

Microeconometría aplicada

Taller 2

Mauricio Romero

20 de septiembre de 2023

1. Instrucciones generales

- Pueden usar R, Stata o cualquier otro programa
- Deben entregar un archivo PDF: Un memo con lo que hicieron y lo que encontraron.
- Adicionalmente, deben entregar un archivo ZIP con los datos crudos (raw data), y todos los códigos (scripts) que usen.
- Yo debería poder replicar sus resultados descomprimiendo el archivo zip y cambiando (únicamente) el directorio principal al principio y corriendo el código.
- Envíen sus respuestas (PDF y ZIP) a: mtromero+microeconometria@itam.mx
- Van a ser evaluados no solo en la respuesta final. También tendré en cuenta el proceso que usen para llegar a sus respuestas (i.e., su código y los comentarios de ese código).
- Es muy importante que escriban su código (y los comentarios) de manera que sea fácil para otros (yo en este caso) seguir el código y entender lo que se hizo.
- Usar recursos adicionales (internet, libros, blogs, ChatGTP, etc.) está bien (y muchas veces es necesario), pero por favor menciones los recursos que usan y cítelos adecuadamente.
- La fecha de entrega es Octubre 5 de 2023 a las 12:00 (medio día)
- La tarea se puede hacer en grupos (del tamaño que quieran)

2. Problema 1 - Experimento en educación

Suponga que el gobierno está considerando las opciones de dar apoyos monetarios a las escuelas o incentivos monetarios individuales a los profesores con base en el rendimiento que hayan tenido sus alumnos en una prueba estandarizada o los dos juntos. Antes de expandir esta política a todas las escuelas públicas, deciden hacer un piloto y que sea evaluado mediante un experimento aleatorio. Uds. están encargados de diseñar el experimento. Suponga que Ud. tiene cierto margen de elegir cuantas escuelas pueden estar en este piloto (entre 100 y 200).

1. Primero debemos diseñar el experimento

- a) Discuta los méritos de asignar maestros aleatoriamente a que se les den incentivos basados en su rendimiento, sin importar la escuela (i.e., asignar el tratamiento a nivel individual). [Hint: ¿Se cumpliría SUTVA?]

- b) Discuta los méritos de asignar escuelas aleatoriamente al tratamiento de que a sus maestros se les den incentivos basados en su rendimiento o no (i.e., asignar el tratamiento a nivel escuela). [Hint: ¿Se pierde algo en términos de poder?]
 - c) ¿Qué diseño preferiría, y por qué?
 - d) Supón ahora que el presupuesto que te había otorgado inicialmente el gobierno federal es aumentado y ahora puedes muestrear más escuelas. ¿Cómo afectaría esto en términos de poder?
 - e) Ahora supón que para que la intervención sea considerada por el gobierno para su planeación de política pública te pide que el efecto de la intervención sea mayor o igual a un valor de referencia establecido por la comunidad académica experta en el tema θ , que refleja si una intervención es relevante. ¿Qué parámetro buscarías calibrar en tus cálculos de poder para cumplir con la condición del gobierno federal? ¿De qué otros parámetros depende y de qué forma lo hace?
 - f) En la evaluación de impacto de esta intervención consideras algunas variables de respuesta para evaluar el efecto del tratamiento. ¿Cómo afecta una alta varianza poblacional de la variable de respuesta en el poder estadístico del experimento?
 - g) ¿De qué manera tendrías que establecer el tamaño de los grupos de tratamiento y control para maximizar el poder estadístico, todo lo demás constante?
2. Suponga que se hizo el segundo diseño (i.e., se asignaron escuelas aleatoriamente). Antes de hacer cualquier análisis, ¿Ud. cree que comparar los resultados de los estudiantes en las escuelas de tratamiento con las de control unos años después del tratamiento nos va a indicar la efectividad de esta política pública? [Pista: Piense en las decisiones de largo plazo de quien se vuelve maestro]
 3. Suponga que efectivamente comparando los resultados de los estudiantes en diferentes escuelas puede estudiar el efecto del tratamiento. Suponga que Ud. tiene una base de datos que tiene resultados a nivel estudiante para diferentes escuelas: ¿Que regresión utilizaría para medir el efecto del tratamiento (indique el nivel de cada variable, como manejaría los errores estándar, si va a incluir controles y cuales, etc.)?
 4. Un experimento muy similar a este se llevó a cabo en Tanzania ver <https://academic.oup.com/qje/article/134/3/1627/5479257?login=false>. En esta intervención se asignaron aleatoriamente a una muestra de 350 escuelas tres tipos de tratamiento: (i) transferencias monetarias incondicionales, (ii) incentivos monetarios a profesores basados en el puntaje que sus alumnos llegaron a conseguir en una prueba estandarizada y (iii) la combinación de los dos tratamientos previamente mencionados. Los datos de este experimento para su replicación se hallan aquí <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/AKVGQ7>. Identifique las bases de datos a nivel escuela, a nivel estudiante y la base con los puntajes de la prueba estandarizada para el 2014. Estime el efecto de tratamiento promedio (ATE) y el p-valor con Mínimos Cuadrados Ordinarios del puntaje promedio de la prueba. También, estime el efecto de tratamiento y el p-valor asociado del puntaje promedio de la prueba usando inferencia de aleatorización. Explique la especificación que decidió utilizar para la estimación del efecto en los dos casos así como los controles que usará. Reporte sus resultados. Recuerde que el diseño es estratificado por lo que debe tomar en cuenta esto de manera explícita en su especificación. La variable del estrato se llama *DistID*.

3. Problema 2 - Diferencias en Diferencias (DiD)

Para el siguiente ejercicio se usaran datos con múltiples períodos y fechas de tratamiento <https://github.com/Rxmirxz13/taller2.git>. Se examinarán los efectos de la expansión del Medicaid en la cobertura de seguros en Estados Unidos usando datos públicos de la Sociedad Americana de Cáncer (ACS por sus siglas en inglés). Si requiere un mayor contexto del ejemplo ver la investigación realizada por Carey et al. llamada *The Impact of Insurance Expansions on the Already Insured: The Affordable Care Act and Medicare* <https://www.aeaweb.org/articles?id=10.1257/app.20190176>.

Los datos que se tendrán que usar contienen una base de datos panel a nivel estado correspondiente a la cobertura de seguros de gastos médicos y la expansión del programa estadounidense federal Medicaid. La base de datos contiene identificadores por estado, año de la observación, proporción de personas adultas sin hijos, con ingreso bajo y que están aseguradas, el año en el que el estado decidió expandir el programa Medicaid bajo la *Affordable Care Act* (el dato es faltante si el estado nunca expandió la ayuda), y una variable de pesos por estado.

3.1. *Non-staggered* DiD

Para lograr una configuración adecuada para un diseño DiD sin variación en el tiempo de tratamiento realice el siguiente procesamiento de la base de datos proporcionada.

- Tire los años estrictamente mayores a 2015.
 - Tire los tres estados que expandieron la ayuda en 2015.
1. Estime los coeficientes de una regresión *Two way fixed effect* y reporte y explique sus resultados junto con la especificación que usó.
 2. Realice un *event study* mostrando un *plot* de *leads & lags* con los coeficientes que obtuviste.

3.2. *Staggered* DiD

Ahora, use la base de datos proporcionada completa (sin ningún procesamiento).

1. Estime el $ATT_{g,t}$ para cada grupo tratado (recordemos que tenemos un grupo por cada fecha de tratamiento distinta en la base) en cada período. Reporte sus resultados y la especificación usada. [Hint: ayúdese de la paquetería `did` función `att_gt` en R o `csdid` función `csdid` en STATA. Si usará la función `att_gt` del paquete `did`, para las observaciones que no tienen fecha de tratamiento (estados que no expandieron Medicaid) impute el año de tratamiento a un número grande, por ejemplo 4000, para que el comando de `att_gt` incorpore correctamente las unidades que nunca fueron tratadas como controles]. *Bonus*: reporte para este ejercicio las versiones cuando considera a las unidades del grupo control como que nunca fueron tratadas y como que todavía no son tratadas. Los comandos citados tienen este parámetro para que sea *tuneado*.
2. Para garantizar que entiende cómo se construyen las estimaciones realizadas en el inciso anterior, verifique a mano las estimaciones de los coeficientes del $ATT(2015, 2013)$, $ATT(2013, 2013)$ y $ATT(2016, 2014)$. [Hint: Concentrémonos en la estimación de $ATT(2015, 2013)$, es decir, el coeficiente del año 2013 del grupo que fue tratado en el 2015. Construya una variable dicotómica D , que indique con 1 si la observación fue tratada en el 2015 y con 0 otro caso. Calcule la media condicional para la variable de proporción de personas adultas sin hijos, con ingreso bajo y que están aseguradas para los años 2012 y 2013, es decir, calcule 4 medias condicionales, una para cada combinación de año y la variable dicotómica D . Manualmente, compute el DiD 2x2 entre 2012 y 2013 y entre $D = 0$ y $D = 1$].

3. Realice un *event study* agregando dinámicamente los coeficientes que obtuviste en el primer inciso con `att_gt`. Muestra el *plot* de *leads & lags* producto de la agregación por grupos. [Hint: lea los archivos *help* de la paquetería que esté usando]. Interprete sus resultados.
4. Estime la siguiente regresión con MCO:

$$Y_{it} = \alpha_i + \lambda_t + D_{it}\beta + \varepsilon_{it}$$

donde D_{it} es una indicadora de si la unidad i fue tratada en el tiempo t . Clusterise sus errores estándar al mismo nivel que ha venido haciendo. ¿Cómo se compara la estimación de esta $\hat{\beta}$ con la dinámica que obtuvo en el inciso anterior?