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

10 de Enero de 2013

PROBLEMA 1: Nivel nutricional de los menores de 6 años

1. Los nutricionistas argumentan que una buena medida del nivel nutricional de los niños pequeños, condicional a su edad y sexo, es su altura. Un modelo económico del hogar sugeriría que la altura del niño, dados su edad y su sexo, debería depender de los recursos del hogar y de la eficiencia con que los hogares traducen esos recursos en una mejor salud de sus miembros. En este ejemplo, se proponen la renta del hogar y la educación de los padres para aproximar los recursos del hogar, y la edad de la madre para aproximar su experiencia en el uso de los recursos del hogar.

   El modelo propuesto es el siguiente:

   $$\begin{aligned}\texttt{ALTED} = \beta_0 + \beta_1\texttt{RENTAH} + \beta_2\texttt{EDADM} + \beta_3\texttt{EDUCM} + \\ \beta_4\texttt{FEM} + \beta_5\texttt{RENTAFEM} + u\end{aligned} \qquad (*)$$

   donde:

   ALTED = altura del niño (en cm) comparada con la altura de un niño bien nutrido de su misma edad y sexo.

   RENTAH = renta del hogar, medida en miles de euros.

   EDADM = edad de la madre.

   EDUCM = años de educación de la madre.

   FEM = variable binaria que toma el valor uno si el menor es una niña y cero si es un niño

   RENTAFEM = RENTAH $\times$ FEM = interacción de la variable RENTAH con la variable binaria FEM.

   Empleando datos de una muestra aleatoria de 269 niños y niñas menores de 6 años, se proporcionan las siguientes estimaciones:

Salida 1: estimaciones MCO utilizando las 269 observaciones 1–269
Variable dependiente: `ALTED`

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
`const`	33.31	6.62	5.03	0.0000
`RENTAH`	2.30	0.89	2.59	0.0101
`EDADM`	-1.15	0.29	-3.95	0.0001
`EDUCM`	1.34	0.29	4.68	0.0000
`FEM`	0.20	0.92	0.22	0.8244
`RENTAFEM`	0.21	0.13	1.60	0.1106

Suma de cuadrados de los residuos	7566.77
R^2	
$\bar{R}^2$ corregido	0.1662
$F(5, 263)$	11.6874

(NOTA: La suma de cuadrados de los residuos de una regresión como la de la *Salida 1* pero que omite tanto `FEM` como `RENTAFEM` es 7715.57).

Salida 2: estimaciones MCO utilizando las 269 observaciones 1–269
Variable dependiente: `ALTED`

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
`const`	7.29	0.77	9.47	0.0000
`RENTAH`	2.51	0.91	2.75	0.0064
`EDUCM`	0.28	0.11	2.68	0.0077
`FEM`	-0.29	0.94	-0.30	0.7607
`RENTAFEM`	0.21	0.14	1.51	0.1333

Suma de cuadrados de los residuos	8016.53
R^2	0.1332
$\bar{R}^2$ corregido	0.1200
$F(4, 264)$	10.1392

Salida 3: Matriz de correlaciones de variables seleccionadas utilizando las 269 observaciones 1–269

	EDADM	EDUCM	RENTAH
EDADM	1.00		0.03
EDUCM		1.00	0.01
RENTAH			1.00

(a) Suponga que el modelo (*) verifica todos los supuestos del modelo de regresión clásico. Contraste la hipótesis nula de que para hogares con el mismo nivel de renta, la altura del menor no depende de su sexo.

Solución: se trata de un contraste de la hipótesis nula $\beta_4 = \beta_5 = 0$ a resolver mendiante un contraste de la F ode Wald usando la nota de la Salida 1. No se rechaza.

(b) Suponga que el modelo (*) verifica todos los supuestos del modelo de regresión clásico. Utilizando las estimaciones apropiadas, manteniendo todo lo demás constante, considerando una niña de 5 años, estimar la diferencia media de altura si vive en un hogar cuya renta es de 12.000 euros respecto a otro cuya renta es de 10.000 euros.

Solución: 5,020cms.

(c) Suponga que el modelo (*) verifica todos los supuestos del modelo de regresión clásico. Contrastar la hipótesis nula de que el efecto de la educación de la madre no depende del sexo del menor.

Solución: el modelo no permite hacer ese contraste (necesitaría un regresor interactuando el sexo del menor con la educación de la madre.

(d) Se podría argumentar que el modelo (*) omite la educación del padre. Suponga que tanto la educación del padre como la de la madre afectan positivamente la altura de sus hijos y que los cónyuges tienden a tener niveles similares de educación. En un modelo que incluyera la variable EDUCP (Educación del Padre) además de las variables ya incluidas en la Salida 1, cómo se vería afectado el coeficiente estimado de EDUCM respecto al de la Salida 1.

Solución: en el modelo con la variable EDUCP, el coeficiente de la variable EDUCM sería menor, habría que argumentarlo con el problema del sesgo por omisión.

(e) En la situación de la cuestión anterior, los estadísticos t de EDUCP y EDUCM son, respectivamente, 0.8 y 0.9. Entonces, ¿podríamos afirmar que la educación del padre y de la madre no son relevantes para la altura? Justifica tu respuesta.

Solución: No se podría afirmar, ya que habría hacer un contraste de la hipótesis nula conjunta, los contrastes individuales miran a otra hipótesis diferente y además pueden estar afectados por un problema de multicolinealidad.

PROBLEMA 2: Estructura familiar y comportamiento laboral de la mujer

2. Estamos interesados en estimar el impacto de las características familiares en el comportamiento laboral de mujeres con hijos. Para ello, consideramos el modelo siguiente:

$$\begin{aligned}\texttt{hourswork} &= \beta_0 + \beta_1\texttt{age} + \beta_2\texttt{age2} + \beta_3\texttt{educ} \\ &\quad + \beta_4\texttt{educ2} + \beta_5\texttt{Marr} + \beta_6\texttt{nchild} + u,\end{aligned}$$

donde

`hourswork` = número de horas trabajados por la mujer en la última semana;

`age` = edad de la mujer (en años);

`age2` = $(\texttt{age})^2$ = edad de la mujer (en años) al cuadrado;,

`educ` = años de educación de la mujer (el número máximo de años de educación es 24);

`educ2` = años de educación de la mujer al cuadrado;

`Marr` = variable binaria que vale 1 si la mujer está actualmente casada y 0 en caso contrario;

`nchild` = número de niños en el hogar.

Es muy posible que $C(\texttt{Marr}, u) \neq 0$ y $C(\texttt{nchild}, u) \neq 0$, por la existencia de posibles variables omitidas correlacionadas con el estado civil y el número de hijos. Para considerar este sesgo potencial de omisión de variables, disponemos de la variable binaria `boy1`, que toma el valor 1 si el primer hijo que tuvo la mujer fue varón y 0 en caso contrario, y la variable binaria `mb1` que toma el valor 1 si la mujer tuvo un parto múltiple en su primer embarazo y 0 en caso contrario.

Salida 1: Estimaciones MCO usando las 662829 observaciones 1–662829
Variable dependiente: `hourswork`

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
`const`	3.3660	1.3668	2.46	0.0140
`age`	1.4447	0.0864	16.73	0.0000
`age2`	−0.0246	0.0013	−18.38	0.0000
`educ`	1.3966	0.0216	64.73	0.0000
`educ2`	−0.0204	0.0017	−12.28	0.0000
`Marr`	−5.6990	0.0514	−110.81	0.0000
`nchild`	−4.7962	0.0264	−181.77	0.0000

Media de la variable dependiente	16.8049
Desv. típica de la variable dependiente	19.0319
Suma de cuadrados de los residuos	2.18125e+08
Desv. típica residual ($\hat{\sigma}$)	18.1407
R^2	0.0914697
R^2 corregido	0.0914615
$F(6, 662822)$	11122.0

Salida 2: Estimaciones MCO usando las 662829 observaciones 1–662829
Variable dependiente: `Marr`
Errores estándar robustos a heterocedasticidad, variante HC1

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
`const`	0.28190	0.0333	8.46	0.0000
`age`	0.0267	0.0021	12.63	0.0000
`age2`	−0.00026	0.00003	−7.95	0.0000
`educ`	−0.0344	0.0005	−69.07	0.0000
`educ2`	0.0013	0.00004	31.37	0.0000
`mb1`	−0.0313	0.0045	−6.91	0.0000
`boy1`	0.0095	0.0011	8.86	0.0000

Media de la variable dependiente	0.732848
Desv. típica de la variable dependiente	0.442473
Suma de cuadrados de los residuos	127184.
R^2	0.0199276
R^2 corregido	0.0199187
$F(6, 662822)$	2264.04

Salida 3: Estimaciones MCO usando las 662829 observaciones 1–662829
Variable dependiente: `Marr`
Errores estándar robustos a heterocedasticidad, variante HC1

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
`const`	0.2875	0.0333	8.62	0.0000
`age`	0.0266	0.0021	12.60	0.0000
`age2`	−0.00026	0.00003	−7.93	0.0000
`educ`	−0.0343	0.0005	−68.94	0.0000
`educ2`	0.00125	0.00004	31.24	0.0000

Suma de cuadrados de los residuos	127209.
R^2	0.0197302
R^2 corregido	0.0197243
$F(4, 662824)$	3365.33

Salida 4: Estimaciones MCO usando las 662829 observaciones 1–662829
Variable dependiente: `nchild`

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
`const`	−0.67952	0.0640	−10.61	0.0000
`age`	0.1512	0.0040	37.39	0.0000
`age2`	−0.00277	0.00006	−44.26	0.0000
`educ`	0.1909	0.0010	195.41	0.0000
`educ2`	−0.00633	0.00008	−81.00	0.0000
`mb1`	0.6616	0.0086	76.71	0.0000
`boy1`	0.0073	0.0021	3.51	0.0004

Media de la variable dependiente	2.06760
Desv. típica de la variable dependiente	0.949626
Suma de cuadrados de los residuos	479046.
R^2	0.198561
R^2 corregido	0.198553
$F(6, 662822)$	27369.6

Salida 5: Estimaciones MCO usando las 662829 observaciones 1–662829
Variable dependiente: `nchild`

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
`const`	−0.6882	0.0643	−10.70	0.0000
`age`	0.1522	0.0041	37.47	0.0000
`age2`	−0.00278	0.00006	−44.15	0.0000
`educ`	0.1901	0.0001	193.68	0.0000
`educ2`	−0.0063	0.00008	−80.40	0.0000

Suma de cuadrados de los residuos	483299.
R^2	0.191445
R^2 corregido	0.191440
$F(4, 662824)$	39235.0

Salida 6: Estimaciones MC2E usando las 662829 observaciones 1–662829
Variable dependiente: texttthourswork
Instrumentos: `mb1 boy1`

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
`const`	9.0637	1.8731	4.84	0.0000
`age`	1.1823	0.1876	6.30	0.0000
`age2`	−0.0173	0.00241	−7.09	0.0000
`educ`	0.3205	0.1333	2.40	0.0162
`educ2`	0.01641	0.0052	3.17	0.0015
`Marr`	−16.7750	4.6990	−3.57	0.0004
`nchild`	−1.1354	0.3646	−3.11	0.0018

Suma de cuadrados de los residuos	2.37224e+08
Desv. típica residual ($\hat{\sigma}$)	18.9183
$F(6, 662822)$	5867.88

Contraste de Hausman –
Estadístico de contraste asintótico: $\chi^2_2 = 234.234$
con valor p = 1.37037e-051

Salida 7: Estimaciones MCO usando las 662829 observaciones 1–662829
Variable dependiente: texttthourswork
(Nota: `u_Marr` y `u_nchild` son los residuos de las *Salidas 2* y *4*, respectivamente)

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
`const`	9.0637	1.7958	5.05	0.0000
`age`	1.1823	0.1799	6.57	0.0000
`age2`	−0.0173	0.0023	−7.40	0.0000
`educ`	0.3205	0.1278	2.51	0.0121
`educ2`	0.01641	0.0050	3.31	0.0009
`Marr`	−16.7750	4.50511	−3.72	0.0002
`nchild`	−1.13540	0.3496	−3.25	0.0012
`u_Marr`	11.0956	4.50541	2.46	0.0138
`u_nchild`	−3.6993	0.3506	−10.55	0.0000

Suma de cuadrados de los residuos	2.18048e+08
Desv. típica residual ($\hat{\sigma}$)	18.1375
R^2	0.0917907
R^2 corregido	0.0917797
$F(8, 662820)$	8373.72

(a) Utilizando las estimaciones apropiadas, manteniendo todo lo demás constante, estime la diferencia promedio en horas semanales trabajadas por una madre de 30 años de edad comparada con una madre de 25 años de edad.

Solución: se debería usar estimaciones MC2E y evaluar las funciones de regresión estimadas para cada caso: $1,18 * 5 - 0,0173 * (30^2 - 25^2)$.

(b) Utilizando las estimaciones apropiadas, comprobar si una mayor educación incrementa la cantidad media de horas trabajadas pero a una tasa decreciente.

Solución: habría que comprobar si la primera derivada de la función de regresión respecto a la educación es positiva y si la segunda es negativa, usando la salida MC2E (no se cumple lo segundo).

(c) Razonar si, dada la información disponible, es posible contrastar la exogeneidad de los instrumentos mb1 y boy1.

Solución: no es posible porque el modelo está exactamente identificado: para hacer el contraste de Sargan-Hansen necesitaríamos sobreidentificación.

(d) Si C(mb1, u) = C(boy1, u) = 0, comprobar usando las salidas adecuadas si mb1 y boy1 son instrumentos válidos.

Solución: contraste de significatividad de los instrumentos en las salidas 2 y 4 de las formas reducidas de las variables endógenas. Claramente significativos ambos.

(e) De acuerdo con la información disponible, comprobar la exogeneidad de nchild y Marr al 5% de significación. ¿Qué consecuencias tiene tu respuesta sobre el análisis del modelo?

Solución: contraste de Hausman en las salidas 6 ó 7. Se rechaza la hipótesis nula de exogeneidad, por lo que los estimadores MCO son inconsistentes y deberemos usar los MC2E.

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 $qF = W$.