

UNIVERSIDAD CARLOS III DE MADRID
EXAMEN DE ECONOMETRÍA I
Convocatoria Ordinaria, 23 de Enero de 2012

Instrucciones. Lea atentamente el enunciado de cada problema antes de contestar a las preguntas relacionadas. Al final del documento con los enunciados de los problemas, se adjuntan algunos valores críticos que podrían necesitarse para completar el examen. Cada problema debe responderse en un pliego de papel. Todas las preguntas valen 1 punto. **Alguna información contenida en las salidas es redundante.**
TIEMPO TOTAL 2 HORAS Y 30 MINUTOS.

PROBLEMA 1

En un estudio reciente se examinó el efecto del atractivo físico sobre la remuneración salarial (*wage*) (medida como dólares por hora) en una población de EE.UU. Para medir el atractivo físico se elaboró un ranking de 1 a 5 de atractivo físico, basado en diferentes características de los individuos como belleza, altura, complexión e incluso simetría facial. En base a este ranking se construyó la variable *belav*, que toma valor 1 si el individuo está por debajo de la media de atractivo físico y cero en caso contrario. Utilizando una muestra representativa de la población analizada se estimó la siguiente ecuación:

$$\begin{aligned} \text{(Modelo 1) } \log(wage) &= \underset{(0.08)}{0.55} - \underset{(0.005)}{0.066educ} - \underset{(0.004)}{0.041Exp} + \underset{(0.0001)}{0.001Exp^2} \\ &\quad - \underset{(0.029)}{0.453female} - \underset{(0.041)}{0.152belavg} \\ R^2 &= 0.3597 \quad n = 1260 \end{aligned}$$

donde *educ* son los años de educación, *Exp* son los años de experiencia laboral y *female* es una variable ficticia que toma valor 1 si el individuo es mujer. (medido en número de miembros). Alternativamente, también se estimó el siguiente modelo:

$$\begin{aligned} \text{(Modelo 2) } \log(wage) &= \underset{(0.08)}{0.57} - \underset{(0.006)}{0.065educ} - \underset{(0.004)}{0.041Exp} + \underset{(0.0001)}{0.001Exp^2} \\ &\quad - \underset{(0.031)}{0.461female} - \underset{(0.074)}{0.208belavg} - \underset{(0.084)}{0.011belavfem} - \underset{(0.014)}{0.005belaveduc} \\ R^2 &= 0.3598 \quad n = 1260 \end{aligned}$$

En este caso, *belavfem* es la interacción entre *belavg* y *female*, mientras que *belaveduc* es la interacción entre *belavg* y *educ*.

- (a) A un nivel de significación del 5%, ¿qué modelo elegiría? Justifique su respuesta.
- (b) Basado en su respuesta al apartado (a), elija el modelo apropiado, y explique en base a qué lo elige, para interpretar el efecto parcial de tener un atractivo físico por debajo de la media sobre el salario. ¿Es este efecto estadísticamente significativo?

PROBLEMA 2

Un investigador social desea analizar si el nivel de ingresos anuales recibidos por una mujer, w , depende de dos factores: si la mujer es una madre trabajadora (medido con la variable binaria m , que toma el valor 1 si lo es), y si está o no casada (c , que toma el valor 1 si está casada). Se han propuesto las siguientes especificaciones:

$$(\text{Modelo 1}) \quad w = \beta_0 + \beta_1 m + \beta_2 c + \beta_3(m \cdot c) + \epsilon$$

$$(\text{Modelo 2}) \quad w = \beta_0 + \beta_1 m + \beta_2 c + \beta_3(m \cdot c) + \beta_4(1 - m) + \epsilon$$

$$(\text{Modelo 3}) \quad w = \beta_1 m + \beta_2 c + \beta_3(m \cdot c) + \beta_4(1 - m) + \epsilon$$

$$(\text{Modelo 4}) \quad w = \beta_1(m \cdot c) + \epsilon$$

- (a) Atendiendo exclusivamente a problemas de multicolinealidad, ¿Cuál o cuáles de las 4 especificaciones planteadas no son estimables por MCO? Justifique su respuesta.

En una segunda parte del mismo estudio se desea analizar los determinantes de la decisión de ser madre trabajadora (m). Se cree que la decisión de ser madre trabajadora depende del salario (w), del salario de la pareja con la que vive (sm), de si está casada o no (c) y de los años de educación (x_1). Se cree que el salario puede estar determinado conjuntamente con la decisión de ser madre y trabajar, con lo que se plantea un modelo de ecuaciones simultáneas como el siguiente:

$$m = \gamma_0 + \gamma_1 w + \gamma_2 sm + \gamma_3 c + \gamma_4 x_1 + u_1$$

$$w = \alpha_0 + \alpha_1 m + \alpha_2 c + \alpha_3(m \cdot c) + u_2$$

- (b) De acuerdo al modelo propuesto, ¿está identificada la ecuación de m ? ¿Por qué sí o por qué no?

- (c) ¿Cuál es la ecuación en forma reducida de w ?

PROBLEMA 3

Utilizando una muestra de mujeres de Bostwana se ha estimado un modelo que relaciona la fertilidad con diferentes variables explicativas. El modelo poblacional es:

$$\text{Modelo (1): } children = \beta_0 + \beta_1 educ + \beta_2 urban + \beta_3 age + u$$

donde los β s son parámetros desconocidos y u es un término de error. La descripción del resto de variables es la siguiente:

Variable	Definición
<i>children</i>	número de hijos vivos
<i>educ</i>	años de educación
<i>urban</i>	Urbana (1 si vive en una ciudad y 0 en otro caso)
<i>age</i>	Edad de la mujer

Estamos interesados principalmente en cuantificar el efecto de la educación de la mujer sobre el número de hijos en este país. Sin embargo, existen serias dudas de que en el modelo propuesto *educ* sea exógena debido a factores omitidos, principalmente la renta familiar (*income*). Por otro lado, las variables *urban* y *age* son exógenas en el modelo (1).

Tenemos una posible variable instrumental para *educ* que es *frsthalf*, que toma valor 1 si la mujer nació entre enero y junio y 0 en caso contrario. Utilizando este instrumento se propone utilizar el estimador *MC2E* en el modelo (1) para solucionar la posible endogeneidad de *educ*.

- ¿Cuáles son las propiedades que ha de cumplir *frsthalf* para ser un instrumento válido? Contraste al 5% de significación una de estas condiciones a partir de las salidas ofrecidas al final del problema.
- Contraste al 5% de significación si *educ* es una variable exógena (no correlacionada con el término de error).
- ¿Cómo afecta, según las estimaciones realizadas, los años de educación sobre el número de niños que hay en el hogar? ¿Es este efecto significativo desde el punto de vista estadístico? Utilice las salidas adecuadas para contestar a esta pregunta.

Salida 1: MCO, usando las observaciones 1–4361

Variable dependiente: children

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	−1,92082	0,0950729	−20,2036	0,0000
educ	−0,0836423	0,00611155	−13,6859	0,0000
urban	−0,218651	0,0457737	−4,7768	0,0000
age	0,174836	0,00269831	64,7946	0,0000
Media de la vble. dep.	2,267828	D.T. de la vble. dep.	2,222032	
Suma de cuad. residuos	9433,222	D.T. de la regresi $\tilde{A}^3n$	1,471419	
R^2	0,561799	R^2 corregido	0,561498	
$F(3, 4357)$	1861,978	Valor p (de F)	0,000000	
Log-verosimilitud	−7870,326	Criterio de Akaike	15748,65	
Criterio de Schwarz	15774,17	Hannan–Quinn	15757,66	

Salida 2: MCO, usando las observaciones 1–4361 ($n = 4356$)

Se han quitado las observaciones ausentes o incompletas: 5

Variable dependiente: children

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	−2,11347	0,0981688	−21,5290	0,0000
educ	−0,0741325	0,00646288	−11,4705	0,0000
urban	−0,144970	0,0470342	−3,0822	0,0021
age	0,177492	0,00272880	65,0440	0,0000
tv	−0,230221	0,0927664	−2,4817	0,0131
bicycle	0,296705	0,0499646	5,9383	0,0000
electric	−0,244708	0,0780355	−3,1358	0,0017
Media de la vble. dep.	2,268365	D.T. de la vble. dep.	2,222073	
Suma de cuad. residuos	9305,786	D.T. de la regresi $\tilde{A}^3n$	1,462789	
R^2	0,567239	R^2 corregido	0,566642	
$F(6, 4349)$	950,0703	Valor p (de F)	0,000000	
Log-verosimilitud	−7834,177	Criterio de Akaike	15682,35	
Criterio de Schwarz	15727,01	Hannan–Quinn	15698,12	

Salida 3: MCO, usando las observaciones 1–4361 ($n = 4356$)

Se han quitado las observaciones ausentes o incompletas: 5

Variable dependiente: educ

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	9,10072	0,194343	46,8282	0,0000
urban	0,839189	0,109096	7,6922	0,0000
age	−0,142201	0,00599619	−23,7151	0,0000
tv	2,64766	0,212871	12,4379	0,0000
bicycle	0,226739	0,116637	1,9440	0,0520
electric	1,94879	0,179828	10,8370	0,0000
frsthalf	−0,671785	0,104140	−6,4508	0,0000
Media de la vble. dep.	5,854224	D.T. de la vble. dep.	3,926256	
Suma de cuad. residuos	50742,99	D.T. de la regresi $\tilde{A}^3n$	3,415807	
R^2	0,244158	R^2 corregido	0,243116	
$F(6, 4349)$	234,1419	Valor p (de F)	6,3e−260	
Log-verosimilitud	−11528,36	Criterio de Akaike	23070,73	
Criterio de Schwarz	23115,38	Hannan–Quinn	23086,49	

Salida 4: MCO, usando las observaciones 1–4361

Variable dependiente: educ

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	9,08319	0,202358	44,8867	0,0000
urban	1,58564	0,110225	14,3855	0,0000
age	−0,132025	0,00634161	−20,8189	0,0000
frsthalf	−0,792281	0,110328	−7,1812	0,0000
Media de la vble. dep.	5,855996	D.T. de la vble. dep.	3,927075	
Suma de cuad. residuos	57287,51	D.T. de la regresión	3,626070	
R^2	0,148009	R^2 corregido	0,147422	
$F(3, 4357)$	252,3012	Valor p (de F)	5,8e−151	
Log-verosimilitud	−11803,61	Criterio de Akaike	23615,22	
Criterio de Schwarz	23640,74	Hannan–Quinn	23624,23	

Salida 5: MCO, usando las observaciones 1–4361

Variable dependiente: children

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	−1,19619	0,496737	−2,4081	0,0161
educ	−0,167121	0,0564996	−2,9579	0,0031
urban	−0,0837805	0,101635	−0,8243	0,4098
age	0,163690	0,00796979	20,5389	0,0000
vhat	0,0844670	0,0568329	1,4862	0,1373
Media de la vble. dep.	2,267828	D.T. de la vble. dep.	2,222032	
Suma de cuad. residuos	9428,441	D.T. de la regresión	1,471215	
R^2	0,562021	R^2 corregido	0,561619	
$F(4, 4356)$	1397,423	Valor p (de F)	0,000000	
Log-verosimilitud	−7869,221	Criterio de Akaike	15748,44	
Criterio de Schwarz	15780,34	Hannan–Quinn	15759,70	

Nota: vhat son los residuos del modelo de la salida 4

VALORES CRÍTICOS

$$Z_{0.025} = 1,96 \quad Z_{0.05} = 1,645 \quad Z_{0.01} = 2,326 \quad Z_{0.005} = 2,576$$

$$Z_{0.1} = 1,282 \quad \chi_{3,0.01}^2 = 11,34 \quad \chi_{3,0.05}^2 = 7,82 \quad \chi_{5,0.05}^2 = 11,07$$

$$\chi_{2,0.05}^2 = 5,99 \quad \chi_{2,0.01}^2 = 9,21 \quad \chi_{6,0.05}^2 = 12,59 \quad \chi_{2,0.1}^2 = 4,61$$

$$\chi_{6,0.01}^2 = 16,81 \quad \chi_{4,0.05}^2 = 9,49 \quad \chi_{3,0.1}^2 = 6,25 \quad \chi_{4,0.01}^2 = 13,28$$

Z es la normal de media cero y varianza uno y χ_q^2 es la chi cuadrado con q grados de libertad, $Pr(Z > Z_\alpha) = \alpha$, $Pr(\chi_q^2 > \chi_{q,\alpha}^2) = \alpha$.

Nótese que la distribución F se puede aproximar por la de la χ_q^2 . Esto es, $\chi_q^2 \sim q \cdot F_{q,n}$ cuando n es grande, $Pr(\chi_q^2 > \chi_{q,\alpha}^2) \simeq Pr(q \cdot F_{q,n} > \chi_{q,\alpha}^2)$, con lo que el estadístico F puede aproximarse como un estadístico de *Wald* (W) tal que $qF = W$.