ECON 480-3
LECTURE 10: RDD AND MATCHING ESTIMATORS

Ivan A. Canay
Northwestern University

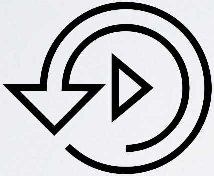
PAST & FUTURE

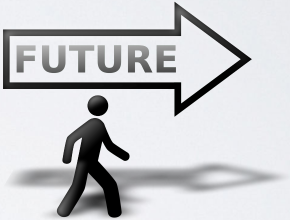
LAST CLASS

- Non-Parametric Regression
- q-NN Estimator
- Nadaraya-Watson Estimator
- Local Linear Regression

TODAY

- The Regression Discontinuity Design
- Sharp and Fuzzy RDD
- Bandwidth Choice
- Matching Estimators





Evaluation Methods

Main goal: learn about treatment effect of policy or intervention on an outcome.

Key requirement: constructing counterfactuals in a convincing way.

Five approaches: to deal with endogenous selection

- Randomized controlled experiments (or RCTs)
  ... exploit controlled/randomized assignment rule.
- Natural experiments
  ... exploit some "natural" randomization - a type of DiD approach.
- Instrumental methods and control function methods
  ...rely on exclusion restrictions or models for the assignment rule.
- Matching methods
  ...attempt to reproduce the treatment group among the non-treated.
- Discontinuity design methods
  ...exploit discreteness in assignment rule.

General notation

- **Potential outcomes**:

$$\begin{array}{ll} Y(0) & \text{potential outcome in the absence of treatment} \\ Y(1) & \text{potential outcome in the presence of treatment} \end{array} .$$

- **Treatment assignment**

$$D \in \{0, 1\}\,.$$

- **Common parameters of interest:**

$$\begin{array}{ll} Y(1) - Y(0) & \text{treatment effect} \\ E[Y(1) - Y(0)] & \text{average treatment effect (ATE)} \\ E[Y(1) - Y(0)|D = 1] & \text{average treatment on the treated (ATT)} \end{array} .$$

- **Problem:** Only one potential outcome is observed for each unit.

RDD: basics

- RD Designs are defined by a triplet: score, cutoff, treatment.
- Units receive a score (e.g., grade in the SAT).
- A treatment is assigned based on the score and a known cutoff (e.g., 2100).
- The treatment (e.g., scholarship):

  is given to units whose score is above the cutoff.

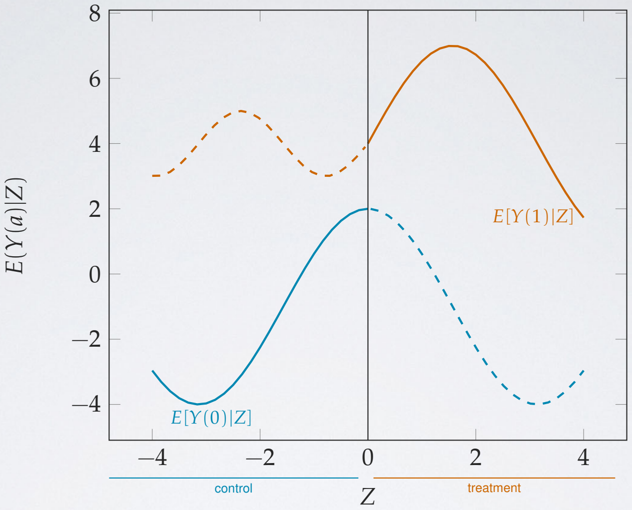
  is withheld from units whose score is below the cutoff.

- The abrupt change in the probability of treatment assignment allows us to learn something about effect of treatment.

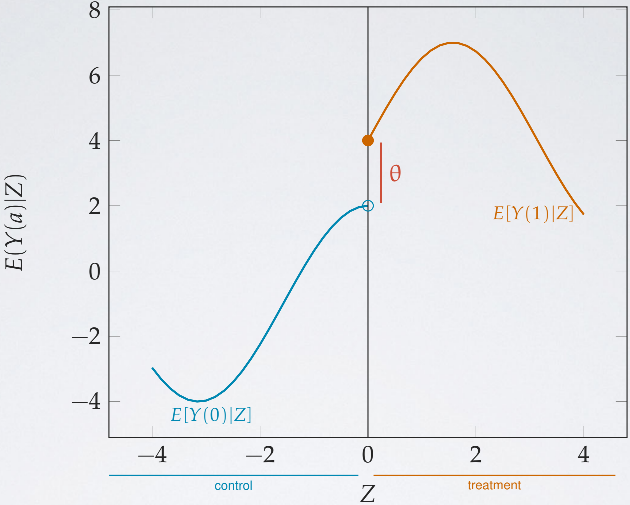
RDD: NOTATION

- **Score or running variable**: $Z_i \in \mathbf{R}$ observed for all units.
- **Cutoff:** a known constant c. We normalize to 0 wlog.
- **Treatment assignment (sharp)**:
$$D_i = I\{Z_i \geqslant 0\} .$$
- **Observed outcome** Y_i is given by (c normalized to 0)
$$Y_i = \begin{cases} Y_i(0) & \text{if } Z_i < 0 \\ Y_i(1) & \text{if } Z_i \geqslant 0 \end{cases}$$
and therefore it follow that
$$E[Y|Z=z] = \begin{cases} E[Y(0)|Z=z] & \text{if } z < 0 \\ E[Y(1)|Z=z] & \text{if } z \geqslant 0 \end{cases} .$$
- **Idea:** Exploit discontinuity in $E[Y|Z]$.

RDD: graphical intuition



RDD: Graphical Intuition



Continuity: key assumption

- A special situation occurs at the cutoff $Z = 0$
  ... only point at which we may "almost" observe both curves .

- Consider two groups of units:

  with score equal to 0, $Z_i = 0 \longrightarrow$ treated

  with with score barely below 0, $Z_i = -\varepsilon \longrightarrow$ control

- If the values of $E[Y(0)|Z = -\epsilon]$ are not abruptly different from $E[Y(0)|Z = 0]$, then units with $Z_i = -\epsilon$ are a valid counterfactual to units with $Z_i = 0$.

- More formally,

$$E[Y(1) - Y(0)|Z = 0] = \underbrace{E[Y(1)|Z = 0]}_{\mathbf{E[Y|Z=0]}} - \underbrace{E[Y(0)|Z = 0]}_{\lim_{\mathbf{z \to 0^-}} \mathbf{E[Y|Z=z]}} \; .$$

- **Requirement**: $E[Y(0)|Z = z]$ is continuous at $z = 0$.

RDD parameter

$$\theta_{\text{srd}} = E[Y(1) - Y(0)|Z = 0] = E[Y|Z = 0] - \lim_{z \uparrow 0} E[Y|Z = z] \ .$$

- **Local**: θ_{srd} is a “local” average treatment effect like the LATE parameter discussed under IV, but is not necessarily the same parameter.

- **Trick**: RDD uses the discontinuous dependence of D on Z to identify a local average treatment effect

$$P\{D = 1|Z = z\} \text{ is discontinuous at } z = 0 \ .$$

- In the so-called Sharp design there is perfect compliance:
  - every unit with score above 0 receives treatment
  - every unit with score below 0 is in the control group

$$P\{D = 1|Z = z\} = \begin{cases} 0 & \text{if } z < 0 \\ 1 & \text{if } z \geqslant 0 \end{cases} \ .$$

- In the so-called Fuzzy design there may be imperfect compliance (Later)

QUESTIONS?

Estimation: sharp RDD

- Nonparametric estimators of $E[Y|Z=0]$ and $\lim_{z\to 0^-} E[Y|Z=z]$ could be constructed using data to the right and left of 0.
- **Boundary issue:** the point of interest, $E[Y|Z=0]$, is always on the boundary.
- LL estimators have better properties at the boundary than kernel estimators.
  - Bias of LL at boundary: order h^2.
  - Bias of NW at boundary: order $h > h^2$.
- **Local linear regression** is especially easy in this case:
  - Only care about estimation at the cut-off $c = 0$.
  - Compute kernel weights based on $c = 0$ and run a weighted least squares regression on observations either above (or below) zero.
- **With uniform kernel**: LL is the same as two unweighted linear regressions on observations with $Z_i \in [-h_n, 0)$ and $Z_i \in [0, h_n]$.

More on Estimation

- **Estimator of** $E[Y|Z=0]$.

$$\{\hat{\beta}_0^+, \hat{\beta}_1^+\} = \underset{(b_0^+, b_1^+)}{\text{argmin}} \sum_{i=1}^{n} k\left(\frac{Z_i}{h}\right) I\{Z_i \geqslant 0\}(Y_i - b_0^+ - b_1^+ Z_i)^2 \; .$$

- **Estimator of** $\lim_{z \to 0^-} E[Y|Z=z]$.

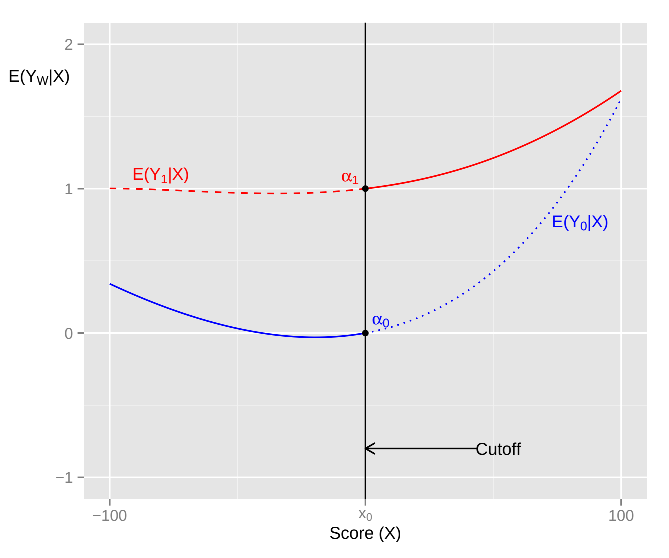
$$\{\hat{\beta}_0^-, \hat{\beta}_1^-\} = \underset{(b_0^-, b_1^-)}{\text{argmin}} \sum_{i=1}^{n} k\left(\frac{Z_i}{h}\right) I\{Z_i < 0\}(Y_i - b_0^- - b_1^- Z_i)^2 \; .$$

- Note: the regressor is $(Z_i - c)$ but $c = 0$ here.

Estimator of θ_{srd}

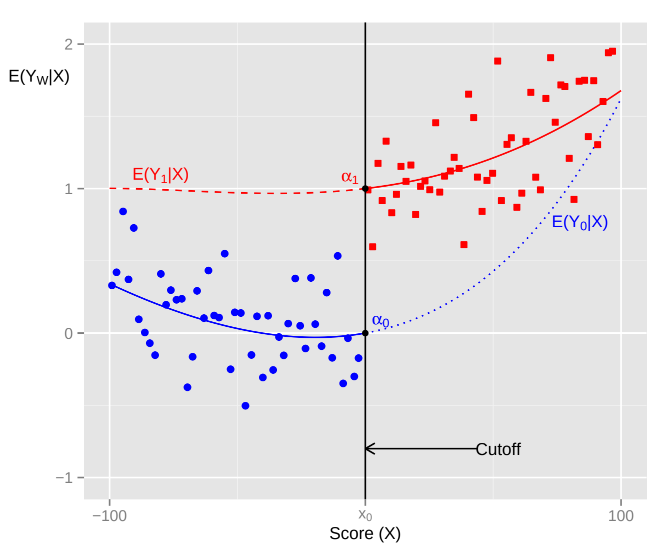
$$\hat{\theta}_{\text{srd}} = \hat{\beta}_0^+ - \hat{\beta}_0^- \; .$$

Local Linear Regression



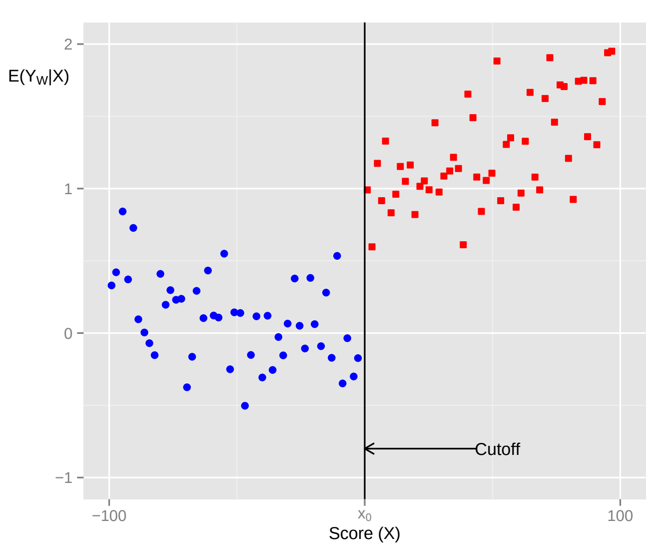
Beware of notation! Z **is** X, D **is** W, **and** β_0 **is** α

Local Linear Regression



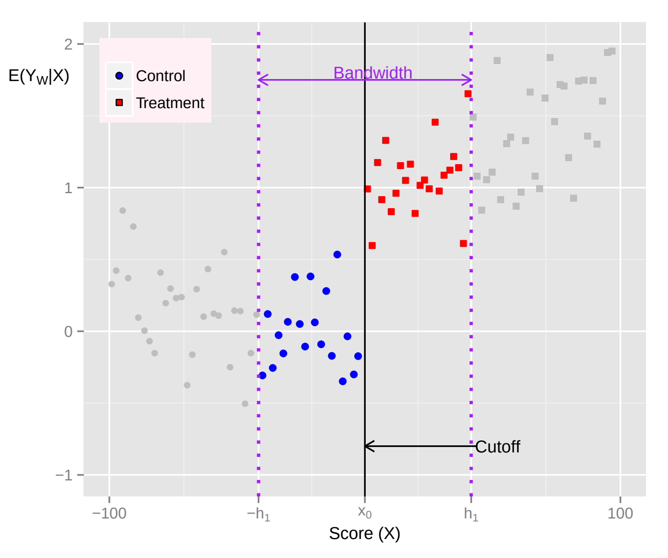
Beware of notation! Z **is** X**,** D **is** W**, and** β_0 **is** α

Local Linear Regression



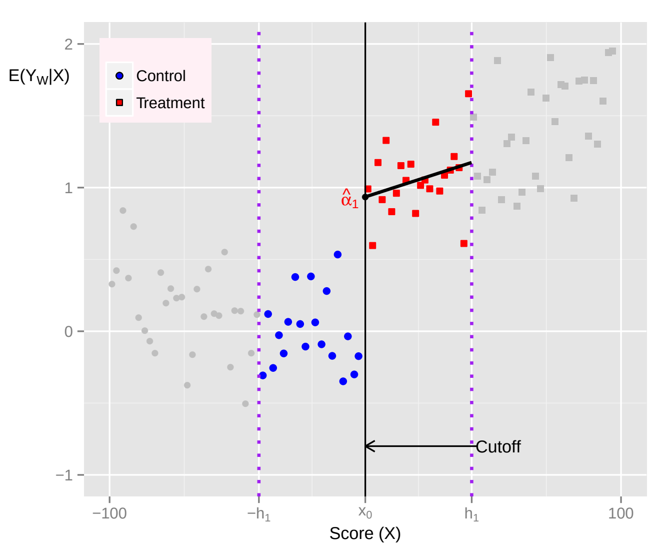
Beware of notation! Z **is** X, D **is** W, **and** β_0 **is** α

Local Linear Regression



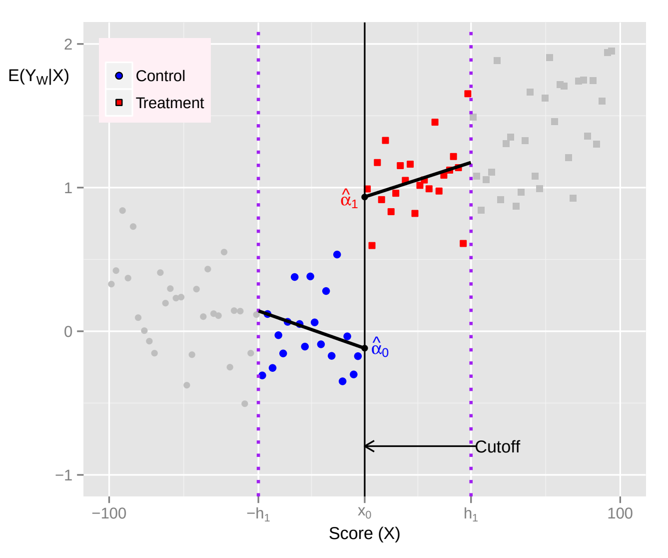
Beware of notation! Z is X, D is W, and β_0 is α

Local Linear Regression



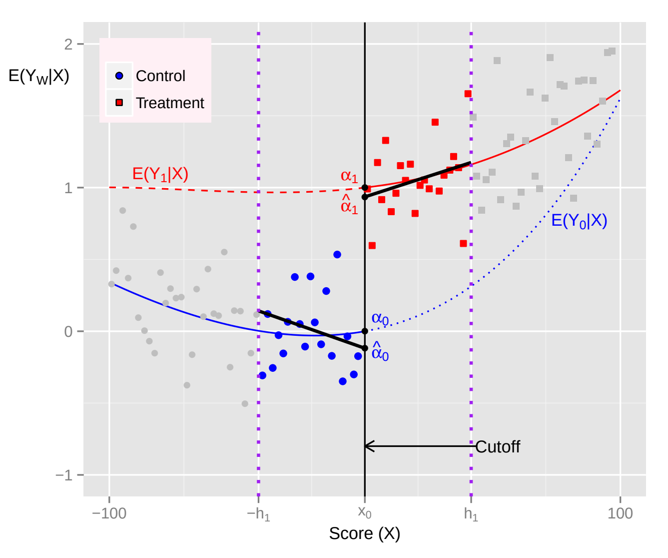
Beware of notation! Z **is** X, D **is** W, **and** β_0 **is** α

Local Linear Regression



Beware of notation! Z is X, D is W, and β_0 is α

Local Linear Regression



Beware of notation! Z **is** X, D **is** W, **and** β_0 **is** α

Bandwidth choice

- How to choose h_n?

- **Key idea**: bias and variance trade-off. Heuristically:

$$\uparrow h_n \Rightarrow \uparrow \text{Bias} \quad \text{but} \quad \uparrow h_n \Rightarrow \downarrow \text{Variance}$$

- Imbens & Kalyanaraman (2012,REStud): "optimal" plug-in,

$$\hat{h}_{\mathtt{IK}} = \hat{C}_{\mathtt{IK}} \cdot n^{-1/5}$$

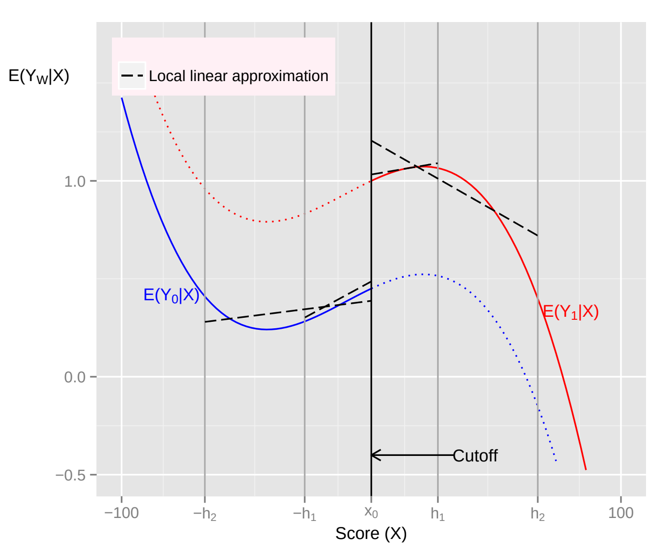
- Calonico, Cattaneo & Titiunik (2014, ECMA): improvement of IK

$$\hat{h}_{\mathtt{CCT}} = \hat{C}_{\mathtt{CCT}} \cdot n^{-1/5} \,.$$

  Important: also propose bias correction methods and new variance estimators that account for the additional noise introduced by estimating bias.

- Still common to see papers based on undersmoothing, i.e., use $nh^5 \to 0$ and ignore asymptotic bias. CCT is a better approach.

Bias in LL Regression - bandwidth



QUESTIONS?

Other RD designs

- **Sharp RD (SRD) and Fuzzy RD (FRD)**
  - Sharp: perfect compliance.
  - Fuzzy: $P\{D = 1|Z = z\}$ discontinuous at c but not necessarily from 0 to 1.

    some units above c may decide not to get treatment (e.g. voting eligibility at 18).
- **Kink RD (KRD) and Kink Fuzzy RD (KFRD)**
  - $P\{D = 1|Z = z\}$ has a kink at c (but it is continuous)
  - Introduces kinks of $E[Y|Z = z]$ at c.
- **Multiple scores RD and Geographic RD**
  - Discontinuity happens in $\mathbf{R}^2$ (or higher).
  - E.g., need high scores in “math” and “English”.
- **Multiple Cutoff RD**

Inference methods use similar tools (LL regression, etc) but are different.

Fuzzy RD

- **Imperfect compliance:**
  - probability of receiving treatment changes at c, but not necessarily from 0 to 1.
  - some units with score above c may decide not to take up treatment
  - E.g., Z is a test score and D is receipt of a scholarship: having a score larger than c makes the application "strong" but does not guarantee a scholarship.
- Allows for identification of **another** local treatment effect.
- Argument is similar to LATE, but a little subtler due to limits.
- **Canonical parameter**:

$$\theta_{\text{frd}} = \frac{E[Y_i(1)|Z_i = c] - E[Y_i(0)|Z_i = c]}{E[D_i(1)|Z_i = c] - E[D_i(0)|Z_i = c]} = \frac{\lim_{z \downarrow c} E[Y_i|Z_i = z] - \lim_{z \uparrow c} E[Y_i|Z_i = z]}{\lim_{z \downarrow c} E[D_i|Z_i = z] - \lim_{z \uparrow c} E[D_i|Z_i = z]}.$$

- **Interpretation**: ATE for units with $Z_i = c$ (by regression discontinuity), and only for compliers (people who are affected by the threshold).

Fuzzy RD: Estimation

- We now need **4 local linear regressions** at c

$$\{\hat\beta_0^+, \hat\beta_1^+\} = \underset{(b_0^+, b_1^+)}{\text{argmin}} \sum_{i=1}^{n} k\left(\frac{Z_i - c}{h}\right) I\{Z_i \geqslant c\}(Y_i - b_0^+ - b_1^+(Z_i - c))^2$$

$$\{\hat\beta_0^-, \hat\beta_1^-\} = \underset{(b_0^-, b_1^-)}{\text{argmin}} \sum_{i=1}^{n} k\left(\frac{Z_i - c}{h}\right) I\{Z_i < c\}(Y_i - b_0^- - b_1^-(Z_i - c))^2$$

$$\{\hat\gamma_0^+, \hat\gamma_1^+\} = \underset{(g_0^+, g_1^+)}{\text{argmin}} \sum_{i=1}^{n} k\left(\frac{Z_i - c}{h}\right) I\{Z_i \geqslant c\}(D_i - g_0^+ - g_1^+(Z_i - c))^2$$

$$\{\hat\gamma_0^-, \hat\gamma_1^-\} = \underset{(g_0^-, g_1^-)}{\text{argmin}} \sum_{i=1}^{n} k\left(\frac{Z_i - c}{h}\right) I\{Z_i < c\}(D_i - g_0^- - g_1^-(Z_i - c))^2$$

- Then estimate θ_{frd} as

$$\hat\theta_{\text{frd}} = \frac{\hat\beta_0^+ - \hat\beta_0^-}{\hat\gamma_0^+ - \hat\gamma_0^-}$$

Fuzzy RD as TSLS

- We can define $T = I\{Z \geqslant c\}$ as the **intention to treat**
- Then, T is a valid instrument for D.
- Conditional on Z, T is exogenous.
- Similar to the situation we found in LATE.

Fuzzy as TSLS

The LL approach with uniform kernels and same bandwidths is numerically equivalent to a TSLS regression:

$$Y_i = \delta_0 + \theta_{\text{frd}} D_i + \delta_1 (Z_i - c) + \delta_2 T_i (Z_i - c) + U_i$$

with T_i as the excluded instrument for D_i on the sample $\{i : c - h_n \leqslant Z_i \leqslant c + h_n\}$.

Validity of RD: manipulation

- RD imposes relatively weak assumptions.
- It also identifies a very specific and local parameter.
- Identification follows from continuity of $E[Y(d)|Z=z]$ at $z=0$.
- Assumption is fundamentally untestable.
- Concern about the following situation:
  - Running variable is a test score.
  - Individuals know the threshold and have the option of re-taking the test, and may do so if scores are just below the threshold.
  - Leads to a discontinuity of the density $f_Z(z)$ of Z at the threshold c, and possibly a discontinuity of $E[Y(d)|Z=z]$ since

$$E[Y(d)|Z=z] = \int y \, f_{Y(d)|Z}(y|z) dy \quad \text{where} \quad f_{Y|Z}(y|z) = \frac{f_{YZ}(y,z)}{f_Z(z)} .$$

  - This invalidates the design!
- This problem is called "manipulation" of the running variable, McCrary (2008).

Validity of RD: discontinuity in covariates

- Suppose there are other observed (X) and unobserved (U) factors that affect potential outcomes,

$$Y(d) = m_d(Z, X) + U \ .$$

- Suppose that the dist. of X is discontinuous at $z = 0$.
- The discontinuity in X at 0 may affect the outcome, and these effects may be attributed erroneously to the treatment of interest.
- **Common practice**: construct a test of the null that

$$H_0 : \lim_{z \uparrow 0} E(X|Z = z) = \lim_{z \downarrow 0} E(X|Z = z)$$

- **Rejection**: suggestive that $E(Y(d)|Z = z)$ may not be continuous either.
- The discontinuity in X may **confound** the effect of the treatment.
- Intuition is about the entire distribution of X; not only its mean: see Canay and Kamat (2018).

Validity of RD: Two tests

Manipulation

- McCrary (2008, JoE) proposes a test for continuity of the density of $f_z(z)$ at the cut-off c.
- In principle, one does not need continuity of the density of Z at c, but a discontinuity is suggestive of violations of the no-manipulation assumption.
- Bugni and Canay (2021) propose a new test based on order statistics that does not require smoothness assumptions.

Continuity of Covariates

- Canay and Kamat (2018) propose a test for the continuity of $F_{X|Z}(x|z)$ at the cut-off value.
- Test is easy to implement and based on permutation tests.
- Novel asymptotic arguments involved.

RD Packages

https://rdpackages.github.io

- **rdrobust package**: estimation, inference and graphical presentation using local polynomials, partitioning, and spacings estimators.
  - `rdrobust`: RD inference (point estimation and CI; classic, bias-corrected, robust).
  - `rdbwselect`: bandwidth or window selection (IK, CV, CCT).
  - `rdplot`: plots data (with “optimal" block length).
- **rddensity package**: discontinuity in density test at cutoff (a.k.a. manipulation testing) using local polynomial density estimator.
  - `rddensity`: manipulation testing using local polynomial density estimation.
  - `rdbwdensity`: bandwidth or window selection.

http://sites.northwestern.edu/iac879/software

- **rdperm package**: approximate permutation test for RDD
- **rdcont package**: approximate sign-test for RDD

QUESTIONS?

Matching Estimators

- Suppose we observe (Y, D, X) and consider the following assumption.

UNCONFOUNDEDNESS

$$(Y(0), Y(1)) \perp\!\!\!\perp D \mid X$$

- Other names: selection on observables, conditional independence, etc.
- Idea: find (or "match") units in the treatment group ($D = 1$) and control group ($D = 0$) with the same value of X, i.e., $X = x$,

$$E[Y|D = 1, X = x] - E[Y|D = 0, X = x] \; .$$

- Unconfoundedness: identifies the "conditional" ATE,

$$\begin{aligned} E[Y(1) - Y(0)|X = x] &= E[Y(1)|D = 1, X = x] - E[Y(0)|D = 0, X = x] \\ &= E[Y|D = 1, X = x] - E[Y|D = 0, X = x] \end{aligned}$$

where the first line follows by unconfoundedness.

On unconfoundedness

- **Unconfoundedness**: for subgroups of agents with the same X there are no unobservable differences between the treatment and control groups.
- Using the "matching" approach then identifies the CATE.
- To be able to "match", we need an overlap assumption:

$$0 < P\{D = 1|X = x\} < 1$$

- Complication arises when X is continuously distributed.
- **Overlap**: Compare with sharp RDD $\Rightarrow$ never holds!

$$P\{D = 1|Z < c\} = 0 \quad \text{and } P\{D = 0|Z \geqslant c\} = 0 \text{ .}$$

- **Unconfoundedness**: Compare with sharp RDD $\Rightarrow$ always holds!

$$D = I\{Z \geqslant c\} \quad \text{so} \quad (Y(0), Y(1)) \perp\!\!\!\perp D \mid Z \quad \text{trivially .}$$

Matching Metric

- If $X \in \mathbf{R}^k$ has continuous components, the event $\{X = x\}$ has measure zero .... so previous strategy is not feasible

- Idea: match X's that are "close" according to some matching metric.

Mahalanobis distance

A common matching metric is given by

$$M_{ij} = (X_i - X_j)'\Sigma^{-1}(X_i - X_j)$$

where $\Sigma = \text{Var}[X]$. Then j is the qth closest to X_i if

$$\sum_{s=1}^{n} I\{M_{is} \leqslant M_{ij}\} = q \; .$$

- Other metrics: Euclidean

$$M_{ij} = |X_i - X_j|$$

or the diagonal version of the Mahalanobis distance,

$$M_{ij} = (X_i - X_j)' \, \text{diag}[\Sigma^{-1}](X_i - X_j) \; .$$

Matching Estimator

- For a fix q, let $j_q(i)$ be the index $j \in \{1, \ldots, n\}$ that solves

Opposing treatment	$D_j = 1 - D_i$
Opposing qth closest to i	$\sum_{s:D_s=1-D_i} I\{M_{is} \leqslant M_{ij}\} = q$.

- $j_q(i)$ is the index of the unit that is the qth closest to unit i in terms of the covariate values, among the units with the treatment opposite to that of unit i.
- Let $\mathcal{J}_q(i)$ denote the set of indices for the first q matches for unit i:

$$\mathcal{J}_q(i) = \{j_1(i), \ldots, j_q(i)\}$$

- The matching estimator of $\theta_{\text{ate}} = E[Y(1) - Y(0)]$ is given by

Matching Estimator

$$\hat{\theta}_{\text{ate}} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i(1) - \hat{Y}_i(0)) \quad \text{where} \quad \hat{Y}(d) = \begin{cases} Y_i & \text{if } D_i = d \\ \frac{1}{q}\sum_{j \in \mathcal{J}_q(i)} Y_j & \text{if } D_i \neq d \end{cases} .$$

Properties of Matching

- **Note:** it is a type of nearest neighbor (NN) estimator.
- As q increases the variance goes down, but the bias goes up.
- Abadie and Imbens (2006,ECMA) study asymptotic properties of $\hat{\theta}_{\text{ate}}$ under a fixed number of matches (as $n \to \infty$):
  - **Consistency**: $\hat{\theta}_{\text{ate}}$ is consistent as $n \to \infty$ for fixed q.
  - **Bias**: is of order $O(n^{-1/k_c})$, where k_c is the dimension of the (cont.) covariates.
  - **Rate of convergence**: variance is of order $O(1/n)$. However, $\sqrt{n}$Bias $\to 0, C$, or ∞ if $k_c = 1$, $k_c = 2$, or $k_c > 2$, respectively. So, if $k_c > 2$, estimator is not $\sqrt{n}$-asympt. normal.
  - **Efficiency**: $\hat{\theta}_{\text{ate}}$ is generally not efficient. Even if the bias is low enough, the estimators are not efficient given a fixed number of matches.
  - **Bootstrap**: AI(08) show that the bootstrap is generally invalid for matching estimators due to nonsmoothness in the matching process.
  - **Subsampling**: valid for $k_c \leqslant 2$.

Propensity Score Matching

- Let $p(x) = P\{D_i = 1|X_i = x\}$ denote the propensity score.
- Unconfoundedness implies that
$$(Y(0), Y(1)) \perp\!\!\!\perp D \mid p(X) \, .$$
- Important observation due to Rosenbaum and Rubin (1983).
- To see this note that
$$\begin{aligned} P\{D = 1|Y(0), Y(1), p(X)\} &= E\left[\, E[D|Y(0), Y(1), P(X), X] \mid Y(0), Y(1), p(X) \,\right] \\ &= E\left[\, E[D|Y(0), Y(1), X] \mid Y(0), Y(1), p(X) \,\right] \\ &= E\left[\, E[D|X] \mid Y(0), Y(1), p(X) \,\right] \\ &= E\left[\, p(X) \mid Y(0), Y(1), p(X) \,\right] \\ &= p(X) \, , \end{aligned}$$
which is the same as $P\{D = 1|p(X)\}$.
- Lesson: all biases due to observable covariates can be removed by conditioning solely on the propensity score.

Propensity Score Matching II

- The Rosenbaum-Rubin result implies that

$$\theta_{\text{ate}} = E[E[Y|D=1, p(X)] - E[Y|D=0, p(X)]] \; .$$

- So: can use the Matching estimator matching on the propensity score only.

- This can be reformulated by noting that

$$E\left[\frac{DY}{p(X)}\right] = E\left[\frac{1}{p(X)} E\left[DY(1)|p(X)\right]\right] = E\left[\frac{1}{p(X)} E[D|p(X)]E[Y(1)|p(X)]\right] = E[Y(1)]$$

  and similarly

$$E\left[\frac{(1-D)Y}{1-p(X)}\right] = E[Y(0)] \; .$$

  Which in turn allows us to write

$$\theta_{\text{ate}} = E\left[\frac{[D_i - p(X_i)]Y_i}{p(X_i)(1-p(X_i))}\right] \; .$$

- **Propensity score weighting**:

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{[D_i - p(X_i)]Y_i}{p(X_i)(1-p(X_i))}\right) \; .$$

Propensity Score Matching III

- Propensity score is a scalar: AI imply that the bias term is of lower order than the variance term and matching leads to a $\sqrt{n}$-consistent, asymptotically normal estimator.

- **Problem**: the propensity score is an unknown function.

- Estimator based on the true propensity score has the same asymp. variance in AI.

- With estimated propensity scores, the asymptotic variance of matching estimators is more involved due to the **generated regressor**

  ...worked out in Hahn and Ridder (2013, ECMA): "The Asymptotic Variance of Semi-parametric Estimators with Generated Regressors".

References

[1] Alberto Abadie and Guido W Imbens. Large sample properties of matching estimators for average treatment effects. *econometrica*, 74(1):235–267, 2006.

[2] Federico A Bugni and Ivan A Canay. Testing continuity of a density via g-order statistics in the regression discontinuity design. *Journal of Econometrics*, 221(1):138–159, 2021.

[3] Sebastian Calonico, Matias D Cattaneo, and Rocio Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014.

[4] Ivan Alexis Canay and Vishal Kamat. Approximate permutation tests and induced order statistics in the regression discontinuity design. *The Review of Economic Studies*, 85(3):1577–1608, 2018.

[5] Jinyong Hahn and Geert Ridder. Asymptotic variance of semiparametric estimators with generated regressors. *Econometrica*, 81(1):315–340, 2013.

[6] Guido Imbens and Karthik Kalyanaraman. Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, (3):933–959, 2012.

[7] David S. Lee. Randomized experiments from non-random selection in u.s. house elections. *Journal of Econometrics*, 142(2):675 – 697, 2008.

[8] Justin McCrary. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698 – 714, 2008.

THE END