**Pontificia Universidad Católica de Chile**
**Instituto de Economía**

Examen de Grado
Econometría y Métodos Cuantitativos
Agosto, 2011

| | | |
|---|---|---|
| Duración | : | 180 minutos |
| Fecha | : | 04 de Agosto de 2011 |
| Hora de comienzo | : | 15 horas |

El examen consta de cuatro preguntas, cada una de ellas de 40 puntos. Todas las partes deben ser respondidas. Usted dispone de 20 minutos para leer el examen y 160 minutos para responderlo, por lo que cada punto corresponde aproximadamente a un minuto.

**I.1 (22 puntos)** Se dispone de una encuesta para el 2000 con la siguiente información: Ingresos por hora (SALARIO), escolaridad en años ( E ), experiencia en años (EXP), sexo (SEXO) y grupos étnicos (pueblos originarios (OR) y resto de la población (PO)). En la encuesta se dispone de antecedentes para 273 personas de pueblos originarios y 2135 personas del resto) Se definen las variables SEXO como 1 para hombres y 0 para mujeres y OR como 1 para personas de pueblos originarios y 0 para los otros. Además se define LSALARIO como el logaritmo natural de SALARIOS. Finalmente, se han construido las siguientes variables de interacción E_OR = E*OR , EX_OR = EXP*OR y S_OR= SEXO*OR.

Se han realizado las siguientes regresiones con LSALARIO como variable dependiente

(1) Variables independientes: E, EXP y SEXO para toda la muestra
(2) Variables independientes: E, EXP, SEXO y OR toda la muestra
(3) Variables independientes: E, EXP y SEXO para la muestra de resto de la población
(4) Variables independientes: E, EXP y SEXO para la muestra de Pueblos originarios solamente
(5) Variables independientes: E, EXP, SEXO, OR, E_OR, EX_OR y S_OR para toda la muestra

Los resultados se presentan en el cuadro siguiente con SCR suma de cuadrados residuales y los errores estándar entre paréntesis.

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| | Muestra total | Muestra total | Sólo Resto | Sólo OR | Muestra total |
| **E** | 0,124 (0,004) | 1,121 (0,004) | 0,122 (0,001) | $b_E$ | 0,122 (0,004) |
| **EXP** | 0,003 (0,002) | 0,032 (0,002) | 0,033 (0,003) | $b_{EXP}$ | 0,033 (0,003) |
| **SEXO** | 0,278 (0,020) | 0,277 (0,020) | 0,306 (0,021) | $b_{SEXO}$ | 0,306 (0,021) |
| **OR** | - | -0,144 (0,032) | - | - | 0,205 (0,225) |
| **E_OR** | - | - | - | - | -0,009 (0,016) |
| **EX_OR** | - | - | - | - | -0,006 (0,007) |
| **S_OR** | - | - | - | - | -0,280 (0,065) |
| **CONSTANT** | 0,390 (0,075) | 0,459 (0,076) | 0,411 (0,084) | Cont | 0,411 (0,082) |
| $R^2$ | 0,335 | 0,341 | 0,332 | 0,321 | 0,347 |
| SCR | 610,0 | 605,1 | 555,7 | $SCR_4$ | 600,0 |
| N | 2.408 | 2.408 | 2.135 | 273 | 2.408 |

a) Interprete el coeficiente de OR en la segunda ecuación (3 puntos)

b) Suponga que los resultados de la ecuación 4 se han perdido y Ud. desea recuperarlos. Determine si es posible calcularlos y en caso afirmativo entregue los resultados para $b_E$, $b_{EXP}$, $b_{SEXO}$, Cont y $SCR_4$ (nota debe encontrar sólo el valor de los coeficientes, no de los errores estándar). (8 puntos)

c) Interprete los coeficientes de OR y de S_OR en la ecuación (5) (6 puntos)

d) Realice un test conjunto para verificar que los parámetros de las variables OR E_OR, EX_OR y S_OR de la ecuación (5) son todos nulos. (6 puntos)

**I.2** (18 puntos) Dado el modelo $y_t = \beta x_t + \mu_t$ $t = 1, \dots, T$

donde $V(\mu_t) = \sigma^2 t^\alpha$ ; $Cov(\mu_t, \mu_{t'}) = 0$ para todo t distinto de t' y $\alpha$ es un parámetro conocido.

Suponga que $\alpha = 0$

a) Si se estima $\beta$ por MICO ¿Que propiedades tiene dicho estimador? (3 puntos)

Suponga que $\alpha > 0$

b) Si se estima $\beta$ por MICO ¿Que propiedades tiene dicho estimador? (3 puntos)

c) Defina el estimador de Mínimos Cuadrados Generalizados y dé una expresión para $b_{MCG}$ en función de los datos observados. (3 puntos)

d) Dé una expresión para la varianza del estimador por Mínimos Cuadrados Generalizados. (5 puntos)

e) ¿Cómo cambiarían sus respuestas c) y d) si el parámetro $\alpha$ fuera desconocido? Explique brevemente. (3 puntos)

**II** (40 puntos) Uno de los temas más estudiados en la Economía del Crimen es el impacto del número de policías en la tasa de criminalidad. En esta pregunta estudiaremos la econometría necesaria para estimar el efecto causal de la disposición de policías sobre tasas de criminalidad.

Suponga que usted dispone de información de tasas de criminalidad, cantidad de policías y una serie de variables que pueden servir para controlar por variables observables. La ecuación que a usted le interesa estimar es de la forma:

$$y_i = \alpha + \beta Policias_i + x_i{}'\varpi + \varepsilon_i \qquad (1)$$

donde *y* es la tasa de criminalidad en el barrio *i*, *Policías* es el número de policías disponibles y *x* es un vector de controles.

1. Mencione los principales problemas econométricos que pueden surgir al estudiar el efecto causal de la disposición de policías en las tasas de criminalidad. Sea lo más formal (en el sentido econométrico y en el sentido económico) que le sea posible. (10 puntos)

2. Ahora consideremos la aproximación del trabajo "Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack" a este problema.[1] Los autores explotan el hecho de que los atentados a la AMIA en Buenos Aires en Julio de 1994 aumentaron la dotación policial que resguardaba las cuadras en que se ubicaban organizaciones judías buscando evitar atentados posibles.

   En concreto los autores corren regresiones de la forma:

   $$y_{it} = \pi + \chi ZonaJudía_i + \phi POST + \psi POST_t * ZonaJudía_i + x_{it}{}'\gamma + u_{it} \qquad (2)$$

   donde $ZonaJudía_i$ es una dummy que toma el valor 1 si en la cuadra *i* hay una institución judía y *POST* es una dummy que toma el valor 1 si el período es posterior al atentado a la AMIA.

   Los autores encuentran que la tasa de criminalidad en delitos menores (asaltos, robos, hurtos) bajó más en las cuadras con instituciones judías después de los atentados que en las cuadras ubicadas a más de un cuadra de las instituciones judías.

   a. Interprete la regresión presentada en la ecuación (2) e indique qué coeficiente permite testear formalmente la hipótesis de que existió un comportamiento diferente en estas zonas luego de los atentados a la AMIA. ¿Este modelo puede resolver los problemas que usted mencionó en la parte 1? ¿Por qué? (10 puntos)

   b. Los autores además presentan en su paper regresiones de la forma:

   $$Policias_{it} = a + bZonaJudía_i + cPOST + dPOST_t * ZonaJudía_i + x_{it}{}'k + v_{it} \qquad (3)$$

[1] R. Di Tella y E. Ernesto Schargrodsky, *American Economic Review*, March 2004.

Explique ahora cómo los resultados de estimar la ecuación (3) pueden ser utilizados para recuperar el parámetro del efecto del número de policías en la criminalidad en una ecuación de la forma:

$$y_{it} = \alpha + \vartheta ZonaJudía_i + oPOST + \beta Policias_{it} + x_{it}'\varpi + \varepsilon_{it} \qquad (1')$$

Sea muy explícito en mencionar bajo qué condiciones puede identificar efectos causales del número de policías en delincuencia. (20 puntos)

III (40 puntos) Considere el siguiente modelo para estimar el efecto sobre el peso de los recién nacidos, de las siguientes variables:

$$\log(bwght) = \beta 0 + \beta 1male + \beta 2parity + \beta 3\log(faminic) + \beta 4packs + u$$

Donde **male** es una variable binaria (dummy) igual a 1 si el recién nacido es hombre y 0 si es mujer, **parity** es el orden de nacimiento del niño, **faminc** es el ingreso familiar y **packs** es el número de cajetillas fumadas al día por la madre durante el embarazo.

En base a los resultados presentados al final en las Tablas 1, 2 y 3, responda las siguientes preguntas:

a)La Tabla 1 muestra la estimación por MCO del modelo anterior. Analice y discuta los resultados, tanto su significancia estadística como su interpretación. (5 puntos)

b)¿Qué pasaría si **packs** está correlacionado con u? Demuestre matemáticamente qué sucede con las propiedades de convergencia del estimador MCO (de un modelo de regresión lineal general) cuando las variables explicativas están correlacionadas con el término de error. (10 puntos)

c) Suponga que tiene datos sobre el precio promedio de los cigarrillos (**cigprice**) en el estado (o ciudad) donde vive cada madre. Discuta si esta información satisface las propiedades estadísticas para ser una variable instrumental de **packs**. Escriba la primera etapa. (10 puntos)

d)Observe las Tablas 1 y 2, las cuales presentan el modelo estimado por MCO y MC2E respectivamente, discuta sobre las diferencias entre ambas estimaciones. Refiérase en particular al signo y error estándar del estimador del parámetro que acompaña a **packs**. (5 puntos)

e)La Tabla 3, presenta la estimación de la primera etapa de packs, ¿qué podría concluir sobre la identificación de β4, usando **cigprice** como un instrumento para packs? ¿Satisface dicha variable instrumental los supuestos requeridos por el método de MC2E? Refiérase a la fortaleza/debilidad del instrumento.(10 puntos)

Tabla 1: Estimación MCO

| Source | SS | df | MS |
|---|---|---|---|
| Model | 1.76664363 | 4 | .441660908 |
| Residual | 48.65369 | 1383 | .035179819 |
| Total | 50.4203336 | 1387 | .036352079 |

| | |
|---|---|
| Number of obs | = 1388 |
| F( 4, 1383) | = 12.55 |
| Prob > F | = 0.0000 |
| R-squared | = 0.0350 |
| Adj R-squared | = 0.0322 |
| Root MSE | = .18756 |

| lbwght | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| male | .0262407 | .0100894 | 2.60 | 0.009 | .0064486 | .0460328 |
| parity | .0147292 | .0056646 | 2.60 | 0.009 | .0036171 | .0258414 |
| lfaminc | .0180498 | .0055837 | 3.23 | 0.001 | .0070964 | .0290032 |
| packs | -.0837281 | .0171209 | -4.89 | 0.000 | -.1173139 | -.0501423 |
| _cons | 4.675618 | .0218813 | 213.68 | 0.000 | 4.632694 | 4.718542 |

Tabla 2: Estimación MC2E

| Source | SS | df | MS |
|---|---|---|---|
| Model | -91.350027 | 4 | -22.8375067 |
| Residual | 141.770361 | 1383 | .102509299 |
| Total | 50.4203336 | 1387 | .036352079 |

| | |
|---|---|
| Number of obs | = 1388 |
| F( 4, 1383) | = 2.39 |
| Prob > F | = 0.0490 |
| R-squared | = . |
| Adj R-squared | = . |
| Root MSE | = .32017 |

| lbwght | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| packs | .7971063 | 1.086275 | 0.73 | 0.463 | -1.333819 | 2.928031 |
| parity | -.0012391 | .0219322 | -0.06 | 0.955 | -.044263 | .0417848 |
| lfaminc | .063646 | .0570128 | 1.12 | 0.264 | -.0481949 | .1754869 |
| male | .0298205 | .017779 | 1.68 | 0.094 | -.0050562 | .0646972 |
| _cons | 4.467861 | .2588289 | 17.26 | 0.000 | 3.960122 | 4.975601 |

Tabla 3: Primera Etapa

| Source | SS | df | MS |
|---|---|---|---|
| Model | 3.76705108 | 4 | .94176277 |
| Residual | 119.929078 | 1383 | .086716615 |
| Total | 123.696129 | 1387 | .089182501 |

| | |
|---|---|
| Number of obs | = 1388 |
| F( 4, 1383) | = 10.86 |
| Prob > F | = 0.0000 |
| R-squared | = 0.0305 |
| Adj R-squared | = 0.0276 |
| Root MSE | = .29448 |

| packs | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| male | -.0047261 | .0158539 | -0.30 | 0.766 | -.0358264 | .0263742 |
| parity | .0181491 | .0088802 | 2.04 | 0.041 | .0007291 | .0355692 |
| lfaminc | -.0526374 | .0086991 | -6.05 | 0.000 | -.0697023 | -.0355724 |
| cigprice | .000777 | .0007763 | 1.00 | 0.317 | -.0007459 | .0022999 |
| _cons | .1374075 | .1040005 | 1.32 | 0.187 | -.0666084 | .3414234 |

IV (40 puntos) Se plantea el siguiente modelo para el precio del cobre:

$\ln(P_t) = \beta_1 + \beta_2 \ln(P_{t-1}) + \beta_3 \ln(P_{t-2}) + \beta_4 \ln(S_t) + u_t \qquad u \sim N(0, \sigma^2 I)$

La estimación por MICO arrojó los siguientes resultados:

$\ln(P_t) = 2 + 0.6 \ln(P_{t-1}) + 0.3 \ln(P_{t-2}) - 0.5 \ln(S_t) + e_t$ (ecuación 1)

donde:

$P_t$ = precio del cobre en el período t

St = stock de cobre en el período t

$e_t$ = residuo de la regresión

a) Bajo el supuesto de que los coeficientes estimados en la ecuación (1) son iguales a los parámetros, demuestre que existe una relación de largo plazo entre precio del cobre y los inventarios del metal. (8 puntos)

b) A la luz de los resultados, ¿cuál sería la relación de largo plazo? (3 puntos)

c) Si en un determinado período, dado el nivel de stocks de cobre vigente, el precio del cobre fuera mayor que el precio de equilibrio el largo plazo, ¿cómo se acercaría el precio de mercado al precio de equilibrio? ¿con ciclos? (3 puntos)

d) Escriba la ecuación (1) en el formato de modelo de corrección de errores y en el formato de cambios y niveles. (8 puntos)

e) Describa un test estadístico para probar o rechazar la existencia de una relación de largo plazo entre $\ln(P_t)$ y $\ln(S_t)$. (8 puntos)

f) Suponga que los coeficientes obtenidos en la regresión (1) son iguales a los parámetros que generaron los datos y el proceso que regula la dinámica de los inventarios es (4 puntos):

   $\ln(S_t) = a + v_t$ (ecuación 2)
   donde,
   $v_t = \rho\, v_{t-1} + \varepsilon_t \qquad \rho < 1$
   $\varepsilon_t \sim N(0, \sigma^2 I)$

   ¿Existiría cointegración entre $\ln(P_t)$ y $\ln(S_t)$?

g) Si, en cambio, los inventarios de cobre siguieran el proceso (4 puntos):
$\ln(S_t) = a + \ln(S_{t-1}) + \varepsilon_t$
$\varepsilon_t \sim N(0, \sigma^2 I)$

¿Habría cointegración entre $\ln(P_t)$ y $\ln(S_t)$?

h) En este último caso, ¿cuál sería el orden de integración de $\ln(P_t)$? (2 puntos)