCO si la correlación tiene lugar. (0,5 punto).
- b. Suponga que tiene observaciones sobre la densidad de población (*POPDEN*) y la tasa de desempleo (*UNEMP*). Suponiendo que la variable observada (*POLICE*) está de hecho correlacionada con el término de error U , ¿cómo estimaría los coeficientes por variables instrumentales? Describa la ecuación de forma reducida para *POLICE* y el estimador de mínimos cuadrados bietápicos (1 punto).

SOLUCIÓN

- a. En principio pueden estar correlacionadas. Por ejemplo, U puede contener la inseguridad en la zona, que depende de diferentes factores y es difícil de medir. El estimador *MCO* será inconsistente.
- b. La ecuación de forma reducida es:

$$POLICE = \pi_0 + \pi_1 MEDAGE + \pi_2 DEATH + \pi_3 POPDEN + \pi_4 UNEMP + V,$$

donde V es un error que no está correlacionado con ninguna de las variables exógenas. Los coeficientes π_j pueden estimarse por *MCO* desde la ecuación de forma reducida, dando lugar a los valores predichos:

$$\widehat{POLICE}_i = \hat{\pi}_0 + \hat{\pi}_1 MEDAGE_i + \hat{\pi}_2 DEATH_i + \hat{\pi}_3 POPDEN_i + \hat{\pi}_4 UNEMP_i.$$

El estimador de mínimos cuadrados bietápicos es el estimador *MCO* en la ecuación:

$$VIOLNTCRIME_i = \beta_0 + \beta_1 \widehat{POLICE}_i + \beta_2 MEDAGE_i + \beta_3 DEATH_i + error.$$