

Difference-in-Difference, DD

Yasmani Vargas

yasmani.vargas@unap.edu.pe

Unidad de Posgrado
Facultad de Ingeniería Económica
Universidad Nacional del Altiplano - Puno

Sesión 6
Ciclo II - Diseño y Evaluación de Políticas Públicas
Noviembre, 2025

Contenido

- 1 Motivación y comparación con otros métodos
- 2 Sesgo de selección y diferencias como contrafactual
- 3 Método DD: teoría y regresión
- 4 Modelo de efectos fijos en panel
- 5 Implementación del método DD
- 6 Ventajas, limitaciones y violaciones a tendencias paralelas
- 7 Modelos alternativos basados en DD
 - PSM con DD
 - Método de triple diferencia (DDD)
 - Ajuste por tendencias temporales diferenciales

Contenido

- 1 Motivación y comparación con otros métodos
- 2 Sesgo de selección y diferencias como contrafactual
- 3 Método DD: teoría y regresión
- 4 Modelo de efectos fijos en panel
- 5 Implementación del método DD
- 6 Ventajas, limitaciones y violaciones a tendencias paralelas
- 7 Modelos alternativos basados en DD
 - PSM con DD
 - Método de triple diferencia (DDD)
 - Ajuste por tendencias temporales diferenciales

DD frente a aleatorización y PSM

- **Aleatorización:** elimina el sesgo de selección al asignar el tratamiento de forma exógena.
- **PSM:** corrige selección *sobre observables*, asumiendo independencia condicional.
- **Doble diferencia (DD):**
 - ▶ Acepta que existe heterogeneidad no observada en la participación.
 - ▶ Supone que esa heterogeneidad no observada es **invariante en el tiempo**.
 - ▶ Utiliza datos **antes y después** para tratados y controles.
- **Intuición:**
 - ▶ Los cambios en los resultados de los no participantes revelan la trayectoria contrafactual de los participantes.
 - ▶ La diferenciación elimina el componente fijo no observado.

Variantes y extensiones del enfoque DD

- **PSM + DD:**

- ▶ Usa PSM para emparejar unidades en la línea de base dentro del soporte común.
- ▶ Luego aplica DD sobre la muestra emparejada.

- **Control por condiciones iniciales de área:**

- ▶ Permite corregir asignación no aleatoria del programa correlacionada con el crecimiento.

- **Triple diferencia (DDD):**

- ▶ Útil cuando no hay línea de base.
- ▶ Introduce un segundo grupo de control post-tratamiento.

Contenido

- 1 Motivación y comparación con otros métodos
- 2 Sesgo de selección y diferencias como contrafactual**
- 3 Método DD: teoría y regresión
- 4 Modelo de efectos fijos en panel
- 5 Implementación del método DD
- 6 Ventajas, limitaciones y violaciones a tendencias paralelas
- 7 Modelos alternativos basados en DD
 - PSM con DD
 - Método de triple diferencia (DDD)
 - Ajuste por tendencias temporales diferenciales

Usando diferencias como contrafactual

- Notación de dos períodos:

$t = 0$ (antes del programa), $t = 1$ (después)

Y_t^T : resultado promedio en tratados, Y_t^C : resultado promedio en controles

- Estimador DD:

$$DD = E(Y_1^T - Y_0^T \mid T_t = 1) - E(Y_1^C - Y_0^C \mid T_t = 0). \quad (1)$$

- Interpretación:

- ▶ Primer término: cambio observado en el grupo tratado.
- ▶ Segundo término: cambio observado en el grupo de control.
- ▶ La diferencia entre ambos aproxima el cambio *debido al programa*.

- Clave: los **cambios** en controles representan el contrafactual dinámico de los tratados.

DD y heterogeneidad no observada

- Supongamos que:
 - ▶ Existen factores no observados (habilidad, normas, calidad institucional, etc.) que difieren entre tratados y controles.
 - ▶ Pero estos factores son **constantes en el tiempo**.
- En ese caso:
 - ▶ Las diferencias de nivel entre grupos se mantienen, pero los **cambios** en el tiempo son comparables.
 - ▶ La doble diferencia elimina el componente fijo no observado.
- DD relaja el supuesto de “selección solo sobre observables” del PSM:
 - ▶ Permite selección sobre no observables, siempre que sea invariante en el tiempo.
 - ▶ Este supuesto se formaliza como **tendencia paralela**.

Contenido

- 1 Motivación y comparación con otros métodos
- 2 Sesgo de selección y diferencias como contrafactual
- 3 Método DD: teoría y regresión**
- 4 Modelo de efectos fijos en panel
- 5 Implementación del método DD
- 6 Ventajas, limitaciones y violaciones a tendencias paralelas
- 7 Modelos alternativos basados en DD
 - PSM con DD
 - Método de triple diferencia (DDD)
 - Ajuste por tendencias temporales diferenciales

Formulación en regresión: modelo básico DD

- Modelo lineal simple:

$$Y_{it} = \alpha + \beta T_i t + \rho T_i + \gamma t + \varepsilon_{it},$$

donde:

- ▶ Y_{it} : resultado de la unidad i en el tiempo t .
 - ▶ T_i : indicador de pertenencia al grupo de tratamiento (constante en el tiempo).
 - ▶ t : indicador temporal (0 = antes, 1 = después).
 - ▶ $T_i t$: interacción tratamiento–tiempo.
- El coeficiente de interés:

$$\beta = DD,$$

es el efecto promedio del programa.

- ρ : diferencia de nivel inicial entre grupos.
- γ : efecto temporal común (choques agregados).

Intuición vía valores esperados

- Para tratados:

$$E(Y_1^T \mid T = 1) = \alpha + DD + \rho + \gamma, \quad (2)$$

$$E(Y_0^T \mid T = 1) = \alpha + \rho. \quad (3)$$

$$E(Y_1^T - Y_0^T \mid T = 1) = DD + \gamma.$$

- Para controles:

$$E(Y_1^C \mid T = 0) = \alpha + \gamma, \quad (4)$$

$$E(Y_0^C \mid T = 0) = \alpha. \quad (5)$$

$$E(Y_1^C - Y_0^C \mid T = 0) = \gamma.$$

- Restando:

$$DD = (DD + \gamma) - \gamma.$$

- Por tanto, el coeficiente de la interacción $T_i t$ recoge la **doble diferencia**.

Comparación con otros estimadores ingenuos

- **Solo tratados, antes vs. después** (reflexivo):

$$\text{Impacto} = DD + \gamma \Rightarrow \text{sesgo} = \gamma$$

(no diferencia cambios agregados en el tiempo).

- **Solo comparación en el período post:**

$$\text{Impacto} = DD + \rho \Rightarrow \text{sesgo} = \rho$$

(no corrige diferencias iniciales entre grupos).

- **DD** combina ambas correcciones:

- ▶ Elimina el efecto común de tiempo γ .
- ▶ Elimina la diferencia fija entre grupos ρ .

Condiciones para interpretar correctamente DD

- Especificación correcta del modelo (estructura aditiva adecuada).
- Condiciones de exogeneidad:

$$\text{Cov}(\varepsilon_{it}, T_i) = 0, \quad \text{Cov}(\varepsilon_{it}, t) = 0, \quad \text{Cov}(\varepsilon_{it}, T_i t) = 0.$$

- La última expresión se conoce como:

Supuesto de tendencia paralela

Las características no observadas que afectan la participación no generan trayectorias de resultados que diverjan diferentemente en tratados y controles en ausencia del programa.

- DD es insesgado solo si la fuente de sesgo de selección es **aditiva** y **constante en el tiempo**.

Contenido

- 1 Motivación y comparación con otros métodos
- 2 Sesgo de selección y diferencias como contrafactual
- 3 Método DD: teoría y regresión
- 4 Modelo de efectos fijos en panel**
- 5 Implementación del método DD
- 6 Ventajas, limitaciones y violaciones a tendencias paralelas
- 7 Modelos alternativos basados en DD
 - PSM con DD
 - Método de triple diferencia (DDD)
 - Ajuste por tendencias temporales diferenciales

Generalización a múltiples períodos: efectos fijos

- Modelo de efectos fijos:

$$Y_{it} = \Phi T_t + \delta X_{it} + \eta_i + \varepsilon_{it},$$

donde:

- ▶ T_t : indicador de período de tratamiento (puede valer 1 para $t \geq t_0$).
 - ▶ X_{it} : covariables que varían en el tiempo.
 - ▶ η_i : heterogeneidad individual invariante en el tiempo (no observada).
- Fuente de endogeneidad:
 - ▶ η_i puede estar correlacionada con el tratamiento o con X_{it} .

Diferenciación y eliminación de heterogeneidad fija

- Diferencia entre t y $t - 1$:

$$(Y_{it} - Y_{i,t-1}) = \Phi(T_t - T_{t-1}) + \delta(X_{it} - X_{i,t-1}) + (\eta_i - \eta_i) + (\varepsilon_{it} - \varepsilon_{i,t-1}),$$

que se resume como:

$$\Delta Y_{it} = \Phi \Delta T_t + \delta \Delta X_{it} + \Delta \varepsilon_{it}.$$

- Resultado:
 - ▶ La heterogeneidad individual no observada η_i desaparece.
 - ▶ OLS en la ecuación diferenciada proporciona un estimador de Φ sin sesgo por η_i .
- Con dos períodos, este modelo es equivalente al DD clásico.
- Con más períodos, permite mayor flexibilidad (y exige correcciones por correlación serial en los errores).

Contenido

- 1 Motivación y comparación con otros métodos
- 2 Sesgo de selección y diferencias como contrafactual
- 3 Método DD: teoría y regresión
- 4 Modelo de efectos fijos en panel
- 5 Implementación del método DD**
- 6 Ventajas, limitaciones y violaciones a tendencias paralelas
- 7 Modelos alternativos basados en DD
 - PSM con DD
 - Método de triple diferencia (DDD)
 - Ajuste por tendencias temporales diferenciales

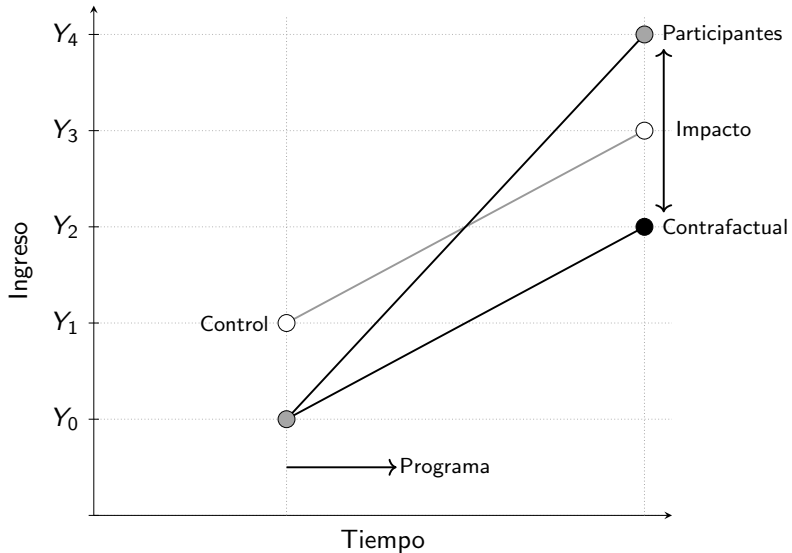
Pasos para implementar DD con datos de panel

- **1. Línea de base:** recolectar datos en áreas de tratamiento y control antes de la intervención.
- **2. Seguimiento:** encuestar a las mismas unidades después del programa.
- **3. Cálculo de cambios:**
 - ▶ Calcular $(Y_1^T - Y_0^T)$ para tratados.
 - ▶ Calcular $(Y_1^C - Y_0^C)$ para controles.
- **4. Doble diferencia:**

$$DD = (Y_1^T - Y_0^T) - (Y_1^C - Y_0^C).$$

- **5. Versión regresión:**
 - ▶ Estimar el modelo con interacción tratamiento–tiempo o un modelo de efectos fijos.
 - ▶ Incorporar covariables para mejorar precisión y controlar factores observados.

Ejemplo gráfico de DD



Interpretación de la figura DD

- La figura ilustra:
 - ▶ Trayectoria observada de tratados: de Y_0 a Y_4 .
 - ▶ Trayectoria observada de controles: de Y_1 a Y_3 .
 - ▶ Trayectoria contrafactual no observada de tratados: de Y_2 a Y_4 .

- Bajo tendencia paralela:

$$(Y_4 - Y_2) = (Y_1 - Y_3).$$

- Sustituyendo en:

$$DD = (Y_4 - Y_0) - (Y_3 - Y_1),$$

se obtiene:

$$DD = (Y_4 - Y_2),$$

que es el verdadero impacto promedio del programa.

Contenido

- 1 Motivación y comparación con otros métodos
- 2 Sesgo de selección y diferencias como contrafactual
- 3 Método DD: teoría y regresión
- 4 Modelo de efectos fijos en panel
- 5 Implementación del método DD
- 6 Ventajas, limitaciones y violaciones a tendencias paralelas**
- 7 Modelos alternativos basados en DD
 - PSM con DD
 - Método de triple diferencia (DDD)
 - Ajuste por tendencias temporales diferenciales

Ventajas y desventajas del método DD

- **Ventajas:**

- ▶ Relaja el supuesto de selección solo sobre observables.
- ▶ Permite manejar heterogeneidad no observada *fija* entre grupos.
- ▶ Tiene interpretación intuitiva y se implementa con modelos lineales estándar.

- **Desventajas:**

- ▶ Requiere el supuesto de **tendencias paralelas**, a menudo fuerte.
- ▶ Si la heterogeneidad no observada varía en el tiempo, DD puede estar sesgado.
- ▶ Puede ser sensible a choques macroeconómicos que afectan de forma diferencial a tratados y controles.

Heterogeneidad no observada cambiante en el tiempo

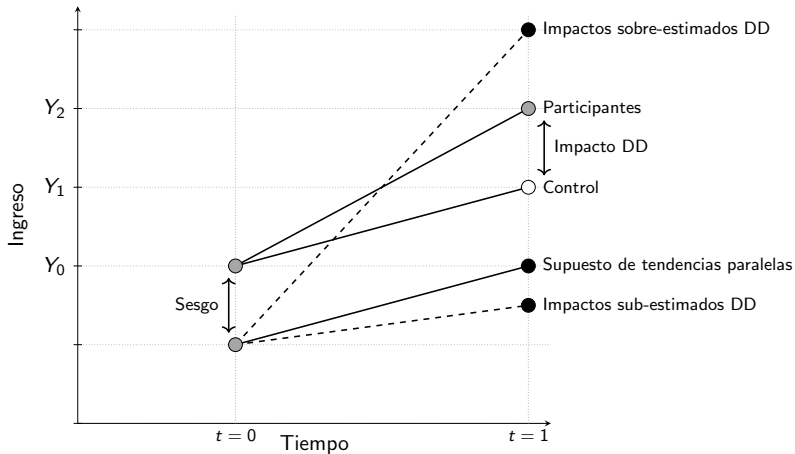


Figure: Heterogeneidad no observada variable en el tiempo

Ejemplos de violaciones a DD

- **Ashenfelter's Dip:**

- ▶ Programas de capacitación laboral donde la inscripción aumenta tras una caída transitoria en ingresos.
- ▶ El grupo tratado habría recuperado ingresos incluso sin el programa.
- ▶ DD tiende a **sobreestimar** el impacto.

- **Diferencias en condiciones iniciales:**

- ▶ Programas focalizados en áreas muy pobres que convergen más rápido.
- ▶ Diferencias agroclimáticas o de infraestructura inicial entre tratamiento y control.
- ▶ Si estas condiciones afectan tanto la ubicación como el crecimiento, generan sesgo en DD.

Control de condiciones iniciales

- Cuando la asignación del programa depende de condiciones iniciales:
 - ▶ No controlarlas implica sesgo por variables omitidas.
- Solución:
 - ▶ Incluir variables que capturen las condiciones iniciales de área (infraestructura, acceso a mercados, pobreza inicial, etc.).
 - ▶ Usar modelos de efectos fijos a nivel de área y DD con múltiples períodos para aislar mejor el impacto.
- Evidencia empírica:
 - ▶ Al controlar estas condiciones, los impactos de largo plazo pueden diferir sustancialmente de las estimaciones DD simples.

Contenido

- 1 Motivación y comparación con otros métodos
- 2 Sesgo de selección y diferencias como contrafactual
- 3 Método DD: teoría y regresión
- 4 Modelo de efectos fijos en panel
- 5 Implementación del método DD
- 6 Ventajas, limitaciones y violaciones a tendencias paralelas
- 7 Modelos alternativos basados en DD**
 - PSM con DD
 - Método de triple diferencia (DDD)
 - Ajuste por tendencias temporales diferenciales

Contenido

- 1 Motivación y comparación con otros métodos
- 2 Sesgo de selección y diferencias como contrafactual
- 3 Método DD: teoría y regresión
- 4 Modelo de efectos fijos en panel
- 5 Implementación del método DD
- 6 Ventajas, limitaciones y violaciones a tendencias paralelas
- 7 Modelos alternativos basados en DD**
 - PSM con DD
 - Método de triple diferencia (DDD)
 - Ajuste por tendencias temporales diferenciales

Combinando PSM y DD

- Motivación:
 - ▶ PSM corrige selección sobre observables.
 - ▶ DD corrige heterogeneidad no observada fija.
- Procedimiento:
 - 1 Estimar el puntaje de propensión en la línea de base (antes del programa).
 - 2 Emparejar unidades tratadas y de control dentro del soporte común.
 - 3 Aplicar DD sobre las unidades emparejadas.
- Estimador DD por área de tratamiento i :

$$DD_i = (Y_{i2}^T - Y_{i1}^T) - \sum_{j \in C} \omega(i, j)(Y_{j2}^C - Y_{j1}^C),$$

donde $\omega(i, j)$ son los pesos PSM (NN, kernel, LLM, etc.).

PSM + DD en marco de regresión

- Hirano, Imbens y Ridder (2003) muestran que una regresión ponderada puede ser totalmente eficiente:

$$\Delta Y_i = \alpha + \beta T_i + \gamma X_i + \varepsilon_i, \quad \beta = DD.$$

- Pesos:
 - ▶ Unidades tratadas: peso 1.
 - ▶ Unidades de comparación: peso $\hat{P}(X)/[1 - \hat{P}(X)]$.
- Ventaja:
 - ▶ Integra selección sobre observables (vía PSM) y corrección por heterogeneidad fija (vía diferencias).

Contenido

- 1 Motivación y comparación con otros métodos
- 2 Sesgo de selección y diferencias como contrafactual
- 3 Método DD: teoría y regresión
- 4 Modelo de efectos fijos en panel
- 5 Implementación del método DD
- 6 Ventajas, limitaciones y violaciones a tendencias paralelas
- 7 Modelos alternativos basados en DD**
 - PSM con DD
 - **Método de triple diferencia (DDD)**
 - Ajuste por tendencias temporales diferenciales

Triple diferencia: motivación

- Problema:
 - ▶ No se dispone de línea de base (por ejemplo, programas creados durante crisis).
- Solución: **triple diferencia (DDD)**:
 - ▶ Además del par tratado–control original, se introduce un **segundo grupo de control** posterior a la intervención.
 - ▶ Este grupo adicional permite netear:
 - ★ Cambios agregados en el tiempo.
 - ★ Diferencias estructurales entre grupos socioeconómicos.
- Requiere:
 - ▶ Múltiples años de datos post-intervención.
 - ▶ Estructura de “dos controles”: uno para el programa, otro para los choques agregados.

Ejemplo de triple diferencia

- Caso de un programa laboral sin línea de base:
 - ▶ Comparar ingresos de “stayers” (permanecen en el programa) y “leavers” (lo abandonan).
 - ▶ Sin grupo adicional, DD entre “stayers” y “leavers” puede estar sesgada:
 - ★ Leavers suelen tener mejores oportunidades de ingreso fuera del programa.
- DDD:
 - ▶ Introduce un grupo de no participantes emparejado.
 - ▶ Resta de la DD entre “stayers” y “leavers” la DD correspondiente en el grupo adicional.
 - ▶ Permite aislar el componente que refleja solo el efecto del programa.

Contenido

- 1 Motivación y comparación con otros métodos
- 2 Sesgo de selección y diferencias como contrafactual
- 3 Método DD: teoría y regresión
- 4 Modelo de efectos fijos en panel
- 5 Implementación del método DD
- 6 Ventajas, limitaciones y violaciones a tendencias paralelas
- 7 Modelos alternativos basados en DD**
 - PSM con DD
 - Método de triple diferencia (DDD)
 - **Ajuste por tendencias temporales diferenciales**

DD ajustado por tendencias temporales diferenciales

- Problema:
 - ▶ En presencia de choques macroeconómicos, tratados y controles pueden responder con tendencias diferentes en el tiempo.
 - ▶ El supuesto de tendencia paralela se rompe.
- Propuesta (Bell, Blundell y van Reenen, 1999):
 - ▶ Utilizar información de otro intervalo temporal previo ($t - 2, t - 1$).
 - ▶ Aproximar la tendencia diferencial pre-tratamiento y descontarla.

Fórmula de DD ajustado por tendencia

- Sea el efecto DD “estándar” entre 0 y 1:

$$DD_{01} = E(Y_1^T - Y_0^T \mid T_1 = 1) - E(Y_1^C - Y_0^C \mid T_1 = 0).$$

- Sea el DD “pretratamiento” entre $t - 2$ y $t - 1$:

$$DD_{t-2,t-1} = E(Y_{t-1}^T - Y_{t-2}^T \mid T_1 = 1) - E(Y_{t-1}^C - Y_{t-2}^C \mid T_1 = 0).$$

- DD ajustado por tendencia:

$$DD^* = DD_{01} - DD_{t-2,t-1} \tag{6}$$

$$= [E(Y_1^T - Y_0^T \mid T_1 = 1) - E(Y_1^C - Y_0^C \mid T_1 = 0)] \tag{7}$$

$$- [E(Y_{t-1}^T - Y_{t-2}^T \mid T_1 = 1) - E(Y_{t-1}^C - Y_{t-2}^C \mid T_1 = 0)]. \tag{8}$$

- Intuición:

- ▶ Se elimina la tendencia diferencial previa entre tratados y controles.
- ▶ El residuo se interpreta como impacto del programa.