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?