EXAMEN EXTRAORDINARIO DE ECONOMETRÍA

Conteste cada pregunta en un cuadernillo diferente en dos horas

1. *Considere el modelo que relaciona las variables Y, X_1, X_2 y Z como sigue*

$$\log Y = \begin{cases} \beta_0 + \beta_1 X_1 + \beta_2 \log X_2 + U \text{ si } Z \geq 0 \\ \gamma_0 + \gamma_1 X_1 + \gamma_2 \log X_2 + U \text{ si } Z < 0, \end{cases}$$

donde U es un término de error cuya esperanza condicional dados X_1, X_2 y Z es cero, pero no observamos Z. En su lugar, observamos la variable

$$W = \begin{cases} 0 \text{ si } Z \geq 0 \\ 1 \text{ si } Z < 0. \end{cases}$$

a) *Expresar el modelo anterior en una sóla ecuación en función de las variables observadas y relacionar los parámetros del nuevo modelo con $(\beta_0, \beta_1, \beta_2)$ y $(\gamma_0, \gamma_1, \gamma_2)$.*

$$\begin{aligned} \log Y &= (\beta_0 + \beta_1 X_1 + \beta_2 \log X_2) * (1 - W) + (\gamma_0 + \gamma_1 X_1 + \gamma_2 \log X_2) * W + U \\ &= \beta_0 + (\gamma_0 - \beta_0) W + \beta_1 X_1 + (\gamma_1 - \beta_1) W \cdot X_1 + \beta_2 \log X_2 + (\gamma_2 - \beta_2) W \cdot \log X_2 + U \\ &= \beta_0 + \theta_0 W + \beta_1 X_1 + \theta_1 W \cdot X_1 + \beta_2 \log X_2 + \theta_2 W \cdot \log X_2 + U, \end{aligned}$$

donde $\theta_0 = \gamma_0 - \beta_0, \theta_1 = \gamma_1 - \beta_1, \theta_2 = \gamma_2 - \beta_2$, y $E[U| X_1, X_2, W] = E[U| X_1, X_2, Z] = 0$.

b) *Proporcione expresiones, en términos de las variables observadas y de los coeficientes cuando X_1 y X_2 toman los valores x_1^0 y x_2^0, respectivamente, para aproximar la elasticidad esperada de Y respecto a X_1 y la elasticidad esperada de Y respecto a X_2.*

La elasticidad esperada de Y respecto a X_1:

$$\begin{aligned} \left.\frac{\partial E[\log Y| X_1, X_2, W]}{\partial \log X_1}\right|_{X_1 = x_1^0, X_2 = x_2^0} &= \frac{\partial E\left[\log Y| X_1 = x_1^0, X_2 = x_2^0, Z\right]}{\partial X_1 / x_1^0} \\ &= (\beta_1 + \theta_1 W) x_1^0 \\ &= \beta_1 x_1^0 \text{ si } W = 0 \ (Z \geq 0) \\ &= (\beta_1 + \theta_1) x_1^0 = \gamma_1 x_1^0 \text{ si } W = 1 \ (Z < 0). \end{aligned}$$

La elasticidad esperada de Y respecto a X_2 :

$$\begin{aligned} \left.\frac{\partial E[\log Y| X_1, X_2, W]}{\partial \log X_2}\right|_{X_1 = x_1^0, X_2 = x_2^0} &= \frac{\partial E\left[\log Y| X_1 = x_1^0, X_2 = x_2^0, Z\right]}{\partial \log X_2} \\ &= \beta_2 + \theta_2 W \\ &= \beta_2 \text{ si } W = 0 \ (Z \geq 0) \\ &= \beta_2 + \theta_2 = \gamma_2 \text{ si } W = 1 \ (Z < 0). \end{aligned}$$

c) *Proporcione una expresión para aproximar la variación porcentual esperada de Y cuando Z pasa de ser positiva a negativa.*

El efecto es el mismo que cuando W pasa de 0 a 1,

$$\begin{aligned} & 100 \frac{\Delta E\left[Y| X_1 = x_1^0, X_2 = x_2^0, W\right]}{E[Y| X_1 = x_1^0, X_2 = x_2^0, W = 0]} \\ \approx\ & 100 \left(E\left[\log Y| X_1 = x_1^0, X_2 = x_2^0, W = 1\right] - E\left[\log Y| X_1 = x_1^0, X_2 = x_2^0, W = 0\right]\right) \\ =\ & 100 \left(\left[\beta_0 + \theta_0 W + \beta_1 x_1^0 + \theta_1 x_1^0 + \beta_2 \log x_2^0 + \theta_2 \log x_2^0\right] - \left[\beta_0 + \beta_1 x_1^0 + \beta_2 \log x_2^0\right]\right) \\ =\ & 100 \left(\theta_0 + \theta_1 x_1^0 + \theta_2 \log x_2^0\right) \\ =\ & 100 \left(\gamma_0 - \beta_0 + (\gamma_1 - \beta_1) x_1^0 + (\gamma_2 - \beta_2) \log x_2^0\right). \end{aligned}$$

2. *Suponga que queremos estudiar si existe discriminación racial respecto a la concesión de hipotecas en EEUU. Con este fin se toma una muestra aleatoria de las solicitudes de hipótecas en el país, de las que se extrae información sobre la variable binaria HIP que toma el valor 1 si se concede y cero en otro caso, $BLACK$ que toma el valor 1 si el solicitante es de color y 0 en otro caso y el ratio del valor de la hipóteca solicitada sobre los ingresos anuales del solicitante, que se denota H/I. Considere el modelo de regresión*

$$HIP = \beta_0 + \beta_1 \cdot BLACK + \beta_2 \cdot H/I + \beta_3 \cdot (BLACK \cdot H/I) + U,$$

donde U tiene media condicional cero dadas todas las variables explicativas.

a) *¿Cuál es el efecto de ser negro sobre la concesión de hipótecas en el este modelo?*
El efecto es

$$\begin{aligned} & E\left[HIP \mid BLACK = 1, H/I\right] - E\left[HIP \mid BLACK = 0, H/I\right] \\ = \; & \Pr\left[HIP = 1 \mid BLACK = 1, H/I\right] - \Pr\left[HIP = 1 \mid BLACK = 0, H/I\right] \\ = \; & \beta_1 + \beta_3 \cdot H/I, \end{aligned}$$

que indica la probabilidad de concesión de la hipoteca aumenta en $\beta_1 + \beta_3 \cdot H/I$ para un solicitante negro comparado con un solicitante no negro, todo lo demás igual.

b) *Suponga que queremos contrastar la significación global de las variables explicativas $BLACK$ y H/I, para lo que contamos con el R^2 del modelo ajustado por mínimos cuadrados ordinarios. ¿Puede realizar un contraste de significación global utilizando esta información? Justifique su respuesta.*
No, por que el modelo lineal de probabilidad es heterocedástico y el estadístico F que se puede calcular a partir del R^2 no proporciona una regla válidad de inferencia (sólo lo haría bajo homocedasticidad).

c) *Obtenga una expresión para la probabilidad condicional de obtener una hipoteca para una persona blanca. Teniendo en cuenta el signo esperado de β_2, ¿es este modelo razonable para cualquier valor de H/I? Justifique su respuesta.*

$$\Pr\left(HIP = 1 \mid BLACK = 0, H/I\right) = E\left[HIP \mid BLACK = 0, H/I\right] = \beta_0 + \beta_2 \cdot H/I,$$

que es una función lineal de H/I, y si suponemos que $\beta_2 < 0$ (menor probabilidad de concesión si la hipoteca aumenta respecto a los ingresos), para valores muy altos *de* H/I la probabilidad puede llegar a ser negativa (o mayor que uno para valores muy pequeños de H/I), indicando que sólo para valores centrales de H/I el modelo lineal de probabilidad puede ser razonable. Extra bonus: el modelo lineal implica que los cambios en la probabilidad de obtener una hipoteca son siempre iguales cuando H/I aumenta en una unidad, independientemente del valor inicial de H/I, lo que no es razonable, sobre todo para valores extremos de H/I.

3. *Para estudiar como afecta la fertilidad a la oferta de trabajo, se dispone de una muestra de las mujeres casadas de entre 21 y 35 años con dos o más hijos en el censo de los EE.UU. de 1980. Para ello se estima por MCO el siguiente modelo de regresión,*

$$weeks = \beta_0 + \beta_1 morekids + \beta_2 age + \beta_3 white + U$$

donde weeks son las semanas trabajadas por las madres en 1979 y morekids es una variable binaria igual a 1 si la madre tiene más de 2 hijos. Se considera que la variable age (edad de la madre en años) y la variable binaria white (igual a 1 si es de raza blanca) son exógenas.

a) *Se sospecha que esta regresión podría no proporcionar un estimador consistente del efecto causal de la fertilidad (morekids) sobre la oferta de trabajo (weeks) por lo que se considera que la variable samesex, que es igual a 1 si los dos primeros hijos son del mismo sexo (chico-chico o chica-chica) e igual a 0 en caso contrario, podría ser un instrumento válido para estimar el modelo por variables instrumentales. Argumente porqué samesex podría ser un instrumento válido para morekids.*
Se debe argumentar que el instrumento propuesto es relevante y exógeno:

Relevancia: el instrumento debe estar correlado (condicional en otras variables exógenas) con el regresor endógeno: en este caso *samesex* tendría una correlación positiva con *morekids* por las preferencias de las mujeres sobre tener hijos de los dos sexos, lo que lleva a aumentar el número de hijos (*morekids*) cuando los dos primeros son del mismo sexo ($samesex = 1$) y no tener más cuando los dos primeros son de sexos diferentes ($samesex = 0$).

Exogeneidad: el valor *samesex* no es una decisión de las mujeres (condicional en tener al menos dos hijos) y por tanto no puede estar correlacionado con factores en U que determinan la oferta de trabajo (*weeks*) aparte de los factores observados.

b) *Investigue si samesex es un instrumento débil y si es un instrumento exógeno con la información proporcionada en la Tabla* 1.

Para comprobar si un instrumento es débil se debe calcular el estadístico F de significación del instrumento en la Forma Reducida (1^a etapa), controlando también por el resto de variable exógenas, es decir contrastar

$$H_0 : \pi_1 = 0 \quad \text{vs.} \quad H_0 : \pi_1 \neq 0$$

en la regresión

$$morekids = \pi_0 + \pi_1 samesex + \pi_2 age + \pi_3 white + v.$$

En este caso podemos usar el estadístico F robusto a la heterocedasticidad tomando el cuadrado del estadístico t para $\hat{\pi}_1$:

$$F = \hat{t}_1^2 = \left(\frac{\hat{\pi}_1}{ee\left(\hat{\pi}_1\right)}\right)^2 = \left(\frac{0{,}136}{0{,}0424}\right)^2 = 10{,}288 > 10$$

por lo que se concluye que el instrumento no es débil.

Se podría calcular también el estadístico F no robusto a la heterocedasticidad comparado R^2 entre las regresiones no restringida y restringida, columnas (2) y (1) de la Tabla 1, resp.,

$$F = \frac{R_{nr}^2 - R_r^2}{1 - R_{nr}^2}\frac{n-k-1}{1} = \frac{0{,}03850 - 0{,}01820}{1 - 0{,}0385}\frac{500-3-1}{1} = 10{,}472$$

que también da un valor superior a 10. Los supuestos utilizados sobre la homocedasticidad deben estar claros en la respuesta en este caso.

Sin embargo no se puede comprobar la exogeneidad del instrumento ya que el modelo está exactamente identificado, el número de instrumentos ($m = 1$) es igual al de regresores endógenos ($k = 1$).

c) *Utilizando la interacción de la variable samesex con la variable exógena white, se propone una nueva variable instrumental adicional: $sswhite = samesex * white$. Investigue si las variables instrumentales $(samesex, sswhite)$ son instrumentos débiles y si son instrumentos exógenos con la información proporcionada en la Tabla 1.*

Para comprobar si un instrumento es débil se debe calcular el estadístico F de significación del instrumento en la Forma Reducida (1^a etapa), controlando también por el resto de variable exógenas, es decir contrastar

$$H_0 : \delta_1 = \delta_2 = 0 \quad \text{vs.} \quad H_0 : \delta_1 \neq 0 \;\text{ ó }\; \delta_2 \neq 0$$

en

$$morekids = \delta_0 + \delta_1 samesex + \delta_2 sswhite + \delta_3 age + \delta_4 white + v.$$

En este caso sólo se puede emplear el F no robusto a la heterocedasticidad, comparado R^2 entre las regresiones no restringida y restringida, columas (3) y (1) de la Tabla 1, resp.,

$$F = \frac{R_{nr}^2 - R_r^2}{1 - R_{nr}^2}\frac{n-k-1}{2} = \frac{0{,}03851 - 0{,}01820}{1 - 0{,}03851}\frac{500-4-1}{2} = 5{,}2281 < 10,$$

por lo que los instrumentos $\{samesex, sswhite\}$ son débiles en conjunto. Los supuestos utilizados sobre la homocedasticidad deben estar claros en la respuesta en este caso.

En este caso sí es posible comprobar la exogeneidad de los instrumento ya que el modelo está sobreidentificado, el número de instrumentos ($m = 2$) es estrictamente mayor al de regresores endógenos ($k = 1$). Para efectúar el contraste de

$$H_0 : \{samesex, sswhite\} \text{ son exógenos}$$

se debe calcular el estadístico $J = mF$, donde el estadístico F se obtiene a partir del contraste de

$$H_0 : \gamma_1 = \gamma_2 = 0 \quad \text{vs.} \quad H_0 : \gamma_1 \neq 0 \text{ ó } \gamma_2 \neq 0$$

en

$$uhat = \gamma_0 + \gamma_1 samesex + \gamma_2 sswhite + \gamma_3 age + \gamma_4 white + error,$$

donde *uhat* son los residuos del modelo (estructural) estimado por MC2E. En este caso sólo se puede emplear el F no robusto a la heterocedasticidad, comparado R^2 entre las regresiones (6) y (5),

$$F = \frac{R^2_{nr} - R^2_r}{1 - R^2_{nr}} \frac{n - k - 1}{2} = \frac{0{,}00030 - 0{,}00001}{1 - 0{,}00030} \frac{500 - 4 - 1}{2} = 0{,}0718,$$

y por tanto $J = 2 * 0{,}0718 = 0{,}1436$, que no es significativo comparado con el valor crítico al 5 % de significación correspondiente a una distribución $\chi^2_{m-k} = \chi^2_1$, 3,86, por lo que no se rechaza la hipótesis nula de exogeneidad.

d) *¿Sería correcto utilizar los errores estándar reportados en la columna (4) de la Tabla 1 para contrastar la hipótesis de que el parámetro β_1 es no negativo con la información proporcionada? Justifique su respuesta.*

No. Para contrastar

$$H_0 : \beta_1 \geq 0 \quad \text{vs.} \quad H_0 : \beta_1 < 0$$

deberíamos usar el estadístico t correspondiente,

$$\hat{t}_1 = \frac{\hat{\beta}_1}{ee\left(\hat{\beta}_1\right)},$$

donde $\hat{\beta}_1$ es el estimador MC2E y el error estándar $ee\left(\hat{\beta}_1\right)$ debe usar la expresión correcta para la estimación MC2E (también teniendo en cuenta la heterocedasticidad respecto a los instrumentos). Sin embargo, la regresión (4) en la Tabla 1 es la 2^a etapa estimada por MCO y por tanto el coeficiente de *morekidshat* (obtenido en la 1^a etapa) es efectivamente igual el estimador MC2E, $\hat{\beta}_1$, pero el error estándar obtenido no es igual a $ee\left(\hat{\beta}_1\right)$ y la inferencia usando esa información no sería válida.

Nota Tabla 1: Todas las regresiones están efectuadas por mínimos cuadrados ordinarios.
La variable morekidshat son las predicciones de la regresión en la columna (3).
La variable uhat se obtiene como

$$uhat = weeks - \hat{\beta}_0 - \hat{\beta}_1 morekids - \hat{\beta}_2 age - \hat{\beta}_3 white$$

donde las estimaciones de los coeficientes corresponden a la regresión de la columna (4) en el mismo orden.

ALGUNOS VALORES CRÍTICOS: $Z_{0,90} = 1{,}282$, $Z_{0,95} = 1{,}645$, $Z_{0,975} = 1{,}96$, $\chi^2_{2,95} = 5{,}99$, $\chi^2_{2,975} = 7{,}378$, $\chi^2_{3,95} = 7{,}81$, $\chi^2_{3,975} = 9{,}3484$, $\chi^2_{4,95} = 9{,}49$, $\chi^2_{4,975} = 11{,}1433$, $\chi^2_{4,95} = 9{,}49$, $\chi^2_{4,975} = 11{,}1433$, $\chi^2_{5,95} = 11{,}07$, $\chi^2_{5,975} = 12{,}8325$, donde $\mathbb{P}(Z \leq Z_\alpha) = \alpha$ y $\mathbb{P}\left(\chi^2_m \leq \chi^2_{m,\alpha}\right) = \alpha$, Z está distribuida como una normal com media cero y varianza uno, y χ^2_m como una chi-cuadrado con m grados de libertad.

Tabla 1: Resultados de Regresión

	(1)	(2)	(3)	(4)	(5)	(6)
Variable Dependiente:	morekids	morekids	morekids	weeks	uhat	uhat
Constante	0.493	0.426	0.421	12.97	0.000000116	1.145
	(0.200)	(0.202)	(0.210)	(10.93)	(8.441)	(8.889)
morekidshat				-7.411		
				(14.39)		
samesex		0.136	0.146			-2.134
		(0.0424)	(0.127)			(5.515)
sswhite			-0.0106			2.455
			(0.135)			(5.897)
age	0.00109	0.000924	0.000928	0.453	-2.18e-10	-0.000862
	(0.00652)	(0.00650)	(0.00651)	(0.281)	(0.278)	(0.279)
white	-0.198	-0.190	-0.185	-4.152	-0.000000122	-1.268
	(0.0686)	(0.0682)	(0.0979)	(4.153)	(2.973)	(4.319)
Observaciones	500	500	500	500	500	500
R^2	0.01820	0.03850	0.03851	0.00610	0.00001	0.00030

Errores estándar robustos en paréntesis.