Lecture 07: Selection on observables

PPHA 34600

Prof. Fiona Burlig

Harris School of Public Policy
University of Chicago

From last time: be cautious with non-experimental findings

We looked at a LaLonde-style evaluation of non-RCT methods:

- These approaches did not nail the experimental result
- The more opportunity for selection, the worse they did

Why did the non-experimental estimators fail?

The root of all ☠ in this class is selection:

1. Selection on observables
   - Are treated and untreated units different in ways we can observe?
2. Selection on unobservables
   - Are treated and untreated units different in ways we can't observe?

Why did the non-experimental estimators fail?

The root of all ☠ in this class is selection:

1. Selection on observables
   - Are treated and untreated units different in ways we can observe?
2. Selection on unobservables
   - Are treated and untreated units different in ways we can't observe?

Suppose we're no longer in a randomized world

We still want to estimate treatment effects

- Our original instinct was the naive estimator: $\bar{Y}(1) - \bar{Y}(0)$

$\rightarrow$ Assumes *all* differences between $D_i = 1$ and $D_i = 0$ are "as good as random"

- We can weaken this assumption if we see other characteristics
- We will turn to a series of designs where we "control for stuff"

We are entering the world of **selection on observables** designs:
We will assume that, conditional on observables, treatment assignment is independent of potential outcomes (☠?)

Food for thought with selection on observables

Selection on observables is a form of last-resort design:

- This section should feel extremely unsatisfying
- That is on purpose!
- These designs are typically not (very) believable

Central assumption underlying SOO designs

$$(Y_i(1), Y_i(0)) \perp D_i | X_i$$

In words:

- Potential outcomes are independent of D_i, conditional on covariates

Central assumption underlying SOO designs

$$(Y_i(1), Y_i(0)) \perp D_i | X_i$$

In words:

- Potential outcomes are independent of D_i, conditional on covariates

In other words:

- "Conditional unconfoundedness"

In different words:

- "Conditional independence"

In other different words:

- "Strongly ignorable treatment assignment"

In other different words:

- Once we control for X_i, treatment is as good as random

In the last set of words:

- Once we control for X_i, we've eliminated selection

We actually need a second assumption too

$$0 < Pr(D_i = 1 | X_i = x) < 1$$

In words:

- The probability that $D_i = 1$ for all levels of X_i is between zero and one

We actually need a second assumption too

$$0 < Pr(D_i = 1|X_i = x) < 1$$

In words:

- The probability that $D_i = 1$ for all levels of X_i is between zero and one

In other words:

- "Commmon support"

In different words:

- There are both treated and untreated units for each level of X

In other different words:

- "Overlap"

What do these assumptions buy us?

Recall that we're trying to estimate the ATE:

$$\tau^{ATE} = E[Y_i(1)] - E[Y_i(0)]$$

...but all we can actually see is $E[Y_i(1)|D_i = 1]$ and $E[Y_i(0)|D_i = 0]$

What do these assumptions buy us?

Recall that we're trying to estimate the ATE:

$$\tau^{ATE} = E[Y_i(1)] - E[Y_i(0)]$$

...but all we can actually see is $E[Y_i(1)|D_i = 1]$ and $E[Y_i(0)|D_i = 0]$

Under random assignment, we had that

$$(Y_i(1), Y_i(0)) \perp D_i$$

This implies that:

$$E[Y_i(1)|D_i = 1] = E[Y_i(1)|D_i = 0]$$

and

$$E[Y_i(0)|D_i = 1] = E[Y_i(0)|D_i = 0]$$

so we could just estimate

$$\tau^{ATE} = \bar{Y}(1) - \bar{Y}(0)$$

What do these assumptions buy us?

Our new assumption says something a bit weaker:

$$(Y_i(1), Y_i(0)) \perp D_i | X_i$$

What do these assumptions buy us?

Our new assumption says something a bit weaker:

$$(Y_i(1), Y_i(0)) \perp D_i | X_i$$

$$E[Y_i(1)|D_i = 1, X_i = x] = E[Y_i(1)|D_i = 0, X_i = x]$$
$$= E[Y_i(1)|D_i, X_i = x] = E[Y_i(1)|X_i = x]$$

and

$$E[Y_i(0)|D_i = 1, X_i = x] = E[Y_i(0)|D_i = 0, X_i = x]$$
$$= E[Y_i(0)|D_i, X_i = x] = E[Y_i(0)|X_i = x]$$

What do these assumptions buy us?

Our new assumption says something a bit weaker:

$$(Y_i(1), Y_i(0)) \perp D_i | X_i$$

$$E[Y_i(1)|D_i = 1, X_i = x] = E[Y_i(1)|D_i = 0, X_i = x]$$
$$= E[Y_i(1)|D_i, X_i = x] = E[Y_i(1)|X_i = x]$$

and

$$E[Y_i(0)|D_i = 1, X_i = x] = E[Y_i(0)|D_i = 0, X_i = x]$$
$$= E[Y_i(0)|D_i, X_i = x] = E[Y_i(0)|X_i = x]$$

So we can write:

$$\tau^{SOO} = E[Y_i(1)|X_i = x] - E[Y_i(0)|X_i = x]$$

Estimating τ^{ATE} under SOO

$$\tau^{SOO} = E[Y_i(1)|X_i = x] - E[Y_i(0)|X_i = x]$$

Now integrate across all values of X_i (take a weighted average):

$$\int (E[Y_i(1)|X_i = x] - E[Y_i(0)|X_i = x])dP(X)$$

Estimating τ^{ATE} under SOO

$$\tau^{SOO} = E[Y_i(1)|X_i = x] - E[Y_i(0)|X_i = x]$$

Now integrate across all values of X_i (take a weighted average):

$$\int (E[Y_i(1)|X_i = x] - E[Y_i(0)|X_i = x])dP(X)$$

$$= \underbrace{E[Y_i(1)] - E[Y_i(0)]}_{\text{by calculus}}$$

Estimating τ^{ATE} under SOO

$$\tau^{SOO} = E[Y_i(1)|X_i = x] - E[Y_i(0)|X_i = x]$$

Now integrate across all values of X_i (take a weighted average):

$$\int (E[Y_i(1)|X_i = x] - E[Y_i(0)|X_i = x])dP(X)$$

$$= \underbrace{E[Y_i(1)] - E[Y_i(0)]}_{\text{by calculus ☠}}$$

$$= \underbrace{\tau^{ATE}}_{\text{by definition}}$$

Estimating τ^{ATE} under SOO

$$\tau^{SOO} = E[Y_i(1)|X_i = x] - E[Y_i(0)|X_i = x]$$

Now integrate across all values of X_i (take a weighted average):

$$\int (E[Y_i(1)|X_i = x] - E[Y_i(0)|X_i = x])dP(X)$$

$$= \underbrace{E[Y_i(1)] - E[Y_i(0)]}_{\text{by calculus ☠}}$$

$$= \underbrace{\tau^{ATE}}_{\text{by definition}}$$

Under conditional independence and common support, we can get from τ^{SOO} to τ^{ATE}!

How do we actually estimate τ^{SOO}?

There are two main SOO designs:

1. Regression adjustment
   - Controlling for stuff
2. Matching
   - Pairing treated and untreated on observables

These are all fancy ways to estimate

$$\tau^{ATE} = \int E[Y_i(1)|X_i = x] - E[Y_i(0)|X_i = x] dP(X)$$

Approach 1: Regression adjustment

Our regression model back in potential outcomes land:

$$Y_i(0) = \alpha + \gamma X_i + \nu_i$$

$$Y_i(1) = Y_i(0) + \tau + \tau_i$$

Approach 1: Regression adjustment

Our regression model back in potential outcomes land:

$$Y_i(0) = \alpha + \gamma X_i + \nu_i$$

$$Y_i(1) = Y_i(0) + \tau + \tau_i$$

Under constant treatment effects:

$$Y_i(1) = Y_i(0) + \tau + \underbrace{0}_{\text{no } i \text{ specific bit}}$$

Approach 1: Regression adjustment

Our regression model back in potential outcomes land:

$$Y_i(0) = \alpha + \gamma X_i + \nu_i$$

$$Y_i(1) = Y_i(0) + \tau + \tau_i$$

Under constant treatment effects:

$$Y_i(1) = Y_i(0) + \tau + \underbrace{0}_{\text{no } i \text{ specific bit}}$$

We can just write this as:

$$Y_i = \alpha + \tau D_i + \gamma X_i + \nu_i$$

Note that we're used to just working with

$$\varepsilon_i = \gamma X_i + \nu_i$$

In randomized land, everything works nicely

Under random assignment, we have that:

- $Y_i \perp D_i$
- AKA $E[\varepsilon_i | D_i] = 0$
- AKA $E[(\gamma X_i + \nu_i) | D_i] = 0$

This lets us estimate:

$$Y_i = \alpha + \tau D_i + \varepsilon_i$$

and have $\hat{\tau} \approx \tau^{ATE}$

$\rightarrow$ Note that by randomization, we don't have to worry about the X_is!

Regression with selection on observables

Under selection on observables, we have that:

- $Y_i \perp D_i | X_i$
- AKA $E[\varepsilon_i | D_i, X_i] = 0$
- AKA $E[(\gamma X_i + \nu_i) | D_i, X_i] = 0$

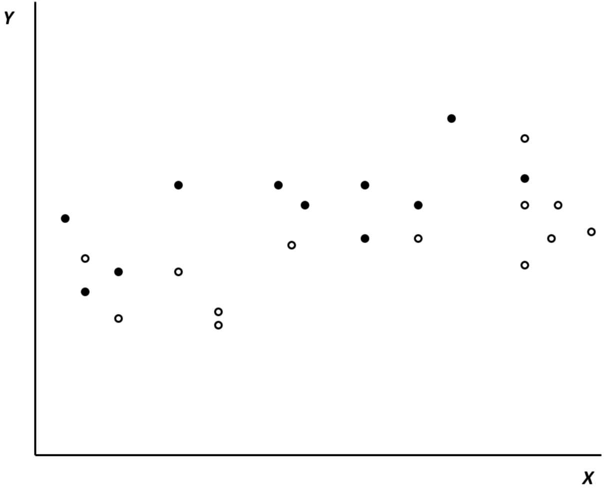
Now we have to estimate:

$$Y_i = \alpha + \tau D_i + \gamma X_i + \nu_i$$

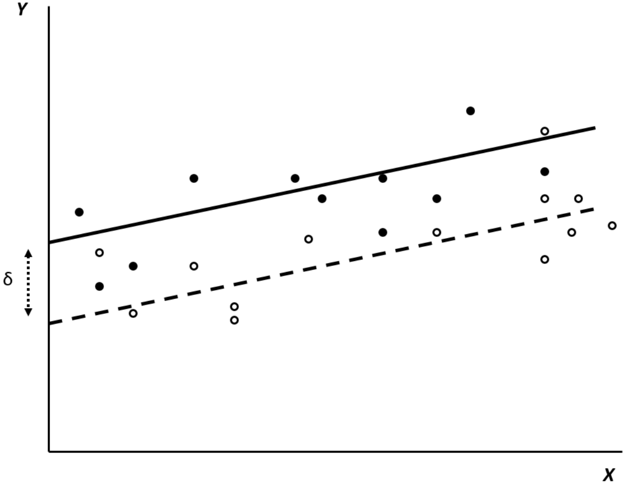
to get $\hat{\tau} \approx \tau^{ATE}$

→ Now we have *conditional* independence: if we leave X_i out, we're in trouble, because $E[\varepsilon_i | D_i]$ is not necessarily zero anymore!

Selection on observables: OLS



Selection on observables: OLS



Concerns with regression adjustment

When we run

$$Y_i = \alpha + \tau D_i + \gamma X_i + \nu_i$$

We can represent this as a difference in means between treated and untreated units:

$$\bar{Y}_U = \alpha + \gamma \bar{X}_U$$

and

$$\bar{Y}_T = \alpha + \tau + \gamma \bar{X}_T$$

Concerns with regression adjustment

When we run

$$Y_i = \alpha + \tau D_i + \gamma X_i + \nu_i$$

We can represent this as a difference in means between treated and untreated units:

$$\bar{Y}_U = \alpha + \gamma \bar{X}_U$$

and

$$\bar{Y}_T = \alpha + \tau + \gamma \bar{X}_T$$

$$\underbrace{(\bar{Y}_T - \bar{Y}_U)}_{\text{subtraction}} = \tau + \gamma(\bar{X}_T - \bar{X}_U)$$

Concerns with regression adjustment

When we run

$$Y_i = \alpha + \tau D_i + \gamma X_i + \nu_i$$

We can represent this as a difference in means between treated and untreated units:

$$\bar{Y}_U = \alpha + \gamma \bar{X}_U$$

and

$$\bar{Y}_T = \alpha + \tau + \gamma \bar{X}_T$$

$$\underbrace{(\bar{Y}_T - \bar{Y}_U)}_{\text{subtraction}} = \tau + \gamma(\bar{X}_T - \bar{X}_U)$$

$$\hat{\tau} = \underbrace{(\bar{Y}_T - \bar{Y}_U) - \hat{\gamma}(\bar{X}_T - \bar{X}_U)}_{\text{rearranged}}$$

Functional form assumptions

We rely heavily on two things:

1. $\bar{X}_T$ being close to $\bar{X}_U$
   - If $|\bar{X}_T - \bar{X}_U|$ is large, our estimate of $\hat{\tau}$ will be biased
   - We need "good overlap" between X_i for control and treatment
   - What does this mean when we have multiple X_is?

Functional form assumptions

We rely heavily on two things:

1. $\bar{X}_T$ being close to $\bar{X}_U$
    - If $|\bar{X}_T - \bar{X}_U|$ is large, our estimate of $\hat{\tau}$ will be biased
    - We need "good overlap" between X_i for control and treatment
    - What does this mean when we have multiple X_is?
2. Our assumed functional form
    - Our regression assumes the true relationship is $Y_i = \alpha + \tau D_i + \gamma X_i$
    - We actually need to control for $E[D_i|X_i]$, not just X_i
    - We should have run: $Y_i = \alpha + \tau D_i + \gamma E[D_i|X_i] + \nu_i$
    - If $X_i \neq E[D_i|X_i]$, then $\gamma(X_i - E[D_i|X_i])$ is in our error term

$\rightarrow$ $E[\nu_i|D_i, X_i]! = 0$ ☠

Approach 2: Matching

We can avoid some concerns by matching:

- We compare untreated units to treated units with **identical** X_is
- Difference in outcomes between treated and untreated is our $\hat{\tau}$

Approach 2: Matching

We can avoid some concerns by matching:

- We compare untreated units to treated units with **identical** X_is
- Difference in outcomes between treated and untreated is our $\hat{\tau}$
- Since we're comparing identical X_is:
  - We guarantee treated and control units have similar X_i

Approach 2: Matching

We can avoid some concerns by matching:

- We compare untreated units to treated units with **identical** X_is
- Difference in outcomes between treated and untreated is our $\hat{\tau}$
- Since we're comparing identical X_is:
  - We guarantee treated and control units have similar X_i
  - Functional form is irrelevant
- Still requires:
  - $Y_i \perp D_i | X_i$
  - $0 < Pr(D_i = 1 | X_i = x) < 1$

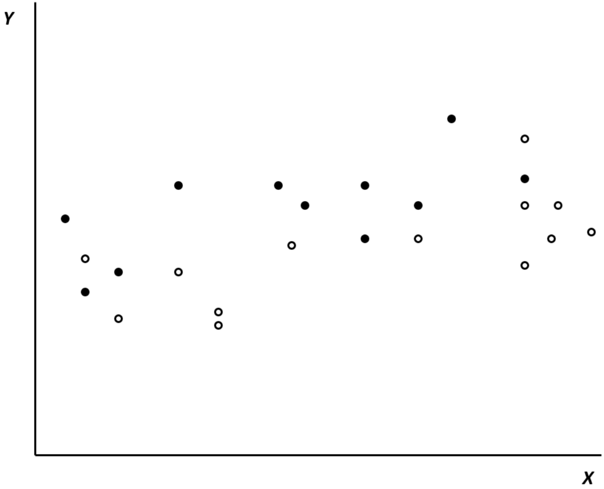
The exact matching estimator

The simplest possible matching estimator is exact matching:

1. Divide data into "cells" uniquely defined by the covariates
2. For each value of $X = x$ (each cell), calculate $\bar{Y}_T$ and $\bar{Y}_U$
3. Calculate $\bar{Y}_T - \bar{Y}_U$ for each $X = x$
4. Estimate τ^{ATE} as a weighted average of (3)

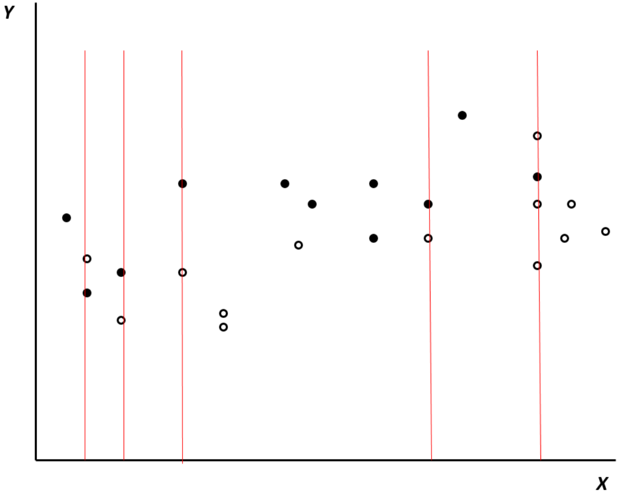
Note: This works for more than one X! See additional slides.

Getting "close"



How would we implement exact matching?

Getting "close"



We would only keep data with identical Xs!

The Curse of Dimensionality

We're often interested in matching on multiple Xs:

- You have to be very lucky (dumb?) to think selection on only one X!
- Much more likely: selection depends on many Xs
- But the more Xs you have, the less likely you are to have a match
- (This same issue bites for regression too)

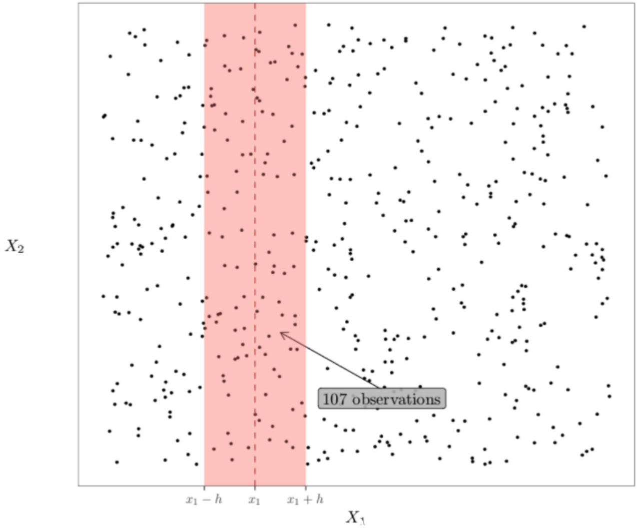
The Curse of Dimensionality

We're often interested in matching on multiple Xs:

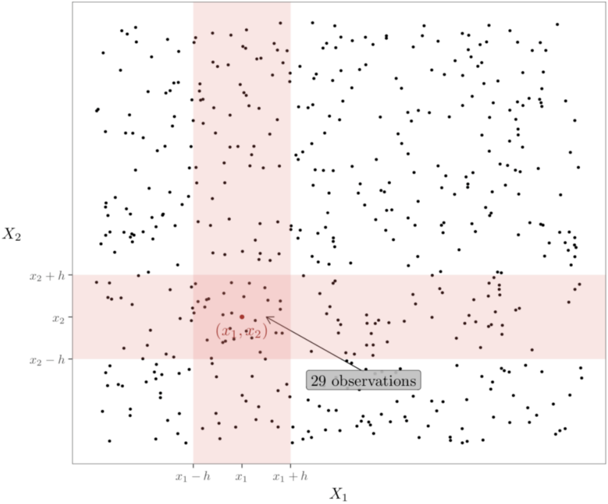
- You have to be very lucky (dumb?) to think selection on only one X!
- Much more likely: selection depends on many Xs
- But the more Xs you have, the less likely you are to have a match
- (This same issue bites for regression too)
- From *my* PhD econometrics class:

  "Although you can sometimes reduce the dimensionality problems by making various parametric assumptions...you can never truly defeat the Curse of Dimensionality. It is, after all, a curse."
  – Michael L. Anderson, UC Berkeley

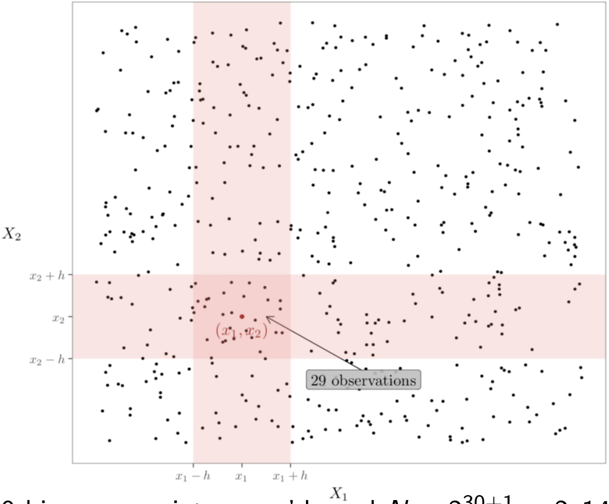
The Curse of Dimensionality



The Curse of Dimensionality



The Curse of Dimensionality



For $K = 30$ binary covariates, you'd need $N = 2^{30+1} = 2,147,483,648$

☠

Examining the exact matching estimator

<u>The good news:</u>

- Creates observably identical treated and untreated comparisons
  - No need to worry about $\bar{X}_T$ and $\bar{X}_U$ being far apart
- Makes no functional form assumptions
  - Don't have to worry about **how** to control for Xs

$\rightarrow$ This is a very flexible estimator!

Examining the exact matching estimator

The good news:

- Creates observably identical treated and untreated comparisons
  - No need to worry about $\bar{X}_T$ and $\bar{X}_U$ being far apart
- Makes no functional form assumptions
  - Don't have to worry about **how** to control for Xs

→ This is a very flexible estimator!

The bad news:

- It doesn't work for *continuous* Xs!
  - How do you define cells of continuous variables?

→ Very flexible, but not super practically useful?

Going beyond the exact matching estimator

<u>What can we do when we have continuous X?</u>

- For each treated unit, we want to estimate its untreated counterfactual:
- We'd like an estimate of $Y_i(0)$ for units with $D_i = 1$
- We can try to go for $Y(0; x)$ for a given $X_i = x$

Going beyond the exact matching estimator

<u>What can we do when we have continuous X?</u>

- For each treated unit, we want to estimate its untreated counterfactual:
- We'd like an estimate of $Y_i(0)$ for units with $D_i = 1$
- We can try to go for $Y(0; x)$ for a given $X_i = x$

☠ What if we don't have any untreated people with $X_i = x$?

→ Find untreated units with X_i *close to* $X_i = x$

- With this population, we can simply take $\bar{Y}(0; x^{\text{close}})$
    - This is still flexible and non-parametric!

Going beyond the exact matching estimator

<u>What can we do when we have continuous X?</u>

- For each treated unit, we want to estimate its untreated counterfactual:
- We'd like an estimate of $Y_i(0)$ for units with $D_i = 1$
- We can try to go for $Y(0; x)$ for a given $X_i = x$

☠ What if we don't have any untreated people with $X_i = x$?

→ Find untreated units with X_i *close to* $X_i = x$

- With this population, we can simply take $\bar{Y}(0; x^{\text{close}})$
  - This is still flexible and non-parametric!

How do we define "close"?

Additional matching estimators

In datasets with continuous Xs, we can:

1. Match to "nearest neighbors"
2. Match within a bandwidth

$\rightarrow$ Different ways of getting "closeness"

$\rightarrow$ Non-parametric: no real functional form assumption on $Y(X)$

Nearest-neighbor matching

For each treated unit $i \in T$, we find its "nearest neighbor" in X:

- Take the untreated unit $j \in U$ with the smallest $|X_j - X_i|$
- Now your "counterfactual" is $\hat{Y}_i(0) = Y_j(0)$
- Repeat this for all treated units $i \in T$

$$\hat{\tau}^{ATT} = \frac{1}{N_T} \sum_{i \in T} (Y_i(1) - \hat{Y}_i(0))$$

Nearest-neighbor matching

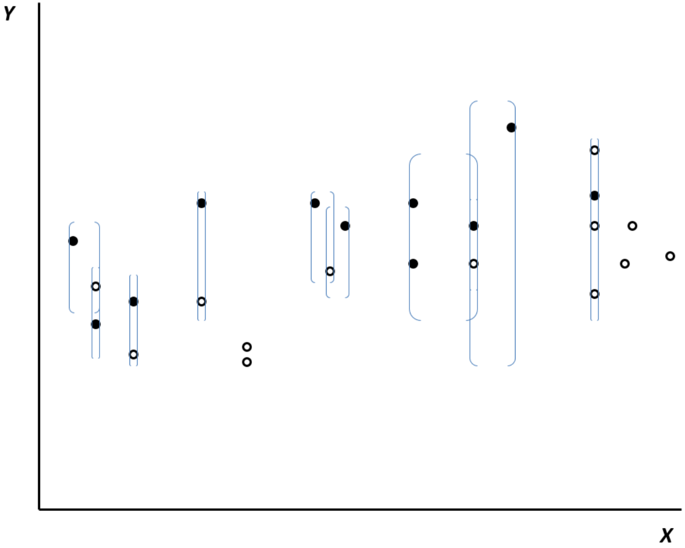
For each treated unit $i \in T$, we find its "nearest neighbor" in X:

- Take the untreated unit $j \in U$ with the smallest $|X_j - X_i|$
- Now your "counterfactual" is $\hat{Y}_i(0) = Y_j(0)$
- Repeat this for all treated units $i \in T$

$$\hat{\tau}^{ATT} = \frac{1}{N_T} \sum_{i \in T} (Y_i(1) - \hat{Y}_i(0))$$

- You can easily do this for an arbitrarily large K nearest neighbors
- With multiple neighbors, just average over the $Y_j(0)$'s to get $\hat{Y}_i(0)$
- Still not picking a functional form, but we are picking K

Getting "close" with nearest neighbors



Bandwidth matching

For each $i \in T$, we find $j \in U$ within a bandwidth h:

- Take all untreated units $j \in U$ with $x_j \in [X_i - h, X_i + h]$
- Now your "counterfactual" is $\hat{Y}_i(0) = \bar{Y}_j(0; X_i - h \leq X_j \leq X_i + h)$
- Repeat this for all treated units $i \in T$

$$\hat{\tau}^{ATT} = \frac{1}{N_T} \sum_{i \in T} (Y_i(1) - \hat{Y}_i(0))$$

Bandwidth matching

For each $i \in T$, we find $j \in U$ within a bandwidth h:

- Take all untreated units $j \in U$ with $x_j \in [X_i - h, X_i + h]$
- Now your "counterfactual" is $\hat{Y}_i(0) = \bar{Y}_j(0; X_i - h \leq X_j \leq X_i + h)$
- Repeat this for all treated units $i \in T$

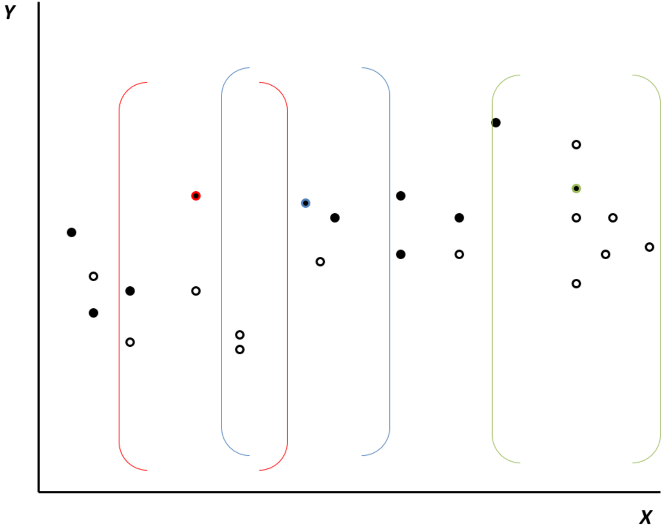
$$\hat{\tau}^{ATT} = \frac{1}{N_T} \sum_{i \in T} (Y_i(1) - \hat{Y}_i(0))$$

- How do you choose a bandwidth?
  - Narrow: we'll get an accurate, but noisy estimate (similar Xs, few observations)
  - Wide: we'll get an inaccurate, but precise estimate (different Xs, many observations)

$\rightarrow$ We face a **bias-variance tradeoff**

$\rightarrow$ There are fancy tools for this (outside this class)

Getting "close" with bandwidths



A note on what we're estimating

For all three matching estimators, we can estimate ATE, ATT, or ATN:

- The trick is to make sure we know which one we're getting
- *Exact matching:*
  - ATE: weight relative to the full sample: $\hat{\Delta}^{ATE} = \sum_{j=1}^{\#\text{ of cells}} \frac{N_j}{N} \hat{\Delta}_j$
  - ATT: weight relative to the treated sample: $\hat{\Delta}^{ATT} = \sum_{k=1}^{\#\text{ of treated cells}} \frac{N_{k,T}}{N_T} \hat{\Delta}_k$
  - ATN: weight relative to the untreated sample: $\hat{\Delta}^{ATN} = \sum_{l=1}^{\#\text{ of untreated cells}} \frac{N_{l,U}}{N_U} \hat{\Delta}_l$
- *Nearest neighbor and bandwidth matching:*
  - ATT: For each treated unit, find untreated matches: $\hat{\tau}^{ATT} = \frac{1}{N_T} \sum_{i \in T} (Y_i(1) - \hat{Y}_i(0))$
  - ATN: For each untreated unit, find treated matches: $\hat{\tau}^{ATN} = \frac{1}{N_U} \sum_{i \in U} (\hat{Y}_i(1) - Y_i(0))$
  - ATE: Weight the ATT and ATN: $\hat{\tau}^{ATE} = \frac{N_T}{N_T + N_U} \hat{\tau}^{ATT} + \frac{N_N}{N_T + N_U} \hat{\tau}^{ATN}$

An example: Appliance replacements in Mexico

Policy issue:

- Energy efficiency is seen as a "win-win":
  - Customers "win" by saving on their power bills
  - The planet "wins" because we reduce GHGs
- But does EE actually work?

Program:

- Mexican government subsidized HVAC and fridge replacements
- Cute title: "Cash for coolers"
- **Non-experimental** program:
  - If you had an old appliance, you were eligible

→ We don't have randomization, so we need an SOO design

Estimating treatment effects of appliance replacements

What happens to energy consumption with a replacement (simplified)?

$$Y_i = \tau D_i + \varepsilon_i$$

where

Y_i is kWh of electricity use at household i

$D_i = \mathbf{1}[\text{New appliance}]_i$ is an indicator for getting an upgrade

ε_i is an error term

Estimating treatment effects of appliance replacements

What happens to energy consumption with a replacement (simplified)?

$$Y_i = \tau D_i + \varepsilon_i$$

where
Y_i is kWh of electricity use at household i
$D_i = \mathbf{1}[\text{New appliance}]_i$ is an indicator for getting an upgrade
ε_i is an error term

This is **not an experiment:** we need a control group! A few options...

1. Treated vs. randomly selected untreated households
2. Only treated households (leverages time comparison – more on that later!)
3. Matched: account numbers
   - Think of this as NN or BW matching
4. Matched: 10 closest account numbers plus consumption
   - BW on accounts, NN on consumption

Mexican appliances: How well does matching work?

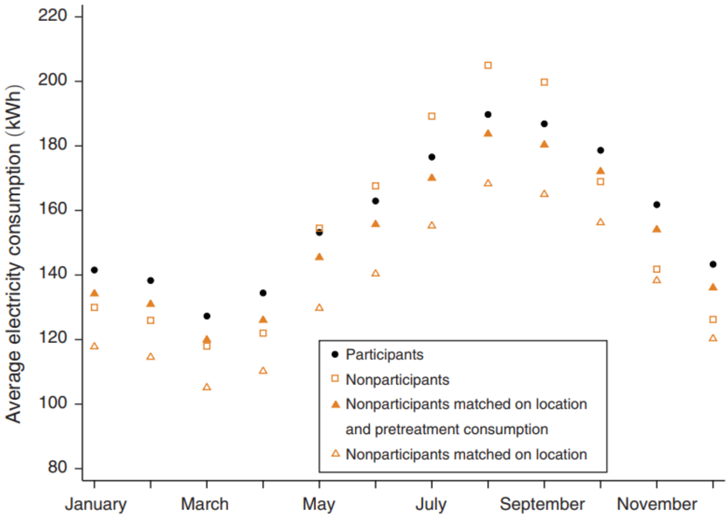
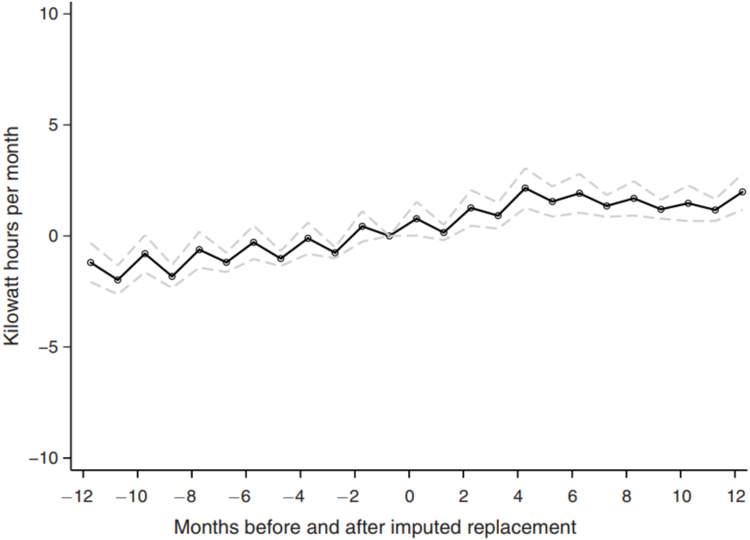
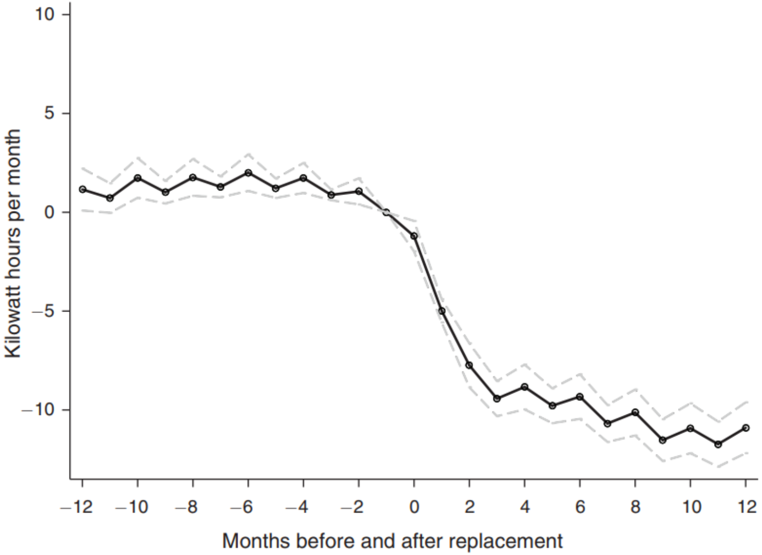


FIGURE 2A. COMPARING PARTICIPANTS TO NONPARTICIPANTS: REFRIGERATORS

Mexican appliances: Untreated consumption patterns



Mexican appliances: Treatment effects



Mexican appliances: Policy implications

Table 6—Electricity Expenditures, Carbon Dioxide Emissions, and Cost-Effectiveness

	Refrigerators (1)	Air conditioners (2)	Both appliances combined (3)
Panel A. Mean per replacement			
Mean annual change in electricity consumption per replacement (kilowatt hours)	−135	91	—
Mean annual change in household expenditure per replacement (2010 US$)	−$13	$9	—
Panel B. Totals			
Total replacements nationwide (between May 2009 and April 2011)	858,962	98,604	957,566
Total annual change in electricity consumption (gigawatt hours)	−115.7	9.0	−106.7
Total annual change in household expenditures (in millions 2010 US$)	−$11.1	$0.9	−$10.2
Total annual change in carbon dioxide emissions (thousands of tons)	−62.2	4.8	−57.4
Panel C. Cost-effectiveness			
Total Direct program cost (in millions 2010 US$)	$129.4	$13.4	$142.7
Program cost per kilowatt hour (2010 US$)	$0.25	—	$0.29
Program cost per ton of carbon dioxide (2010 US$)	$457	—	$547

Notes: Mean annual change in electricity consumption per replacement comes from column 4 of Table 4. Change in expenditures is calculated using an average price of $0.096 per kilowatt hour. Carbon dioxide emissions are calculated using 0.538 tons of carbon dioxide per megawatt hour (538 tons per gigawatt hour) following Johnson et al. (2009). Direct program cost is the dollar value of the cash subsidies and excludes administrative costs. In calculating the program cost per kilowatt hour and program cost per ton of carbon dioxide we assumed that the program accelerated replacement by five years and used a 5 percent annual discount rate.

Wrapping up SOO

We've covered the two main ways of doing SOO

1. Regression adjustment
   - Controlling for stuff
   - Makes parametric assumptions
2. Matching
   - Pairing observations
   - Less parametric

Wrapping up SOO

We've covered the two main ways of doing SOO

1. Regression adjustment
   - Controlling for stuff
   - Makes parametric assumptions
2. Matching
   - Pairing observations
   - Less parametric

A few last words:

- There are other, fancier ways to do this
- All make the extremely strong conditional independence assumption

→ This is generally not reasonable in real life!

→ We will end our treatment of SOO here

Recap

TL;DR:

1. Selection on observables designs are **dubious**
2. They require extremely strong assumptions!
3. But as a last resort, matching can be useful