EXAMEN FINAL DE ECONOMETRÍA

Conteste cada pregunta en un cuadernillo diferente en dos horas y media. Todas los ejercicios tienen la misma puntuación.

1. Se desea estimar la siguiente ecuación de salarios

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + abil + u, \tag{1}$$

donde $wage$ es el salario mensual, $educ$ son los años de educación, $exper$ son los años de experiencia laboral y u satisface los supuestos habituales del modelo de regresión lineal múltiple, pero no se puede observar la capacidad del trabajador ($abil$).

Se dispone de observaciones para las calificaciones de dos exámenes ($test1$ y $test2$) que son indicadores de la capacidad ($abil$). Se supone que las calificaciones se pueden escribir como

$$test1 = \gamma_1 abil + e1, \quad \mathrm{Cov}\,(abil, e1) = 0$$

y

$$test2 = \delta_1 abil + e2, \quad \mathrm{Cov}\,(abil, e2) = 0,$$

donde $\gamma_1 > 0$ y $\delta_1 > 0$. Dado que la capacidad es la que afecta al salario, se puede suponer que $test1$ y $test2$ no están correlados con u, y también se supone que $e1$ y $e2$ no están correlados con ninguna de las variables explicativas en (1).

(a) Explique por qué una regresión MCO de (1) con $abil$ omitida producirá estimadores inconsistentes y razonar si $test1$ y $test2$ son instrumentos válidos.

(b) Si se escribe $abil$ en términos de la calificación del primer examen y se inserta el resultado en (1), se obtiene

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \alpha_1 test1 + v. \tag{2}$$

Determine el valor de α_1, escriba v en función de u y $e1$, y demuestre que $test1$ es endógena en esta ecuación. ¿Producirá una regresión MCO de (2) estimadores consistentes de β_1?

(c) Si además se supone que $e1$ y $e2$ no están correlados entre sí, ¿usaría $test2$ preferentemente como una variable de control adicional o como un instrumento para $test1$ en (2)? Razone la respuesta.

(d) Considere la ecuación (2) y las estimaciones de la Tabla 1. Contraste, si es posible, si $test2$ es un instrumento relevante para $test1$. Contraste, si es posible, si $test2$ es exógena. ¿Qué información nos proporcionan las estimaciones sobre si $test1$ es endógena o exógena (suponiendo que $test2$ es exógena)?

Table 1: Tabla de Regresiones

	(1)	(2)	(3)	(4)	(5)	(6)
Var. Dependiente:	log($wage$)	log($wage$)	log($wage$)	test1	test2	log($wage$)
educ	0.0780	0.0573	0.0478	2.637	1.258	0.00965
	(0.00680)	(0.00792)	(0.00860)	(0.243)	(0.116)	(0.0178)
exper	0.0163	0.0157	0.0179	0.239	-0.290	0.0145
	(0.0140)	(0.0140)	(0.0138)	(0.398)	(0.222)	(0.0154)
exper^2	0.000152	0.000165	-0.0000685	-0.0181	0.0307	0.000194
	(0.000588)	(0.000591)	(0.000587)	(0.0167)	(0.00916)	(0.000656)
test1		0.00579	0.00468		0.146	0.0191
		(0.000984)	(0.000999)		(0.0155)	(0.00424)
test2			0.00758	0.524		
			(0.00206)	(0.0614)		
Constante	5.517	5.214	5.194	47.02	2.672	4.514
	(0.125)	(0.131)	(0.128)	(3.874)	(2.336)	(0.239)
Observaciones	935	935	935	935	935	935
R^2	0.131	0.162	0.176	0.322	0.267	.

Errores estándar robustos entre paréntesis
Todas las regresiones se ajustan por MCO, excepto (6), que se ajusta por MC2E con test2 como instrumento para test1.

2. Se desea estimar esta ecuación

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + \beta_4 age^2 + \beta_5 yngkid + u$$

para explicar los minutos de sueño nocturnos (por semana), *sleep*, de una muestra de trabajadores, hombres y mujeres, en términos de *totwrk* (minutos trabajados por semana), *educ* (años de educación), *age* (edad en años) y *yngkid* (que es una variable binaria igual a uno si en el hogar hay niños menores de tres años). Se supone que u satisface los supuestos habituales del modelo de regresión, incluyendo homoscedasticidad condicional. Usando las estimaciones pertinentes de la Tabla 2, responder a las siguientes preguntas.

(a) Contraste si el mismo modelo de regresión es apropiado tanto para hombres como para mujeres y si hay discriminación en contra de las mujeres en el reparto de las tareas de cuidado de los niños.

(b) Contraste si el efecto de *age* sobre *sleep* depende del género y encuentre el nivel de *age* para el que el valor esperado de *sleep* es mínimo para las mujeres, todos los demás factores fijos.

(c) Construya e interprete un intervalo de confianza al 95% para el efecto sobre *sleep* del incremento de un año de educación para un hombre.

(d) Contraste si el efecto medio sobre *sleep* de un año adicional de edad (*age*) es igual al efecto de un año menos de educación (*educ*) para hombres de 20 años de edad, todos los demás factores fijos.

Table 2: Tabla de Regresiones

	(1)	(2)	(3)	(4)	(5)
Variable Dependiente:	sleep	sleep	sleep	sleep	sleep
totwrk	-0.146	-0.163	-0.182	-0.183	-0.182
	(0.0191)	(0.0207)	(0.0293)	(0.0291)	(0.0293)
educ	-11.14	-11.71	-13.05	-13.87	-7.731
	(5.747)	(5.748)	(7.767)	(7.646)	(11.58)
age	-8.124	-8.697	7.157	-9.230	
	(11.86)	(11.79)	(13.63)	(11.78)	
age^2	0.126	0.128	-0.0448	0.133	
	(0.137)	(0.136)	(0.156)	(0.136)	
yngkid	17.15	-0.0228	60.38	39.54	60.38
	(53.93)	(53.91)	(64.52)	(62.57)	(64.52)
female		-87.75	590.5	-226.2	590.5
		(35.54)	(541.6)	(162.4)	(541.6)
totwrkf*female			0.0422	0.0381	0.0422
			(0.0412)	(0.0406)	(0.0412)
educ*female			2.847	5.748	2.847
			(11.53)	(11.01)	(11.53)
age*female			-37.51		-37.51
			(24.91)		(24.91)
age^2*female			0.413		0.413
			(0.289)		(0.289)
yngkid*female			-178.7	-128.0	-178.7
			(117.6)	(109.6)	(117.6)
$age - educ$					7.157
					(13.63)
$age^2 - 41 * educ$					-0.0448
					(0.156)
Constante	3825.4	3928.6	3648.2	4010.0	3648.2
	(259.3)	(257.9)	(323.0)	(278.7)	(323.0)
Observaciones	706	706	706	706	706
R^2	0.115	0.123	0.131	0.126	0.131

Errores estándar entre paréntesis
Todas las regresiones se ajustan por MCO.

3. Considere un modelo de regresión lineal simple,

$$Y_i = \beta_0 + \beta_1 X_i + u_i,$$

y un instrumento binario Z_i para X_i.

(a) Demuestre que el estimador MC2E de β_1 se puede escribir como

$$\hat{\beta}_1^{MC2E} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{X}_1 - \bar{X}_0}$$

donde $\bar{Y}_1$ y $\bar{X}_1$ son las medias muestrales de Y_i y X_i (respectivamente) para la parte de la muestra en la que $Z_i = 1$ y $\bar{Y}_0$ y $\bar{X}_0$ son las medias muestrales de Y_i y X_i (respectivamente) para la parte de la muestra en la que $Z_i = 0$.

Pista: definiendo n_1 como el número de observaciones para las que $Z_i = 1$ y n_0 como el número de observaciones para las que $Z_i = 0$, $n = n_1 + n_0,$ se puede escribir

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\left(\sum_{i:Z_i=1} Y_i + \sum_{i:Z_i=0} Y_i\right) = \frac{n_1}{n}\bar{Y}_1 - \frac{n_0}{n}\bar{Y}_0.$$

Considere un modelo simple para estimar el efecto de poseer un ordenador personal (PC) sobre la nota media del grado para alumnos universitarios de último año,

$$GPA_i = \beta_0 + \beta_1 PC_i + u_i,$$

donde PC_i es una variable binaria indicando la posesión de un ordenador personal.

(b) ¿Por qué podría estar correlada la posesión de un PC con u_i? Explique porque PC_i está probablemente correlada con los ingresos anuales de los padres. ¿Esto significa que los ingresos de los padres es una buena variable instrumental para PC_i? ¿Por qué sí o por qué no?

(c) Suponga que, hace cuatro años, la universidad concedió becas para comprar ordenadores a la mitad de los nuevos estudiantes, y que los estudiantes que recibieron la beca fueron seleccionados aleatoriamente. Explique cómo usaría esta información para construir una variable instrumental para PC_i.

En particular, si le dicen que

• entre los estudiantes que recibieron las becas, un 90% de ellos tiene un PC y el grupo tiene un GPA medio de 3.05

• entre los estudiantes que no recibieron las becas, un 75% de ellos tiene un PC y el grupo tiene un GPA medio de 2.75.

¿Cuál sería su estimador $\hat{\beta}_1^{MC2E}$?

(d) Ahora imagine que la universidad sólo dió becas a estudiantes (seleccionados aleatoriamente) cuyos padres tenían una renta por debajo de un determinado nivel (y se dispone de la lista de estudiantes que cumplían este requisito, aunque no se disponen de datos sobre la renta de los padres). ¿Cómo modificaría su modelo y/o estrategia de estimación para obtener estimadores consistentes de β_1?

ALGUNOS VALORES CRÍTICOS: $Z_{0.90} = 1.282$, $Z_{0.95} = 1.645$, $Z_{0.975} = 1.96$, $\chi^2_{2,0.95} = 5.99$, $\chi^2_{2,0.975} = 7.378$, $\chi^2_{3,0.95} = 7.815$, $\chi^2_{3,0.975} = 9.348$, $\chi^2_{4,0.95} = 9.488$, $\chi^2_{4,0.975} = 11.143$, $\chi^2_{5,0.95} = 11.071$, $\chi^2_{5,0.975} = 12.833$, $\chi^2_{6,0.95} = 12.592$, $\chi^2_{6,0.975} = 14.449$, 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 estándar con media cero y varianza uno, y χ^2_m como una chi-cuadrado con m grados de libertad.