

1a. REVISIÓN DE ECONOMETRÍA II
17 de mayo 2012

Ejercicio 1 (25 puntos)

En Card, D. (1995), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," se propone la estimación por variables instrumentales de un modelo para estimar los retornos de la educación a partir de la siguiente ecuación:

$$\ln(w_i) = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{exper}_i + \beta_3 \text{exper}_i^2 + x_i' \gamma + u_i$$

Donde:

w: salario por hora
educ: años de educación completados
exper: experiencia potencial (edad-educ-6)
*exper*²: experiencia potencial al cuadrado sobre 100

El vector *x* incluye un conjunto de características observables tales como: Indicador de raza y dummies por lugar de residencia en el año 66 (son 9 regiones: reg661 a reg669).

Estimación 1

```
. reg lnw educ exper exper2 Afro reg661 reg662 reg663 reg664 reg665 reg666 reg667 reg668
```

Source	SS	df	MS	Number of obs = 3010		
Model	158.799165	12	13.2332637	F(12, 2997) = 91.42		
Residual	433.84248	2997	.144758919	Prob > F = 0.0000		
				R-squared = 0.2680		
				Adj R-squared = 0.2650		
Total	592.641645	3009	.196956346	Root MSE = .38047		

	lnw	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ		.0790114	.0035504	22.25	0.000	.0720499 .0859729
exper		.0875037	.0067655	12.93	0.000	.0742381 .1007692
exper2		-.2440668	.0323233	-7.55	0.000	-.3074449 -.1806887
Afro		-.1828211	.018577	-9.84	0.000	-.219246 -.1463962
reg661		-.139752	.0396339	-3.53	0.000	-.2174643 -.0620397
reg662		-.0191121	.0288544	-0.66	0.508	-.0756886 .0374643
reg663		.0158491	.0279321	0.57	0.570	-.0389189 .0706172
reg664		-.1197227	.0358249	-3.34	0.001	-.1899666 -.0494789
reg665		-.1546054	.0286622	-5.39	0.000	-.2108049 -.0984059
reg666		-.1540476	.0332663	-4.63	0.000	-.2192746 -.0888205
reg667		-.1512786	.0317948	-4.76	0.000	-.2136206 -.0889367
reg668		-.198882	.0472951	-4.21	0.000	-.2916162 -.1061479
_cons		4.798294	.0725941	66.10	0.000	4.655954 4.940633

```
. est store MCO1
```

Estimación 2

```
. reg educ cercania exper exper2 Afro reg661 reg662 reg663 reg664 reg665 reg666 reg667 reg668
```

Source	SS	df	MS	Number of obs = 3010		
Model	10197.5689	12	849.79741	F(12, 2997) = 224.10		
Residual	11364.5112	2997	3.79196235	Prob > F = 0.0000		
				R-squared = 0.4729		
				Adj R-squared = 0.4708		
Total	21562.0801	3009	7.16586243	Root MSE = 1.9473		

	educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cercania		.4511111	.0804158	5.61	0.000	.2934354 .6087868
exper		-.4122875	.033806	-12.20	0.000	-.4785728 -.3460021
exper2		.0616767	.1654383	0.37	0.709	-.2627073 .3860608
Afro		-.917234	.0937926	-9.78	0.000	-1.101138 -.7333297
reg661		-.2626612	.2028006	-1.30	0.195	-.6603036 .1349813
reg662		-.2826386	.14774	-1.91	0.056	-.5723206 .0070435
reg663		-.2426988	.1429935	-1.70	0.090	-.5230741 .0376764
reg664		-.2062228	.1839477	-1.12	0.262	-.5668994 .1544538
reg665		-.5978626	.1473585	-4.06	0.000	-.8867966 -.3089287
reg666		-.6319188	.173194	-3.65	0.000	-.9715099 -.2923277
reg667		-.4815461	.1636825	-2.94	0.003	-.8024876 -.1606046
reg668		.2755212	.2423875	1.14	0.256	-.1997416 .750784
_cons		17.11109	.2037857	83.97	0.000	16.71151 17.51066

```
. est store MCO2
```

Estimación 3

```
. reg cercania educ exper exper2 Afro reg661 reg662 reg663 reg664 reg665 reg666 reg667 reg668
```

Source	SS	df	MS	Number of obs = 3010		
Model	72.4409605	12	6.0367467	F(12, 2997) = 31.18		
Residual	580.290269	2997	.193623713	Prob > F = 0.0000		
				R-squared = 0.1110		
				Adj R-squared = 0.1074		
Total	652.731229	3009	.216926298	Root MSE = .44003		

	cercania	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ		.0230345	.0041062	5.61	0.000	.0149833 .0310856
exper		.0092538	.0078245	1.18	0.237	-.0060882 .0245958
exper2		-.0202356	.0373828	-0.54	0.588	-.0935342 .053063
Afro		.0760105	.0214848	3.54	0.000	.033884 .118137
reg661		.0208837	.0458377	0.46	0.649	-.0689929 .1107602
reg662		.0824481	.0333709	2.47	0.014	.0170158 .1478804
reg663		-.0670166	.0323043	-2.07	0.038	-.1303574 -.0036757
reg664		-.1883116	.0414325	-4.55	0.000	-.2695507 -.1070726
reg665		-.2192266	.0331486	-6.61	0.000	-.284223 -.1542302
reg666		-.4178311	.0384734	-10.86	0.000	-.4932681 -.3423942
reg667		-.2438311	.0367717	-6.63	0.000	-.3159313 -.1717309
reg668		-.1675116	.0546982	-3.06	0.002	-.2747614 -.0602619
_cons		.4244655	.0839573	5.06	0.000	.2598459 .5890852

```
. est store MCO3
```

Estimación 4

```
. ivreg lnw exper exper2 Afro reg661 reg662 reg663 reg664 reg665 reg666 reg667 reg668 (educ =
cercania)
```

```
Instrumental variables (2SLS) regression
```

Source	SS	df	MS	Number of obs =	3010
Model	-86.6485253	12	-7.22071044	F(12, 2997) =	34.25
Residual	679.29017	2997	.226656713	Prob > F =	0.0000
				R-squared =	.
				Adj R-squared =	.
Total	592.641645	3009	.196956346	Root MSE =	.47608

lnw	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	.2252076	.0435823	5.17	0.000	.1397534 .3106618
exper	.1477947	.0197825	7.47	0.000	.1090059 .1865834
exper2	-.2518298	.0405116	-6.22	0.000	-.3312632 -.1723964
Afro	-.0523826	.0451294	-1.16	0.246	-.1408703 .0361052
reg661	-.1023405	.0508197	-2.01	0.044	-.2019855 -.0026955
reg662	.0171478	.0376728	0.46	0.649	-.0567193 .0910149
reg663	.0561695	.0369402	1.52	0.128	-.0162613 .1286003
reg664	-.0767075	.0466074	-1.65	0.100	-.1680932 .0146781
reg665	-.0516725	.0470966	-1.10	0.273	-.1440174 .0406725
reg666	-.0328478	.0549963	-0.60	0.550	-.1406821 .0749865
reg667	-.0638896	.0474812	-1.35	0.179	-.1569885 .0292094
reg668	-.2284216	.0598252	-3.82	0.000	-.3457243 -.111119
_cons	2.242164	.7634572	2.94	0.003	.7452112 3.739118

```
Instrumented: educ
```

```
Instruments: exper exper2 Afro reg661 reg662 reg663 reg664 reg665 reg666
reg667 reg668 cercania
```

```
. est store V11
```

Contraste de Hausman

```
. hausman V11 MCO1
```

	---- Coefficients ----			
	(b)	(B)	(b-B)	sqrtdiag(V_b-V_B)
	V11	MCO1	Difference	S.E.
educ	.2252076	.0790114	.1461961	.0434374
exper	.1477947	.0875037	.060291	.0185897
exper2	-.2518298	-.2440668	-.007763	.0244212
Afro	-.0523826	-.1828211	.1304385	.0411286
reg661	-.1023405	-.139752	.0374115	.0318088
reg662	.0171478	-.0191121	.0362599	.0242211
reg663	.0561695	.0158491	.0403204	.0241739
reg664	-.0767075	-.1197227	.0430152	.0298132
reg665	-.0516725	-.1546054	.1029329	.0373707
reg666	-.0328478	-.1540476	.1211997	.0437944
reg667	-.0638896	-.1512786	.0873891	.035264
reg668	-.2284216	-.198882	-.0295396	.0366365

Valor del estadístico del contraste = 11.33

p-valor de una χ^2 con 1 grado de libertad en 11.33 = .00076265

p-valor de una χ^2 con 2 grados de libertad en 11.33 = .00346515

p-valor de una χ^2 con 12 grados de libertad en 11.33 = .50086928

p-valor de una χ^2 con 13 grados de libertad en 11.33 = .58319361

Se pide:

1) Defina "regresor endógeno" y señale qué fuentes de endogeneidad podrían estar afectando la variable *educ* en el modelo antes definido. Indique los potenciales sesgos y la dirección (signo) de los mismos.

2) Indique que características debe tener una variable para ser un instrumento válido y que contrastes conoce para corroborar dicha validez.

3) En Card (1995) se propone utilizar la variable *cercanía* como instrumento para la variable *educ*, siendo la variable *cercanía* una dummy que indica si, al momento de tomar la decisión de realizar estudios universitarios (aprox. a los 18 años), los individuos residían en una localidad en la que había una institución terciaria que ofreciera carreras de 4 años. Discuta la validez de dicho instrumento, utilice para ello la evidencia que se ofrece en las estimaciones disponibles.

4) Realice un contraste para someter a prueba la hipótesis de que la variable *educ* es exógena. Indique la hipótesis nula y la alternativa del contraste, la forma del estadístico del contraste, su distribución asintótica y su conclusión. Analice las implicancias de dicha conclusión.

5) En base a la evidencia analizada señale cuál es la medida que le resulta más apropiada para medir cuánto se espera se incrementen (en promedio) los salarios ante un incremento de 1 año de educación.

Ejercicio 2 (25 puntos)

Se considera un modelo para la probabilidad de que un individuo realice al menos una visita médica al año utilizando datos de USA. Las variables a considerar son:

Ddocvis: variable binaria que adopta el valor 1 si la persona realizó al menos una visita,
priv: variable binaria que toma el valor 1 si la atención de salud de que dispone es privada,
cron: variable binaria que toma el valor 1 si el individuo *i* tiene una afección crónica,
muj: variable binaria que toma valor 1 si el individuo es mujer,
edad: edad en años dividido 10
edad2: la variable edad al cuadrado
ing: ingreso en miles de dólares
tam: tamaño de la familia
educ: los años de educación del individuo
casado: variable binaria que toma el valor 1 si el individuo es casado(a)

A continuación se informa la estimación Logit y de los efectos parciales valorados en la media de todos los regresores.

```
. logit Ddocvis priv cron muj edad edad2 ing tam educ casado
Logistic regression               Number of obs   =       30124
                                LR chi2(9)       =       7036.05
                                Prob > chi2      =       0.0000
Log likelihood = -16109.856       Pseudo R2    =       0.1792
```

Ddocvis	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
priv	.9846743	.0372183	26.46	0.000	.9117277	1.057621
cron	1.613106	.0341729	47.20	0.000	1.546128	1.680083
muj	.9562305	.0283331	33.75	0.000	.9006986	1.011762
edad	-.1619542	.1151828	-1.41	0.160	-.3877084	.0638
edad2	.0330747	.0136565	2.42	0.015	.0063085	.0598409
ing	.0048832	.0006294	7.76	0.000	.0036496	.0061167
tam	-.125555	.0098428	-12.76	0.000	-.1448465	-.1062635
educ	.0590256	.0051823	11.39	0.000	.0488685	.0691826
casado	.4796543	.0321152	14.94	0.000	.4167097	.5425989
_cons	-1.783356	.2375673	-7.51	0.000	-2.248979	-1.317733

```
. mfx
```

```
Marginal effects after logit
      y = Pr(Ddocvis) (predict)
      = .68285094
```

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
priv*	.2300533	.00899	25.60	0.000	.212439	.247668	.819712	
cron*	.306815	.00527	58.24	0.000	.29649	.31714	.336443	
muj*	.202822	.00578	35.10	0.000	.191498	.214146	.472215	
edad	-.0350737	.02494	-1.41	0.160	-.083955	.013807	4.07427	
edad2	.0071628	.00296	2.42	0.015	.001368	.012957	17.619	
ing	.0010575	.00014	7.77	0.000	.000791	.001324	31.8973	
tam	-.0271909	.00213	-12.75	0.000	-.03137	-.023012	3.15788	
educ	.0127829	.00112	11.38	0.000	.010582	.014984	12.7243	
casado*	.1063211	.00724	14.68	0.000	.092124	.120518	.658014	

Se pide:

Parte 1

- 1) Analice los resultados de la estimación Logit.
- 2) En base a las salidas STATA ofrezca una medida (exacta o aproximada) de cómo cambia la probabilidad de realizar una visita médica en los siguientes casos considerando un individuo "promedio" en todas las características excepto por la variable respecto a la cual se quiere medir el efecto:
 - a. entre aquellos que cuentan con un seguro privado y aquellos que no
 - b. entre hombres y mujeres
 - c. entre un individuo con 15 años de educación frente a otro con 10.

Importante: Indique la fórmula que corresponde a sus respuestas y señale en cada caso si el cálculo es exacto o aproximado, si considerara que existe una fórmula más apropiada que la que utilizó en alguno de los casos a. b. o c. indíquelo y escriba dicha fórmula.
- 3) ¿Cuál sería la fórmula para obtener el efecto parcial asociado a la edad del individuo en la especificación estimada?

Parte 2

- 1) Escriba la fórmula de la probabilidad condicional $\Pr(y = 1|x)$ en el caso del modelo Logit y exprese la correspondiente ecuación en forma de error $y = \Lambda(x'\beta) + u$.
- 2) Explique cómo procedería para estimar el modelo como un problema de Mínimos Cuadrados no Lineales. Obtenga las condiciones de primer orden del estimador MCNL para el modelo Logit (considere el caso en el cual en vector x incluye una constante y un solo regresor).