

Matching in Social Science

Du CHEN

Antai College of Economics and Management
Shanghai Jiao Tong University

chendu2017@sjtu.edu.cn

September 23, 2019

Overview

1 Selection Bias

- Selection Bias
- Random Assignment Solves the Selection Problem

2 Matching

- Key Assumption
- Estimating Causal Effects by Stratification
- Propensity Score Matching

Hospital

Do hospitals make people healthier?

- If-then

Hospital

Do hospitals make people healthier?

- Compare the health status of those who have been to the hospital to the health of those who have not.
- The National Health Interview Survey (NHIS) contains the information needed to make this comparison.

Group	Sample Size	Mean health status	Std. Error
Hospital	7774	3.21	0.014
No Hospital	90049	3.93	0.003

Hospital

Do hospitals make people healthier?

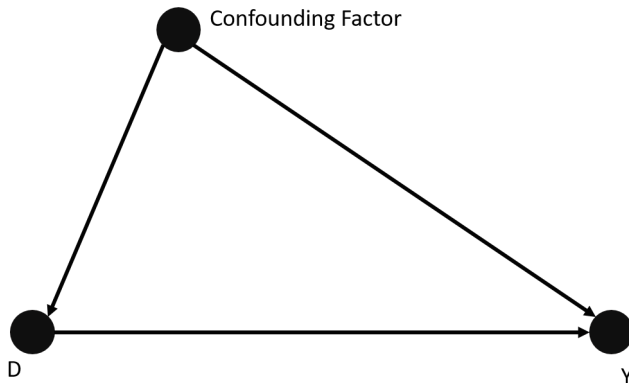
- The naive estimation is biased.
- How do you design an experiment to support your opinion?

University

Do people go to university gain a higher salary?

- Confounding factor?

University



Selection Bias

- Think about hospital treatment as described by a binary random variable: $D_i = \{0, 1\}$.
- The outcome of interest, a measure of health status, is denoted by Y_i .
- The question is whether Y_i is affected by hospital care.
- **Counterfactual:** To address this question, we assume we can imagine what might have happened to someone who went to the hospital if they had not gone and vice versa.

Selection Bias

$$\text{potential outcome} = \begin{cases} Y_{1i}, & \text{if } D_i = 1 \\ Y_{0i}, & \text{if } D_i = 0 \end{cases}$$

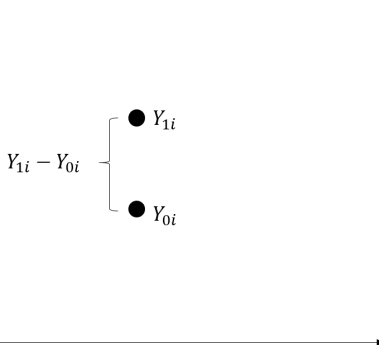
- The observed outcome, Y_i , can be written in terms of potential outcomes as

$$\begin{aligned} Y_i &= \begin{cases} Y_{1i}, & \text{if } D_i = 1 \\ Y_{0i}, & \text{if } D_i = 0 \end{cases} \\ &= Y_{0i} + (Y_{1i} - Y_{0i})D_i \end{aligned}$$

Selection Bias

Potential outcome:

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$



$Y_{1i} - Y_{0i}$ is the causal effect of hospitalization for an individual.

Selection Bias

But:

- The treatment effect can be different for different people.
- We never see both potential outcomes for any one person.

Selection Bias

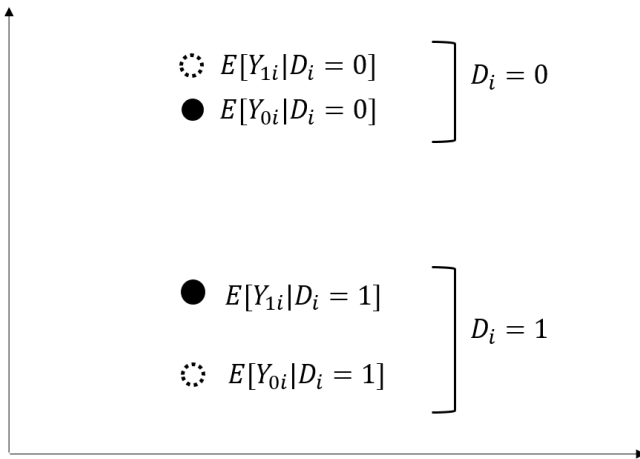
The comparison of average health conditional on hospitalization status is formally linked to the average causal effect by the equation below:

$$\underbrace{E[Y_i|D_i = 1] - E[Y_i|D_i = 0]}_{\text{Observed difference in average health}} = \underbrace{E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]}_{\text{average treatment effect on the treated}} + \underbrace{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]}_{\text{selection bias}}$$

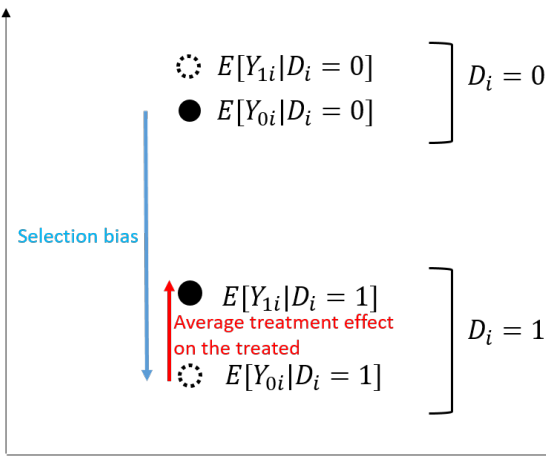
where,

- **Average Treatment effect on the Treated (ATT):**
 $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] = E[Y_{1i} - Y_{0i}|D_i = 1]$
- **Average Treatment effect on the Control (ATC):**
 $E[Y_{1i}|D_i = 0] - E[Y_{0i}|D_i = 0] = E[Y_{1i} - Y_{0i}|D_i = 0]$

Selection Bias



Selection Bias



ATE, ATT, ATC

Selection Bias

- Because the sick are more likely than the healthy to seek treatment, those who were hospitalized have worse y_{0i} 's, making selection bias negative in this example.
- The selection bias may be so large (in absolute value) that it completely masks a positive treatment effect.
- The goal of most empirical economic research is to overcome selection bias, and therefore to say something about the causal effect of a variable like D_i .

Random Assignment

Random assignment of D_i solves the selection problem because random assignment makes D_i independent of potential outcomes, i.e.:

$$E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$$

Therefore:

$$\begin{aligned} E[Y_i|D_i = 1] - E[Y_i|D_i = 0] &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\ &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] \end{aligned}$$

Random Assignment

In fact, given random assignment, this simplifies further to:

$$\begin{aligned} E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] &= E[Y_{1i} - Y_{0i}|D_i = 1] \\ &= E[Y_{1i} - Y_{0i}]. \end{aligned}$$

- That is: the effect of randomly-assigned hospitalization on the hospitalized is the same as the effect of hospitalization on a randomly chosen patient.
- The main thing is that random assignment of D_i eliminates selection bias.
- Random assignment does *NOT* mean that samples are equally divided.

Random Assignment

Try to select observational data such that the treatment is almostly randomly assigned.

Complete Information Assumption

- There are no systematic differences between the treated and the control in terms of unobserved characteristics that may influence outcomes.
- **All** the variables that affect simultaneously D and Y are observed.

Perfect Stratification

Suppose that those who take the treatment and those who do not are very much unlike each other, and yet the ways in which they differ are captured exhaustively by a set of observed treatment assignment/selection variables S .

Stratification

That is to say, if conditioned on S , we would be able to assert:
No selection bias:

$$E[Y_{0i}|D_i = 1, S_i = s] = E[Y_{0i}|D_i = 0, S_i = s]$$

We further assume homogeneous response holds:

$$E[Y_{1i}|D_i = 1, S_i = s] = E[Y_{1i}|D_i = 0, S_i = s]$$

Stratification

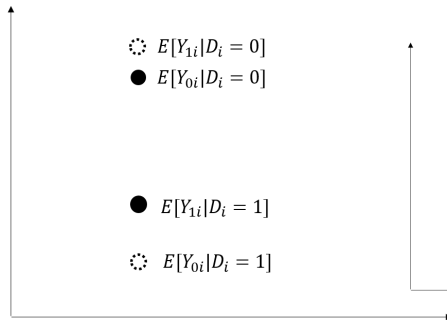


Figure: No stratification

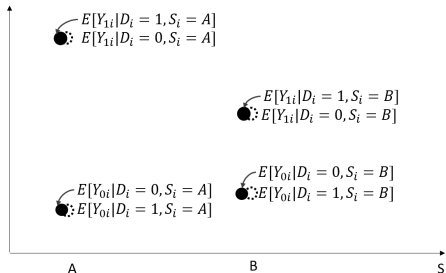


Figure: stratified

Stratification

Then:

$$\begin{aligned} E[Y_{1i} - Y_{0i}] &= \sum_s \text{Prob}\{S = s\} E[Y_{1i} - Y_{0i} | S = s] \\ &= \sum_s \text{Prob}\{S = s\} (E[Y_{1i} | D_i = 1, S = s] - E[Y_{0i} | D_i = 0, S = s]) \end{aligned}$$

Example

Table 5.1 The Joint Probability Distribution and Conditional Population Expectations for Matching Demonstration 1

Joint probability distribution of S and D			
	$D = 0$	$D = 1$	
$S = 1$	$\Pr[S = 1, D = 0] = .36$	$\Pr[S = 1, D = 1] = .08$	$\Pr[S = 1] = .44$
$S = 2$	$\Pr[S = 2, D = 0] = .12$	$\Pr[S = 2, D = 1] = .12$	$\Pr[S = 2] = .24$
$S = 3$	$\Pr[S = 3, D = 0] = .12$	$\Pr[S = 3, D = 1] = .2$	$\Pr[S = 3] = .32$
	$\Pr[D = 0] = .6$	$\Pr[D = 1] = .4$	
Potential outcomes			
	Under the control state	Under the treatment state	
$S = 1$	$E[Y^0 S = 1] = 2$	$E[Y^1 S = 1] = 4$	$E[Y^1 - Y^0 S = 1] = 2$
$S = 2$	$E[Y^0 S = 2] = 6$	$E[Y^1 S = 2] = 8$	$E[Y^1 - Y^0 S = 2] = 2$
$S = 3$	$E[Y^0 S = 3] = 10$	$E[Y^1 S = 3] = 14$	$E[Y^1 - Y^0 S = 3] = 4$
	$E[Y^0 D = 0]$ $= \frac{.36}{.6}(2) + \frac{.12}{.6}(6)$ $+ \frac{.12}{.6}(10)$ $= 4.4$	$E[Y^1 D = 1]$ $= \frac{.08}{.4}(4) + \frac{.12}{.4}(8)$ $+ \frac{.2}{.4}(14)$ $= 10.2$	

$$ATT = 2 * 0.2 + 2 * 0.3 + 4 * 0.5 = 3$$

$$ATC = 2 * 0.6 + 2 * 0.2 + 4 * 0.2 = 2.4$$

$$ATE = 2 * 0.44 + 2 * 0.24 + 4 * 0.32 = 2.64 = 3 * 0.4 + 2.4 * 0.6; \text{ selec bias?}$$

What if

What if there is a stratum of the population in which no member of the stratum ever receives the treatment, i.e. common support assumption is violated.

- In this case, the ATE is ill-defined.
- And the analyst will only be able to generate a consistent and unbiased estimate of the ATT.

Example

Table 5.3 The Joint Probability Distribution and Conditional Population Expectations for Matching Demonstration 2

Joint probability distribution of S and D			
	$D = 0$	$D = 1$	
$S = 1$	$\Pr[S = 1, D = 0] = .4$	$\Pr[S = 1, D = 1] = 0$	$\Pr[S = 1] = .4$
$S = 2$	$\Pr[S = 2, D = 0] = .1$	$\Pr[S = 2, D = 1] = .13$	$\Pr[S = 2] = .23$
$S = 3$	$\Pr[S = 3, D = 0] = .1$	$\Pr[S = 3, D = 1] = .27$	$\Pr[S = 3] = .37$
	$\Pr[D = 0] = .6$	$\Pr[D = 1] = .4$	
Potential outcomes			
	Under the control state	Under the treatment state	
$S = 1$	$E[Y^0 S = 1] = 2$		
$S = 2$	$E[Y^0 S = 2] = 6$	$E[Y^1 S = 2] = 8$	$E[Y^1 - Y^0 S = 2] = 2$
$S = 3$	$E[Y^0 S = 3] = 10$	$E[Y^1 S = 3] = 14$	$E[Y^1 - Y^0 S = 3] = 4$
	$E[Y^0 D = 0]$ $= \frac{.4}{.6}(2) + \frac{.1}{.6}(6)$ $+ \frac{.1}{.6}(10)$ $= 4$	$E[Y^1 D = 1]$ $= \frac{.13}{.4}(8) + \frac{.27}{.4}(14)$ $= 12.05$	

$$ATT = 2 * 0.325 + 4 * 0.675 = 3.35$$

no consistent and unbiased estimates of the ATC or the ATE are available

Propensity Score Matching

Suppose all the variables that affect simultaneously D and Y are observed as variables X_i , and CIA holds.

- PROPENSITY SCORE: Probability of participating in the intervention, conditional on the characteristics X_i .
- Find participants and non-participants with equal/similar propensity score.

Problem:

- As the number of characteristics determining selection increases, it is more and more difficult to find comparable individuals (curse of dimensionality).

Propensity Score Matching step by step

- 1 Estimation of the propensity score
- 2 Check the assumptions: common support
- 3 Match participants with non-participants
- 4 Check the assumptions: covariates' balance
- 5 Compute the average treatment effect

Internships-Job Abroad Example

- The intervention offers unemployed people the opportunity to do an internship abroad or to work with a short-term contract in another EU country.
- Intervention takes place between t_1 and t_2 , where t is time.

Internships-Job Abroad Example

What we observe:

- 1 $Y_i = 1$, if employed at t_3 ; $= 0$, o.w.
- 2 $D_i = 1$, if treated, $= 0$, o.w.
- 3 X_i , individual characteristics/controls (e.g age, gender, education)

Internships-Job Abroad Example

Assumption:

- ① Age is observed.
- ② Age of the individual is the **only** factor determining selection into the intervention.

Is there self-selection?

Internships-Job Abroad Example



Internships-Job Abroad Example

	D=1	D=0	Diff
Mean	34.4	43.4	9.04***
St.d	(9.05)	(10.4)	
N	128	272	

- Participation to intervention is *NOT* random.
- Young people are more likely to participate maybe because they are less affected by family ties.

Step 1: Estimation of the Propensity Score

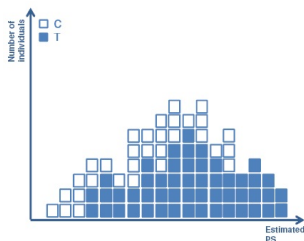
- 1 Use a binary model to estimate γ_0 and γ_1

$$P(D_i = 1|X_i) = G(\gamma_0 + \gamma_1 X_i)$$

- 2 $G(\cdot)$ can be a logit or probit link
- 3 Compute the predicted values

$$\hat{P}(D_i = 1|X_i) = G(\hat{\gamma}_0 + \hat{\gamma}_1 X_i) = \hat{P}\hat{S}_i$$

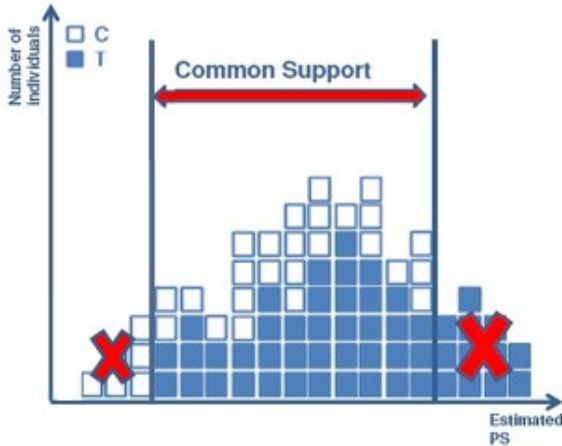
Step 2: Check common support



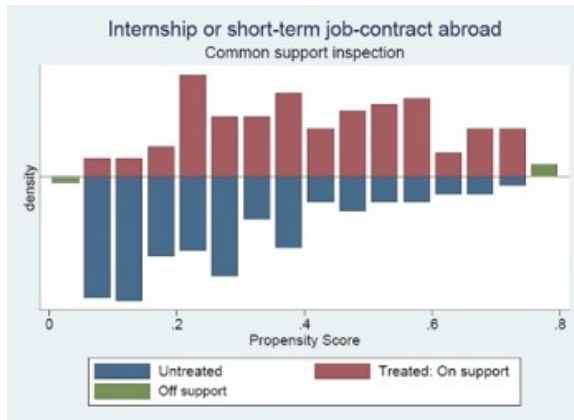
Compare only similar individuals, that is with similar propensity score.

Drop treated units that have no units with similar PS in the control group.

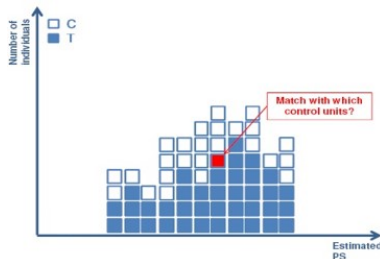
Step 2: Check common support



Step 2: Check common support



Step 3: Matching



Which control units to choose?

- Number of control units.
- Weight attributed to each control.
- With or without replacement.

Step 3: Matching

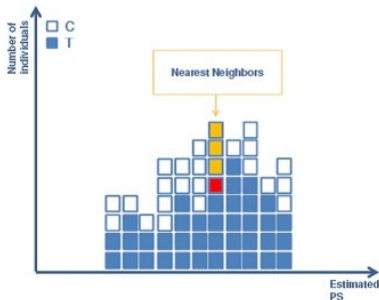


Figure: Nearest Neighbors Matching

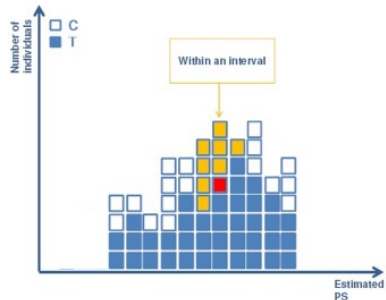


Figure: Radius Matching

Step 3: Matching

- Is there a matching algorithm better than the others? Not clear.
- Try several algorithms and options for robustness.

Step 4: Check X's balance

- Assess this property of PS, for instance, by checking if the distribution of all covariates is the same for participants and matched non-participants.
- If covariate balance is not satisfactory it may indicate a fundamental lack of comparability between the two groups.

Step 4: Check X's balance

	D=1	D=0	Diff
Before Matching	34.4	43.4	9.04***
After Matching	34.69	34.7	0

- No difference in age between treated and matched control.
- The PS successfully balances the covariates.

Step 5: the Average Treatment Effect

Treatment effect for treated i:

$$\hat{T}_i = Y_i D_i - \sum_{j \in C_i} w_{i,j} Y_j (1 - D_j)$$

Results:

	Options	Effect	s.e.
Nearest Neighbors	2 controls	0.167**	(0.078)
	4 controls	0.218***	(0.059)
	2 + max	0.174**	(0.084)
Radius	0.02 radius	0.243***	(0.067)
	0.04 radius	0.247***	(0.054)

Pros and Cons

Pros:

- More focus on the selection process and on the underlying assumptions.
- Imposition of the common support ensures comparability.
- **Versatility:**
 - Allows to estimate **heterogeneous** effects (by sub-group)
 - Allows to put more emphasis on specific variables, on which exact matching can be done (e.g. region, gender)

Pros and Cons

Cons:

- “Data-hungry” method.
- Requires strong robustness and sensitivity analysis
- **CIA** is a Strong assumption:
 - Impossible to verify, so bias stemming from unobservables can never be ruled out.
 - Matching is only as good as the characteristics used for matching.

PSM versus Regression

PSM:

$$Y_i = \alpha + \gamma T_i + \beta D_i + v_i \quad (i \in \text{chosen set})$$

Regression:

$$Y_i = \alpha + \gamma T_i + \beta f(X_i) + u_i \quad (i \in \text{sample set})$$