

Sesión 26 (B) — ¿Compra votos el gasto electoral? Las elecciones a la Cámara de Representantes de 1988 con `vote1`

Marcos Bujosa

Descripción de la práctica

¿Obtiene más votos el candidato que más gasta? Los datos son `vote1`, de Wooldridge (pestaña Wooldridge de los ficheros de muestra de Gretl): 173 elecciones a la Cámara de Representantes de Estados Unidos en 1988, una por distrito, con dos candidatos en cada una. Para el candidato A, uno de los dos en cada distrito, el fichero da el porcentaje de voto (`voteA`), el gasto de campaña en miles de dólares (`expendA`) y el de su rival (`expendB`), el porcentaje de voto que el partido de A obtuvo en el distrito en las elecciones presidenciales más recientes (`prtystrA`, la fuerza del partido), el porcentaje del gasto total de la elección que corresponde a A (`shareA`) y la indicadora `democA`, que vale 1 si A es demócrata.

La práctica estima la relación entre el voto y el gasto con las herramientas de las lecciones 13, 15 y 16, y termina con la pregunta que importa para la política: si el coeficiente del gasto mide lo que ganaría un candidato gastando más, o solo cómo varía el voto entre candidatos que gastan distinto.

Objetivos

1. Leer un modelo nivel-log con dos gastos: cuántos puntos de voto acompañan a un 1 % más de gasto propio y a un 1 % más de gasto del rival.
2. Contrastar con `restrict`, y por sustitución, que un aumento proporcional de los dos gastos no cambia el reparto del voto.
3. Leer el modelo con la cuota de gasto y compararlo con el modelo nivel-log de la Actividad 2.
4. Distinguir la asociación que el modelo mide del efecto causal que la pregunta pide.

Comandos nuevos de esta práctica. No hay ninguno: `open`, `summary --simple`, `series`, `printf` (sesiones 4 y 7), `ols` y sus accesorios (sesiones 11 a 18), `restrict` y `$test` ([actividad 2 \(html\)](#) de la práctica A de la sesión 20), `$vcv` y `critical` ([actividad 2 \(html\)](#) de la práctica B de la sesión 25).

Actividad 1 - Los datos

Abra `vote1` (Archivo -> Abrir datos -> Archivo de muestra, pestaña Wooldridge), mire los estadísticos de las variables y cree el logaritmo del gasto de A.

`open` carga el fichero y `summary --simple` da los estadísticos principales. El fichero trae `lexpendB`, el logaritmo del gasto de B, pero no el de A: `series lexpendA = log(expendA)` lo crea. El `printf` cuenta las elecciones y las que ganó A, con la comparación `voteA > 50`, que vale 1 cuando es cierta.

o bien teclee en línea de comandos:

```
open vote1 --quiet
```

```
summary voteA expendA expendB prtystA shareA --simple
series lexpendA = log(expendA)
printf "Elecciones: %d ; ganadas por A (voteA > 50): %d\n", $nobs, sum(voteA > 50)
```

	Media	Mediana	D. T.	Mín	Máx
voteA	50,50	50,00	16,78	16,00	84,00
expendA	310,6	242,8	281,0	0,3000	1471
expendB	305,1	221,5	306,3	0,9300	1548
prtystA	49,76	50,00	9,984	22,00	71,00
shareA	51,08	50,85	33,48	0,09000	99,50

Elecciones: 173 ; ganadas por A (voteA > 50): 83

Lo que debe observar

Son 173 elecciones y A ganó 83, así que A no es siempre el ganador; el fichero no dice con qué criterio se designa A en cada distrito ni cuál de los dos candidatos es el titular del escaño. El porcentaje de voto de A tiene media 50,5. El gasto es muy asimétrico: la media de `expendA` es 311 mil dólares, la mediana 243 y el máximo 1471; la mitad de los candidatos gasta menos de un cuarto de millón y unos pocos gastan más de un millón. Esa asimetría es la razón para medir el gasto en logaritmos (lección 15, “Lo que cambia y lo que no”): lo que cabe esperar que importe es gastar el doble, no gastar mil dólares más, y el logaritmo mide variaciones relativas. La fuerza del partido va de 22 a 71 puntos.

Actividad 2 - El voto y los dos gastos, en logaritmos

Estime el porcentaje de voto de A sobre el logaritmo de su gasto, el logaritmo del gasto de su rival y la fuerza del partido.

El `ols` estima el modelo nivel-log (Modelo $\rightarrow$ Mínimos cuadrados ordinarios). El `printf` traduce los dos coeficientes del gasto a puntos de voto por un 1 % más de gasto, dividiendo por 100 (lección 15, “Nivel-log: la superficie de la vivienda”), y calcula cuántos puntos acompañan a duplicar el gasto de A, multiplicando el coeficiente por $\log 2$.

```
ols voteA const lexpendA lexpendB prtystA
printf "1% mas de gasto de A: %.3f puntos ; 1% mas de gasto de B: %.3f puntos ; duplicar el gasto de A: %.1f puntos\n",
↪ $coeff(lexpendA)/100, $coeff(lexpendB)/100, $coeff(lexpendA)*log(2)
```

Modelo 1: MCO, usando las observaciones 1-173
Variable dependiente: voteA

	coeficiente	Desv. típica	Estadístico t	valor p
const	45,0860	3,92679	11,48	6,34e-23 ***
lexpendA	6,08133	0,382119	15,91	1,95e-35 ***
lexpendB	-6,61527	0,378876	-17,46	1,11e-39 ***
prtystA	0,152014	0,0620267	2,451	0,0153 **
Media de la vble. dep.	50,50289	D.T. de la vble. dep.	16,78476	
Suma de cuad. residuos	10055,08	D.T. de la regresión	7,713464	
R-cuadrado	0,792496	R-cuadrado corregido	0,788812	
F(3, 169)	215,1472	Valor p (de F)	1,81e-57	
Log-verosimilitud	-596,8862	Criterio de Akaike	1201,772	
Criterio de Schwarz	1214,386	Crit. de Hannan-Quinn	1206,890	

1% mas de gasto de A: 0,061 puntos ; 1% mas de gasto de B: -0,066 puntos ; duplicar el gasto de A: 4,2 puntos

Lo que debe observar

El coeficiente de `lexpendA` es 6,08 con error estándar 0,38 y el de `lexpendB`, -6,62 con error estándar 0,38. El modelo es nivel-log: si el gasto de A varía en la fracción p con el gasto de B y la fuerza del partido fijos, el voto de A varía en $6,08p$ puntos; un 1 % más de gasto se acompaña de 0,06 puntos más, un 10 % de 0,6 puntos, y duplicar el gasto, de 4,2 puntos ($6,08 \times \log 2$). El gasto del rival resta en una magnitud parecida: un 1 % más de gasto de B, 0,07 puntos menos para A. Los t son 15,9 y -17,5: datos como estos serían muy raros si alguno de los dos coeficientes fuese cero. A la fuerza del partido le corresponden 0,15 puntos de voto por cada punto de voto presidencial, con $t = 2,45$. El R^2 es 0,79.

La constante, 45,1, es el valor ajustado en el punto en que los dos logaritmos y la fuerza del partido valen cero. Los logaritmos cero corresponden a un gasto de mil dólares, dentro del rango del fichero, pero ninguna elección tiene `prtystrA` igual a cero, el mínimo es 22: la constante no se lee.

Actividad 3 - ¿Se anula un aumento proporcional de los dos gastos?

Si los dos candidatos aumentan su gasto en la misma fracción, el voto de A cambia en $(\beta_2 + \beta_3)p$ puntos. Contraste $H_0: \beta_2 + \beta_3 = 0$, con `restrict` y por sustitución.

El primer bloque calcula la suma de los dos coeficientes, su error estándar con las varianzas y la covarianza que da `$vcv`, el t de la suma frente a cero y el intervalo al 95 %; después, `restrict` impone la restricción (Contrastes $\rightarrow$ Restricciones lineales) y `$test` y `$pvalue` recogen el F y su valor p . El segundo bloque impone la restricción por sustitución, como en la lección 13: con $\beta_3 = -\beta_2$, el modelo se escribe con un solo regresor de gasto, la diferencia de logaritmos $\log(\text{expendA}) - \log(\text{expendB})$, que es el logaritmo del cociente de gastos; antes se guarda en `SRC` la suma de cuadrados del modelo sin restringir.

```
scalar suma = $coeff(lexpendA) + $coeff(lexpendB)
scalar eeSuma = sqrt($vcv(lexpendA, lexpendA) + $vcv(lexpendB, lexpendB) + 2*$vcv(lexpendA, lexpendB))
scalar tc = critical(t, $df, 0.025)
printf "b2 + b3 = %.3f ; error estandar %.3f ; t = %.2f ; intervalo al 95%: [%.2f, %.2f]\n", suma, eeSuma, suma/eeSuma,
↪ suma - tc*eeSuma, suma + tc*eeSuma
scalar SRC = $ess
restrict --quiet
    b[lexpendA] + b[lexpendB] = 0
end restrict
printf "F(1,%d) = %.2f    valor p = %.3f\n", $df, $test, $pvalue
```

```
series ldif = lexpendA - lexpendB
ols voteA const ldif prtystrA
printf "SRC sin restringir %.1f ; SRC restringido %.1f\n", SRC, $ess
```

b2 + b3 = -0,534 ; error estandar 0,533 ; t = -1,00 ; intervalo al 95%: [-1,59, 0,52]

Restricción:

b[lexpendA] + b[lexpendB] = 0

Estadístico de contraste: F(1, 169) = 1,0031, con valor p = 0,317994

F(1,169) = 1,00 valor p = 0,318

Modelo 2: MCO, usando las observaciones 1-173

Variable dependiente: voteA

	coeficiente	Desv. típica	Estadístico t	valor p
const	42,7016	3,12285	13,67	3,45e-29 ***
ldif	6,35062	0,271524	23,39	5,14e-55 ***
prtystrA	0,146439	0,0617769	2,370	0,0189 **

Media de la vble. dep. 50,50289 D.T. de la vble. dep. 16,78476
Suma de cuad. residuos 10114,76 D.T. de la regresión 7,713534

R-cuadrado	0,791264	R-cuadrado corregido	0,788808
F(2, 170)	322,2133	Valor p (de F)	1,46e-58
Log-verosimilitud	-597,3981	Criterio de Akaike	1200,796
Criterio de Schwarz	1210,256	Crit. de Hannan-Quinn	1204,634

SRC sin restringir 10055,1 ; SRC restringido 10114,8

Lo que debe observar

La suma estimada es $-0,53$ con error estándar $0,53$: $t = -1,00$, $F(1, 169) = 1,00$ y valor p $0,32$. Datos como estos no serían raros si los dos coeficientes del gasto sumasen cero, es decir, si un aumento proporcional de los dos gastos no cambiase el voto, y el contraste al 5% no rechaza H_0 . El intervalo al 95% para la suma va de $-1,59$ a $0,52$: si los dos gastos se duplicasen, deja sin descartar desde $1,1$ puntos menos para A ($-1,59 \times \log 2$) hasta $0,4$ puntos más.

En el modelo restringido, el coeficiente del logaritmo del cociente de gastos es $6,35$ con error estándar $0,27$: lo que cuenta es cuántas veces gasta A lo que gasta B, no cuánto gasta cada uno. La suma de cuadrados de los residuos sube de 10055 a 10115 , un $0,6\%$, y ese aumento dividido por 5^2 es el F .

Actividad 4 - La cuota de gasto

El fichero trae `shareA`, el porcentaje del gasto total de la elección que corresponde a A. Estime el voto de A sobre la cuota de gasto.

El `ols` estima la regresión simple de la lección 8. El `printf` imprime el valor ajustado con la mitad del gasto, `shareA` igual a 50 .

```
ols voteA const shareA
printf "Con la mitad del gasto (shareA = 50): voto ajustado %.1f%%\n", $coeff(const) + 50*$coeff(shareA)
```

Modelo 3: MCO, usando las observaciones 1-173
Variable dependiente: voteA

	coeficiente	Desv. típica	Estadístico t	valor p
const	26,8125	0,887189	30,22	1,72e-70 ***
shareA	0,463824	0,0145393	31,90	6,61e-74 ***

Media de la vble. dep.	50,50289	D.T. de la vble. dep.	16,78476
Suma de cuad. residuos	6970,774	D.T. de la regresión	6,384729
R-cuadrado	0,856146	R-cuadrado corregido	0,855305
F(1, 171)	1017,704	Valor p (de F)	6,61e-74
Log-verosimilitud	-565,1968	Criterio de Akaike	1134,394
Criterio de Schwarz	1140,700	Crit. de Hannan-Quinn	1136,952

Con la mitad del gasto (shareA = 50): voto ajustado 50,0%

Lo que debe observar

El coeficiente de `shareA` es $0,464$ con error estándar $0,015$: a un punto más de cuota de gasto le corresponde casi medio punto más de voto, y con la mitad del gasto el voto ajustado es el $50,0\%$: el modelo reparte el voto a medias cuando el gasto se reparte a medias. El R^2 es $0,856$, mayor que el $0,792$ de la Actividad 2, y los dos sí son comparables, porque el regresando es el mismo, `voteA`. La cuota de gasto, una sola variable, ajusta mejor que los dos logaritmos y la fuerza del partido; es una forma funcional distinta de la misma idea de la Actividad 3, que lo que importa es el reparto del gasto.

Actividad 5 - ¿Efecto del gasto o asociación con el gasto?

Añada la indicadora **democA** al modelo de la Actividad 2 y, con la tabla delante, responde a la pregunta de la práctica.

El **ols** añade **democA**; su coeficiente es la diferencia de voto entre un demócrata y un republicano con el mismo gasto, el mismo gasto del rival y la misma fuerza de partido (lección 16).

```
ols voteA const lexpendA lexpendB prtystA democA
```

Modelo 4: MCO, usando las observaciones 1-173

Variable dependiente: voteA

	coeficiente	Desv. típica	Estadístico t	valor p
const	37,6661	4,73669	7,952	2,57e-13 ***
lexpendA	5,77732	0,391783	14,75	4,13e-32 ***
lexpendB	-6,23759	0,397514	-15,69	9,51e-35 ***
prtystA	0,251999	0,0713016	3,534	0,0005 ***
democA	3,79398	1,40671	2,697	0,0077 ***
Media de la vble. dep.	50,50289	D.T. de la vble. dep.	16,78476	
Suma de cuad. residuos	9637,781	D.T. de la regresión	7,574150	
R-cuadrado	0,801108	R-cuadrado corregido	0,796372	
F(4, 168)	169,1694	Valor p (de F)	8,29e-58	
Log-verosimilitud	-593,2197	Criterio de Akaike	1196,439	
Criterio de Schwarz	1212,206	Crit. de Hannan-Quinn	1202,836	

Lo que debe observar

El coeficiente de **democA** es 3,79 con error estándar 1,41 y $t = 2,70$: en 1988, entre dos candidatos con el mismo gasto, el mismo rival y la misma fuerza de partido, al demócrata le corresponden 3,8 puntos más. Los coeficientes del gasto apenas cambian (5,78 y $-6,24$).

Ahora la pregunta. El coeficiente de **lexpendA** describe cómo varía el voto entre candidatos que gastan distinto, con lo demás fijo. Para leerlo como lo que ganaría un candidato dado si gastase un 1 % más hace falta el supuesto de exogeneidad de la lección 9, $E[U | \mathbf{X}] = 0$: que lo que el modelo no recoge, la calidad del candidato, si es el titular del escaño, lo reñida que se espera la elección, sea ortogonal a las funciones de los regresores. Y hay razones para pensar que no lo es: las donaciones llegan a los candidatos con posibilidades, y el gasto sube en las elecciones reñidas. El modelo mide una asociación; qué parte de ella es efecto del gasto no se puede saber con estos datos y este modelo. Esta cautela es la advertencia de la lección 9, “Cuidado: Esta descomposición ortogonal no implica causalidad”, en un caso en el que la pregunta de política es la causal.

Preguntas de interpretación para la clase

1. El coeficiente de **lexpendA** es 6,08. ¿Cuántos puntos de voto acompañan a un 1 % más de gasto, a un 10 % y a duplicarlo? ¿Por qué duplicar el gasto no acompaña a 100 veces los puntos de un 1 % más?
2. ¿Por qué el gasto entra en logaritmos y el voto en niveles? ¿Qué cambiaría en la lectura con el gasto en niveles?
3. ¿Qué hipótesis contrasta la Actividad 3 y qué significa que no se rechace? ¿Qué dice el intervalo de $-1,59$ a $0,52$ que el valor p no dice?
4. La cuota de gasto tiene un R^2 de 0,856 y el modelo de la Actividad 2, 0,792. ¿Son comparables? ¿Lo serían si el regresando de uno de los dos fuese el logaritmo del voto?
5. ¿Qué mide el coeficiente de **democA** y qué no mide?

6. Un partido lee la tabla y concluye que, duplicando el gasto de sus candidatos, ganará 4 puntos en cada distrito. ¿Qué supuesto está haciendo y por qué es dudoso?

Para profundizar

- Wooldridge, J. M. (2020), *Introductory Econometrics*, cap. 2 (el ejemplo de la cuota de gasto) y cap. 4, sección 4.4 (el ejemplo del gasto de campaña y el contraste de $\beta_2 = -\beta_3$); los datos proceden de la Federal Election Commission.
- Jacobson, G. C. (1978), “The effects of campaign spending in congressional elections”, *American Political Science Review* 72, 469–491: el artículo clásico sobre la pregunta, con la discusión de por qué el gasto del titular del escaño y el del aspirante no se leen igual.
- Wooldridge, J. M. (2020), cap. 15 (variables instrumentales): la herramienta que los economistas usan cuando el regresor no es exógeno; queda fuera de este curso.

Código completo de la práctica

```
# -----
# Copyright (C) 2026 Marcos Bujosa
# Licencia: GNU General Public License v3.0 o posterior
# Este código es software libre y puede ser redistribuido y/o modificado bajo los términos de la GPL.
# Ver el archivo LICENSE del repositorio para más detalles.
# -----

# Los dos primeros comandos son necesarios para que Gretl guarde los resultados de la práctica en el
↳ directorio de trabajo
# al ejecutar lo siguiente desde un terminal (use los nombres y ruta que correspondan)
#
#   DIRECTORIO="Nombre_Directorio_trabajo" gretlcli -b ruta/nombre_fichero_de_la_practica.inp
#
# Si esto no le funciona en su sistema, comente las siguientes dos líneas y sitúese en el directorio de
↳ trabajo de gretl
# que corresponda (configure dicho directorio de trabajo desde la ventana principal de Gretl).

string directory = getenv("DIRECTORIO")
set workdir "@directory"

set verbose off

open vote1 --quiet
outfile --quiet datos.txt
    summary voteA expendA expendB prtystA shareA --simple
    series lexpendA = log(expendA)
    printf "Elecciones: %d ; ganadas por A (voteA > 50): %d\n", $nobs, sum(voteA > 50)
end outfile

outfile --quiet nivelLog.txt
    ols voteA const lexpendA lexpendB prtystA
    printf "1% mas de gasto de A: %.3f puntos ; 1% mas de gasto de B: %.3f puntos ; duplicar el gasto de
↳ A: %.1f puntos\n", $coeff(lexpendA)/100, $coeff(lexpendB)/100, $coeff(lexpendA)*log(2)
end outfile

outfile --quiet sumaCero.txt
    scalar suma = $coeff(lexpendA) + $coeff(lexpendB)
    scalar eeSuma = sqrt($vcv(lexpendA, lexpendA) + $vcv(lexpendB, lexpendB) + 2*$vcv(lexpendA, lexpendB))
    scalar tc = critical(t, $df, 0.025)
    printf "b2 + b3 = %.3f ; error estandar %.3f ; t = %.2f ; intervalo al 95%: [%.2f, %.2f]\n", suma,
↳ eeSuma, suma/eeSuma, suma - tc*eeSuma, suma + tc*eeSuma
    scalar SRC = $ess
    restrict --quiet
        b[lexpendA] + b[lexpendB] = 0
    end restrict
    printf "F(1,%d) = %.2f valor p = %.3f\n", $df, $test, $pvalue
    series ldif = lexpendA - lexpendB
    ols voteA const ldif prtystA
    printf "SRC sin restringir %.1f ; SRC restringido %.1f\n", SRC, $ess
end outfile

outfile --quiet cuota.txt
    ols voteA const shareA
    printf "Con la mitad del gasto (shareA = 50): voto ajustado %.1f%\n", $coeff(const) +
↳ 50*$coeff(shareA)
end outfile

outfile --quiet democrata.txt
    ols voteA const lexpendA lexpendB prtystA democA
end outfile
```

Respuestas

1. Un 1 % más de gasto, 0,06 puntos; un 10 %, 0,6 puntos con la aproximación $\log(1 + p) \approx p$; duplicar el gasto, $6,08 \times \log 2 = 4,2$ puntos. Para variaciones grandes la aproximación falla: duplicar es $p = 1$, que daría 6,1 puntos, pero el logaritmo cambia en $\log 2 = 0,69$, y el coeficiente multiplica al cambio en el logaritmo, no a la fracción (lección 15).
2. El gasto es muy asimétrico y lo que cabe esperar que importe es la variación relativa: la misma cantidad de dinero pesa distinto para quien gasta 50 mil y para quien gasta un millón. El voto es un porcentaje acotado entre 0 y 100 y se lee bien en puntos. Con el gasto en niveles, el coeficiente serían puntos de voto por cada mil dólares, los mismos para el primer millar que para el último.
3. Contrasta que la suma de los dos coeficientes del gasto sea cero, es decir, que un aumento proporcional de los dos gastos deje el voto como estaba. No rechazar no prueba que la suma sea cero: datos como estos no serían raros con esa hipótesis, pero tampoco con cualquier suma entre $-1,59$ y $0,52$, que es lo que el intervalo añade al valor p .
4. Sí, porque los dos modelos tienen el mismo regresando, `voteA`, y el R^2 mide la fracción de su variación que el ajuste reproduce. Con el logaritmo del voto en uno de ellos dejarían de serlo: habría que pasar las previsiones a la misma variable y comparar correlaciones, como en la lección 15 y en la [actividad 2 \(html\)](#) de la práctica A de la sesión 25.
5. Mide la diferencia de voto entre un demócrata y un republicano con el mismo gasto, el mismo gasto del rival y la misma fuerza de partido en el distrito, en las elecciones de 1988: 3,8 puntos. No mide lo que un candidato ganaría cambiando de partido ni dice nada de otros años.
6. Supone que el coeficiente del gasto es un efecto causal: que si un candidato dado gastase el doble, su voto subiría 4,2 puntos. Para eso hace falta que lo que el modelo no recoge sea ortogonal a las funciones de los regresores, y el gasto de un candidato depende de sus posibilidades y de lo reñido del distrito. La tabla dice que los candidatos que gastan el doble obtienen, con lo demás fijo, 4,2 puntos más; no dice que gastar el doble los produzca. Y supone que el rival no responde: si los dos gastos se duplicasen, el intervalo de la Actividad 3 deja sin descartar desde 1,1 puntos menos para A hasta 0,4 más.