Universidad Torcuato Di Tella
Maestrías en Economía y Econometría
2022

Microeconometría II
Práctica 1
Modelo de Resultados Potenciales

---

1. Resultados potenciales, sesgo de selección y el médico perfecto

Considere el ejemplo hipotético simple del Cuadro 1. Este ejemplo involucra una población de once pacientes, cada uno de los cuales está infectado con COVID-19. Hay dos tratamientos: **ventiladores y reposo en cama**. El Cuadro 1 muestra los resultados potenciales de cada paciente en términos de años de supervivencia después del tratamiento con cada tratamiento. Los valores de resultado más grandes corresponden a mejores resultados de salud.

Cuadro 1: Resultados Potenciales

Paciente	$Y_T(u)$	$Y_C(u)$	TE	S	Y
1	1	10			
2	1	5			
3	1	4			
4	5	6			
5	5	1			
6	9	7			
7	6	8			
8	7	10			
9	8	2			
10	9	6			
11	10	7			

1. Calcule el efecto del tratamiento para cada paciente (columna TE)
2. Cuál es el efecto tratamiento promedio (ATE) para ventiladores (Y_T) comparado con reposo en cama (Y_C)? Qué tipo de intervención es más efectiva en promedio?
3. Supongamos que el "médico perfecto" conoce los resultados potenciales de cada paciente y, como resultado, elige el mejor tratamiento para cada paciente. Si asigna a cada paciente el tratamiento más beneficioso para ese paciente, qué pacientes recibirán ventiladores y cuáles recibirán reposo en cama (columna S)? Complete la última columna en función de lo que elija el médico perfecto.
4. Cuál es el efecto tratamiento promedio para ventiladores comparado con reposo en cama en el caso del médico perfecto (calculado con los datos observados)?
5. Cuál es la diferencia entre el ATE del médico perfecto y el ATE calculado en el punto b)? Explique con sus palabras por qué el resultado es diferente en ambos casos y justifique cual de los resultados sería el correcto.

2. Simulación de Monte Carlo

En este ejercicio se busca mostrar, utilizando simulaciones de Monte Carlo, que bajo el supuesto de independencia, la diferencia de medias para los individuos que recibieron el tratamiento y aquellos que no lo recibieron identifica el ATE. Considere los siguientes datos.

1. Calcule el ATE.
2. Genere una variable que, para cada observación, obtenga una realización de una normal estándar. Ordene las observaciones de menor a mayor en función del valor de esta normal estándar.

Cuadro 2: Resultados Potenciales

Paciente	$Y_T(u)$	$Y_C(u)$
1	7	1
2	5	6
3	5	1
4	7	8
5	4	2
6	10	1
7	1	10
8	5	6
9	3	7
10	9	8

3. Genere una variable d de otorgamiento del tratamiento que valga 1 para las primeras 5 observaciones ordenadas y que valga 0 para las restantes.
4. Compute la diferencia de medias en los promedios muestrales.
5. Repita el procedimiento anterior 10000 veces y reporte la media de las diferencias de medias de cada simulación.

3. Efectos de Tratamiento y Heterogeneidad

Sean Y_T, Y_C los resultados potenciales y D la variable de otorgamiento del tratamiento. Se define el el efecto promedio de tratamiento como

$$ATT = E[Y_T - Y_C], \tag{1}$$

el efecto promedio de tratamiento sobre los tratados como

$$ATT = E[Y_T - Y_C | D = 1] \tag{2}$$

y el efecto promedio de tratamiento sobre los no tratados como

$$ATT = E[Y_T - Y_C | D = 0]. \tag{3}$$

En este ejercicio se propone una descomposición del ATE diferente que la analizada en clase. La misma sigue el capítulo 4 de *Causal Inference: The Mixtape* de Scott Cunnningham.

1. Muestre que $ATT = \pi ATT + (1-\pi)ATU$, y describa los ponderadores π.
2. En base al libro mecionado arriba, interprete que

$$\bar{y}_1 - \bar{y}_0 = ATE + \text{Sesgo de Selección } + \text{Sesgo por heterogeneidad en tratamiento.}$$

en términos de efectos de tratamiento.
3. Inicialice una muestra con 100 observaciones. Genere resultados potenciales de no recibir el tratamiento como

$$Y_0 \sim \mathcal{N}(100, 30)$$

.
4. Genere ahora un efecto de tratamiento constante e igual a 20 para todos, es decir, $TE_i = 20$ para todo $i = 1, ..., n$. Genere una variable aleatoria normal estándar. Genere una variable de tratamiento D_i igual a 1 para aquellas observaciones que poseen un valor positivo en la variable aleatoria normal. Genere la variable Y observada como $Y = DY_1 + (1-D)Y_0$. Compute la diferencia de medias y el test t. Luego, calculo ATE, ATT y ATU.
5. Repita el inciso anterior pero utilice que $TE \sim \mathcal{N}(20, 10)$.
6. Repita el inciso anterior pero genere una variable aleatoria normal estándar $rand$ y genere $W = 1\{rand > 0\}$. Genere el tratamiento como $TE \sim \mathcal{N}(20, 10)$ si $W = 1$ y $TE \sim \mathcal{N}(10, 10)$ si $W = 0$. Utilice W como la variable utilizada para asignar el tratamiento.
7. Repita el inciso anterior pero ahora luego de generar los efectos de tratamiento en función de W asigne el tratamiento aleatoriamente como en el primer inciso.
8. Comente las conclusiones obtenidas con respecto a la heterogeneidad del tratamiento.

Universidad Torcuato Di Tella
Maestrías en Economía y Econometría
2022

Microeconometría II
Práctica 2
Matching

1. Exact Matching

(Basado en Cunningham)Este ejercicio propone utilizar el procedimiento de matching sobre una variable para pensar en términos simples de dónde provienen los estimadores de Matching. Hoy en día es conocimiento común que fumar aumenta la tasa de mortalidad. Sin embargo, esta afirmación no proviene de datos experimentales.

1. Considere los datos de mortalidad y condición de fumador de la Tabla 1. ¿Qué resultado da la diferencia de medias entre grupos de fumadores en términos de tasa de mortalidad? ¿Qué pide el supuesto de independencia? En particular, comente qué espera que pase con otras variables observables.

2. Considere la edad de las personas. ¿Qué puede decir sobre el supuesto de independencia?

3. Estime el efecto correcto.

Table 5.1: Death rates per 1,000 person-years (Cochran 1968)

Smoking group	Canada	UK	US
Non-smokers	20,2	11,3	13,5
Cigarettes	20,5	14,1	13,5
Cigars/pipes	35,5	20,7	17,4

Table 5.2: Mean ages, years (Cochran 1968).

Smoking group	Canada	UK	US
Non-smokers	54,9	49,1	57,0
Cigarettes	50,5	49,8	53,2
Cigars/pipes	65,9	55,7	59,7

Table 5.2: Mean ages, years (Cochran 1968).

Smoking group	Canada	UK	US
Non-smokers	54,9	49,1	57,0
Cigarettes	50,5	49,8	53,2
Cigars/pipes	65,9	55,7	59,7

Solution:

1. La diferencia de medias simple sugiere que fumar pipas y habanos está asociado a una mayor tasa de mortalidad. El supuesto de independencia sugiere que la media los resultados potenciales es iguales entre grupos,

$$\begin{aligned} E\left[Y^1 \mid \text{ Cigarette }\right] &= E\left[Y^1 \mid \text{ Pipe }\right] = E\left[Y^1 \mid \text{ Cigar }\right] \\ E\left[Y^0 \mid \text{ Cigarette }\right] &= E\left[Y^0 \mid \text{ Pipe }\right] = E\left[Y^0 \mid \text{ Cigar }\right]. \end{aligned} \tag{1}$$

   Es decir, si se cumple, debería suceder que la mortalidad si se deciden a fumar no debería variar entre no fumadores, fumadores de cigarrillos y fumadores de habanos y pipas, tanto si fumaran como si no fumaran. También debería ser cierto que que las características observables entre fumadores no deberían ser, en promedio, diferentes. Es decir, los grupos deberían estar balanceados.

2. La diferencia en edades por grupo (variable observada) está relacionada al no cumplimiento del supuesto de independencia.

2. Teorema de Rosenbaum y Rubin

Demuestre el siguiente teorema. Sean $Y(0), Y(1)$ resultados potenciales, X un tratamiento binario, W un vectr de características observables. Suponiendo que vale el supuesto de *unconfoundedness*, se define $e(W)$ como la probabilidad de recibir el tratamiento en función de variables observables. Luego $(Y(0), Y(1) \perp X | e(W))$. Como corolario,

$$ATE = E\left[E[Y^{obs}|X=1, e(W)] - E[Y^{obs}|X=0, e(W)]\right]. \tag{2}$$

Solution: Por la definición de independencia, es suficiente probar que

$$P(X=1|Y(0), Y(1), e(W)) = P(X=1|e(W)) \tag{3}$$

Como X es binaria,

$$\begin{aligned} P(X=1|Y(0),Y(1),e(W)) &= E(X|Y(0),Y(1),e(W)) \\ &= E\left(E(X|Y(0),Y(1),e(W),W)|Y(0),Y(1),e(W)\right) \qquad \text{LIE} \\ &= E\left(E(X|Y(0),Y(1),W)|Y(0),Y(1),e(W)\right) \\ &= E\left(E(X|W)|Y(0),Y(1),e(W)\right) \qquad \text{unconfoundedness} \\ &= E\left(e(W)|Y(0),Y(1),e(W)\right) \qquad e: W \to [0,1] \\ &= e(W) \end{aligned}$$

Por lo tanto,

$$P(X=1|Y(0),Y(1),e(W)) = P(X=1|e(W)) = P(X=1|W) = e(W) \tag{4}$$

El resultado implica que los resultados potenciales son independientes del tratamiento habiendo condicionado en la probabilidad de recibir el tratamiento, $e(W)$ (el llamado *propensity score*). Como tenemos el resultado demostrado al nivel de la probabilidad, esto implica el resultado deseado para las esperanzas, y podemos calcular el ATE habiendo condicionado en el puntaje de propensión en vez de en todas las características W.

$$ATE = E\left[E[Y^{obs}|X=1, e(W)] - E[Y^{obs}|X=0, e(W)]\right]. \tag{5}$$

Notar que

1. Si hay 2 unidades en tratamiento y control con igual probabilidad de recibir el tratamiento, en promedio tienen los mismos resultados potenciales. Es decir, si antes *unconfoundedness* pedía

$$E\left[Y(d)|X=1, W\right] = E\left[Y(d)|X=0, W\right], \quad d \in \{0,1\}, \tag{6}$$

ehora podemos relajar la condición a

$$E\left[Y(d)|X=1, e(W)\right] = E\left[Y(d)|X=0, e(W)\right], \quad d \in \{0,1\}. \tag{7}$$

2. El teorema de Rosenbaum y Rubin sugiere un procedimiento aplicado para estimar los efectos promedios de trataminto: calcular la probabilidad $e(W)$ de recibir el tratamiento en función de las características W, y hacer un matching en esta variable que resume las anteriores. Esto, además, simplifica el problema de dimensionalidad de emparejar muchas variables con algunas potencialmente continuas.

3. El supuesto de *unconfoundedness* es necesario para el teorema de Rosenbaum y Rubin. No se puede reemplazar fallas en este supuesto y esperar que se solucione haciendo Propensity Score Matching.

3. Propensity Score Matching

Para solucionar el problema de la dimensionalidad utilizando matching sobre múltiples variables se puede reducir el conjunto de variables en una sola, la llamada *propensity score*, puntaje de propensión o probabilidad de recibir el tratamiento. En este ejercicio se propone realizar una evaluación del impacto de fumar durante el embarazo sobre el peso de los bebés en base a datos observacionales. Utilizando la base de datos `cattaneo2.dta` que utiliza Cattaneo (XX), responda las siguientes preguntas.

1. Compruebe si los grupos de control y de tratamiento están balanceados.

2. ¿Cuál es la diferencia de medias simple?

3. Estime la probabilidad de recibir el tratamiento con un modelo Logit. Utilice como variables explicativas `mmarried deadkids nprenatal monthslb prenatal fbaby alcohol`. Defina una sección de soporte común.

4. Estime el efecto promedio de tratamiento sobre los tratados utilizando Propensity Score Matching implementando nearest-neighbor matching, radius matching, kernel matching y stratification matching.

5. Estime el efecto promedio de tratamiento utilizando nearest-neighbour matching.

6. Estime el efecto promedio de tratamiento utilizando ajuste por regresión.

7. Estime el efecto promedio de tratamiento utilizando *inverse probability weighting*.

8. Estime el efecto promedio de tratamiento utilizando el estimador doblemente robusto.

4. Estimador de matching sin características oservables

Suponga que vale el supuesto de *unconfoundedness*. En este ejercicio se propone computar la expresión analítica del estimador de PSM cuando la probabilidad de recibir el tratamiento no depende de características oservables. Utilice el corolario del ejercicio 2.

Solution: Si vale *unconfoundedness* sin condicionar en otras variables explicativas, entonces $X \perp Y$. El ATE es

$$ATE = E\left[E[Y^{obs}|X=1] - E[Y^{obs}|X=0]\right]. \tag{8}$$

Definimos

$$\mu_X = E(Y^{obs}|X=x) \tag{9}$$

Como estimador, tomamos el análogo muestral, porque converge en probabilidad al parámetro de interés. Entonces,

$$\hat{\mu}_X = \frac{n^{-1}\sum_{i=1}^n Y_i^{obs}\mathbb{I}\left\{X_i = x\right\}}{n^{-1}\sum_{i=1}^n \mathbb{I}\left\{X_i = x\right\}} \tag{10}$$

Luego, el estimador del ATE es la diferencia de medias.

$$\widehat{ATE} = \hat{\mu}_1 - \hat{\mu}_0.$$

Utilizando la expresión del estimador de las medias,

$$\begin{aligned}
\widehat{ATE} &= \hat{\mu}_1 - \hat{\mu}_0 \\
&= \frac{n^{-1}\sum_{i=1}^n Y_i^{obs}x_i}{n^{-1}\sum_{i=1}^n x_i} - \frac{n^{-1}\sum_{i=1}^n Y_i^{obs}(1-x_i)}{n^{-1}\sum_{i=1}^n (1-x_i)} \\
&= \frac{n^{-1}\sum_{i=1}^n Y_i^{obs}x_i}{\overline{x}} - \frac{n^{-1}\sum_{i=1}^n Y_i^{obs}(1-x_i)}{(1-\overline{x})} \\
&= n^1 \sum_{i=1}^n Y_i^{obs}\left(\frac{x_i}{\overline{x}} - \frac{1-x_i}{1-\overline{x}}\right) \\
&= n^{-1}\sum_{i=1}^n Y_i^{obs}\frac{x_i - \overline{x}}{\overline{x}(1-\overline{x})}
\end{aligned}$$

Universidad Torcuato Di Tella
Maestrías en Economía y Econometría
2022

Microeconometría II
Práctica 3
Diferencias-en-Diferencias

---

1. Evaluación de Impacto con DiD

Una práctica común en la evaluación de un programa cuando se tienen datos de panel para dos períodos es la siguiente: sea y_{it} el resultado observado para i en el período t. En $t = 1$ nadie está en el programa. En $t = 2$ algunos están en el grupo de control y otros en el grupo de tratamiento. Sea $prog_{it}$una variable que vale 1 si el individuo i está en el grupo de tratamiento en el período t; y cero en caso contrario. Note que $prog_{i1} = 0$ para todo i. Se puede plantear el siguiente modelo:

$$y_{it} = \theta_1 + \theta_2 d2_t + \delta_1 prog_{it} + c_i + u_{it}$$

Con $E(u_{it}/prog_{i2}, c_i) = 0$. En el que $d2_t$ es una variable dummy que vale 1 si $t = 2$, cero si $t = 1$; c_i es el efecto no observado. Usando el método de primeras diferencias, muestre que $\hat{\theta}_2 = \overline{\Delta y_c}$ y $\hat{\delta}_1 = \overline{\Delta y}_t - \overline{\Delta y}_c$, donde $\overline{\Delta y}_c$ es el cambio promedio en y a lo largo de los dos períodos para el grupo con $prog_{i2} = 0$, y $\overline{\Delta y}_t$ es el cambio promedio en y a lo largo de los dos períodos para el grupo con $prog_{i2} = 1$.

Solution:
Tenemos el siguiente modelo

$$y_{it} = \theta_1 + \theta_2 d2t + \delta_1 prog_{it} + c_i + u_{it}$$

donde y_{it} es el resultado observado para i en el período t, $prog_{it}$ vale 1 si el individuo i está en el grupo de tratamiento en t y $d2t$ es una variable dicotómica que vale 1 si $t = 2$. c_i es el efecto no observado por individuo, que asumimos constante en el tiempo.
El modelo en $t = 1$ es

$$y_{i1} = \theta_1 + \theta_2 * 0 + \delta_1 * 0 + c_i + u_{i1} = \theta_1 + c_i + u_{i1}$$

El modelo en $t = 2$ es

$$y_{i2} = \theta_1 + \theta_2 * 1 + \delta_1 * prog_{i2} + c_i + u_{i2} = \theta_1 + \theta_2 + \delta_1 * prog_{i2} + c_i + u_{i2}$$

Tomando primeras diferencias

$$y_{i2} - y_{i1} = (\theta_1 + \theta_2 + \delta_1 * prog_{i2} + c_i + u_{i2}) - (\theta_1 + c_i + u_{i1}) = \theta_2 + \delta_1 * prog_{i2} + u_{i2} - u_{i1}$$

Reordenando

$$\Delta y_i = \theta_2 + \delta_1 * prog_{i2} + \Delta u_i$$

Los estimadores OLS de θ_2 y δ_1 son

$$\delta_1^{OLS} = \frac{Cov\,(prog_{i2}, \Delta y_i)}{Var\,(prog_{i2})} \qquad \theta_2^{OLS} = \overline{\Delta y} - \delta_1^{OLS}\overline{prog}_{i2}$$

Sea N_1 la cantidad de veces que $prog_{i2} = 1$. Computemos los distintos elementos. Utilizaremos algunos atajos de la Práctica 1. En particular, del Ejercicio 0, Ejercicio 7.

$$
\begin{aligned}
Cov\left(prog_{i2}, \Delta y_i\right) &= \sum_{i=1}^{N} prog_{i2}\Delta y_i - N\overline{prog}_{i2}\overline{\Delta y} \\
&= \sum_{prog_{i2}=1} \Delta y_i - N\left(\frac{\sum_{i=1}^{N} prog_{i2}}{N}\right)\overline{\Delta y} \\
&= N_1\overline{\Delta y_t} - N_1\overline{\Delta y} \\
&= N_1\left(\frac{\sum_{prog_{i2}=1}\Delta y_i}{N_1} - \frac{\sum_{i=1}^{N}\Delta y_i}{N}\right) \\
&= N_1\left(\frac{N\sum_{prog_{i2}=1}\Delta y_i - N_1\sum_{i=1}^{N}\Delta y_i}{N_1 N}\right) \\
&= \frac{(N-N_1)\sum_{prog_{i2}=1} y_i - N_1\sum_{prog_{i2}=0}\Delta y_i}{N} \\
&= \frac{(N-N_1)N_1\overline{\Delta y_t} - (N-N_1)N_1\overline{\Delta y_c}}{N} \\
&= \left(1-\frac{N_1}{N}\right)N_1\left(\overline{\Delta y_t} - \overline{\Delta y_c}\right)
\end{aligned}
$$

$$
\begin{aligned}
Var\left(prog_{i2}\right) &= \sum_{i=1}^{N} prog_{i2}^2 - N\overline{prog}_{i2}^2 \\
&= N_1 - N\left(\frac{N_1}{N}\right)^2 = N_1 - \frac{N_1^2}{N} \\
&= N_1\left(1-\frac{N_1}{N}\right)
\end{aligned}
$$

Luego

$$
\delta_1^{OLS} = \overline{\Delta y_t} - \overline{\Delta y_c}
$$

Computamos θ_2^{OLS}

$$
\begin{aligned}
\theta_2^{OLS} &= \overline{\Delta y} - \delta_1^{OLS}\overline{prog}_{i2} \\
&= \overline{\Delta y} - \left(\overline{\Delta y_t} - \overline{\Delta y_c}\right)\frac{N_1}{N} \\
&= \frac{\sum_{i=1}^{N}\Delta y_i}{N} - \frac{N_1}{N}\left(\frac{\sum_{prog_{i2}=1}\Delta y_i}{N_1} - \frac{\sum_{prog_{i2}=0}\Delta y_i}{N-N_1}\right) \\
&= \frac{\sum_{i=1}^{N}\Delta y_i}{N} + \frac{\sum_{prog_{i2}=1}\Delta y_i}{N} - \frac{N_1\sum_{prog_{i2}=0}\Delta y_i}{N(N-N_1)} \\
&= \frac{(N-N_1)\sum_{i=1}^{N}\Delta y_i - (N-N_1)\sum_{prog_{i2}=1}\Delta y_i + N_1\sum_{prog_i 2=0}\Delta y_i}{N(N-N_1)} \\
&= \frac{N\sum_{prog_{i2}=0}\Delta y_i}{N(N-N_1)} \\
&= \frac{\sum_{prog_{i2}=0}\Delta y_i}{N(N-N_1)} \\
&= \overline{\Delta y_c}
\end{aligned}
$$

2. DiD Simple en Stata

1. Abra la base `Panel101.dta` y genere las siguientes variables de resultado, de tiempo y de tratamiento. Los "países" tratados son los países 5 a 7 y el tratamiento se otorgó en 1994.

$$\begin{aligned} Y &= y/1000000000 \\ time &= \mathbb{I}\{year \geq 1994\} \\ treated &= \mathbb{I}\{country > 4\} \end{aligned}$$

2. Compute e interprete el estimador de diferencias en diferencias utilizando una regresión lineal.
3. Compute e interprete el estimador de diferencias en diferencias utilizando una especificación con efectos fijos.
4. Compute e interprete el estimador de diferencias en diferencias utilizando el paquete de Stata `diff`.

3. Card & Krueger (1994)

Este ejercicio se basa en el artículo *Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania.*

1. ¿Qué efecto intentan estimar los autores en el artículo?
2. ¿Cuál es la estrategia de identificación?
3. Utilice el archivo `CardKrueger1994.dta`. Utilizando `diff`, compute el estimador de diferencias-en-diferencias.
4. Repita el inciso (c) utilizando errores estándar de *bootstrap*.
5. Repita el inciso (c) utilizando la cadena de restaurantes como variables explicativas.

4. `didregress`

Hasta ahora estimamos utilizando la especificación

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 T_{it} + \beta_3 (D \times T)_{it} + \mathbf{z}_{it}\theta + u_{it} \tag{1}$$

O en el caso de datos longitudinales, ampliamos la ecuación con efectos fijos. En Stata, usamos los comandos `regress` y/o `xtreg`. Stata 17 incluyó nuevos comandos para estimar modelos con la siguiente forma

$$Y_{ist} = \gamma_s + \gamma_t + \mathbf{z}_{ist}\beta + \delta D_{st} + u_{ist}$$

con el comando `didregress` o incluyendo efectos fijos por individuo en el caso de datos longitudinales con el comando `didregress`

1. Explique en qué difieren estos comandos con respecto al *setup* usual de DiD.
2. ¿Puede replicar las regresiones de los ejercicios anteriores con estos comandos?
3. Un proveedor de salud está interesado en estudiar el efecto de un nuevo procedimiento de ingreso hospitalario en la satisfacción de los pacientes. El proveedor dispone de datos mensuales de pacientes de enero a julio. El nuevo procedimiento de admisiones fue implementado en abril por hospitales que estaban bajo nueva administración. De los 46 hospitales del estudio, 18 implementaron el nuevo procedimiento.

   El proveedor de salud utilizará una regresión DID para analizar el efecto del nuevo procedimiento de admisiones en los hospitales que participaron en el programa. El resultado de interés es la satisfacción del paciente, `satis`, que se registra como un promedio de las respuestas a un conjunto de cuatro preguntas realizadas a los pacientes. `satis` puede tomar valores entre 0 y 10, donde 10 es el mayor nivel de satisfacción posible y 0 es la decepción total. La variable procedimiento marca las observaciones tratadas; es 1 si una persona encuestada ingresó al hospital utilizando el nuevo procedimiento después de marzo y 0 en caso contrario.

   Los datos están en la base `hospdd.dta`. Evalúe el imapcto del nuevo procedimiento sobre la satisfacción de los pacientes.
4. ¿Cómo interpreta el coeficiente obtenido? ¿Se cumple el supuesto de tendencias paralelas?
5. Comente sobre los errores estándar utilizados y estudie las distintas opciones que el comando tiene pre-configuradas para usar. ¿Hay diferencias en la inferencia?

Universidad Torcuato Di Tella
Maestrías en Economía y Econometría
2022

Microeconometría II
Práctica 4
Extensiones de Diferencias-en-Diferencias

---

1. Static Two-Way Fixed Effects

Una forma de extender el *framework* de diferencias-en-diferencias cuando hay más períodos temporales disponibles es utilizar efectos fijos temporales adicionalmente a los efectos fijos por individuo. Es decir, si el resultado de interés es Y y el tratamiento está definido por D, entonces se puede estimar la ecuación

$$Y_{it} = \alpha_i + \phi_t + D_{it}\beta_{post} + \epsilon_{it}. \tag{1}$$

α_i son efectos fijos por individuo, phi_t son efectos fijos temporales y

1. ¿Cómo debe definirse D_{it} para que esto sea la generalización de la ecuación de DiD 2x2?
2. Escriba como queda la variable si $T = 3$ $N = 3$ con la unidad 1 sin tratar, la unidad 2 tratada en el período 1, la unidad 2 tratada en el período 2.
3. ¿Cuál es la expresión del estimador de β^{post}?
4. ¿Cuándo podría este estimador ser bueno para recuperar el efecto promedio de tratamiento sobre los tratados? ¿Qué problemas podría tener esta especificación?
5. Utilizando la base de datos `organ_donations.dta` de Kessler & Roth (2014) estime el coeficiente β^{post} de esta especificación.

2. Dynamic Two-Way Fixed Effects

Con la especificación anterior se estima un único efecto de tratamiento.Sin embargo, si se posee información de varios períodos, ésta podría utilizarse para evaluar, por un lado, cómo cambian los efectos en el tiempo, y por otro, evaluar cómo se comportaba la variable de interés previo del otorgamiento del tratamiento. La especificación dinámica de TWFE es

$$Y_{it} = \alpha_i + \phi_t + \sum_{\substack{r \neq 0 \\ -\underline{T} \leqslant r \leqslant \bar{T}}} 1\left[R_{it} = r\right]\beta_r + \epsilon_{it}.$$

1. Escriba como quedan las variables si $T = 3$ $N = 3$ con la unidad 1 sin tratar, la unidad 2 tratada en el período 1, la unidad 2 tratada en el período 2.
2. Compute los efectos de tratamiento dinámicos β_r. Notar que se normaliza $\beta_0 = 0$. (¿Por qué?')
3. Grafique los efectos de tratamiento pre y post otorgamiento junto con sus intervalos de confianza.
4. Utilice los datos de los períodos anteriores al tratamiento para hacer un *"test de placebo"*:
   - Utilice solo los datos que llegaron antes de que el tratamiento entrara en vigor.
   - Elija un período de tratamiento falso.
   - Calcule el mismo modelo de diferencias en diferencias que planeaba usar (por ejemplo), pero cree la variable igual a 1 si está en el grupo tratado y después de la fecha de tratamiento falso que eligió.
   - Si encuentra un "efecto" para esa fecha de tratamiento donde realmente no debería haberlo, eso es evidencia de que hay algo mal con su diseño, lo que puede implicar una violación de tendencias paralelas.

3. Callaway & Sant'Anna y `csdid`

Una solución a los problemas de TWFE es la que proponen Callaway & Sant'Anna (2020). Ellos proponen computar todos los ATT válidos y ponderarlos adecuadamente. En Stata esto se puede hacer con el comando `csdid`. Utilizando la base de datos `mpdta.dta`, se busca estimar el impacto de una suba del salario mínimo en el empleo joven.

1. Estime todos los $ATT(g,t)$ sin variables explicativas.

2. Evalúe si es plausible el efecto de tendencias paralelas en base a las tendencias previas al otorgamiento del tratamiento. ¿Considera que puede haber habido factores que hayan afectado la evolución del empleo en todos los estados tratados que **no** se deba al otorgamiento del tratamiento? Reflexione acerca del rol de la forma funcional de las variables (por ejemplo, en niveles vs. en logaritmos).

3. Compute el efecto agregado simple, el efecto agregado por grupos, el efecto agregado por período y el efecto agregado por períodos tras el otorgamiento del tratamiento.

4. Repita los incisos anteriores condicionando en la variable de población.

5. Hasta ahora se utilizaron los nunca tratados como grupo de control Repita los incisos anteriores utilizando el grupo de los no tratados todavía. ¿Observa cambios?

6. Compare con los resultados que surgen de hacer TWFE estático y dinámico.

4. Enfoque de Wooldridge y `jwdid`

Frente a las críticas de la literatura a la presunta incapacidad de la especificación de TWFE para estimar los efectos promedio de tratamiento sobre los tratados, Wooldridge (2021) propone que el problema no es intrínseco a que la ecuación de estimación sea lineal con efectos fijos ni a que se estime con los métodos tradicionales de datos de panel.[1] En particular, con T períodos temporales y el tratamiento otorgado en cada período desde el período q hasta el último se propone estimar la siguiente ecuación:

$$Y_{it} = \alpha + \lambda_q d_{iq} + \cdots + \lambda_T d_{iT} + \sum_{r=q}^{T} \sum_{s=r}^{T} \tau_{rs} \left(d_{ir} \cdot 1 \{t = s\} \right) + \theta_t + \epsilon_{it}$$

donde τ_{rs} representa el ATT en el período s para el grupo r, α es una constante, d_{ir} es una dummy que vale 1 para el grupo tratado en r, con $r = q, \ldots, T$. Replique el ejercicio anterior con el comando `jwdid`.[2]

[1] En particular, el artículo plantea una equivalencia entre una especificación de efectos fijos y una especificación two-way Mundlak, dejando el caso de DiD como una aplicación.

[2] Debe agregar los archivos .ado a su directorio. No va a poder replicar el test de tendencias previas ni los gráficos, al menos no al momento en el que se escribe esto.

Universidad Torcuato Di Tella
Maestrías en Economía y Econometría
2022

Microeconometría II
Práctica 5
Variables instrumentales

---

1. Estimador de Wald

Suponga un modelo de regresión simple

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

donde x_i es potencialmente endógena. Además, suponga que que el instrumento, z_i es una variables binaria. Muestre que el estimador IV en este caso es

$$\beta_1^{IV} = \frac{\overline{y}_1 - \overline{y}_0}{\overline{x}_1 - \overline{x}_0}$$

donde $\overline{y}_1, \overline{x}_1 \left(\overline{y}_0, \overline{x}_0\right)$ representan las medias cuando $z = 1 \left(z = 0\right)$.

2. Estimador de Wald con datos simulados

En este ejercicio se propone extender la simulación del Problem Set 1 a un marco en el que la asignación del tratamiento y quienes resultan tratados no son iguales.

1. Inicialice una muestra con 100 observaciones. Genere resultados potenciales de no recibir el tratamiento como

$$Y_0 \sim \mathcal{N}(100, 30)$$

.

2. Genere ahora un efecto de tratamiento constante e igual a 20 para todos, es decir, $TE_i = 20$ para todo $i = 1, ..., n$. Genere una variable aleatoria normal estándar. Genere una variable de tratamiento D_i igual a 1 para aquellas observaciones que poseen un valor positivo en la variable aleatoria normal.
3. Genere una variable aleatoria uniforme en el intervalo $[0, 1]$. Con ella, genere variables que indiquen el tipo de individuo. Utilice: *always taker* si la variable es menor a 0.25, *never taker* si la variable está entre 0.25 y 0.5, *defier* si la variable está entre 0.5 y 0.75 y *complier* si la variable es mayor a 0.75. *G*enere la variable de si los individuos toman el tratamiento o no dependiendo del grupo en el que están.
4. Genere la variable Y observada como $Y = DY_1 + (1 - D)Y_0$.
5. Estime el LATE y compare con el ATE.

3. Galiani & Schargrodsky (2010)

Lea el artículo *"Property rights for the poor: Effects of land titling"* de Galiani & Schargrodsky.

1. ¿Qué efectos intentan estimar en el paper?
2. ¿Cuál es la estrategia de identificación? ¿Por qué no funciona la diferencia de medias simple?
3. Replique las resultados del paper.