

ECONOMETRIA II – CURSO 2016

PRÁCTICO 4

Variables instrumentales y modelos binarios

En el país “A”, durante el año 2009 se implementó un programa para apoyar la inserción laboral de personas que enfrentaban problemas para acceder al sector formal. Por restricciones presupuestales, sólo había disponibilidad para 3022 cupos, los cuales eran sorteados. En el programa se inscribieron 8328 personas. Quienes no salían sorteados como titulares, quedaban como suplentes (5306 personas). Algunos de los beneficiarios del sorteo que salieron titulares, por motivos personales eligieron no participar del programa. En ese caso, fueron sustituidos. Terminado el programa, se identificaron a aquellas personas que durante el 2010 y 2011 obtuvieron al menos un empleo formal. Se cuenta con la siguiente información:

Definición de variables

Cotiza	Cotiza=1 si la persona registra al menos una cotización en un empleo formal entre 2010 y 2011, y Cotiza=0 en otro caso
Participa Programa (Participa)	Participa=1 si la persona participó efectivamente del programa, y Participa= 0 en otro caso.
Cotización Previa (Cprevia)	Cprevia=1 si la persona registra haber tenido un empleo formal en 2007 o 2008 ; y Cprevia=0 en otro caso

Para evaluar si los participantes del programa tenían una mayor probabilidad de haber ingresado a un empleo formal en este período se especificó el siguiente modelo de probabilidad lineal: $Cotiza = \beta_0 + \beta_1 Participa + u$

MODELO I: Estimación MCO de la variable cotiza con participa del programa como regresor.

```
. reg cotizacion participa, robust
```

Linear regression

Number of obs = 8328
F(1, 8326) = 1.42
Prob > F = 0.2327
R-squared = 0.0002
Root MSE = .49323

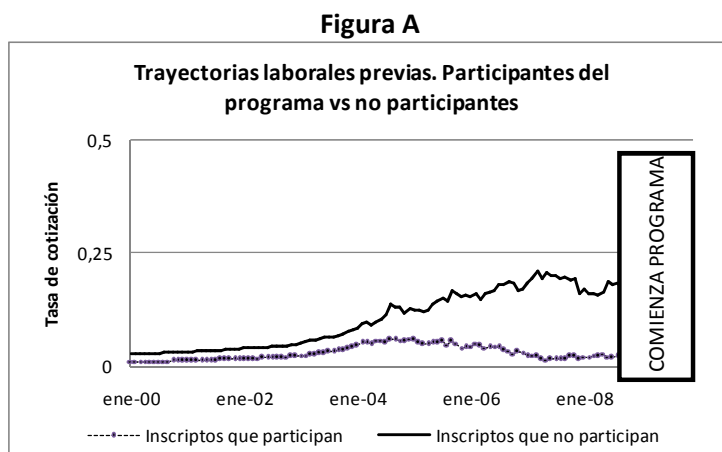
cotizacion	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
participa	.0133974	.0112262	1.19	0.233	-.0086088 .0354036
_cons	.577271	.0067825	85.11	0.000	.5639756 .5905664

Se pide I

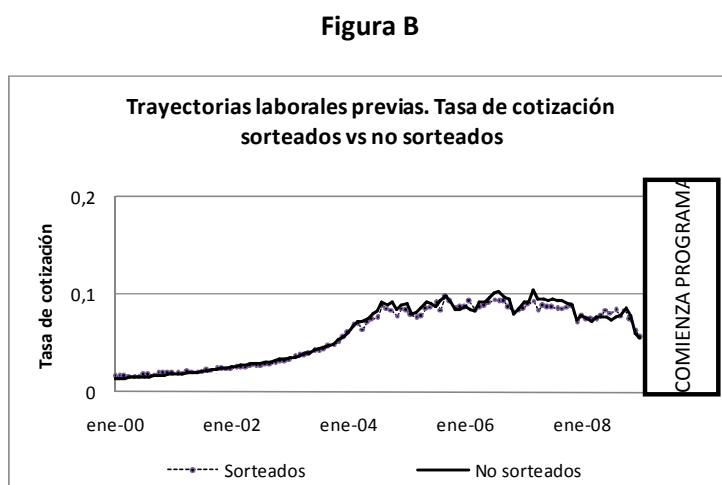
1. Comente los resultados. ¿Qué se puede concluir sobre la participación en el programa? ¿Tuvo una incidencia significativa en el acceso a un empleo formal?; de ser así ¿Cuánto se espera cambie la probabilidad de acceder a un empleo formal debido a la participación en el programa?.
2. ¿Bajo qué supuestos $\hat{\beta}_1$ es un estimador consistente del impacto de la participación en el programa?

Parte II

El equipo de evaluadores se propuso avanzar en una caracterización de las personas que efectivamente participaron del programa en relación al total de inscriptos. En primer lugar, se analizó la historia laboral previa el programa. La figura A describe las trayectorias de empleo formal de los inscriptos, distinguiendo entre quienes aceptaron participar del programa de aquellos que no. Por otra parte, la figura B presenta la historia laboral distinguiendo entre quienes salieron sorteados y los que no.



** Nota: Inscriptos que no participan son las personas que salieron sorteadas pero optaron por no realizar al programa*



** Nota: Sorteados, se refieren a todos los sorteados, independientemente si participan del programa o si optan por no participar. No sorteados, se refiere a los inscriptos que quedaron fuera del sorteo.*

Para evaluar si la trayectoria laboral previa tenía alguna incidencia en la probabilidad de participar del programa se especificó el **MODELO II**:

$$P(\text{Participa} = 1 | \text{cotización previa}) = F(PP, a_0, a_1) = \frac{1}{1 + e^{-(a_0 + a_1 \text{Cotización Previa})}}$$

Se utilizó el método de máxima verosimilitud para estimar los parámetros del modelo anterior. En un segundo paso, se estimó un modelo restringido donde $a_1 = 0$.

(1) Modelo II sin restricción

```
. logit participa cprevia, robust
```

```
Iteration 0: log pseudolikelihood = -5455.2805
Iteration 1: log pseudolikelihood = -5360.8401
Iteration 2: log pseudolikelihood = -5360.3773
Iteration 3: log pseudolikelihood = -5360.3772
```

```
Logistic regression               Number of obs   =      8328
                                Wald chi2(1)       =      179.99
                                Prob > chi2        =      0.0000
Log pseudolikelihood = -5360.3772 Pseudo R2        =      0.0174
```

participa	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
cprevia	-.7101458	.0529333	-13.42	0.000	-.8138932	-.6063984
_cons	-.3642929	.0266859	-13.65	0.000	-.4165964	-.3119894

(2) Modelo II con restricción

```
. logit participa , robust
```

```
Iteration 0: log pseudolikelihood = -5455.2805
Iteration 1: log pseudolikelihood = -5455.2805
```

```
Logistic regression               Number of obs   =      8328
                                Wald chi2(0)       =      .
                                Prob > chi2        =      .
Log pseudolikelihood = -5455.2805 Pseudo R2        =     -0.0000
```

participa	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
_cons	-.5629194	.0227911	-24.70	0.000	-.6075892	-.5182496

Se pide II

1. Calcule la probabilidad de haber participado del programa:
-Entre todos los inscriptos
-Entre los inscriptos que tuvieron al menos un empleo formal en los 2 años previos.
2. En base a la información disponible ¿Qué se puede concluir sobre las personas que aceptaron participar del programa en relación al total de sorteados que no participaron?
¿Es estadísticamente relevante la trayectoria previa?; de ser así ¿Cómo cambia la probabilidad de haber participado el tener un empleo formal en los 2 años previos?
3. En base a los resultados presentados en la parte II y la evidencia de la figura A: ¿identifica algún problema que podría estar sesgando los resultados de la estimación del Modelo I? ¿Cuál(es) de los supuestos presentados en se pide I.2 no se estaría cumpliendo?
4. En caso de que su respuesta sea positiva, proponga alguna hipótesis sobre qué sesgo podría estar operando; ¿Podría esto explicar que el parámetro que mide la incidencia de la participación en el programa sea no significativa?
5. Presente las condiciones que debe cumplir un instrumento para que la aplicación del método de variables instrumentales arroje estimadores consistentes.

Parte III

Para avanzar en esta estrategia se siguieron las recomendaciones del evaluador 1 y se estimó un modelo por mínimo cuadrado en dos etapas. A continuación se presentan los resultados:

Definición de variables

Cotiza	=1 si la persona registra haber tenido al menos una cotización en un empleo formal entre 2010 y 2011, y Cotiza=0 en otro caso
Participa Programa (Participa)	=1 si la persona participó efectivamente del programa, 0 en otro caso.
Cotización Previa (Cprevia)	=1 si la persona registra haber tenido un empleo formal en 2007 o 2008 ; y 0 en otro caso
Sorteo	=1 si la persona salió sorteada para participar en el programa; =0 en otro caso
Hombre	=1 si la persona es hombre; =0 en otro caso

MODELO III: Estimación por variables instrumentales de la variable *Cotiza*. Se utiliza como regresor *participa del programa (PP)* y como instrumento “sorteo”

```
. ivreg cotizacion ( participa = sorteo)
```

Instrumental variables (2SLS) regression

Source	SS	df	MS	Number of obs =	8328
Model	-12.1402713	1	-12.1402713	F(1, 8326) =	8.37
Residual	2037.9616	8326	.24477079	Prob > F =	0.0038
				R-squared =	.
				Adj R-squared =	.
Total	2025.82133	8327	.243283455	Root MSE =	.49474

cotizacion	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
participa	.0939258	.0324752	2.89	0.004	.0302663 .1575852
_cons	.5480495	.0129716	42.25	0.000	.522622 .5734771

Instrumented: participa
Instruments: sorteo

TEST DE HAUSMAN (Notar que el modelo MCO1 se basa en especificación del modelo I, ejercicio1)

```
. hausman VI MCO
```

	Coefficients		(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
	(b) VI	(B) MCO		
participa	.0939258	.0133974	.0805283	.0304678

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

chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
 = 6.99
 Prob>chi2 = 0.0082

MODELO IV: Estimación por MCO de la variable PARTICIPA. Se utiliza como regresor “sorteo”

```
. reg participa sorteo , robust
```

Linear regression

Number of obs = 8328
 F(1, 8326) = 1048.63
 Prob > F = 0.0000
 R-squared = 0.1205
 Root MSE = .45097

participa	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
sorteo	.3508664	.0108351	32.38	0.000	.3296269 .3721058
_cons	.2413666	.0058003	41.61	0.000	.2299967 .2527366

Se pide III

- En base a la evidencia presentada en la información disponible discuta si la variable “sorteo” satisface las condiciones que debe cumplir un instrumento.
- Comente las argumentaciones de los dos evaluadores presentadas a continuación.
*Ante los potenciales sesgos que podría tener la estimación de θ_1 presentada en el ejercicio 1, los evaluadores buscaron estrategias alternativas y surgieron dos posiciones. El **evaluador 1** estaba convencido de que el modelo I tenía problemas de inconsistencia. Observó que si bien algunos sorteados deciden no participar del programa, existía una correlación positiva muy alta entre participar del programa y salir sorteado. En base a eso propuso utilizar la variable “sorteo” como instrumento de la participación en el programa.*
*El **evaluador 2**, si bien estaba de acuerdo con que deberían definir una estrategia alternativa, no estuvo de acuerdo con utilizar el método de variables instrumentales, fundamentando que no se aplica a esta situación. Además, si bien estaba de acuerdo que “sorteo” y “participar” estaban correlacionados, de la figura B este evaluador desprendía que el sorteo no era un buen instrumento*
- Plantee el test de Hausman (supuestos, hipótesis, estadístico). ¿Cuál es su utilidad en el contexto de este análisis? Fundamente. Analice los resultados de este test y sus implicancias en el problema bajo estudio. ¿Qué aporta a la discusión entre los dos evaluadores?
- Compare los resultados de la estimación por VI con los del modelo I del ejercicio