Capítulo 5

Experimentos naturales o cuasi experimentos

¿Por qué usar experimentos naturales?

- Dado que los experimentos controlados presentan algunas dificultades, como:
  - Altos costos.
  - Conflictos éticos.
  - Corta duración.

Existe un marco no experimental que produce una asignación entre tratamiento y control con características similares a las que se obtendrían en un experimento aleatorio controlado.

Experimentos naturales o cuasi experimentos.

Ejemplo 1

- Supongamos que debemos analizar el efecto de una caída en el ingreso del hogar en el bienestar de los niños.
- Puede ocurrir que aquellos hogares para los que cae el ingreso sean sistemáticamente distintos de los hogares en los que el ingreso no cae (p. ej., los primeros son más inestables que los segundos).

Contenido

Ejemplo 1

- Suponga que en este país pasó un huracán.
- Los sitios arrasados por el huracán están esparcidos de manera aleatoria.
- Los damnificados por el huracán experimentaron una caída en el ingreso *presumiblemente* aleatoria y, por lo tanto, no está correlacionada con las características del hogar.
- Entonces, el huracán produce un experimento natural.

Salir

Ejemplo 2. *Canasta* "Hallazgo de petróleo"

- Suponga que algunos municipios del país encontraron petróleo de manera inesperada.

↓

Aumento en los ingresos municipales por regalías.

- Ahora el gobierno expide un decreto según el cual el municipio con un crecimiento de sus ingresos mayor a *x%* debe destinar un punto porcentual de éste a programas de infancia temprana, en particular a *Canasta.*

Ejemplo 2. *Canasta* "Experimento natural"

- En este caso, no se exigirá a los beneficiarios que apliquen o asistan a control médico, sino que la oficina administradora del programa se encarga de entregarlo.

- El hallazgo de petróleo no está correlacionado con:

  1. La decisión de participar en el programa.
  2. Los indicadores nutricionales.

- **El experimento natural**: Compara municipios que encontraron petróleo con los que no lo hicieron.

Tipos de experimentos naturales

1. Experimentos naturales en los que una variación aleatoria exógena (evento fortuito) genera una asignación totalmente aleatoria del tratamiento.

2. Experimentos naturales en los que una variación aleatoria exógena determina parcialmente la asignación del tratamiento.

Tipo 1

Cuando la asignación es completamente aleatoria.

El efecto se puede estimar por:

- Modelo de diferencias.
- Modelo de diferencias ampliado.
- Modelo de diferencias en diferencias (éste también aplica en el tipo 2).

Tipo 2: Modelo de diferencias en diferencias (Dif en Dif)

En el caso de experimentos naturales:

La aleatorización generada por el evento fortuito generalmente no es perfecta.

→

Sí pueden existir diferencias sistemáticas entre el grupo de control y de tratamiento.

Dif en Dif es una forma de controlar por estas diferencias.

Modelo de diferencias en diferencias

Definamos:

	Tratamiento	Control
$t = 1$	$Y_1 \mid D = 1$	$Y_1 \mid D = 0$
$t = 2$	$Y_2 \mid D = 1$	$Y_2 \mid D = 0$

$t = 1$: Período anterior a la implementación del tratamiento.

$t = 2$: Período posterior a la implementación del tratamiento.

Modelo de diferencias en diferencias

- ¿Qué estima?

$$\tau_{dif-en-dif} =$$

$$[E(Y_2|D=1) - E(Y_1|D=1)] \quad (1)$$

$$-[E(Y_2|D=0) - E(Y_1|D=0)] \quad (2)$$

(1) Cambio esperado entre el período anterior y el posterior a la implementación del tratamiento en el grupo de tratamiento.

(2) Cambio esperado entre el período anterior y el posterior a la implementación en el tratamiento en el grupo de control.

Análogo muestral

$$\hat{\tau}_{dif-en-dif} =$$

$$[(\bar{Y}_2|D=1)-(\bar{Y}_1|D=1)]$$

$$-[(\bar{Y}_2|D=0)-(\bar{Y}_1|D=0)]$$

Si el cambio temporal (longitudinal) lo defino como ΔY:

$$\hat{\tau}_{dif-en-dif} = (\Delta\bar{Y}|D=1)-(\Delta\bar{Y}|D=0)$$

Análogo muestral

Reescribiendo:

$$\hat{\tau}_{dif-en-dif} =$$

$$[(\bar{Y}_2|D = 1) - (\bar{Y}_2|D = 0)] \quad (1)$$

$$-[(\bar{Y}_1|D = 1) - (\bar{Y}_1|D = 0)] \quad (2)$$

(1) Diferencia promedio entre el grupo de tratamiento y el de control en el período *t = 2*.

(2) Diferencia promedio entre el grupo de tratamiento y el de control en el período *t = 1*.

¿Cuándo usar el estimador de Dif en Dif?

¿Cuándo el estimador Dif en Dif es más eficiente que el de diferencias?

Cuando algunos determinantes no observables de la variable de interés son persistentes en el tiempo.

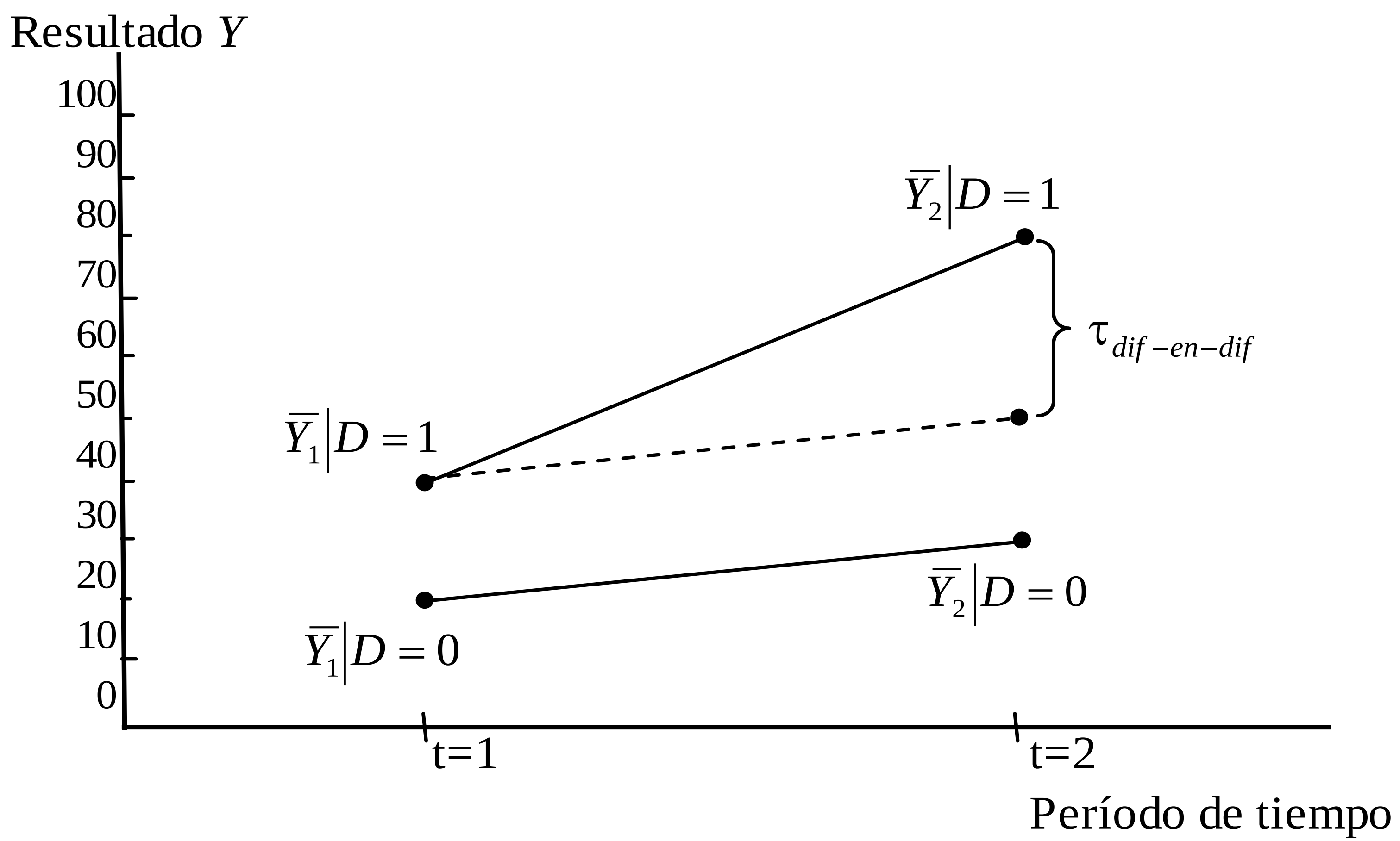
La eficiencia del estimador de Dif en Dif depende de la fracción de la varianza de Y que esté explicada por factores no observados.

¿Cuándo usar el estimador de Dif en Dif?

Cuando se observan diferencias preexistentes entre el grupo de control y el de tratamiento. Generalmente en presencia de un experimento natural que genera asignación parcialmente aleatoria.

Si el tratamiento está correlacionado con el nivel inicial de *Y* antes de la asignación del tratamiento, el estimador de diferencias es sesgado; pero el de Dif en Dif no lo es.

Ejemplo gráfico



Fuente: Stock y Watson (2003), página 387, gráfica 11.1

Ejemplo 1. STAR Descripción

- Objetivo: Evaluar el efecto de pertenecer a una clase pequeña en el aprendizaje de los estudiantes.
- Duración: Cuatro años, en la década de los ochenta.
- Tipos de tratamiento
  - Clases normales → 22-25 estudiantes.
  - Clases normales con asistente → 22-25 estudiantes.
  - Clases pequeñas → 13-17 estudiantes.

Ejemplo 1. STAR Descripción

- Variable de interés: Puntajes en los exámenes estandarizados de lectura y matemáticas (SAT), realizados anualmente.
- Asignación ideal: Aleatoria para estudiantes y profesores, sin reasignación desde kínder a tercero.
- Problemas

  10% de reasignaciones

  - Por conflictos personales (puede no estar correlacionado con el experimento).
  - Por preocupación de los padres (puede estar correlacionado con el experimento).

Ejemplo: Programa STAR

	Tratamiento	Control
$t = 1$	$[\bar{Y}_1 \mid D = 1] = 40$	$[\bar{Y}_1 \mid D = 0] = 20$
$t = 2$	$[\bar{Y}_2 \mid D = 1] = 80$	$[\bar{Y}_2 \mid D = 0] = 30$

$$\hat{\tau}_{dif-en-dif} =$$

$$[(\bar{Y}_2|D = 1) - (\bar{Y}_2|D = 0)]$$

$$-[(\bar{Y}_1|D = 1) - (\bar{Y}_1|D = 0)]$$

$$(80 - 40) - (30 - 20) = 30$$

Supuesto de tendencia paralela

La variable de resultado Y evoluciona de manera natural en el tiempo de la misma forma que entre el grupo de control y el de tratamiento.

$\Downarrow$

Permite utilizar $[(\bar{Y}_1|D=1)-(\bar{Y}_1|D=0)]$ como control apropiado de las diferencias preexistentes entre el grupo de tratamiento y el de control.

Entonces:

$\hat{\tau}_{dif-en-dif} \longrightarrow$ es insesgado

Modelo de regresión

Si tenemos datos de panel

$$Y_{it2} - Y_{it1} = \beta_0 + \beta_1 D_i + (u_{it2} - u_{it1})$$

$$\Delta Y_i = \beta_0 + \beta_1 D_i + v_i$$

Impacto del programa por el método de diferencias en diferencias.

Tomando esperanza condicional

$$E(\Delta Y|D=1) = \beta_0 + \beta_1 + E(v|D=1)$$
$$= \beta_0 + \beta_1$$
$$E(\Delta Y|D=0) = \beta_0 + E(v|D=0)$$
$$= \beta_0$$

Por independencia condicional $\longrightarrow$ $E(u_i|D_i) = 0$

$$E(\Delta Y|D=1) - E(\Delta Y|D=0) = \beta_1 = \tau_{dif-en-dif}$$

Diferencias en diferencias con regresores adicionales

Incluye características de los individuos antes de la asignación del tratamiento y que no son afectadas directamente por el tratamiento.

$$\Delta Y_i = \beta_0 + \beta_1 D_i + \underbrace{\gamma_1 X_{1i} + \cdots + \gamma_K X_{Ki}}+u_i$$

Insesgado siempre que $E(u_i|D_i) = 0$

- Controla diferencias entre el grupo de tratamiento y el de control.
- Mejora la eficiencia del estimador.
- Evalúa la validez del proceso de aleatorización.

Diferencias en diferencias para múltiples períodos

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \gamma_2 G_2 + \cdots + \gamma_N G_N + \delta_2 T_2 + \cdots + \delta_T T_T + \upsilon_{it}$$

Donde:

G → Controla por características individuales no observadas.

T → Controla por las diferencias entre períodos que afectan la variable de resultado, independiente de si el individuo recibe o no el tratamiento.

Dif en Dif para múltiples períodos

- Donde:

$$D_{it} \begin{cases} 1 & \text{si recibe el individuo } i \text{ tratamiento en el período } t \\ 0 & \text{de lo contrario} \end{cases}$$

$$G_i \begin{cases} 1 & \text{si corresponde al } i\text{-ésimo individuo} \\ 0 & \text{de lo contrario} \end{cases}$$

Efecto fijo por individuo.

$$T_i \begin{cases} 1 & \text{para el } t\text{-ésimo período.} \\ 0 & \text{de lo contrario} \end{cases}$$

Efecto fijo de tiempo.

Dif en Dif con datos de corte transversal repetidos

- Se usa cuando no se tienen datos de panel para la evaluación.
- Los datos de corte transversal repetidos deben ser <u>una muestra representativa de la misma población.</u>

Base de datos inicial
Antes del experimento natural
↓
Sustitutos del grupo de tratamiento
Sustitutos del grupo de control

Base de datos seguimiento
Después del experimento natural
↓
Grupo de tratamiento
Grupo de control

Ejemplo: *Canasta* Descripción

- Siguiendo con la idea de municipios petroleros y no petroleros, asignados aleatoriamente.

- **Datos**: Encuesta de hogares (repetidas).

  1. Línea de base: Antes del hallazgo de los yacimientos.

  2. Seguimiento: Después del experimento natural (luego del tratamiento).

- **Grupo de control**: Residentes de municipios sin petróleo pertenecientes al Sisben 1 y 2.
- **Grupo de tratamiento:** Residentes de municipios con petróleo pertenecientes al Sisben 1 y 2.

Ejemplo: *Canasta*

Acerca de las observaciones

No tenemos un panel, pero sí datos sobre los individuos residentes en municipios petroleros (que después son tratados por el decreto de regalías) y municipios no petroleros (que no serán tratados porque no tienen regalías).

Podemos utilizar los datos de corte transversal.

Para el grupo de control.	Datos de corte transversal de municipios que posteriormente no encuentran petróleo.	Observaciones pretratamiento
Para el grupo de tratamiento.	Datos de corte transversal de municipios que posteriormente encuentran petróleo.	

Modelo de regresión

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 I[t=2] + \beta_3 (D_i \times I[t=2]) + u_i$$

Donde:

$I[\cdot]$
- 1 si la condición [·] se cumple.
- 0 de lo contrario.

$D_i * I[\cdot] \longrightarrow$ Interacción entre la *dummy* de tratamiento y la *dummy* del período de seguimiento.

Efecto del programa por Dif en Dif

$$[E(Y|D_i = 1, t = 2) - E(Y|D_i = 1, t = 1)]$$

$$-[E(Y|D_i = 0, t = 2) - E(Y|D_i = 0, t = 1)]$$

$$=\big[[\beta_0 + \beta_1 + \beta_2 + \beta_3] - [\beta_0 + \beta_1]\big] - \big[[\beta_0 + \beta_2] - \beta_0\big]$$

$$= \beta_3$$

Efecto del programa estimado por el método diferencias en diferencias.

Ejemplo 1. Efecto de la migración de los "marielitos"

Pregunta: El efecto de la migración en los salarios. Pero… migrar es una decisión endógena.

Experimento natural:

Cambio inesperado en la ley estadounidense para los migrantes cubanos que generó una asignación del tratamiento (migrar) que *podría parecer aleatoria*.

Período:

1980 entre mayo y septiembre.

Resultado:

125.000 migrantes conocidos como "marielitos" llegaron a Estados Unidos.

Alrededor de 60.000 se establecieron en Miami, lo que incrementó la oferta laboral en 7%.

Ejemplo 1. Efecto de la migración de los “marielitos”

Objetivo:

Evaluar el efecto causal de la migración sobre los salarios del sitio de destino.

¿Cómo?:

Comparar los cambios salariales de los trabajadores poco calificados en Miami (tratados-migrantes) con los cambios en otras ciudades comparables (controles-no migrantes).

Datos:

Usa una serie de base de datos de corte transversal repetidos provenientes de encuestas preexistentes.

Ejemplo 1. Efecto de la migración de los "marielitos"

Estimador del efecto del programa:

$$[(\bar{Y}_2|D = 1) - (\bar{Y}_1|D = 1)]$$

$$-[(\bar{Y}_2|D = 0) - (\bar{Y}_1|D = 0)]$$

Regresión:

$$Y_i = \beta_0 + \beta_1 Miami_i + \beta_2 Post_i + \beta_3 (Miami_i \times Post_i) + \beta_4 X_i + u_i$$

Efecto del programa

Ejemplo 1. Efecto de la migración de los "marielitos"

- Donde:

$$Miami \begin{cases} 1 & \text{si el trabajador poco calificado está en Miami} \\ 0 & \text{de lo contrario} \end{cases}$$

$$Post_i \begin{cases} 1 & \text{si la observación es de un período posterior a la migración de los “marielitos”} \\ 0 & \text{de lo contrario} \end{cases}$$

X_i — Matriz de características observables de los trabajadores

Ejemplo 1. Efecto de la migración de los "marielitos"

Efectos de la inmigración en los salarios

Log(salarios) de negros			Diferencia en log(salarios) de negros Miami-ciudades comparables	
Año	Miami	Ciudades comparables	Actual	Ajustada
1979	1.59	1.74	-0.15	-0.12
1980	1.55	1.70	-0.16	-0.12
1981	1.61	1.72	-0.11	-0.10
1982	1.48	1.71	-0.24	-0.20
1983	1.48	1.69	-0.21	-0.15
1984	1.57	1.67	-0.10	-0.05
1985	1.60	1.65	-0.05	-0.01

Ejemplo 1. Efecto de la migración de los “marielitos”

Resultado:

Los salarios no empeoraron significativamente en Miami a raíz de la llegada de los inmigrantes cubanos.