Lecs 15-16

10/11/17 11:35

Recap: Regression Discontinuity Designs

$Z_i = \mathbb{I}[X_i > c]$

$\mathbb{E}[Y(z) / X = x]$ is continuous in a nbhd at $x = c$ $\forall z \in \{0,1\}$

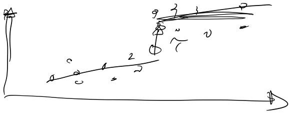
$\Rightarrow$ $\mathbb{E}[Y / X = x]$ is continuous around $x = c$
except for one
discontinuity at $x = c$
corresponds to effect at c

$$\tau = CATE(c) = \mathbb{E}[TE \mid X = c]$$
$$= \mathbb{E}[Y(1) \mid X = c] - \mathbb{E}[Y(0) \mid X = c]$$
$$= \lim_{x \downarrow c} \mathbb{E}[Y / X = x] - \lim_{x \uparrow c} \mathbb{E}[Y / X = x]$$

$$\hat{\tau} = \lim_{x \downarrow c} \hat{\mathbb{E}}[Y / X = x] - \lim_{x \uparrow c} \hat{\mathbb{E}}[Y / X = x]$$

a few options to do

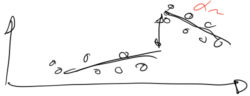
- Regress $Y \sim \alpha + \beta X + \tau \mathbb{I}[X > c]$



- Use different slopes

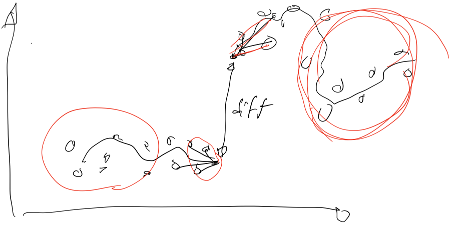
Regress $Y \sim \alpha_1 + \beta_1 X$ on $X < c$

$Y \sim \alpha + \tau + \beta_2 X$ on $X > c$



$\tau = \alpha_2 - \alpha_1$

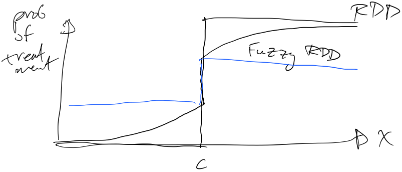
- Do some local nonparametric regression on each side



Fuzzy Regression Discontinuity

What if threshold is not exact?

Prob of treatment doesn't sharply from 0 to 1 at the cutoff.



Fuzzy RDDs look like encouragement designs

Just like noncompliance

Still need a jump:

$$\lim_{x \downarrow c} \mathbb{P}(Z_i=1 | X_i=x) \neq \lim_{x \uparrow c} \mathbb{P}(Z_i=1 | X_i=x)$$

The Wald-RDD estimate:

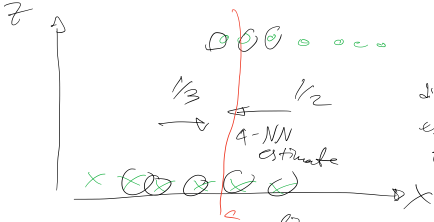
$$\hat{\tau} = \frac{\lim_{x \downarrow c} \hat{E}[Y | X=x] - \lim_{x \uparrow c} \hat{E}[Y | X=x]}{\lim_{x \downarrow c} \hat{\mathbb{P}}(Z_i=1 | X_i=x) - \lim_{x \uparrow c} \hat{\mathbb{P}}(Z_i=1 | X_i=x)}$$

← ITT

← Prob of Comp

What does $\hat{\tau}$ estimate?

$$\mathbb{E}\left[Y(1) - Y(0) \;\middle|\; \begin{matrix} \text{compliers} \\ X=c \end{matrix} \right]$$



diff $= \frac{1}{6}$
estimate for denominator

$\lim_{x \uparrow c} \mathbb{P}(Z=1|X=x)$ $\lim_{x \downarrow c} \mathbb{P}(Z=1|X=x)$

Alternative: $\mathbb{P}(Z=1|X=x) = \text{logit}(\alpha+\beta x+\theta \mathbb{I}[x>c])$

Next topic: Observational Studies Under Ignorability

In the kind of obs study we'll discuss, we have no control on treatment (as in RCTs) / instrument (as in IVs) or any a priori understanding of treatment assignment mechanism (as in RDDs)

Some questions RCTs cannot answer:

① What is the effect of smoking on lung cancer incidence?

② What is the effect of lead paint in a child's home on their intelligence/academic achievement?

③ What is the effect of UN peace-keeping missions on incidence of violence?

④ What is the effect of diabetes

on microbiome bisbiosis?

⑤ What is the effect of being a cis-woman on wages or college admissions?

⑥ What is the effect of ballot measures on voter turnout?

Many situations w/ causal effects of interest but where RCT/ /RCT w/ noncomp/ RDD is not possible:

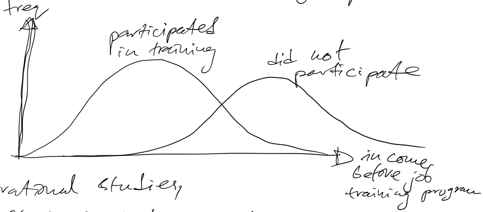
- Treatment is plausibly harmful or unethical to administer (smoking, exposure to Radon gas)
- Treatment may be beyond reach of the experimenter like those controlled by a political or biological process or management decisions

Sometimes we can find a natural experiment (IVs) but usually this is hard.

Instead, we conduct an observational study: we observe rather than design/influence the assignment to treatment & the corresponding outcome.

A key feature of observational studies is systematic differences b/w treated & control groups

Ex:



In observational studies, groups differ prior to treatment in ways that matter for outcome of interest.

In controlled trials, we eliminated systematic differences using randomization & eliminated random finite-sample differences using balancing.

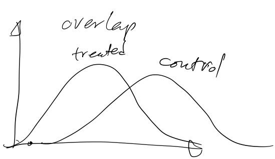
What to do?

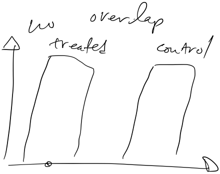
We can try to control for differences:

Compare like to like

- When will comparing like to like work?

a) Overlap: need like to be available





b) Ignorability: "like" needs to account for all differences that are ① systematic
② matter for the outcome of interest

Set-up

Again use potential outcomes

Subjects $i=1,\dots,n$

Treatment $Z_i \in \{0,1\}$ Covars $X_i \in \mathcal{X}$

Outcomes $Y_i(1), Y_i(0) \in \mathbb{R}$

$$Y_i = Y_i(Z_i)$$

Quantities of interest

$$\text{SATE} = \frac{1}{n}\sum_i \left(Y_i(1)-Y_i(0)\right) \qquad \text{PATE} = \mathbb{E}\left[Y_i(1)-Y_i(0)\right] = \mathbb{E}\,\text{SATE}$$

$$\text{SATT} = \frac{1}{n_1}\sum_{i:Z_i=1}\left(Y_i(1)-Y_i(0)\right)$$

$$\text{PATT} = \mathbb{E}\,\text{SATT} = \mathbb{E}\left[Y(1)-Y(0) \mid Z=1\right]$$

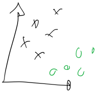
interest in obs studies b/c usually there are plenty of controls

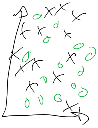
Assumption (Weak overlap) (note diff to $\mathbb{P}(Z_i=1 \mid X_i=x)<1 \ \forall x \in \mathbb{R}^d$)

$$\Uparrow \qquad 0 < \mathbb{P}(Z_i=1 \mid X_i) < 1 \quad \text{a.s.}$$

both sides needed for SATE

only this side needed for SATT





Assumption (Strong overlap)

$$0 < p_{min} \le \mathbb{P}(Z_i=1 \mid X_i) \le p_{max} < 1 \quad \text{a.s.}$$

Assumption (Strong ignorability)

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \perp\!\!\!\perp Z_i \mid X_i$$

$\Downarrow$

notice difference to $Y_i \perp\!\!\!\perp Z_i \mid X_i$

} all are untestable

Assumption (Weak ignorability)

$$\forall z \in \{0,1\} \qquad Y_i(z) \perp\!\!\!\perp Z_i \mid X_i$$

$$\Downarrow$$

Assumption (Weak-mean ignorability)

$$\forall z \in \{0,1\},\ \mathbb{E}[Y_i(z) \mid Z_i, X_i] = \mathbb{E}[Y_i(z) \mid X_i]$$

Under weak-mean ignorability

$$\begin{aligned}
\mathbb{E}[Y_i(z)] &= \mathbb{E}\big[\mathbb{E}[Y_i(z) \mid X_i]\big] \\
&= \mathbb{E}\big[\mathbb{E}[Y_i(z) \mid X_i, Z_i = z]\big] \\
&= \mathbb{E}\big[\mathbb{E}[Y_i(Z_i) \mid X_i, Z_i = z]\big] \\
&= \boxed{\mathbb{E}\big[\mathbb{E}[Y_i \mid X_i, Z_i = z]\big]} \\
&= \int_x \mathbb{E}[Y \mid X = x, Z = z]\, d\mathbb{P}(X = x)
\end{aligned}$$

Ignorability aka unconfoundedness

aka no unobserved confounders

aka selection on observables

SATT version $z' = 1$

$$\begin{aligned}
\mathbb{E}[Y_i(z) \mid Z_i = z'] &= \mathbb{E}\big[\mathbb{E}[Y_i \mid X_i, Z_i = z] \mid Z_i = z'\big] \\
&= \int_x \mathbb{E}[Y \mid X = x, Z = z]\, d\mathbb{P}(X = x \mid Z = z')
\end{aligned}$$

Regression Approaches to Obs Studies

If we can fit $\mathbb{E}[Y \mid X = x, Z = z]$

then we can plug in

w/ $\mathbb{P}(X = x) = \mathbb{P}_n(X = x)$

Linear model:

Posit $\mathbb{E}[Y_i(z) \mid X_i = x] = \alpha + \beta^T x + \tau z$

OLS regress of Y_i on X_i, Z_i

set $\hat{\tau}$ to coeff on Z

More common: $\mathbb{E}[Y_i(z) | X_i = x] = \alpha_z + \beta_z^T x$

for ATE: $\hat{\tau} = \frac{1}{n} \sum_i (\hat{\alpha}_1 + \hat{\beta}_1^T X_i - \hat{\alpha}_0 - \hat{\beta}_0^T X_i)$

$= (\hat{\alpha}_1 + \hat{\beta}_1^T \bar{X}) - (\hat{\alpha}_0 + \hat{\beta}_0^T \bar{X})$

equiv: OLS regress Y_i on $X_i, Z_i, Z_i(X_i - \bar{X})$

set $\hat{\tau}$ to coeff on Z_i

for ATT: $\hat{\tau} = \frac{1}{n_1} \sum_{i:Z_i=1} (\hat{\alpha}_1 + \hat{\beta}_1^T X_i - \hat{\alpha}_0 - \hat{\beta}_0^T X_i)$

$= (\hat{\alpha}_1 + \hat{\beta}_1^T \bar{X}_1) - (\hat{\alpha}_0 + \hat{\beta}_0^T \bar{X}_1)$

equiv: OLS regress Y_i on $X_i, Z_i, Z_i(X_i - \bar{X}_1)$

More generally: Fit any predictive model

$\hat{f}_z(x)$ to $\{(X_i, Y_i) : Z_i = z\}$

for ATE: $\hat{\tau} = \frac{1}{n} \sum_i (\hat{f}_1(X_i) - \hat{f}_0(X_i))$

for ATT: $\hat{\tau} = \frac{1}{n_1} \sum_{i:Z_i=1} (\hat{f}_1(X_i) - \hat{f}_0(X_i))$

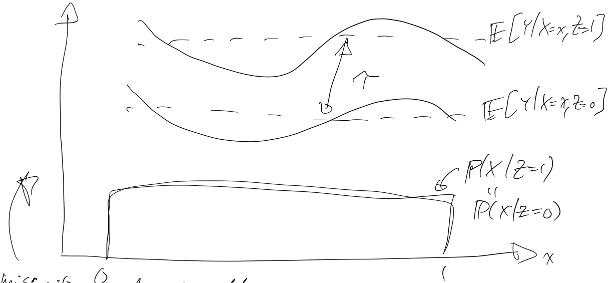
$= \frac{1}{n_1} \sum_{i:Z_i=1} (Y_i - \hat{f}_0(X_i))$

note that if $\hat{f}_z$ is fit w/ a free constant term to minimize sum of squared errors then we will always have

$$\frac{1}{n_z} \sum_{i:Z_i=z} \hat{f}_z(X_i) = \frac{1}{n_z} \sum_{i:Z_i=z} Y_i$$

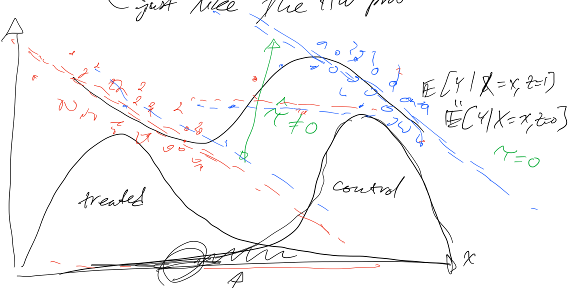
$$= \frac{n_1}{n} \cdot \left(\frac{1}{n_1} \sum_{i:Z_i=1} (Y_i - \hat{f}_0(X_i)) \right) + \frac{n_0}{n} \left(\frac{1}{n_0} \sum_{i:Z_i=0} (\hat{f}_1(X_i) - Y_i) \right)$$

Covariate Shift & Misspecification



misspecification doesn't matter if no systematic diffs b/w groups

↖ just like the HW prob



if we limit to area where like looks like like