

# Ejercicios variables instrumentales

Alejandro Trujillo y Raquel Badillo

El Colegio de México

9 Noviembre 2018

- 1 Ejercicio C15.1
  - Rendimiento educación hombre
  - Ejercicio
- 2 Ejercicio C15.2
  - Educación e hijos
- 3 Ejercicio C15.3
  - Salarios y educación
- 4 Ejercicio C15.4
  - Tasa de interés e inflación

# Rendimiento educación hombres

- Queremos estimar el rendimiento de la educación de los hombres respecto a su salario (en log)
- Utilizamos el número de hermanos como un instrumento de educación y hacemos una regresión simple en el primer paso obtenemos:

Cuadro 1: Primera etapa educación-núm de hermanos

	Educación
# hermanos	-0.228*** (0.030)
Observaciones	935
R <sup>2</sup>	0.057
Residual Std. Error	2.134 (df = 933)
F	56.667*** (df = 1; 933)

*Nota:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## IV con hermanos

Ahora aplicamos IV para observar cómo es el coeficiente de escolaridad

Asumimos que hermanos es un buen instrumento porque no está correlacionado con el error (aunque podría estarlo) y ya vimos que sí esta correlacionada con la variable endógena.

	Salario
Educación	0.122*** (0.026)
Constante	5.130*** (0.355)
Observaciones	935
R <sup>2</sup>	-0.009
Residual Std. Error	0.423 (df = 933)
Note:	*p<0.1; **p<0.05; ***p<0.01

# Ejercicio C15.1

- i) Estimar un mco de **número de hermanos** en vez de **educación** (es decir no podemos ocupar el instrumento como variable explicativa simplemente)

	Salario
# Hermanos	-0.028*** (0.006)
Constante	6.861*** (0.022)
Observaciones	935
R <sup>2</sup>	0.023
Residual Std. Error	0.416 (df = 933)
F Statistic	22.307*** (df = 1; 933)
Note:	* p<0.1; ** p<0.05; *** p<0.01

ii)

Busca cómo el orden de nacimiento y la educación están negativamente correlacionadas

Los padres podrían ser menos exigentes conforme más hijos tienen //  
Los padres pueden tener menos recursos y tiempo para preocuparse por la educación de sus hijos

- Realiza una regresión de educación y orden de nacimiento

	Educación
Orden de nacimiento	-0.283*** (0.046)
Constante	14.149*** (0.129)
Observaciones	852
R <sup>2</sup>	0.042
Residual Std. Error	2.155 (df = 850)
F Statistic	37.285*** (df = 1; 850)

Note:

\* p<0.1; \*\* p<0.05; \*\*\* p<0.01

iii) Utiliza el orden de nacimiento como una variable instrumental (en vez de número de hermanos) Reporte e interprete los resultados

	Salario
Educación	0.131*** (0.032)
Constante	5.030*** (0.433)
Observaciones	852
R <sup>2</sup>	-0.029
Adjusted R <sup>2</sup>	-0.030
Residual Std. Error	0.422 (df = 850)
Note: * p<0.1; ** p<0.05; *** p<0.01	

## Interpretación

El coeficiente de educación encontrado con la variable instrumental orden de nacimiento es muy similar al del número de hermanos. Podemos interpretarlo como un año más de educación da un rendimiento de 13.1 por ciento más de salario, pues está en logaritmos.

- iv) Ahora supon que el número de hermanos es una variable explicativa. Corre el primer paso de la ecuación ocupando como instrumento el orden de nacimiento (ols)

	Educación
# de hermanos	-0.153*** (0.040)
Orden de nacimiento	-0.153*** (0.057)
Constante	14.296*** (0.133)
Observaciones	852
R <sup>2</sup>	0.058
Residual Std. Error	2.137 (df = 849)
F Statistic	26.293*** (df = 2; 849)
<i>Note:</i> * p<0.1; ** p<0.05; *** p<0.01	

¿cumple supuesto de identificación?

Podemos observar que cumple con el supuesto de identificación porque el orden de nacimiento sigue siendo estadísticamente significativa



v) Estime por IV la ecuación y analiza los errores estándares de las betas de las variables estimadas

	Salario
Educación	0.137* (0.075)
# Hermanos	0.002 (0.017)
Constante	4.939*** (1.056)
Observaciones	852
R <sup>2</sup>	-0.054
Residual Std. Error	0.427 (df = 849)
Note:	* p<0.1; ** p<0.05; *** p<0.01

## Interpretación

Podemos observar que los errores estándares de ambas variables crecieron, lo que provoca que dejen de ser estadísticamente significativas al 5 por ciento

vi) ¿cuál es la correlación entre la educación predicha ( $\hat{educ}$ ) y el número de hermanos?

### Respuesta

La correlación es  $-0,9294818$  la cual es altamente negativa. Asumir que  $brthour$  es exogena no es un supuesto creíble tomando en cuenta la alta autorrelación entre las variables.

# Educación e hijos

- Estimaremos el número de hijos de las mujeres de Bostwana durante 1988 con predictores como edad y educación al cuadrado
- Estima una regresión del numero de hijos contra educacion edad y edad al cuadrado

	# de niños
Educación	-0.091*** (0.006)
Edad	0.332*** (0.017)
Edad2	-0.003*** (0.0003)
Constante	-4.138*** (0.241)
Observaciones	4,361
R <sup>2</sup>	0.569
Residual Std. Error	1.460 (df = 4357)
F Statistic	1,915.196*** (df = 3; 4357)
Note:	* p<0.1; ** p<0.05; *** p<0.01

Podemos decir que un año adicional de educación manteniendo la edad fija reduciría el numero de hijos en 0.091 (por cada 100 mujeres habrá 9.0575 menos niños)

ii) Suponiendo que la variable nació en la primera mitad del año es exógena de la decisión de tener hijos demuestra que es un instrumento válido

	Educación
Nació 1ra mitad	-0.938*** (0.118)
Constante	6.363*** (0.087)
Observaciones	4,361
R <sup>2</sup>	0.014
Residual Std. Error	3.900 (df = 4359)
F Statistic	62.620*** (df = 1; 4359)
Note: * p<0.1; ** p<0.05; *** p<0.01	

## Respuesta

Podemos observar que es relevante la para ecuación por lo tanto si suponemos que también es exógeno podemos confirmar que la variable nació en la primera mitad del año es un instrumento válido

iii) Estime el modelo del inciso i) pero con primera mitad como instrumento, compare

	# de hijos
Educación	-0.171*** (0.053)
Edad	0.324*** (0.018)
Edad2	-0.003*** (0.0003)
Constante	-3.388*** (0.548)
Observaciones	4,361
R <sup>2</sup>	0.550
Residual Std. Error	1.491 (df = 4357)
Note: *p<0.1; **p<0.05; ***p<0.01	

## Respuesta

Encontramos que el efecto limpio es mucho mayor con variables instrumentales. El efecto casi se duplica con el inciso 1

iv) Agregue las variables electricidad, tv y bicicleta compare los coeficientes

	Hijos	
	<i>MCO</i>	<i>Instrumental</i>
	(1)	(2)
Educación	−0.077*** (0.006)	−0.164** (0.066)
Edad	0.340*** (0.016)	0.328*** (0.019)
Edad2	−0.003*** (0.0003)	−0.003*** (0.0003)
Electric	−0.303*** (0.076)	−0.107 (0.166)
TV	−0.253*** (0.091)	−0.003 (0.209)
Bicicleta	0.318*** (0.049)	0.332*** (0.052)
Constante	−4.390*** (0.240)	−3.591*** (0.645)
Observaciones	4,356	4,356
R <sup>2</sup>	0.576	0.558
Adjusted R <sup>2</sup>	0.575	0.557
Residual Std. Error (df = 4349)	1.448	1.479
F Statistic	984.921*** (df = 6; 4349)	
<i>Note:</i> * p<0.1; ** p<0.05; *** p<0.01		

- Queremos estimar el salario de los trabajadores respecto a la educación que recibieron.
- Utilizamos el crecer cerca de un colegio como un instrumento de educación.

i) ¿podría la variable instrumental estar correlacionada con objetos en el término de error?

Es de tener cuidado porque uno podría argumentar que los colegios eligen su ubicación geográfica en lugares donde se sabe hay más habilidad entre estudiantes.

- ii) Corre el modelo en MCO entre IQ y crecer cerca de la universidad para ver si este es un buen instrumento

	IQ
Cerca de la universidad	2.596*** (0.745)
Constante	100.611*** (0.627)
Observaciones	2,061
R <sup>2</sup>	0.006
Residual Std. Error	15.382 (df = 2059)
F Statistic	12.128*** (df = 1; 2059)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	



- iii) Agrega las dummies de ubicación geográfica. ¿Sigue siendo significativa la variable de crecer cerca de la universidad?

	IQ
Cerca de la universidad	0.348 (0.814)
smsa66	1.089 (0.809)
reg662	1.099 (1.650)
reg663	-1.559 (1.623)
...	
reg668	-2.925 (2.345)
reg669	-2.892 (1.797)
Constante	104.773*** (1.625)
Observaciones	2,061
R <sup>2</sup>	0.063
Residual Std. Error	14.969 (df = 2050)
F Statistic	13.700*** (df = 10; 2050)
Note: * p<0.1; ** p<0.05; *** p<0.01	

iv) Dados los incisos anteriores, ¿qué concluye respecto a la importancia de incluir las dummies?

En efecto el incluirlas le quita significancia a la variable de crecer cerca de la universidad. Esto implica que, en efecto, esta estaba correlacionada con el error.

# Tasa de interés e inflación

- i) Corre una regresión simple entre ambas variables, omitiendo la primera observación.

	Tasa de interés
Inflación en t	0.698*** (0.087)
Constante	2.321*** (0.423)
Observaciones	55
R <sup>2</sup>	0.546
Residual Std. Error	1.934 (df = 53)
F Statistic	63.858*** (df = 1; 53)
Note:	* p<0.1; ** p<0.05; *** p<0.01

- ii) utiliza la variable de inflación rezagada como variable instrumental para la inflación en t

	Tasa de interés
Inflación en t (VI: inflación en t-1)	0.902*** (0.134)
Constante	1.543** (0.579)
Observaciones	55
R <sup>2</sup>	0.500
Residual Std. Error	2.032 (df = 53)
<i>Note:</i> * p<0.1; ** p<0.05; *** p<0.01	

- iii) Vuelve a correr la primera regresión, pero en primeras diferencias.

	$\Delta Tasadeinterés$
$\Delta Inflación_t$	0.221*** (0.072)
Constante	0.023 (0.172)
Observaciones	55
R <sup>2</sup>	0.152
Adjusted R <sup>2</sup>	0.136
Residual Std. Error	1.271 (df = 53)
F Statistic	9.533*** (df = 1; 53)
Note:	* p<0.1; ** p<0.05; *** p<0.01

iv) puedes usar la variable  $\Delta Inflaci3n_{t-1}$  como instrumento de  $\Delta Inflaci3n_t$

No. La siguiente regresi3n simple muestra que estas no est3n correlacionadas.

	$\Delta Inflaci3n_t$
$\Delta Inflaci3n_{t-1}$	-0.010 (0.118)
Constant	0.064 (0.286)
Observations	54
R <sup>2</sup>	0.0001
Residual Std. Error	2.096 (df = 52)
F Statistic	0.008 (df = 1; 52)
Note:	* p<0.1; ** p<0.05; *** p<0.01