

# Inferencia, Causalidad y Políticas Públicas

## ECO-60116

Week 08: Metodo de Emparejamiento (Propensity Score Matching)

Eduard F. Martinez Gonzalez, Ph.D.

Departamento de Economía, Universidad Icesi

October 17, 2025

## Recap: Método de Event Study

- El **Event Study** extiende el enfoque de Diferencias en Diferencias (*DiD*) permitiendo analizar la **dinámica temporal** del efecto del tratamiento.
- En lugar de una sola interacción ( $D_i \times Post_t$ ), el modelo introduce una serie de **dummies relativas al evento**:

$$y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \tau_k \cdot \text{Lead/Lag}_{it}^{(k)} + \varepsilon_{it}$$

donde:

- $k < 0$ : *Leads* (periodos pre-tratamiento)  $\rightarrow$  permiten **verificar tendencias paralelas**.
- $k > 0$ : *Lags* (periodos post-tratamiento)  $\rightarrow$  capturan la **evolución del efecto** a lo largo del tiempo.
- El coeficiente de referencia ( $k = -1$ ) se omite para identificar el modelo. Los  $\tau_k$  se interpretan como el **efecto del tratamiento en el periodo  $k$**  relativo al periodo base.

## Recap: Event Study

Los coeficientes  $\tau_k$  se grafican con sus intervalos de confianza, lo que permite evaluar:

- **Pre-trends:** validación del supuesto de **tendencias paralelas**.
- **Dinámica del efecto:** si el impacto persiste, se acumula o se disipa en el tiempo.

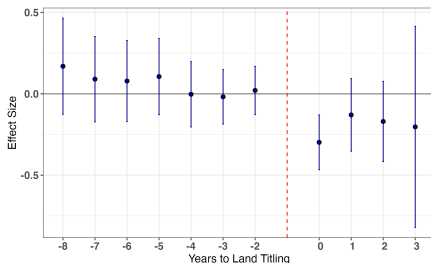


Figure: Tomado de Martínez et al. (2023).

## Intuición

El método de **Event Study** nos permite observar **cuándo** aparece el efecto, **cuánto dura** y si **desaparece o se acumula** con el tiempo.

# 1 Metodo de Emparejamiento (PSM)

- Motivación

- Etapas del PSM

- Supuestos de identificación

# 2 Combinación PSM-D&D

# 3 Aplicación en R

# ¿Por qué usar Emparejamiento (PSM)?

- **Contexto:** Hasta ahora hemos estudiado distintos diseños cuasi-experimentales:
  - ▶ **Regresión discontinua (RD):** identifica efectos cuando existe una regla clara de asignación a tratamiento.
  - ▶ **Variables instrumentales (IV):** identifica efectos cuando existe una fuente exógena de variación en el tratamiento.
  - ▶ **Diferencias en Diferencias (DiD):** identifica efectos cuando grupos tratados y de control siguen *tendencias paralelas*.

# ¿Por qué usar Emparejamiento (PSM)?

- **Contexto:** Hasta ahora hemos estudiado distintos diseños cuasi-experimentales:
  - ▶ **Regresión discontinua (RD):** identifica efectos cuando existe una regla clara de asignación a tratamiento.
  - ▶ **Variables instrumentales (IV):** identifica efectos cuando existe una fuente exógena de variación en el tratamiento.
  - ▶ **Diferencias en Diferencias (DiD):** identifica efectos cuando grupos tratados y de control siguen *tendencias paralelas*.
- **El problema:** ¿Qué ocurre cuando:
  - ▶ Tenemos unidades tratadas y de control, pero **no hay una regla de asignación observada** (como en RD)?
  - ▶ Tampoco contamos con un **instrumento válido** que genere variación exógena?
  - ▶ Y además, las **tendencias paralelas no se cumplen** o no tenemos datos en panel?

# ¿Por qué usar Emparejamiento (PSM)?

- **Contexto:** Hasta ahora hemos estudiado distintos diseños cuasi-experimentales:
  - ▶ **Regresión discontinua (RD):** identifica efectos cuando existe una regla clara de asignación a tratamiento.
  - ▶ **Variables instrumentales (IV):** identifica efectos cuando existe una fuente exógena de variación en el tratamiento.
  - ▶ **Diferencias en Diferencias (DiD):** identifica efectos cuando grupos tratados y de control siguen *tendencias paralelas*.
- **El problema:** ¿Qué ocurre cuando:
  - ▶ Tenemos unidades tratadas y de control, pero **no hay una regla de asignación observada** (como en RD)?
  - ▶ Tampoco contamos con un **instrumento válido** que genere variación exógena?
  - ▶ Y además, las **tendencias paralelas no se cumplen** o no tenemos datos en panel?
- **Motivación**
  - ▶ En esos escenarios, necesitamos construir un *grupo de control comparable* a partir de los datos disponibles.
  - ▶ El método de **Emparejamiento** busca aproximar un contrafactual creíble seleccionando observaciones similares a las tratadas.

# ¿Qué es el PSM?

Recordamos que en la introducción vimos la posibilidad de hacer una estimación de una simple diferencia:

$$\hat{\tau}_{DIF} = E[Y_i^1 | D_i = 1] - E[Y_i^0 | D_i = 0]$$

Si añadimos y sustraemos el contrafactual, obtenemos la descomposición siguiente:

$$\hat{\tau}_{DIF} = \underbrace{E[Y_i^1 | D_i = 1] - E[Y_i^0 | D_i = 1]}_{\tau_{ATT}} + \underbrace{E[Y_i^0 | D_i = 1] - E[Y_i^0 | D_i = 0]}_{\text{Sesgo de Selección}}$$

## Idea central

La lógica del Método de Emparejamiento consiste en emparejar a cada tratado con la o las personas más similares en el grupo de control, de forma que estas sirvan como contrafactuales válidos y se elimine el sesgo de selección.



# ¿Qué es el PSM?

- Idealmente quisiéramos asociar cada tratado con un individuo idéntico en todas las características, pero esto no es factible por la **maldición de la dimensionalidad**.
- Rosenbaum y Rubin (1983) mostraron que es equivalente emparejar usando la probabilidad de ser tratado  $Pr(D_i = 1|X_i)$ , lo cual es más factible que usar todos los  $X_i$ .

## Nota

Es una solución apropiada cuando se puede argumentar que no existen características no observables que afecten simultáneamente el resultado y la probabilidad de estar tratado. En la mayoría de los casos, ¡es un supuesto muy fuerte!

# Ejemplos donde se puede aplicar PSM

- **Minería y desarrollo cognitivo** Suponga que se quiere analizar si la presencia de minerías que contaminan los ríos con mercurio afecta el desarrollo cognitivo de los niños (por ejemplo, medido con las pruebas *Saber 5*).
  - ▶ El posicionamiento de las minerías no es completamente aleatorio: puede depender de condiciones geográficas, institucionales o económicas.
  - ▶ Sin embargo, condicional en esas variables, la decisión de ubicar una minería no estaría muy relacionada con el nivel educativo de los niños.
  - ▶ Este contexto podría prestarse para aplicar PSM.

# Ejemplos donde se puede aplicar PSM

- **Minería y desarrollo cognitivo** Suponga que se quiere analizar si la presencia de mineras que contaminan los ríos con mercurio afecta el desarrollo cognitivo de los niños (por ejemplo, medido con las pruebas *Saber 5*).
  - ▶ El posicionamiento de las mineras no es completamente aleatorio: puede depender de condiciones geográficas, institucionales o económicas.
  - ▶ Sin embargo, condicional en esas variables, la decisión de ubicar una mina no estaría muy relacionada con el nivel educativo de los niños.
  - ▶ Este contexto podría prestarse para aplicar PSM.
- **Redistribución de tierras y movilidad intergeneracional** Suponga que el gobierno implementa una política de redistribución de tierras y desea conocer su efecto sobre la riqueza de los beneficiarios y la de sus hijos.
  - ▶ Una institución define quiénes son prioritarios con base en criterios observables (pobreza, victimización, condiciones sociales, etc.).
  - ▶ Si el investigador tiene acceso a las mismas variables que usó la institución, puede argumentarse que los no observables relevantes son similares entre grupos.
  - ▶ En ese caso, PSM puede ofrecer un contrafactual razonable.

## 1 Metodo de Emparejamiento (PSM)

- Motivación
- Etapas del PSM
- Supuestos de identificación

## 2 Combinación PSM-D&D

## 3 Aplicación en R

# Las etapas del PSM

**El procedimiento del método de PSM se desarrolla en cinco pasos:**

- ➊ **Estimar** la probabilidad de estar tratado  $Pr(D_i = 1|X_i)$  y obtener su predicción para cada observación.
- ➋ **Restringir** la muestra al soporte común y verificar que exista balance entre tratados y controles.
- ➌ **Emparejar** las observaciones tratadas con las no tratadas según la similitud en su probabilidad estimada.
- ➍ **Estimar el impacto** como un promedio ponderado de las diferencias entre tratados y controles seleccionados.
- ➎ **Evaluar la significancia estadística** e intervalos de confianza del estimador.

## Interpretación

Cada paso busca aproximar el contrafactual que permitiría medir el efecto causal del tratamiento, reduciendo progresivamente el sesgo de selección.

# 1) Estimar la probabilidad de estar tratado $P(D|X)$

El método requiere, en primer lugar, estimar la probabilidad de estar tratado como función de las variables de control:

$$\pi(X_i) = Pr(D_i = 1 | X_i)$$

- Puede estimarse mediante una función lineal, logit o probit.
- **Lineal:** puede predecir probabilidades fuera del rango  $[0, 1]$ .
- **Logit:** requiere supuestos más fuertes si se incluyen efectos fijos.
- **Probit:** suele preferirse por su interpretación y estabilidad.

# 1) Estimar la probabilidad de estar tratado $P(D|X)$

El método requiere, en primer lugar, estimar la probabilidad de estar tratado como función de las variables de control:

$$\pi(X_i) = Pr(D_i = 1 | X_i)$$

- Puede estimarse mediante una función lineal, logit o probit.
- **Lineal:** puede predecir probabilidades fuera del rango  $[0, 1]$ .
- **Logit:** requiere supuestos más fuertes si se incluyen efectos fijos.
- **Probit:** suele preferirse por su interpretación y estabilidad.

## Selección de covariables

Las variables  $X_i$  deben ser predeterminadas —medidas antes del tratamiento— y teóricamente relevantes tanto para la probabilidad de tratamiento como para el resultado. Nunca deben incluirse variables afectadas por el tratamiento.

# 1) Estimar la probabilidad de estar tratado $P(D|X)$

El método requiere, en primer lugar, estimar la probabilidad de estar tratado como función de las variables de control:

$$\pi(X_i) = Pr(D_i = 1 | X_i)$$

- Puede estimarse mediante una función lineal, logit o probit.
- **Lineal:** puede predecir probabilidades fuera del rango  $[0, 1]$ .
- **Logit:** requiere supuestos más fuertes si se incluyen efectos fijos.
- **Probit:** suele preferirse por su interpretación y estabilidad.

## Selección de covariables

Las variables  $X_i$  deben ser predeterminadas —medidas antes del tratamiento— y teóricamente relevantes tanto para la probabilidad de tratamiento como para el resultado. Nunca deben incluirse variables afectadas por el tratamiento.

## Resultado de esta etapa

La estimación genera una predicción  $\pi(X_i)$  para cada observación  $i$ , tratada o no tratada, que servirá para el emparejamiento posterior.



## 2) Soporte común y restricción de la muestra

Una vez estimada la probabilidad de tratamiento  $\pi(X_i)$ , se analizan sus distribuciones para los grupos de tratados y de control.

- El **soporte común** es el rango de valores de  $\pi(X_i)$  donde existen observaciones en ambos grupos.
- Las unidades fuera del soporte común deben excluirse, pues no tienen un contrafactual comparable.
- Es necesario verificar que el soporte no sea demasiado estrecho y que exista **balance adecuado** entre las covariables.
- Si las densidades no se solapan adecuadamente, no es posible construir un contrafactual válido.

## 2) Soporte común y restricción de la muestra

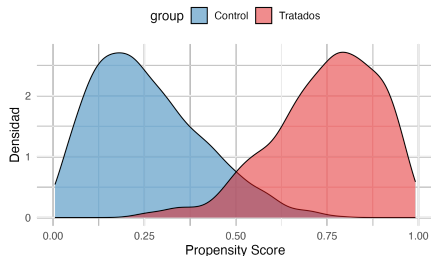
Una vez estimada la probabilidad de tratamiento  $\pi(X_i)$ , se analizan sus distribuciones para los grupos de tratados y de control.

- El **soporte común** es el rango de valores de  $\pi(X_i)$  donde existen observaciones en ambos grupos.
- Las unidades fuera del soporte común deben excluirse, pues no tienen un contrafactual comparable.
- Es necesario verificar que el soporte no sea demasiado estrecho y que exista **balance adecuado** entre las covariables.
- Si las densidades no se solapan adecuadamente, no es posible construir un contrafactual válido.

### Interpretación

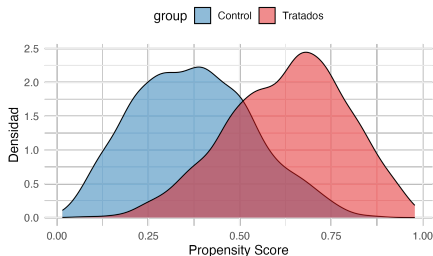
El efecto estimado bajo soporte común se interpreta como un **LATE** (efecto local promedio del tratamiento) dentro de la subpoblación comparable.

# Soporte común: suficiente vs. insuficiente



Falta de soporte común.

Las distribuciones apenas se traslapan → los grupos no son comparables.



Soporte común amplio.

Las distribuciones se traslapan ampliamente → los grupos son comparables.

## Takeaway

El soporte común no elimina el problema del sesgo de selección por no observables; simplemente reduce la parte del sesgo atribuible a diferencias en los observables.

### 3) El emparejamiento

Una vez definido el soporte común, el siguiente paso es emparejar cada observación tratada con una o varias observaciones del grupo de control que tengan valores similares en su probabilidad estimada de tratamiento  $\pi(X_i)$ .

### 3) El emparejamiento

Una vez definido el soporte común, el siguiente paso es emparejar cada observación tratada con una o varias observaciones del grupo de control que tengan valores similares en su probabilidad estimada de tratamiento  $\pi(X_i)$ .

| Método   | Descripción  |
|--|--|
| <b>Vecinos más cercanos</b><br>( <i>Nearest Neighbor</i> ) | Selecciona los $n$ controles cuya diferencia en $\pi(X_i)$ es mínima: $ \pi(X_i) - \pi(X_j) $ .                              |
| <b>Distancia máxima</b>                                    | Incluye todos los controles con una distancia menor o igual a un umbral $k$ : $ \pi(X_i) - \pi(X_j)  \leq k$ .               |
| <b>Kernel Matching</b>                                     | Asigna ponderaciones decrecientes según la distancia ( <i>Kernel</i> triangular, gaussiano o Epanechnikov).                  |
| <b>Estratificación o Sub-clasificación</b>                 | Divide el rango de $\pi(X_i)$ en bloques y compara tratados y controles dentro de cada bloque, garantizando balance interno. |

### 3) El emparejamiento

Una vez definido el soporte común, el siguiente paso es emparejar cada observación tratada con una o varias observaciones del grupo de control que tengan valores similares en su probabilidad estimada de tratamiento  $\pi(X_i)$ .

| Método   | Descripción  |
|--|--|
| <b>Vecinos más cercanos</b><br>( <i>Nearest Neighbor</i> ) | Selecciona los $n$ controles cuya diferencia en $\pi(X_i)$ es mínima: $ \pi(X_i) - \pi(X_j) $ .                              |
| <b>Distancia máxima</b>                                    | Incluye todos los controles con una distancia menor o igual a un umbral $k$ : $ \pi(X_i) - \pi(X_j)  \leq k$ .               |
| <b>Kernel Matching</b>                                     | Asigna ponderaciones decrecientes según la distancia ( <i>Kernel</i> triangular, gaussiano o Epanechnikov).                  |
| <b>Estratificación o Sub-clasificación</b>                 | Divide el rango de $\pi(X_i)$ en bloques y compara tratados y controles dentro de cada bloque, garantizando balance interno. |

#### Trade-off

Existe un equilibrio entre **poder estadístico** y **sesgo**: aumentar el número de vecinos reduce varianza pero puede disminuir comparabilidad. El método de Kernel suele ofrecer un buen balance.

## 4) Estimación del impacto promedio del tratamiento

Una vez emparejadas las observaciones tratadas y de control, el efecto del tratamiento se calcula como la diferencia promedio entre cada tratado y el promedio ponderado de sus contrafactuales emparejados.

## 4) Estimación del impacto promedio del tratamiento

Una vez emparejadas las observaciones tratadas y de control, el efecto del tratamiento se calcula como la diferencia promedio entre cada tratado y el promedio ponderado de sus contrafactuales emparejados.

$$\hat{\tau}_{PSM} = \frac{1}{N_T} \sum_{i \in (D=1)} \left[ Y_i - \sum_{j \in (D=0)} w_{ij} Y_j \right]$$

- $N_T$ : número total de observaciones tratadas.
- $Y_i$ : resultado de la observación tratada  $i$ .
- $Y_j$ : resultado de la observación de control  $j$ .
- $w_{ij}$ : peso asignado al control  $j$  como contrafactual de  $i$ .



## 4) Estimación del impacto promedio del tratamiento

Una vez emparejadas las observaciones tratadas y de control, el efecto del tratamiento se calcula como la diferencia promedio entre cada tratado y el promedio ponderado de sus contrafactuales emparejados.

$$\hat{\tau}_{PSM} = \frac{1}{N_T} \sum_{i \in (D=1)} \left[ Y_i - \sum_{j \in (D=0)} w_{ij} Y_j \right]$$

- $N_T$ : número total de observaciones tratadas.
- $Y_i$ : resultado de la observación tratada  $i$ .
- $Y_j$ : resultado de la observación de control  $j$ .
- $w_{ij}$ : peso asignado al control  $j$  como contrafactual de  $i$ .

### Intuición

Cada unidad tratada se compara con un “promedio sintético” de controles similares en su probabilidad de tratamiento  $\pi(X_i)$ .

# Interpretación de las ponderaciones y métodos

## Formas de asignar ponderaciones $w_{ij}$

- **Vecino más cercano:**  $w_{ij} = \frac{1}{n}$  para los  $n$  controles más cercanos a  $i$ ;  $w_{ij} = 0$  para los demás.
- **Kernel:** los controles más próximos tienen mayor peso; los más lejanos reciben ponderaciones decrecientes según la función de kernel.
- **Estratificación:** el efecto se calcula dentro de cada bloque del  $\pi(X_i)$  y luego se promedia ponderando por el número de tratados en cada bloque.

# Interpretación de las ponderaciones y métodos

## Formas de asignar ponderaciones $w_{ij}$

- **Vecino más cercano:**  $w_{ij} = \frac{1}{n}$  para los  $n$  controles más cercanos a  $i$ ;  $w_{ij} = 0$  para los demás.
- **Kernel:** los controles más próximos tienen mayor peso; los más lejanos reciben ponderaciones decrecientes según la función de kernel.
- **Estratificación:** el efecto se calcula dentro de cada bloque del  $\pi(X_i)$  y luego se promedia ponderando por el número de tratados en cada bloque.

## Interpretación práctica

El estimador  $\hat{\tau}_{PSM}$  es un **promedio ponderado de diferencias**, que mide cuánto difieren las observaciones tratadas de sus controles más comparables.

# Interpretación de las ponderaciones y métodos

## Formas de asignar ponderaciones $w_{ij}$

- **Vecino más cercano:**  $w_{ij} = \frac{1}{n}$  para los  $n$  controles más cercanos a  $i$ ;  $w_{ij} = 0$  para los demás.
- **Kernel:** los controles más próximos tienen mayor peso; los más lejanos reciben ponderaciones decrecientes según la función de kernel.
- **Estratificación:** el efecto se calcula dentro de cada bloque del  $\pi(X_i)$  y luego se promedia ponderando por el número de tratados en cada bloque.

## Interpretación práctica

El estimador  $\hat{\tau}_{PSM}$  es un **promedio ponderado de diferencias**, que mide cuánto difieren las observaciones tratadas de sus controles más comparables.

## Nota

El resultado depende del método de emparejamiento y del número de vecinos utilizados. Un exceso de controles puede aumentar el sesgo; un número muy bajo reduce la precisión.

## 5) Estimación de la significancia e intervalos de confianza

- El emparejamiento produce un estimador  $\hat{\tau}_{PSM}$ , pero necesitamos cuantificar su incertidumbre.

## 5) Estimación de la significancia e intervalos de confianza

- El emparejamiento produce un estimador  $\hat{\tau}_{PSM}$ , pero necesitamos cuantificar su incertidumbre.
- En algunos casos es posible derivar la varianza del estimador de forma analítica, aunque esto suele ser complejo para métodos como el Kernel Matching.

## 5) Estimación de la significancia e intervalos de confianza

- El emparejamiento produce un estimador  $\hat{\tau}_{PSM}$ , pero necesitamos cuantificar su incertidumbre.
- En algunos casos es posible derivar la varianza del estimador de forma analítica, aunque esto suele ser complejo para métodos como el Kernel Matching.

### Método de Bootstrapping

El enfoque más usado consiste en estimar la varianza mediante **re-muestreo con reemplazo**:

- 1 Tomar  $B$  muestras aleatorias (con reemplazo) del conjunto original.
- 2 Calcular  $\hat{\tau}_{PSM}^{(b)}$  para cada muestra  $b = 1, 2, \dots, B$ .
- 3 Estimar la varianza empírica:

$$\widehat{Var}(\hat{\tau}_{PSM}) = \frac{1}{B-1} \sum_{b=1}^B \left( \hat{\tau}_{PSM}^{(b)} - \bar{\tau}_{PSM} \right)^2$$

## 5) Estimación de la significancia e intervalos de confianza

- El emparejamiento produce un estimador  $\hat{\tau}_{PSM}$ , pero necesitamos cuantificar su incertidumbre.
- En algunos casos es posible derivar la varianza del estimador de forma analítica, aunque esto suele ser complejo para métodos como el Kernel Matching.

### Método de Bootstrapping

El enfoque más usado consiste en estimar la varianza mediante **re-muestreo con reemplazo**:

- 1 Tomar  $B$  muestras aleatorias (con reemplazo) del conjunto original.
- 2 Calcular  $\hat{\tau}_{PSM}^{(b)}$  para cada muestra  $b = 1, 2, \dots, B$ .
- 3 Estimar la varianza empírica:

$$\widehat{Var}(\hat{\tau}_{PSM}) = \frac{1}{B-1} \sum_{b=1}^B \left( \hat{\tau}_{PSM}^{(b)} - \bar{\tau}_{PSM} \right)^2$$

**El intervalo de confianza** se obtiene a partir de la distribución de los estimadores bootstrap. Permite evaluar la significancia estadística de  $\hat{\tau}_{PSM}$  sin depender de supuestos paramétricos fuertes.



## 1 Metodo de Emparejamiento (PSM)

- Motivación
- Etapas del PSM
- Supuestos de identificación

## 2 Combinación PSM-D&D

## 3 Aplicación en R

## Supuesto de Identificación: Independencia Condicional

Para que el emparejamiento identifique un efecto causal válido, debemos asumir que, condicional en las variables observables  $X_i$ , la asignación al tratamiento es independiente de los resultados potenciales.

$$E[Y_i^0 \mid D_i = 1, X_i] = E[Y_i^0 \mid D_i = 0, X_i]$$

# Supuesto de Identificación: Independencia Condicional

Para que el emparejamiento identifique un efecto causal válido, debemos asumir que, condicional en las variables observables  $X_i$ , la asignación al tratamiento es independiente de los resultados potenciales.

$$E[Y_i^0 \mid D_i = 1, X_i] = E[Y_i^0 \mid D_i = 0, X_i]$$

- Este es el **Supuesto de Independencia Condicional** (CIA, por sus siglas en inglés).
- Al condicionar en  $X_i$ , el cambio en  $D_i$  no afecta  $Y_i^0$ , excepto a través del tratamiento mismo.
- En otras palabras: si los tratados y los controles son observacionalmente iguales, sus resultados esperados sin tratamiento también lo serían.

# Supuesto de Identificación: Independencia Condicional

Para que el emparejamiento identifique un efecto causal válido, debemos asumir que, condicional en las variables observables  $X_i$ , la asignación al tratamiento es independiente de los resultados potenciales.

$$E[Y_i^0 \mid D_i = 1, X_i] = E[Y_i^0 \mid D_i = 0, X_i]$$

- Este es el **Supuesto de Independencia Condicional** (CIA, por sus siglas en inglés).
- Al condicionar en  $X_i$ , el cambio en  $D_i$  no afecta  $Y_i^0$ , excepto a través del tratamiento mismo.
- En otras palabras: si los tratados y los controles son observacionalmente iguales, sus resultados esperados sin tratamiento también lo serían.

## Advertencia

El supuesto falla si existen variables no observables que influyen simultáneamente en la probabilidad de recibir el tratamiento y en el resultado  $Y_i$ .

# Comparación con la regresión con controles

## ● Relación entre PSM y regresión con controles

- ▶ El método de regresión lineal controla por las variables  $X_i$  haciendo un supuesto de forma funcional, usualmente que los efectos de  $X_i$  sobre  $Y_i$  sean lineales y aditivos.
- ▶ PSM no impone una forma funcional específica, pero requiere que el efecto del tratamiento sea identificable dentro del soporte común.
- ▶ Al restringir la muestra al soporte común y emparejar observaciones similares, el PSM reduce la necesidad de **extrapolar** a regiones donde no hay comparabilidad.

# Comparación con la regresión con controles

## ● Relación entre PSM y regresión con controles

- ▶ El método de regresión lineal controla por las variables  $X_i$  haciendo un supuesto de forma funcional, usualmente que los efectos de  $X_i$  sobre  $Y_i$  sean lineales y aditivos.
- ▶ PSM no impone una forma funcional específica, pero requiere que el efecto del tratamiento sea identificable dentro del soporte común.
- ▶ Al restringir la muestra al soporte común y emparejar observaciones similares, el PSM reduce la necesidad de **extrapolar** a regiones donde no hay comparabilidad.

## ● Pregúntese siempre:

- ▶ ¿Por qué algunas unidades fueron tratadas y otras no?
- ▶ ¿Podría esa decisión estar relacionada con  $Y_i$ ?

# Comparación con la regresión con controles

## ● Relación entre PSM y regresión con controles

- ▶ El método de regresión lineal controla por las variables  $X_i$  haciendo un supuesto de forma funcional, usualmente que los efectos de  $X_i$  sobre  $Y_i$  sean lineales y aditivos.
- ▶ PSM no impone una forma funcional específica, pero requiere que el efecto del tratamiento sea identificable dentro del soporte común.
- ▶ Al restringir la muestra al soporte común y emparejar observaciones similares, el PSM reduce la necesidad de **extrapolar** a regiones donde no hay comparabilidad.

## ● Pregúntese siempre:

- ▶ ¿Por qué algunas unidades fueron tratadas y otras no?
- ▶ ¿Podría esa decisión estar relacionada con  $Y_i$ ?

## ● Conclusión:

- ▶ El PSM no es un método cuasi-experimental por sí solo.
- ▶ El emparejamiento controla por diferencias observables, que mejora la estimación observacional bajo supuestos fuertes.
- ▶ No existe una prueba empírica definitiva para validar el supuesto de CIA.

# ¿Cuándo es más creíble PSM?

- **PSM es más creíble cuando:**

- ▶ La asignación al tratamiento está influenciada por criterios técnicos u observables, como la presencia de ciertos recursos naturales o características geográficas.
- ▶ La decisión de asignar el tratamiento es institucional, basada en información disponible en los datos del investigador.
- ▶ Se puede argumentar que los factores no observables no afectan simultáneamente la probabilidad de tratamiento y el resultado.



# ¿Cuándo es más creíble PSM?

## ● PSM es más creíble cuando:

- ▶ La asignación al tratamiento está influenciada por criterios técnicos u observables, como la presencia de ciertos recursos naturales o características geográficas.
- ▶ La decisión de asignar el tratamiento es institucional, basada en información disponible en los datos del investigador.
- ▶ Se puede argumentar que los factores no observables no afectan simultáneamente la probabilidad de tratamiento y el resultado.

## ● Aplicaciones problemáticas

- ▶ Cuando la participación es una decisión individual (o del hogar), potencialmente influenciada por motivación o información no observable.
- ▶ Si algunos agentes fueron informados y otros no, lo cual puede generar endogeneidad en la decisión de participar.
- ▶ Cuando la institución que asigna el tratamiento utiliza criterios no observables que también afectan el resultado  $Y_i$ .

# ¿Cuándo es más creíble PSM?

## ● PSM es más creíble cuando:

- ▶ La asignación al tratamiento está influenciada por criterios técnicos u observables, como la presencia de ciertos recursos naturales o características geográficas.
- ▶ La decisión de asignar el tratamiento es institucional, basada en información disponible en los datos del investigador.
- ▶ Se puede argumentar que los factores no observables no afectan simultáneamente la probabilidad de tratamiento y el resultado.

## ● Aplicaciones problemáticas

- ▶ Cuando la participación es una decisión individual (o del hogar), potencialmente influenciada por motivación o información no observable.
- ▶ Si algunos agentes fueron informados y otros no, lo cual puede generar endogeneidad en la decisión de participar.
- ▶ Cuando la institución que asigna el tratamiento utiliza criterios no observables que también afectan el resultado  $Y_i$ .

La credibilidad del PSM depende del contexto y de la capacidad de justificar que las diferencias entre tratados y controles se explican solo por variables observables.

## 1 Metodo de Emparejamiento (PSM)

- Motivación
- Etapas del PSM
- Supuestos de identificación

## 2 Combinación PSM-D&D

## 3 Aplicación en R

# Combinación de PSM con Diferencias en Diferencias

## ● **Intuición:**

- ▶ El método de emparejamiento puede reducir el sesgo dado que compara a personas relativamente similares.
- ▶ Usar el método de Diferencias en Diferencias (DiD) reduce el sesgo porque controla por las diferencias entre tratados y control que no cambian a través del tiempo.

# Combinación de PSM con Diferencias en Diferencias

## ● **Intuición:**

- ▶ El método de emparejamiento puede reducir el sesgo dado que compara a personas relativamente similares.
- ▶ Usar el método de Diferencias en Diferencias (DiD) reduce el sesgo porque controla por las diferencias entre tratados y control que no cambian a través del tiempo.

## ● **Ventajas:**

- ▶ Combinar los dos métodos puede aprovechar las ventajas de cada uno.
- ▶ Es más exigente en datos, ya que requiere al menos dos rondas de información: una antes y otra después del tratamiento.

# Combinación de PSM con Diferencias en Diferencias

## ● Intuición:

- ▶ El método de emparejamiento puede reducir el sesgo dado que compara a personas relativamente similares.
- ▶ Usar el método de Diferencias en Diferencias (DiD) reduce el sesgo porque controla por las diferencias entre tratados y control que no cambian a través del tiempo.

## ● Ventajas:

- ▶ Combinar los dos métodos puede aprovechar las ventajas de cada uno.
- ▶ Es más exigente en datos, ya que requiere al menos dos rondas de información: una antes y otra después del tratamiento.

## Procedimiento

Suponiendo solo dos periodos,  $t = 0$  antes del tratamiento y  $t = 1$  después, reemplazamos el resultado  $Y_i$  por el cambio observado:

$$\Delta_i = Y_{i1} - Y_{i0}$$

y aplicamos el mismo procedimiento de emparejamiento sobre  $\Delta_i$ .

## Estimador combinado PSM-D&D

El estimador de PSM-D&D se aplica sobre los cambios en el tiempo, en lugar de usar niveles de  $Y_i$ :

## Estimador combinado PSM-D&D

El estimador de PSM-D&D se aplica sobre los cambios en el tiempo, en lugar de usar niveles de  $Y_i$ :

$$\hat{\tau}_{PSM-DiD} = \frac{1}{N_T} \sum_{i \in (D=1)} \left[ \Delta_i - \sum_{j \in (D=0)} w_{ij} \Delta_j \right]$$



# Estimador combinado PSM-D&D

El estimador de PSM-D&D se aplica sobre los cambios en el tiempo, en lugar de usar niveles de  $Y_i$ :

$$\hat{\tau}_{PSM-DiD} = \frac{1}{N_T} \sum_{i \in (D=1)} \left[ \Delta_i - \sum_{j \in (D=0)} w_{ij} \Delta_j \right]$$

## Interpretación

- El estimador compara la evolución temporal de los resultados ( $\Delta_i$ ) entre tratados y controles emparejados.
- Esta combinación puede ser útil cuando el método de Diferencias en Diferencias no cumple plenamente el supuesto de **tendencias paralelas**.
- Mantener grupos tratados y de control más similares aumenta la probabilidad de que dicho supuesto se cumpla.
- Si los grupos son similares y las tendencias paralelas se cumplen dentro del soporte común, el estimador identifica un efecto causal creíble.

## 1 Metodo de Emparejamiento (PSM)

- Motivación
- Etapas del PSM
- Supuestos de identificación

## 2 Combinación PSM-D&D

## 3 Aplicación en R

# Contexto: Evaluar el impacto de *Ser Pilo Paga*

- **Contexto:**

- ▶ *Ser Pilo Paga (SPP)* fue un programa del Gobierno de Colombia (2014–2018) que otorgó créditos–beca a estudiantes de bajos recursos con alto desempeño en el examen Saber 11, para financiar su educación superior en universidades acreditadas.

# Contexto: Evaluar el impacto de *Ser Pilo Paga*

- **Contexto:**

- ▶ *Ser Pilo Paga (SPP)* fue un programa del Gobierno de Colombia (2014–2018) que otorgó créditos–beca a estudiantes de bajos recursos con alto desempeño en el examen Saber 11, para financiar su educación superior en universidades acreditadas.

- **Pregunta de evaluación:**

- ▶ ¿El acceso al programa *Ser Pilo Paga* incrementó la probabilidad de que los beneficiarios ingresaran y permanecieran en la educación superior?

# Contexto: Evaluar el impacto de *Ser Pilo Paga*

## ● Contexto:

- ▶ *Ser Pilo Paga* (SPP) fue un programa del Gobierno de Colombia (2014–2018) que otorgó créditos–beca a estudiantes de bajos recursos con alto desempeño en el examen Saber 11, para financiar su educación superior en universidades acreditadas.

## ● Pregunta de evaluación:

- ▶ ¿El acceso al programa *Ser Pilo Paga* incrementó la probabilidad de que los beneficiarios ingresaran y permanecieran en la educación superior?

## ● Desafío de identificación

- ▶ La asignación al programa dependía del puntaje Saber 11 y del nivel socioeconómico.
- ▶ Los beneficiarios ( $D_i = 1$ ) y los no beneficiarios ( $D_i = 0$ ) pueden diferir sistemáticamente en características observables y no observables.
- ▶ No existe una asignación completamente aleatoria ni una regla de corte continua que permita RD pura.

# Contexto: Evaluar el impacto de *Ser Pilo Paga*

## ● Contexto:

- ▶ *Ser Pilo Paga (SPP)* fue un programa del Gobierno de Colombia (2014–2018) que otorgó créditos–beca a estudiantes de bajos recursos con alto desempeño en el examen Saber 11, para financiar su educación superior en universidades acreditadas.

## ● Pregunta de evaluación:

- ▶ ¿El acceso al programa *Ser Pilo Paga* incrementó la probabilidad de que los beneficiarios ingresaran y permanecieran en la educación superior?

## ● Desafío de identificación

- ▶ La asignación al programa dependía del puntaje Saber 11 y del nivel socioeconómico.
- ▶ Los beneficiarios ( $D_i = 1$ ) y los no beneficiarios ( $D_i = 0$ ) pueden diferir sistemáticamente en características observables y no observables.
- ▶ No existe una asignación completamente aleatoria ni una regla de corte continua que permita RD pura.

## ● Estrategia

- ▶ Usar **Propensity Score Matching (PSM)** para emparejar beneficiarios y no beneficiarios con características similares antes del ingreso a la universidad.
- ▶ Extender a **PSM–DiD** cuando se disponga de información antes y después del programa (por ejemplo, matrícula o desempeño académico en el tiempo).

# Aplicaciones en R

A continuación aplicaremos el método de **Emparejamiento (PSM)** y su extensión combinada con **Diferencias en Diferencias (PSM-DiD)** en dos contextos empíricos inspirados en políticas públicas reales:

❶ **Aplicación 1 — PSM:**

Evaluación del impacto del programa *Ser Pilo Paga (SPP)* sobre el acceso a la educación superior en Colombia.

*Basado en un caso sintético diseñado por Martínez-González (2025).*

❷ **Aplicación 2 — PSM-D&D:**

Extensión del análisis anterior considerando información antes y después del programa, para evaluar la evolución temporal del efecto sobre el ingreso a la universidad.

*Basado en: Martínez-González (2025).*

Puedes acceder al **script de R** [aquí](#).

¡Gracias!



# Referencias

- Rosenbaum, P. R., Rubin, D. B. (1983). *The central role of the propensity score in observational studies for causal effects*. Biometrika, 70(1), 41–55.
- Bernal, R., Peña, X. (2011). *Guía práctica para la evaluación de impacto*. Ediciones Uniandes.
- Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.