

Microeconometría II

Lecture 2

Microeconometría II

1 Problemas al Utilizar Datos Observacionales

- Sesgos por características observables y no observables

2 Sesgos y problema de selección

- Selección sobre observables

3 Matching

- Matching a través de características observables
- Estimación de la varianza
- Estimadores de matching sugeridos por Abadie and Imbens (2002)
- Matching a través del propensity score
- Propensity Score Matching en Stata
- Estimadores de Heckman, Ichimura y Todd

4 EPEM

- Emparejamiento con Características Observadas

5 Inversa del Propensity Score como Ponderador

- Método de la IPW

6 Propensity Score en el Ajuste por Regresión

Agenda

- 1 Problemas al Utilizar Datos Observacionales
 - Sesgos por características observables y no observables
- 2 Sesgos y problema de selección
 - Selección sobre observables
- 3 Matching
 - Matching a través de características observables
 - Estimación de la varianza
 - Estimadores de matching sugeridos por Abadie and Imbens (2002)
 - Matching a través del propensity score
 - Propensity Score Matching en Stata
 - Estimadores de Heckman, Ichimura y Todd
- 4 EPEM
 - Emparejamiento con Características Observadas
- 5 Inversa del Propensity Score como Ponderador
 - Método de la IPW
- 6 Propensity Score en el Ajuste por Regresión

Ejemplo

- Supongamos que se desea conocer el efecto de un programa educativo que se implementa en niños de 5 años, a partir de los resultados de un test cognitivo realizado cuando los niños cumplen diez años de edad. El programa constituye un *tratamiento* mientras que el test realizado a la edad de 10 es la variable respuesta. ¿Cómo podemos determinar si el tratamiento es efectivo?
- Es necesario comparar los dos puntajes *potenciales* del test a la edad 10, el primer resultado (identificado como y_1) sería el obtenido si se aplicó el tratamiento y el segundo (identificado como y_0) si no se aplicó el mismo. Si $y_1 - y_0 > 0$ podemos decir que el programa funciona. Sin embargo, nunca podemos observar y_1 e y_0 para el mismo niño ya que es imposible ir hacia atrás en el tiempo y “deshacer” el tratamiento.
- En la implementación de esta política no es posible/ético utilizar una aleatorización.

Ejemplo

- Alternativa a la aleatorización: utilizar **datos observacionales** donde los niños (sus padres) auto-seleccionen el tratamiento.
- Supongamos que el tratamiento es muy bueno y por lo tanto se requiera un pago elevado. Entonces el grupo T será marcadamente diferente al grupo C : los niños del grupo T podrían tener un nivel cognitivo inferior a los 5 años y padres más ricos.
- Sea W un conjunto de **variables observables** mientras que ε representa **variables no observables** que pudieran ser relevantes para y . Por ejemplo, W pueden ser características de habilidad cognitivas a los 5 años de edad (*baseline variables*) y el nivel de ingreso de los padres, mientras que ε representa características genéticas de los niños y su estilo de vida.

Ejemplo

- Supongamos que se ignoran las diferencias entre los grupos tratamiento y control tanto en W como en ε y se comparan los resultados de los tests a los 10 años.
- Como el grupo T probablemente estaba conformado por niños con menores habilidades cognitivas, los resultados de los tests (a los 10 años) en el grupo T pueden llegar a ser menores a los obtenidos en el grupo C .
- El programa puede haber funcionado, pero no lo suficientemente bien. Podemos concluir erróneamente que no hay efecto tratamiento o más aún que tiene un efecto negativo.

Ejemplo

- Claramente esto no es correcto ya que se han comparado individuos que no son comparables en el sentido de que ambos grupos eran diferentes tanto en términos de W como de ε .
- La diferencia de medias entre grupos $E(y|s = 1) - E(y|s = 0)$ no tiene porque ser equivalente a $E(y_1) - E(y_0)$ ya que:

$$\begin{aligned} E(y|s = 1) - E(y|s = 0) &= E(y_1|s = 1) - E(y_0|s = 0) \\ &\neq E(y_1) - E(y_0) \end{aligned}$$

- $E(y_1|s = 1)$ es la respuesta media del grupo tratamiento para un grupo menos hábil y más rico de niños, por lo tanto probablemente sea diferente a $E(y_1)$, que representa la respuesta media de los grupos T y C combinados.
- Del mismo modo, $E(y_0|s = 0) \neq E(y_0)$.

Ejemplo

- Las diferencias en las características observables W a través de los dos grupos pueden generar un **sesgo por características observables** con respecto a $E(y_1 - y_0)$
- Las diferencias en términos de variables no observables ε pueden causar **sesgos por características no observables**.
- Tratar con las diferencias en las W y ε es una de las principales tareas para encontrar efectos tratamientos a partir de datos observacionales.
- Si no hay diferencias en términos de ε , la manera clásica de remover las diferencias en w es elegir los individuos que conformarán T y C de manera tal que compartan las mismas w , lo que habitualmente se denomina **matching**. En nuestro ejemplo, se compararán niños con las mismas habilidades cognitivas iniciales (a los 5 años) y el mismo nivel de ingreso de los padres:

$$\begin{aligned} E(y|w, s = 1) - E(y|w, s = 0) &= E(y_1|w, s = 1) - E(y_0|w, s = 0) \\ &= E(y_1|w) - E(y_0|w) \end{aligned}$$

- Se cumplen los supuestos 1 y 2.

Ejemplo. Selección Basada en Observables

- La variable s en $E(y_j|w, s)$ desaparece una vez que se condiciona por w como si s fuera aleatorizada dado o condicional a w .
- El supuesto $E(y_j|w, s) = E(y_j|w)$ se denomina 'selección basada en observables' o *ignorabilidad del tratamiento*.
- Si el *efecto condicional* $E(y_1 - y_0|w)$ está identificado podemos calcular un promedio "w-ponderado". La elección de la función de ponderación refleja la importancia de la subpoblación caracterizada por w .
- Por ejemplo, si los niños con menores ingresos son más importantes para el programa educativo, entonces puede asignarse un mayor peso que el actual a la subpoblación de niños en esta situación.

Emparejamiento (Matching) Exacto

- Recuerde que nuestros supuestos 1 y 2 implicaban que dos unidades de diferentes grupos pero con las mismas variables observables deberían tener el mismo resultado potencial esperado: $E[Y(s) | S = 1, W] = E[Y(s) | S = 0, W]$.
- Esta observación sugiere la siguiente estrategia para calcular $E[Y(0) | S = 1, W = w]$:
 - 1 Encuentre las unidades del grupo de control con las mismas w . Para estas observaciones $Y_i = Y(0)$.
 - 2 Tome el promedio para estas unidades: \bar{Y}_C . A medida que el tamaño muestral crece, este promedio converge a $E[Y(0) | S = 1, W = w]$.
- La estrategia para calcular $E[Y(1) | S = 1, W = w]$ es la misma pero buscando las unidades en el grupo de tratamiento.

Emparejamiento (Matching) Exacto

- Este procedimiento sugiere emparejar las unidades del grupo control (tratamiento) con otras unidades del grupo de tratamiento (control) que tengan las mismas características observables, w .
- La técnica de estimación basada en esta estrategia se denomina en la literatura **Matching (emparejamiento)**.
- Existen dos potenciales problemas con el matching:
 - ▶ **Problema de la dimensión:** si w es de dimensión alta es difícil encontrar individuos para el grupo tratamiento y control que compartan exactamente el mismo w .
 - ▶ **Problema de soporte:** Los grupos T y C no estarán superpuestos en w . Por ejemplo, supongamos que w representa el ingreso de la familia por año y $s = 1[w > \tau]$ donde $\tau = 100.000$. Entonces, el grupo T está compuesto por niños que provienen de hogares de mayores ingresos que los niños del grupo C y por lo tanto no hay superposición de ambos grupos en términos de w .

Matching

- Para que las características observables W generen sesgo es necesario que W cambie la probabilidad de recibir el tratamiento.
- Una alternativa para evitar el problema de dimensión del matching sobre w : en lugar de realizar un matching considerando los valores de w , hacerlo sobre una variable unidimensional denominada **propensity score**:

$$\pi(w) \equiv P(s = 1|w) = E(s|w)$$

- De esta manera, se debe calcular $\pi(w)$ para ambos grupos y realizar el matching solamente en base a $\pi(w)$.
- En la práctica, $\pi(w)$ se puede calcular utilizando modelos Logit o Probit.

Matching

- El problema del soporte está dado por la combinación de dos factores:

$$s = 1[w \geq \tau], \quad \text{y} \quad w \text{ afecta a } (y_0, y_1)$$

- De este modo debería controlarse el efecto de w sin embargo es imposible porque no hay superposición en términos de w . Debido a que $s = 1[w \geq \tau]$, $E(y_0|w)$ y $E(y_1|w)$ tienen una discontinuidad en $w = \tau$.
- Caso especial: análisis del tipo antes-después si w representa el tiempo.
- El problema del soporte no se puede evitar pero los individuos que se encuentran cercanos al punto de la discontinuidad probablemente sean similares y por lo tanto comparables.
- Esta comparabilidad conduce a una aleatorización “de borde” que identifica $E(y_1 - y_0|w \simeq \tau)$, el efecto medio para la subpoblación $w \simeq \tau$.

Matching

- Supongamos que no hay problemas de dimensión o soporte y que queremos encontrar elementos para conformar un grupo de control comparable para cada elemento del grupo tratamiento a través del matching. Los individuos del grupo control “matcheados” conforman el grupo de comparación. Se deberá decidir:
- ① Cuantos controles se necesitan para cada individuo tratado (“matching por pares” o “matching múltiple”).
- ② En el caso de matching múltiple, exactamente cuantos y si ese número es el mismo para todos los tratados.
- ③ Se debe determinar si un control se asociará una o varias veces.
- ④ Cuando descartar un individuo tratado si no se pudo realizar un buen matching con un individuo de control.
- ⑤ Determinar que es un buen matching, que distancia debe considerarse para $|w_0 - w_1|$.

Matching

- A partir de estas decisiones habrá nuevos grupos T y C .
- Habrá un nuevo grupo de tratamiento T' si se descartaron individuos que no tienen su correspondiente matching en el grupo de comparación.
- El éxito del matching estará medido al chequear el balance de w a través de los individuos de los nuevos grupos.
- Algunas veces, el matching puede remover el sesgo por factores no observables:
 - ▶ por ejemplo, si dos gemelos idénticos se asignan uno a T y el otro a C , entonces el efecto no observable de los genes se controla.
 - ▶ si se consideran dos hermanos de la misma familia y se asignan uno a T y el otro a C , entonces la influencia no observable de los padres o el ambiente familiar se controlan, por lo menos en alguna medida.

Agenda

- 1 Problemas al Utilizar Datos Observacionales
 - Sesgos por características observables y no observables
- 2 Sesgos y problema de selección
 - Selección sobre observables
- 3 Matching
 - Matching a través de características observables
 - Estimación de la varianza
 - Estimadores de matching sugeridos por Abadie and Imbens (2002)
 - Matching a través del propensity score
 - Propensity Score Matching en Stata
 - Estimadores de Heckman, Ichimura y Todd
- 4 EPEM
 - Emparejamiento con Características Observadas
- 5 Inversa del Propensity Score como Ponderador
 - Método de la IPW
- 6 Propensity Score en el Ajuste por Regresión

Sesgo y selección

- Cuando la asignación del tratamiento depende de características observables, la estimación del efecto tratamiento promedio a partir de la diferencia de medias estará sesgada.
- El sesgo por características observables se puede evitar controlando por w .
- Aún en el contexto de corte transversal, si bien las variables de cada individuo se observan una sola vez, las mismas se refieren a distintos momentos del tiempo.
- La asignación del tratamiento debe ser anterior al resultado. La distinción entre variables pre y post-tratamiento es importante para controlar w .
- En general se controlan las variables pre-tratamiento observadas.

Selección sobre observables

- Consideremos la variable y_j y su distribución f .
- Para que la selección sobre observables sea válida, se debe verificar que s está determinado:
 - 1 por los valores observados de w
 - 2 y posiblemente por algunas variables no observables independientes de y_j dado w .
- De este modo, s se vuelve irrelevante para y_j una vez que se condiciona por w (es decir, $y_j \perp\!\!\!\perp s|w$, $\perp\!\!\!\perp \equiv$ “es independiente de”).
- La condición de selección sobre observables es la siguiente:

$$f(y_j|s) \neq f(y_j) \quad \text{pero} \quad f(y_j|s, w) = f(y_j|w)$$

Selección sobre no observables

- Una situación más general que la anterior es la que se da cuando la asignación del tratamiento depende tanto de características observables como no observables, llamada **selección sobre no observables**:

$$f(y_j|s, w) \neq f(y_j|w) \quad \text{pero} \quad f(y_j|s, w, \varepsilon) = f(y_j|w, \varepsilon), \text{ para algún } \varepsilon$$

- Para que la selección sobre no observables sea válida, s debería estar determinada:
 - posiblemente por los valores observados de w ,
 - por las variables no observables ε relacionadas con y_j dado w ,
 - y posiblemente por algunas variables no observables independientes de y_j dado w y ε .
- En esta situación, s se vuelve irrelevante para y_j solo si tanto w como ε fueron condicionadas (es decir, $y_j \perp\!\!\!\perp s | (w, \varepsilon)$)

Selección sobre observables y no observables

- En general el objetivo del análisis es estimar el efecto medio de un tratamiento (ATE). En ese caso se pueden redefinir las nociones de selección de la siguiente manera:
- Selección sobre observables:

$$E(y_j|s) \neq E(y_j) \quad \text{pero} \quad E(y_j|s, w) = E(y_j|w)$$

- De acuerdo a esta definición, la selección sobre observables es equivalente a la noción de ignorabilidad del tratamiento, condicional a w : $y_j \perp s | w$ ($\perp \equiv$ “no está correlacionado con”).
- Note que para estimar el ATE solo necesitamos ausencia de correlación, \perp , y no el supuesto más fuerte de independencia, II .
- Selección sobre no observables:

$$E(y_j|s, w) \neq E(y_j|w) \quad \text{pero} \quad E(y_j|s, w, \varepsilon) = E(y_j|w, \varepsilon)$$

Efectos identificados

- Ya hemos mencionado que si se controla por variables observables se elimina el sesgo por variables observables.
 - La pregunta natural es ¿cuál es el efecto identificado si se remueve ese sesgo?
- 1 Si $y_0 \perp s | w$ entonces es posible identificar $E(y_1 - y_0 | s = 1)$, el efecto sobre los tratados.
 - 2 Si $y_1 \perp s | w$ entonces es posible identificar $E(y_1 - y_0 | s = 0)$, el efecto sobre los no tratados.
 - 3 Si $y_0 \perp s | w$ e $y_1 \perp s | w$ entonces es posible identificar $E(y_1 - y_0)$, el efecto sobre la población.

Efectos identificados

Para verificar (1):

$$\begin{aligned} E(y|s = 1, w) - E(y|s = 0, w) &= E(y_1|s = 1, w) - E(y_0|s = 0, w) \\ &= E(y_1|s = 1, w) - E(y_0|s = 1, w) \\ &\quad + [E(y_0|s = 1, w) - E(y_0|s = 0, w)] \\ &= E(y_1 - y_0|s = 1, w) + \text{sesgo} \end{aligned}$$

donde el sesgo proviene de la comparación de y_0 a través de los grupos T y C .

- La diferencia de medias condicional entre grupos identifica el efecto tratamiento (medio) condicional sobre los individuos que efectivamente recibieron el tratamiento **si el sesgo de comparación de y_0 entre grupos es igual a 0**.
- Esto es equivalente a la condición de selección sobre observables para y_0 : $y_0 \perp s | w$

Efectos identificados

Para verificar (2):

$$\begin{aligned} E(y|s = 1, w) - E(y|s = 0, w) &= E(y_1|s = 1, w) - E(y_0|s = 0, w) \\ &= E(y_1|s = 1, w) - E(y_1|s = 0, w) \\ &\quad + [E(y_1|s = 0, w) - E(y_0|s = 0, w)] \\ &= E(y_1 - y_0|s = 0, w) + \text{sesgo} \end{aligned}$$

donde, ahora, el sesgo proviene de la comparación de y_1 a través de los grupos T y C .

- La diferencia de medias condicional entre grupos identifica el efecto tratamiento (medio) condicional sobre los individuos que no recibieron el tratamiento **si el sesgo de comparación de y_1 entre grupos es igual a 0**.
- Esto es equivalente a la condición de selección sobre observables para y_1 : $y_1 \perp s | w$

Efectos identificados

Para verificar (3) si ambos sesgos (en (1) y (2)) son cero; esto es $y_0 \perp s|w$ e $y_1 \perp s|w$ entonces el efecto sobre los tratados y sobre los no tratados es el mismo y como:

$$E(y_1 - y_0|w) = E(y_1 - y_0|s = 0, w)P(s = 0|w) + E(y_1 - y_0|s = 1, w)P(s = 1|w)$$

los tres efectos son iguales y,

$$E(y|s = 1, w) - E(y|s = 0, w) = E(y_1 - y_0|w)$$

identifica el efecto sobre la población.

Efectos identificados

- En los tres casos integrando sobre w se obtiene el efecto deseado.
- Por ejemplo, el efecto sobre los tratados es:

$$\int E(y_1 - y_0 | s = 1, w) f(w | s = 1) dw \underset{\uparrow}{=} E(y_1 - y_0 | s = 1)$$

por expectativas iteradas

- En los tres casos la pregunta relevante es
- ¿Cuán posible es el supuesto de sesgo por comparación entre grupos igual a 0?.

Efectos identificados

- Intuitivamente, si w es lo suficientemente 'detallado', podría valer.
- Sin embargo, se podría argumentar que si hay dos individuos con el mismo w que eligen tratamientos diferentes, eso puede indicar que difieren en características no observables que afectan a y_j .

Consideremos el modelo:

$$y_{ji} = w_i' \beta_j + u_{ji}, \quad j = 0, 1 \quad s_i = 1[w_i' \alpha + e_i > 0], \quad e_i \perp (u_0, u_1) | w$$

Como u_j es independiente de s dado w la condición de sesgo por comparación entre grupos igual a 0 es válida para $j = 0, 1$:

$$\begin{aligned} E(y_j | s = 1, w) &= w_i' \beta_j + E(u_j | s = 1, w) = w_i' \beta_j \\ &= w_i' \beta_j + E(u_j | s = 0, w) = E(y_j | s = 0, w) \end{aligned}$$

Efectos identificados

- En palabras, imagine a dos personas con el mismo w pero diferentes e .
- Una persona tiene un gran e positivo y la otra un gran e negativo.
- Como resultado, la primera persona tiene $s = 1$ y la otra tiene $s = 0$.
- Si e está relacionado con u_j , entonces, una gran diferencia en e tendrá consecuencias sobre y_j .
- Esto daría lugar a un sesgo por comparación entre grupos distinto de cero.
- Sin embargo, siempre que $e_j \perp\!\!\!\perp (u_0, u_1) | w$ el sesgo por comparación entre grupos es cero.

Aún cuando los efectos estén identificados, aquí aparecen los mismos problemas que en emparejamiento de grupos.

Agenda

- 1 Problemas al Utilizar Datos Observacionales
 - Sesgos por características observables y no observables
- 2 Sesgos y problema de selección
 - Selección sobre observables
- 3 Matching
 - Matching a través de características observables
 - Estimación de la varianza
 - Estimadores de matching sugeridos por Abadie and Imbens (2002)
 - Matching a través del propensity score
 - Propensity Score Matching en Stata
 - Estimadores de Heckman, Ichimura y Todd
- 4 EPEM
 - Emparejamiento con Características Observadas
- 5 Inversa del Propensity Score como Ponderador
 - Método de la IPW
- 6 Propensity Score en el Ajuste por Regresión

Matching: implementación

- Si los grupos T y C son diferentes en términos de variables observadas w , la diferencia en los resultados de y entre ambos grupos no puede atribuirse a las diferencias en el tratamiento.
- Se entiende por *matching* la selección de individuos de los grupos T y C que sean similares en w .
- Si w es de dimensión grande puede ser difícil encontrar individuos "matcheables" en ambos grupos.
- Una manera de evitar el problema de dimensión es realizar el matching a partir del *propensity score*.
- Comenzamos por el emparejamiento a través de las características observables.

Matching: implementación

- Si $E(y_0|s, w) = E(y_0|w)$ es posible identificar el efecto tratamiento de los tratados $E(y_1 - y_0|s = 1)$ como:

$$\begin{aligned} & \int [E(y|w, s = 1) - E(y|w, s = 0)]f(w|s = 1)dw = \\ &= E(y|s = 1) - \int E(y|w, s = 0)f(w|s = 1)dw \\ &= E(y - E(y_0|w, s = 0)|s = 1) \\ &= E(y_1 - E(y_0|w, s = 1)|s = 1) = E((y_1 - y_0)|s = 1) \end{aligned}$$

- El segundo término es un contrafactual que debe construirse.

Matching: implementación

- La forma de construirlo es considerando

$$E(y_0|w, s = 0) = E(y_0|w, s = 1)$$

para un elemento i del grupo T con $w = w_i$ buscamos en el grupo C controles cuyo(s) w es el mismo que w_i .

- Así queda definido el grupo C_i de controles denominado grupo de comparación para el individuo tratado i .
- Supongamos un conjunto de N individuos, N_1 de los cuales pertenecen al grupo tratamiento y los $N_0 = N - N_1$ restantes pertenecen al grupo de control.

Matching: implementación

- EJEMPLO: supongamos que hay $N_1 = 3$ elementos en el grupo T y $N_0 = 4$ elementos en el grupo de control:

$$T_1 : c_1, c_2, c_3 \Rightarrow C_1 = \{c_1, c_2, c_3\}, \quad |C_1| = 3$$

$$T_2 : c_4 \Rightarrow C_2 = \{c_4\}, \quad |C_2| = 1$$

$$T_3 : \text{sin matching} \Rightarrow C_3 = \emptyset, \quad |C_3| = 0$$

- El número de elementos *matcheados* o emparejados con éxito es:

$$N_{1u} = \sum_{i \in T} 1[C_i \neq \emptyset]$$

- Luego veremos cómo seleccionar C_i de la reserva de controles.

Matching: implementación

- Para la observación i , estimaremos la $E(y_0|w_i, s = 0)$ como:

$$\hat{y}_{0i} = \frac{\sum_{m \in C_i} y_{mi}}{|C_i|}$$

donde y_{mi} se refiere a la variable respuesta en C_i .

- Un estimador del efecto tratamiento promedio sobre los tratados es el siguiente:

$$T_N = \frac{\sum_{i \in T_u} (y_i - \sum_{m \in C_i} |C_i|^{-1} y_{mi})}{N_{1u}} = \frac{\sum_{i \in T_u} (y_i - \hat{y}_{0i})}{N_{1u}}$$

donde T_u es el conjunto de los elementos del grupo tratamiento “usados” (es decir, emparejados con éxito)

$$T_u = \{i : |C_i| \neq \emptyset, i \in T\}$$

Matching: implementación

- Sea X_M el conjunto de valores de w que son emparejables (matcheables).
- La versión poblacional de T_N se puede escribir como:

$$\begin{aligned} E(y - E(y_0|w, s = 0)|s = 1, w \in X_M) &= \\ &= E(y_1 - E(y_0|w, s = 1)|s = 1, w \in X_M) = \\ &= E(y_1|s = 1, w \in X_M) - E(y_0|s = 1, w \in X_M) \end{aligned}$$

- Si $|C_i| = 1, \forall i \in T$ se denomina matching de a pares sin eliminación.
- Si $|C_i| = 0, 1, \forall i \in T$:

$$T_{pM} = \frac{\sum_{i \in T_u} (y_i - y_{mi})}{N_{1u}}$$

Matching: implementación

- En el otro extremo, cuando todos los elementos del grupo de control se utilizan como grupo comparación para cada elemento del grupo tratamiento:

$$T_M = \frac{\sum_{i \in T} (y_i - \sum_{j \in C} k_{ij} y_{mi})}{N_1}$$

donde

$$\sum_{j \in C} k_{ij} = 1 \quad \forall i, \quad k_{ij} > 0 \quad \forall i, j$$

Matching: implementación

- Estimador de Matching Ponderado

$$T_{wM} = (N_1)^{-1} \sum_{i \in T} \left(y_i - \frac{\sum_{j \in C} K((w_j - w_i)/h) y_j}{\sum_{j \in C} K((w_j - w_i)/h)} \right)$$

- Estimador de Matching Generalizado

$$T_{gM} = (N_{1u})^{-1} \sum_{i \in T_u} (y_i - E_N(y|w, s = 0))$$

- Donde $E_N(y|w, s = 0)$ es una estimación no paramétrica de $E(y|w, s = 0)$ usando C_i

Matching: implementación

- Efecto sobre los no tratados. Definamos T_j a los sujetos del grupo de tratamiento emparejados con el control j y:

$$N_{0u} = \sum_{i \in C} 1[T_j \neq \emptyset]; \quad C_u \equiv \{j : |T_j| \neq \emptyset, j \in C\}$$

- un estimador del efecto sobre los no tratados, $E[E(y|w, s = 1) - y|s = 0]$, que es análogo a T_N es:

$$U_N = \frac{\sum_{j \in C_u} (\sum_{m \in T_j} |T_j|^{-1} y_{mj} - y_j)}{N_{0u}}$$

Se define $N_u = N_{0u} + N_{1u}$

$$\begin{aligned} E(y_1 - y_0 | w \in X_M) &= P(s = 1 | w \in X_M) \cdot E(y_1 - y_0 | s = 1, w \in X_M) \\ &+ P(s = 0 | w \in X_M) \cdot E(y_1 - y_0 | s = 0, w \in X_M) \end{aligned}$$

Un estimador del efecto sobre toda la subpoblación emparejable (matcheable) es:

$$\begin{aligned} T_{UN} &= \frac{N_{1u}}{N_u} \cdot T_N + \frac{N_{0u}}{N_u} \cdot U_N \\ &= \frac{N_{1u}}{N_u} \cdot \frac{\sum_{i \in T_u} y_i - \sum_{m \in C_i} |C_i|^{-1} y_{mi}}{N_{1u}} \\ &+ \frac{N_{0u}}{N_u} \cdot \frac{\sum_{j \in C_u} (\sum_{m \in T_j} |T_j|^{-1} y_{mj} - y_j)}{N_{0u}} \\ &= \frac{\sum_{i \in T_u} y_i - \sum_{m \in C_i} |C_i|^{-1} y_{mi}}{N_u} \\ &+ \frac{\sum_{j \in C_u} (\sum_{m \in T_j} |T_j|^{-1} y_{mj} - y_j)}{N_u} \end{aligned}$$

Matching en la práctica

- El grupo comparación C_i para el individuo tratado i debería elegirse de manera tal que los valores de w sean lo más cercano posibles a w_i (idealmente iguales).
- Si w es muy grande o toma muchos valores diferentes, el matching exacto es muy difícil de alcanzar.
- Se define un pequeño entorno de w_i y cualquier elemento del grupo control cuyo w se encuentre dentro del entorno es considerado un matching.

Matching en la práctica

Hay dos métodos polares:

- Definir un entorno fijo para w_i , que llamaremos N_i , y todos aquellos controles cuyo w caigan en N_i serán considerados un matching. En este caso el entorno es fijo y el tamaño del grupo comparación para i , $|C_i|$ es aleatorio.
- Nearest Neighbor matching: consiste en fijar la cantidad de elementos “vecinos cercanos a i ” pertenecientes al grupo de control que conformarán el grupo de comparación, más allá de su distancia con respecto a w_i . En este caso $|C_i|$ es fijo.

Matching en la práctica

- Para un individuo tratado i , la similitud de un individuo del grupo de control con respecto al primero se puede medir por alguna distancia, por ejemplo la distancia de mahalanobis:

$$(w_i - w_m)' C_N^{-1} (w_i - w_m)$$

donde el subíndice m representa a los elementos del grupo de control y C_N^{-1} es una matriz de covarianzas muestral para w .

- Siguiendo el primer método, para un individuo tratado i incluiremos en C_i sólo aquellos controles que se encuentren a una distancia menor a un cierto umbral (a veces denominado “caliper”).

Matching en la práctica

- Si se desea evitar que la construcción de los grupos de comparación sea sensible a la presencia de outliers o transformaciones monótonas de las variables, en lugar de utilizar distancias es posible utilizar diferencias en rankings:

$$\sum_{j=1}^N |q_{ij} - q_{mj}|$$

donde q_{ij} es el ranking de la j -ésima componente w_{ij} de w_i en la muestra conjunta, $j = 1, 2, \dots, k$ y q_{mj} es el ranking de la j -ésima componente w_{mj} de w_m en la muestra conjunta.

Matching en la práctica

- Algunas veces el matching se realiza en dos etapas. Por ejemplo, consideremos que $w = (w'_1, w'_2)'$ donde w_1 es un vector de variables que deberán emparejarse exactamente .
- En la primera etapa se realiza una estratificación asignando todos los elementos del grupo T y C a los estratos definidos por los valores de las variables en w_1 .
- En la segunda etapa, para el individuo i del grupo T se realiza el matching buscando elementos del grupo C que pertenezcan al mismo estrato.
- Por ejemplo, $w_1 = (\text{género}, \text{edad})$ y $w_2 = (\text{años de educación}, \text{experiencia})$.
- Si consideramos un individuo del grupo T hombre de 40 años, se elegirán uno o más controles del mismo estrato (hombres de 40 años) realizando un matching a partir de w_2 .

Matching en la práctica

- En la práctica hay varios esquemas:
- **MATCHING DE A PARES:** para cada individuo tratado i el grupo comparación consiste de un único control.
- **MATCHING MÚLTIPLE:** el grupo de comparación para cada individuo tratado i está compuesto por más de un control. Puede ser la misma cantidad de controles para cada individuo tratado como en NN o puede diferir a través de los individuos.
- Si cada individuo tratado tiene asociado (matcheado) al menos un control y cada control tiene asociado al menos un individuo tratado entonces hay *full matching*.
- En principio es posible que un mismo individuo del grupo control esté matcheado con múltiples individuos del grupo tratamiento. Cuando un control es matcheado a lo sumo a un individuo tratado se denomina **GREEDY MATCHING**

- **NON-GREEDY SEQUENTIAL MATCHING:** las observaciones son consideradas de a una por vez. Una vez que un individuo del grupo T fue matcheado (o descartado), queda fuera del análisis y la reserva de elementos de control se mantiene igual.
- **GREEDY SEQUENTIAL MATCHING:** las observaciones son consideradas de a una por vez. Una vez que un individuo del grupo T fue matcheado (o descartado), queda fuera del análisis y también lo hace el/los elementos del control. De este modo, la reserva de elementos de control se reduce a medida que avanza el proceso.
- El problema en este caso es que el control matcheado al individuo tratado i puede ser más importante para otro individuo tratado i' que se encuentre más abajo en el proceso.
- Una alternativa superior viene dada por los procedimientos no secuenciales que consideran todos los individuos tratados y de control al mismo tiempo pero desde el punto de vista del cálculo es bastante más difícil.

Agenda

- 1 Problemas al Utilizar Datos Observacionales
 - Sesgos por características observables y no observables
- 2 Sesgos y problema de selección
 - Selección sobre observables
- 3 Matching
 - Matching a través de características observables
 - Estimación de la varianza
 - Estimadores de matching sugeridos por Abadie and Imbens (2002)
 - Matching a través del propensity score
 - Propensity Score Matching en Stata
 - Estimadores de Heckman, Ichimura y Todd
- 4 EPEM
 - Emparejamiento con Características Observadas
- 5 Inversa del Propensity Score como Ponderador
 - Método de la IPW
- 6 Propensity Score en el Ajuste por Regresión

Estimación de la varianza

- Es difícil encontrar la varianza asintótica de los estimadores de matching ya que la selección del grupo comparación para cada individuo i es un proceso que involucra a todas las observaciones, lo que genera dependencia a través de las observaciones de la muestra.
- El problema es más serio a medida que el matching es más elaborado.
- Considerando el estimador más simple bajo matching de a pares, si se asume que las diferencias $(y_i - y_{mi})$ son iid, se podría calcular la varianza como:

$$V_{pN} = N_{1u}^{-2} \sum_{i \in T_u} \left((y_i - y_{mi}) - N_{1u}^{-1} \sum_{i \in T_u} (y_i - y_{mi}) \right)^2$$

Estimación de la varianza

- Sin embargo en la selección de y_{mi} participan los demás controles lo que genera que la presunción de independencia no sea cierta.
- En aquellos procedimientos de matching donde un control se asociado como máximo a un individuo tratado, el supuesto iid no parece muy problemático.
- En aquellos casos donde el grupo control es muy grande con respecto al grupo tratamiento, la posibilidad de que un mismo control se asocie muchas veces es pequeña

Estimación de la varianza

- Una posibilidad para obtener intervalos de confianza asintóticos es utilizar bootstrap:
 - 1 Seleccionar una muestra aleatoria con reemplazo de N_0 elementos a partir del grupo C .
 - 2 Seleccionar una muestra aleatoria con reemplazo de N_1 elementos a partir del grupo T .
 - 3 Aplicar el mismo procedimiento de matching a la pseudo muestra total de $N = N_0 + N_1$ elementos para obtener una pseudo estimación del efecto tratamiento de interés.
 - 4 Repetir este procedimiento B veces para obtener la misma cantidad de replicaciones del efecto tratamiento estimado.
 - 5 Tomar los percentiles correspondientes al 2.5% y 97.5% para obtener un intervalo de confianza del 95% para el efecto tratamiento y comprobar si ese efecto es igual a 0.

Agenda

- 1 Problemas al Utilizar Datos Observacionales
 - Sesgos por características observables y no observables
- 2 Sesgos y problema de selección
 - Selección sobre observables
- 3 Matching
 - Matching a través de características observables
 - Estimación de la varianza
 - Estimadores de matching sugeridos por Abadie and Imbens (2002)
 - Matching a través del propensity score
 - Propensity Score Matching en Stata
 - Estimadores de Heckman, Ichimura y Todd
- 4 EPEM
 - Emparejamiento con Características Observadas
- 5 Inversa del Propensity Score como Ponderador
 - Método de la IPW
- 6 Propensity Score en el Ajuste por Regresión

Abadie and Imbens (2002)

- Para asegurar que los estimadores de matching identifiquen y estimen de manera consistente el efecto tratamiento de interés, se asume que la asignación del tratamiento es independiente de los resultados, condicional a x (Abadie-Imbens usan x en lugar de w):
- Supuesto 2.1: para los x pertenecientes al soporte de X ,
 - ① s es independiente de (y_0, y_1) condicional a x . Este supuesto representa que la asignación del tratamiento es "aleatoria" entre individuos similares.
 - ② $c < P(s = 1|x) < 1 - c$ para algún $c > 0$. Este supuesto corresponde al de selección sobre observables.

Abadie and Imbens (2002)

- Consideremos $\|x\|_V = (x' V x)^{1/2}$ la norma del vector x considerando una matriz de ponderaciones V definida positiva. Se define $\|x - z\|_V$ la distancia entre los vectores x y z .
- Sea $d_M(i)$ la distancia entre las covariables del individuo i , x_i con respecto al M -ésimo match más cercano en el grupo asignado al tratamiento contrario. Considerando la posibilidad de que haya empates, esta distancia es tal que hay estrictamente menos de M elementos más cerca al individuo i que $d_M(i)$ y al menos M elementos tan cercanos a i como $d_M(i)$.
- Formalmente:

$$\sum_{l: s_l = 1 - s_i} 1[\|x_l - x_i\|_V < d_M(i)] < M$$

y

$$\sum_{l: s_l = 1 - s_i} 1[\|x_l - x_i\|_V \leq d_M(i)] \geq M$$

Abadie and Imbens (2002)

- Se define $\mathcal{J}_M(i)$ el conjunto de índices para los matches del individuo i que están al menos tan cerca como el M -ésimo match.

$$\mathcal{J}_M(i) = \{l = 1, 2, \dots, N \mid s_l = 1 - s_i, \|x_l - x_i\|_V \leq d_M(i)\}.$$

- Si no hay empates, el conjunto $\mathcal{J}_M(i)$ tiene M elementos, en general contiene más de M .
- La cantidad de elementos en este conjunto se representa como $\#\mathcal{J}_M(i)$.

Abadie and Imbens (2002)

- Se define $K_M(i)$ como el número de veces que se usa i como emparejamiento de todas las observaciones l del grupo opuesto, cada una de estas veces ponderada por el total de emparejamientos de l .
- $K'_M(i)$ es una medida comparable en la que el cuadrado del número de emparejamientos se usa como ponderador.

$$K_M(i) = \sum_{l=1}^N 1[i \in \mathcal{J}_M(l)] (\#\mathcal{J}_M(l))^{-1}.$$

$$K'_M(i) = \sum_{l=1}^N 1[i \in \mathcal{J}_M(l)] (\#\mathcal{J}_M(l))^{-2}.$$

- Se verifica que

$$\sum_i K_M(i) = N, \quad \sum_{i:s_i=1} K_M(i) = N_0, \quad \sum_{i:s_i=0} K_M(i) = N_1$$

Abadie and Imbens (2002)

- Estimador simple

$$\hat{y}_{0i} = \begin{cases} y_i, & \text{si } s_i = 0; \\ \frac{1}{\#\mathcal{J}_M(i)} \sum_{l \in \mathcal{J}_M(i)} y_l, & \text{si } s_i = 1. \end{cases}$$

$$\hat{y}_{1i} = \begin{cases} \frac{1}{\#\mathcal{J}_M(i)} \sum_{l \in \mathcal{J}_M(i)} y_l, & \text{si } s_i = 0; \\ y_i, & \text{si } s_i = 1. \end{cases}$$

y

Abadie and Imbens (2002)

- Estimador simple para la muestra total:

$$T_M^{sm} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_{1i} - \hat{y}_{0i}) = \frac{1}{N} \sum_{i=1}^N (2s_i - 1)(1 + K_M(i))y_i$$

- Estimador simple para los tratados:

$$T_M^{sm,t} = \frac{1}{N_1} \sum_{i:s_i=1} (y_i - \hat{y}_{0i}) = \frac{1}{N_1} \sum_{i=1}^N (s_i - (1 - s_i)K_M(i))y_i$$

- Estimador simple para los no tratados:

$$T_M^{sm,c} = \frac{1}{N_0} \sum_{i:s_i=0} (\hat{y}_{1i} - y_i) = \frac{1}{N_0} \sum_{i=1}^N (s_i \cdot K_M(i) - (1 - s_i))y_i$$

Estimador corregido por sesgo

- El estimador simple será sesgado en muestras finitas si el matching no es exacto.
- Abadie and Imbens (2002) demostraron que con k covariables continuas el estimador tendrá un término correspondiente a las diferencias en el matching que será del orden $O_p(N^{-1/k})$

$$bias = E \left(\frac{1}{N} \sum_{i=1}^N ((\hat{y}_{1i} - \hat{y}_{0i}) - (y_{1i} - y_{0i})) | \{x_i, s_i\}_{i=1}^N \right)$$

- Si el matching no es exacto (es decir $x_i \neq x_{mi}$) introduce sesgo.
- El estimador corregido por sesgo ajusta las diferencias en el matching por las diferencias en las covariables.

Estimador corregido por sesgo

- El estimador está basado en la estimación de dos funciones de regresión $\mu_j(x) = E(y_j|x)$, con $j=0,1$.
- Supongamos que deseamos estimar el efecto tratamiento promedio (global). En este caso estimamos las funciones de regresión utilizando los datos de la muestra matchheada:

$$\hat{\mu}_j(x) = \hat{\beta}_{j0} + \hat{\beta}'_{j1}x, \quad j = 0, 1$$

donde

$$(\hat{\beta}_{j0}, \hat{\beta}'_{j1}) = \underset{i:s_i=j}{\operatorname{argmin}} \sum K_M(i)(y_i - \beta_{j0} - \beta'_{j1}x_i)^2$$

- Si nos interesa estimar el efecto tratamiento de los tratados sólo necesitamos estimar la regresión para los controles ($\hat{\mu}_0(x)$) y si nos interesa estimar el efecto tratamiento de los no tratados sólo necesitaremos estimar la regresión para los tratados ($\hat{\mu}_1(x)$).

Estimador corregido por sesgo

- Estimador corregido por sesgo

$$\tilde{y}_{0i} = \begin{cases} y_i, & \text{si } s_i = 0; \\ \frac{1}{\#\mathcal{J}_M(i)} \sum_{l \in \mathcal{J}_M(i)} y_l + \hat{\mu}_0(x_i) - \hat{\mu}_0(x_l), & \text{si } s_i = 1. \end{cases}$$

$$\tilde{y}_{1i} = \begin{cases} \frac{1}{\#\mathcal{J}_M(i)} \sum_{l \in \mathcal{J}_M(i)} y_l + \hat{\mu}_1(x_i) - \hat{\mu}_1(x_l), & \text{si } s_i = 0; \\ y_i, & \text{si } s_i = 1. \end{cases}$$

y

Estimador corregido por sesgo

- Estimador corregido para la muestra total:

$$T_M^{bcm} = \frac{1}{N} \sum_{i=1}^N (\tilde{y}_{1i} - \tilde{y}_{0i})$$

- Estimador corregido para los tratados:

$$T_M^{bcm,t} = \frac{1}{N_1} \sum_{i:s_i=1} (y_i - \tilde{y}_{0i})$$

- Estimador corregido para los no tratados:

$$T_M^{bcm,c} = \frac{1}{N_0} \sum_{i:s_i=0} (\tilde{y}_{1i} - y_i)$$

Estimación de la varianza

- Abadie and Imbens (2002) derivaron las propiedades asintóticas de su estimador de matching.

$$\hat{V}^{sample} = \frac{1}{N^2} \sum_{i=1}^N (1 + K_M(i))^2 \cdot \hat{\sigma}_{s_i}^2(x_i)$$

- Estimador corregido para los tratados:

$$\hat{V}^{sample, T} = \frac{1}{N_1^2} \sum_{i=1}^N (s_i - (1 - s_i)K_M(i))^2 \cdot \hat{\sigma}_{s_i}^2(x_i)$$

- Estimador corregido para los no tratados:

$$\hat{V}^{sample, C} = \frac{1}{N_0^2} \sum_{i=1}^N (s_i \cdot K_M(i) - (1 - s_i))^2 \cdot \hat{\sigma}_{s_i}^2(x_i)$$

Estimación de la varianza

- Los estimadores de la varianza para el estimador de matching del efecto tratamiento promedio poblacional son:

$$\begin{aligned}\hat{V}^{POP} &= \frac{1}{N^2} \sum_{i=1}^N \left((\hat{y}_{1i} - \hat{y}_{0i}) - \hat{\tau} \right) \\ &\quad + \frac{1}{N^2} (K_M^2(i) + 2K_M(i) - K'_M(i)) \hat{\sigma}_{s_i}^2(x_i)\end{aligned}$$

- El cálculo de esta varianza requiere estimar la varianza de los resultados potenciales condicional a x , $\hat{\sigma}_{s_i}^2(x_i)$.
- Próximamente veremos que este cálculo se puede realizar bajo dos supuestos: varianza constante para ambos grupos tratamiento y control y todas las covariables o no.

Estimación de la varianza

- Los estimadores de la varianza para el estimador de matching del efecto tratamiento promedio poblacional para los grupos condicional a $s = 1$ y $s = 0$ son:

$$\begin{aligned}\hat{V}^{POP,T} &= \frac{1}{N_1^2} \sum_{i=1}^N \left(s_i(y_{1i} - \hat{y}_{0i}) - \hat{T}^T \right) \\ &+ \frac{1}{N_1^2} (1 - s_i)(K_M^2(i) - K'_M(i)) \hat{\sigma}_{s_i}^2(x_i)\end{aligned}$$

$$\begin{aligned}\hat{V}^{POP,C} &= \frac{1}{N_0^2} \sum_{i=1}^N \left((1 - s_i)(\hat{y}_{1i} - y_{0i}) - \hat{T}^C \right) \\ &+ \frac{1}{N_0^2} (s_i)(K_M^2(i) - K'_M(i)) \hat{\sigma}_{s_i}^2(x_i)\end{aligned}$$

Ejemplo

TABLE 3.2
A MATCHING ESTIMATOR WITH SEVEN OBSERVATIONS, $m = 1$

i	W_i	X_i	Y_i	$\mathcal{J}_1(i)$	$\hat{Y}_i(0)$	$\hat{Y}_i(1)$	$K_M(i)$
1	0	2	7	{5}	7	8	3
2	0	4	8	{4,6}	8	$7\frac{1}{2}$	1
3	0	5	6	{4,6}	6	$7\frac{1}{2}$	0
4	1	3	9	{1,2}	$7\frac{1}{2}$	9	1
5	1	2	8	{1}	7	8	1
6	1	3	6	{1,2}	$7\frac{1}{2}$	6	1
7	1	1	5	{1}	7	5	0

Ejemplo

- En Stata el comando es `nnmatch`

```
nnmatch depvar treatvar varlist_nnmatch [weight] [if exp] [in range] [,  
tc({ate | att | atc}) m(#) biasadj(bias | varlist_adj)]
```

- **nnmatch** estima el efecto medio del tratamiento sobre *vardep* comparando los resultados entre las observaciones tratadas y el control (definidas por *treatvar*), usando el método del vecino más cercano a través de las variables definidas en *varlist_nnmatch*.
- **biasadj(bias | varlist_adj)**: esta opción ajusta (usando regresión) los resultados con la(s) variable(s) de emparejamiento original(es) si se especifica **biasadj(bias)** o con nuevas variables si se especifican en **biasadj(varlist_adj)**.

Ejemplo

```
. nnmatch y w x, m(1)
```

Matching estimator: Average Treatment Effect

Number of obs = 7
Number of matches (m) = 1

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
SATE	.1428571	.9407699	0.15	0.879	-1.701018	1.986732

Matching variables: x

```
. nnmatch y w x, m(1) tc(att)
```

Matching estimator: Average Treatment Effect for the Treated

Number of obs = 7
Number of matches (m) = 1

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
SATT	-.25	1.032669	-0.24	0.809	-2.273995	1.773995

Matching variables: x

```
. nnmatch y w x, m(1) tc(ato)
```

Matching estimator: Average Treatment Effect for the Controls

Number of obs = 7
Number of matches (m) = 1

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
SATC	.6666667	.860663	0.77	0.439	-1.020202	2.353535

Matching variables: x

Agenda

- 1 Problemas al Utilizar Datos Observacionales
 - Sesgos por características observables y no observables
- 2 Sesgos y problema de selección
 - Selección sobre observables
- 3 Matching
 - Matching a través de características observables
 - Estimación de la varianza
 - Estimadores de matching sugeridos por Abadie and Imbens (2002)
 - Matching a través del propensity score
 - Propensity Score Matching en Stata
 - Estimadores de Heckman, Ichimura y Todd
- 4 EPEM
 - Emparejamiento con Características Observadas
- 5 Inversa del Propensity Score como Ponderador
 - Método de la IPW
- 6 Propensity Score en el Ajuste por Regresión

Matching vía Propensity Score

- Definimos $\pi(x) = P(s = 1|x)$ que representa la probabilidad de recibir el tratamiento $s = 1$ condicional a las características observadas x y se denomina *propensity score*.

- Teorema 1 [Rosenbaum & Rubin(1983)]:**

La asignación del tratamiento s y las variables observables x son condicionalmente independientes dado el propensity score:

$$x \perp s | \pi(x).$$

- Se define un *balancing score* $b(x)$ como una función de las variables observables x tal que la distribución condicional de x dado $b(x)$ es la misma para el grupo T ($s = 1$) y el grupo C ($s = 0$):

$$x \perp s | b(x)$$

- El balancing score trivial viene dado por $b(x) = x$.

Matching vía Propensity Score

- **Teorema 3 [Rosenbaum & Rubin(1983)]:**

Si la asignación del tratamiento es fuertemente ignorable dado x , entonces es fuertemente ignorable dado cualquier balancing score $b(x)$:

$$y_j \perp s|x, \quad 0 < P(s = 1|x) < 1 \quad \forall x, j = 0, 1$$

implica

$$y_j \perp s|b(x), \quad 0 < P(s = 1|b(x)) < 1 \quad \forall b(x), j = 0, 1.$$

Matching vía Propensity Score

- **Teorema 4 [Rosenbaum & Rubin(1983)]:**

Supongamos que la asignación del tratamiento es fuertemente ignorable y además $b(x)$ es un balancing score. Entonces la esperanza de la diferencia en las respuestas observadas de los dos grupos en $b(x)$ es igual al efecto tratamiento promedio en $b(x)$:

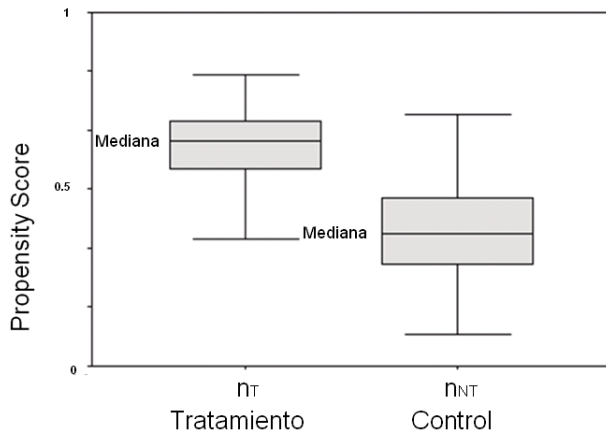
$$E\{y_1|b(x), s = 1\} - E\{y_0|b(x), s = 0\} = E\{(y_1 - y_0)|b(x)\}$$

- En el caso de estudios observacionales, la especificación del propensity score, $\pi(x)$, no es conocida y puede estimarse a partir de los datos observados.
- Los métodos de matching restringen el análisis solo a los casos del grupo tratamiento que tienen un propensity score que cae entre el mínimo y máximo del propensity score del grupo de control (y viceversa).

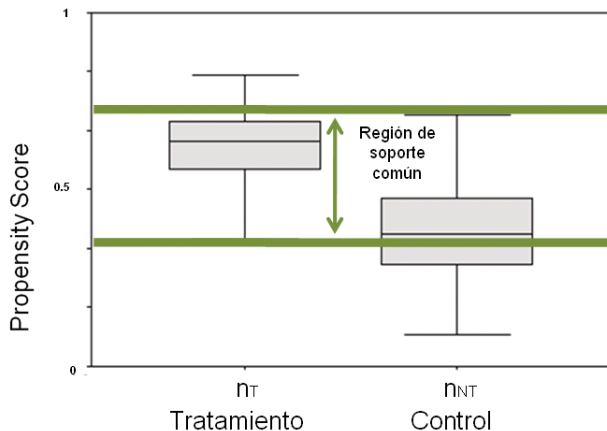
Implementación práctica del Matching

- Para evitar la influencia de observaciones atípicas se define una región de soporte común.
- Usualmente este soporte común se restringe.
- Los casos del grupo de control cuyos propensity score son menores que el propensity score de los casos del grupo de tratamiento al percentil a de la distribución del propensity score del tratamiento se excluyen del análisis.
- Los casos del grupo de tratamiento cuyo propensity score es mayor que el percentil $(100 - a)$ de la distribución en el grupo de control también se excluyen.

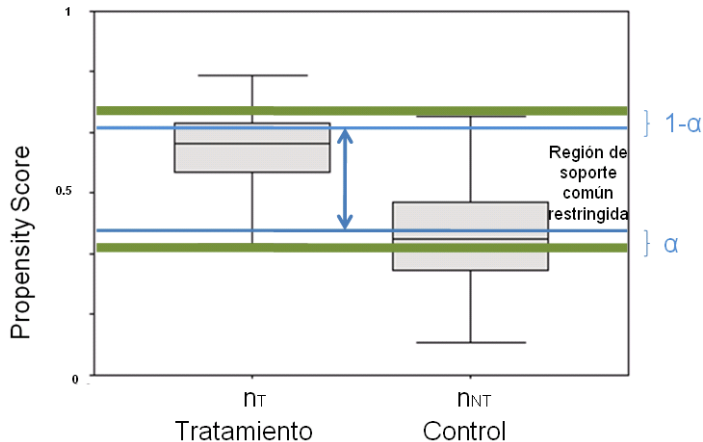
La Región de Soporte Común



La Región de Soporte Común



La Región de Soporte Común



Implementación práctica del Matching

- La distribución de las variables observadas, x , que afectan los resultados podría ser diferente en ambos grupos.
- El sesgo potencial en la estimación del ATE, debido a distribuciones diferentes de las x entre individuos de ambos grupos, se elimina re-ponderando las observaciones del grupo de control.
- Esta re-ponderación puede hacerse mediante una variedad de algoritmos diferentes que incluyen: Nearest-Neighbor Matching, Kernel Matching, Radius Matching, Stratification Matching...

Agenda

- 1 Problemas al Utilizar Datos Observacionales
 - Sesgos por características observables y no observables
- 2 Sesgos y problema de selección
 - Selección sobre observables
- 3 Matching
 - Matching a través de características observables
 - Estimación de la varianza
 - Estimadores de matching sugeridos por Abadie and Imbens (2002)
 - Matching a través del propensity score
 - **Propensity Score Matching en Stata**
 - Estimadores de Heckman, Ichimura y Todd
- 4 EPEM
 - Emparejamiento con Características Observadas
- 5 Inversa del Propensity Score como Ponderador
 - Método de la IPW
- 6 Propensity Score en el Ajuste por Regresión

Propensity score matching usando *pscore*

- *Propensity score matching* en Stata.
- Es importante tener en cuenta que estos procedimientos permiten reducir pero no eliminar el sesgo generado por factores no observables que puedan estar confundidos en la determinación del efecto de un cierto tratamiento de interés.
- La reducción del sesgo estará relacionada con la calidad de las variables observables utilizadas para la estimación del propensity score y la calidad del matching.
- El sesgo sólo se elimina si puede asumirse que la exposición al tratamiento es aleatoria entre los individuos que tienen el mismo propensity score.
- El comando **pscore** realiza la estimación del propensity score y verifica que se cumpla el supuesto de balance ($s \perp x | \pi(x)$) utilizando el siguiente algoritmo:

Propensity score matching usando *pscore*

- **(1)** Estimar el modelo probit (logit):

$$P(s_i = 1|x_i) = \Phi(h(x_i))$$

donde $\Phi(\cdot)$ representa la función de distribución normal (logística) y $h(\cdot)$ representa una especificación inicial del modelo que incluye todas las variables como términos lineales sin interacciones o términos de mayor orden.

- **(2)** Dividir la muestra en p intervalos equiespaciados del propensity score, donde p está definido por el usuario (por default $p = 5$).
- **(3)** Dentro de cada intervalo, comprobar que el propensity score promedio entre los grupos T y C no difiere.

Propensity score matching usando *pscore*

- **(4)** Si en algún intervalo se encuentran diferencias en el propensity score entre ambos grupos, se divide ese intervalo por la mitad y se realiza nuevamente el test. Este procedimiento se repite hasta que en todos los intervalos no se encuentren diferencias en el PS entre grupos.
- **(5)** Dentro de cada intervalo, comprobar que los promedios de todas las características no difieren entre los grupos T y C . Esta es una condición necesaria pero no suficiente para la hipótesis de balance.
- **(6)** Si la media de una o más características difiere, se informa al usuario que la propiedad de balance no se cumple y que es preciso una especificación menos parsimoniosa de $h(\cdot)$.

Propensity score matching usando Stata

- Los pasos 2 a 6 se pueden restringir al soporte común de los grupos T y C .
- Es importante notar que, aún cuando los grupos se formen utilizando el propensity score, esto no es suficiente para estimar el efecto tratamiento. La probabilidad de encontrar dos elementos con exactamente el mismo valor de PS es 0.
- Cuatro de los procedimientos propuestos en la práctica son:
 - ▶ Nearest Neighbor Matching
 - ▶ Radius Matching
 - ▶ Kernel Matching
 - ▶ Stratification Matching
- Estos procedimientos permiten realizar el matching y además estimar el efecto tratamiento en los tratados.

PS matching usando nearest neighbor

- Consideremos los grupos T y C y consideremos que y_i^T es el resultado de interés observado para el individuo i del grupo T mientras que y_j^C es el resultado de interés observado para el individuo j del grupo C .
- Expresamos con $C(i)$ el conjunto comparación para el elemento $i \in T$ cuyo propensity score estimado es p_i .
- El conjunto de controles correspondientes a Nearest Neighbor Matching se define como:

$$C(i) = \min_j \| p_i - p_j \|$$

- Para cada i del grupo T , $C(i)$ es un conjunto simple a menos que haya empates.

PS matching usando radius matching

- En el caso de radius matching, el grupo de comparación para cada observación se determina como:

$$C(i) = \{p_j \mid \|p_i - p_j\| < r\}$$

- El grupo de comparación de i estará compuesto por todos los elementos del grupo control cuyo propensity score se encuentre dentro de un radio r con respecto a p_i .

Nearest neighbor y radius matching

- El efecto tratamiento se calcula como:

$$\begin{aligned}T_M &= \frac{1}{N_T} \sum_{i \in T} \left(y_i^T - \sum_{j \in C(i)} w_{ij} y_j^C \right) \\&= \frac{1}{N_T} \left(\sum_{i \in T} y_i^T - \sum_{i \in T} \sum_{j \in C(i)} w_{ij} y_j^C \right) \\&= \frac{1}{N_T} \sum_{i \in T} y_i^T - \frac{1}{N_T} \sum_{j \in C} w_j y_j^C\end{aligned}$$

- Las ponderaciones asociadas a cada elemento del grupo C se obtienen como $w_j = \sum_i w_{ij}$.

Nearest neighbor y radius matching

- Para determinar la variancia del estimador, los pesos w_j se consideran fijos y se asume que los resultados de la variable de interés son independientes a través de los individuos.

$$\begin{aligned} \text{Var}(T_M) &= \frac{1}{N_T^2} \left(\sum_{i \in T} \text{Var}(y_i^T) + \sum_{j \in C(i)} w_j^2 \text{Var}(y_j^C) \right) \\ &= \frac{1}{N_T} \text{Var}(y_i^T) + \frac{1}{N_T^2} \sum_{j \in C(i)} w_j^2 \text{Var}(y_j^C) \end{aligned}$$

- Notar que el procedimiento de matching en ambos casos es con reemplazo.
- También es posible calcular los errores estándar a través de bootstrapping.

Nearest neighbor matching: comandos *attnd* y *attnw*

- El procedimiento de nearest neighbor matching consiste en ordenar todos los individuos en función del propensity score y luego realizar una búsqueda de arriba hacia abajo y abajo hacia arriba para encontrar el (los) control(es) más próximo(s).
- Si para un individuo del grupo tratamiento los matchings hacia arriba y hacia abajo son igualmente buenos hay dos opciones:
 - 1 Comando **attnw**: Se le da el mismo peso (ponderación) a los matching hacia arriba y hacia abajo.
 - 2 Comando **attnd**: Se selecciona al azar uno de los dos matchings.

Radius matching: comando *attr*

- El procedimiento del radio consiste en emparejar a cada individuo del grupo de tratamiento con todos los controles cuyos propensity score cae dentro de una determinada vecindad del propensity score de la unidad de tratamiento.
- Denotando al número de controles emparejados con la observación $i \in T$ por N_i^C , el comando **attr** define los ponderadores como

$$\begin{aligned}w_{ij} &= \frac{1}{N_i^C} \text{ si } j \in C(i) \\ &= 0 \text{ en otro caso.}\end{aligned}$$

Kernel matching: comando *attk*

- Cuando se implementa Kernel matching todos los individuos del grupo tratamiento están matcheados con un promedio ponderado de todos los elementos del grupo control.
- Las ponderaciones que se utilizan son inversamente proporcionales a la distancia entre el propensity score de los controles y el propensity score de cada individuo tratado.
- El estimador del efecto tratamiento a partir de Kernel matching es el siguiente:

$$T_{KM} = \frac{1}{N_T} \sum_{i \in T} \left(y_i^T - \frac{\sum_{j \in C} K((p_j - p_i)/h) y_j^C}{\sum_{k \in C} K((p_k - p_i)/h)} \right)$$

- Los errores estándar se calculan utilizando bootstrapping.

Stratification matching: comando atts

- Este método está basado en la misma estratificación utilizada para la estimación del propensity score a partir de *pscore*.
- Supongamos que el subíndice q identifica los intervalos definidos a partir del propensity score, dentro de un mismo bloque se calcula:

$$T_{qS} = \frac{1}{N_T^{(q)}} \sum_{i \in I(q)} y_i^T - \frac{1}{N_C^{(q)}} \sum_{j \in I(q)} y_j^C$$

donde $I(q)$ identifica el conjunto de individuos en el intervalo q , $N_T^{(q)}$ y $N_C^{(q)}$ representan la cantidad de individuos de los grupos T y C que pertenecen al intervalo q respectivamente.

Stratification matching: comando atts

- Luego, el efecto tratamiento se calcula como:

$$T_{SM} = \sum_{q=1}^Q T_{qS} w_q$$

donde w_q representa la fracción de individuos del grupo tratamiento en el intervalo q y Q es el número total de intervalos.

- Asumiendo independencia de los resultados a través de las observaciones, la variancia del estimador anterior se obtiene como:

$$Var(T_{SM}) = \frac{1}{N_T} \left(Var(y_i^T) + \sum_{q=1}^Q \frac{N_T^{(q)}}{N_T} \frac{N_T^{(q)}}{N_C^{(q)}} Var(y_j^C) \right)$$

- Además es posible calcular los errores estándar a través de bootstrapping.

Especificación del Comando en Stata

```
pscore treatment [varlist] [weight] [if exp] [in range] , pscore(newvar) [  
    blockid(newvar) detail logit comsup level(#) numblo(#) ]  
  
attnd outcome treatment [varlist] [weight] [if exp] [in range] [ ,  
    pscore(scorevar) logit index comsup detail bootstrap reps(#) noisily  
    dots ]  
  
attnw outcome treatment [varlist] [weight] [if exp] [in range] [ ,  
    pscore(scorevar) logit index comsup detail bootstrap reps(#) noisily  
    dots ]  
  
attr outcome treatment [varlist] [weight] [if exp] [in range] [ ,  
    pscore(scorevar) logit index radius(#) comsup detail bootstrap  
    reps(#) noisily dots ]  
  
attk outcome treatment [varlist] [weight] [if exp] [in range] [ ,  
    pscore(scorevar) logit index epan bwidth(#) comsup detail bootstrap  
    reps(#) noisily dots ]  
  
atts outcome treatment [varlist] [if exp] [in range] , pscore(scorevar)  
    blockid(blockvar) [ comsup detail bootstrap reps(#) noisily dots ]
```

Propensity score matching usando *psmatch2*

```
psmatch2 depvar [indepvars] [if exp] [in range] [,  
    outcome(varlist)  
    pscore(varname) logit odds index  
    neighbor(integer) ties  
    noreplacement descending  
    caliper(real)  
    radius  
    kernel  
    llr  
    kerneltype(type) bwidth(real)  
    spline nknots(integer)  
    mahalanobis(varlist) add pcaliper(real)  
    common trim(real)  
    ate  
    ai]
```

Agenda

- 1 Problemas al Utilizar Datos Observacionales
 - Sesgos por características observables y no observables
- 2 Sesgos y problema de selección
 - Selección sobre observables
- 3 Matching
 - Matching a través de características observables
 - Estimación de la varianza
 - Estimadores de matching sugeridos por Abadie and Imbens (2002)
 - Matching a través del propensity score
 - Propensity Score Matching en Stata
 - Estimadores de Heckman, Ichimura y Todd
- 4 EPEM
 - Emparejamiento con Características Observadas
- 5 Inversa del Propensity Score como Ponderador
 - Método de la IPW
- 6 Propensity Score en el Ajuste por Regresión

Estimadores de Heckman, Ichimura y Todd

- Heckman, Ichimura y Todd (1997) “extienden” los resultados de Abadie e Imbens que vimos en el emparejamiento usando variables observables al emparejamiento por propensity score.
- “Extienden” en el contexto de nuestro curso ya que vimos primero el emparejamiento por variables observables y después por el propensity score.
- En realidad su trabajo es anterior al de Abadie e Imbens.
- Usando el método de regresión muestran que el ATT se puede estimar con

$$\tau_{ATT} = \bar{Y}_1 - \frac{1}{N_1} \sum_{i:s_i=1} \hat{\mu}_0(X_i) \quad (1)$$

donde $\bar{Y}_1 = \frac{1}{N_1} \sum_{i:s_i=1} Y_i$ y $\hat{\mu}_0(X_i)$ es el estimador basado en la función de regresión como en Abadie e Imbens.

Agenda

- 1 Problemas al Utilizar Datos Observacionales
 - Sesgos por características observables y no observables
- 2 Sesgos y problema de selección
 - Selección sobre observables
- 3 Matching
 - Matching a través de características observables
 - Estimación de la varianza
 - Estimadores de matching sugeridos por Abadie and Imbens (2002)
 - Matching a través del propensity score
 - Propensity Score Matching en Stata
 - Estimadores de Heckman, Ichimura y Todd
- 4 EPEM
 - Emparejamiento con Características Observadas
- 5 Inversa del Propensity Score como Ponderador
 - Método de la IPW
- 6 Propensity Score en el Ajuste por Regresión

EPEM: Evaluación de Impacto

- Como mencionamos, en el caso del EPEM Honduras, los grupos ya estaban armados con $n_T = 1151$ y $n_C = 792$.
- Recordemos que los grupos debieran mostrar características observables similares antes de la implementación del programa.
- La siguiente tabla muestra los contrastes de diferencias de medias de distintas variables observadas.

EPEM: Evaluación de Impacto

	Tratamiento	Control	Diferencia
Edad	22,84 (3,31)	22,25 (3,17)	0,59*** (0,16)
Género	0,42 (0,49)	0,47 (0,50)	-0,04* (0,02)
Casado	0,12 (0,33)	0,07 (0,26)	0,05*** (0,01)
Jefe de hogar	0,42 (0,49)	0,36 (0,48)	0,06** (0,02)
Nivel educativo	4,64 (0,94)	4,81 (0,93)	-0,17*** (0,05)
Hijos	0,44 (0,50)	0,33 (0,47)	0,11*** (0,02)
Salario real	24,07 (13,25)	25,09 (15,11)	1,02 (0,79)
Desocupado	0,30 (0,46)	0,35 (0,48)	-0,05** (0,02)

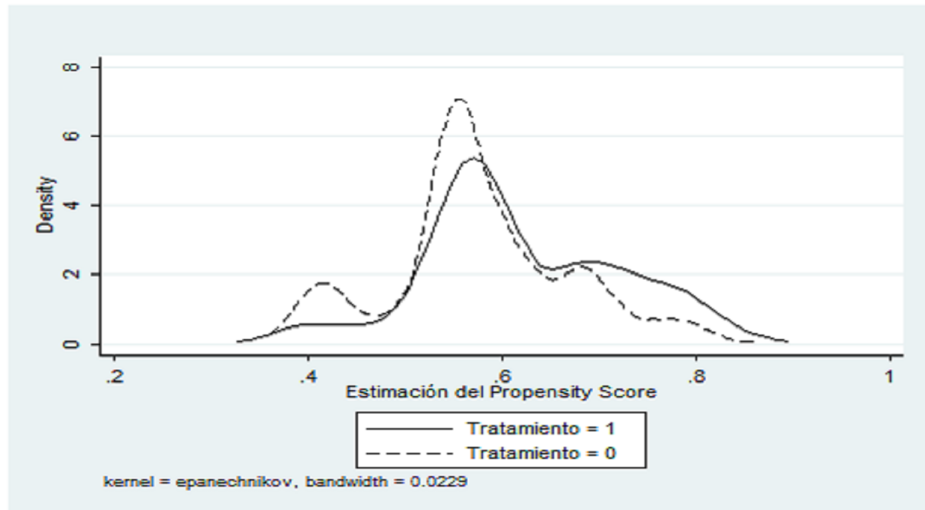
EPEM: propensity score matching

- La evidencia presentada en la tabla sugiere que los grupos de tratamiento y de control no resultan comparables en la mayoría de las características sociodemográficas y económicas analizadas.
- Esto indicaría que la estimación del efecto causal del entrenamiento del programa EPEM no daría los resultados esperados de la comparación directa de los promedios de las variables de resultado en cada uno de los grupos así considerados.
- Se decidió realizar un emparejamiento (matching) de ambos grupos.
- El emparejamiento de los grupos se realizó a partir de la estimación de un modelo logit en la línea de base.
- Como variable dependiente se utilizó una variable indicadora de si el joven pertenecía al grupo que participó del entrenamiento del EPEM (tratamiento = 1) o al grupo de control (tratamiento = 0).

EPEM: propensity score matching

- Las variables explicativas de la fueron: edad y su cuadrado, género, y variables binarias de si el joven estaba casado, tenía hijos, era jefe de hogar, tenía estudios primarios o secundarios, vivía en Tegucigalpa o en San Pedro Sula y si estaba desocupado. Además, se incluyeron interacciones entre las variables desocupado e hijos y desocupado y género.
- El siguiente gráfico muestra la estimación de las funciones de densidad de recibir el tratamiento (propensity score) de ambos grupos

EPEM: propensity score matching

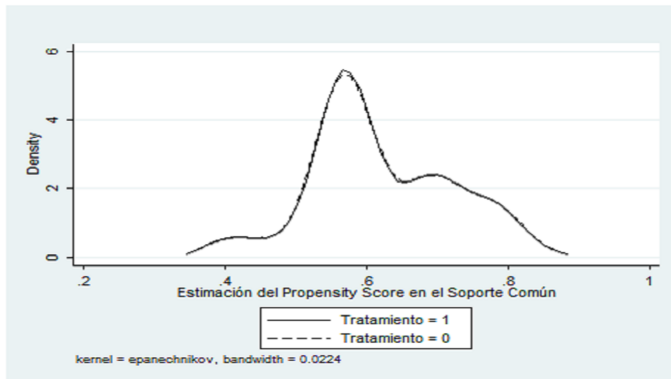


EPEM: propensity score matching

- Como se observa, a pesar de las diferencias en las densidades, existe un amplio rango de soporte común a ambas funciones.
- Se emparejaron los grupos, asignándosele a cada tratamiento el control más parecido posible con el procedimiento del radio.
- Cada individuo del grupo de tratamiento se emparejó con aquellos controles que tenían una probabilidad de recibir el tratamiento que cayera en un radio determinado del puntaje de propensión (propensity score) del individuo.
- Se utilizó un radio de 0.01, lo que significó que cada joven del grupo de tratamiento se emparejó con aquellos controles cuyos puntajes de propensión estuvieran a una distancia menor o igual a 0.01 del puntaje de propensión del joven del grupo de tratamiento.
- Se determinó el soporte común eliminando del grupo de tratamiento (control) a aquellos jóvenes con un puntaje de propensión mayor (menor) al máximo (mínimo) puntaje obtenido en el grupo de control (tratamiento).

EPEM: propensity score matching

- Una vez realizado el emparejamiento y obtenido el soporte común quedaron definidos los nuevos grupos de tratamiento y control.
- Sobre estos grupos se estimaron las densidades del propensity score que se muestran en la siguiente figura.



EPEM: propensity score matching

- La siguiente tabla muestra los contrastes de diferencias de medias.

	Tratam.	Control	Dif.
Edad	24.73 (0.151)	24.45 (0.231)	-0.286 (0.275)
Género	0.36 (0.024)	0.33 (0.038)	-0.034 (0.045)
Casado	0.26 (0.022)	0.22 (0.038)	-0.041 (0.044)
Jefe de hogar	0.72 (0.023)	0.74 (0.039)	0.018 (0.045)
Nivel educativo	4.38 (0.052)	4.29 (0.088)	-0.093 (0.103)
Hijos	1.00 (0.000)	1.00 (0.000)	0.00 (0.000)
Salario real	23.36 (0.623)	22.53 (1.178)	-0.83 (1.332)
Desocupado	0.29 (0.023)	0.32 (0.034)	0.029 (0.041)

EPEM: propensity score matching

- En Stata el comando para realizar los contrastes de diferencias de medias es `ttest`
- Dos opciones:
 - 1 Grupos Independientes (diseño experimental)
`ttest variable, by(variable binaria que define los grupos)`
 - 2 Grupos Dependientes (diseño cuasi-experimental)
`ttest variable, by(variable binaria que define los grupos) unequal welch`
- En la tabla anterior,
`ttest Salario real, by(tratamiento) unequal welch`

EPEM: Evaluación de Impacto

- Para realizar la evaluación de impacto se utilizó el *Stata*.
- Antes de la implementación hay que asegurarse que el *Stata* tenga instalados todos los comandos.
- En la línea de comandos del programa escribir `findit psmatch2`
- En la línea de comandos del programa escribir `findit pscore`
- En la línea de comandos del programa escribir `findit rbounds`
- En la línea de comandos del programa escribir `findit nnmatch`
- Copiar en el subdirectorio `Stata11\ado\base\o` el archivo `optselect.ado`

Agenda

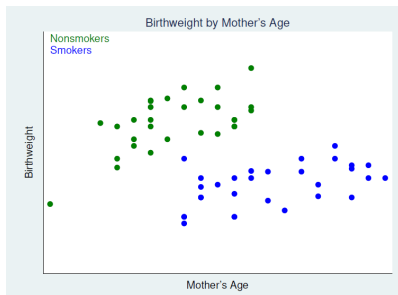
- 1 Problemas al Utilizar Datos Observacionales
 - Sesgos por características observables y no observables
- 2 Sesgos y problema de selección
 - Selección sobre observables
- 3 Matching
 - Matching a través de características observables
 - Estimación de la varianza
 - Estimadores de matching sugeridos por Abadie and Imbens (2002)
 - Matching a través del propensity score
 - Propensity Score Matching en Stata
 - Estimadores de Heckman, Ichimura y Todd
- 4 EPEM
 - Emparejamiento con Características Observadas
- 5 Inversa del Propensity Score como Ponderador
 - Método de la IPW
- 6 Propensity Score en el Ajuste por Regresión

Inversa del propensity score como ponderador

- La estimación del ATE usando la inversa del propensity score como ponderador (IPW) trata a este problema como uno de observaciones faltantes.
- A diferencia de la estimación de las esperanzas condicionales de la variable de resultado usando un modelo de regresión múltiple, IPW utiliza esperanzas condicionales ponderadas.
- Como mencionamos anteriormente, uno podría calcular el ATE si observara todos los resultados potenciales.
- Como esto no es posible, los estimadores de IPW ven a los resultados potenciales no observados como observaciones faltantes y usan ponderadores para corregir las esperanzas matemáticas condicionales por estos datos faltantes.
- Describamos el estimador con un ejemplo.

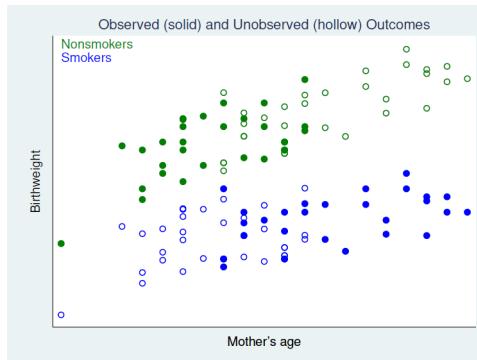
Inversa del propensity score como ponderador

- Considere un caso hipotético como el de Cattaneo (2010).
- La pregunta que se quiere responder es si fumar durante el embarazo afecta el peso de un recién nacido.
- Las unidades en este caso son mujeres embarazadas, algunas de las cuales fumaron durante el embarazo.
- La variable de resultado es el peso del bebé al nacer.
- La figura muestra el peso del bebé al nacer para madres fumadoras y no fumadoras como función de la edad de la madre.



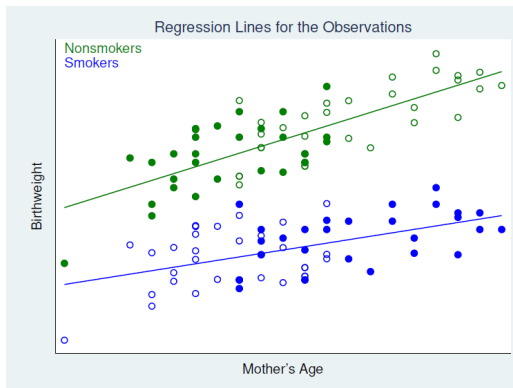
Inversa del propensity score como ponderador

- La figura sugiere que las mujeres fumadoras tienden a tener mayor edad que las no fumadoras.
- Para las mujeres fumadoras de mayor edad y las no fumadoras más jóvenes no parece haber un soporte común.
- Supongamos que observamos los resultados potenciales de cada mujer embarazada:



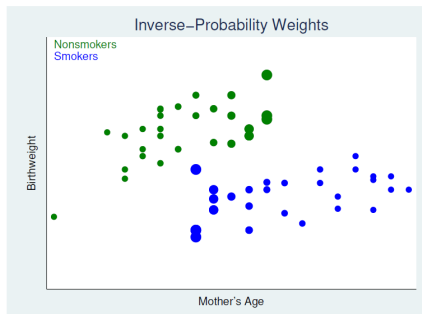
Inversa del propensity score como ponderador

- Lo que hacía el método de regresión era estimar una regresión (con los datos observados) del peso del bebé sobre la edad de la madre para el grupo de madres fumadoras y para el grupo de madres no fumadoras.
- Luego usaba la línea de regresión de las madres fumadoras como resultado contrafáctico de las madres no fumadoras y viceversa.



Inversa del propensity score como ponderador

- El IPW es un estimador que usa esperanzas ponderadas en lugar de no ponderadas como el método de regresión.
- El método del propensity score ve a los resultados potenciales no observados (círculos sin relleno) como observaciones faltantes y usa ponderadores para corregir las estimaciones de las esperanzas para las unidades tratadas y no tratadas.
- El método aplica mayor ponderación a los círculos verdes sólidos correspondientes a madres de mayor edad y menor ponderación a los correspondientes a madres más jóvenes.



Inversa del propensity score como ponderador

- En términos formales, una forma de estimar el ATE es utilizando la **inversa del propensity score como ponderador**.
- Recordando que $sy = sy_1$ tenemos

$$\begin{aligned} E \left[\frac{sy}{\pi(x)} \middle| x \right] &= E \left[\frac{sy_1}{\pi(x)} \middle| x \right] = E \left\{ E \left[\frac{sy_1}{\pi(x)} \middle| x, s \right] \middle| x \right\} \\ &= E \left\{ \frac{sE(y_1|x, s)}{\pi(x)} \middle| x \right\} = E \left\{ \frac{sE(y_1|x)}{\pi(x)} \middle| x \right\} \\ &= E \left\{ \frac{s}{\pi(x)} \middle| x \right\} E(y_1|x) = E(y_1|x) \end{aligned}$$

- Haciendo uso de $E(s|x) = \pi(x)$.
- Usando un argumento similar se puede mostrar que:

$$E \left[\frac{(1-s)y}{1-\pi(x)} \middle| x \right] = E(y_0|x)$$

Inversa del propensity score como ponderador

- Usando álgebra se puede mostrar que,

$$E(y_1 - y_0|x) = E \left[\frac{[s - \pi(x)]y}{\pi(x)[1 - \pi(x)]} \middle| x \right]$$

- Y usando expectativas iteradas,

$$\tau_{ATE} = E(y_1 - y_0) = E \left[\frac{[s - \pi(x)]y}{\pi(x)[1 - \pi(x)]} \right] \quad (2)$$

Inversa del propensity score como ponderador

- El argumento para identificar el efecto sobre los tratados es un poco diferente.
- Escribamos

$$\begin{aligned}[s - \pi(x)]y &= [s - \pi(x)][y_0 + s(y_1 - y_0)] \\ &= [s - \pi(x)]y_0 + s[s - \pi(x)](y_1 - y_0) \\ &= [s - \pi(x)]y_0 + s[1 - \pi(x)](y_1 - y_0).\end{aligned}$$

donde la última igualdad surge de $s^2 = s$.

- Entonces

$$\frac{[s - \pi(x)]y}{1 - \pi(x)} = \frac{[s - \pi(x)]y_0}{1 - \pi(x)} + s(y_1 - y_0)$$

- Note que la esperanza matemática del primer término del lado derecho de arriba es cero. Por lo tanto

$$E\left\{\frac{[s - \pi(x)]y}{1 - \pi(x)} \middle| x\right\} = E[s(y_1 - y_0)|x]$$

Inversa del propensity score como ponderador

- Aplicando la ley de expectativas iteradas

$$E \left\{ \frac{[s - \pi(x)]y}{1 - \pi(x)} \right\} = E[s(y_1 - y_0)]$$

- Usando la definición de esperanza,

$$\begin{aligned} E[s(y_1 - y_0)] &= P(s = 0)E[s(y_1 - y_0)|s = 0] \\ &+ P(s = 1)E[s(y_1 - y_0)|s = 1] \\ &= 0 + P(s = 1)E[s(y_1 - y_0)|s = 1] \\ &= \rho E[y_1 - y_0|s = 1] = \rho \tau_{ATT} \end{aligned}$$

donde $\rho \equiv P(s = 1)$ es la probabilidad marginal de recibir el tratamiento.

- Juntando los resultados se tiene,

$$\tau_{ATT} = E \left\{ \frac{[s - \pi(x)]y}{\rho [1 - \pi(x)]} \right\} \quad (3)$$

Inversa del propensity score como ponderador

- Estimaciones de (2) y (3)

$$\hat{\tau}_{ATE} = \frac{1}{N} \sum_{i=1}^N \left[\frac{[s_i - \hat{\pi}(x_i)]y_i}{\hat{\pi}(x_i)[1 - \hat{\pi}(x_i)]} \right] \quad (4)$$

$$\hat{\tau}_{ATT} = \frac{1}{N} \sum_{i=1}^N \left\{ \frac{[s_i - \hat{\pi}(x_i)]y_i}{\hat{\rho}[1 - \hat{\pi}(x_i)]} \right\} \quad (5)$$

donde $\hat{\rho} = N_1/N$ es la fracción de unidades en el grupo de tratamiento.

- Definamos con \mathbf{d}_i a la derivada primera del logaritmo de la función de verosimilitud de la estimación del propensity score y llamemos

$$k_i = \frac{[s_i - \hat{\pi}(x_i)]y_i}{\hat{\pi}(x_i)[1 - \hat{\pi}(x_i)]}$$

Inversa del propensity score como ponderador

- La varianza asintótica de $\hat{\tau}_{ATE}$ es $Var(e_i)$, con e_i los residuos de una regresión de k_i sobre una constante y \mathbf{d}_i .
- El error estándar asintótico es:

$$\hat{\sigma}_{\hat{\tau}_{ATE}} = \left[N^{-1} \sum_{i=1}^N e_i^2 \right]^{0.5} / \sqrt{N}$$

- Similarmente,

$$\hat{\sigma}_{\hat{\tau}_{ATT}} = \hat{\rho}^{-1} \left[N^{-1} \sum_{i=1}^N (r_i - \hat{\tau}_{ATT} s_i)^2 \right]^{0.5} / \sqrt{N}$$

donde r_i son los residuos de una regresión de $\hat{q}_i = \frac{[s_i - \hat{\pi}(x_i)]y_i}{[1 - \hat{\pi}(x_i)]}$ sobre una constante y \mathbf{d}_i (Wooldridge, 2002).

Inversa del propensity score como ponderador

- Note que el estimador del ATE (ATT) no es más que un **M-Estimator** de dos pasos.
- Siguiendo nuestra notación de M-Estimation, en este caso

$$\min_{\theta \in \Theta} \frac{s_i}{\hat{\pi}(x, \hat{\gamma})} \times [y_i - m_1(\mathbf{w}_i, \theta)]^2$$

donde $\hat{\gamma}$ son los estimadores del propensity score obtenidos en un primer paso.

- Tenemos una expresión similar para los no tratados.
- La estimación de la varianza del ATE (ATT) sigue los pasos detallados en la lectura de M-Estimation para los estimadores de dos pasos.

Propensity score en regresión

- Un uso diferente del propensity score es en el ajuste por regresión.
- El estimador más sencillo consiste en estimar por MCC:

$$y_i \quad \text{on} \quad 1, s_i, \hat{\pi}(\mathbf{x}_i), \quad i = 1, \dots, N$$

- El coeficiente estimado sobre la variable s_i , $\hat{\pi}_{ate,psreg}$, es el ATE estimado.
- La idea detrás de esta estimación es que el propensity score estimado debiera ser suficiente como para controlar por la correlación entre el tratamiento s_i y las variables en \mathbf{x}_i .
- En la práctica, los errores estándar de los coeficientes de esta estimación se obtienen via bootstrapping.
- La regresión en el propensity score parece atractiva porque, en comparación con el ajuste de regresión usando \mathbf{x} , reduce el problema de controlar por un posible gran número de variables para controlar por una sola $\hat{\pi}(\mathbf{x})$.

Combinación del ajuste por regresión y el IPW

- Hemos analizado la estimación del ATE basados en dos estrategias: la primera estimando $E(y_g | \mathbf{x})$ for $g = 0, 1$ usando un modelo de regresión lineal y la segunda ponderando por la inversa del propensity score como en (4).
- Se pueden combinar estos dos enfoques para obtener un estimador del ATE que es doblemente robusto porque solo requiere que uno de los dos modelos, el de la esperanza condicional o el del propensity score esté bien especificado (Robins y Rotnitzky, 1995; Robins, Rotnitzky y Zhao, 1995).
- Para describir esta estimación definamos con $m_0(\cdot, \delta_0)$ y $m_1(\cdot, \delta_1)$ a las funciones paramétricas de $E(y_g | \mathbf{x})$ for $g = 0, 1$ y con $p(\cdot, \gamma)$ al modelo paramétrico del propensity score.

Combinación del ajuste por regresión y el IPW

- En un primer paso se estima el propensity score $p(x_i, \hat{\gamma})$ usando un modelo Probit o Logit.
- En el segundo paso, para estimar $\delta_1 = (\alpha_1, \beta_1')'$, usamos un modelo de regresión ponderando por la inversa del propensity score estimado en el primer paso:

$$\min_{\alpha_1, \beta_1} \sum_{i=1}^N s_i (y_i - \alpha_1 - \mathbf{x}_i \beta_1)^2 / p(\mathbf{x}_i, \hat{\gamma}) \quad (6)$$

- Para estimar $\delta_0 = (\alpha_0, \beta_0')'$ ponderamos por $1/[1 - \hat{p}(\mathbf{x}_i)]$ y usamos la muestra con $s_i = 0$.
- La estimación del ATE es;

$$\hat{\tau}_{\text{ate, pswreg}} = N^{-1} \sum_{i=1}^N \left[\left(\hat{\alpha}_1 + \mathbf{x}_i \hat{\beta}_1 \right) - \left(\hat{\alpha}_0 + \mathbf{x}_i \hat{\beta}_0 \right) \right]. \quad (7)$$

Combinación del ajuste por regresión y el IPW

- La estimación de la varianza de este estimador sigue los pasos de la M-Estimation de dos pasos.
- Una alternativa es hacer bootstrapping el procedimiento de dos pasos para obtener $\hat{\tau}_{ate, pswreg}$ y obtener errores estándar que son válidos para hacer inferencia estadística.

Estimadores por Emparejamiento: Resumen

- La característica fundamental es que cualquiera que sea el método que haya utilizado, la suposición es que después de haber emparejado las unidades tratadas y de control usando variables observables, cualquier diferencia en los resultados refleja el efecto causal del tratamiento, y no la influencia de alguna variable no observada.
- El problema es que en la mayoría de las políticas públicas a evaluar hay cierta autoselección en el tratamiento y esta autoselección casi siempre está relacionada con características no observables.
- El enfoque estándar para pensar si el emparejamiento brindará estimaciones causales confiables es a través de pruebas de plausibilidad estadística.

Estimadores por Emparejamiento: Resumen

- Pruebas *placebo* donde se estima el impacto del programa en dos momentos **antes** de la implementación del mismo; evaluar la sensibilidad de las estimaciones usando distintos métodos de emparejamiento (o diferentes variables para emparejar); y estimar el efecto de un pseudo-tratamiento. Por ejemplo, si tiene dos potenciales grupos de control (solicitantes no seleccionados para el programa y no solicitantes) estimar el impacto en estos dos grupos y ver si hay una diferencia en los resultados.
- Note que estos son controles útiles para ayudar a argumentar la plausibilidad desde el punto de vista estadístico,
- Sin embargo, estas pruebas no responden la pregunta de porque se cumple el supuesto de independencia condicional en media.
- Esto es, no responden “¿Por qué, si las unidades son tan similares, una fue tratada y otra no?”.

Estimadores por Emparejamiento: Resumen

- En la práctica estas pruebas estadísticas se suelen acompañar de una argumentación retórica de porque el emparejamiento es razonable.
- Describir, ¿cuándo parece más plausible que, condicionando en observables, la razón por la que una unidad recibe tratamiento y otra no, es aleatoria?
- O al menos, tratar de justificar que no hay no observables que también están correlacionados con la variable de resultado.

Bibliografía

- Matching:

- Abadie, A., Drukker, D., Herr, J. L. and Imbens, G. (2001): *Implementing Matching Estimators for Average Treatment Effects in Stata*, The Stata Journal, 1 (1), 1-18.
- Becker, S. and Ichino, A. (2002): *Estimation of Average Treatment Effects Based on Propensity Scores*, The Stata Journal , 2 (4), 358-377.
- Gonzalez-Rozada, M. (2012): *Evaluación de Impacto del Programa Entrenamiento para el Empleo en Honduras*, Nota Técnica.
- Robins, J. M., and A. Rotnitzky (1995): "Semiparametric Efficiency in Multivariate Regression Models," Journal of the American Statistical Association 90, 122 – 129.
- Robins, J. A., A. Rotnitzky, and L. Zhao (1995): "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data," Journal of the American Statistical Association 90, 106 – 121.
- Wooldridge, J. (2002): *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge, Massachusetts.