UNIVERSIDAD CARLOS III DE MADRID
ECONOMETRÍA I
Curso 2012/13
EXAMEN FINAL (Convocatoria Extraordinaria)

25 de Junio de 2013

PROBLEMA 1: Determinantes del test SAT

La variable sat es la puntuación en el test SAT de aptitud escolar, $hsize$ es el tamaño de la promoción (medido en cientos de alumnos) a la que pertenece el alumno, $female$ es una variable binaria de sexo (que toma el valor 1 si el estudiante es una mujer y 0 en caso contrario), y $black$ es una variable binaria racial (que toma el valor 1 si el estudiante es de raza negra y 0 en caso contrario). Se propone el modelo siguiente para estimar los efectos de varios factores sobre los resultados del test SAT de aptitud escolar,

$$sat = \beta_0 + \beta_1 hsize + \beta_2 hsize2 + \beta_3 female + \beta_4 black + \beta_5 femaleblack + u \quad (1.1)$$

donde $hsize2$ es el cuadrado de la variable $hsize$ y la variable $femaleblack$ es el término de interacción $female \times black$.

Se considera también un modelo más general que incluye el efecto adicional de que el alumno sea deportista, mediante la variable $athlete$ (que toma el valor 1 si la observación corresponde a un estudiante deportista y 0 en caso contrario), así como la variable interacción $athleteblack = athlete \times black$.

Salida 1: estimaciones MCO utilizando las 4137 observaciones 1–4137
Variable dependiente: sat

	Coeficiente	Desv. típica	estadístico t	valor p
$const$	1028.1000	6.2902	163.44	0.0000
$hsize$	19.2971	3.8323	5.03	0.0000
$hsize2$	−2.1948	0.5272	−4.16	0.0000
$female$	−45.0910	4.2911	−10.51	0.0000
$black$	−169.8100	12.7131	−13.36	0.0000
$femaleblack$	62.3064	18.1542	3.43	0.0006

Media de la var. dependiente	1030.33
D.T. de la variable dependiente	139.401
Suma de cuadrados de los residuos	7.34791e+07
Desviación típica de la regresión ($\hat{\sigma}$)	133.369
R^2	0.0858
$\bar{R}^2$ corregido	0.0847
$F(5, 4131)$	77.52

Salida 2: estimaciones MCO utilizando las 4137 observaciones 1–4137
Variable dependiente: $usq1$ (residuos de la *Salida 1* al cuadrado)

	Coeficiente	Desv. típica	estadístico t	valor p
$const$	19456.50	1195.80	16.27	0.0000
$hsize$	25.94	728.54	0.04	0.9716
$hsize2$	−43.98	100.22	−0.44	0.6608
$female$	−3226.80	815.76	−3.96	0.0001
$black$	7445.69	2416.85	3.08	0.0021
$femaleblack$	−9217.30	3451.23	−2.67	0.0076

Suma de cuadrados de los residuos	3.2907e+12
R^2	0.0357
$\bar{R}^2$ corregido	0.0346
$F(5, 4131)$	30.61

Salida 3: estimaciones MCO utilizando las 4137 observaciones 1–4137
Variable dependiente: $usq1$ (residuos de la *Salida 1* al cuadrado)

	Coeficiente	Desv. típica	estadístico t	valor p
$const$	2024.31	607.05		
$female$	−3892.49	907.94		
$black$	26726.56	2689.55		
$femaleblack$	−11921.12	3838.87		

R^2	0.0355
$F(3, 4131)$	50.78

Salida 4: estimaciones MCO utilizando las 4137 observaciones 1–4137
Variable dependiente: *sat*

	Coeficiente	Desv. típica	estadístico t	valor p
$const$	1033.6200	6.2360	165.75	0.0000
$hsize$	18.0214	3.7787	4.77	0.0000
$hsize2$	−1.8931	0.5202	−3.64	0.0003
$female$	−48.4500	4.2450	−11.41	0.0000
$black$	−130.2500	14.1122	−9.23	0.0000
$femaleblack$	36.7119	18.3839	2.00	0.0459
$athlete$	−97.5820	11.0091	−8.86	0.0000
$athleteblack$	−59.1710	25.2181	−2.35	0.0190

1. .

   (a) *Suponiendo que el modelo (1.1) cumple los supuestos clásicos del modelo de regresión, contrastar la hipótesis de que sat depende del tamaño de la promoción.*

   Se contrasta la hipótesis

   $$H_0 : \beta_1 = \beta_2 = 0$$

   mediante un test de la F. Sin embargo no está disponible el R^2 del modelo restringido, por lo que no se puede realizar. Los contrastes t individuales que posiblemente se rechace H_0.

   (b) *Suponiendo que el modelo (1.1) cumple los supuestos clásicos del modelo de regresión, ¿es el efecto de hsize sobre sat siempre positivo?*

   No, ya que $\hat{\beta}_2$ es negativa, a partir de $hsize$ =4.39 el efecto parcial se hace negativo.

   (c) *Interpretar el coeficiente de $female$ en el modelo (1.1), suponiendo que cumple los supuestos clásicos del modelo de regresión. Contrastar si, dado un tamaño de la promoción, un hombre negro y un hombre blanco obtienen el mismo sat en promedio.*

   El coeficiente de $female$ mide el efecto sobre sat de ser mujer blanca (comparado con los hombres), ya que el efecto de ser mujer negra se mide por ese coeficiente más el de $femaleblack$.

   Hay que hacer un contraste de significatividad de la t sobre $\hat{\beta}_4$, $H_0 : \beta_4 = 0$.

   (d) *Contrastar la presencia de heterocedasticidad en el modelo (1.1) y comentar las consecuencias de los resultados en los métodos utilizados en (a).*

   Se haría un contraste de significativida global (F o LM) sobre la salida 2 (o incluso la 3).

(e) *Contrastar al nivel de significación del 1% si el efecto de la raza sobre el sat, depende del hecho de ser atleta o no.*

Contraste de significatividad de la t sobre $\hat{\beta}_7$, $H_0 : \beta_7 = 0$.

PROBLEMA 2: Rendimiento de la educación

Estamos interesados en estimar una ecuación salarial con datos de individuos varones, utilizando el logaritmo neperiano del salario mensual, $lwage$, como variable dependiente. Las variables explicativas para cada individuo son los años de educación, $educ$, su edad, age, una variable binaria que indica su estado civil, $married$ (que toma el valor uno si el individuo está casado y cero en caso contrario), y su habilidad, $abil$.

$$lwage = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 married + \beta_4 abil + e \tag{2.1}$$

donde $E(\varepsilon | \, educ, age, married, abil) = 0$. Esperaríamos que $\beta_4 > 0$.

Dado que $abil$ es inobservable, la ecuación que podemos estimar omite dicha variable, de manera que estimamos el siguiente modelo de determinación salarial:

$$lwage = \delta_0 + \delta_1 educ + \delta_2 age + \delta_3 married + u \tag{2.2}$$

Tenemos razones sólidas para creer que los años de educación, $educ$, están correlacionados con la habilidad omitida, $abil$. Por el contrario, podemos asumir que dicha variable omitida no está correlacionada con las otras dos variables explicativas observables, age y $married$.

Como posibles instrumentos para $educ$, disponemos de dos variables observables no incluidas en el modelo, que sabemos que no están correlacionadas con la habilidad:

- $urban$ (una variable binaria que toma el valor uno si el individuo reside en una ciudad de más de 50000 habitantes y cero en caso contrario);
- $feduc$ (años de educación del padre).

Nuestro objetivo es obtener estimadores consistentes de los parámetros de (2.1), β_1, β_2, β_3, a partir de la información observable.

Salida 1: estimaciones MCO utilizando las 663 observaciones 1–663
Variable dependiente: $lwage$

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	5.0578	0.1831	27.62	0.0000
$educ$	0.0597	0.0066	9.03	0.0000
age	0.0228	0.0048	4.71	0.0000
$married$	0.2101	0.0494	4.25	0.0000

Suma de cuadrados de los residuos	94.4723
R^2	0.1601
$\bar{R}^2$ corregido	0.1563
$F(3, 659)$	41.88

Salida 2: estimaciones MCO utilizando las 663 observaciones 1–663
Variable dependiente: *educ*

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	12.4497	0.9632	12.92	0.0000
age	0.0383	0.0282	1.35	0.1761
married	−0.4256	0.2890	−1.47	0.1413
urban	0.4888	0.4161	1.17	

Suma de cuadrados de los residuos	3244.91
R^2	0.0156
$F(3, 659)$	3.47

Salida 3: estimaciones MCO utilizando las 663 observaciones 1–663
Variable dependiente: *educ*

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	9.1774	0.9140	10.04	0.0000
age	0.0565	0.0257	2.20	0.0281
married	−0.3868	0.2623	−1.47	0.1407
feduc	0.2907	0.0239	12.18	0.0000

Suma de cuadrados de los residuos	2674.57
R^2	0.1886
$F(3, 659)$	51.06

Salida 4: estimaciones MC2E utilizando las 663 observaciones 1–663
Variable dependiente: *lwage*
Instrumentos: *urban*

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−0.2134	2.2910	−0.09	0.9258
educ	0.4700	0.1753	2.68	0.0073
age	0.0073	0.0142	0.51	0.6087
married	0.3964	0.1515	2.62	0.0089

Suma de cuadrados de los residuos	646.232
$F(3, 659)$	29.27

Contraste de Hausman –
Hipótesis nula: Los estimadores de MCO son consistentes
Estadístico de contraste asintótico: $\chi^2_1 = 40.42$
con valor p = 2.04972e-010

Salida 5: estimaciones MC2E utilizando las 663 observaciones 1–663
Variable dependiente: $lwage$
Instrumentos: $feduc$

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	4.4176	0.2669	16.55	0.0000
$educ$	0.1095	0.0161	6.81	0.0000
age	0.0209	0.0051	4.13	0.0000
$married$	0.2327	0.0519	4.49	0.0000

Suma de cuadrados de los residuos	102.610
$F(3, 659)$	38.79

Contraste de Hausman –
Hipótesis nula: Los estimadores de MCO son consistentes
Estadístico de contraste asintótico: $\chi^2_1 = 13.11$
con valor p = 0.000293096

2. .

(a) *Contrastar la validez de los instrumentos propuestos.*

En la estimación de las formas reducidas en las salidas 2 y 3 vemos que $urban$ no es un buen instrumento, pero $feduc$ sí, dado que son exógenas.

(b) *Explicar si se podría contrastar si los instrumentos propuestos son exógenos y como se realizaría dicho contraste.*

Contraste de sobreespecificación de Sargan/Hansen. En principio sí que podría usar porque tenemos dos instrumentos y un regresor endógeno, pero uno de ellos no es válido.

(c) *Comentar la validez de los estimadores de MCO usando la información adecuada.*

Contraste de Hausman (exogeneidad del regresor posiblemente endógeno). Rechaza, por lo que MCO no es consistente.

(d) *¿Cuál sería el signo de la correlación entre habilidad y educación? Explicar la respuesta.*

Usando la fórmula para el sesgo por variables omitidas, en ausencia de otros regresores,

$$\hat{\beta}_1^{simple} = \hat{\beta}_1 + \hat{\beta}_2 \frac{Cov\,(educ, abil)}{Var\,(educ)}$$

podemos pensar $\hat{\beta}_2 \approx \beta_2 > 0$ y por $\hat{\beta}_1 \approx \hat{\beta}_1^{MC2E} \approx \beta_1$ por lo que el modelo de regresión simple, $\hat{\beta}_1^{simple} = 0.0597$, infraestima el valor verdadero, $\beta_1 \approx \hat{\beta}_1^{MC2E} = 0.1095$, y el sesgo es negativo, lo que sólo puede ocurrir si $Cov\,(educ, abil) < 0$.

(e) *Construir e interpretar un intervalo de confianza al 5% para el efecto sobre el salario del incremento de 10 años de educación, todos los demás factores constantes, usando la información adecuada.*

IC: 10*$\left(\hat{\beta}_1^{MC2E} \pm 1.96 s.e.\left(\hat{\beta}_1^{MC2E}\right)\right)$, usando la salida 5.

.

VALORES CRÍTICOS

$Z_{0.025} = 1,96$	$Z_{0.05} = 1,645$	$Z_{0.01} = 2,326$	$Z_{0.005} = 2,576$
$Z_{0.1} = 1,282$	$\chi^2_{3,0.01} = 11,34$	$\chi^2_{3,0.05} = 7,82$	$\chi^2_{5,0.05} = 11,07$
$\chi^2_{2,0.05} = 5,99$	$\chi^2_{2,0.01} = 9,21$	$\chi^2_{6,0.05} = 12,59$	$\chi^2_{2,0.1} = 4,61$
$\chi^2_{6,0.01} = 16,81$	$\chi^2_{4,0.05} = 9,49$	$\chi^2_{3,0.1} = 6,25$	$\chi^2_{4,0.01} = 13,28$

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

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