Diferencias en Diferencias

2025-06-19

Motivación

Hemos visto cómo estimar efectos causales cuando el tratamiento es casi aleatorio, condicionado a características observables.

Pero a menudo nos preocupan características no observables (variables confusoras).

Veremos cómo tratar ciertos tipos de confusores no observados usando datos de panel.

¿Qué son los datos de panel?

Datos de panel: observaciones repetidas de unidades i (personas, estados, etc.) a lo largo del tiempo t.

Utilidad: permite comparar resultados entre unidades tratadas y no tratadas *antes* del tratamiento.

Diferencias previas al tratamiento pueden revelar factores confusos **no observados**.

Hastings (2004)

En 2004, Justine Hastings escribió un estudio que analiza cómo las fusiones en la industria de la gasolina afectan los precios del combustible.

En particular, estudió un episodio en California donde la refinería ARCO compró una de las mayores cadenas de estaciones de servicio, Thrifty.

¿Cómo crees que una fusión así podría afectar los precios?

Hastings (2004)

En 2004, Justine Hastings escribió un estudio que analiza cómo las fusiones en la industria de la gasolina afectan los precios del combustible.

En particular, estudió un episodio en California donde la refinería ARCO compró una de las mayores cadenas de estaciones de servicio, Thrifty.

¿Cómo crees que una fusión así podría afectar los precios?

- Por un lado, podría reducir la competencia y aumentar los precios.
- Por otro lado, la fusión podría reducir los costos de proveer gasolina y, gracias a las sinergias, disminuir los precios.

Hastings (2004)

Hastings intentó responder a esta pregunta de forma empírica usando datos de precios de gasolina por vecindario en California.

El conjunto de datos contenía información sobre vecindarios tanto con como sin estaciones Thrifty.

¿Si sólo tuviéramos datos posteriores a la fusión?

Supongamos primero que sólo contáramos con datos de precios de la gasolina *posteriores* a la fusión.

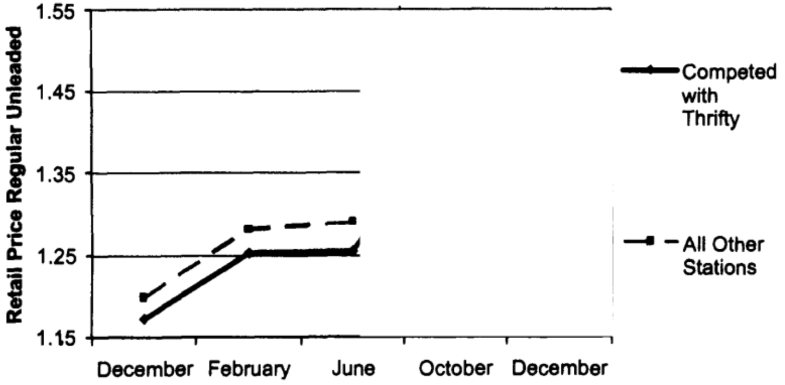
Podríamos comparar los precios en zonas que tenían una Thrifty antes ($D_i = 1$) con zonas que no la tenían ($D_i = 0$) para estimar el efecto causal de convertir una Thrifty.

¿Por qué esto podría no darnos el efecto causal? ¡Variables omitidas!

Los lugares que ya tenían una Thrifty antes probablemente tenían mayor competencia que los lugares sin Thrifty, por lo que esperaríamos que sus precios fueran más bajos.

Con datos de panel, podemos comprobar esto empíricamente examinando los precios *antes* de la fusión.

Diferencias pre-tratamiento



(a) Los Angeles

Diferencias pre-tratamiento

Antes de la fusión, las estaciones en mercados que competían con Thrifty tenían precios de la gasolina ≈ 3 centavos más bajos en cada período.

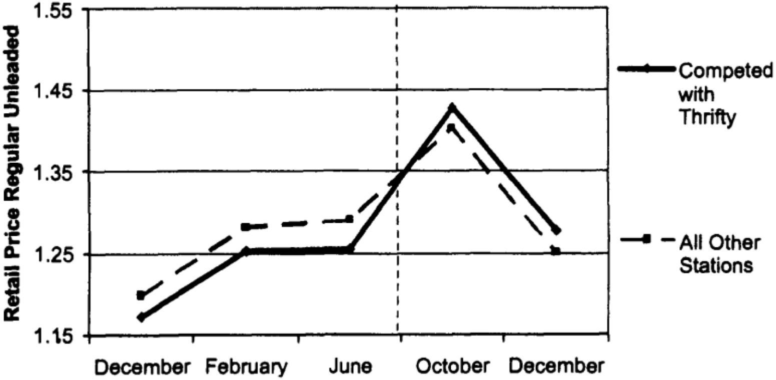
¿Es razonable asumir ausencia de confusión después de la fusión?

¡No!

Un supuesto más creíble es que la brecha hubiera permanecido en 3 c si la fusión no hubiera ocurrido.

¡Esta es la idea de la *diferencia-en-diferencias* (DiD)!

Después de la fusión



(a) Los Angeles

Después de la fusión

- Después de la fusión, las estaciones ubicadas en zonas con una Thrifty presentaron precios aproximadamente 2 centavos más altos.
- Si suponemos que habrían tenido precios 3 centavos más bajos (como antes de la fusión), ello implica un efecto del tratamiento de $2 - (-3) = 5$ centavos.
- Esta cifra corresponde a la diferencia post-tratamiento (2 centavos) entre tratadas y controles menos la diferencia pre-tratamiento (-3 centavos); es decir, una *diferencia-en-diferencias*.

Formalizando los supuestos de DiD

Suponga que hay dos períodos, $t = 1, 2$.

Las unidades tratadas ($D_i = 1$) reciben el tratamiento en el período 2; las unidades de control nunca son tratadas.

Sea Y_{it} el resultado observado para la unidad i en el período t.

Suponemos

$$Y_{it} = D_i Y_{it}(1) + (1 - D_i) Y_{it}(0)$$

Supuesto de no anticipación:

$$Y_{i1}(0) = Y_{i1}(1)$$

El tratamiento en el período 2 no afecta el resultado en el período 1.

Formalizando los supuestos de DiD

Supuesto de *tendencias paralelas*:

$$\underbrace{\mathbb{E}[\, Y_{i2}(0) - Y_{i1}(0) \mid D_i = 1\,]}_{\text{Cambio en } Y(0) \text{ para tratados}} = \underbrace{\mathbb{E}[\, Y_{i2}(0) - Y_{i1}(0) \mid D_i = 0\,]}_{\text{Cambio en } Y(0) \text{ para controles}}$$

Equivalente a

$$\underbrace{\mathbb{E}[\, Y_{i2}(0) \mid D_i = 1\,] - \mathbb{E}[\, Y_{i2}(0) \mid D_i = 0\,]}_{\text{Sesgo de selección en el período 2}} = \underbrace{\mathbb{E}[\, Y_{i1}(0) \mid D_i = 1\,] - \mathbb{E}[\, Y_{i1}(0) \mid D_i = 0\,]}_{\text{Sesgo de selección en el período 1}}$$

Demostración

$$\underbrace{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1]}_{\text{Cambio observado (tratados)}} \; - \; \underbrace{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0]}_{\text{Cambio observado (controles)}} =$$

$$= \mathbb{E}[Y_{i2}(1) - Y_{i1}(1) \mid D_i = 1] \; - \; \mathbb{E}[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 0] \quad \text{(regla de datos observados)}$$

$$= \mathbb{E}[Y_{i2}(1) - Y_{i1}(0) \mid D_i = 1] \; - \; \mathbb{E}[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 0] \quad \text{(sin anticipación)}$$

$$= \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1]$$

$$+ \; \big[\mathbb{E}[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 1] \; - \; \mathbb{E}[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 0]\big] \quad \text{(sumar y restar)}$$

$$= \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1] \quad \text{(tendencias paralelas)}$$

$$\implies \boxed{\tau_{\text{ATT}} = \mathbb{E}[\, Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1\,]}.$$

Interpretación

τ_{ATT} es el **efecto medio del tratamiento sobre los tratados**: el efecto promedio en el período 2 para las unidades que recibieron el tratamiento.

Estimando el ATT

Demostramos que, bajo los supuestos de DiD (tendencias paralelas y sin anticipación), el ATT se identifica como

$$\tau_{\text{ATT}} = \underbrace{\mathbb{E}[\, Y_{i2} - Y_{i1} \mid D_i = 1\,]}_{\text{Cambio en la media poblacional (tratados)}} - \underbrace{\mathbb{E}[\, Y_{i2} - Y_{i1} \mid D_i = 0\,]}_{\text{Cambio en la media poblacional (controles)}} .$$

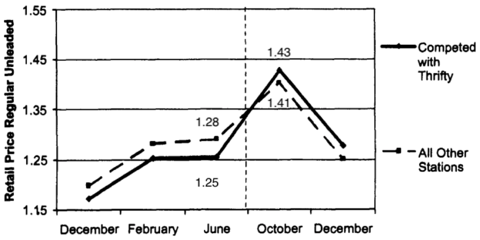
¿Cómo lo estimamos?

¡Sustituyendo medias muestrales!

Nuestro estimador es

$$\hat{\tau}_{\text{ATT}} = \underbrace{\bar{Y}_{12} - \bar{Y}_{11}}_{\text{Cambio en la media muestral (tratados)}} - \underbrace{\bar{Y}_{02} - \bar{Y}_{01}}_{\text{Cambio en la media muestral (controles)}} ,$$

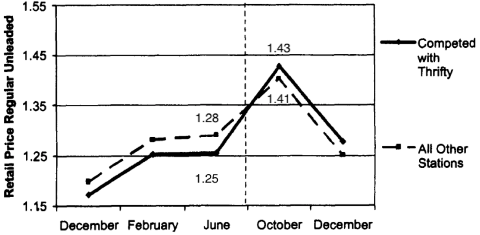
Ejemplo numérico



(a) LOS ANGELES

$$\hat{\tau}_{\text{ATT}} = \underbrace{\bar{Y}_{12} - \bar{Y}_{11}}_{\text{Cambio en la media muestral (tratados)}} - \underbrace{\bar{Y}_{02} - \bar{Y}_{01}}_{\text{Cambio en la media muestral (controles)}} =$$

Ejemplo numérico



(a) LOS ANGELES

$$
\begin{aligned}
\hat{\tau}_{\text{ATT}} &= \underbrace{\bar{Y}_{12} - \bar{Y}_{11}}_{\text{Cambio en la media muestral (tratados)}} - \underbrace{\bar{Y}_{02} - \bar{Y}_{01}}_{\text{Cambio en la media muestral (controles)}} = \\
&= (1.43 - 1.25) - (1.41 - 1.28) = \\
&= 0.05.
\end{aligned}
$$

DiD como regresión

Considere la regresión

$$Y_{it} = \beta_0 + \beta_1 \, \text{Post}_t + \beta_2 D_i + \beta_3 \, D_i \times \text{Post}_t + \varepsilon_{it}$$

donde

$$\text{Post}_t = \mathbf{1}[t = 2].$$

El coeficiente poblacional β_3 es igual a τ_{ATT} bajo los supuestos de DiD.

DiD como regresión

El modelo anterior implica la CEF:

$$
\begin{aligned}
\mathbb{E}[Y_{it} \mid D_i = 0, \text{Post}_t = 0] &= \beta_0, \\
\mathbb{E}[Y_{it} \mid D_i = 0, \text{Post}_t = 1] &= \beta_0 + \beta_1, \\
\mathbb{E}[Y_{it} \mid D_i = 1, \text{Post}_t = 0] &= \beta_0 + \beta_2, \\
\mathbb{E}[Y_{it} \mid D_i = 1, \text{Post}_t = 1] &= \beta_0 + \beta_1 + \beta_2 + \beta_3.
\end{aligned}
$$

Por lo tanto,

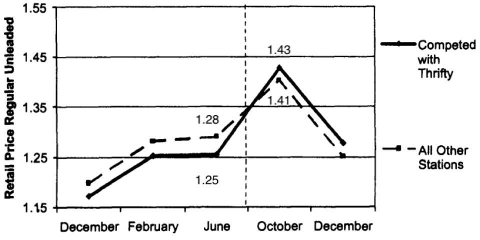
$$
\begin{aligned}
\beta_3 = \big(&\mathbb{E}[Y_{it} \mid D_i = 1, \text{Post}_t = 1] - \mathbb{E}[Y_{it} \mid D_i = 1, \text{Post}_t = 0]\big) \\
-\big(&\mathbb{E}[Y_{it} \mid D_i = 0, \text{Post}_t = 1] - \mathbb{E}[Y_{it} \mid D_i = 0, \text{Post}_t = 0]\big) = \tau_{\text{ATT}}.
\end{aligned}
$$

DiD como regresión

De manera análoga con datos muestrales:

$$\hat{\beta}_3 \; = \; (\bar{Y}_{12} - \bar{Y}_{11}) \; - \; (\bar{Y}_{02} - \bar{Y}_{01}) \; = \; \hat{\tau}_{\mathsf{ATT}}.$$

Ejemplo con Datos de Hastings



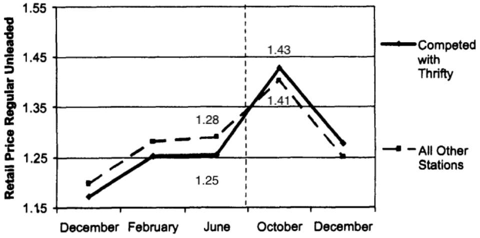
(a) LOS ANGELES

Estimamos la siguiente regresión vía MCO:

$$Y_{it} = \beta_0 + \beta_1 \cdot \text{Post}_t + \beta_2 D_i + \beta_3 (D_i \cdot \text{Post}_t) + \varepsilon_{it}$$

donde $\text{Post}_t = 1$ para octubre y 0 para junio. Los coeficientes estimados son:

Ejemplo con Datos de Hastings



(a) Los Angeles

- Constante ($\hat{\beta}_0$) = 1.28
- Post ($\hat{\beta}_1$) = 0.13
- Tratado ($\hat{\beta}_2$) = -0.03
- Tratado $\times$ Post ($\hat{\beta}_3$) = 0.05

DID: Resumen hasta ahora

El estimador de Diferencias en Diferencias (DID) se puede escribir como: la diferencia entre grupos, de la diferencia a lo largo del tiempo:

$$\hat{\tau}_{\text{ATT}} = (\bar{Y}_{12} - \bar{Y}_{11}) - (\bar{Y}_{02} - \bar{Y}_{01})$$

Los supuestos clave son:

- 1) No anticipación: el efecto del tratamiento en el período cero es nulo.
- 2) Tendencias paralelas: el grupo tratado habría tenido el mismo *cambio* en los resultados que el grupo de control, en ausencia del tratamiento.

Podemos estimar el efecto del tratamiento por MCO (Mínimos Cuadrados Ordinarios):

$$Y_{it} = \beta_0 + \beta_1 \cdot \text{Post}_t + \beta_2 D_i + \beta_3 (D_i \cdot \text{Post}_t) + \varepsilon_{it}$$

El efecto del tratamiento corresponde a β_3.

DID con múltiples períodos

DID con múltiples períodos

El DID con dos períodos y dos grupos es bastante sencillo. Pero a menudo queremos usar datos de muchos períodos.

Supongamos que tenemos un grupo tratado y uno de control observados desde $t = -t_a$ hasta $+t_b$. El tratamiento ocurre en el período 1 (la notación anterior tenía el tratamiento en 2).

Para cada período s, podemos estimar un efecto del tratamiento usando la estrategia DID de 2x2, fijando $t = 0$ como referencia.

$$\beta_s = \underbrace{(\bar{Y}_{1s} - \bar{Y}_{0s})}_{\text{Dif en el período s}} - \underbrace{(\bar{Y}_{10} - \bar{Y}_{00})}_{\text{Dif en el período 0}}$$

DID con múltiples períodos

Convenientemente, podemos estimar los mismos efectos con una regresión:

$$Y_{it} = \phi_t + \gamma D_i + \sum_{s \neq 0} \beta_s \times D_i \times 1[t = s] + \varepsilon_{it}$$

Este tipo de regresión a veces se denomina "efectos fijos de dos vías" (two-way fixed effects), porque tenemos dummies para el tiempo (ϕ_t) y el grupo (γ).

Alternativamente, podríamos sustituir γD_i por efectos fijos individuales λ_i y obtener las mismas estimaciones.

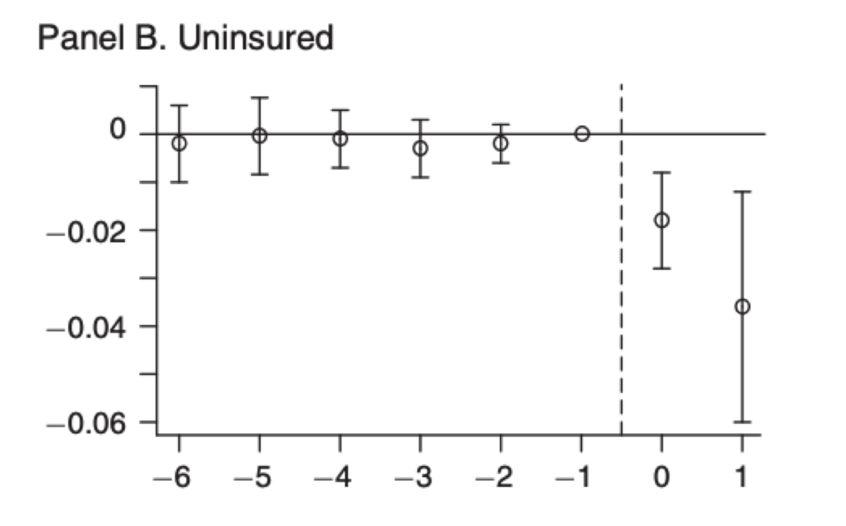
Ejemplo: Expansión de Medicaid

La Ley del Cuidado de Salud a Bajo Precio (Affordable Care Act) amplió la cobertura médica para personas de bajos ingresos.

La expansión entró en vigencia en 2014, pero algunos estados optaron por no participar y no tuvieron cambios.

Carey, Miller y Wherry (2020) estudian el impacto de la expansión de Medicaid mediante un análisis DID. Comparan los primeros estados en adoptar la medida con aquellos que no lo hicieron.

Veamos las estimaciones de β_s con sus IC para cada punto en el tiempo:



Tendencias Paralelas

Nótese que el gráfico anterior incluye “efectos del tratamiento” estimados para períodos anteriores al tratamiento. ¿Por qué es importante mostrarlos?

Tendencias Paralelas

Nótese que el gráfico anterior incluye "efectos del tratamiento" estimados para períodos anteriores al tratamiento. ¿Por qué es importante mostrarlos?

El supuesto de tendencias paralelas es que, sin el tratamiento, las diferencias entre los grupos deberían permanecer constantes a lo largo del tiempo. Por lo tanto, en los períodos anteriores al tratamiento deberíamos ver que el coeficiente es igual a cero.

Tendencias Paralelas

Una forma de pensar en el supuesto de tendencias paralelas es en términos de una ecuación en diferencias.

Si tomamos la forma de regresión del DID simple de 2x2 y restamos Y_{i0}:

$$(Y_{i1} - Y_{i0}) = \beta_1 + \beta_3 D_i + (\varepsilon_{i1} - \varepsilon_{i0})$$

Esto es simplemente una comparación de medias entre los grupos tratado y de control **en diferencias**.

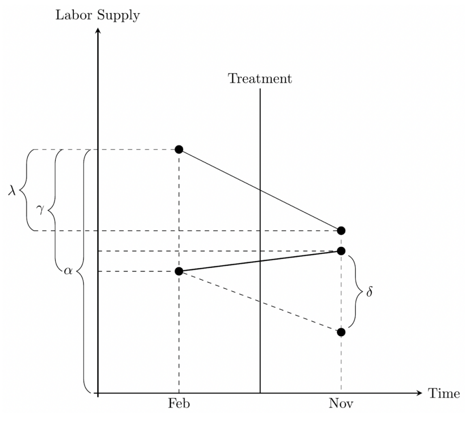
Tendencias Paralelas

En la forma de regresión en diferencias, el supuesto de tendencias paralelas es simplemente la independencia del tratamiento.

Estamos *asumiendo* que no hay factores de confusión, solo que en lugar de los niveles, lo estamos asumiendo para los cambios.

Tendencias Paralelas

$$Y_{it} = \alpha_0 + \gamma \cdot \mathrm{D}_t + \lambda Post_i + \delta(D_i \cdot \mathrm{Post}_t) + \varepsilon_{it}$$



Tendencias Paralelas

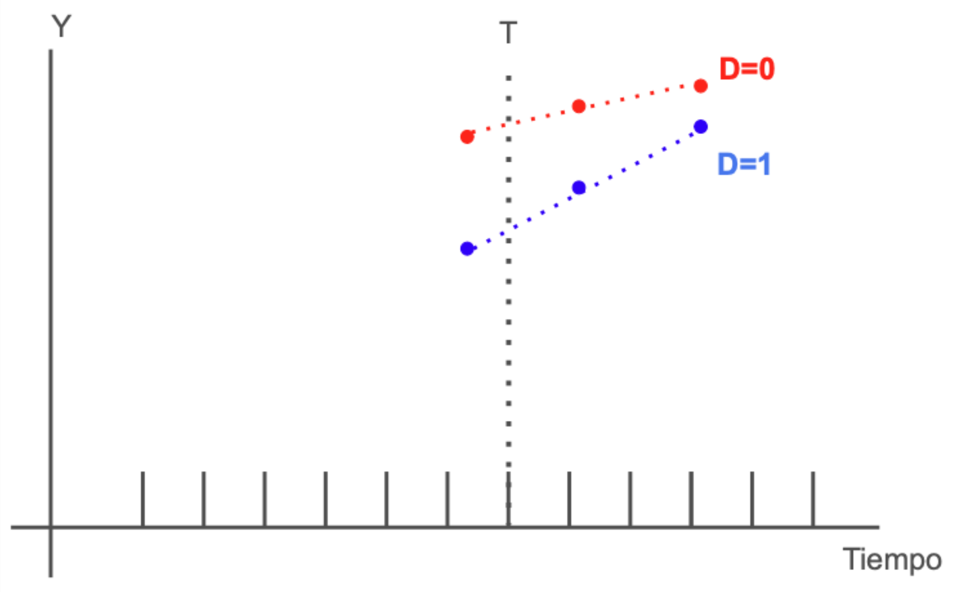
Podemos pensar en el método DID como un control efectivo para cualquier **factor de confusión invariante en el tiempo**, así como para cualquier factor que sea **exactamente común a los grupos tratado y no tratado**.

Por ejemplo:

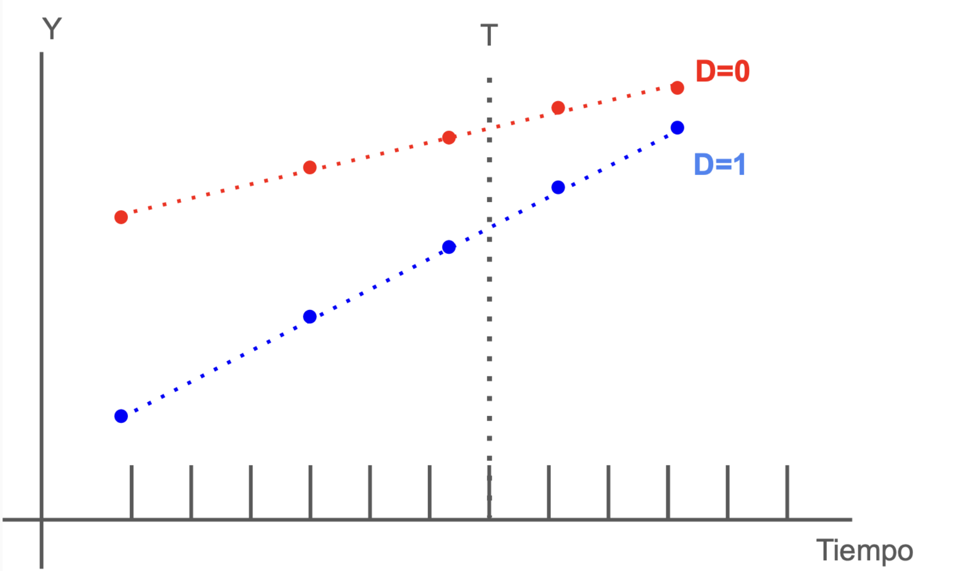
- Los estados que adoptaron la expansión de Medicaid son más demócratas, más educados, tienen ingresos promedio más altos.
  - Los factores que son relativamente "fijos" son controlados por las diferencias en el tiempo.
- El período "posterior" incluye diferentes condiciones macroeconómicas.
  - Los factores que cambian con el tiempo, pero que son iguales para todos los estados, son controlados por las diferencias entre grupos.

El supuesto de tendencias paralelas se violaría si, por ejemplo, las condiciones macroeconómicas tendieran a beneficiar a los estados republicanos en comparación con

Tendencias paralelas



Tendencias paralelas



Tendencias Paralelas

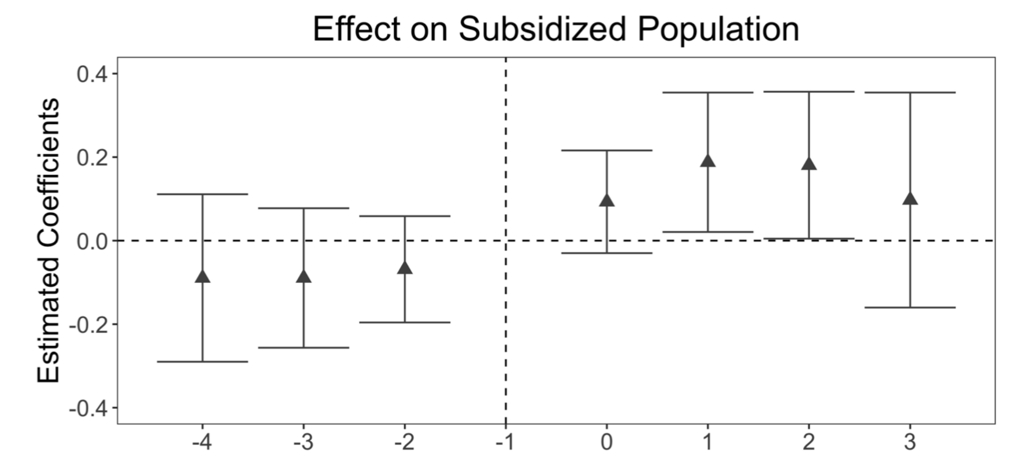
En este caso, ¡estimaríamos "efectos del tratamiento" para cada período, incluso antes de que comenzara el tratamiento!

Por lo tanto, queremos ver efectos nulos en los períodos "anteriores" para asegurarnos de que los grupos seguían una tendencia paralela antes del tratamiento.

Tendencias Paralelas

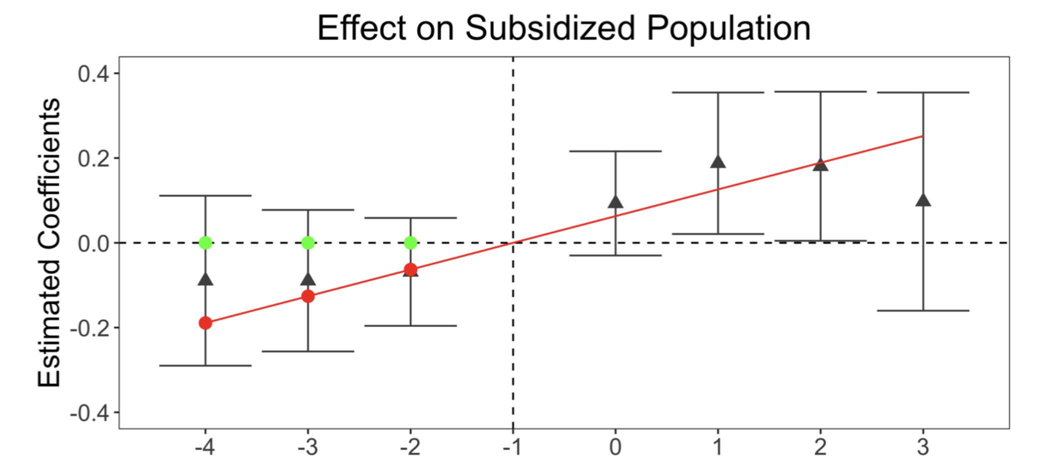
También deberíamos considerar lo que los intervalos de confianza descartan.

¿Este efecto parece convincente?

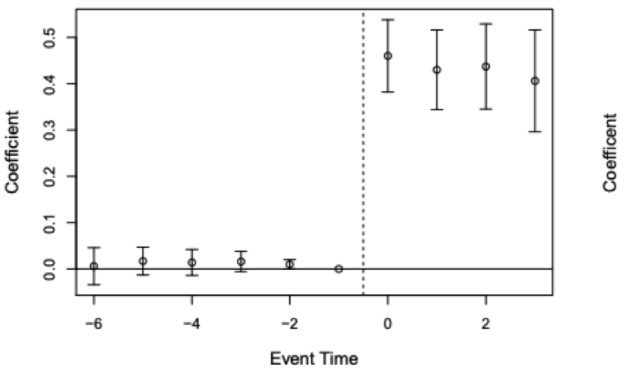


También deberíamos considerar lo que los intervalos de confianza descartan.

¿Este efecto parece convincente?



¿Qué tal este?



(a) Medicaid Eligibility

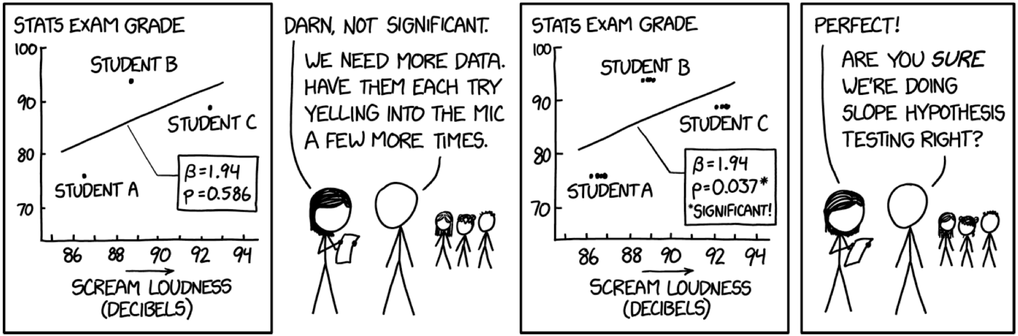
Inferencia

Inferencia

Una preocupación muy importante al estimar un DID es que los errores deben ser agrupados (clustered).

Cuando discutimos la agrupación, vimos que cuando los errores están correlacionados, tenemos que ajustar los errores para reflejar eso. De lo contrario, tendremos errores estándar que son demasiado pequeños.

En datos de panel, tenemos muchas observaciones de un mismo individuo a lo largo del tiempo. Naturalmente, hay características que están correlacionadas a lo largo del tiempo.


STATS EXAM GRADE
100
90
80
70
STUDENT B
STUDENT C
STUDENT A
β=1.94
p=0.586
86 88 90 92 94
SCREAM LOUDNESS (DECIBELS)
DARN, NOT SIGNIFICANT.
WE NEED MORE DATA. HAVE THEM EACH TRY YELLING INTO THE MIC A FEW MORE TIMES.
STATS EXAM GRADE
100
90
80
70
STUDENT B
STUDENT C
STUDENT A
β=1.94
p=0.037*
*SIGNIFICANT!
86 88 90 92 94
SCREAM LOUDNESS (DECIBELS)
PERFECT!
ARE YOU *SURE* WE'RE DOING SLOPE HYPOTHESIS TESTING RIGHT?

Inferencia

The Quarterly Journal of Economics, February 2004

HOW MUCH SHOULD WE TRUST DIFFERENCES-IN-DIFFERENCES ESTIMATES?*

MARIANNE BERTRAND
ESTHER DUFLO
SENDHIL MULLAINATHAN

Most papers that employ Differences-in-Differences estimation (DD) use many years of data and focus on serially correlated outcomes but ignore that the resulting standard errors are inconsistent. To illustrate the severity of this issue, we randomly generate placebo laws in state-level data on female wages from the Current Population Survey. For each law, we use OLS to compute the DD estimate of its "effect" as well as the standard error of this estimate. These conventional DD standard errors severely understate the standard deviation of the estimators: we find an "effect" significant at the 5 percent level for up to 45 percent of the placebo interventions. We use Monte Carlo simulations to investi-

Inferencia

Lección: siempre agrupar (cluster) los errores estándar a nivel del individuo.

DID con múltiples tratamientos

DID con múltiples tratamientos

A veces tenemos diferentes unidades que son tratadas en diferentes momentos en el tiempo.

Por ejemplo: varios estados pueden adoptar una determinada política en diferentes años.

Supongamos que, una vez que se adopta el tratamiento, permanece “activo”. Todas las unidades comienzan con el tratamiento “desactivado”.

DID con múltiples tratamientos

En este caso, es usual estimar el efecto del tratamiento usando una regresión como esta:

$$Y_{it} = \phi_i + \lambda_t + \beta D_{it} + e_{it}$$

Esto generalmente se llama DID con "adopción escalonada" (staggered adoption).

DID con múltiples tratamientos

Sin embargo, trabajos recientes han demostrado que esta estrategia a veces no funciona muy bien.

Por razones que son un poco complicadas, esta regresión puede dar lugar a resultados que no corresponden a un promedio de los efectos del tratamiento, especialmente si:

- Hay **efectos dinámicos**: los efectos del tratamiento aumentan/disminuyen con el tiempo.
- Hay heterogeneidad en el efecto del tratamiento: diferentes unidades tienen diferentes efectos del tratamiento.

DID con múltiples tratamientos

Hay muchas formas de abordar el problema, pero la mayoría se basa en hacer comparaciones claras.

Entonces, una estrategia básica que sí funciona es:

- 1) Separar los diferentes grupos: el grupo que nunca es tratado y los grupos tratados en cada momento del tiempo.
- 2) Usar el enfoque de regresión utilizando solo el grupo nunca tratado frente al primer grupo tratado. Obtener los efectos del tratamiento.
- 3) Repetir para cada grupo diferente, siempre comparando con el grupo nunca tratado.
- 4) Promediar los efectos del tratamiento.

Este es esencialmente el enfoque de Callaway y Sant'Anna.

DID resumen

Usamos diferencias en diferencias cuando tenemos un grupo de comparación que **no es necesariamente** igual que el tratado.

Si obervamos los dos grupos antes y después del tratamiento, podemos usar la hipótesis de tendencias paralelas para crear el contrafactual del grupo de tratamiento.

Podemos estimar el efecto de tratamiento con diferencias de medias, o usando regresión.

Las hipótesis clave son:

- No Antecipación
- Tendencias Paralelas.

Siempre agrupar los errores estándar.

Cuándo tenemos adopción escalonada, es posible utilizar la estratégia, pero tener más cuidado.