

UNIVERSIDAD DE LA REPÚBLICA
FACULTAD DE CIENCIAS ECONÓMICAS Y DE ADMINISTRACIÓN
1era. REVISIÓN DE ECONOMETRÍA II
17 de mayo de 2011

Ejercicio 1

Un investigador pretende explicar los elementos que están detrás del número de hijos que tienen las mujeres en Botswana. En particular desea analizar el efecto de la educación. Para ello cuenta con una base de datos de 4.361 observaciones que incluye las variables:

- **Children:** cantidad de hijos de cada mujer
- **Educ:** años de educación de la mujer
- **Age:** edad en años de la mujer
- **Agesq:** edad al cuadrado
- **Frsthalf:** Toma valor 1 si la mujer nació durante los primeros seis meses del año

***Estimación 1:** En una primera instancia se estima el modelo que sigue por MCO:

. reg children educ age agesq					
Source	SS	df	MS	Number of obs = 4361	
Model	12243.0295	3	4081.00985	F(3, 4357) = 1915.20	
Residual	9284.14679	4357	2.13085765	Prob > F = 0.0000	
Total	21527.1763	4360	4.93742577	R-squared = 0.5687	
				Adj R-squared = 0.5684	
				Root MSE = 1.4597	
children	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	-.0905755	.0059207	-15.30	0.000	-.102183 -.0789679
age	.3324486	.0165495	20.09	0.000	.3000032 .364894
agesq	-.0026308	.0002726	-9.65	0.000	-.0031652 -.0020964
_cons	-4.138307	.2405942	-17.20	0.000	-4.609994 -3.66662

Luego, se realizaron las siguientes estimaciones:

***Estimación 2**

. reg educ frsthalf age agesq					
Source	SS	df	MS	Number of obs = 4361	
Model	7238.42472	3	2412.80824	F(3, 4357) = 175.21	
Residual	60001.141	4357	13.7712052	Prob > F = 0.0000	
Total	67239.5657	4360	15.4219187	R-squared = 0.1077	
				Adj R-squared = 0.1070	
				Root MSE = 3.711	
educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
frsthalf	-.8522854	.1128296	-7.55	0.000	-1.073489 -.6310821
age	-.1079504	.0420402	-2.57	0.010	-.1903706 -.0255302
agesq	-.0005056	.0006929	-0.73	0.466	-.0018641 .0008529
_cons	9.692864	.5980686	16.21	0.000	8.520346 10.86538

*Estimación 3

```
. ivreg children age agesq (educ = frsthalf)
```

Instrumental variables (2SLS) regression

Source	SS	df	MS	Number of obs =	4361
Model	11844.96	3	3948.32001	F(3, 4357) =	1765.12
Residual	9682.2163	4357	2.22222086	Prob > F =	0.0000
Total	21527.1763	4360	4.93742577	R-squared =	0.5502
				Adj R-squared =	0.5499
				Root MSE =	1.4907

children	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	-.1714989	.0531796	-3.22	0.001	-.2757581 -.0672398
age	.3236052	.0178596	18.12	0.000	.2885913 .3586191
agesq	-.0026723	.0002797	-9.55	0.000	-.0032206 -.0021239
_cons	-3.387805	.5481502	-6.18	0.000	-4.462459 -2.313152

Instrumented: educ

Instruments: age agesq frsthalf

*Contraste 1

```
hausman VI_1 MCO
```

	---- Coefficients ----			
	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
	VI_1	MCO	Difference	S.E.
educ	-.1714989	-.0905755	-.0809235	.052849
age	.3236052	.3324486	-.0088434	.0067142
agesq	-.0026723	-.0026308	-.0000415	.0000626

b = consistent under Ho and Ha; obtained from ivreg
B = inconsistent under Ha, efficient under Ho; obtained from regress

Test: Ho: difference in coefficients not systematic

$\chi^2(3) = (b-B)'[(V_b-V_B)^{-1}](b-B)$
= 2.34

Prob>chi2 = 0.5040

Se pide

- 1) Defina qué entiende por regresor endógeno y explique cuáles pueden ser las causas de dicho problema. Razone en el ejemplo considerado qué regresor(es) podría(n) ser endógeno(s). Describa qué propiedades estadísticas del estimador dejan de cumplirse al estimar un modelo por MCO ante la existencia de regresores endógenos. Fundamente su respuesta.
- 2) Escriba el modelo asociado a la primera estimación y analice breve pero detalladamente sus resultados.
- 3) La estimación 3 utiliza el método de variables instrumentales.
 - a. Describa en qué consiste el método de estimación por mínimos cuadrados en dos etapas y qué rol juegan las variables instrumentales en él.
 - b. ¿Qué características debe tener una variable para ser un buen instrumento? ¿De esas características, qué se puede decir sobre el instrumento que se ha elegido para la variable educación?
 - c. Comente las propiedades de los estimadores hallados a través de variables instrumentales. Suponga que el/los instrumentos utilizados son válidos.

- d. Compare resultados de la estimación 3 con los obtenidos en la estimación 1.
- 4) Explique en qué consiste el test de Hausman (detallando hipótesis, estadístico de prueba, su distribución y regla de decisión) e indique por qué se realiza en este caso. ¿Qué puede concluir a partir de los resultados obtenidos en el contraste 1?

Ejercicio 2 (25 puntos)

(Fuente: <http://www.ats.ucla.edu/stat/stata/dae/logit.htm>)

Una prestigiosa universidad de los EE.UU. recibe anualmente gran cantidad de aplicaciones por parte de estudiantes locales para poder cursar estudios de posgrado en el área de economía. En ese sentido, se desea estimar la probabilidad de ser admitido por dicha casa de estudios, para lo cual se procedió a estimar dos modelos en base a la siguiente información:

- admit:** variable binaria que toma el valor 1 si el estudiante es admitido por la universidad, y 0 en caso contrario.
- gre:** puntaje obtenido en la prueba “*Graduate Record Exam*”.
- gpa:** puntaje promedio con el que se egresa del grado
- rank:** variable discreta que recoge el prestigio de la universidad donde se cursaron los estudios de grado. Toma los valores: 1, 2, 3 y 4. El valor 1 está asociado a las universidades más reconocidas, y el 4 a las menos reconocidas. Para estimar se definen 3 variables binarias: *rank j*, que adoptan el valor 1 si rank es igual a *j* (*j*=2,3,4) y 0 en caso contrario.

Estadísticos descriptivos

. summarize gre gpa

Variable	Obs	Mean	Std. Dev.	Min	Max
gre	400	587.7	115.5165	220	800
gpa	400	3.3899	.3805668	2.26	4

. tab rank

rank	Freq.	Percent	Cum.
1	61	15.25	15.25
2	151	37.75	53.00
3	121	30.25	83.25
4	67	16.75	100.00
Total	400	100.00	

. tab admit rank

admit	rank				Total
	1	2	3	4	
0	28	97	93	55	273
1	33	54	28	12	127
Total	61	151	121	67	400

Modelo 1: Estimación por MCO

```
. reg admit gre gpa i.rank,
```

Source	SS	df	MS	Number of obs	=	400
Model	8.70247579	5	1.74049516	F(5, 394)	=	8.79
Residual	77.9750242	394	.197906153	Prob > F	=	0.0000
				R-squared	=	0.1004
				Adj R-squared	=	0.0890
Total	86.6775	399	.217236842	Root MSE	=	.44487

admit	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
gre	.0004296	.0002107	2.04	0.042	.0000153	.0008439
gpa	.155535	.0639618	2.43	0.015	.0297859	.2812842
rank						
2	-.1623653	.0677145	-2.40	0.017	-.2954922	-.0292385
3	-.2905705	.0702453	-4.14	0.000	-.428673	-.1524679
4	-.3230264	.0793164	-4.07	0.000	-.4789626	-.1670902
_cons	-.2589103	.2159904	-1.20	0.231	-.6835481	.1657275

Modelo 2: Estimación Logit

```
. logit admit gre gpa i.rank
```

```
Iteration 0: log likelihood = -249.98826
```

```
Iteration 1: log likelihood = -229.66446
```

```
Iteration 2: log likelihood = -229.25955
```

```
Iteration 3: log likelihood = -229.25875
```

```
Iteration 4: log likelihood = -229.25875
```

Logistic regression	Number of obs	=	400
	LR chi2(5)	=	-
	Prob > chi2	=	-
Log likelihood = -229.25875	Pseudo R2	=	-

admit	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
gre	.0022644	.001094	2.07	0.038	.0001202	.0044086
gpa	.8040377	.3318193	2.42	0.015	.1536838	1.454392
rank						
2	-.6754429	.3164897	-2.13	0.033	-1.295751	-.0551346
3	-1.340204	.3453064	-3.88	0.000	-2.016992	-.6634158
4	-1.551464	.4178316	-3.71	0.000	-2.370399	-.7325287
_cons	-3.989979	1.139951	-3.50	0.000	-6.224242	-1.755717

Tabla 1

. estat class

Logistic model for admit

Classified	----- Observado-----		Total
	1	0	
1	30	19	49
0	97	254	351
Total	127	273	400

Se Pide 1

- 1) Comente las estadísticas descriptivas presentadas y los resultados obtenidos en ambas estimaciones.
- 2) ¿Cuál de los dos modelos estimados le parece más apropiado para modelizar la probabilidad antes descrita? Justifique describiendo las propiedades estadísticas de los estimadores en ambos casos.
- 3) Plantee la forma de la función de verosimilitud del modelo logit.
- 4) Analice si el modelo 2 es significativo en su conjunto utilizando el test asintótico LR.

Se Pide 2

- 1) Utilizando la información proporcionada por la tabla 1, analice la capacidad predictiva del modelo 2, indique una medida de su bondad de ajuste y justifique su aplicación.
- 2) Para ambos modelos, obtenga el efecto parcial de la variable gpa e indique en signo y magnitud el efecto que genera obtener un punto adicional en dicha variable sobre la probabilidad de ser aceptado por la universidad para cursar el posgrado.
- 3) En base al modelo 2, indique en cuanto se reduce la probabilidad de ser aceptado por la universidad por el hecho de haber cursado el grado en una universidad con rank igual a 3, respecto a otra con rank igual a 1.
- 4) Suponga que un estudiante rinde la prueba "Graduate Record Exam" y obtiene 300 puntos. Conciente de su bajo puntaje estudia más, y al darla nuevamente obtiene 700 puntos. Calcule en cuanto mejora la probabilidad de ser aceptado por la universidad para hacer el posgrado.