

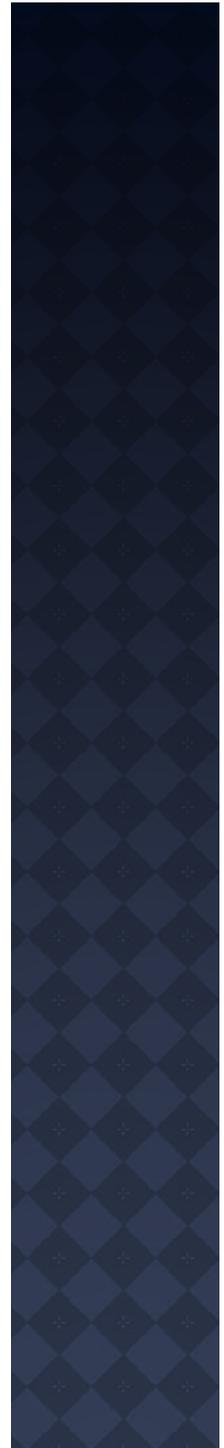
Evaluación de políticas públicas con Microsimulaciones

Karina Caballero y Jimmy Ferrer

Febrero 2011

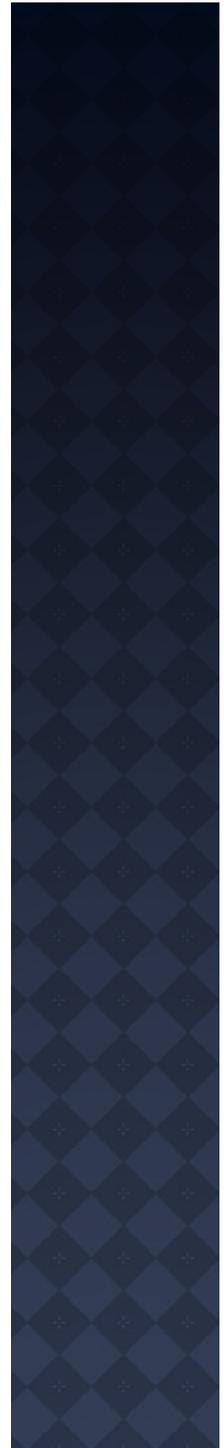
Introducción

- ◉ Las evaluaciones de impacto de programas públicos buscan medir los cambios en el bienestar de los individuos que pueden ser atribuidos a un programa o a una política específica y sus objetivos son, entonces, proveer información y ayudar a mejorar su eficacia.
- ◉ Asimismo, el conocimiento obtenido de estos estudios que evalúan el impacto proporciona información decisiva para el diseño adecuado de programas al proporcionar respuesta a preguntas tales como:



Introducción

- ¿Logra el programa las metas propuestas?,
- ¿Son los cambios producidos un resultado directo del programa, o son resultado de otros factores que ocurrieron simultáneamente?,
- ¿Cambia el impacto del programa dependiendo del grupo al que se está tratando de beneficiar (hombres, mujeres, pueblos indígenas), o de la región, o a través del tiempo?;
- ¿Tuvo el programa efectos inesperados, ya sean positivos o negativos?,
- ¿Qué tan eficiente es el programa en comparación con intervenciones alternativas?
- ¿Justifica el valor del programa su costo?



Introducción

- ◉ Una cuestión clave en la evaluación de políticas es identificar la causalidad de la correlación.
 - **Correlación:** dos variables se mueven juntas
 - **Causalidad:** una de las variables causa el movimiento de la otra
- ◉ Es fundamental para las políticas públicas entender la diferencia, de otra forma la política puede no tener el impacto deseado.
- ◉ Para propósitos de política lo que importa es la relación de **causalidad**

Bases teóricas de la evaluación de impacto

- ⦿ Preguntas de causa - efecto:
 - X causa a Y?
 - Si X causa a Y, qué tan grande es el efecto en Y?
 - Este efecto es mayor que el generado por otra causas?

- ⦿ Este tipo de preguntas es la motivación de mucho trabajo empírico en las ciencias sociales

Metodología básica

- ◉ Las evaluaciones tratan de establecer los resultados que se produjeron o causaron por un programa (política).
- ◉ **PREGUNTA DE INVESTIGACIÓN:** ¿Cuál es el efecto causal del programa (política) en los participantes, comparado con la no implementación del programa?
- ◉ Aproximación: mediante la construcción de un escenario hipotético.

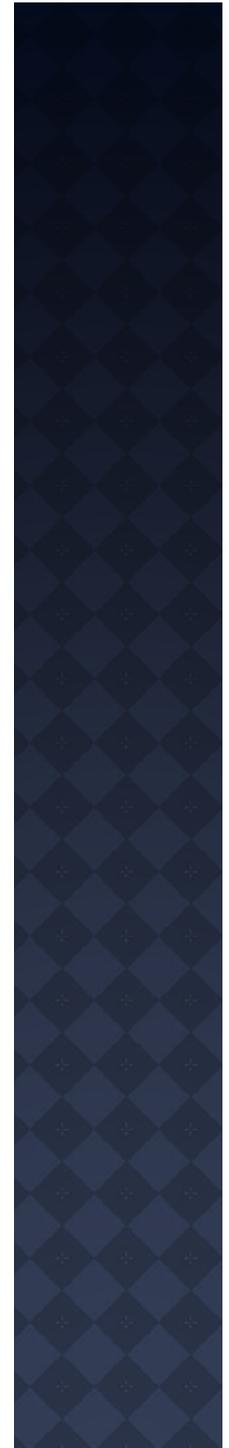
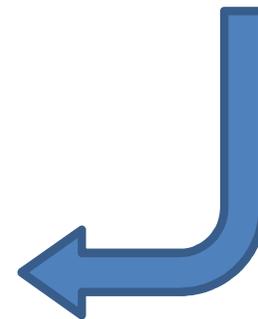
Metodología básica

- Opciones para estimar el efecto causal de un programa (Aproximación a la evaluación de programas):

Opciones	Dificultades
Comparar resultados “antes” y “después”	VARIABLES ECONÓMICAS SE MUEVEN
Incluir un grupo control	EXISTEN DIFERENCIAS EN LAS CARACTERÍSTICAS OBSERVADAS Y NO OBSERVADAS DE LOS INDIVIDUOS

Grupo control seleccionado de forma aleatoria
(EXPERIMENTALES)

Grupo control manipulado con técnicas estadísticas
(CUASI-EXPERIMENTALES)

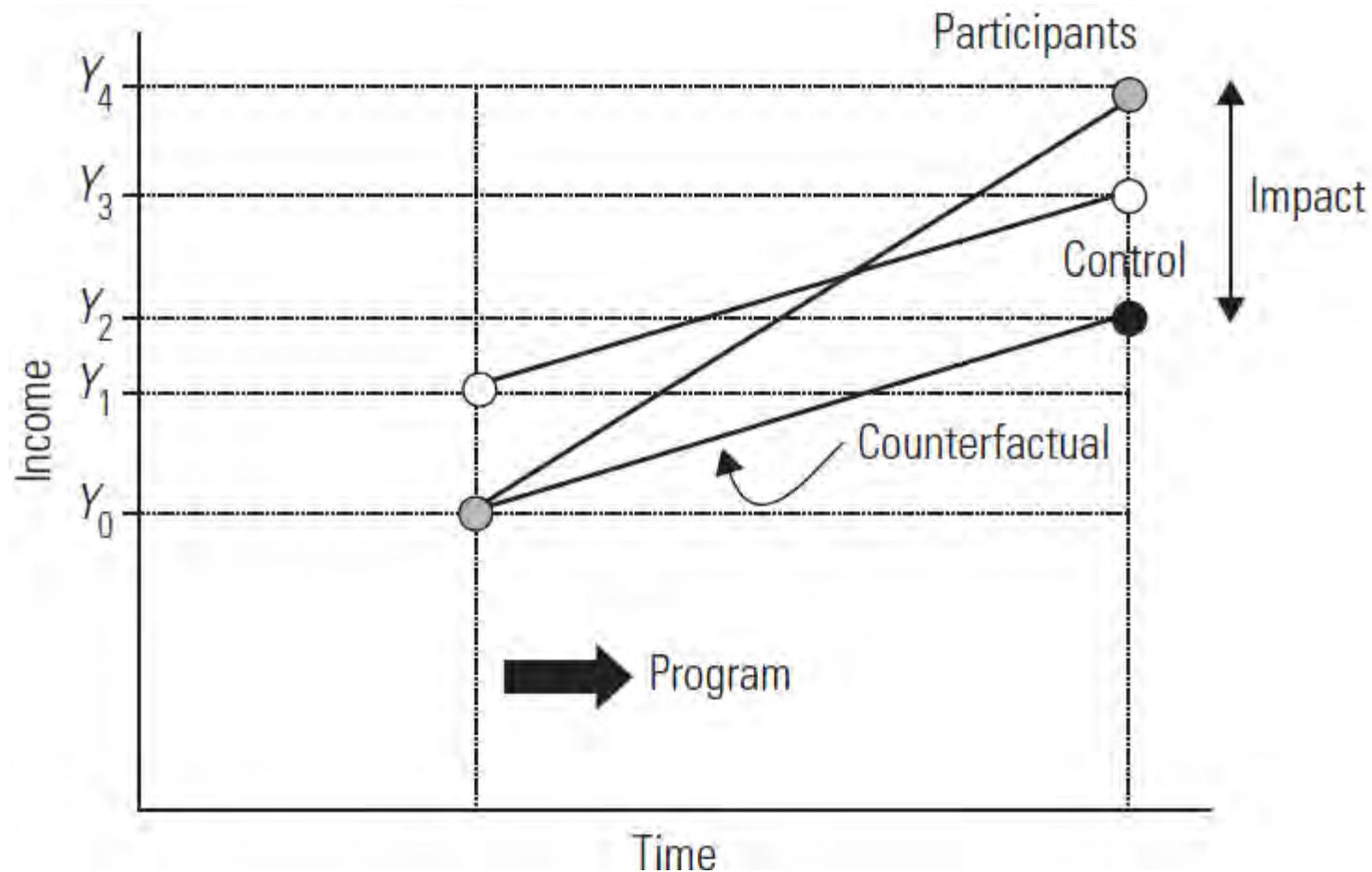


Metodología básica

- ◉ El **grupo de control** está compuesto de individuos que presenten características muy similares a los beneficiarios del programa (Caliendo y Kopeining, 2005).
- ◉ Un grupo de control es comparable en el caso en que la única diferencia en los resultados promedio del grupo de tratamiento y el de control, se explica por el efecto de la intervención.

Metodología básica

- CONTRAFACTUAL: el resultado de un beneficiario con la ausencia del programa



Fuente: Khandker, et al. 2010

Metodología básica

ENFOQUE EXPERIMENTAL

- ⦿ Enfoques experimentales a menudo se describen como “el estándar de oro” de la evaluación (Campbell y Stanley, 1973).
- ⦿ Esencialmente, se trata de reproducir el tipo de condiciones en las que los comportamientos de los “fenómenos de las ciencias naturales” son observados y entendidos a un contexto social.
- ⦿ La característica fundamental es la **asignación aleatoria** del tratamiento, tanto al grupo de control como al grupo de tratamiento, en consecuencia cada unidad de análisis tiene la misma probabilidad de ser elegida.

Metodología básica

- ◉ El principal objetivo de las técnicas experimentales es reunir suficiente evidencia confiable acerca de un fenómeno particular en una situación particular.
- ◉ VENTAJA: representan el más riguroso y "científico" enfoque para evaluar los resultados e impactos, y para inferir "causalidad" entre las intervenciones y los resultados e impactos.
- ◉ DESVENTAJAS:
 - 1) la dificultad de aplicar el método clásico de las ciencias naturales a las ciencias sociales
 - 2) alto costo de los métodos clásicos de la ciencia natural

Metodología básica

ENFOQUE CUASI-EXPERIMENTAL

- ◉ Métodos cuasi-experimentales flexibilizan las condiciones de probabilidad y la distribución de la población impuestas por los diseños de investigación experimental.
- ◉ Desarrollo de nuevas herramientas estadísticas que permitieron el control estadístico.
- ◉ Frecuentemente utilizan técnicas estadísticas como análisis probit, análisis de supervivencia y análisis de regresión jerárquica.

EL PROBLEMA DE EVALUACIÓN

Para identificar el impacto del programa, se estima la diferencia en la variable resultado (Y) que registra el individuo i con y sin la intervención del programa (Heckman y Smith, 1996)

$$\Delta_i = Y_{1i} - Y_{0i} \quad (1)$$

Donde Δ_i es el impacto del programa sobre el individuo i:

$Y_{0i} = \mu(X) + U_0$ resultado potencial que registraría el productor i sin el tratamiento,

$Y_{1i} = \mu(X) + U_1$ resultado potencial que registraría el productor i con el tratamiento.

EL PROBLEMA DE EVALUACION

Debido al “problema fundamental de identificación”, que establece que es imposible conocer ambos resultados para el mismo productor, entonces, los distintos métodos de evaluación buscan construir el contrafactual (Caliendo y Kopeining, 2006).

Para ello se redefine entonces el problema, pasando del nivel individual al nivel poblacional y, así, se estima el valor medio de Δ o algún aspecto de su distribución (Heckman *et al.*, 1999).

Metodología básica

Existen dos parámetros básicos disponibles en la literatura para medir el impacto promedio (Caliendo y Kopening, 2006).

1. Efecto promedio del tratamiento (*average treatment effect*, ATE): la diferencia esperada de los resultados de participar y no participar.
 - ⦿ Este estadístico expresa el efecto esperado en el resultado en el caso que los individuos de la población sean asignados aleatoriamente al tratamiento.
 - ⦿ Este caso puede ser no relevante en la medida en que desde la óptica de la política pública importa más el caso donde las personas son seleccionadas, específicamente para participar en el programa (Heckman, 1997)

Metodología básica

2. Efecto promedio del tratamiento en la población tratada (*average treatment effect on the treated, ATT*): los efectos de aquellos donde el programa es efectivamente aplicado.
- ◉ El valor esperado de ATT es la diferencia entre los valores del resultado esperado con y sin tratamiento para aquellos que participan en el tratamiento.

$$ATT = E(Y_{1i} - Y_{0i} | D=1) = E(Y_{1i} | D=1) - E(Y_{0i} | D=1)$$

- ◉ Los beneficios obtenidos se pueden comparar directamente el costo para, de este modo, analizar los costos y beneficios del programa específico (Heckman, Lalonde y Smith, 1999).

Metodología básica

- ◉ Existen varias alternativas metodológicas (Todd, 2006) aunque todas coinciden en intentar simular las condiciones de un diseño experimental a partir de bases de datos de beneficiarios y no-beneficiarios y utilizando avanzadas técnicas estadísticas.
- ◉ Estas técnicas se aplican una vez que la intervención del programa ya tuvo lugar, es decir, una vez que se seleccionaron los beneficiarios del programa usando métodos no aleatorios.
- ◉ Por esta razón, en un diseño cuasi-experimental los grupos de “control” y “tratamiento” no son iguales entre sí y, por lo tanto, el corazón de esta metodología de evaluación consiste en aplicar sofisticados controles estadísticos para eliminar o minimizar estas diferencias o sesgo de selección (el cual resulta del hecho que los individuos no se seleccionan en una forma que no este relacionada con el resultado, inciden factores como motivación habilidad natural, etc).

Metodología básica

MÉTODOS PARA RESOLVER EL PROBLEMA DEL CONTRAFACTUAL

- ◉ Evaluaciones aleatorias
- ◉ Métodos de Matching, Propensity Score Matching (PSM)
- ◉ Métodos de dobles diferencias
- ◉ Métodos de variables instrumentales
- ◉ Regresiones discontinuas
- ◉ Impactos distribucionales
- ◉ Modelos estructurales

Metodología básica

- Uno de los métodos más utilizados para la medición de impacto de un programa, es el de *matching* de datos que corrige las diferencias observables entre el grupo de tratamiento y el grupo de control, buscando para cada individuo de la muestra del grupo de tratamiento al individuo el más parecido de la muestra del grupo de comparación (Rosenbaum y Rubin, 1983, Todd, 1999)

Metodología básica

- ◉ Matching es una aproximación estadística que resuelve el problema de la evaluación mediante la búsqueda en un grupo de no participantes a aquellos individuos que son similares a los participantes en **todas** las características X pre-tratamiento relevantes.
- ◉ Matching asume que no hay sesgo de selección basado en características no observadas
- ◉ El método de matching estima el contrafactual reponderizando el grupo de control para que tenga la misma distribución de variables X que el grupo de tratamiento

Metodología básica

SUPUESTOS DE IDENTIFICACION

- ⦿ **Supuesto de independencia condicional:** la participación es independiente de los resultados una vez que se ha controlado por las características observables X
- ⦿ **Condición de soporte común:** se comparan individuos comparables

Metodología básica

Supuesto de independencia condicional (SIC):

- ◉ Supone que la selección del tratamiento es "en los observables" y por lo tanto puede ser eliminado por el **condicionamiento de un amplio conjunto de covariables observadas.**
- ◉ Para el parámetro ATT: $E(Y_0 | X, D=1) = E(Y_0 | X, D=0)$
- ◉ El SIC se cumplirá si X incluye todas las variables que afectan tanto a la participación y los resultados.
- ◉ Teoría y conocimiento institucional puede indicar si el SIC es probable que dispongan de un determinado X.
- ◉ **EL SIC IMPLICA LA AUSENCIA DE SESGO DE SELECCIÓN**

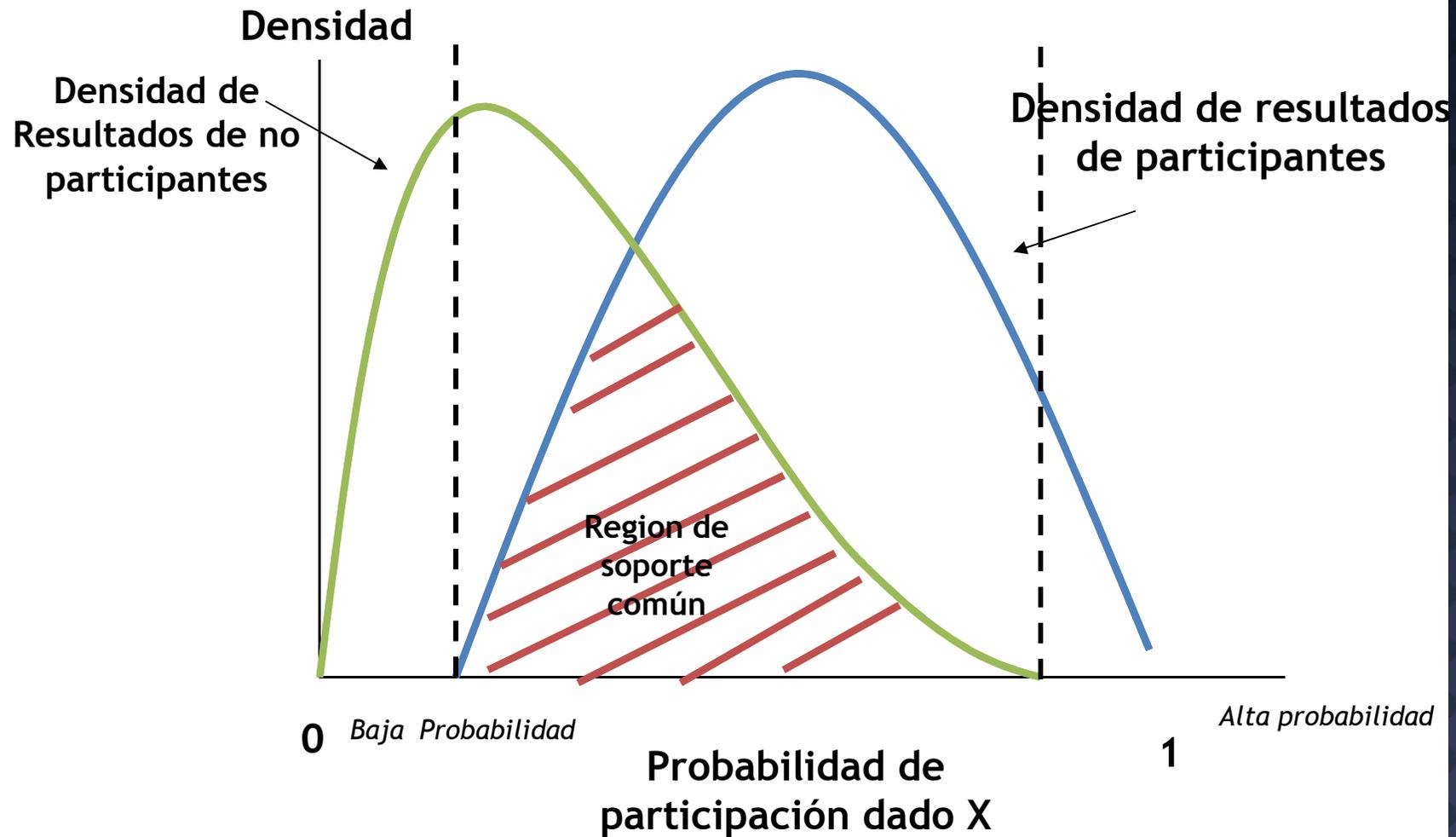
Metodología básica

Condición de Soporte Común:

- ⦿ Los individuos con los mismos valores de X tienen una probabilidad positiva de ser los participantes y no participantes.
- ⦿ Para el parámetro ATT: $P(D = 1 | X) \leq 1$
- ⦿ El soporte asegura la existencia, en la población, de observaciones de no tratados que se "parecen" a las observaciones sin tratar

Metodología básica

Soporte Común



Propensity Score Matching (PSM)

- ◉ ¿Si tenemos muchas variables X , cómo se realiza el matching?
- ◉ De acuerdo a Rosenbaum y Rubin (1983):
 - Si se cumple: $E(Y_0 | X, D=1) = E(Y_0 | X, D=0)$
 - Entonces también se cumple:
 $E(Y_0 | P(X), D=1) = E(Y_0 | P(X), D=0)$

Dónde $P(x)$ es el propensity score o la probabilidad de participar

Propensity Score Matching (PSM)

- ◉ El *Propensity Score Matching* (PSM) consiste en modelar estadísticamente la participación en el programa por medio de una regresión logística, para luego estimar la probabilidad de participar en el programa dadas sus características de elegibilidad de los individuos de ambas muestras (Rosenbaum y Rubin, 1983):

$$P_i = \varphi + \delta X_i + u_i \quad (3)$$

- ◉ Donde: P_i es igual a 1 si el individuo i participa en el programa y 0 de lo contrario. X_i son las variables que afectan la participación en el programa y u_i es un término de error aleatorio.

Propensity Score Matching (PSM)

- ◉ El SIC en el PS implica que es necesario que se cumpla la propiedad de balanceo (Rosenbaum y Rubin, 1983):

$$D \perp X \mid p(X) \quad (2)$$

- ◉ Es decir, para aquellas observaciones con el mismo propensity score, la distribución de las características antes del tratamiento debe ser la misma para los grupos de tratamiento y control.
- ◉ Esto implica que, cada individuo tiene la misma probabilidad de asignación al tratamiento, condicionada en el PS, tal como en un experimento aleatorio.

Propensity Score Matching (PSM)

PROCEDIMIENTO

1. Tener una base de datos representativa y comparable de participantes y no-participantes.
2. Estimar la probabilidad de participación del programa como una función de características observables
3. Usar los valores estimados para generar el propensity score $P(x)$, para ambos grupos.

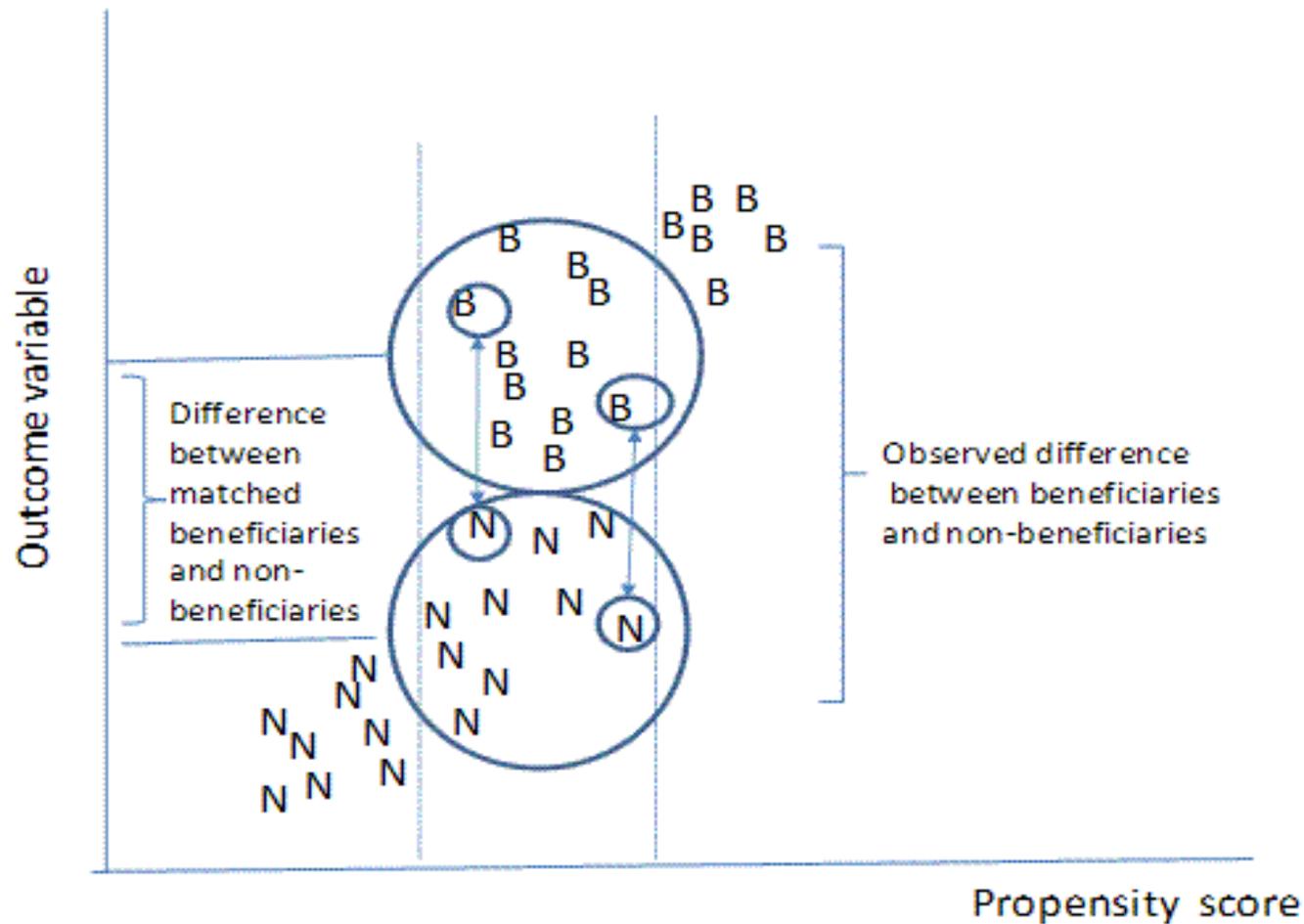
Propensity Score Matching (PSM)

PROCEDIMIENTO

4. Emparejar a los participantes: encontrar una muestra de no-participantes con $p(x)$ similar. Restringir la muestra para asegurar el soporte común.
5. Una vez encontrado el matching calcular el impacto con los promedios de los resultados en todos los participantes

Propensity Score Matching (PSM)

Figure 1. A graphical representation of matching on the propensity score



Propensity Score Matching (PSM)

Los distintos métodos se pueden clasificar de una manera general en dos grupos (Todd, 1999):

- *Estimadores matching de corte transversal.* Este estimador compara los resultados para las personas beneficiarios y el grupo de comparación, medidos en el mismo período de tiempo después del programa.
- *Estimadores matching diferencias en diferencias.* Compara los cambios en los resultados para los beneficiarios con respecto a los cambios en los resultados en el grupo de comparación. En este caso los cambios son medidos respecto a un período de referencia antes del programa.

Propensity Score Matching (PSM)

ATT en PSM

$$ATT = E(Y_1 - Y_0 | X, D=1) - E(Y_0 | X, D=1)$$

$$ATT = E \{ Y_1 - Y_0 | X, D=1, p(X) \}$$

$$ATT = E \{ E \{ Y_1 | D=1, p(X) \} - E \{ Y_0 | D=0, p(X) \} | D=1 \}$$

- Donde D es una variable *dummy* que indica la participación (1) o no participación (0) en el programa. El parámetro ATT mide la ganancia media para los individuos que eligieron participar con respecto de la situación que habrían experimentado sin participar.

Estimadores de Matching

- Todos los estimadores de matching son estimadores ponderados en los cuales los individuos no tratados que estén más cercanos a los tratados (en términos de X) reciben el mayor peso.

$$\Delta_{ATT} = \frac{1}{n_i} \left[\sum_{i \in C} Y_i - \left[\sum_{j \in C} W(\rho_i, \rho_j) Y_j \right] \right]$$

donde n_i es el número de individuos tratados, y $w(\rho_i, \rho_j)$ es el peso de la observación sin tratamiento j en la construcción del contrafactual para la observación del i -ésimo individuo de tratamiento.

Diferentes estimadores de matching difieren en cómo construir el peso $w(i, j)$.

Estimadores de Matching

- ◉ *Nearest Neighbor Matching*

Con este método se compara el resultado que obtiene cada beneficiario tratado con el grupo de control que tenga el propensity score más cercano. Así se calcula la diferencia entre cada par de unidades emparejadas en la variable de interés y se promedian todas las diferencias para calcular el ATT.

Esto es: $C(i) = \min \| p_i - p_j \|$

Donde $C(i)$ representa el conjunto de unidades de control emparejadas a la unidad tratada i , con un valor estimado del propensity score p_i

Así la fórmula del estimador es:

$$ATT^{NN} = \frac{1}{N^T} \sum_{i \in T} \left[Y_i^T - \sum_{j \in C(i)} w_{ij} Y_j^c \right]$$
$$= \frac{1}{N^T} \left[\sum_{i \in T} Y_i^T - \sum_{i \in T} \sum_{j \in C(i)} w_{ij} Y_j^c \right]$$

Donde N^T es el número de unidades tratadas en la muestra, y las $w_{ij} = \frac{1}{N_i^c}$ son ponderaciones si $j \in C(i)$ y $w_{ij} = 0$ de otra forma.

Estimadores de Matching

Nearest Neighbor Matching

- ⦿ Con o sin reemplazo
- ⦿ Con reemplazo mejora la calidad del pareo pero puede aumentar la varianza
- ⦿ Desventaja, puede ser que el N vecino más cercano esté demasiado lejos, lo que disminuye la calidad del pareo

Estimadores de Matching

- ◉ *Kernel Matching*

En el método de Kernel Matching todas las observaciones tratadas son emparejadas con un promedio ponderado de todas las unidades de control.

Las ponderaciones empleadas son inversamente proporcionales a la distancia entre los propensity scores de las unidades tratadas y de control. El estimador se calcula de la forma siguiente:

$$ATT^K = \frac{1}{N^T} \sum_{i \in T} \left\{ Y_i^T - \frac{\sum_{j \in C} Y_j^C K\left(\frac{p_j - p_i}{h_n}\right)}{\sum_{k \in C} K\left(\frac{p_k - p_i}{h_n}\right)} \right\}$$

donde $K(\cdot)$ es una función kernel y h_n es un parámetro de ancho de banda.

Estimadores de Matching

Kernel Matching

- ⦿ Le da un peso decreciente a observaciones que se encuentran más lejos de cada observación tratamiento
- ⦿ Suele utilizar más observaciones que los métodos anteriores (aunque en realidad depende del tipo de Kernel), por ello hay que tener especial cuidado con determinar cuales observaciones están en el soporte común

Estimadores de Matching

- ◉ *Estimador Caliper*

El estimador Caliper es una variación del Nearest Neighbor que permite a los matches con tolerancia δ en la distancia

$\|p_i - p_j\|$ con la finalidad de evitar que diferencias del grupo del control este demasiado lejos del grupo de tratamiento.

Para estimar el caliper matching del neighborhood del grupo de tratamiento i :

$$C(P_i) = \left\{ j : \delta > \|P_i - P_j\| = \min_{k \in I_0} \|P_i - P_k\| \right\}$$

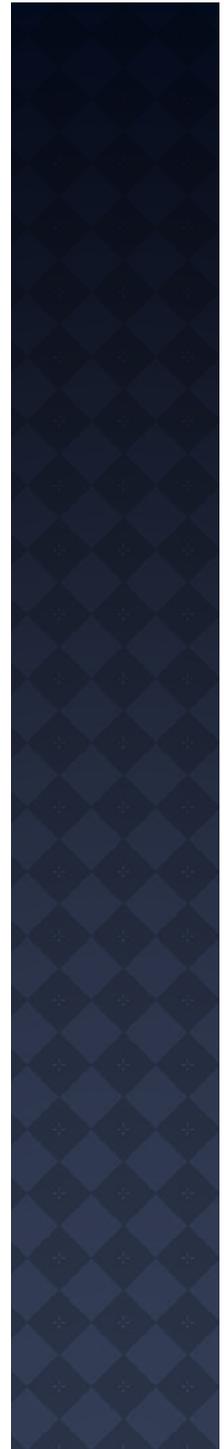
y un solo individuo de comparación es igual cuando su distancia al individuo de tratamiento es la más baja y por debajo de la tolerancia δ . Si ninguno de los individuos de comparación se encuentra dentro del criterio de la tolerancia, entonces el tratamiento i es unmatched.

Los pesos están dados por $W(i, j) = 1[j = k]$

Estimadores de Matching

Estimador Caliper

- ⦿ NN vecinos más cercanos sólo si están en una distancia suficientemente cercana
- ⦿ Es una manera de imponer la condición de soporte común



Estimadores de Matching

- Local-Linear matching

El estimador Local-Linear *matching* es una versión generalizada del Kernel *matching* que añade un término lineal P_i a la constante en una regresión no paramétrica de la variable de resultado.

Los pesos están dados por:

$$W(i, j) = \frac{G_{ij} \left(\sum_{k \in I_0} G_{ik} (P_k - P_i)^2 \right) - (G_{ij} (P_k - P_i)) \left(\sum_{k \in I_0} G_{ik} (P_k - P_i) \right)}{\sum_{j \in I_0} G_{ij} \sum_{k \in I_0} G_{ik} (P_k - P_i)^2 - \left(\sum_{j \in I_0} G_{ij} (P_k - P_i) \right)^2}$$

Donde $G_{ij} = G\left(\frac{P_j - P_i}{h_n}\right)$

PSTEST

- ◉ La propiedad de equilibrio asegura que no hay diferencias significativas en la distribución de las características de pre-tratamiento observables entre las observaciones de los grupos de tratamiento y control con matching, es decir, que las observaciones con el tratamiento y control sean lo más similar posible.
- ◉ Esta propiedad se prueba por medio de el PSTEST (Leuve y Sianesi, 2003), esta prueba viene originalmente de Rosenbaum y Rubin (1985) y se basa en el examen de las diferencias estandarizadas. La prueba calcula para cada variable coincidente la medida de la reducción de sesgos:

$$d = \frac{100(\bar{X}_{\text{trat}} - \bar{X}_{\text{control}})}{\sqrt{\frac{S^2_{\text{trat}} + S^2_{\text{control}}}{2}}}$$

- ◉ Dónde \bar{X} es la media y S^2 la desviación estándar.

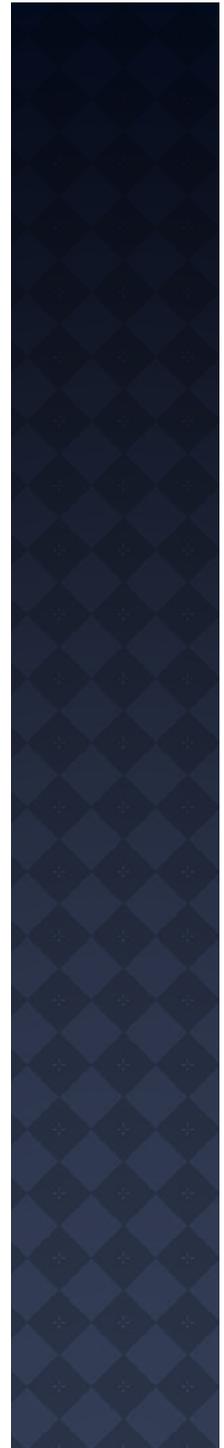
Ventajas de usar PSM

- ⦿ El grado en el cual características observadas guían la participación en el programa.
- ⦿ No requiere necesariamente una línea base o una encuesta de panel
- ⦿ Es un método semiparamétrico, por lo que impone pocas restricciones en la forma funcional del modelo de tratamiento, y también pocos supuestos sobre el término de error.

Consideraciones prácticas

Indicadores de Impacto

- ◉ Se clasifican en 2 grupos:
 - Indicadores finales: miden los resultados de los programas (ej. mayor ingreso per cápita).
 - Indicadores intermedios: miden los insumos de los programas (ej. esquemas de subsidio) y los productos del programa (ej. Carreteras construidas)

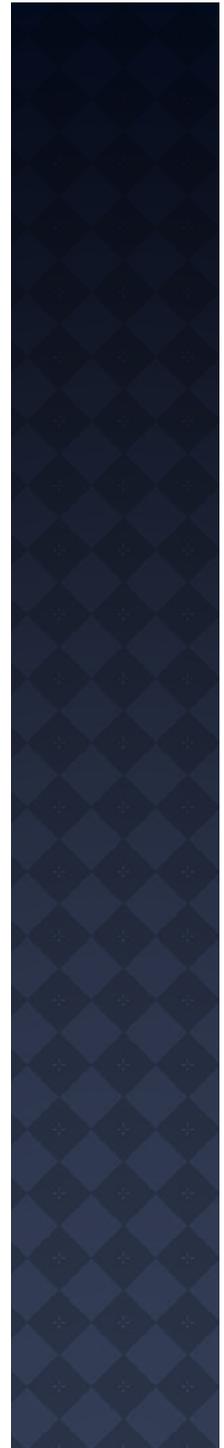


Consideraciones prácticas

Indicadores de Impacto

- ⦿ Se definen de manera más precisa si:
 - Existe un buen entendimiento de los insumos, productos y la etapa del proyecto.
 - Se especifica el nivel y unidad de medida de los indicadores
 - Levantamiento frecuente de información para ver como los productos intermedios están evolucionando y si es necesario revisar el indicador

- ⦿ Para el levantamiento de información tiene que estar definido:
 - Timing del levantamiento
 - Tipo de instrumento
 - Nivel de recolección



Consideraciones prácticas

¿Cómo elegir las X?

- ⦿ Se deben incluir todas las variables que afecten la decisión de participación y la variable de resultado
 - Teórica económica y conocimiento del entorno
 - Criterios basados sólo en el propensity score (significancia estadística)
- ⦿ Entre más variables condicionales, con mayor facilidad se puede cumplir el SIC, pero incluir variables irrelevantes puede disminuir el soporte común.

Cuando usar PSM

- ◉ Matching es plausible cuando se tiene una base de datos rica en características observables.
- ◉ Matching no es apropiado cuando la selección en el tratamiento se basa en características no observables que se correlacionan con los resultados de interés.
- ◉ Matching no es apropiado cuando los efectos de equilibrio general son importantes. Cuando GE están presentes, el tratamiento también (indirectamente) afecta a las observaciones de los no-tratados, lo que dificulta la estimación de los resultados del contrafactual.
- ◉ En resumen, la importancia de hacer Matching depende de las instituciones y la disponibilidad de datos para cada programa en particular.

Rutinas de STATA para realizar Matching

psmatch2.do

E. Leuven y B. Sianesi

<http://ideas.repec.org/c/boc/bocode/s432001.html>

pscore.do; attnw.do; attnd.do; attr.do; atts.do

Sasha Becker y Andrea Ichino

<http://www.iue.it/Personal/Ichino>.

match.do

G. Imbens y Abadie.

<http://elsa.berkeley.edu/~imbens/estimators.shtml>

Evaluación de políticas públicas con Microsimulaciones

Karina Caballero y Jimmy Ferrer

Febrero 2011