

Control Sintético

2026-08-03

Del DID al control sintético

¿Dónde estábamos?

DID compara la **evolución** de un grupo tratado con la de un control, apoyado en **tendencias paralelas**.

Pero eso supone que tenemos un **buen grupo de control**: uno que se movía igual que el tratado antes del tratamiento.

¿Qué hacemos cuando...

- el tratamiento afecta a **una sola unidad grande** — un estado, un país — y
- **ningún control individual** se le parece lo suficiente?

El caso motivador: **Proposición 99**

En **1988**, California aprobó la **Proposición 99**: un fuerte impuesto al tabaco (25¢ por cajetilla) + campañas antitabaco.

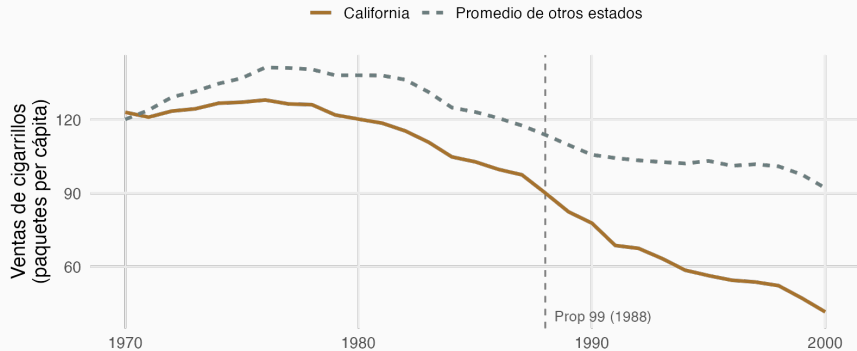
Queremos el efecto sobre el consumo de cigarrillos (paquetes per cápita).

El problema de siempre, pero agudo: **¿cuál es la California contrafactual?**

- No podemos aleatorizar una ley estatal.
- Ningún otro estado es “como California sin Prop 99”.

Un solo control no sirve

¿Y si comparamos California con el **promedio** de los demás estados (un DID ingenuo)?



Un solo control no sirve

El promedio simple **no sigue a California** antes de 1988: California ya venía más abajo y con otra pendiente.

Si los grupos no eran paralelos **antes**, no podemos creerles **después**.

Idea: en vez de un control cualquiera (o un promedio parejo), **construyamos** uno a la medida.

La idea del control sintético

Una California “sintética”

Combinemos varios estados donantes en un **promedio ponderado** — la **California sintética** — eligiendo los pesos para que **reproduzca a California antes** de la Prop 99.

$$\text{California sintética}_t = \sum_{j \in \text{donantes}} w_j Y_{jt}, \quad w_j \geq 0, \quad \sum_j w_j = 1$$

- Abadie & Gardeazabal (2003); Abadie, Diamond & Hainmueller (2010).
- Los pesos se eligen con los datos **pre-tratamiento**, no a mano.

El puente con DID

El control sintético es un **DID con pesos elegidos**:

DID:	control = una unidad (o promedio parejo); <i>suponemos</i> tendencias paralelas.
Control sintético:	control = promedio ponderado de donantes; <i>elegimos</i> los pesos para que las trayectorias pre <i>de hecho</i> coincidan.

En vez de **suponer** un buen contrafactual, lo **construimos** y lo mostramos.

Dos restricciones que importan

$$w_j \geq 0, \quad \sum_j w_j = 1$$

- **Pesos no negativos y que suman 1** $\Rightarrow$ la sintética es una **combinación convexa** de estados reales.
- Esto **prohíbe extrapolar**: la sintética vive *dentro* del rango de los donantes, nunca fuera.
- Transparencia: sabemos **qué estados** y con **qué peso** forman el contrafactual.

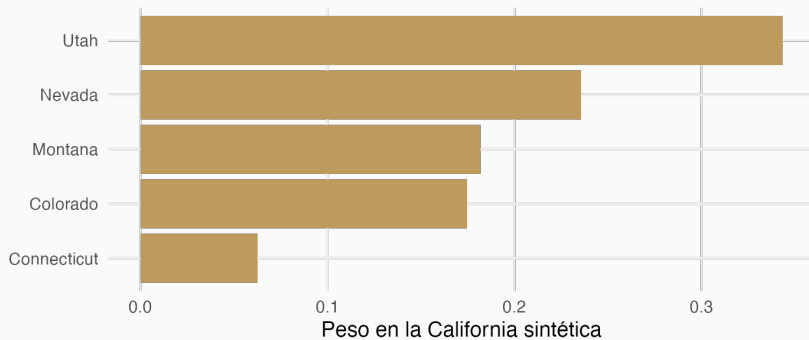
Cómo se construye

Los ingredientes

1. **Unidad tratada:** California.
2. **Grupo de donantes:** estados **sin** su propia política antitabaco fuerte (38 estados).
3. **Predictores X :** variables pre-1988 que predicen el consumo — ingreso, precio, % jóvenes, cerveza, y **algunos rezagos del propio consumo**.
4. **Pesos W :** elegidos para minimizar la distancia pre-tratamiento

$$\min_W \|X_{\text{California}} - X_{\text{donantes}} W\|_V \quad \text{s.a. } w_j \geq 0, \sum_j w_j = 1$$

¿Qué estados forman la California sintética?



Cinco estados cargan casi todo el peso: **Utah, Nevada, Montana, Colorado, Connecticut**. El resto: peso cero.

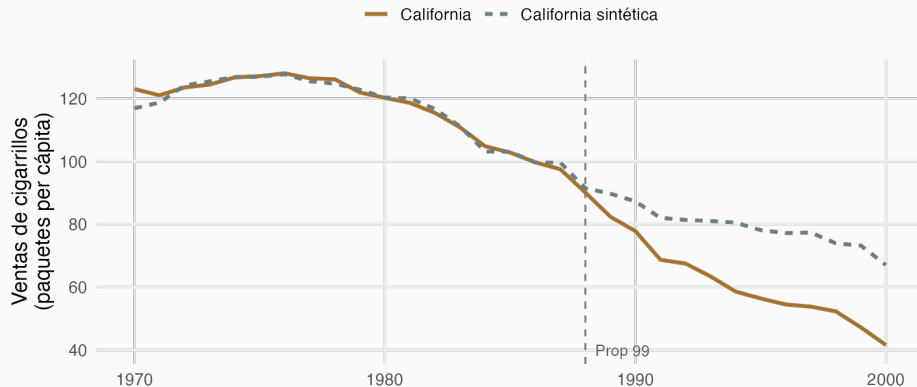
¿Quedó bien emparejada?

Predictor	California	Sintética	Prom. donantes
Ingreso (log)	10.1	10.1	9.8
Precio minorista	89.4	89.3	87.3
% 15–24 años	0.17	0.17	0.17
Cerveza per cápita	24.3	24.1	23.7
Ventas 1975	127	127	137
Ventas 1980	120	120	138
Ventas 1988	90	91	114

KEY POINT: la **sintética** calza a California; el **promedio simple** de donantes (última columna) no — sobre todo en el consumo, que es lo que importa.

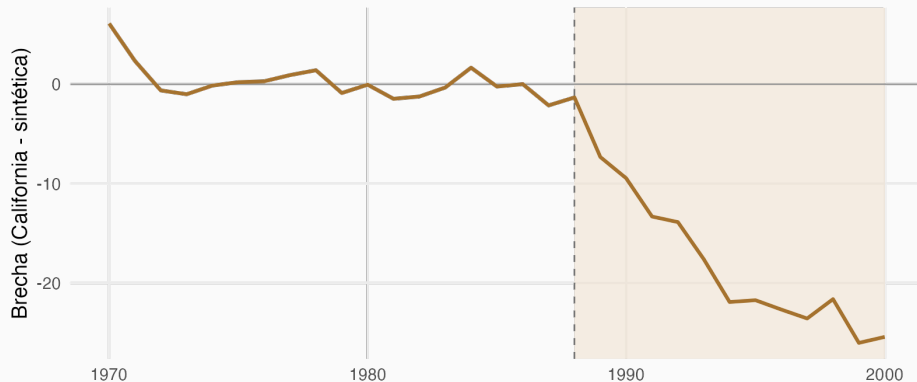
Resultados

California vs. California sintética



Antes de 1988: casi idénticas (buen ajuste). Después: California cae **muy por debajo** de su sintética.

El efecto estimado = la brecha



La brecha (California — sintética) es plana en cero antes, y se abre después: el efecto de la Prop 99.

El efecto estimado = la brecha

Lectura

- Brecha casi cero antes de 1988 $\Rightarrow$ el contrafactual es creíble.
- Brecha media post-1988: ≈ -19 paquetes per cápita al año.
- Hacia 2000: ≈ -26 paquetes — casi un tercio del consumo.

El estimador es simplemente

$$\hat{\tau}_t = Y_{\text{California},t} - \sum_j w_j^* Y_{jt}, \quad t > 1988.$$

Inferencia

El problema: una sola unidad tratada

Solo hay **una** California. No hay muestra de unidades tratadas $\Rightarrow$ los errores estándar clásicos no aplican.

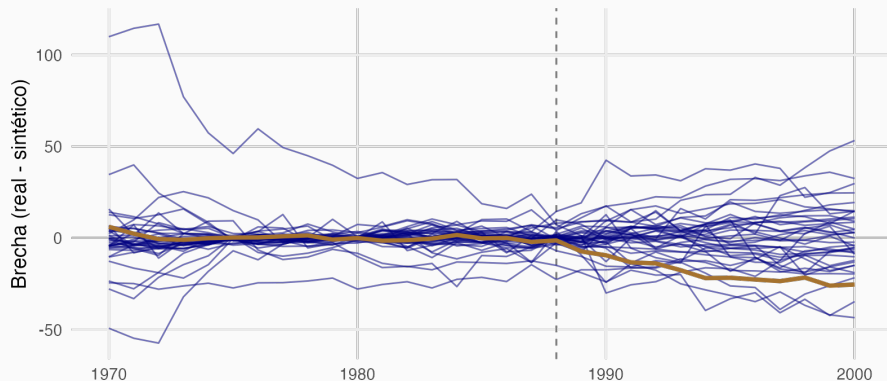
Pregunta que **sí** podemos responder: ¿podría una brecha así de grande aparecer **por azar**?

Inferencia por permutación (placebos)

Idea: démosle el “tratamiento” a **cada donante** (que no fue tratado) y estimemos su brecha.

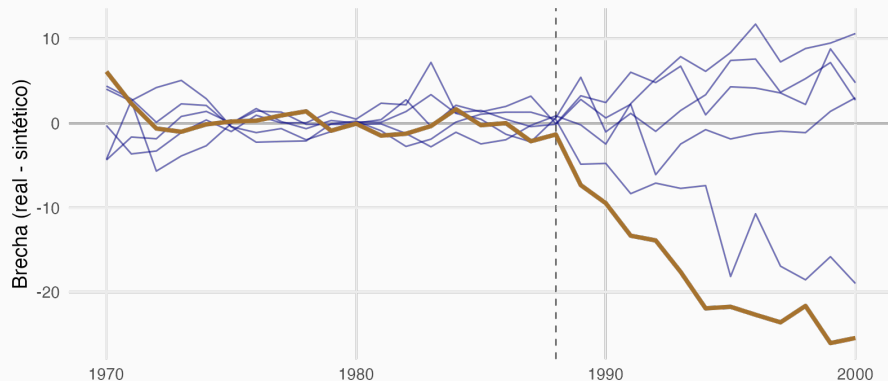
- Corremos el control sintético para cada estado, fingiendo que su Prop 99 fue en 1988.
- Obtenemos una **distribución de brechas placebo**: cómo se ve el efecto *cuando no hay efecto*.
- Si California es **más extrema** que casi todos los placebos, difícilmente es azar.

Placebos: todos los estados



California (línea gruesa) cae mucho... pero algunos placebos también saltan: son estados que la sintética **nunca ajustó bien** antes de 1988.

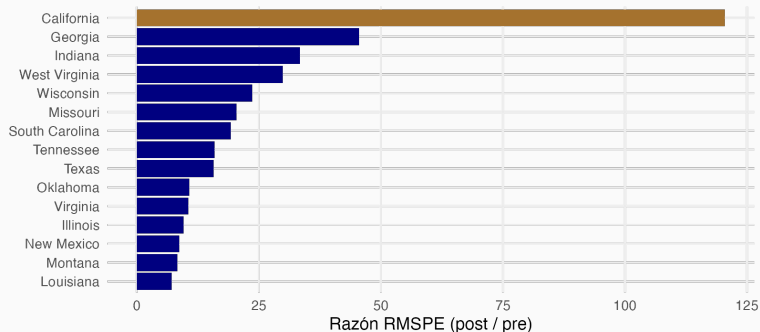
Placebos: descartando el mal ajuste



Quitando los donantes con mal ajuste pre-1988 ($\text{RMSPE pre} > 2 \times \text{la de California}$), California queda **claramente sola** en el extremo.

Un p-valor por ranking

Resumimos cada estado por su **razón RMSPE** = ajuste post / ajuste pre. Un valor alto = brecha grande que **no** venía de un mal ajuste previo.



Un p-valor por ranking

California tiene la razón RMSPE **más alta de los 39 estados** (rank 1).

$$p = \frac{\text{rank}}{N} = \frac{1}{39} \approx 0,026$$

KEY POINT: la inferencia es por **permutación**, no por muestreo: preguntamos qué tan raro es el resultado de California dentro de los propios datos.

Cuándo (y cuándo no) usarlo

- **El ajuste pre lo es todo.** Si la sintética no reproduce el pasado, no le creas el futuro.
- **No extrapola.** La combinación convexa vive dentro del rango de los donantes: si el tratado es extremo, ningún peso lo alcanza (**sesgo de interpolación**).
- **Período pre largo** y donantes “**limpios**” (sin su propia política ni contagio del tratamiento).
- Ideal con **una o pocas** unidades agregadas (estados, países).

Control sintético en una frase

Un contrafactual **ponderado y transparente**, con pesos elegidos para calzar el pasado; el efecto es la brecha, y la inferencia es por permutación.

- Es un **DID con pesos elegidos**: resuelve “¿y si no hay un buen control?”.
- Extensiones modernas (mención): *augmented SC*, control sintético generalizado, *matrix completion*.
- Y como siempre en este curso: **confía en tus ojos** — un buen control sintético se ve.