EXAMEN FINAL DE ECONOMETRÍA

Conteste cada pregunta en un cuadernillo diferente en dos horas y media. El valor de cada apartado es un punto sobre diez.

1. Sean (Y, X, W) tres variables aleatorias relacionadas por el siguiente modelo de regresión lineal

$$Y = \beta_0 + \beta_1 X + \beta_2 W + u, \tag{1}$$

donde se sabe que

$$\mathbb{E}\left(u \mid X, W \right) = \mathbb{E}\left(u \mid W \right) = \gamma_0 + \gamma_2 W. \tag{2}$$

(a) Escriba la ecuación (2) como un modelo de regresión lineal para explicar la variable u en función de W donde deberá demostrar que el nuevo término de error v satisface $\mathbb{E}\left(v \mid X, W \right) = 0$. Sustituya ese modelo para u en (1) e interprete los coeficientes del modelo obtenido que relaciona Y con X, W y v.

(b) Se observa una muestra aleatoria simple $\{Y_i, X_i, W_i\}_{i=1}^{n}$ de (Y, X, W) y se considera el estimador de mínimos cuadrados ordinarios (MCO) de los coeficientes en el modelo (1) donde Y es la variable explicada y (X, W) las variables explicativas. Justifique si este estimador es sesgado o insesgado para los parámetros $(\beta_0, \beta_1, \beta_2)$ cuando se cumple (2). ¿Qué ocurriría si en la regresión no se incluye el regresor W?

(c) Un estudio posterior confirma que la condición (2) no es sostenible, ya que en realidad se cumple que

$$\mathbb{E}\left(u \mid X, W \right) = \alpha_0 + \alpha_1 X + \alpha_2 W, \quad \alpha_1 \neq 0, \tag{3}$$

pero se dispone de datos de otra variable Z que satisface

$$\mathbb{E}\left(u \mid Z, W \right) = \mathbb{E}\left(u \mid W \right) = \delta_0 + \delta_2 W, \quad \delta_2 \neq 0. \tag{4}$$

¿Cuál sería ahora su estrategia de estimación del modelo (1) con las restricciones (3)-(4)? ¿Qué coeficientes de ese modelo se podrían estimar consistentemente? Justifique su respuesta.

2. Utilizando datos de sección cruzada en Estados Unidos para todos los estados, se quiere estudiar la relación entre el consumo de cannabis y los ingresos de los individuos. Con este fin se considera el sistema de ecuaciones

$$\log(\mathit{ingresos}) = \beta_0 + \beta_1 \mathit{cannabis} + u_1 \tag{5}$$

$$\mathit{cannabis} = \gamma_0 + \gamma_1 \log(\mathit{ingresos}) + \gamma_2 \mathit{multa} + \gamma_3 \mathit{prisión} + u_2, \tag{6}$$

donde *cannabis* es el consumo mensual de cannabis, *multa* es la multa estándar impuesta por posesión de cannabis en el estado de residencia y *prisión* es una variable binaria que toma el valor 1 si en el estado de residencia se puede condenar a un individuo a prisión por posesión de *cannabis* para consumo personal. Se supone que tanto *multa* como *prisión* varían entre estados y $\gamma_1 \neq 0$.

(a) Si las variables *multa* y *prisión* son exógenas (no están correlacionadas con ningún término de error), encuentre la forma reducida de la variable *cannabis* y demuestre que *cannabis* es endógena en la ecuación (5).

(b) ¿Qué restricciones deben cumplir los coeficientes $\gamma' s$ en la ecuación (6) para poder estimar sin sesgo los coeficientes $\beta' s$ en la ecuación (5)? Explique si sería posible contrastar esas restricciones, directa o indirectamente.

(c) Explique cómo estimaría los coeficientes $\beta's$ de la ecuación (5) y cómo obtendría errores estándard válidos bajo heterocedasticidad.

(d) Explique cómo contrastaría que los instrumentos para estimar los parámetros $\beta's$ en la ecuación (5) son realmente exógenos; esto es, que ninguno está correlacionado con el término de error. En particular, explique como obtendría el estadístico del contraste y sus valores críticos o p-valores, para implementarlo.

3. Se quiere investigar la función de demanda de subscripciones de revistas científicas por parte de las bibliotecas universitarias. Para ello se dispone de datos para 180 revistas de economía referentes al año 2000. Se considera el modelo de regresión lineal

$$\log(subs) = \beta_0 + \beta_1 \log(ppcite) + \beta_2 \left(\log(ppcite)\right)^2 + \beta_3 \log(age) + \beta_4 \log(ppcite) * \log(age) + u, \quad (7)$$

donde *subs* es el número de bibliotecas que están subscritas a una determinada revista, *ppcite* es el precio de la subscripción divido por el número total de citas de los artículos publicados en esa revista y *age* es la antigüedad de la revista en años. $\log(x)$ es el logaritmo neperiano de x. El término de error u satisface los supuestos habituales del modelo de regresión múltiple clásico que garantizan que las estimaciones MCO son insesgadas.

(a) Proporcione una expresión en términos de los coeficientes del modelo (7) que aproxime la variación esperada en el número de subscripciones cuando el precio por cita aumenta un 1% para una revista de 20 años con $ppcite = 10$. Explique como contrastaría que dicho efecto es igual o mayor que -1 y bajo qué condiciones es válido el contraste propuesto.

(b) Contraste si la elasticidad de la demanda de subscripciones a los precios depende de la antigüedad de la revista en el modelo (7) usando las estimaciones de la Tabla 1. ¿Bajo que condiciones es este contraste válido?

(c) Alternativamente se decide estimar el siguiente modelo usando la variable binaria $age40$ que es igual a 1 si la antigüedad de la revista es igual o inferior a 40 años, y 0 en caso contrario,

$$\begin{aligned}\log(subs) &= \beta_0 + \beta_1 \log(ppcite) + \beta_2 \left(\log(ppcite)\right)^2 + \beta_3 age40 \\ &\quad + \beta_4 \log(ppcite) * age40 + \beta_5 \left(\log(ppcite)\right)^2 * age40 + u.\end{aligned} \quad (8)$$

Contraste si la elasticidad de la demanda de subscripciones a los precios es constante para todos los niveles de precios en el modelo (8) usando las estimaciones de la Tabla 1. ¿Bajo que condiciones es este contraste válido?

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.

Table 1: Tabla de Regresiones

	(1)	(2)	(3)	(4)	(5)	(6)
Variable Dependiente:	log($subs$)	log($subs$)	log($subs$)	log($subs$)	log($subs$)	log($subs$)
log($ppcite$)	-0.561	-0.460	-1.468	-0.382	-0.455	-0.351
	(0.0449)	(0.0446)	(0.385)	(0.0369)	(0.0474)	(0.0597)
log($ppcite$)^2	-0.0291		0.00721		-0.0142	0.0164
	(0.0177)		(0.0219)		(0.0170)	(0.0197)
log(age)		0.531	0.556			
		(0.222)	(0.247)			
log($ppcite$) $*$ log(age)			0.257			
			(0.106)			
$age40$				-0.534	-0.624	-0.491
				(0.117)	(0.126)	(0.146)
log($ppcite$) $*$ $age40$				-0.205		-0.228
				(0.0805)		(0.0854)
log($ppcite$)^2 $*$ $age40$						-0.0272
						(0.0415)
Constant	4.839	2.724	2.699	5.054	5.050	5.028
	(0.0680)	(0.841)	(0.934)	(0.0766)	(0.0827)	(0.0822)
Observaciones	180	180	180	180	180	180
R^2	0.564	0.584	0.607	0.632	0.620	0.633

Errores Estándard en paréntesis