

# Propensity Score Matching, PSM

Yasmani Vargas

yasmani.vargas@unap.edu.pe

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

Sesión 5  
Ciclo II - Diseño y Evaluación de Políticas Públicas  
Noviembre, 2025

# Contenido

- 1 Motivación y concepto de PSM
- 2 PSM en la teoría: ATE, TOT y supuestos
- 3 Supuestos clave: Independencia condicional y soporte común
  - Independencia condicional
  - Soporte común
- 4 Estimador TOT con PSM
- 5 Pasos prácticos del PSM
  - Paso 1: Modelo de participación y puntaje de propensión
  - Paso 2: Soporte común y balanceo
  - Paso 3: Algoritmos de emparejamiento
- 6 Métodos de emparejamiento
- 7 Errores estándar y crítica al PSM
- 8 PSM y métodos basados en regresión

# Objetivos de la sesión

- Entender qué es el emparejamiento por puntaje de propensión (PSM) y qué problema resuelve.
- Formular rigurosamente el PSM dentro del marco de resultados potenciales.
- Discutir los supuestos clave: independencia condicional y soporte común.
- Presentar estimadores de TOT con PSM y su interpretación.
- Describir los pasos prácticos de implementación (modelo de participación, soporte común, emparejamiento).
- Analizar diferentes algoritmos de emparejamiento y el uso de DD+PSM.
- Discutir estimación de errores estándar (bootstrap) y críticas al método.
- Introducir enfoques híbridos PSM + regresión ponderada.

# Contenido

- 1 Motivación y concepto de PSM
- 2 PSM en la teoría: ATE, TOT y supuestos
- 3 Supuestos clave: Independencia condicional y soporte común
  - Independencia condicional
  - Soporte común
- 4 Estimador TOT con PSM
- 5 Pasos prácticos del PSM
  - Paso 1: Modelo de participación y puntaje de propensión
  - Paso 2: Soporte común y balanceo
  - Paso 3: Algoritmos de emparejamiento
- 6 Métodos de emparejamiento
- 7 Errores estándar y crítica al PSM
- 8 PSM y métodos basados en regresión

## Motivación: “imitar” un experimento aleatorizado

- Cuando el tratamiento no puede ser aleatorizado, el objetivo es construir un **análogo observacional** de un experimento.
- Idea central:
  - ▶ Encontrar, dentro de un gran grupo de no participantes, individuos *observacionalmente similares* a los participantes, en términos de características previas al programa.
- Cada participante se empareja con uno o varios no participantes similares; la diferencia media en resultados aproxima el efecto del tratamiento.
- Problema: alta dimensionalidad del vector de covariables  $X$  (maldición de la dimensionalidad).
- Solución: reducir  $X$  a un índice unidimensional: el **puntaje de propensión**.

# Definición básica de PSM

- Variable de tratamiento:

$$T_i = \begin{cases} 1 & \text{si la unidad } i \text{ participa en el programa} \\ 0 & \text{si no participa} \end{cases}$$

- Vector de covariables observadas no afectadas por el programa:  $X_i$ .
- **Puntaje de propensión** (Rosenbaum y Rubin, 1983):  
$$P(X_i) = \Pr(T_i = 1 | X_i)$$
- El PSM:
  - ▶ Estima  $P(X_i)$  para todos los individuos.
  - ▶ Empareja participantes ( $T_i = 1$ ) con no participantes ( $T_j = 0$ ) con puntajes de propensión similares.

# Contenido

- 1 Motivación y concepto de PSM
- 2 PSM en la teoría: ATE, TOT y supuestos
- 3 Supuestos clave: Independencia condicional y soporte común
  - Independencia condicional
  - Soporte común
- 4 Estimador TOT con PSM
- 5 Pasos prácticos del PSM
  - Paso 1: Modelo de participación y puntaje de propensión
  - Paso 2: Soporte común y balanceo
  - Paso 3: Algoritmos de emparejamiento
- 6 Métodos de emparejamiento
- 7 Errores estándar y crítica al PSM
- 8 PSM y métodos basados en regresión

# Resultados potenciales y efectos del tratamiento

- Resultados potenciales:

$Y_i(1)$  : resultado si  $i$  participa,     $Y_i(0)$  : resultado si  $i$  no participa

- Resultado observado:

$$Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$$

- **TOT** (efecto del tratamiento sobre los tratados):

$$TOT = E[Y_i(1) - Y_i(0) \mid T_i = 1]$$

- **ATE** (efecto promedio del tratamiento):

$$ATE = E[Y_i(1) - Y_i(0)]$$

- En aplicaciones con validez externa limitada, el foco suele ser el TOT.

# ¿Qué hace exactamente el PSM?

- **Paso conceptual:**

- 1 Estimar  $P(X_i) = \Pr(T_i = 1 | X_i)$  para cada individuo.
- 2 Construir para cada participante  $i$  un “gemelo estadístico”  $j$  (o conjunto de gemelos) no participante con  $P(X_j) \approx P(X_i)$ .
- 3 Comparar resultados  $Y_i$  vs.  $Y_j$ .

- El efecto de tratamiento sobre los tratados (TOT) se obtiene como:

$$TOT_{PSM} = E_{X|T=1} [E(Y^T | T = 1, P(X)) - E(Y^C | T = 0, P(X))]$$

- En práctica, esto se implementa mediante ponderaciones  $\omega(i, j)$  sobre los controles.

# Contenido

- 1 Motivación y concepto de PSM
- 2 PSM en la teoría: ATE, TOT y supuestos
- 3 **Supuestos clave: Independencia condicional y soporte común**
  - Independencia condicional
  - Soporte común
- 4 Estimador TOT con PSM
- 5 Pasos prácticos del PSM
  - Paso 1: Modelo de participación y puntaje de propensión
  - Paso 2: Soporte común y balanceo
  - Paso 3: Algoritmos de emparejamiento
- 6 Métodos de emparejamiento
- 7 Errores estándar y crítica al PSM
- 8 PSM y métodos basados en regresión

# Contenido

- 1 Motivación y concepto de PSM
- 2 PSM en la teoría: ATE, TOT y supuestos
- 3 Supuestos clave: Independencia condicional y soporte común
  - Independencia condicional
  - Soporte común
- 4 Estimador TOT con PSM
- 5 Pasos prácticos del PSM
  - Paso 1: Modelo de participación y puntaje de propensión
  - Paso 2: Soporte común y balanceo
  - Paso 3: Algoritmos de emparejamiento
- 6 Métodos de emparejamiento
- 7 Errores estándar y crítica al PSM
- 8 PSM y métodos basados en regresión

# Supuesto de independencia condicional (CIA)

- Supuesto central para identificación con PSM:

$$(Y_i^T, Y_i^C) \perp T_i | X_i$$

- También llamado:
  - ▶ *Unconfoundedness* o *no confusión*.
- Interpretación:
  - ▶ Dado  $X_i$ , la participación en el programa es “como si” fuera aleatoria.
  - ▶ No hay selección en función de características no observadas que afecten el resultado.
- Para el TOT, basta un supuesto más débil:

$$Y_i^C \perp T_i | X_i$$

# Limitaciones del supuesto de independencia condicional

- No es directamente verificable.
- Depende de:
  - ▶ Reglas de focalización del programa.
  - ▶ Mecanismos de auto-selección.
  - ▶ Calidad y amplitud de los datos previos al programa.
- Si hay selección importante sobre variables no observadas:
  - ▶ El PSM **no** es apropiado por sí solo.
  - ▶ Deben considerarse alternativas: IV, DD, RDD, modelos de selección, etc.

# Contenido

- 1 Motivación y concepto de PSM
- 2 PSM en la teoría: ATE, TOT y supuestos
- 3 **Supuestos clave: Independencia condicional y soporte común**
  - Independencia condicional
  - Soporte común
- 4 Estimador TOT con PSM
- 5 Pasos prácticos del PSM
  - Paso 1: Modelo de participación y puntaje de propensión
  - Paso 2: Soporte común y balanceo
  - Paso 3: Algoritmos de emparejamiento
- 6 Métodos de emparejamiento
- 7 Errores estándar y crítica al PSM
- 8 PSM y métodos basados en regresión

# Supuesto de soporte común

- Condición de *overlap*:

$$0 < P(T_i = 1 | X_i) < 1$$

- Asegura que:
  - ▶ Para cada tratado, existe al menos un no tratado similar en términos de  $P(X)$ .
- Para el TOT:
$$P(T_i = 1 | X_i) < 1$$
- Implicaciones:
  - ▶ Deben eliminarse observaciones fuera del soporte común.
  - ▶ Algunas unidades tratadas pueden también quedar fuera si no tienen comparables.

## Ejemplo de soporte común adecuado

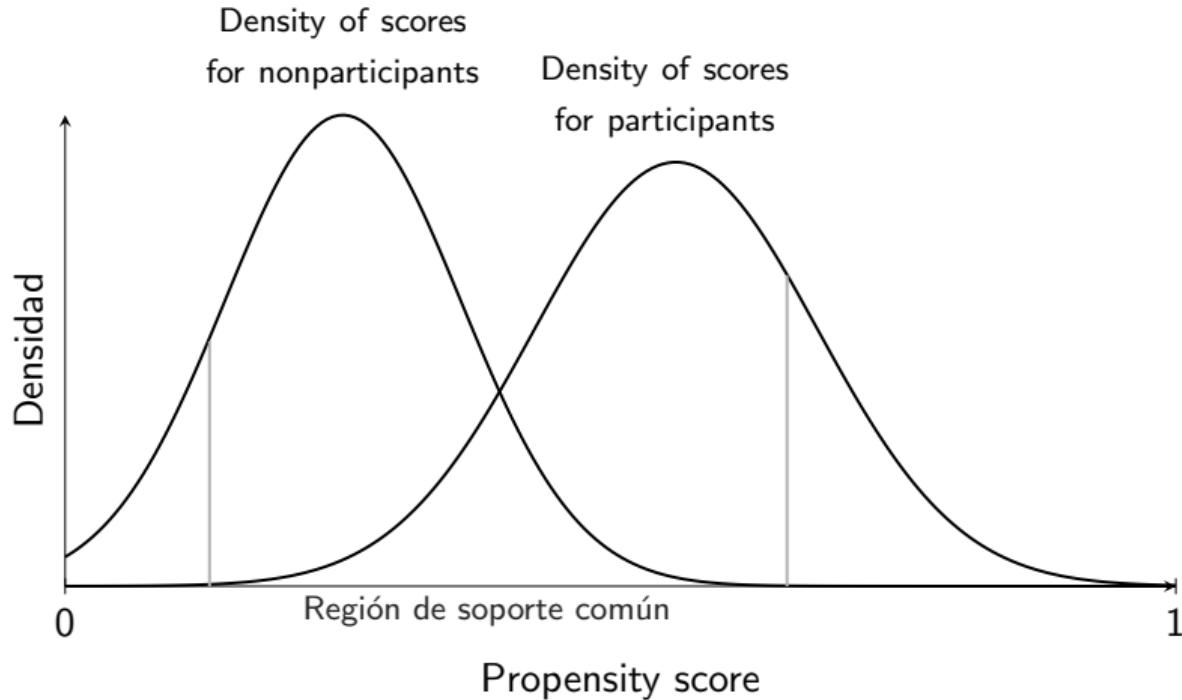


Figure: Ejemplo de Soporte Común

## Ejemplo de soporte común débil

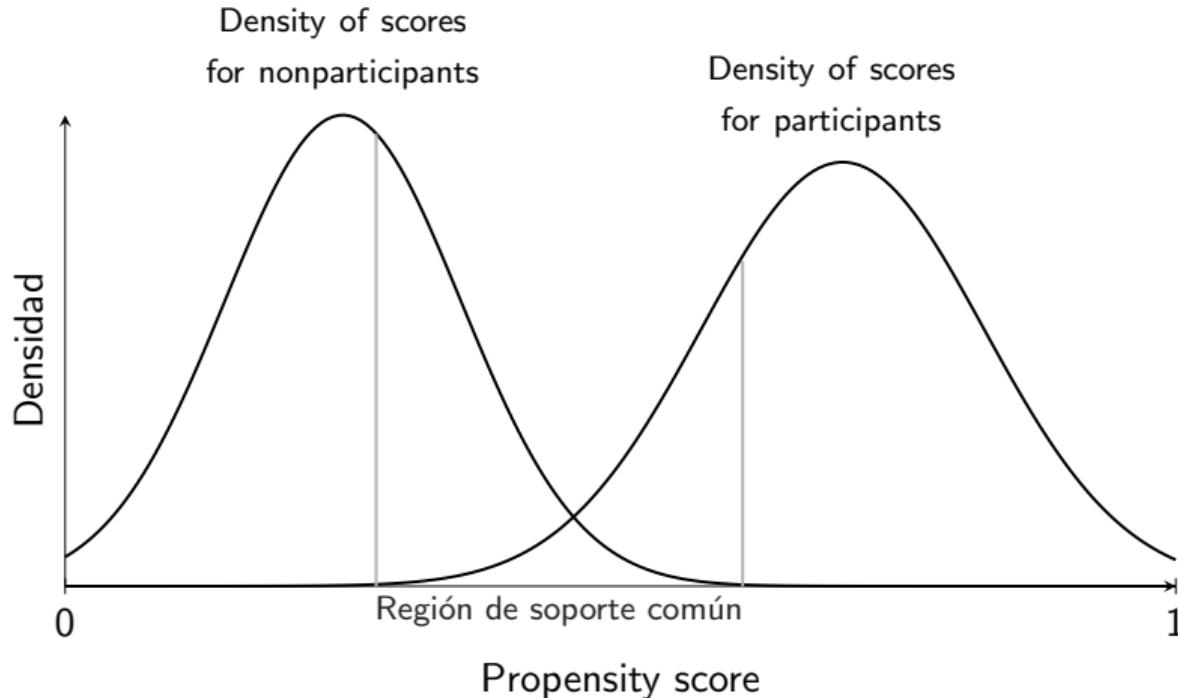


Figure: Ejemplo de un Deficiente Balanceo y Débil Soporte Común

# Consecuencias del soporte común débil

- Un área reducida de solapamiento implica:
  - ▶ Eliminación de muchas unidades de control.
  - ▶ Posible eliminación de unidades tratadas sin “gemelo” comparable.
- Esto puede generar:
  - ▶ **Sesgo de muestreo** en el efecto del tratamiento.
- Recomendación:
  - ▶ Examinar las características de las unidades eliminadas.
  - ▶ Interpretar el efecto estimado como válido para la submuestra en soporte común.

# Contenido

- 1 Motivación y concepto de PSM
- 2 PSM en la teoría: ATE, TOT y supuestos
- 3 Supuestos clave: Independencia condicional y soporte común
  - Independencia condicional
  - Soporte común
- 4 Estimador TOT con PSM
- 5 Pasos prácticos del PSM
  - Paso 1: Modelo de participación y puntaje de propensión
  - Paso 2: Soporte común y balanceo
  - Paso 3: Algoritmos de emparejamiento
- 6 Métodos de emparejamiento
- 7 Errores estándar y crítica al PSM
- 8 PSM y métodos basados en regresión

## Expresión general del TOT con PSM

- Bajo independencia condicional y soporte común:

$$TOT_{PSM} = E_{X|T=1} \left[ E(Y^T | T = 1, P(X)) - E(Y^C | T = 0, P(X)) \right]$$

- A nivel de datos:

$$TOT_{PSM} = \frac{1}{N_T} \left[ \sum_{i \in T} Y_i^T - \sum_{j \in C} \omega(i, j) Y_j^C \right]$$

- Donde:

- ▶  $N_T$ : número de tratados.
- ▶  $\omega(i, j)$ : peso asignado al control  $j$  en el emparejamiento del tratado  $i$ .

# Contenido

- 1 Motivación y concepto de PSM
- 2 PSM en la teoría: ATE, TOT y supuestos
- 3 Supuestos clave: Independencia condicional y soporte común
  - Independencia condicional
  - Soporte común
- 4 Estimador TOT con PSM
- 5 Pasos prácticos del PSM
  - Paso 1: Modelo de participación y puntaje de propensión
  - Paso 2: Soporte común y balanceo
  - Paso 3: Algoritmos de emparejamiento
- 6 Métodos de emparejamiento
- 7 Errores estándar y crítica al PSM
- 8 PSM y métodos basados en regresión

# Contenido

- 1 Motivación y concepto de PSM
- 2 PSM en la teoría: ATE, TOT y supuestos
- 3 Supuestos clave: Independencia condicional y soporte común
  - Independencia condicional
  - Soporte común
- 4 Estimador TOT con PSM
- 5 Pasos prácticos del PSM
  - **Paso 1: Modelo de participación y puntaje de propensión**
  - Paso 2: Soporte común y balanceo
  - Paso 3: Algoritmos de emparejamiento
- 6 Métodos de emparejamiento
- 7 Errores estándar y crítica al PSM
- 8 PSM y métodos basados en regresión

## Paso 1: Estimación del modelo de participación

- Combinar muestras de participantes y no participantes.
- Estimar:

$$\Pr(T_i = 1 | X_i) = P(X_i)$$

- ▶ Usualmente con modelos *logit* o *probit* binarios.
- ▶ Para tratamientos multinivel: modelos multinomiales o series de modelos binarios.
- Obtener el puntaje predicho:  
 $\hat{P}(X_i)$
- En esta etapa, interesa la **correlación** entre  $X$  y  $T$ , no la causalidad.

# Selección de covariables $X$

- El PSM será sesgado si faltan covariables relevantes que determinan:
  - ▶ La participación en el programa.
  - ▶ El resultado de interés.
- Recomendaciones prácticas:
  - ▶ Usar la misma fuente de datos para tratados y controles.
  - ▶ Incluir una muestra representativa de no participantes elegibles.
  - ▶ Asegurar tamaño de muestra suficientemente grande para encontrar buenos emparejamientos.
- Evitar:
  - ▶ Incluir demasiadas covariables que generen predicción casi perfecta ( $\hat{P}(X) = 0 \text{ o } 1$ ).
  - ▶ Modelos sobre especificados que inflan errores estándar de  $\hat{P}(X)$ .

# Contenido

- 1 Motivación y concepto de PSM
- 2 PSM en la teoría: ATE, TOT y supuestos
- 3 Supuestos clave: Independencia condicional y soporte común
  - Independencia condicional
  - Soporte común
- 4 Estimador TOT con PSM
- 5 Pasos prácticos del PSM
  - Paso 1: Modelo de participación y puntaje de propensión
  - **Paso 2: Soporte común y balanceo**
  - Paso 3: Algoritmos de emparejamiento
- 6 Métodos de emparejamiento
- 7 Errores estándar y crítica al PSM
- 8 PSM y métodos basados en regresión

## Paso 2: Definir soporte común y pruebas de balanceo

- Definir la región de soporte común:
  - ▶ Por ejemplo, usando los percentiles mínimos y máximos del puntaje en cada grupo.
- Eliminar observaciones fuera del soporte común (tratados o controles).
- Realizar **pruebas de balanceo**:
  - ▶ Verificar que, dentro de cada estrato de  $\hat{P}(X)$ , las medias de  $X$  sean similares entre tratados y controles.
  - ▶ Verificar que las distribuciones de  $\hat{P}(X)$  sean similares entre  $T = 1$  y  $T = 0$ .
- Balance adecuado implica:
  - ▶  $\hat{P}(X | T = 1) \approx \hat{P}(X | T = 0)$  en la región de soporte común.

# Contenido

- 1 Motivación y concepto de PSM
- 2 PSM en la teoría: ATE, TOT y supuestos
- 3 Supuestos clave: Independencia condicional y soporte común
  - Independencia condicional
  - Soporte común
- 4 Estimador TOT con PSM
- 5 Pasos prácticos del PSM
  - Paso 1: Modelo de participación y puntaje de propensión
  - Paso 2: Soporte común y balanceo
  - **Paso 3: Algoritmos de emparejamiento**
- 6 Métodos de emparejamiento
- 7 Errores estándar y crítica al PSM
- 8 PSM y métodos basados en regresión

## Paso 3: Criterios de emparejamiento y pesos $\omega(i,j)$

- El efecto del programa depende de cómo se construyen los pesos  $\omega(i,j)$  para los controles.
- Diferentes algoritmos implican diferentes pesos:
  - ▶ *Nearest-neighbor*.
  - ▶ Caliper/radius.
  - ▶ Estratificación.
  - ▶ Kernel y lineal local (LLM).
  - ▶ PSM + DD.
- Comparar resultados entre métodos ayuda a evaluar la robustez del TOT estimado.

# Contenido

- 1 Motivación y concepto de PSM
- 2 PSM en la teoría: ATE, TOT y supuestos
- 3 Supuestos clave: Independencia condicional y soporte común
  - Independencia condicional
  - Soporte común
- 4 Estimador TOT con PSM
- 5 Pasos prácticos del PSM
  - Paso 1: Modelo de participación y puntaje de propensión
  - Paso 2: Soporte común y balanceo
  - Paso 3: Algoritmos de emparejamiento
- 6 Métodos de emparejamiento
- 7 Errores estándar y crítica al PSM
- 8 PSM y métodos basados en regresión

# Neighbor Matching (vecino más cercano)

- Cada tratado se empareja con el control con puntaje  $\hat{P}(X)$  más cercano.
- Puede usarse:
  - ▶ Un solo vecino.
  - ▶ Varios vecinos ( $k$ -vecinos más cercanos, típicamente  $k \leq 5$ ).
- Emparejamiento con o sin reemplazo:
  - ▶ Con reemplazo: el mismo control puede ser usado para varios tratados.
  - ▶ Sin reemplazo: cada control se usa una sola vez.

# Caliper / Radius Matching

- Problema del vecino más cercano:
  - ▶ El vecino más cercano puede estar “lejos” en términos de  $\hat{P}(X)$ .
- Solución:
  - ▶ Establecer un **caliper** (radio máximo) de diferencia en el puntaje.
  - ▶ Emparejar solo controles con  $|\hat{P}_i - \hat{P}_j|$  dentro del caliper.
- Consecuencias:
  - ▶ Mejor calidad de emparejamiento.
  - ▶ Mayor número de observaciones eliminadas ( posible sesgo de muestreo).

## Estratificación / Intervalos

- Dividir el soporte común de  $\hat{P}(X)$  en varios estratos o “bloques”.
- Dentro de cada estrato:
  - ▶ Estimar la diferencia promedio en resultados entre tratados y controles.
- Efecto total:
  - ▶ Promedio ponderado de los efectos en cada estrato, usando como peso la participación de tratados en cada uno.
- Ventaja:
  - ▶ Implementación sencilla.
  - ▶ Relación directa con pruebas de balanceo por estratos.

# Kernel Matching

- Para cada tratado  $i$ , se usan todos los controles  $j$  dentro del soporte común.
- Los pesos dependen de la distancia en puntaje de propensión:

$$\omega(i,j)_{KM} = \frac{K\left(\frac{P_i - P_j}{a_n}\right)}{\sum_{j \in C} K\left(\frac{P_i - P_j}{a_n}\right)}$$

- Donde:
  - ▶  $K(\cdot)$  es una función kernel (por ejemplo, gaussiano, Epanechnikov).
  - ▶  $a_n$  es el parámetro de ancho de banda.
- Ventaja:
  - ▶ Usa toda la información de los controles.
  - ▶ Reduce varianza respecto al matching discreto.

## Local Linear Matching (LLM)

- Extiende el kernel matching incorporando no solo un nivel, sino una pendiente local.
- El peso LLM es:

$$\omega(i,j)_{LLR} = \frac{K_{ij} \sum_{k \in C} K_{ik}(P_k - P_i)^2 - [K_{ij}(P_j - P_i)] \sum_{k \in C} K_{ik}(P_k - P_i)}{\sum_{k \in C} K_{ik} \sum_{k \in C} K_{ik}(P_k - P_i)^2 - (\sum_{k \in C} K_{ik}(P_k - P_i))^2}$$

- Mejora el comportamiento cerca de los extremos de la distribución (*boundary bias* menor).
- Es un estimador localmente lineal del contrafactual.

## PSM combinado con DD

- Si se dispone de datos antes y después del programa, se puede usar:

$$TOT_{PSM}^{DD} = \frac{1}{N_T} \left[ \sum_{i \in T} (Y_i^{T2} - Y_i^{T1}) - \sum_{j \in C} \omega(i, j) (Y_j^{C2} - Y_j^{C1}) \right]$$

- Con solo cortes transversales en dos períodos:

$$TOT_{PSM}^{DD} = \left( \bar{Y}^{T_2} - \sum_{j \in C} \omega(i, j) Y_j^{C_2} \right) - \left( \bar{Y}^{T_1} - \sum_{j \in C} \omega(i, j) Y_j^{C_1} \right)$$

- Ventaja:

- ▶ Permite controlar características no observadas constantes en el tiempo.
- ▶ Combina selección sobre observables (PSM) con corrección parcial por no observables (DD).

# Contenido

- 1 Motivación y concepto de PSM
- 2 PSM en la teoría: ATE, TOT y supuestos
- 3 Supuestos clave: Independencia condicional y soporte común
  - Independencia condicional
  - Soporte común
- 4 Estimador TOT con PSM
- 5 Pasos prácticos del PSM
  - Paso 1: Modelo de participación y puntaje de propensión
  - Paso 2: Soporte común y balanceo
  - Paso 3: Algoritmos de emparejamiento
- 6 Métodos de emparejamiento
- 7 Errores estándar y crítica al PSM
- 8 PSM y métodos basados en regresión

# Estimación de errores estándar: bootstrap

- La varianza del estimador PSM incluye:
  - ▶ Variación por muestreo.
  - ▶ Variación por estimación del puntaje de propensión.
  - ▶ Variación por definición de soporte común.
  - ▶ Variación por el orden de emparejamiento (si es sin reemplazo).
- Usar fórmulas estándar ignora parte de esta variación.
- Solución práctica:
  - ▶ **Bootstrap:** repetir el proceso completo (estimación de  $P(X)$ , soporte común, emparejamiento) en muchas muestras re-muestreadas.
  - ▶ Estimar errores estándar e intervalos de confianza a partir de la distribución empírica de los TOT re-estimados.

# Crítica al método PSM

- Principal ventaja y a la vez debilidad:
  - ▶ Depende de que la selección esté completamente explicada por características observadas.
- Si variables relevantes no están medidas o incluidas en  $X$ :
  - ▶ El supuesto de independencia condicional se viola.
  - ▶ El efecto estimado puede estar sesgado.
- Requiere soporte común amplio:
  - ▶ Si hay pocos controles comparables, el sesgo y la varianza aumentan.
- Aun así, comparado con una simple regresión:
  - ▶ PSM fuerza explícitamente comparabilidad en la distribución de  $X$  entre tratados y controles.

# Contenido

- 1 Motivación y concepto de PSM
- 2 PSM en la teoría: ATE, TOT y supuestos
- 3 Supuestos clave: Independencia condicional y soporte común
  - Independencia condicional
  - Soporte común
- 4 Estimador TOT con PSM
- 5 Pasos prácticos del PSM
  - Paso 1: Modelo de participación y puntaje de propensión
  - Paso 2: Soporte común y balanceo
  - Paso 3: Algoritmos de emparejamiento
- 6 Métodos de emparejamiento
- 7 Errores estándar y crítica al PSM
- 8 PSM y métodos basados en regresión

# PSM + regresión ponderada por puntaje de propensión

- Hirano, Imbens y Ridder (2003) proponen combinar:

- ▶ Estimación no paramétrica de  $P(X)$ .
  - ▶ Regresión ponderada por  $P(X)$ .

- Especificación típica:

$$Y_i = \alpha + \beta T_i + \gamma X_i + \varepsilon_i$$

- Pesos:

- ▶ Para estimar TOT: peso 1 para tratados y  $\hat{P}(X)/(1 - \hat{P}(X))$  para controles.
  - ▶ Para estimar ATE:  $1/\hat{P}(X)$  para tratados y  $1/(1 - \hat{P}(X))$  para controles.

- Ventaja:

- ▶ Combina la mejora de comparabilidad del PSM con la eficiencia de los modelos de regresión.