

Diferencias en diferencias

Oficina de Planificación Estratégica y Efectividad en el Desarrollo
Semana de Diseño de Evaluación de Impacto

De conformidad con la Política de Acceso a Información, esta presentación está sujeta a divulgación pública.

Copyright © 2018 Banco Interamericano de Desarrollo. Esta obra está bajo una licencia Creative Commons IGO 3.0 Reconocimiento-No Comercial-Sin Obra Derivada (CC-IGO BY-NC-ND 3.0 IGO) (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) y puede ser reproducida para cualquier uso no-comercial otorgando crédito al BID. No se permiten obras derivadas.

Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la CNUDMI. El uso del nombre del BID para cualquier fin que no sea para la atribución y el uso del logotipo del BID, estará sujeta a un acuerdo de licencia por separado y no está autorizado como parte de esta licencia CC-IGO.

Notar que el enlace URL incluye términos y condicionales adicionales de esta licencia.

Las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.

El contenido de esta presentación está basado parcialmente en Gertler et al (2011).



Organización

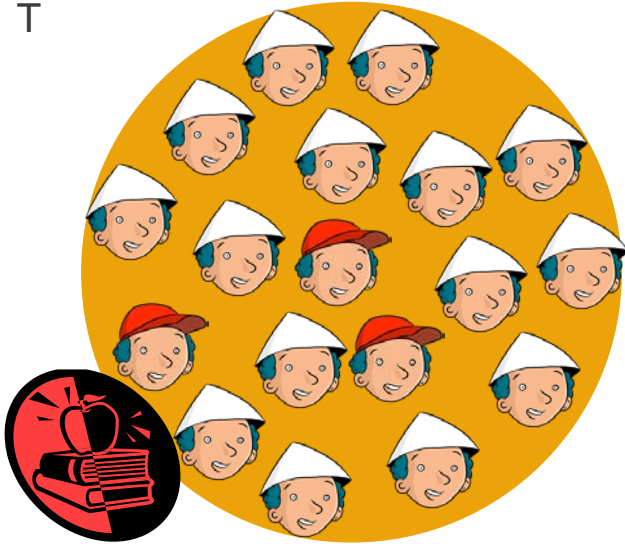
- I. Intuición
- II. Definición
- III. Ejercicios

I. Intuición



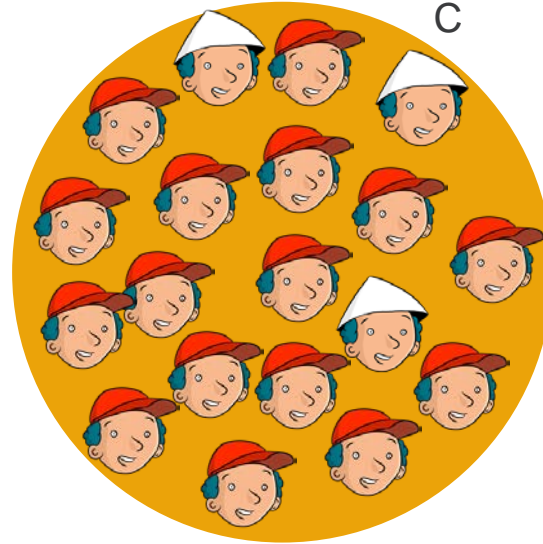
Problema de evaluación

T



$$E[Y_i|T] = 630$$

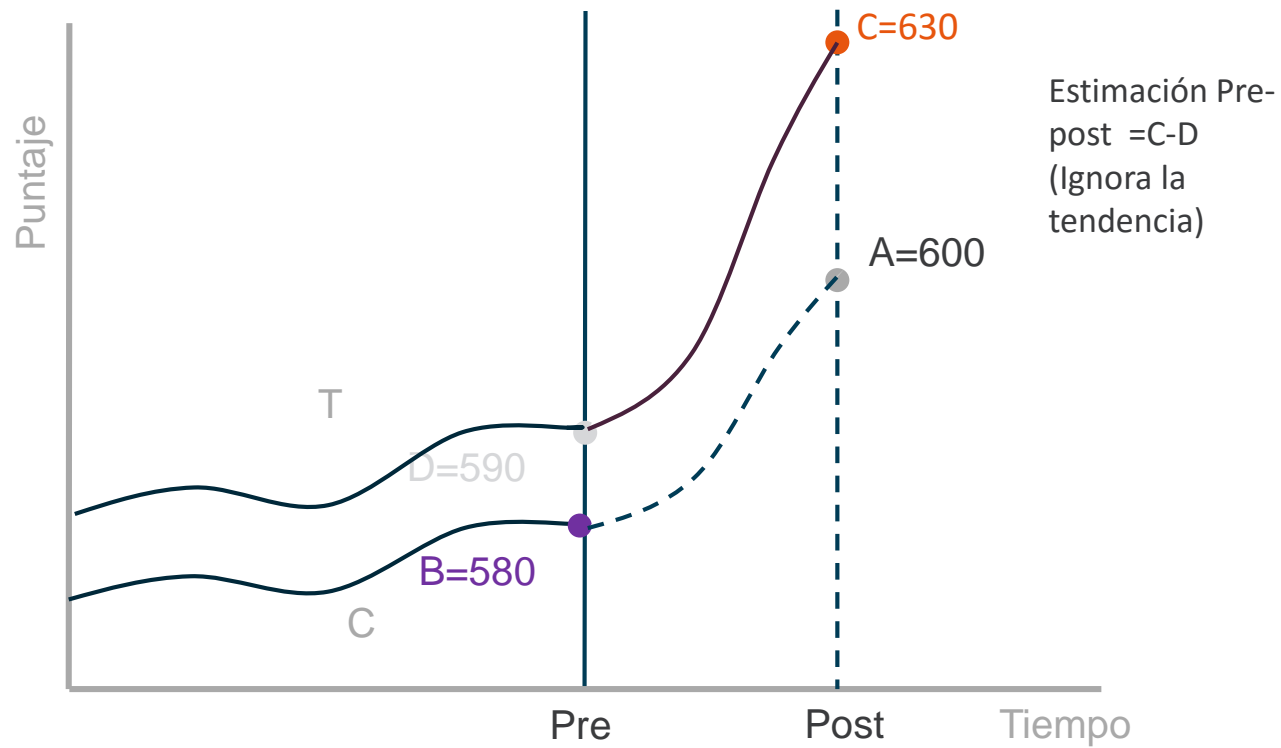
C

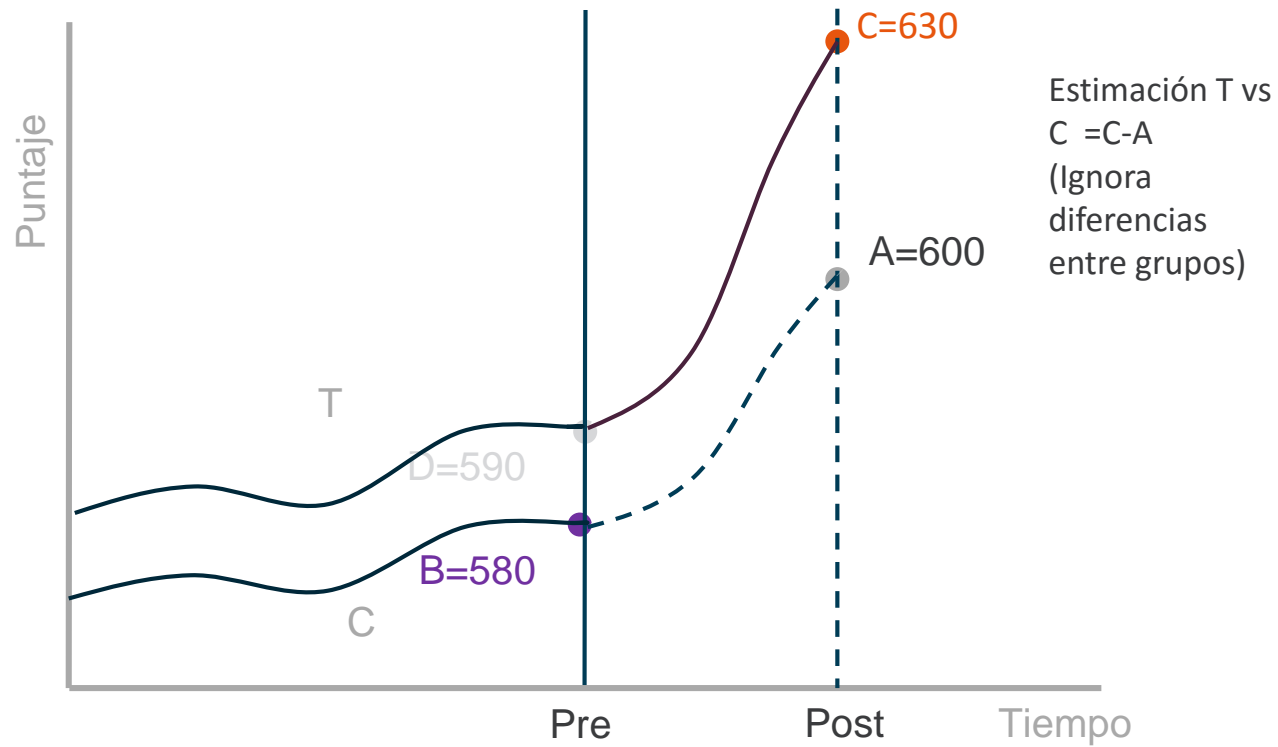


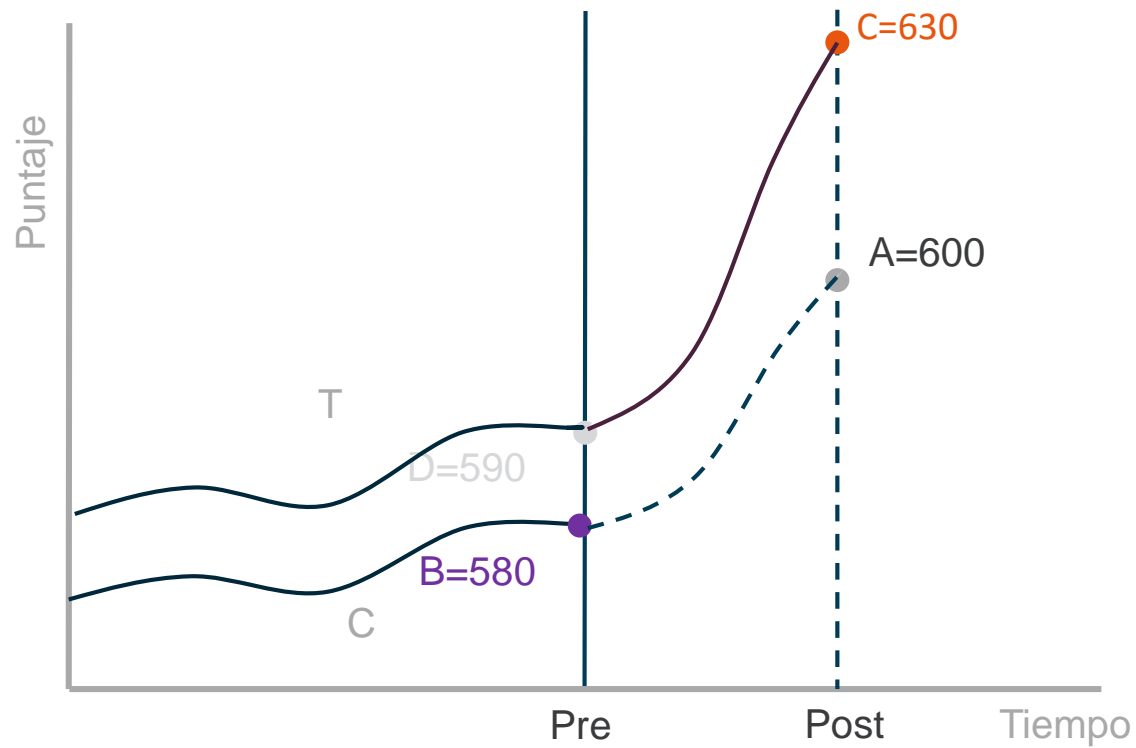
$$\hat{E}[Y_i|C] = 600$$

$$\text{IMPACTO} = 630 - 600 = 30$$

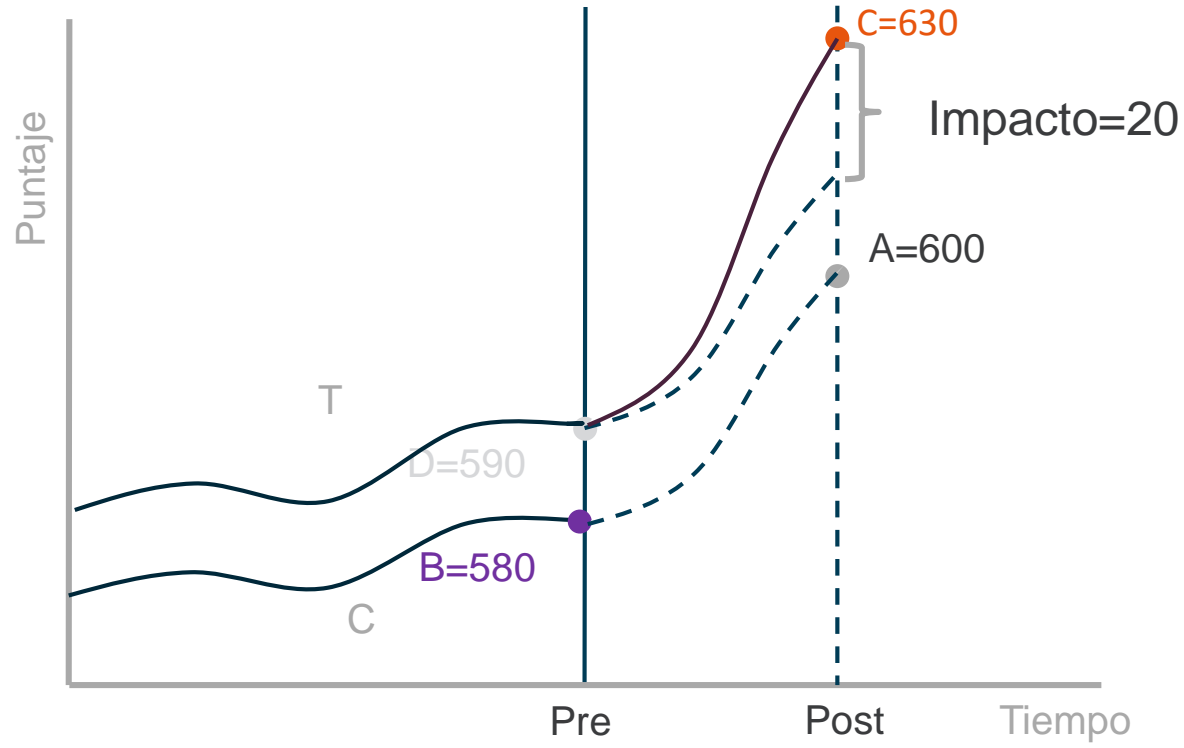
puntos

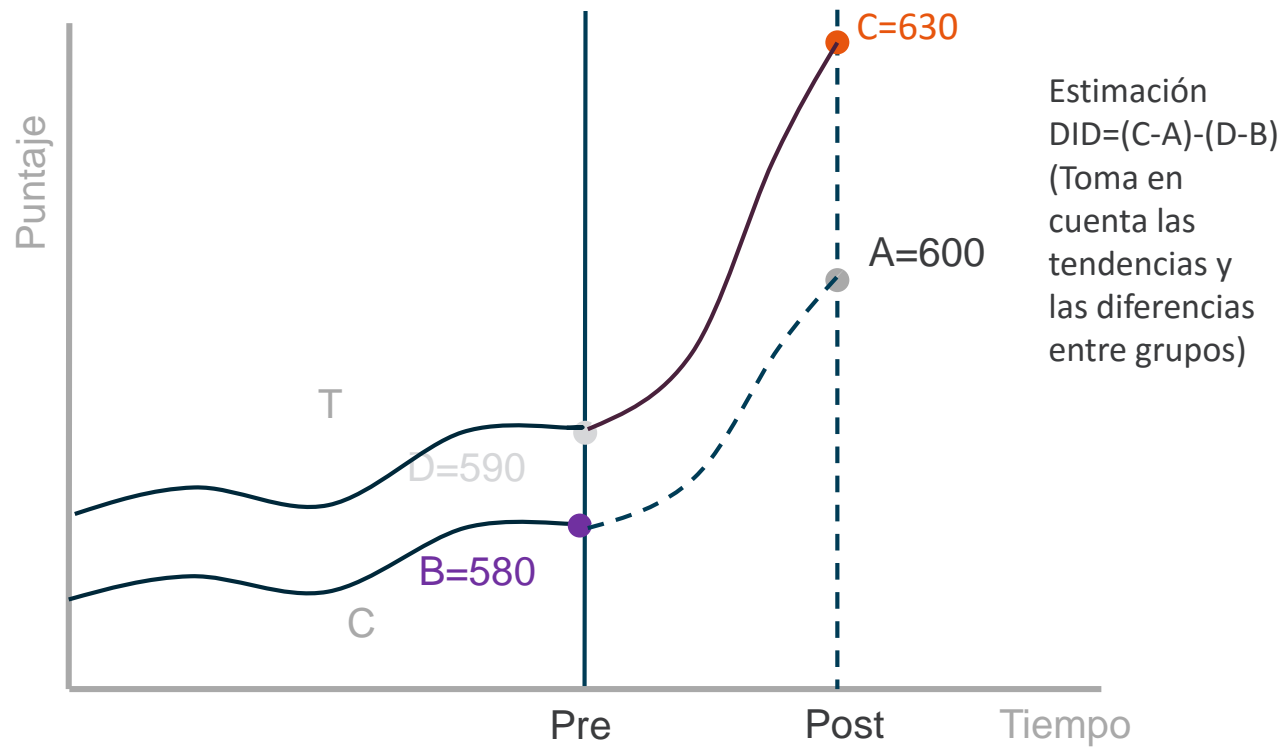




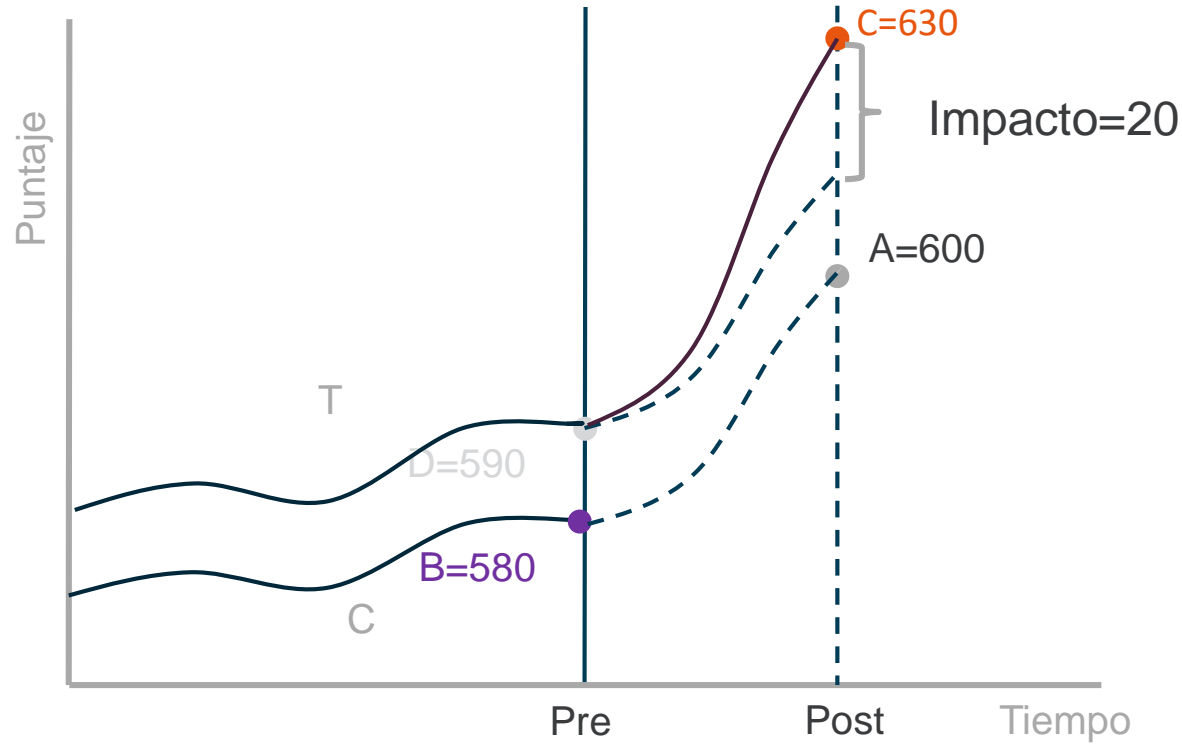


$$\text{Impacto} = (C-A) - (D-B)$$



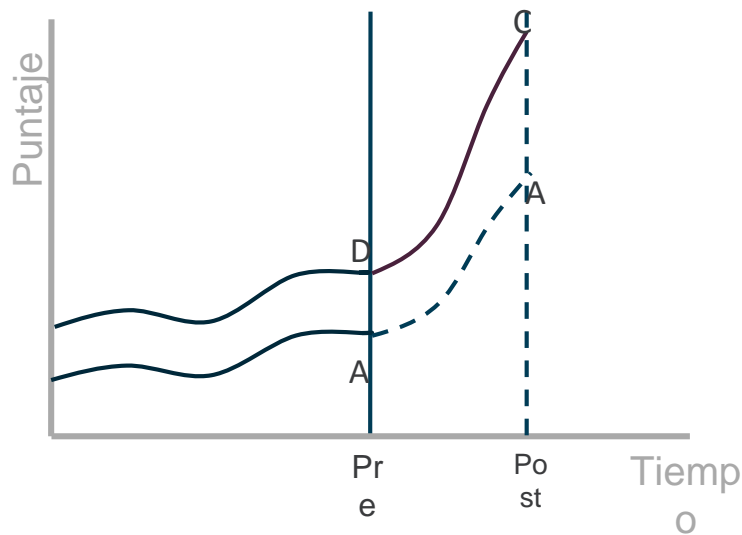


$$\text{Impacto} = (C-A) - (D-B) = (C-D) - (A-B)$$



Notación alternativa

$$Y_i = \alpha + \phi T_i + \eta D2_i + \beta T_i D2_i + \epsilon_i$$



$$C = E[Y|T = 1, D2 = 1] = \alpha + \phi + \eta + \beta$$

$$A = E[Y|T = 0, D2 = 1] = \alpha + \eta$$

$$D = E[Y|T = 1, D2 = 0] = \alpha + \phi$$

$$B = E[Y|T = 0, D2 = 0] = \alpha$$

Donde $D2 = \begin{cases} 1 & \text{si post} \\ 0 & \text{si no} \end{cases}$

$T = \begin{cases} 1 & \text{si tratamiento} \\ 0 & \text{if no} \end{cases}$

II. Definición

El método de *diferencias en diferencias* compara los cambios en los resultados a lo largo del tiempo entre unidades inscritas en un programa (el grupo de tratamiento) y unidades que no lo están (el grupo de comparación). Esto permite corregir cualquier diferencia entre los grupos de tratamiento y comparación que sea constante a lo largo del tiempo. (Gertler et al. 2017)

Tips prácticos para DID

- » Muestra evidencia de que las tendencias no son diferentes antes del tratamiento (obligado!)
 - Prueba de diferencias en tendencias antes del T
 - Cuando hay más de dos periodos, haz pruebas para ver si se puede predecir el tratamiento
 - Encuentra evidencia de que los grupos tampoco difieren en niveles (útil, pero no necesario)
- » Restrinje la muestra a grupos que sean similares y donde te preocupa que hay diferencias en tendencias.
 - Combina con pareamiento
 - Combina con DID en otras dimensiones
 - Combina con otros métodos
- » Cuando hay más de dos periodos haz pruebas placebo
 - Pretende que los datos post-intervención no existen y revisa si encuentras efectos en periodos antes de la intervención

Resumen

El supuesto clave para DID

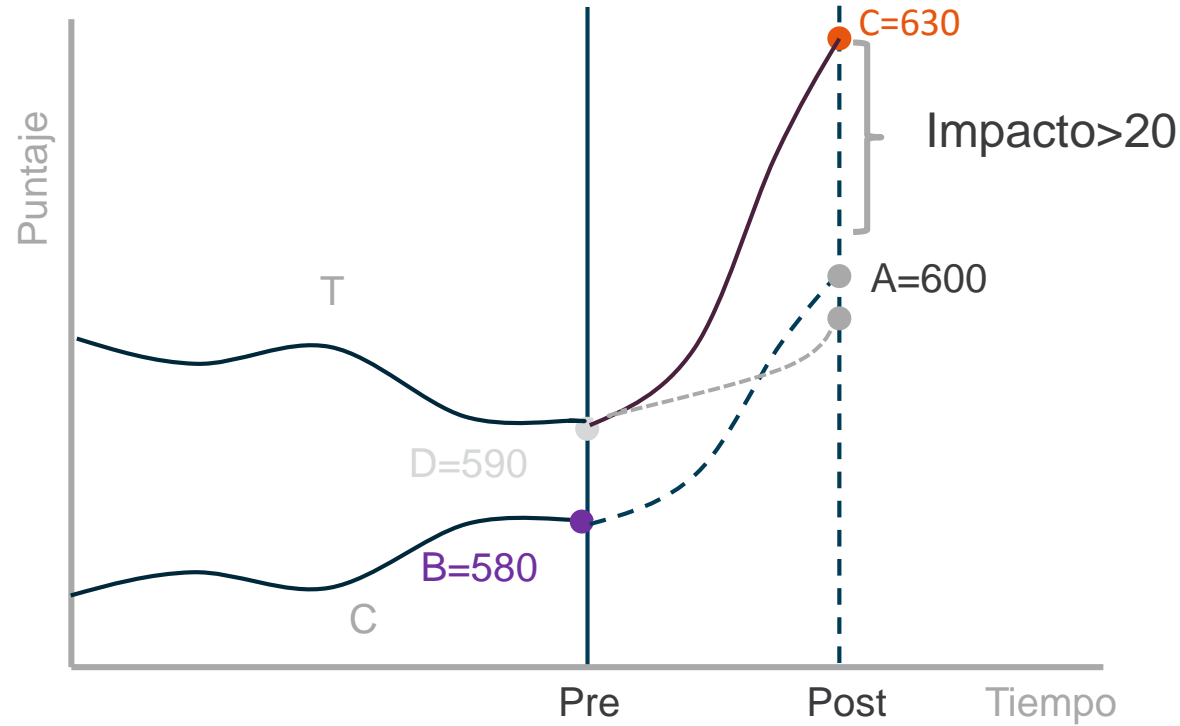
Las tendencias de los grupos de T y C en ausencia del tratamiento hubieran sido iguales en ausencia del tratamiento.

$$T_i, \Delta Z_{it} \perp \Delta \epsilon_{it}$$

Requiere muchas observaciones para poder estimar correctamente la matriz de varianza-covarianza (conglomerados)



Supuesto clave: Tendencias similares antes de la intervención



Resumen

- **DID:** cambio pre-post entre un grupo participante y otro no-participante
- Asume que en la ausencia de programa, **tendencias** entre participantes y no-participantes serian **iguales**

III. Ejercicio

Referencias

Bertrand, M., Duflo, M. and Mullainathan, S. (2004) "How Much Should We Trust Differences-in-Differences Estimates?"; Quarterly Journal of Economics, 119(1), pp. 249-75.

Cameron, C. and Trivedi, P. (2005) Microeconometrics. Methods and Applications, Cambridge University Press. Chapter 21.

Duflo, E. (2001) "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment". The American Economic Review. Vol. 91, No. 4. September, pp. 795-813.

Gertler P., Martinez, M., Premand, P., Rawlings, L. and Vermeersch, C. (2011) Impact Evaluation in Practice. Washington DC: The World Bank. Chapter 6.

Reto de evaluación

$$\begin{aligned}\beta &= E[Y_i^T | T] - E[Y_i^C | C] \\ &= E[Y_i^T | T] - E[Y_i^C | T] - E[Y_i^C | C] + E[Y_i^C | T] \\ &= \underbrace{E[(Y_i^T - Y_i^C) | T]}_{\text{Efecto por tratamiento}} + \underbrace{E[Y_i^C | T] - E[Y_i^C | C]}_{\text{Sesgo de selección}}\end{aligned}$$

Supongamos que hay datos un periodo ...Cómo nos puede ayudar esta información?

$$Y_i = a + bT_i + cX_i + e_i$$

$$ATE = \mathbb{E}[Y_i|T_i = 1] - \mathbb{E}[Y_i|T_i = 0]$$

$$C = \mathbb{E}[Y_i|T_i = 1] = a + b + c\mathbb{E}[X_i|T_i = 1] + \mathbb{E}[e_i|T_i = 1]$$

$$A = \mathbb{E}[Y_i|T_i = 0] = a + 0 + c\mathbb{E}[X_i|T_i = 0] + \mathbb{E}[e_i|T_i = 0]$$

$$ATE = b + c(\mathbb{E}[X_i|T_i = 1] - \mathbb{E}[X_i|T_i = 0])$$

Supongamos tenemos un periodo antes de la intervención

$$D = \mathbb{E}[Y_{i0}|T_i = 1] = a + b(0) + c\mathbb{E}[X_{i0}|T_{i1} = 1]$$

$$B = \mathbb{E}[Y_{i0}|T_i = 0] = a + 0 + c\mathbb{E}[X_{i0}|T_{i1} = 0]$$

$$\mathbb{E}[Y_{i0}|T_i = 1] - \mathbb{E}[Y_{i0}|T_i = 0] = c(\mathbb{E}[X_{i0}|T_{i1} = 1] - \mathbb{E}[X_{i0}|T_{i1} = 0])$$

Y ahora, DID:

$$\mathbb{E}[Y_i|T_i = 1] - \mathbb{E}[Y_i|T_i = 0] - (\mathbb{E}[Y_{i0}|T_i = 1] - \mathbb{E}[Y_{i0}|T_i = 0])$$

$$= b + c(\mathbb{E}[X_i|T_i = 1] - \mathbb{E}[X_i|T_i = 0]) - c(\mathbb{E}[X_{i0}|T_{i1} = 1] - \mathbb{E}[X_{i0}|T_{i1} = 0])$$

$$= b$$

Sea Z notación para controles

$$Y_{it} = \alpha + \phi T_{it} + \eta D2_{it} + \beta T_{it} D2_{it} + \gamma Z_{it} + \epsilon_{it}$$

$$Y_{it-1} = \alpha + \phi T_{it-1} + \eta D2_{it-1} + \beta T_{it-1} D2_{it-1} + \gamma Z_{it-1} + \epsilon_{it-1}$$

$$\Delta Y_{it} = \eta + \beta T_{it} + \Delta Z_{it} + \Delta \epsilon_{it}$$

$$\text{Recuerda } \Delta T_{it} = T_{i1} - T_{i0} = T_{i1} - 0 = T_{i1}$$

Estimado insesgado si

$$T_i, \Delta Z_{it} \perp \Delta \epsilon_{it}$$

Estimador eficiente si

$$V[\varepsilon_{it}] = \sigma^2 I$$

Pero lo mas probable es que

$$\varepsilon_{it} = \rho \varepsilon_{it-1} + u_{it}$$

Así que supongamos

$$COV[\varepsilon_{it}, \varepsilon_{is}] \neq 0$$

$$COV[\varepsilon_{it}, \varepsilon_{js}] = 0$$

Sea $\mathbf{x}=[T, D2, TD2]$ y \mathbf{x} denota controles, entonces

Se necesita un número grande de “clusters” (>50) para estimar la matriz de varianza-covarianza

$$\hat{V}[\hat{\beta}_{OLS}] = (\mathbf{x}'\mathbf{x})^{-1}(\sum_{i=1}^N \mathbf{u}'_i \mathbf{u}_i)(\mathbf{x}'\mathbf{x})^{-1}$$

$$\mathbf{u}_i = \sum_{t=1}^T \varepsilon_{it} \mathbf{x}_{it}$$

