



Examen de Grado
Econometría y Métodos Cuantitativos
Agosto, 2016

Duración : 120 minutos
Fecha : 5 de agosto de 2016
Hora de comienzo : 8:30 horas

INSTRUCCIONES

Una vez leído en voz alta por el profesor en la sala, usted dispone de dos horas para responder este examen. No empiece a responder hasta que se le indique hacerlo.

Responda **solamente 2** preguntas de las 3 preguntas propuestas. Si responde más de 2 preguntas solamente se considerarán las 2 peores respuestas.

Pregunta 1

Suponga que Ud. está interesada en determinar el efecto de la participación en un curso de capacitación laboral en el salario. Para ello Ud. plantea el siguiente modelo de regresión lineal $y_i = \beta_0 + x_1\beta_1 + x_2\beta_2 + u_i$ donde y_i es el logaritmo del salario por hora, x_1 representa los años de escolaridad del individuo y x_2 toma el valor 1 si la persona participó en un programa de capacitación y 0 si no lo hizo. Además, β_0 , β_1 y β_2 son los parámetros de la regresión y u_i es un componente aleatorio "no observable" para cada individuo.

1. Explique formalmente cómo podría determinar el efecto de participar en un curso de capacitación en el salario para un nivel de educación dado, interpretando económicamente dicho efecto. Indique los supuestos necesarios para ello. Recuerde que x_2 es binaria. **(6 puntos)**
2. Una vecina le indica que su regresión sufre de omisión de variables relevantes por cuanto se ha excluido el nivel de habilidad de las personas. Explique cómo afecta las propiedades de los estimadores MCO la omisión de variables relevantes, en general, y en particular explique si tiene razón su vecina en este ejercicio. Ayuda: puede trabajar matricialmente. **(6 puntos)**
3. Un compañero de Yoga le indica que la única forma de solucionar el problema mencionado por su vecina y estimar consistentemente el efecto de participar en un curso de capacitación en el salario es mediante la realización de un "experimento aleatorio controlado" (EAC). En este, los cursos de participación se asignan aleatoriamente entre los trabajadores. Explique formalmente cómo un EAC puede solucionar el problema planteado en la parte 2. **(6 puntos)**
4. Su prima le indica que existe otra posibilidad de estimar consistentemente el efecto de la capacitación en el salario. Ella le explica que si Ud. tiene una variable z_2 correlacionada con x_2 y no correlacionada con u_i puede estimar consistentemente β_2 por "Mínimos Cuadrados en Dos Etapas" (MC2E). Suponiendo que x_1 es exógena, explique si su prima tiene razón y como implementaría MC2E. **(6 puntos)**
5. Una geógrafa le sugiere usar la distancia del trabajo al centro de capacitación más cercano como instrumento de la participación en un curso de capacitación. Discuta si ésta podría ser una variable instrumental válida. **(6 puntos)**
6. Complete las siguientes tablas. La Tabla 1 presenta una estimación por MCO y la Tabla 2 por MC2E en la que se usa distancia del trabajo al centro de capacitación más cercano como variable instrumental para x_2 . Asumiendo que el instrumento z_2 satisface los supuestos de MC2E, compare el efecto de participar en un curso de capacitación obtenido por MCO y MC2E y explique la dirección del sesgo del

estimador MCO (asumiendo que la muestra es suficientemente grande). (4 puntos)

7. Si los resultados de la Tabla 2 fueran ciertos, los años de escolaridad promedio iguales a 11 años y el porcentaje de capacitados un 40 %, calcule cuánto es el máximo que un individuo estaría dispuesto a pagar por el programa de capacitación. Trabaje en flujos mensuales. Para ello suponga que el trabajador tiene una jornada de 45 horas a la semana, un mes tiene 4 semanas y la tasa de descuento mensual es 1%. (6 puntos)

Cuadro 1: Estimación MCO

. reg y x_1 x_2

Source	SS	df	MS	Number of obs	=	1,000
Model	323.54111	2	161.770555	F(2, 997)	=	
Residual	314.293797	997	.315239516	Prob > F	=	0.0000
				R-squared	=	0.5072
				Adj R-squared	=	0.5063
Total	637.834907	999	.63847338	Root MSE	=	.56146

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x_1	.094576	.0031121	30.39	0.000	.088469 .1006829
x_2		.036244	10.44	0.000	.3073251 .4495717
_cons	6.769047	.0406682	166.45	0.000	6.689242 6.848852

Cuadro 2: Estimación MC2E

. ivregress 2sls y x_1 (x_2=z_2)

Instrumental variables (2SLS) regression

Number of obs	=	1,000
Wald chi2(2)	=	867.96
Prob > chi2	=	0.0000
R-squared	=	0.4756
Root MSE	=	.57836

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
x_2	.0882489	.0473181		0.062	-.0044928 .1809907
x_1	.0943263	.0032058	29.42		.088043 .1006097
_cons	6.94551	.0454692	152.75	0.000	6.856392 7.034628

Instrumented: x_2
 Instruments: x_1 z_2

Pregunta 2

Considere el siguiente modelo simple de series de tiempo:

$$y_t = \beta_0 + \beta_1 x_t^* + u_t,$$

donde la variable explicativa x_t^* no es observable pero en su lugar se cuenta con una medida x_t tal que $x_t = x_t^* + e_t$, de modo que e_t un el error de medida. Suponga que u_t , x_t^* y e_t tienen media cero, y que no correlacionan entre sí.

- (6 puntos)** ¿Es consistente el estimador de MCO que utiliza a x_t en lugar de x_t^* ? Demuestre formalmente, y refiérase a las implicancias de su resultado.
- (6 puntos)** Suponga ahora que la variable y_t tampoco es observable pero se cuenta con una medida y_t^* tal que $y_t = y_t^* + \theta_t$, de modo que θ_t un el error de medida. ¿Es consistente el estimador de MCO que utiliza a y_t^* en lugar de y_t ? Demuestre formalmente, y refiérase a las implicancias de su resultado.
- (10 puntos)** Sabemos que el procedimiento de variables instrumentales puede resolver los problemas de errores en las variables. ¿Qué condiciones debe cumplir x_{t-1} para ser un instrumento válido de x_t ? Considere que el R^2 de la regresión de x_t en x_{t-1} con $T = 27$ es igual a 0.12. ¿Es x_{t-1} un instrumento relevante? Haga su análisis usando un nivel de significancia del 5%. ¿Cambia en algo su respuesta si T fuera igual a 42? Explique.
- (6 puntos)** ¿Es el estimador de variables instrumentales eficiente en muestra pequeña y/o grande? Explique cuidadosamente su respuesta.
- (5 puntos)** Considere ahora la siguiente estimación por MC2E:

$$\hat{y}_t = -0.39 + 2.24 x_t^* + 0.09 \omega_t$$

Covariance matrix of coefficients of ivregress model

e (V)	x_t	w_t	_cons
x_t	.05245751		
w_t	.00680763	.00565345	
_cons	-.14621701	-.01175799	.45677479

donde se supone que ω_t es una variable exógena. Evalúe la hipótesis nula que $\beta_1 = 5\beta_2$ a un nivel de significancia del 1%.

- (7 puntos)** En la práctica suele utilizarse como instrumento no sólo a x_{t-1} sino que también se incluye x_{t-2} , x_{t-3} y x_{t-4} en la regresión. Explique en detalle cómo se podría implementar un contraste (test) de sobre-identificación en dicho contexto, e indique cómo se interpreta el resultado.

Pregunta 3

La municipalidad de Valdivia lo contrata a usted para determinar si la planta de celulosa que se construyó en la orilla del río Cruces ha afectado los precios de las viviendas del balneario situado sobre el río. Los datos contienen precios (en dólares) de venta de 321 casas vendidas durante el año 2015 en esta ciudad, así como algunas características de las mismas (número de dormitorios, número de baños, metros cuadrados de terreno) y una variable dummy que toma valor 1 si hay piscina.

En la municipalidad han estimado la siguiente regresión:

$$\log(\text{Precio}_i) = \beta_1 + \beta_2 \text{dist}_i + \mu_i$$

donde *dist* es la distancia de cada casa a la planta en metros.

1. Realice un análisis crítico de los resultados presentados más abajo en la Regresión 1 **(8 puntos)**.
2. Para tratar de entender los determinantes del precio de las viviendas, usted hace la siguiente regresión:

$$\log(\text{Precio}_i) = \beta_1 + \beta_2 \log \text{sup}_i + \beta_3 \text{Dorms}_i + \beta_4 \text{Baños}_i + \beta_5 \text{Piscina}_i + \mu_i$$

Donde *sup* es la superficie del terreno en metros cuadrados, *Dorms* es el número de dormitorios, *Baños* el número de baños y *Piscina* es una variable muda que toma valor 1 si la casa tiene piscina y 0 si no la tiene. Evalúe e interprete los resultados de la regresión 2. **(6 puntos)**

Para responder las siguientes preguntas use la información de las regresiones 2 a 5:

3. (a) En base a la regresión 2, estime la hipótesis nula que el coeficiente asociado al número de dormitorios es igual al del número de baños contra la alternativa de que son distintos. Indique cuáles cambios son necesarios para testear la misma hipótesis nula, contra la alternativa $\beta_4 > \beta_3$. **(6 puntos)**

(b) Haga un test para la presencia de heterocedasticidad. **(6 puntos)**

(c) Elija entre las regresiones 2 y 5 la que a su juicio es la mejor. Justifique su elección. ¿Cambiaría con esto su respuesta a 1? ¿Cuál será el precio de una vivienda de 1000 mts cuadrados de terreno, con 2 baños y 3 dormitorios, que no tiene piscina? **(6 puntos)**

4. Finalmente, usted se siente en condiciones de estimar el efecto de la planta sobre el precio de las vivienda

$$\log(\text{Precio}_i) = \beta_1 + \beta_2 \log \text{sup}_i + \beta_3 \text{Dorms}_i + \beta_4 \text{Baños}_i + \beta_5 \text{Piscina}_i + \beta_5 \text{Dist}_i + \mu_i$$

como se reporta en la regresión 6. ¿qué problema tiene la metodología empleada por usted? **(8 puntos)**

Regresión 1.

Dependent Variable: LOG(PRECIO)

Method: Least Squares

Sample(adjusted): 1 321

Included observations: 321 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.07468	0.061893	178.9312	0.0000
DIST	1.46E-05	2.76E-06	5.298871	0.0000
R-squared	0.080898	Mean dependent var		11.37812
Adjusted R-squared	0.078017	S.D. dependent var		0.438174
S.E. of regression	0.420735	Akaike info criterion		1.112583
Sum squared resid	56.46868	Schwarz criterion		1.136081
Log likelihood	-176.5696	F-statistic		28.07803
Durbin-Watson stat	1.485745	Prob(F-statistic)		0.000000

*ln(Y)***Regresión 2.**

Dependent Variable: LOG(PRECIO)

Method: Least Squares

Date: 12/04/07 Time: 18:11

Sample(adjusted): 1 321

Included observations: 321 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.573910	0.226231	42.31923	0.0000
LOG(SUP)	0.083163	0.024344	3.416180	0.0007
DORMS	0.087845	0.017554	5.004394	0.0000
BAÑOS	0.258632	0.027549	9.388190	0.0000
PISCINA	0.083125	0.044859	1.853048	0.0648
R-squared	0.538422	Mean dependent var		11.37812
Adjusted R-squared	0.532579	S.D. dependent var		0.438174
S.E. of regression	0.299572	Akaike info criterion		0.442529
Sum squared resid	28.35888	Schwarz criterion		0.501274
Log likelihood	-66.02590	F-statistic		92.15203
Durbin-Watson stat	1.938050	Prob(F-statistic)		0.000000

→ *rechazar al 5%
aceptar al 10%.*

Regresión 3.

Dependent Variable: LOG(PRECIO)

Method: Least Squares

Date: 12/04/07 Time: 18:14

Sample(adjusted): 1 321

Included observations: 321 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.409585	0.230528	40.81751	0.0000
LOG(SUP)	0.104380	0.024668	4.231335	0.0000
BAÑOS+DORMITORIOS	0.144922	0.012899	11.23520	0.0000
PISCINA	0.090369	0.046255	1.953713	0.0516
R-squared	0.507082	Mean dependent var		11.37812
Adjusted R-squared	0.502417	S.D. dependent var		0.438174
S.E. of regression	0.309086	Akaike info criterion		0.501991
Sum squared resid	30.28440	Schwarz criterion		0.548987
Log likelihood	-76.56950	F-statistic		108.7029
Durbin-Watson stat	1.877841	Prob(F-statistic)		0.000000

Regresión 4.

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Sample: 1 321

Included observations: 321

$nD^2 \sim \chi^2_{j-1}$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.086717	0.879068	1.236214	0.2173
LOG(SUP)	-0.332912	0.184771	-1.801748	0.0726
(LOG(SUP))^2	0.023082	0.009937	2.322766	0.0208
(LOG(SUP))*DORMS	-0.014667	0.010841	-1.352917	0.1771
(LOG(SUP))*BAÑOS	-0.059533	0.019939	-2.985795	0.0031
(LOG(SUP))*PISCINA	0.080028	0.040106	1.995419	0.0469
DORMS	0.146396	0.101977	1.435590	0.1521
DORMS^2	-0.002833	0.004758	-0.595570	0.5519
DORMS*BAÑOS	0.010183	0.017270	0.589618	0.5559
DORMS*PISCINA	0.016958	0.029115	0.582440	0.5607
BAÑOS	0.560196	0.180334	3.106442	0.0021
BAÑOS^2	0.012071	0.020377	0.592377	0.5540
BAÑOS*PISCINA	-0.070578	0.047698	-1.479678	0.1400
PISCINA	-0.692798	0.361137	-1.918380	0.0560
R-squared	0.075603	Mean dependent var		0.088345
Adjusted R-squared	0.036459	S.D. dependent var		0.139833
S.E. of regression	0.137260	Akaike info criterion		-1.091239
Sum squared resid	5.784011	Schwarz criterion		-0.926753
Log likelihood	189.1439	F-statistic		1.931410
Durbin-Watson stat	1.722364	Prob(F-statistic)		0.026336

Regresion 5

Dependent Variable: LOG(PRECIO)

Method: Least Squares

Date: 12/04/07 Time: 18:09

Sample(adjusted): 1 321

Included observations: 321 after adjusting endpoints

White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.573910	0.245413	39.01138	0.0000
LOG(SUP2)	0.083163	0.027594	3.013829	0.0028
DORMS	0.087845	0.017430	5.039858	0.0000
BAÑOS	0.258632	0.027334	9.461779	0.0000
PISCINA	0.083125	0.046657	1.781616	0.0758
R-squared	0.538422	Mean dependent var		11.37812
Adjusted R-squared	0.532579	S.D. dependent var		0.438174
S.E. of regression	0.299572	Akaike info criterion		0.442529
Sum squared resid	28.35888	Schwarz criterion		0.501274
Log likelihood	-66.02590	F-statistic		92.15203
Durbin-Watson stat	1.938050	Prob(F-statistic)		0.000000

Regression 6

Dependent Variable: LOG(PRECIO)

Method: Least Squares

Sample(adjusted): 1 321

Included observations: 321 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.458467	0.257115	36.78688	0.0000
LOG(SUP2)	0.099119	0.029626	3.345709	0.0009
BANOS	0.260810	0.027649	9.432745	0.0000
DORMS	0.086904	0.017585	4.942010	0.0000
PISCINA	0.081049	0.044920	1.804305	0.0721
DIST	-2.42E-06	2.56E-06	-0.945398	0.3452
R-squared	0.539728	Mean dependent var		11.37812
Adjusted R-squared	0.532422	S.D. dependent var		0.438174
S.E. of regression	0.299622	Akaike info criterion		0.445926
Sum squared resid	28.27865	Schwarz criterion		0.516420
Log likelihood	-65.57114	F-statistic		73.87560
Durbin-Watson stat	1.938744	Prob(F-statistic)		0.000000

Tabla 3: Valores Críticos para una distribución t-Student

n-k	90 %	95 %	97.50 %	99 %	99.50 %
1	3.078	6.314	12.71	31.82	63.66
2	1.886	2.92	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.44	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
.					
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
.					
78	1.292	1.664	1.991	2.375	2.641
∞	1.281	1.644	1.959	2.326	2.576

