Econometría
Universidad Carlos III de Madrid
Soluciones Examen Final
27 de Mayo de 2013

1. **[6 puntos/sobre 10]** *Estamos interesados en estudiar el impacto del tamaño de la familia (número de hijos) en la cantidad de recursos invertidos en estos. Específicamente, nuestra variable dependiente, private, es una variable binaria que toma un valor de uno en caso de que los padres hayan enviado a un niño a una escuela privada y cero en caso contrario. Con el fin de responder a esta pregunta tenemos acceso a los datos del censo de los Estados Unidos para el año 1980. En concreto, queremos estimar el siguiente modelo,*

$$private = \beta_0 + \beta_1 kids + \beta_2 dropout + \beta_3 dropout * black + \beta_4 black + \beta_5 Age + u$$

donde kids es el número de niños en el hogar; droput es una variable binaria que toma un valor de uno en caso de que una madre no haya terminado la secundaria y el cero, en caso contrario; black es una variable binaria que toma un valor de uno en caso la madre sea afroamericana y cero, de lo contrario, y Age representa la edad de la madre en años. Nuestro conocimiento del problema nos hace creer que $Cov(kids, u) \neq 0$. *Para hacer frente a este problema hemos construido dos potenciales instrumentos (Angrist and Evans, AER 1998):* $mb2$ *es una variable binaria que toma un valor de uno en caso la madre haya tenido gemelos, y cero en caso contrario, y ssex que es una variable binaria que toma un valor uno en caso de que los dos primeros partos de una mujer hayan terminado siendo dos niños o dos niñas.*

Use la siguiente tabla para responder las preguntas:

Tabla 1.

	(1)	(2)	(3)	(4)	(5)
			DEPENDENT VARIABLE		
	private	kids	private	private	vmc2e
VARIABLES	OLS	OLS	2SLS	OLS	OLS
ssex		0.077800			0.000580
		(0.001940)			(0.000744)
mb2		0.850000			-0.001480
		(0.009120)			(0.003670)
dropout	-0.047200	0.289000	-0.044200	-0.044200	-0.000005
	(0.001540)	(0.006160)	(0.001920)	(0.001920)	(0.001530)
dropout*black	0.008150	0.291000	0.011400	0.011400	0.000002
	(0.002420)	(0.017600)	(0.002720)	(0.002720)	(0.002420)
black	-0.049700	0.313000	-0.046400	-0.046400	0.000004
	(0.000966)	(0.003620)	(0.001570)	(0.001570)	(0.000941)
age	0.001110	0.019900	0.001200	0.001200	0.000001
	(0.000065)	(0.000151)	(0.000072)	(0.000072)	(0.000065)
kids	0.001130		-0.009360	-0.009360	
	(0.000408)		(0.004020)	(0.004020)	
v				0.010600	
				(0.004010)	
Constant	0.067200	1.813000	0.091600	0.091600	-0.000303
	(0.002480)	(0.005030)	(0.009610)	(0.009610)	(0.002310)
F-statistic para H0 mb2 y ssex conjuntamente insignificantes		5022.77			0.4
Observaciones	647565	647565	647565	647565	647565
R^2	0.00513	0.05283	0.00412	0.00412	0

En la parte superior de cada columna se reporta la variable dependiente para cada uno de los modelos estimados. MCO y MC2E es sinónimo de mínimos cuadrados ordinarios y mínimos cuadrados dos etapas, respectivamente. OLS en la parte superior de una determinada columna indica que el modelo estimado en dicha columna se ha estimado por MCO; MC2E, que el modelo de esa columna se ha estimado por MC2E.v y vmc2e corresponden a los residuos de los modelos estimados en la columna (2) y (3), respectivamente.

1. *¿Cuál es la interpretación de β_1? [1 punto/sobre 10]*

   [0/1: 0 puntos ó 1 punto, todo bien o todo mal, sin decimales ni fracciones] Como tenemos un modelo lineal de probabilidad, β_1 indica el cambio en la probabilidad de que un niño asista a una escuela privada cuando se incrementa el número de hermanos en uno (ceteris paribus).

2. *¿Son los instrumentos ssex y mb2 instrumentos relevantes para la variable kids? Explique. ¿Tenemos evidencia de instrumentos débiles? Explique. [1 punto/sobre 10]*

   [0/1] Para comprobar la relevancia de *ssex* y *mb*2, deberíamos comprobar la significatividad de esas variables en la primera etapa de MC2E (columna 2), es decir, $H0 : \pi_{ssex} = \pi_{mb2} = 0$. El F estadístico F proporcionado en la columna 2 asociado a esta hipótesis nula es 5022.77, mucho mayor que cualquier valor crítico habitual. Es decir, no sólo tenemos evidencia de que los instrumentos son relevantes, pero también podemos descartar cualquier sospecha de instrumentos débiles porque observamos un F mayor que 10.

3. *¿Podemos probar la exogeneidad de los instrumentos en este problema? Explicar y dar una respuesta en el caso de que la exogeneidad puede ser probada. Dada la información anterior, ¿son ssex and mb2 instrumentos válidos? [1 punto/sobre 10]*

   [0/0.5 explicación; 0/0.5 realización del contraste y comprobación debilidad instrumentos] Dado el hecho de que tenemos una variable endógena ($k = 1$) y dos instrumentos excluidos ($m = 2$), el modelo de interés está sobre identificado. Por tanto, podemos hacer un contraste de sobre-identificación para comprobar la exogeneidad de los instrumentos. El contaste de sobre-identificación que hemos estudiado usa el estadístico J que tienen distribución χ^2_{m-k} bajo la nula. En concreto, $J = mF$ donde F corresponde al estadístico F para la hipótesis nula $H0 : \delta_{ssex} = \delta_{mb2} = 0$ en el modelo que tiene variable dependiente los residuos de MC2E (columna 3) y como regresores los instrumentos excluidos ($ssex$ y $mb2$) y el resto de controles ($dropout$, $dropout * black$, $black$, age, y $Immigrant$) que aparece en la columna 5. De esta forma, el estadístico J es 0.8, que está por debajo del valor crítico del 5%, 3.8. Es decir, no podemos rechazar la nula y concluimos que no hay evidencia de que los instrumentos estén correlados con el término de error.

   La validez del instrumento se define por la relevancia y exogeneidad de esos instrumentos. En la pregunta anterior hemos demostrados que los instrumentos están correlados fuertemente con la variable endógena susceptible de endogeneidad ($kids$) y no hemos encontrado correlación con el término de error. Por lo tanto tenemos evidencia apoyando la validez de los dos instrumentos.

4. *¿Tenemos evidencia que apoye la endogeneidad del tamaño de la familia? [1 punto/sobre 10]*

   [0/1] Para comprobar la endogeneidad de $kids$ necesitamos hacer un contraste de Hausman, que implica contrastar la nula $H0 : \beta_v = 0$ en la regresión auxiliar de la columna 4. El estadístico t para este contraste es aproximadamente 2.6 que es mayor que el valor crítico al nivel de significación del 5% (1.96). Es decir, rechazamos la nula y hay evidencia apoyando la endogeneidad de $kids$.

5. *Usando el modelo apropiado, ¿Tenemos evidencia de que los padres con menos niños invierten más en la educación de sus hijos? [1 punto/sobre 10]*

   [0/1] Dada la evidencia encontrada en la pregunta anterior sobre la evidencia de endogeneidad del tamaño de la familia (más le hecho de disponer de instrumentos válidos) deberíamos usar para esta pregunta y la siguiente la información de la columna 3. Dada esta salida, un niño adicional reduce la probabilidad de asistir a una escuela privada en 0.93 puntos porcentuales. Además, necesitamos contrastar la significatividad de este impacto. El estadístico t para la nula $H_0 : \beta_1 = 0$ es aproximadamente 2.3, que es mayor que el valor crítico para un nivel de significación del 5%. Por tanto, tenemos evidencia apoyando que las familias que tienen más niños en media reducen su inversión por niño medida por la probabilidad de atender una escuela privada.

6. *Contraste que el impacto el impacto de que una madre no termine secundaria es el mismo para las madres afroamericanas y las madres que no son afroamericanas. [1 punto/sobre 10]*

   [0/1] Para que el impacto de que una madre que no ha finalizado al escuela secundaria sea igual entre madres afroamericanas y no-afroamericas, debería ocurrir que $\beta_3 = 0$. Usando el modelo apropiado, columna 3, el estadístico para la nula $H0 : \beta_3 = 0$ es aproximadamente 4.2, que es mayor que el valor crítico al nivel d significación el 5%. Por tanto, tenemos evidencia de que el impacto de educación medida por la variable $dropout$ es diferente entre madres afroamericanas y el resto.

2. *[4 puntos/sobre 10] Estamos interesados en explicar el salario de un trabajador en términos de los años de educación (educ) y años de experiencia (exper) usando el siguiente modelo:*

$$log\,(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u,$$

donde asumimos que u satisface los supuestos clásicos de MCO y es homocedástico. Los parámetros estimados por MCO para una muestra de $n = 935$ observaciones se aparecen en la primera columna de la Tabla 2.

Tabla 2. Estimadores MCO. Variable Dependiente: log($wage$)

Variables	**Modelo 1**	**Modelo 2**	**Modelo 3**	**Modelo 4**	**Modelo 5**	**Modelo 6**
educ	0.07778	0.05316	0.07815	0.07192	0.071984	0.05571
	(0.00669)	(0.02085)	(0.00653)	(0.00666)	(0.00677)	(0.00600)
exper	0.01977	0.00038	0.01829	0.01791	0.01678	
	(0.00330)	(0.01066)	(0.00330)	(0.00327)	(0.01389)	
$educ * married$		0.02813				
		(0.02194)				
$exper * married$		0.01952				
		(0.01120)				
married		-0.38069	0.20926	0.18886	0.18873	0.21311
		(0.36818)	(0.04272)	(0.04764)	(0.04763)	(0.04709)
black				-0.24167	-0.24128	-0.22500
				(0.08391)	(0.08417)	(0.08212)
$married * black$				0.03599	0.03543	0.01071
				(0.09387)	(0.09404)	(0.09224)
$exper^2$					4.86e-05	
					(0.00058)	
const	5.50271	5.85694	5.32796	5.46166	5.46653	5.86609
	(0.11427)	(0.34889)	(0.11574)	(0.12017)	(0.12914)	(0.09445)
Observaciones	935	935	935	935	935	935
R^2	0.13086	0.15705	0.15420	0.18131	0.18132	0.15417

Se han considerado diferentes extensiones de este modelo para describir los efectos de estar casado (con la variable binaria married) y/o ser de raza negra (con la binaria black) o la posible no linealidad del efecto de años de experiencia. Usando la salidas adecuadas de la Tabla 2 conteste a las siguientes preguntas:

1. *Contraste si las regresiones para trabajadores casados y no-casados son iguales. Escriba las regresiones estimadas para cada uno de los grupos e interprete el coeficiente de educ. [1 punto/sobre 10]*

   [0/0.5] Para este contraste hay que considerar el modelo no restringido (Modelo 2)

   $$log\,(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 educ * married + \beta_4 exper * married + \beta_5 married + u$$

   y contrastar

   $$\begin{aligned} H_0 &: \beta_3 = \beta_4 = \beta_5 = 0 \\ H_1 &: H_0 \text{ falsa} \end{aligned}$$

con el Modelo restringido (Modelo 1) mediante el estadístico F bajo homocedasticidad:

$$\begin{aligned} F &= \frac{R^2_{nores} - R^2_{res}}{1 - R^2_{nores}} \frac{n-k-1}{q} \\ &= \frac{0.15705 - 0.13086}{1 - 0.15705} \frac{935 - 5 - 1}{3} = 9.6212 \end{aligned}$$

donde el estadístico F se distribuye bajo la nula como una $\chi^2_3/3$, con valor crítico al 5% igual a $7.81/3 = 2.603$. Por tanto se rechaza la hipótesis nula de que ambas regresiones sean iguales.

[0/0.5] Las dos regresiones estimadas serían:

$$\begin{aligned} \text{no-casados: } \log\widehat{(wage)} &= (5.85694+) + (0.05316)\, educ + (0.00038)\, exper \\ \text{casados: } \log\widehat{(wage)} &= (5.85694 - 0.36818) + (0.05316 + 0.02813)\, educ + (0.00038 + 0.01952)\, exper \\ &= 5.4888 + 0.08129 educ + 0.0199 exper. \end{aligned}$$

2. *Contraste si los efectos de la educación y experiencia dependen del estado civil. [1 p/s 10]*

   [0/1] Para el contraste hay que considerar de nuevo el modelo no restringido (Modelo 2)

   $$log\,(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 educ * married + \beta_4 exper * married + \beta_5 married + u$$

   y contrastar

   $$\begin{aligned} H_0 &: \beta_3 = \beta_4 = 0 \\ H_1 &: H_0 \text{ falsa} \end{aligned}$$

   con el Modelo restringido (Modelo 3) mediante el estadístico F bajo homocedasticidad:

   $$\begin{aligned} F &= \frac{R^2_{nores} - R^2_{res}}{1 - R^2_{nores}} \frac{n-k-1}{q} \\ &= \frac{0.15705 - 0.15420}{1 - 0.15705} \frac{935 - 5 - 1}{2} = 1.5705, \end{aligned}$$

   donde el estadístico F se distribuye bajo la nula como una $\chi^2_2/2$, con valor crítico al 5% igual a 3.00. Por tanto no se rechaza la hipótesis nula al 5%.

3. *Dado un nivel de años de educación y experiencia, ¿cuál es el grupo definido por las variables married y black con el valor medio de* $\log(wage)$ *más bajo? ¿Tiene este grupo un salario significativamente más bajo que el grupo de trabajadores no-negros y no-casados? [Use el Modelo 4 para esta pregunta] [1 punto/sobre 10]*

   [0/0.5] Para la primera pregunta nos fijamos en los coeficientes de $married$, $black$ y $married * black$ en el Modelo 4,

   $$E\left[\log(wage) \backslash educ, exper, married, black\right] = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_m married + \beta_b black + \beta_{mb} married * b$$

   y los diferentes grupos

   $$\begin{aligned} E\left[\log(wage) \backslash educ, exper, married = 0, black = 0\right] &= \beta_0 + \beta_1 educ + \beta_2 exper \\ E\left[\log(wage) \backslash educ, exper, married = 1, black = 0\right] &= \beta_0 + \beta_1 educ + \beta_2 exper + \beta_m \\ E\left[\log(wage) \backslash educ, exper, married = 0, black = 1\right] &= \beta_0 + \beta_1 educ + \beta_2 exper + \beta_b \\ E\left[\log(wage) \backslash educ, exper, married = 1, black = 1\right] &= \beta_0 + \beta_1 educ + \beta_2 exper + \beta_m + \beta_b + \beta_{mb} \end{aligned}$$

para un nivel *educ* y *exper* fijos, y la combinación que proporciona un valor predicho de $\log(wage)$ más bajo es $black = 1$ y $married = 0$.

[0/0.5] Para la comparación, se obtiene que

$$\begin{aligned} E\left[\log(wage)\,|educ, exper, married = 0, black = 1\right] &= \beta_0 + \beta_1 educ + \beta_2 exper + \beta_b \\ E\left[\log(wage)\,|educ, exper, married = 0, black = 0\right] &= \beta_0 + \beta_1 educ + \beta_2 exper \end{aligned}$$

por lo que el contraste unilateral que se solicita es

$$\begin{aligned} H_0 &: \beta_b = 0 \\ H_1 &: \beta_b < 0 \end{aligned}$$

que se realiza mediante un contraste t,

$$t = \frac{\hat{\beta}_b}{e.e.\left(\hat{\beta}_b\right)} = \frac{-0.24167}{0.08391} = -2.8801$$

que es significativo al 1% en el contraste de una cola y la hipótesis nula de igualdad de salarios se rechaza en favor de que los negros tienen un salario más bajo dentro del grupo de los no-casados.

4. *¿Qué podemos concluir sobre la posible no linealidad de la relación entre* $\log(wage)$ *y los años de experiencia? ¿Podemos concluir que los años de experiencia no tienen un efecto significativo sobre* $\log(wage)$ *en el Modelo 5? Realice dos contrastes estadísticos para responder a estas preguntas. [1 punto/sobre 10]*

   [0/0.5] Para la primera pregunta hay que contrastar en el Modelo 5 si

$$\begin{aligned} H_0 &: \beta_{exper^2} = 0 \\ H_1 &: \beta_{exper^2} \neq 0 \end{aligned}$$

mediante un contraste t,

$$t = \frac{\hat{\beta}_{exper^2}}{e.e.\left(\beta_{exper^2}\right)} = \frac{4.86e - 05}{0.00058} = 0.084$$

que no es significativo al 5% , por lo que el término no lineal no es significativo una vez que se ha incluido el término lineal.

[0/0.5] Para la segunda pregunta debemos contrastar en el Modelo 5 si

$$\begin{aligned} H_0 &: \beta_{exper} = \beta_{exper^2} = 0 \\ H_1 &: H_0 \text{ es falsa} \end{aligned}$$

comparando con el Modelo restringido (Modelo 6), con un contraste F bajo homocedasticidad:

$$\begin{aligned} F &= \frac{R^2_{nores} - R^2_{res}}{1 - R^2_{nores}} \frac{n-k-1}{q} \\ &= \frac{0.18132 - 0.15417}{1 - 0.18132} \frac{935 - 6 - 1}{2} = 15.388, \end{aligned}$$

donde el estadístico F se distribuye bajo la nula como una $\chi^2_2/2$, con valor crítico al 5% igual a 3.00, por lo que se rechaza la hipótesis nula de que años de experiencia no tiene efecto sobre $\log(wage)$, a pesar de que los dos estadísticos t no son individualmente significativos.