Examen de Econometría I
Universidad Carlos III de Madrid
1aa Convocatoria (28 de enero 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. El examen se corrige sobre 10 según las puntuaciones mostradas en cada pregunta o apartado.

1. (1 punto) Suponga que:

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

y

$$PLO(Y|X) = 3 + 2X.$$

Sabemos que: $E(X) = E\left[X^3\right] = 0$ y $V(X) = 1$. Teniendo en cuenta que $\beta_2 = 1$, calcule β_0 y β_1.

SOLUCIÓN:

$$3 = E(Y) - 2E(X) = E(Y)$$

$$\begin{aligned} \beta_0 &= E(Y) - \beta_1 E(X) - E\left[X^2\right] = 3 - 1 = 2 \\ C\left[X, X^2\right] &= E\left[X^3\right] = 0 \Longrightarrow \beta_1 = 2. \end{aligned}$$

2. Utilizando una muestra de individuos con estudios secundarios de la Encuesta sobre Estructura Salarial en España para el año 1995, se ha estimado el siguiente modelo para la oferta de trabajo:

$$E(H|S,W) = \beta_0 + \beta_1 S + \beta_2 \ln W + \beta_3 (S \times \ln W),$$

donde H es el número de horas trabajadas durante el año, S es una variable ficticia que recoge el hecho de ser hombre o mujer (que toma el valor 1 si el individuo es un hombre y 0 si es mujer) y W es el salario-hora en euros. Utilizando una muestra aleatoria se ha ajustado el modelo por MCO con el siguiente resultado (errores estándar en paréntesis):

$$\begin{array}{rccccc} \widehat{H}_i = & 1429,66 & +224,96 S_i & +91,67 \ln W_i & -66,41 (S_i \times \ln W_i) \\ & (64,63) & (20,68) & (6,14) & (7,08) \end{array}$$

a. (1 punto) ¿Cual será el efecto medio sobre las horas trabajadas provocado por un aumento del 1% en el salario para la población de hombres y para la población de mujeres respectivamente?.

b. (0.75 puntos) Teniendo en cuenta que la media muestral de $\ln W$ es $\overline{\ln W_H} = 3,38$ para hombres y $\overline{\ln W_M} = 2,48$ para mujeres, proporcione los valores estimados de las horas medias trabajadas para hombres y para mujeres a partir del ajuste reportado.

SOLUCIÓN:

a. La variación media ceteris paribus para hombres será

$$\Delta E(H|S=1,W) \approx \frac{(\beta_2 + \beta_3)}{100} \%\Delta \ln W$$

y estimamos que para estos un aumento del 1% en el salario implica un aumento medio en las horas trabajadas al año de $(91,67 - 66,41)/100 = 0,25$ horas. La variación media ceteris paribus para mujeres será

$$\Delta E(H|S=0,W) \approx \frac{\beta_2}{100} \%\Delta \ln W$$

y estimamos que para las mujeres un aumento del 1% en el salario implica un aumento medio en las horas trabajadas al año de $91,67/100 = 0,91$ horas.

b. La media de horas para hombres y mujeres, según el modelo, satisface las expresiones:

$$\begin{aligned} E(H|S=1) &= \beta_0 + \beta_1 \cdot 1 + \beta_2 E(\ln W|S=1) + \beta_3 (1 \cdot E(\ln W|S=1)) \\ &= (\beta_0 + \beta_1) + (\beta_2 + \beta_3) E(\ln W|S=1) \end{aligned}$$

$$\begin{aligned} E(H|S=0) &= \beta_0 + \beta_1 \cdot 0 + \beta_2 E(\ln W|S=0) + \beta_3 (0 \cdot E(\ln W|S=0)) \\ &= \beta_0 + \beta_2 E(\ln W|S=0) \end{aligned}$$

respectivamente. Por tanto la media estimada de horas trabajadas para un hombre es:

$$\begin{aligned} \hat{E}(H|S=1) &= (1.429,66 + 224,96) + (91,67 - 66,41) \cdot 3,38 \\ &= 1.740 \; horas, \end{aligned}$$

y el estimador para mujeres es:

$$\begin{aligned} \hat{E}(H|S=0) &= 1.429,66 + 91,67 \cdot 2,48 \\ &= 1.657 \; horas \end{aligned}$$

3. Empleando una muestra de 500 empresas se ha estimado el modelo lineal que se presenta a continuación en las salidas de E-Views. Las variables consideradas son:

$$\begin{aligned} VENTAS &= \text{Ventas (cientos de euros)} \\ PUBLIC. &= \text{Gastos en publicidad (cientos de euros)} \\ SECTOR &= \begin{array}{l} 1 \text{ si la empresa pertenece al sector servicios} \\ 0 \text{ si la empresa es industrial} \end{array} \end{aligned}$$

SALIDA 1:

Dependent Variable: VENTAS
Method: Least Squares
Sample: 1 500
Included observations: 500

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	320.0229	0.402607	794.8758	0.0000
SECTOR	250.7840	0.531199	472.1096	0.0000
PUBLIC.	0.699146	0.008418	83.04965	0.0000
PUBLIC.*SECTOR	-0.021591	0.011506	-1.876562	0.0612

R-squared	0.998491	Mean dependent var	487.6030
Adjusted R-squared	0.998482	S.D. dependent var	126.0017
S.E. of regression	4.909086	Akaike info criterion	6.028020
Sum squared resid	11953.17	Schwarz criterion	6.061737
Log likelihood	-1503.005	F-statistic	109414.6
Durbin-Watson stat	1.793330	Prob(F-statistic)	0.000000

SALIDA 2:

Dependent Variable: VENTAS
Method: Least Squares
Sample: 1 500
Included observations: 500
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
c	320.0229	0.086691	3691.521	0.0000
SECTOR	250.7840	0.455807	550.1974	0.0000
PUBLIC.	0.699146	0.001846	378.6482	0.0000
PUBLIC.*SECTOR	-0.021591	0.009434	-2.288747	0.0225

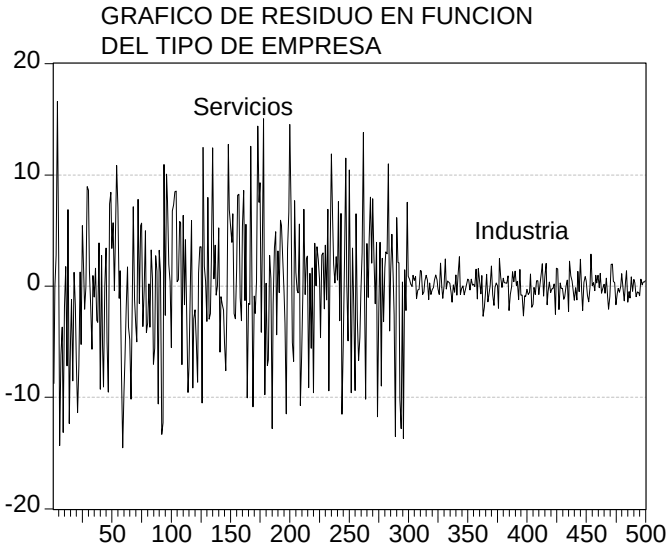

Figure 1:

R-squared	0.998491	Mean dependent var	487.6030
Adjusted R-squared	0.998482	S.D. dependent var	126.0017
S.E. of regression	4.909086	Akaike info criterion	6.028020
Sum squared resid	11953.17	Schwarz criterion	6.061737
Log likelihood	-1503.005	F-statistic	109414.6
Durbin-Watson stat	1.793330	Prob(F-statistic)	0.000000

El gráfico adjunto presenta los residuos obtenidos ordenados en función de si la empresa pertenece al sector servicios (primeras 300 observaciones) o al sector industrial (200 últimas observaciones)

a. (1 punto) Teniendo en cuenta las salidas y gráfico adjuntos, donde se observa una mayor variabilidad de los residuos en las empresas del sector servicios, contraste la hipótesis de que un aumento de una unidad en los gastos de publicidad produce en media el mismo efecto sobre las ventas en el sector servicios y el sector industrial.

b. (0.75 puntos) Las ventas de una empresa dependen también de un conjunto de variables que las podemos englobar en un cajón de sastre que denominamos REPUTACIÓN como son: servicio postventa, atención al cliente, implantación dentro del sector etc. Sabiendo que la REPUTACIÓN no ha sido incluida en el modelo por no ser observada ¿En que circustancias serán los estimadores MCO del modelo considerado en a. consistentes?

SOLUCIÓN:

a. A la vista del gráfico reportado, la varianza es claramente heterogenea entre sectores, existe heterocedasticidad. Consecuentemente, debemos utilizar la Salida 2, que proporciona un p-valor para el contraste de significación del coeficiente del efecto interactivo de 0,0225. Por lo tanto, podemos rechazar la hipótesis nula a los niveles de significación habituales.

b. Cuando las variables explicativas que se han incluido en el modelo en a. no estén correlacionadas con la variable REPUTACIÓN.

4. Empleando datos de 342 viviendas se han estimado los modelos que se presentan a continuación en los que:

PRECIO = Precio de la vivienda (en decenas de miles de euros)

M2 = metros cuadrados de la vivienda

DISTANCIA = distancia al centro de la ciudad en kilómetros

SALIDA 1

Dependent Variable: PRECIO
Method: Least Squares
Sample: 1 342
Included observations: 342

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.411209	0.132188	3.110790	0.0020
M2	0.170525	0.001044	163.3059	0.0000
DISTANCIA	-0.746720	0.103172	-7.237651	0.0000
DISTANCIA^2	0.115279	0.024852	4.638671	0.0000

R-squared	0.987683	Mean dependent var	18.80569
Adjusted R-squared	0.987574	S.D. dependent var	7.283984
S.E. of regression	0.811956	Akaike info criterion	2.432885
Sum squared resid	222.8340	Schwarz criterion	2.477737
Log likelihood	-412.0234	F-statistic	9034.914
Durbin-Watson stat	2.054491	Prob(F-statistic)	0.000000

SALIDA 2

Dependent Variable: PRECIO
Method: Least Squares
Sample: 1 342
Included observations: 342

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.269966	0.132469	2.037963	0.0423
M2	0.170656	0.001075	158.7582	0.0000
DISTANCIA	-0.325217	0.050319	-6.463066	0.0000

R-squared	0.986899	Mean dependent var	18.80569
Adjusted R-squared	0.986822	S.D. dependent var	7.283984
S.E. of regression	0.836166	Akaike info criterion	2.488754
Sum squared resid	237.0198	Schwarz criterion	2.522392
Log likelihood	-422.5769	F-statistic	12768.81
Durbin-Watson stat	2.035314	Prob(F-statistic)	0.000000

a. (1 punto) Contraste si la distancia ha de entrar de forma lineal en el modelo.

b. (0.75 puntos) Sean dos viviendas de 100 m^2 situadas a 2 y 3 kilómetros del centro de la ciudad respectivamente. ¿Cuál es la estimación de la diferencia esperada entre el precio de las dos viviendas?

SOLUCIÓN:

a. La salida 1 es contundente, el p-valor correspondiente a $DISTANCIA \wedge 2$ es cero, por lo que rechazamos la hipótesis nula de linealidad en la dirección del modelo cuadrático que aparece en la Salida 1 a cualquier nivel de significación.

b. El modelo regresión considerado es:

$$E(PRECIO|\, M2, DISTANCIA) = \beta_0 + \beta_1 M2 + \beta_2 DISTANCIA + \beta_3 DISTANCIA^2$$

$\Longrightarrow$La diferencia esperada en los precios de dos viviendas con los mismos metros cuadrados y situadas a 2 y 3 kilómetros del centro de la ciudad es:

$$\begin{aligned} & E(PRECIO|\, M2 = 100, DISTANCIA = 3) - E(PRECIO|\, M2 = 100, DISTANCIA = 2) \\ = \;& \beta_2 (3-2) + \beta_3 \left[3^2 - 2^2\right], \end{aligned}$$

cuyo valor estimado es:

$$-0,747 \cdot (3-2) + 0,115 \cdot \left[3^2 - 2^2\right] = 0,172.$$

La diferencia estimada es de $0,172 \times 10.000 = 1.720$ euros.

5. Considere la siguiente función de demanda para vino en España:

$$Q = \beta_0 + \beta_1 PV + \beta_2 PC + \beta_3 R + \beta_4 PUB + U$$

donde los $\beta's$ son parámetros, U un error de media cero, y

Q	=consumo per cápita de vino de equilibrio
PV	= precio del vino de equilibrio
PC	=precio de la cerveza
R	=renta disponible per cápita
PUB	=gasto per cápita en publicidad.

Las variables PC, R y PUB son de carácter exógeno, pero el precio PV se determina simultaneamente con la cantidad de equilibrio y, por tanto, es endógeno, correlacionado con el error U. Sabemos que los $COSTES$ de producción del vino son exógenos, se determinan fuera del modelo. Esto es, PC, R, PUB y $COSTES$ son variables exógenas que no están correlacionadas con el término de error U, mientras que Q y PV son endógenas y, por tanto, correlacionadas con U. Suponga que tenemos una muestra aleatoria.

a. (0,5 puntos) ¿Son los estimadores MCO de β en la ecuación de demanda consistentes? Justifique su respuesta.

b. (1.5 puntos) ¿Cómo contrastaría que la variable exógena $COSTES$ es un instrumento válido para estimar los parámetros β en la ecuación de demanda?. ¿Cómo computaría el estimador de mínimos cuadrados bietápicos ($MC2$)?

SOLUCIÓN:

a. Los estimadores de β son inconsistentes, al estar PV correlacionado con el término de error.

b. Para contrastar que los costes son un buen instrumento de PV, consideramos la forma reducida de PV,

$$PV = \pi_0 + \pi_1 PC + \pi_2 R + \pi_3 PUB + \pi_4 COSTES + V,$$

donde V no está correlacionada con ninguna de las variables exógenas, y realizamos el contraste:

$$H_0 : \pi_4 = 0 \text{ versus } H_1 : \pi_4 \neq 0.$$

Si rechazamos H_0, concluimos que los $COSTES$ son un instrumento válido.

Para computar el vector de estimadores $MC2$, seguimos dos estapas:
1aa Etapa: Computamos los precios ajustados del vino para cada observación por MCO :

$$\widehat{PV}_i = \hat{\pi}_0 + \hat{\pi}_1 PC_i + \hat{\pi} R_i + \hat{\pi}_3 PUB_i + \hat{\pi}_4 COSTES_i.$$

2aa Etapa: Estimamos por MCO los parámetros del modelo de demanda utilizando los valores ajustados de PV por MCO, $\widehat{PV}_i$. Esto es ajustamos por MCO el modelo lineal:

$$Q_i = \beta_0 + \beta_1 \widehat{PV}_i + \beta_2 PC_i + \beta_3 R_i + \beta_4 PUB_i + error,$$

y los valores estimados de los parámetros β son las estimaciones de $MC2$.

6. Un investigador quiere estimar una función de producción para las empresas industriales españolas. Para ello, utiliza una muestra de 891 empresas manufactureras de la Central de Balances del Banco de España. El investigador considera dos modelos alternativos:

$$\text{Translog} \qquad \ln Y = \beta_0 + \beta_l \ln L + \beta_k \ln K + \beta_{lk}(\ln L \times \ln K) + \beta_{ll}(\ln L)^2 + \beta_{kk}(\ln K)^2 + U$$

$$\text{Cobb-Douglas} \quad \ln Y_i = \delta_0 + \delta_l \ln L + \delta_k \ln K + U$$

donde Y representa la producción en términos reales, L el número de unidades de trabajo y K es el stock de capital en términos reales. El ajuste de ambos modelos por MCO es el siguiente:

Translog

$$\widehat{\ln Y_i} = \underset{(0.25)}{2.40} + \underset{(0.12)}{0.64 \ln L_i} + \underset{(0.07)}{0.32 \ln K_i} - \underset{(0.03)}{0.07(\ln L_i \times \ln K_i)} + \underset{(0.02)}{0.05(\ln L_i)^2} + \underset{(0.01)}{0.02(\ln K_i)^2}$$

$$R_1^2 = 0.799$$

Cobb-Douglas

$$\widehat{\ln Y_i} = \underset{(0.08)}{2.39} + \underset{(0.03)}{0.71 \ln L_i} + \underset{(0.02)}{0.28 \ln K_i}$$

$$R_2^2 = 0.797$$

donde R_1^2 y R_2^2 son los coeficientes de determinación en cada uno de los modelos.

a. (1 punto) Contraste al 5% de significación la especificación Cobb-Douglas frente a la Translog.

b. (0.75 puntos) Suponga que en la especificación Cobb-Douglas, el capital no se observa, de manera que se estima el modelo con $\ln L$ como única variable explicativa, y omitiendo $\ln K$. ¿Cuál esperaría que fuera el efecto sobre el coeficiente del trabajo?. Justifique su respuesta.

SOLUCIÓN:

a.

$$H_0 : \beta_{lk} = \beta_{ll} = \beta_{kk} = 0 \text{ versus } H_1 : \beta_{lk} \neq 0 \text{ó } \beta_{ll} \neq 0 \text{ó } \beta_{kk} \neq 0.$$

Suponemos que se cumple el supuesto de homoscedasticidad y Utilizamos el estadístico de Wald:

$$W^0 = \frac{\Delta R_1^2}{\left(1 - R_1^2\right)/n - k - 1} = \frac{R_1^2 - R_2^2}{\left(1 - R_1^2\right)/n - k - 1} = \frac{0,799 - 0,797}{(1 - 0,797)/(891 - 5 - 1)} = 8,719$$

$\Longrightarrow$Como $\chi^2_{3;0.95} = 7,82$, se rechaza la hipótesis nula al 5%. Se puede utilizar el contraste de la F si así se desea.

b. Si

$$PLO\left(\ln Y | \ln L\right) = \gamma_0 + \gamma_1 \ln L,$$

$\Longrightarrow$

$$\gamma_1 = \delta_l + \delta_k \frac{C\left(\ln L, \ln K\right)}{V\left(\ln L\right)} > \delta_l$$

El coeficiente poblacional del trabajo γ_{l}, que es el que se estima inconsistententemente por MCO cuando se omite $\ln K$, sería mayor que δ_{l}, el parámetro que se estima en el modelo con $\ln L$ y $\ln K$. Si $\hat{\gamma}_{\mathsf{l}}$, $\hat{\delta}_{\mathsf{l}}$ y $\hat{\delta}_{\mathsf{k}}$ son los valores estimados por MCO de γ_{l}, δ_{l} y δ_{k}, la relación poblacional se cumpliría también en la muestra:

$$\hat{\gamma}_{\mathsf{l}} = \hat{\delta}_{\mathsf{l}} + \hat{\delta}_{\mathsf{k}} \frac{S_{\ln \mathsf{L},\ln \mathsf{K}}}{S^2_{\ln \mathsf{L}}}.$$

Entonces $\hat{\gamma}_{\mathsf{l}} > \hat{\delta}_{\mathsf{l}}$ si $S_{\ln \mathsf{L},\ln \mathsf{K}} > 0$ en la muestra considerada, lo que es de esperar. Si se argumenta de forma lógica que $S_{\ln \mathsf{L},\ln \mathsf{K}} < 0$ se dará por válida la pregunta.