Universidad Carlos III de Madrid
Estadística II: Introduccion a la Econometría
Examen Final, Convocatoria Ordinaria, Curso 2010-2011.
Duración del examen: 2 horas y 30 minutos.

17 de enero de 2011

**PREGUNTA 1:** 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,

$$sat = \beta_0 + \beta_1 hsize + \beta_2 hsize2 + \beta_3 female + \beta_4 black + \beta_5 femaleblack + u \tag{1}$$

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 |

Suma de cuadrados de los residuos 7.34791e+07
$R^2$ 0.0858

*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

*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

*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 |

| | |
|---|---|
| $R^2$ | 0.112819 |
| $F(3,4131)$ | 75.00976 |

**a.** Enumere los supuestos que han de cumplirse para que el método de mínimos cuadrados proporcione estimadores insesgados de los coeficientes del modelo (1). Entonces, ¿es el supuesto de homocedasticidad necesario para que el estimador sea insesgado?.

**b.** Utilizando la salida 1, ¿cuál es la diferencia estimada en el test de aptitud escolar, *sat*, entre un chico blanco y una chica negra de la misma promoción? Utilizando la salida 2, ¿cuál es la diferencia estimada en el test de aptitud escolar, *sat*, entre una chica deportista y un chico no deportista, ambos negros y de la misma promoción.

**c.** ¿Existe evidencia de heterocedasticidad en el modelo (1)? Realice un contraste para justificar su respuesta?

**d.** A la luz de su respuesta al apartado anterior, explique de forma razonada si se puede contrastar, utilizando las salidas disponibles, que el hecho de ser deportista no tiene nigún efecto sobre *sat*.

**PREGUNTA 2** El siguiente es un modelo de ecuaciones simultáneas para estudiar si la apertura de la economía (*open*) lleva a menores tasas de inflación ($inf$),

$$
\begin{aligned}
inf &= \delta_{10} + \gamma_{12} open + \delta_{11} \log(pcinc) + u_1 \\
open &= \delta_{20} + \gamma_{21} inf + \delta_{21} \log(pcinc) + \delta_{22} \log(land) + u_2.
\end{aligned}
$$

Se asume que (los logaritmos de) *pcinc* (renta per cápita) y *land* (tierra de cultivo) son exógenos en todo el ejercicio. Se han obtenido diferentes estimaciones por MCO y MC2E que se proporcionan a continuación de las cuestiones.

**a.** Obtener la forma reducida del sistema. ¿Cómo se estimarían consistentemente los parámetros de esta forma reducida?

**b.** Estudiar la identificación del sistema de ecuaciones simultáneas con las estimaciones proporcionadas. ¿Hay alguna ecuación potencialmente sobreidentificada?

**c.** Contrastar si *open* es un regresor endógeno en la primera ecuación. En función del resultado comentar sobre la consistencia del estimador de mínimos cuadrados ordinarios de los parámetros $(\gamma_{12}, \delta_{11})$.

**d.** Si se sabe que $\delta_{11} = 0$, ¿cómo cambiaría tu respuesta a la pregunta (b)? ¿Y si en cambio se sabe que $\delta_{21} = 0$ (pero se ignora el valor de $\delta_{11}$ o de cualquier otro parámetro)?

**Salida 1:** Estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: inf

| Variable | Coeficiente | Desv. típica | Estadístico $t$ | valor p |
|---|---|---|---|---|
| const | 25,1040 | 15,2052 | 1,6510 | 0,1016 |
| open | -0,215070 | 0,0946289 | -2,2728 | 0,0250 |
| lpcinc | 0,0175673 | 1,97527 | 0,0089 | 0,9929 |

| | |
|---|---|
| Suma de cuadrados de los residuos | $62127,5$ |
| $R^2$ | $0,0452$ |

**Salida 2:** Estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: open

| Variable | Coeficiente | Desv. típica | Estadístico $t$ | valor p |
|---|---|---|---|---|
| const | 116,226 | 15,8808 | 7,3187 | 0,0000 |
| inf | -0,0680353 | 0,0715556 | -0,9508 | 0,3438 |
| lpcinc | 0,559501 | 1,49395 | 0,3745 | 0,7087 |
| lland | -7,3933 | 0,834814 | -8,8563 | 0,0000 |

| | |
|---|---|
| Suma de cuadrados de los residuos | $34865,3$ |
| $R^2$ | $0,4531$ |

**Salida 3:** Estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: inf

| Variable | Coeficiente | Desv. típica | Estadístico $t$ | valor p |
|---|---|---|---|---|
| const | -12,615 | 21,0313 | -0,5998 | 0,5498 |
| lpcinc | 0,191394 | 1,98158 | 0,0966 | 0,9232 |
| lland | 2,55380 | 1,08049 | 2,3635 | 0,0198 |

| | |
|---|---|
| Suma de cuadrados de los residuos | $61903,2$ |
| $R^2$ | $0,0487$ |

**Salida 4:** Estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: open

| Variable | Coeficiente | Desv. típica | Estadístico $t$ | valor p |
|---|---|---|---|---|
| const | 117,085 | 15,8483 | 7,3878 | 0,0000 |
| lpcinc | 0,546479 | 1,49324 | 0,3660 | 0,7151 |
| lland | -7,5671 | 0,814216 | -9,2937 | 0,0000 |

| | |
|---|---|
| Suma de cuadrados de los residuos | $35151,8$ |
| $R^2$ | $0,4486$ |

**Salida 5:** Estimaciones MC2E utilizando las 114 observaciones 1–114
Variable dependiente: inf
Instrumentos: land

| Variable | Coeficiente | Desv. típica | Estadístico $t$ | valor p |
|---|---|---|---|---|
| const | 26,8993 | 15,4012 | 1,7466 | 0,0807 |
| open | -0,337487 | 0,144121 | -2,3417 | 0,0192 |
| lpcinc | 0,375823 | 2,01508 | 0,1865 | 0,8520 |

| | |
|---|---|
| Suma de cuadrados de los residuos | $63064,2$ |

Contraste de Hausman –
Hipótesis nula: Los estimadores de MCO son consistentes
Estadístico de contraste asintótico: $\chi^2_1 = 1{,}35333$
con valor p = 0,244697

**PREGUNTA 3:** Para estudiar la relación entre el precio de los alquileres y la población universitaria, en ciudades que albergan una universidad, se obtienen datos de Estados Unidos sobre las siguientes variables: *rent* es el alquiler mensual medio de los pisos en dólares en la ciudad, *pop* es el número total de habitantes de la ciudad, *avginc* es la renta media de los habitantes de la ciudad y *pctstu* es el cociente entre la población estudiantil y la población total de la ciudad. Se considera el siguiente modelo econométrico:

$$log(rent) = \beta_0 + \beta_1 log(pop) + \beta_2 log(avginc) + \beta_3 pctstu + U$$

La ecuación estimada, utilizando datos del año 1990 de 64 ciudades con Universidad es:

$$\begin{aligned} log(rent) &= \underset{(0.464)}{-3.368} + \underset{(0.027)}{0,031} log(pop) + \underset{(0.041)}{0.877} log(avginc) + \underset{(0.120)}{0.658} pctstu \\ n &= 64,\ R^2 = 0.458. \end{aligned}$$

Tenemos el ajuste por mínimos cuadrados ordinarios de la siguiente regresión auxiliar:

$$\begin{aligned} log(rent) &= \underset{(0.464)}{-3.368} \underset{(0,054)}{-0,845}\, log(pop) \\ &\quad + \underset{(0.041)}{0.877} log(pop * avginc) + \underset{(0.120)}{0.658} pctstu \\ n &= 64, R^2 = 0.458 \end{aligned}$$

**a.** Interprete los valores estimados de $\beta_1$ y $\beta_3$.

**b.** Escriba formalmente la hipótesis nula de que el porcentaje de población estudiantil no tiene efecto ceteris-paribus sobre los alquileres mensuales. Realice el contrate al 1% contra una alternativa unilateral.

**c.** Contraste la hipótesis nula al 5% de significación de que el tamaño de la población tiene el mismo efecto que la renta media en la ciudad sobre los alquileres mensuales contra una alternativa bilateral.

**PROBLEMA 4:** 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 abilidad, *abil*.

$$lwage = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 married + \beta_4 abil + U,$$

donde $E\left(U|\, educ, age, married, abil\right) = 0$. 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:

$$lwage = \delta_0 + \delta_1 educ + \delta_2 age + \delta_3 married + \varepsilon.$$

Tenemos razones sólidas para creer que los años de educación, *educ*, están correlacionados con la abilidad 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 (1), $\beta_1$, $\beta_2$, $\beta_3$, 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 |

*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 | |

*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 |

*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 |

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 |

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

**a.** Supongamos que la variable *abil* se sustituye por la variable $IQ$, el coeficiente de inteligencia, dando lugar a la ecuación

$$lwage = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 married + \beta_4 IQ + error.$$

¿Qué propiedades ha de cumplir $IQ$ para que los estimadores de mínimos cuadrados de $\beta_1$, $\beta_2$ y $\beta_3$ sean consistentes? Puede proporcionar, alternativamente, condiciones más restrictivas para que sean insesgados.

**b.** Suponiendo que tanto *urban* como *feduc* no están correlacionados con el término de error, compruebe si son buenos instrumentos utilizando contrastes de significación al 5%.

**c.** ¿Por qué razón los estimadores de MC2E en la salida 5 tienen errores estándar más pequeños que los de la salida 4?

*VALORES CRÍTICOS*:

| $N(0,1) \sim t_\infty$ |
|---|
| $\Pr(N(0,1) > 2,576) = 0,005$ |
| $\Pr(N(0,1) > 2,326) = 0,01$ |
| $\Pr(N(0,1) > 1,960) = 0,025$ |
| $\Pr(N(0,1) > 1,645) = 0,05$ |
| $\Pr(N(0,1) > 1,282) = 0,10$ |

| $\chi^2_1 \sim F_{1,\infty}$ | $\chi^2_2 \sim 2F_{2,\infty}$ | $\chi^2_3 \sim 3F_{3,\infty}$ | $\chi^2_5 \sim 5F_{5,\infty}$ |
|---|---|---|---|
| $\Pr(\chi^2_1 > 6,63) = 0,01$ | $\Pr(\chi^2_2 > 9,21) = 0,01$ | $\Pr(\chi^2_3 > 11,34) = 0,01$ | $\Pr(\chi^2_5 > 15,09) = 0,01$ |
| $\Pr(\chi^2_1 > 3,84) = 0,05$ | $\Pr(\chi^2_2 > 5,99) = 0,05$ | $\Pr(\chi^2_3 > 7,81) = 0,05$ | $\Pr(\chi^2_5 > 11,07) = 0,05$ |
| $\Pr(\chi^2_1 > 2,71) = 0,10$ | $\Pr(\chi^2_2 > 4,60) = 0,10$ | $\Pr(\chi^2_3 > 6,251) = 0,10$ | $\Pr(\chi^2_5 > 9,24) = 0,10$ |

Recordamos que una $t$ de Student con $n$ grados de libertad se comporta como una $N(0,1)$ para $n$ razonablemente grande ($n > 30$). Por otro lado, una $F$ de *Fisher* con $q$ grados de libertad en el numerador y $n$ grados de libertad en el denominador se comporta aproximadamente para $n$ grande como una $\chi^2_{(q)}/q$. Utilice esta aproximación si lo estima necesario.