



Pontificia Universidad Católica de Chile
Instituto de Economía

Examen de Grado
Econometría
Otoño, 2009

Duración : 150 minutos
Fecha : 29 de julio de 2009
Hora de comienzo : 15 horas

El examen consta de cuatro preguntas, cada una de ellas de 30 puntos. Todas las partes deben ser respondidas. Usted dispone de 15 minutos para leer el examen y 135 minutos para responderlo, por lo que cada punto corresponde aproximadamente a un minuto.

Pregunta 1 (30 puntos)

Un investigador obtiene una muestra de 2.185 hombres entre 35 y 42 años en el año 2009. La muestra incluye años de escolaridad (ESC), edad (EDA), y sus ingresos en pesos (ING). El investigador hace una regresión entre el $\log(\text{ING})$ y ESC, obteniendo los resultados que se presentan en la columna (1) del cuadro adjunto (errores estándares entre paréntesis, RSS = suma de residuos al cuadrado).

Variable	(1)	(2)	(3)
ESC	0.1032 (0.0058)	0.1134 (0.0086)	0.1224 (0.0059)
EXPOT	–	0.0103 (0.0065)	–
EXEFE	–	–	0.0435 (0.0038)
Constante	6.0131 (0.0787)	5.6700 (0.2305)	5.0603 (0.1125)
R ²	0.1274	0.1284	0.1776
RSS	1048.0	1046.8	987.7

Luego el investigador se da cuenta que, posiblemente, la experiencia es un importante determinante de los ingresos, pero que la variable no está disponible en la muestra. Por ello, construye una medida de experiencia potencial, definida como:

$$\text{EXPOT} = \text{EDA} - \text{ESC} - 6$$

y hace una regresión de $\log(\text{ING})$ contra ESC y EXPOT, obteniendo los resultados de la columna (2). Posteriormente descubre que en la muestra sí existía una medida de la experiencia efectiva, EXEFE, por lo cual el investigador hace una tercera regresión, que usted considerará como la correcta, esta vez $\log(\text{ING})$ contra ESC y EXEFE, obteniendo los resultados de la columna (3). La matriz de correlación de regresores es:

	ESC	EXPOT	EXEFE
ESC	1.00		
EXPOT	-0.74	1.00	
EXEFE	0.28	0.47	1.00

- (a) ¿En cuánto cambia el ingreso de un individuo si éste tiene un año más de educación? ¿En cuánto difiere el ingreso de dos individuos si uno de ellos tiene un año más de educación? **5 puntos**
- (b) Explique intuitiva y matemáticamente por qué el coeficiente estimado de S es más pequeño en la columna (1) que en la columna (3). **5 puntos**

- (c) Suponga que el investigador no sabía que existían datos de EXEFE
- a. Explique cuáles son los potenciales beneficios que, en teoría, el investigador deseaba obtener al incluir EXPOT en la especificación. **5 puntos**
 - b. Evalúe si la inclusión en EXPOT en la regresión efectivamente mejora los resultados de la forma cómo el investigador esperaba. Defina lo que es una “mejor estimación”. **5 puntos**
- (d) La reducción en la RSS cuando se añade EXPOT en la especificación de la columna (2) es mucho menor que la reducción cuando se añade EXEFE en la columna (3). Muestre la relación que existe entre la RSS de cada ecuación y los respectivos estadísticos t para EXPOT y EXEFE. **5 puntos**
- (e) Suponga que el largo del periodo de experiencia laboral está correlacionado con determinantes no observables del ingreso, tales como motivación y responsabilidad. ¿Cómo podría esto afectar la estimación de los coeficientes de la tercera especificación? Si el largo del periodo de experiencia laboral estuviese efectivamente correlacionado con determinantes no observables del ingreso, indique un procedimiento de estimación adecuado. **5 puntos**

Pregunta 2 (30 puntos)

1. **(8 puntos)** Un economista se basa en la teoría del Q de Tobin y propone el siguiente modelo para la inversión en viviendas en Chile:

$$I = \beta_1 + \beta_2 Q$$

donde I es el nivel de inversión bruta en viviendas (en pesos del año 2003), Q es el coeficiente propuesto por Tobin (que se define como el valor de mercado de la inversión en viviendas dividido por el costo de construcción de las viviendas).

En equilibrio, el valor del Q de Tobin debiera fluctuar en torno a la unidad más un porcentaje por riesgo (si el Q sube (baja), tiende a aumentar (bajar) la inversión y la oferta de casas, lo que termina haciendo caer (subir) el valor de mercado de las casas y el valor del Q).

La muestra contiene datos anuales desde 1960 hasta 2008.

- a. **(5 puntos)** ¿Qué esperaría Ud. respecto de la significancia estadística del estimador de β_2 ? ¿Por qué?
 - b. **(3 puntos)** ¿En torno a qué valor debería estar el estadígrafo de Durbin y Watson? ¿Por qué razones?
2. **(5 puntos)** Para hacer un análisis de las variables que más influyeron en los puntajes en la prueba PSU del año 2007 se estimó el siguiente modelo:

$$PSU = \beta_1 + \beta_2 NEM + \beta_3 CPP + \beta_4 CPS + \beta_5 AEP + \beta_6 AEM$$

donde, NEM son las notas de enseñanza media, CPP es una variable muda (dummy) que toma el valor uno si el alumno proviene de un colegio particular pagado, CPS es una variable muda que toma el valor uno si el alumno proviene de un colegio particular subvencionado, AEP son los años de educación del padre y AEM son los años de educación de la madre.

Los resultados del estudio que se reportaron fueron el R^2 , el error estándar del modelo, el test de heterocedasticidad y el de normalidad de los residuos, pero se omitieron los estadígrafos de autocorrelación y los tests CUSUM, CUSUM cuadrado y coeficientes recursivos. ¿Cuán importante es la omisión señalada? ¿Por qué?

3. **(6 puntos)** ¿Qué tan importante es averiguar si hay colinealidad entre las variables de un modelo? ¿Cuándo la colinealidad se puede calificar de "seria"?
4. **(5 puntos)** En varios estudios en ciencias sociales se usa la siguiente metodología. Se calculan los coeficientes de correlación simple entre la variable dependiente y cada variable explicativa y se ordenan de mayor a menor. Se comienza estimando una regresión simple entre la variable dependiente y la variable explicativa con mayor r^2 , siempre que éste sea significativo. Luego se añade aquella variable que ocupa el segundo lugar en el ranking y se hace un test t de significancia respecto a ésta. Se repite el mismo proceso de incorporación de variables adicionales hasta que la última variable incorporada no supera un determinado porcentaje de error tipo I. ¿Qué comentario le merece el procedimiento descrito?

5. **(6 puntos)** Se estima el siguiente modelo

$$Y = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{2t-1} + \alpha_4 Z_{2t} + \alpha_5 Z_{2t-1} + \alpha_6 Y_{t-1} + u_t$$

El R^2 es de 0.97. En un segundo modelo, para la misma variable, se estimó

$$\Delta Y_t = \beta_1 + \beta_2 \Delta X_{2t} + \beta_3 \Delta Z_{2t} + \beta_4 X_{2t-1} + \beta_5 Z_{2t-1} + \beta_6 Y_{t-1} + u_t$$

donde Δ es el operador primera diferencia. El R^2 obtenido fue de 0.70.

- a. **(4 puntos)** ¿Cuál modelo tiene “mejor ajuste”? ¿Por qué? ¿Cuál ecuación debiera tener un valor mayor para el estimador de la varianza de los errores?
- b. **(2 puntos)** ¿Qué valor debiera tener α_6 y β_6 para que el modelo no sea explosivo? (el modelo sería explosivo si ante un shock en u_t , la variable dependiente empieza a crecer o decrecer indefinidamente).

Pregunta 3 (30 puntos)

- A. Se estudia la relación entre consumo e ingreso contemporáneo usando datos trimestrales. Una inspección casual de los residuos de la regresión sugiere que la varianza de los errores del cuarto trimestre sería mayor que la de otros trimestres. Explique de manera precisa cómo verificar estadísticamente este hecho. **(5 puntos)**
- B. Una regresión con información anual de 16 años entre la variable "participación del trabajo en el valor agregado" (representada por Y_t) y el tiempo (t) entrega los siguientes resultados:

$$(1) \quad Y_t = 0,45 - 0,004 t \quad R^2 = 0,53 \\ \quad \quad \quad \quad \quad \quad (-3,96) \quad \quad \quad \quad \quad \quad DW = 0,83$$

La cifra entre paréntesis es el estadístico t . ¿Qué problemas econométricos tendría este modelo y qué efectos producen dichos problemas? **(5 puntos)**.

Otro economista considera que Y_t presenta un comportamiento no lineal y que, por lo tanto, quedaría mejor descrita por el siguiente modelo:

$$(2) \quad Y_t = 0,479 - 0,013 t + 0,0005 t^2 \quad R^2 = 0,66 \\ \quad \quad \quad \quad \quad \quad (-3,27) \quad (2,78) \quad \quad \quad \quad \quad \quad DW = 1,9$$

Estadísticos t entre paréntesis. ¿Qué opina de este segundo modelo? ¿Cómo cambia su opinión, si es que cambia, respecto de los problemas presentes en el modelo (1)? **(5 puntos)**

- C. Para calcular el efecto del ingreso (Y) en el consumo de alimentos (CA) Ud. estima el modelo $CA = C + \gamma Y + \pi$ que satisface todos los supuestos del modelo de regresión lineal. Cuando Ud. presenta su estudio a otro economista, éste le aconseja separar la estimación en los dos grupos de que dispone de información: *Pan* y *Resto de los Alimentos*. En este caso se estiman los modelos: $CP = A + \alpha Y + \varepsilon$ y $CR = B + \beta Y + \mu$ que también satisfacen los supuestos del modelo de regresión lineal, con ε y μ independientes, con varianza σ_ε^2 y σ_μ^2 . Además, todos los hogares consumen ambas categorías de alimentos, por lo que la base de datos del modelo conjunto tiene el mismo número de observaciones que las de los modelos para CP y CR. Este economista afirma que $\hat{\alpha} + \hat{\beta}$ es un mejor estimador del efecto ingreso que $\hat{\gamma}$. Explique si está de acuerdo con esta afirmación y discuta la noción de mejor estimador. **(6 puntos)**.
- D. Para estudiar la relación entre ahorro e ingreso se plantea el modelo $A = B + \beta Y + \mu$ y se estima con datos de una población de hogares del país. Como se teme que este caso puede haber heteroscedasticidad, se aplica el test respecto del ingreso y se obtiene que no se puede aceptar la hipótesis nula de homoscedasticidad. Sin embargo, otro experto en el tema indica que lo que sucede es que los datos mezclan hogares urbanos y rurales que tienen diferentes patrones de ahorro. Al hacer un test de cambio estructural urbano-rural tampoco se puede aceptar la hipótesis de que un solo modelo explique el comportamiento de ambos sectores.
- Si se acepta la existencia de heteroscedasticidad ¿Qué opina de la aplicabilidad del test de cambio estructural? **(3 puntos)**
 - Si acepta la existencia de cambio estructural ¿Qué opina del modelo y de la presencia de heteroscedasticidad? **(3 puntos)**
 - ¿Qué propone en definitiva? **(3 puntos)**

Pregunta 4 (30 puntos)

Considere una muestra aleatoria de individuos provenientes de un corte transversal de tamaño N en que usted observa si el individuo está trabajando o no. Denote la observación del estatus laboral del individuo i con y_i , que toma el valor de 0 si no trabaja y 1 si trabaja.

1. **(5 puntos)** Supongamos inicialmente que los datos provienen de una distribución Bernoulli. ¿Cómo estimaría la proporción de individuos que están trabajando? ¿Cómo calcularía la varianza de ese estimador?. Note que se espera que las respuestas consideren estimadores insesgados y eficientes.

Supongamos ahora que usted descubre que existe más estructura en este problema de modo que nota que la decisión de participar o no depende de un conjunto de variables observables para el economista que denotamos por el vector x_i y un shock idiosincrático que denotamos ε_i , donde i se refiere a cada individuo del corte transversal. Sabemos que este shock distribuye Normal con media 0 y varianza 1. Además, usted sabe que el individuo trabaja si $y_i > 0$, donde $y_i = x_i' \beta + \varepsilon_i$.

2. **(10 puntos)** Un amigo suyo estima simplemente una regresión de y_i contra el vector x_i por mínimos cuadrados ordinarios. Discuta las ventajas y desventajas de este estimador de (que, por si no sabe, se llama modelo de probabilidad lineal) en términos de consistencia, eficiencia y capacidad predictiva.
3. **(10 puntos)** Otro amigo argumenta que una mejor manera de obtener el estimador es usando el estimador de máxima verosimilitud. Nuevamente el interés se relaciona tanto con obtener estimadores consistentes y eficientes del vector β , como con generar buenas predicciones ¿Cómo estimaría usando máxima verosimilitud? ¿Por qué? Presente las funciones que le permitirían obtener estos estimadores y discuta sus propiedades.
4. **(5 puntos)** Un conspicuo economista le hace notar que ambos estimadores (o sea los derivados de las preguntas 2 y 3 tienen similares implicancias en términos del efecto de las variables del vector x_i sobre la probabilidad de trabajar, pero que los estimadores derivados de la pregunta 3 tienen errores estándares más bajos. ¿Cómo puede explicar esta situación? ¿Es razonable este resultado?