

## Replicación #1: Washington, *AER*, 2008

### Metodología Política Avanzada

#### Instrucciones:

- Cada grupo debe preparar: (a) un reporte en formato .pdf respondiendo las preguntas listadas más abajo; y (b) el código de R ó Stata con el que reproducir los resultados. Corresponde enviar ambos a [adrian.lucardi@itam.mx](mailto:adrian.lucardi@itam.mx) no después de las 12:00pm del viernes 12 de octubre. Tanto el trabajo como la calificación son grupales.
- La claridad y organización del reporte no es determinante pero será valorada. Salvo en caso de usar Markdown, knitr o similares, se recomienda enviar el código y las respuestas de forma separada.
- También se valora el cuidado en la presentación de tablas y gráficos, pero los únicos requisitos indispensables son (a) no usar más de tres decimales; y (b) no incluir asteriscos (\*) para indicar coeficientes significativos.

El texto a replicar es

Ebonya L. Washington. Female Socialization: How Daughters Affect Their Legislator Fathers' Voting on Women's Issues. *The American Economic Review*, 98(1):311–332, 2008.

Los datos y un archivo con la descripción de las variables están disponibles en la carpeta del curso.

## 1 Replicación

- (1) *Congresos*. La base original contiene información sobre cuatro Congresos estadounidenses (105<sup>to</sup>-108<sup>vo</sup>, de 1997-2004). Sin embargo, cada modelo se centra en los legisladores de un solo Congreso. Comienza creando 4 bases distintas, una para cada Congreso.
- (2) *Estadísticas descriptivas*.
  - (a) *Relación entre los puntajes*. Como variables dependientes, la autora emplea los puntajes de tres organizaciones: NOW, AAUW y NRLC. Las tres clasifican el voto de los legisladores según apoyen distintas causas relacionadas con los derechos de las mujeres. Dado que las tres organizaciones miran a los mismos legisladores durante el mismo período de tiempo, esperaríamos que sus medidas se parezcan mucho. Compruébalo (i) visualizando la relación de las tres medidas entre sí; y (ii) calculando la correlación entre ellas. ¿Dirías que las tres medidas capturan más o menos lo mismo?
  - (b) Haz dos gráficos replicando los paneles de la Figura 1. No es necesario reproducir las barras de la izquierda, que indican el promedio de todos los legisladores; alcanza con las de los demócratas y republicanos. Tampoco es necesario incluir el número de observaciones en cada grupo que aparece encima de “Democrats” o “Republicans.” (*Tip*: averigua cómo usar la función `geom_bar()` en `ggplot`, especialmente las opciones de `position="dodge"`, `stat="summary"` y `fun.y="mean"`.)

- (c) La Figura 1 presenta los puntajes promedio de la *National Organization for Women* (NOW) para cada legislador. ¿Qué pasa si reemplazas esos valores por los de la *American Association of University Women* (AAUW)?
  - (d) Repite el ejercicio anterior pero con los datos de otro Congreso, a elección. Nota que para los Congresos posteriores no hay datos de NOW.
  - (e) *Crédito extra*. Replica los valores de la Tabla 1, que muestra el promedio de las principales variables para el 105<sup>to</sup> Congreso. Averigua cómo hacerlo con `stargazer()`. Puedes tomarte algunas licencias: (i) no es necesario crear una tabla con 3 columnas, puedes crear 3 tablas (*Full Sample*, Demócratas, Republicanos); y (ii) no es necesario calcular los desvíos estándar de las variables dependientes.
- (3) *Tabla 2*.
- (a) Corre los modelos correspondientes y luego replica la Tabla 2. Deberías obtener exactamente los mismos resultados que en el texto original. Asegúrate de leer la nota al pie de la tabla para incluir todas las variables relevantes.
  - (b) ¿Qué pasa con el coeficiente de *Number of female children* si usas errores robustos o los agrupas por estado? Discute brevemente.
  - (c) Corre los mismos modelos pero controlando por % voto demócrata, región y número de hijos exclusivamente. Elige los errores estándar que te parezcan más apropiados. Comenta cómo (si) cambian los resultados.
  - (d) En la Tabla 2 hay un modelo distinto para cada Congreso. ¿Qué pasa si incluyes los datos de *todos* los Congresos simultáneamente? Compruébalo corriendo tres versiones distintas del modelo correspondiente: (i) con errores convencionales; (ii) con errores robustos; y (iii) con errores agrupados por legislador, para controlar por el hecho de que muchos legisladores sirvieron en múltiples Congresos. Comenta brevemente lo que encuentras.
- (4) *Tabla 3 (versión libre)*. En la Tabla 3, la autora examina en qué medida los resultados varían por género o afiliación partidaria. Para eso, corrió cuatro versiones de cada modelo incluido en la Tabla 2: el primero para el subconjunto de legisladores hombres, el segundo para el subconjunto de legisladoras mujeres, el tercero para los demócratas, y el cuarto para los republicanos.
- (a) Haz una versión más simple de la segunda y tercera columnas de la Tabla 3, incluyendo una interacción entre *ngirls* y *female*. ¿Qué indica el coeficiente de *ngirls*? ¿Cómo puedes calcular el efecto del número de hijas mujeres entre las legisladoras mujeres? ¿Qué tanto difieren tus resultados de los reportados por Washington?
  - (b) Repite el ejercicio anterior pero con la afiliación partidaria de los legisladores. ¿Qué indica el coeficiente de *ngirls*? ¿Cómo puedes calcular el efecto del número de hijas mujeres entre los legisladores republicanos? ¿Encuentras diferencias sustantivas respecto a los resultados reportados por Washington?
- (5) *Crédito extra: Tabla A1*. Replica la Tabla 1 del apéndice, que muestra que (para los miembros del 108<sup>o</sup> Congreso), el sexo del primer hijo determina el número de hijas mujeres, pero no el número total de hijos:

- (a) La información sobre el sexo del primer hijo de un subconjunto de 227 legisladores del 108<sup>vo</sup> Congreso se encuentra en el archivo `geno1d108.dta`. Debes averiguar cómo usar la función `merge()` para unir estos datos con los legisladores del 108<sup>vo</sup> Congreso en la base original.
- (b) Replica exactamente los resultados de la Tabla A1.
- (c) Repite el ejercicio pero quitando los controles, i.e. incluyendo únicamente la variable *First child female* como predictor. ¿Observas algún cambio sustantivo?

## 2 Afirmaciones

La autora ofrece una serie de afirmaciones para las cuales no ofrece evidencia directa. Sin embargo, como contamos con los mismos datos que la autora, podemos verificar la exactitud de las mismas.

- (1) *Efectos fijos por hijo*. La variable independiente de interés es el número de hijas mujeres, pero todos los modelos incluyen efectos fijos por el número de hijos con que cuenta un legislador (i.e., *dummies* indicando que un legislador tiene 1 hijo, 2, 3, 4, etc). La razón de esto es que “*Failure to include these child fixed effects would yield an estimate of  $\beta_1$  which combines both the impact of parenting an additional daughter and the impact of parenting an additional child.*” (p. 315).
  - (a) La afirmación anterior sugiere que, si bien estimar el efecto de tener una hija adicional no es problemático –al contrario, es lo que el *paper* pretende estudiar–, el efecto de tener un hijo adicional (de cualquier sexo) puede ser diferente, i.e. puede ser una *confounding variable*. Ello sería un problema porque, por ejemplo, los legisladores más conservadores y/o más religiosos podrían tener más hijos. Comprueba si es el caso corriendo los modelos correspondientes para examinar si los legisladores (i) republicanos; (ii) que votan de forma más conservadores; y/o (iii) que pertenecen a distintas religiones, tienen más hijos (hombres o mujeres) en promedio.
  - (b) Vuelve a correr los modelos de la Tabla 2 pero sin incluir los efectos fijos por el número de hijos. ¿Cambian los resultados?
- (2) *Legisladores sin hijos*. En la nota 13 (p. 315) leemos que “*Results are robust to the exclusion of members of Congress without children.*” Comprueba que efectivamente es el caso.
- (3) *Proporción de mujeres por partido político*. “*Republicans have a slightly smaller proportion of girls than their Democratic counterparts,*” pero “*this difference is not statistically significant.*” (p. 316 y nota 17). ¿Es verdad que la proporción de hijas mujeres es “*slightly smaller*” pero no significativo entre los legisladores republicanos? Demuéstralo.
- (4) *Puntajes de la NRLC*. En las Tablas 2 a 4, la variable dependiente son los puntajes de NOW o AAUW, pero en la página 322 leemos que “*In specifications akin to those in Table 2, I find that parenting daughters has a significant impact on NRLC scores for all four focal Congresses,*” y en la nota 30 encontramos los coeficientes correspondientes. Comprueba que efectivamente sea el caso. Recuerda cambiar el valor de los puntajes NRLC para que valores más altos indiquen una orientación más liberal.