Inferencia, Causalidad y Políticas Públicas ECO-60116

Week 06: Diferencias en Diferencias (DiD)

Eduard F. Martinez Gonzalez, Ph.D.

Departamento de Economía, Universidad Icesi

October 3, 2025

Roadmap

Estructura de Datos Panel

Definición: Un **panel de datos** contiene observaciones de múltiples unidades (individuos, regiones, empresas, etc.) a lo largo del tiempo.

Notación:

- $i = 1, \ldots, N$: unidades (e.g., distritos)
- $t = 1, \ldots, T$: periodos de tiempo
- y_{it}: variable dependiente
- x_{it}: covariables (pueden ser variantes o invariantes en el tiempo)

Tipos:

- **Balanceado:** todas las unidades observadas en todos los periodos
- **Desbalanceado:** faltan observaciones para algunas unidades o periodos

Ejemplo: Mortalidad por cólera en distritos de Londres para 1849 y 1854

Modelo Lineal Básico en Datos Panel

Modelo base:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \epsilon_{it}$$

Asunción clave: $\mathbb{E}[\epsilon_{it} \mid x_{it}] = 0$

Problema típico: Si hay variables omitidas correlacionadas con x_{it} y constantes en el tiempo para cada unidad $\rightarrow$ sesgo de variable omitida.

Soluciones posibles:

- Modelo de efectos fijos (captura heterogeneidad no observada constante en el tiempo)
- Modelo de efectos aleatorios (si la heterogeneidad es no correlacionada con los regresores)

Motivación: Necesitamos separar efectos de interés (e.g., política) de diferencias estructurales entre unidades.

Modelo con Efectos Fijos

Modelo con efectos fijos unitarios:

$$y_{it} = \alpha_i + \beta x_{it} + \epsilon_{it}$$

donde α_i captura características inobservables constantes en el tiempo para cada unidad i

Interpretación:

- α_i: heterogeneidad inobservable (e.g., calidad del sistema de agua en un distrito)
- El modelo utiliza solo **variación dentro de la unidad** para identificar β

Estimación:

- **Transformación within:** resta del promedio temporal
- **Dummy variables:** una por cada unidad (menos una)

Extensión: Efectos fijos también pueden incluir el tiempo:

$$y_{it} = \alpha_i + \lambda_t + \beta x_{it} + \epsilon_{it}$$

donde λ_t capta shocks agregados comunes (e.g., epidemia o cambio institucional)

Roadmap

1. Recap: Datos Panel
2. Diferencias En Diferencias
   - Motivación
   - Estimador de D&D
   - Supuestos de D&D
3. Metodos de Estimación: D&D
4. Extensiones de D&D

Motivación

- A mediados del siglo XIX, Londres enfrentó múltiples epidemias de cólera. En ese momento, los médicos desconocían su origen y no sabían cómo detener su propagación.

Motivación

- A mediados del siglo XIX, Londres enfrentó múltiples epidemias de cólera. En ese momento, los médicos desconocían su origen y no sabían cómo detener su propagación.
- La explicación dominante era la **teoría miasmática**, según la cual la enfermedad se transmitía a través de gases fétidos o tóxicos provenientes de materia en descomposición.

Motivación

- A mediados del siglo XIX, Londres enfrentó múltiples epidemias de cólera. En ese momento, los médicos desconocían su origen y no sabían cómo detener su propagación.
- La explicación dominante era la **teoría miasmática**, según la cual la enfermedad se transmitía a través de gases fétidos o tóxicos provenientes de materia en descomposición.
- El médico y epidemiólogo **John Snow** (1813–1858) cuestionó esta teoría, argumentando que las medidas tomadas para evitar la exposición a los “miasmas” no habían sido efectivas.

Motivación

- A mediados del siglo XIX, Londres enfrentó múltiples epidemias de cólera. En ese momento, los médicos desconocían su origen y no sabían cómo detener su propagación.

- La explicación dominante era la **teoría miasmática**, según la cual la enfermedad se transmitía a través de gases fétidos o tóxicos provenientes de materia en descomposición.

- El médico y epidemiólogo **John Snow** (1813–1858) cuestionó esta teoría, argumentando que las medidas tomadas para evitar la exposición a los "miasmas" no habían sido efectivas.

- Snow propuso una hipótesis alternativa: el cólera se transmitía a través del consumo de agua contaminada del sistema de acueducto.
    - Sostenía que las excreciones humanas contenían un agente infeccioso (aunque aún no identificado), el cual contaminaba las fuentes de agua potable.
    - El ciclo de consumo y reinfección a partir de esa agua explicaba la propagación sostenida de la epidemia.

El experimento natural

- ¿Cómo probar empíricamente la hipótesis de transmisión por agua contaminada?

El experimento natural

- ¿Cómo probar empíricamente la hipótesis de transmisión por agua contaminada?
- La **estrategia de identificación** de John Snow se basó en lo que hoy conocemos como un **experimento natural**.

El experimento natural

- ¿Cómo probar empíricamente la hipótesis de transmisión por agua contaminada?

- La **estrategia de identificación** de John Snow se basó en lo que hoy conocemos como un **experimento natural**.
  - Un **experimento natural** ocurre cuando factores fuera del control del investigador –a diferencia de un experimento aleatorizado– generan **asignaciones al tratamiento** de forma potencialmente **exógena**.

El experimento natural

- ¿Cómo probar empíricamente la hipótesis de transmisión por agua contaminada?
- La **estrategia de identificación** de John Snow se basó en lo que hoy conocemos como un **experimento natural**.
  - Un **experimento natural** ocurre cuando factores fuera del control del investigador –a diferencia de un experimento aleatorizado– generan **asignaciones al tratamiento** de forma potencialmente **exógena**.
- En la Londres de mediados del siglo XIX, el suministro de agua estaba a cargo de múltiples compañías privadas.

El experimento natural

- ¿Cómo probar empíricamente la hipótesis de transmisión por agua contaminada?

- La **estrategia de identificación** de John Snow se basó en lo que hoy conocemos como un **experimento natural**.
  - Un **experimento natural** ocurre cuando factores fuera del control del investigador –a diferencia de un experimento aleatorizado– generan **asignaciones al tratamiento** de forma potencialmente **exógena**.

- En la Londres de mediados del siglo XIX, el suministro de agua estaba a cargo de múltiples compañías privadas.

- La mayoría de ellas obtenían el agua del río Támesis, altamente contaminado con aguas residuales.

El experimento natural

- ¿Cómo probar empíricamente la hipótesis de transmisión por agua contaminada?
- La **estrategia de identificación** de John Snow se basó en lo que hoy conocemos como un **experimento natural**.
  - Un **experimento natural** ocurre cuando factores fuera del control del investigador –a diferencia de un experimento aleatorizado– generan **asignaciones al tratamiento** de forma potencialmente **exógena**.
- En la Londres de mediados del siglo XIX, el suministro de agua estaba a cargo de múltiples compañías privadas.
- La mayoría de ellas obtenían el agua del río Támesis, altamente contaminado con aguas residuales.
- Pero había una excepción clave...

El experimento natural (cont.)

- En 1852, una de las compañías proveedoras de agua, *Lambeth Company*, trasladó su punto de captación río arriba del Támesis, a una zona más limpia, ubicada por encima del vertimiento de aguas residuales.

El experimento natural (cont.)

- En 1852, una de las compañías proveedoras de agua, *Lambeth Company*, trasladó su punto de captación río arriba del Támesis, a una zona más limpia, ubicada por encima del vertimiento de aguas residuales.

- En contraste, otras compañías —particularmente la *Southwark and Vauxhall Company*— mantuvieron su punto de recolección río abajo, justo donde el río estaba contaminado con desechos humanos.

El experimento natural (cont.)

- En 1852, una de las compañías proveedoras de agua, *Lambeth Company*, trasladó su punto de captación río arriba del Támesis, a una zona más limpia, ubicada por encima del vertimiento de aguas residuales.
- En contraste, otras compañías —particularmente la *Southwark and Vauxhall Company*— mantuvieron su punto de recolección río abajo, justo donde el río estaba contaminado con desechos humanos.
- La hipótesis de Snow era clara: si el cólera se transmitía por el agua, los hogares abastecidos por *Lambeth* deberían tener tasas de mortalidad por cólera más bajas que aquellos atendidos por *Southwark and Vauxhall*, bajo condiciones similares.

El experimento natural (cont.)

- En 1852, una de las compañías proveedoras de agua, *Lambeth Company*, trasladó su punto de captación río arriba del Támesis, a una zona más limpia, ubicada por encima del vertimiento de aguas residuales.
- En contraste, otras compañías —particularmente la *Southwark and Vauxhall Company*— mantuvieron su punto de recolección río abajo, justo donde el río estaba contaminado con desechos humanos.
- La hipótesis de Snow era clara: si el cólera se transmitía por el agua, los hogares abastecidos por *Lambeth* deberían tener tasas de mortalidad por cólera más bajas que aquellos atendidos por *Southwark and Vauxhall*, bajo condiciones similares.
- Su estrategia: comparar tasas de mortalidad por cólera entre distritos abastecidos por cada compañía durante el brote de 1854.

Evidencia: diferencias en tasas de mortalidad por cólera

Tasas de mortalidad por cólera (por 100.000 habitantes)

Compañía de agua	Tasas 1849	Tasas 1854
Lambeth Company	847	193
Southwark & Vauxhall Company	1,349	1,466
Diferencia entre distritos		-1,273

Fuente: Adaptado de John Snow (1854).

- Los distritos que recibían el suministro de agua de Southwark and Vauxhall Company tuvieron una tasa de mortalidad 7.5 veces mayor relativo a los distritos que recibían el suministro de Lambeth company.
- En primera medida esto parece una evidencia contundente, pero no se puede descartar problemas de selección.

Roadmap

Resultados potenciales: caso John Snow

Supuesto de identificación de Snow en términos de resultados potenciales:

$$y_d = \begin{cases} y_d^1 & \text{tasa de mortalidad del distrito } d \text{ si recibe agua limpia} \\ y_d^0 & \text{tasa de mortalidad del distrito } d \text{ si recibe agua contaminada} \end{cases}$$

Resultados potenciales: caso John Snow

Supuesto de identificación de Snow en términos de resultados potenciales:

$$y_d = \begin{cases} y_d^1 & \text{tasa de mortalidad del distrito } d \text{ si recibe agua limpia} \\ y_d^0 & \text{tasa de mortalidad del distrito } d \text{ si recibe agua contaminada} \end{cases}$$

Asignación al tratamiento:

$D_d = 1$ si el proveedor de agua del distrito d es *Lambeth company* y
$D_d = 0$ si el proveedor es *Southwark and Vauxhall Company*.

Resultados potenciales: caso John Snow

Supuesto de identificación de Snow en términos de resultados potenciales:

$$y_d = \begin{cases} y_d^1 & \text{tasa de mortalidad del distrito } d \text{ si recibe agua limpia} \\ y_d^0 & \text{tasa de mortalidad del distrito } d \text{ si recibe agua contaminada} \end{cases}$$

Asignación al tratamiento:

$D_d = 1$ si el proveedor de agua del distrito d es *Lambeth company* y
$D_d = 0$ si el proveedor es *Southwark and Vauxhall Company*.

El efecto causal de interés:

$$\tau = y_d^1 - y_d^0$$

Resultados potenciales: caso John Snow

Supuesto de identificación de Snow en términos de resultados potenciales:

$$y_d = \begin{cases} y_d^1 & \text{tasa de mortalidad del distrito } d \text{ si recibe agua limpia} \\ y_d^0 & \text{tasa de mortalidad del distrito } d \text{ si recibe agua contaminada} \end{cases}$$

Asignación al tratamiento:

$D_d = 1$ si el proveedor de agua del distrito d es *Lambeth company* y
$D_d = 0$ si el proveedor es *Southwark and Vauxhall Company*.

El efecto causal de interés:

$$\tau = y_d^1 - y_d^0$$

Problema fundamental de la inferencia causal: para cada distrito d sólo observamos uno de los dos estados:

$$y_d = D_d \cdot y_d^1 \quad + \quad (1 - D_d) \cdot y_d^0$$

Resultados potenciales: caso John Snow

Estimador Ingenuo de Snow:

$$\tau_{\text{Snow}} = \mathbb{E}[y_d | D_d = 1] - \mathbb{E}[y_d | D_d = 0]$$

Resultados potenciales: caso John Snow

Estimador Ingenuo de Snow:

$$\tau_{\text{Snow}} = \mathbb{E}[y_d|D_d = 1] - \mathbb{E}[y_d|D_d = 0]$$

Pero sabemos que esto es igual a:

$$\mathbb{E}[y_d|D_d = 1] - \mathbb{E}[y_d|D_d = 0] = \underbrace{\mathbb{E}[y_d^1|D_d = 1] - \mathbb{E}[y_d^0|D_d = 1]}_{\tau} + \underbrace{\mathbb{E}[y_d^0|D_d = 1] - \mathbb{E}[y_d^0|D_d = 0]}_{\text{Sesgo de selección}}$$

Resultados potenciales: caso John Snow

Estimador Ingenuo de Snow:

$$\tau_{\text{Snow}} = \mathbb{E}[y_d | D_d = 1] - \mathbb{E}[y_d | D_d = 0]$$

Pero sabemos que esto es igual a:

$$\mathbb{E}[y_d | D_d = 1] - \mathbb{E}[y_d | D_d = 0] = \underbrace{\mathbb{E}[y_d^1 | D_d = 1] - \mathbb{E}[y_d^0 | D_d = 1]}_{\tau}$$
$$+ \underbrace{\mathbb{E}[y_d^0 | D_d = 1] - \mathbb{E}[y_d^0 | D_d = 0]}_{\text{Sesgo de selección}}$$

Interpretación: $\tau_{\text{Snow}} = \tau$ únicamente si no hay diferencias sistemáticas entre los distritos que son abastecidos por cada una de las compañías.

$$\mathbb{E}[y_d^0 \mid D_d = 1] = \mathbb{E}[y_d^0 \mid D_d = 0]$$

Intuición de diferencias en diferencias

- **No** podemos asumir que los distritos tratados y no tratados eran comparables en 1854.

Intuición de diferencias en diferencias

- **No** podemos asumir que los distritos tratados y no tratados eran comparables en 1854.

- Pero podemos usar información previa al tratamiento (por ejemplo, año 1849, antes del cambio de fuente de agua en 1852) para controlar por diferencias estructurales entre grupos.

Intuición de diferencias en diferencias

- **No** podemos asumir que los distritos tratados y no tratados eran comparables en 1854.

- Pero podemos usar información previa al tratamiento (por ejemplo, año 1849, antes del cambio de fuente de agua en 1852) para controlar por diferencias estructurales entre grupos.

- *"Ser tratado"* significa que el distrito pasó a recibir agua limpia a partir de 1852 (*Lambeth Company*). *"Control"* significa que el distrito continuó recibiendo agua contaminada (*Southwark & Vauxhall Company*).

- **Supuesto clave de identificación:** en ausencia de tratamiento, la evolución de las tasas de mortalidad habría sido la misma para ambos grupos:

$$\mathbb{E}[y_d^0(1854) - y_d^0(1849) \mid D_d = 1] = \mathbb{E}[y_d^0(1854) - y_d^0(1849) \mid D_d = 0]$$

- *Este supuesto se conoce como el supuesto de* **tendencias paralelas**.

Intuición de diferencias en diferencias

- **No** podemos asumir que los distritos tratados y no tratados eran comparables en 1854.

- Pero podemos usar información previa al tratamiento (por ejemplo, año 1849, antes del cambio de fuente de agua en 1852) para controlar por diferencias estructurales entre grupos.

- *"Ser tratado"* significa que el distrito pasó a recibir agua limpia a partir de 1852 (*Lambeth Company*). *"Control"* significa que el distrito continuó recibiendo agua contaminada (*Southwark & Vauxhall Company*).

- **Supuesto clave de identificación:** en ausencia de tratamiento, la evolución de las tasas de mortalidad habría sido la misma para ambos grupos:

$$\mathbb{E}[y_d^0(1854) - y_d^0(1849) \mid D_d = 1] = \mathbb{E}[y_d^0(1854) - y_d^0(1849) \mid D_d = 0]$$

- *Este supuesto se conoce como el supuesto de* **tendencias paralelas**.

- **Implicación:** podemos comparar la evolución en el tiempo del grupo tratado (agua limpia desde 1852) con la del grupo control (agua contaminada) para identificar τ.

Estimador de diferencias en diferencias (D&D)

Observaciones: disponemos de tasas de mortalidad en dos momentos del tiempo (1849 y 1854), para dos grupos de distritos:

- $D_d = 1$: distritos abastecidos por *Lambeth Company* (agua limpia en 1854)
- $D_d = 0$: distritos abastecidos por *Southwark & Vauxhall Company* (agua contaminada en ambos años)

Estimador de DiD:

$$\tau_{\text{DiD}} = \underbrace{\mathbb{E}[y_d \mid D_d = 1, t = 1854] - \mathbb{E}[y_d \mid D_d = 1, t = 1849]}_{\text{Cambio en el grupo tratado}} - \underbrace{(\mathbb{E}[y_d \mid D_d = 0, t = 1854] - \mathbb{E}[y_d \mid D_d = 0, t = 1849])}_{\text{Cambio en el grupo control}}$$

Bajo el supuesto de tendencias paralelas:

$$\tau_{\text{DiD}} = \mathbb{E}[y_d^1(1854) - y_d^0(1854) \mid D_d = 1] = \tau$$

Estimador de D&D para tasas de mortalidad por cólera

Tasas de mortalidad por cólera (por 100.000 habitantes)

Compañía de agua	Tasas 1849	Tasas 1854	Cambio (1854-1849)
Lambeth Company	847	193	-653
Southwark & Vauxhall Company	1,349	1,466	118
Diferencia en Diferencias			-771

Fuente: Adaptado de John Snow (1854).

Estimador de diferencias en diferencias:

$$\begin{aligned}\tau_{\text{DiD}} &= (193 - 847) - (1466 - 1349) \\ &= -653 - 118 = \boxed{-771}\end{aligned}$$

Interpretación: el acceso a agua limpia está asociado con una reducción de 771 muertes por cólera por cada 100.000 habitantes.

Roadmap

Supuesto de tendencias paralelas

- Implica que, en ausencia del tratamiento, la evolución de la variable de interés habría sido la misma en ambos grupos.
- Puede **validarse parcialmente** si contamos con datos para varios **periodos pre-tratamiento**.
- Supongamos que los resultados potenciales en ausencia de tratamiento se pueden escribir como:

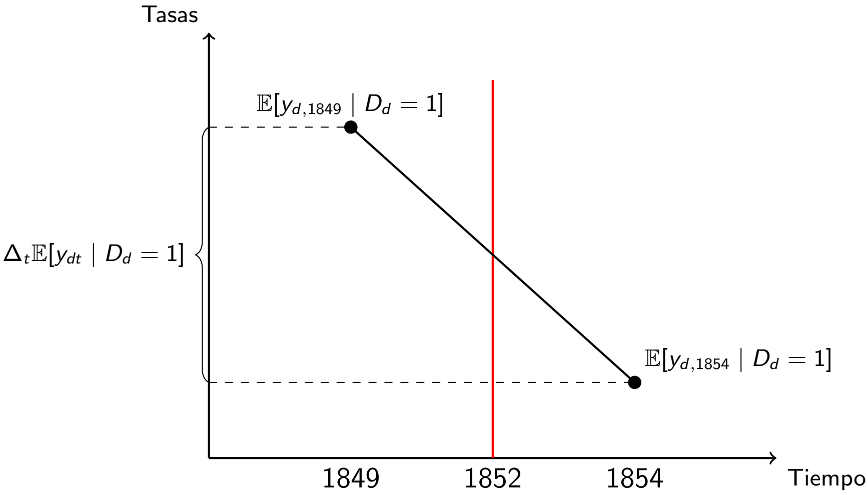
$$\mathbb{E}[y^0_{d,st} \mid s, t] = \phi_t + \eta_s$$

  donde:
  - ϕ_t: efecto común a todos los grupos en el tiempo (cambio agregado)
  - η_s: diferencias fijas entre unidades (estados, distritos, etc.)
- Entonces, para cualquier periodo t pre-tratamiento:

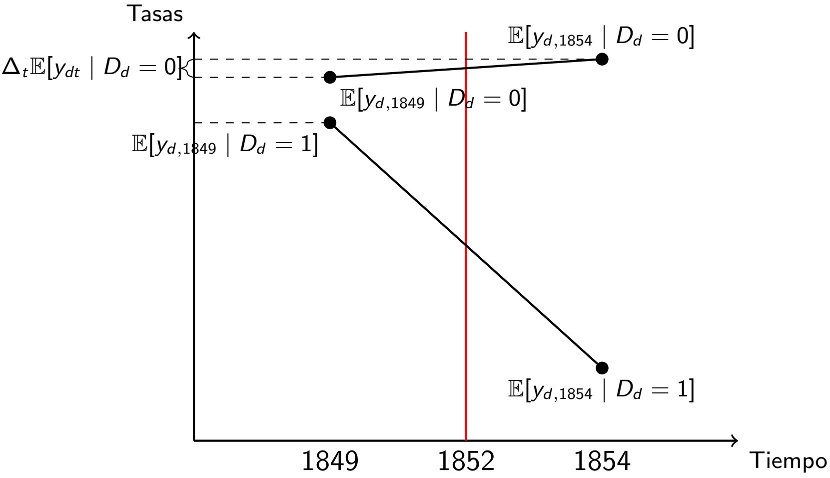
$$\mathbb{E}[y^0_{ist} \mid s = \mathsf{L}] - \mathbb{E}[y^0_{ist} \mid s = \mathsf{SV}] = (\phi_t + \eta_{\mathsf{L}}) - (\phi_t + \eta_{\mathsf{SV}}) = \eta_{\mathsf{L}} - \eta_{\mathsf{SV}}$$

- **Conclusión:** la diferencia entre los grupos es constante en el tiempo antes del tratamiento. Eso es lo que se denomina **tendencias paralelas**.

Representación gráfica: Grupo tratado (Lambeth)

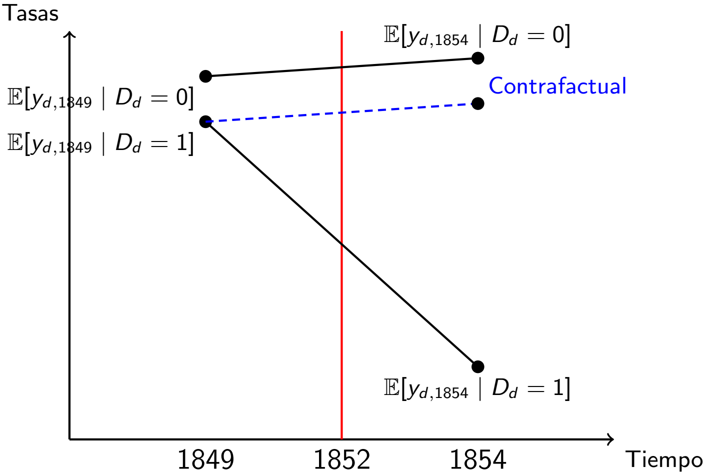


Representación gráfica: Grupo tratado vs. control



El valor esperado de la tasa de mortalidad en el periodo pre no es igual en los dos grupos.

Representación gráfica: Contrafactual



En ausencia del tratamiento, el cambio esperado en la variable dependiente del grupo de tratamiento es igual al cambio observado en esa variable para el grupo de control.

Roadmap

Modelo de regresión lineal

$$y_{dt} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 D_d + \beta_3 (D_d \times \text{Post}_t) + \epsilon_{dt}$$

- y_{dt}: Tasa de mortalidad por cólera en el distrito d y periodo t.
- d: Distritos de Londres (e.g., Lambeth o Southwark & Vauxhall).
- t: Puede tomar dos valores: 1849 y 1854.
- Post_t: Indicador temporal. Toma valor 1 si el año es posterior al tratamiento (1854), y 0 si es antes (1849).
- D_d: Indicador de tratamiento. Toma valor 1 si el distrito es tratado (Lambeth), y 0 si es control (Southwark & Vauxhall).
- $D_d \times \text{Post}_t$: Término de interacción que captura el efecto diferencial del tratamiento en el tiempo, es decir, la estimación DiD.

Modelo de regresión lineal

$$y_{dt} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 D_d + \beta_3 (D_d \times \text{Post}_t) + \epsilon_{dt}$$

Valores esperados del modelo:

1. $\mathbb{E}[y_{dt} | \text{Post} = 1, D_d = 1] = \beta_0 + \beta_1 + \beta_2 + \beta_3$
2. $\mathbb{E}[y_{dt} | \text{Post} = 0, D_d = 1] = \beta_0 + \beta_2$
3. $\mathbb{E}[y_{dt} | \text{Post} = 1, D_d = 0] = \beta_0 + \beta_1$
4. $\mathbb{E}[y_{dt} | \text{Post} = 0, D_d = 0] = \beta_0$

Modelo de regresión lineal

$$y_{dt} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 D_d + \beta_3 (D_d \times \text{Post}_t) + \epsilon_{dt}$$

Valores esperados del modelo:

1. $\mathbb{E}[y_{dt} | \text{Post} = 1, D_d = 1] = \beta_0 + \beta_1 + \beta_2 + \beta_3$
2. $\mathbb{E}[y_{dt} | \text{Post} = 0, D_d = 1] = \beta_0 + \beta_2$
3. $\mathbb{E}[y_{dt} | \text{Post} = 1, D_d = 0] = \beta_0 + \beta_1$
4. $\mathbb{E}[y_{dt} | \text{Post} = 0, D_d = 0] = \beta_0$

Diferencias en el tiempo para cada grupo ($\Delta_t \mathbb{E}[Y_{dt} | D_d]$):

a. $(1) - (2) = \beta_1 + \beta_3$

b. $(3) - (4) = \beta_1$

Modelo de regresión lineal

$$y_{dt} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 D_d + \beta_3 (D_d \times \text{Post}_t) + \epsilon_{dt}$$

Valores esperados del modelo:

1. $\mathbb{E}[y_{dt} | \text{Post} = 1, D_d = 1] = \beta_0 + \beta_1 + \beta_2 + \beta_3$
2. $\mathbb{E}[y_{dt} | \text{Post} = 0, D_d = 1] = \beta_0 + \beta_2$
3. $\mathbb{E}[y_{dt} | \text{Post} = 1, D_d = 0] = \beta_0 + \beta_1$
4. $\mathbb{E}[y_{dt} | \text{Post} = 0, D_d = 0] = \beta_0$

Diferencias en el tiempo para cada grupo ($\Delta_t \mathbb{E}[Y_{dt} | D_d]$):

a. (1) – (2) $= \beta_1 + \beta_3$

b. (3) – (4) $= \beta_1$

Estimador de D&D ($\Delta_t \mathbb{E}[Y_{dt} | D_d = 1] - \Delta_t \mathbb{E}[Y_{dt} | D_d = 0]$):

- $\tau_{\text{DiD}} \equiv (a) - (b) = \beta_3$

Algunas ventajas del modelo de regresión

1. Podemos agregar controles si son necesarios.
   - Controles que hacen que el supuesto de **tendencias paralelas** sea más plausible.
   - Reducen la varianza residual $\Longrightarrow$ mejora la precisión del estimador.
   - Evitar usar variables de control que puedan estar afectadas por el tratamiento.
2. Permite calcular errores estándar directamente.
3. Se puede generalizar a más de dos periodos. Esto a su vez permite testear **parcialmente** el supuesto de tendencias paralelas (próxima clase).
4. Se puede generalizar a tratamientos que no son dicotómicos. Por ejemplo, diferentes intensidades del tratamiento.

Modelo TWFE (Two-Way Fixed Effects)

Extensión del modelo DiD con múltiples unidades y periodos:

$$y_{it} = \alpha_i + \lambda_t + \tau \cdot D_{it} + \varepsilon_{it} \tag{1}$$

- y_{it}: resultado de interés para la unidad i en el periodo t
- α_i: efectos fijos por unidad (capturan diferencias constantes entre unidades)
- λ_t: efectos fijos por periodo (capturan shocks comunes en el tiempo)
- D_{it}: indicador de tratamiento (1 si la unidad i está tratada en t, 0 si no)
- τ: estimador del efecto promedio del tratamiento
- ε_{it}: término de error idiosincrático

Supuesto clave: **Tendencias paralelas entre unidades tratadas y no tratadas** antes del tratamiento.

Estimación: por Mínimos Cuadrados Ordinarios (OLS) con efectos fijos de unidad y de tiempo.

Roadmap

D&D con variación en el momento del tratamiento

En muchos casos el tratamiento no ocurre al mismo tiempo para todas las unidades:

- Departamentos que implementan una reforma laboral en distintos años.
- Distritos escolares que adoptan tecnología en diferentes momentos.
- Ciudades que reciben infraestructura o inversión de manera gradual.

El modelo TWFE tradicional combina comparaciones entre unidades tratadas en distintos momentos. Esto puede llevar a estimaciones sesgadas si tratamiento es:

- **Heterogéneos en el tiempo**: el efecto cambia según cuánto tiempo ha pasado desde la intervención.
- **Heterogéneos entre unidades**: algunas unidades responden más que otras.

Solución: Usar estimadores que separan los efectos por cohorte de tratamiento y tiempo (e.g. Callaway & Sant'Anna, 2021; Sun & Abraham, 2021). Ver Callaway & Sant'Anna, 2021

D&D con múltiples períodos de tiempo

En estudios empíricos aplicados, a menudo se dispone de datos panel con varios periodos antes y después del tratamiento.

Ventajas:

- Permite validar el supuesto de tendencias paralelas observando la evolución pre-tratamiento.
- Facilita la estimación de efectos dinámicos del tratamiento (e.g., efectos acumulados, persistentes o retardados).
- Mejora la eficiencia al aprovechar más información temporal.

Ejemplos:

- Evaluar el impacto de una política fiscal usando datos de ingresos municipales entre 2000 y 2020.
- Estimar efectos de una reforma educativa implementada en 2012 observando puntajes escolares desde 2005.

Precaución: Si el tratamiento varía en el tiempo entre unidades, se deben considerar los sesgos del estimador TWFE. En ese caso, usar métodos alternativos (e.g., Callaway-Sant'Anna, event studies). Ver Callaway-Sant'Anna.

TWFE con varios tratamientos

En algunos estudios, las unidades pueden recibir más de un tratamiento a lo largo del tiempo, o diferentes tipos de tratamiento:

- Una ciudad implementa primero un toque de queda y luego una política de cierre de escuelas.
- Un país adopta diferentes componentes de una reforma tributaria en distintos años.

Problema: El estimador TWFE tradicional puede:

- Promediar **efectos de tratamientos diferentes** de forma no interpretable.
- Generar **comparaciones inválidas** entre grupos ya tratados vs. recién tratados.

Alternativas:

- Modelos de efectos heterogéneos por tipo y momento del tratamiento.
- Métodos como Callaway-Sant'Anna (2021) o Gardner (2022) que permiten descomposición adecuada. Ver