Pontificia Universidad Católica de Chile
Instituto de Economía

Examen de Grado
Econometría y Métodos Cuantitativos
Marzo, 2012

Duración	:	180 minutos
Fecha	:	2 de Marzo de 2012
Hora de comienzo	:	9:30 horas

Usted dispone de 30 minutos para leer el examen y 150 minutos para responderlo. Cada punto corresponde aproximadamente a un minuto. Debe responder todas las preguntas del examen.

Parte I (45 puntos)

Pregunta I.1 (25 puntos):

Usted está interesado en estimar la elasticidad de la demanda por cigarrillos con el fin de proponer una política pública para reducir en un 10% la demanda. Para eso cuenta con datos de consumo promedio de cigarrillos (Q_i) y de precio promedio de los cigarrillos (P_i) para 48 estados de USA en 1995. Las tablas que se anexan contienen información acerca de distintas regresiones asociadas a estos datos.

a) (8 puntos) Explique por qué la siguiente regresión puede sufrir problemas de endongeneidad:

$$Ln[Q_i] = \beta_0 + \beta_0 Ln[P_i] + u_i \tag{1}$$

En particular, mencione dos ejemplos concretos de variables en u_i que pueden crear endogeneidad. Discuta cuál es el efecto de tener endogeneidad en la regresión.

b) (8 puntos) Suponga que además usted cuenta con datos de ingreso promedio (Y_i) y de las tasas de impuesto a las ventas ($SalesTax_i$) para los 48 estados en 1995. Discuta si estas variables son buenos instrumentos para $Ln[P_i]$.

c) (5 puntos) Suponga que la tasa de impuesto a las ventas y el ingreso promedio están correlacionados. Seleccione un modelo de los que se adjuntan en las tablas que esté bien especificado.

d) (4 puntos) Utilizando el modelo seleccionado anteriormente, determine en qué porcentaje debe aumentar el precio de los cigarrillos para reducir su demanda en 10%.

```
. regress LogQ LogP if year==1995

Linear regression                                      Number of obs =      48
                                                       F(  1,    46) =   38.86
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4058
                                                       Root MSE      =  .18962

------------------------------------------------------------------------------
             |
        LogQ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        LogP |  -1.213057   .1945896    -6.23   0.000    -1.604746   -.8213685
       _cons |   10.33892   .9348201    11.06   0.000     8.457229    12.22062
------------------------------------------------------------------------------

. regress LogP SalesTax LogY if year==1995

Linear regression                                      Number of obs =      48
                                                       F(  2,    45) =   48.98
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.6389
                                                       Root MSE      =  .07848

------------------------------------------------------------------------------
             |
        LogP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    SalesTax |   .0273948   .0040961     6.69   0.000     .0191449    .0356447
        LogY |   .3892824   .0653914     5.95   0.000     .2575774    .5209874
       _cons |   3.590811   .1727273    20.79   0.000      3.24292    3.938702
------------------------------------------------------------------------------
```

```
. regress LogY SalesTax LogP if year==1995

Linear regression                                      Number of obs =      48
                                                       F(  2,    45) =   13.43
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3392
                                                       Root MSE      =  .11357

------------------------------------------------------------------------------
             |
        LogY |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    SalesTax |  -.0164893   .0071945    -2.29   0.027    -.0309796   -.0019989
        LogP |   .8153188   .1653164     4.93   0.000     .4823544    1.148283
       _cons |  -1.129018   .7630289    -1.48   0.146    -2.665837    .4078011
------------------------------------------------------------------------------

. ivreg LogQ (LogP = LogY SalesTax) SalesTax if year==1995

Instrumental variables (2SLS) regression               Number of obs =      48
                                                       F(  2,    45) =    7.62
                                                       Prob > F      =  0.0014
                                                       R-squared     =  0.3308
                                                       Root MSE      =  .20344

------------------------------------------------------------------------------
             |
        LogQ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        LogP |  -.5923217   .5991155    -0.99   0.328    -1.799002    .6143589
    SalesTax |   -.015096   .0229064    -0.66   0.513    -.0612319    .0310399
       _cons |   7.451929   2.765619     2.69   0.010     1.881687    13.02217
------------------------------------------------------------------------------
Instrumented:  LogP
Instruments:   SalesTax LogY
------------------------------------------------------------------------------

. ivreg LogQ (LogP = LogY SalesTax) LogY if year==1995

Instrumental variables (2SLS) regression               Number of obs =      48
                                                       F(  2,    45) =    8.19
                                                       Prob > F      =  0.0009
                                                       R-squared     =  0.4189
                                                       Root MSE      =  .18957

------------------------------------------------------------------------------
             |
        LogQ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        LogP |  -1.043375   .3723028    -3.07   0.004    -1.893231    -.393519
        LogY |   .2245154   .1117468     2.01   0.047      .005492     .443539
       _cons |   9.430658   1.259393     7.49   0.000     6.894111    11.96721
------------------------------------------------------------------------------
Instrumented:  LogP
Instruments:   LogY SalesTax
------------------------------------------------------------------------------
```

Pregunta I.2 (20 puntos):

Ud. no sabe si los datos de una variable Y_t han sido generados por uno de los dos siguientes modelos:

Modelo 1:

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

$$E(u \mid X) = 0, E(uu' \mid X) = \sigma^2 I$$

Modelo 2:

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

$$u_t = \rho\, u_{t-4} + v_t$$

$$E(v \mid X) = 0, E(vv' \mid X) = \sigma^2 I$$

Explique en detalle como implementaría un test del Multiplicador de Lagrange para testear la hipótesis nula de que el primer modelo es el correcto, es decir, que los errores u_t son bien comportados.

Parte II (42 puntos)

Pregunta II.1 (22 puntos):

Con el objeto de estudiar el poder predictivo de la prueba Simce se elige una muestra de 104 trabajadores y se construye el siguiente modelo:

$$[M1] \quad Salario = \underset{(1{,}78)}{1{,}72} + \underset{(0{,}005)}{0{,}050}\, E + \underset{(0{,}001)}{0{,}006}\, P \qquad \text{con} \quad \text{SCR} = 2045$$

En que E es años de escolaridad y P el promedio de las pruebas Simce de matemáticas y lenguaje de segundo medio normalizado sobre 100. Los números entre paréntesis son errores estándares y SCR es la suma de cuadrados residuales.

Se especula que las pruebas verbal y de matemáticas podrían tener un impacto diferente y se estima un segundo modelo:

$$[M2] \quad Salario = \underset{(1{,}8)}{2{,}02} + \underset{(0{,}007)}{0{,}063}\, E + \underset{(0{,}0011)}{0{,}0044}\, M + \underset{(0{,}001)}{0{,}0026}\, L \qquad \text{con} \quad \text{SCR}=2000$$

En este caso M indica el resultado del alumno en la prueba de matemáticas y L en la prueba de lenguaje.

Finalmente se estima un modelo que combina el promedio con el valor de una de las pruebas: la de lenguaje en este caso. Los resultados son:

$$[M3] \quad Salario = \underset{(1{,}80)}{2{,}02} + \underset{(0{,}007)}{0{,}063}\, E + \underset{(0{,}0023)}{0{,}0088}\, P - \underset{(0{,}00121)}{0{,}0018}\, L \qquad \text{con} \quad \text{SCR} = 2000$$

a) (7 puntos) Muestre que el modelo 1 (M1) es una versión restringida del modelo 2 (M2) y verifique si la restricción es significativa.
b) (3 puntos) ¿Que beneficios tiene el modelo 1 en relación con el modelo 2?
c) (7 puntos) Muestre que el modelo 3 es una nueva representación del modelo 2 y verifique en este nuevo modelo la restricción del modelo 1.
d) (5 puntos) ¿Puede usar la información de los modelos anteriores para verificar la hipótesis que la prueba de matemática tiene un efecto mayor en salarios que la prueba de lectura? En caso contrario indique un test para hacerlo.

Pregunta II.2 (8 puntos):

Un investigador que estima los determinantes del peso al nacer (PN) ha estimado el siguiente modelo:

$$\log(PN) = \underset{(1{,}22)}{4{,}66} - \underset{(0{,}0009)}{0{,}0044}cig + \underset{(0{,}0059)}{0{,}0013}\log(Y) + \underset{(0{,}006)}{0{,}016}\,paridad + \underset{(0{,}01)}{0{,}027}H + \underset{(0{,}013)}{0{,}055}blanco$$

Con n= 1388 y $R^2 = 0{,}0472$

En que *cig* es el número promedio de cigarrillos fumados a la semana durante el embarazo, Y es el ingreso familiar, *paridad* es el número del hijo de esta mujer, *H* es el sexo del recién nacido (Hombre=1) y *blanco* su raza (Blanco=1), *n* número de observaciones.

Al presentar este trabajo en un seminario se le comenta que sería interesante incluir la educación del padre y la de la madre. Para analizar este comentario, el investigador estima un nuevo modelo que agrega al anterior la educación del padre y de la madre. Sin embargo, como no dispone de dicha información para todas las observaciones lo estima con sólo 1191 observaciones.

$$\log(PN) = \underset{(0{,}38)}{4{,}66} - \underset{(0.0010)}{0{,}0052}cig + \underset{(0{,}0085)}{0{,}0010}\log(Y) + \underset{(0{,}006)}{0{,}017}\,paridad + \underset{(0{,}011)}{0{,}034}H + \underset{(0{,}015)}{0{,}045}blanco - \underset{(0{,}0030)}{0{,}0030}EM + \underset{(0{,}0026)}{0{,}0032}EP$$

con n=1,191 R2=0.0493

En que *EM* y *EP* es el número de años de educación del padre y de la madre respectivamente.

a) (3 puntos) Interprete el coeficiente que acompaña la variable *cig* en la primera regresión.
b) (5 puntos) Explique cómo evaluaría la hipótesis conjunta de que la educación del padre y la de la madre no afectan el peso al nacer.

Pregunta II.3 (12 puntos):

En un estudio de precios hedónicos para un condominio muy elegante se ha corrido el siguiente modelo:

$$PE = \beta_0 + \beta_1\, sup + \beta_2\, dormit + \beta_3\, dist + \beta_4\, patio + \mu$$

En que *PE* es el valor del arriendo mensual, *sup* es la superficie de la vivienda en m^2, *dormit* es el número de dormitorios, *dist* es la distancia al centro en cuadras y *patio* es una variable muda que toma el valor 1 si la vivienda tiene un patio de más de 350 m2 y es cero en otro caso.

Para facilitar la clasificación de las viviendas se ha construido un índice (Índice de calidad) que es igual al cuadrado del impacto de la distancia dividido por la suma del impacto de las variables superficie y patio. El valor óptimo de este índice es 2 que se refiere a las viviendas de la mejor calidad.

Explique cómo se puede verificar estadísticamente la hipótesis que en el condominio las viviendas son de la mejor calidad. Use un nivel de significancia de 5%. Explique claramente la hipótesis y el procedimiento para testearla.

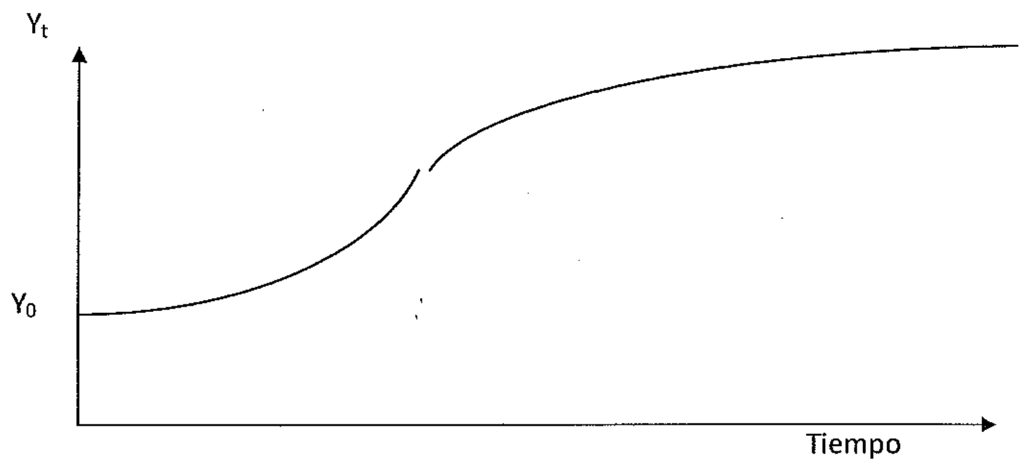
Parte III (43 puntos)

Se tiene un modelo con datos trimestrales de la tasa de inversión respecto al stock de capital (Y_t) la que depende positivamente de diferenciales entre la tasa de rentabilidad y la tasa de interés real (X_t). Así:

$$Y_t = \mu + \delta_0 X_t + \delta_1 X_{t-1} \ldots\ldots\ldots\ldots\ldots\ldots + \delta_h X_{t-h} + v_t \qquad \text{(ecuación 1)}$$

$$E(v \mid X) = 0, E(vv' \mid X) = \sigma^2 I$$

El efecto de X_t opera con rezagos de tal modo que cuando X_t sube en forma discreta a un nuevo nivel, Y_t reacciona en forma creciente durante los primeros trimestres y a partir de cierto período el efecto positivo sobre el crecimiento de Y_t empieza a declinar. Gráficamente:



a) (5 puntos) Describa claramente cómo deben comportarse los parámetros δ_i.
b) (5 puntos) ¿Qué restricción deben cumplir los coeficientes del modelo para que no haya una conducta explosiva de la inversión?
c) (5 puntos) ¿Cuál es el efecto de largo plazo de un aumento del diferencial entre tasa de rentabilidad y tasa de interés?
d) (10 puntos) Explique cómo estimaría el modelo si Ud. tiene 40 datos y 35 rezagos para evitar una pérdida excesiva de grados de libertad. (*Pista: asuma que* $\delta_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2$)
e) (6 puntos) Suponga ahora que Ud. decide remplazar los rezagos de X_t por un rezago de la variable dependiente, proponiendo estimar:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 Y_{t-1} + u_t \qquad \text{(ecuación 2)}$$

$$u_t = \rho u_{t-1} + v_t$$

$$E(v \mid X) = 0, E(vv' \mid X) = \sigma^2 I$$

¿Es una buena opción de especificación para reflejar el comportamiento dinámico de la tasa de inversión descrita en la ecuación 1? ¿por qué?

f) (5 puntos) Suponga ahora que el modelo que genera los datos es uno tal que la ecuación 2 está bien especificada, ¿qué propiedad en cuanto sesgo y consistencia tendrían los estimadores de mínimos cuadrados ordinarios? ¿por qué?

g) (7 puntos) Proponga un estimador consistente distinto al de máxima verosimilitud y explique cómo lo implementaría.