Experimental Analysis of Public Policies

Sessions 7–8: three-hour teaching version

Thresholds, Synthetic Controls, and Evidence

Etienne Dagorn
University of Lille

Roadmap

Sharp RDD: A Rule and a Local Counterfactual

Local Estimation and Credibility Checks

Fuzzy RDD: Reconnecting the Threshold to IV

Synthetic Control: Constructing an Untreated Path

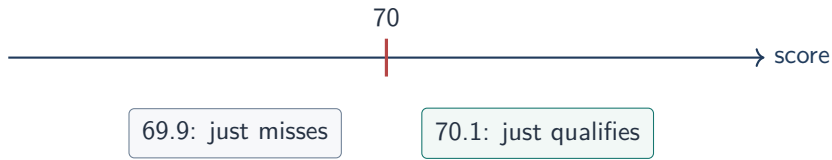
Heterogeneity, Replication, and the Final Method Map

Our Three Hours: From a Rule to an Evidence Audit

Clock	Question or activity	Minutes
0–10	Opening question and links to earlier designs	10
10–45	Sharp RDD: rule, missing counterfactual, equations	35
45–65	Bandwidth, diagnostics, and an 8-minute R activity	20
65–95	Fuzzy RDD, IV, and class-size evidence	30
95–105	Break	10
105–145	Synthetic control: construct and challenge a comparison	40
145–170	Heterogeneity and a concrete replication case	25
170–180	Method map and exit ticket	10

A Scholarship for Pupils Scoring at Least 70

The ministry asks: **does the scholarship increase university enrolment?**



Opening poll: two minutes

Would you compare all recipients with all non-recipients, or focus on pupils near 70? What might still make the local comparison misleading?

Roadmap

Sharp RDD: A Rule and a Local Counterfactual

Local Estimation and Credibility Checks

Fuzzy RDD: Reconnecting the Threshold to IV

Synthetic Control: Constructing an Untreated Path

Heterogeneity, Replication, and the Final Method Map

Audit the Rule Before Choosing an Estimator

Our scholarship example

- Running variable: exam score X_i.
- Cutoff: $c = 70$.
- Treatment: scholarship received D_i.
- Outcome: university enrolment Y_i.

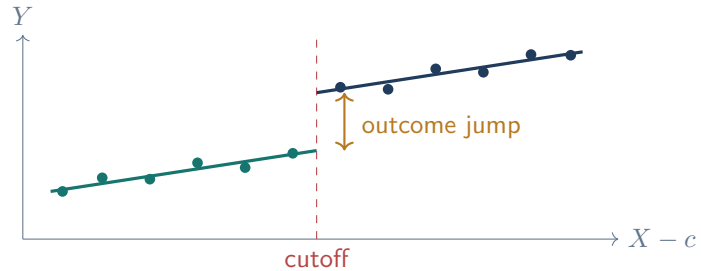
Sharp assignment rule

$$D_i = \mathbf{1}\{X_i \geq c\}.$$

Everyone qualifying receives the scholarship; nobody below the cutoff does.

Notation: here X is the running variable; other baseline covariates are W.

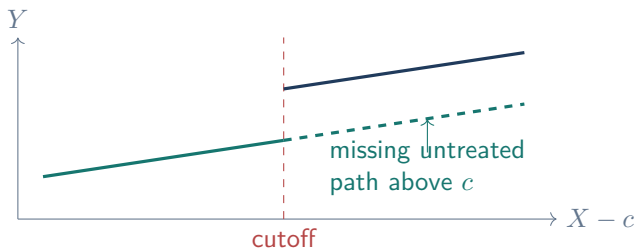
Read the Jump Before Reading a Regression Table



The running variable can predict outcomes smoothly. The design uses the **discontinuity at the rule**.

Illustrative schema; not estimates from a paper.

Draw the Missing Counterfactual



Illustrative schema; not estimates from a paper.

Just below c, we observe pupils without the scholarship. Can their outcome path approximate what just-qualifying pupils would have experienced **without it**?

The Assumption: No Counterfactual Jump at 70

Continuity in plain language

Without the scholarship rule, expected outcomes would change smoothly through the cutoff. No other policy, sorting process or institutional change creates a jump there.

Formally, for each treatment state $d \in \{0, 1\}$:

$$\lim_{x \downarrow c} E[Y_i(d) \mid X_i = x] = \lim_{x \uparrow c} E[Y_i(d) \mid X_i = x].$$

- $x \downarrow c$: approach from above; $x \uparrow c$: from below.
- We need smooth **potential outcomes**, not identical observed outcomes.

From the Observed Jump to the Causal Target

Observed discontinuity:

$$\Delta_Y(c) = \lim_{x \downarrow c} E[Y_i \mid X_i = x] - \lim_{x \uparrow c} E[Y_i \mid X_i = x].$$

Causal cutoff estimand:

$$\tau_{SRD}(c) = E[Y_i(1) - Y_i(0) \mid X_i = c].$$

Identification statement

Under the sharp rule and continuity, $\Delta_Y(c) = \tau_{SRD}(c)$. The jump identifies an effect **at the cutoff**, not an ATE for all pupils.

Why Continuity Turns a Jump into an Effect

The sharp rule tells us which potential outcome is observed on each side:

$$\begin{aligned}\Delta_Y(c) &= \lim_{x \downarrow c} E[Y_i(1) \mid X_i = x] - \lim_{x \uparrow c} E[Y_i(0) \mid X_i = x] \\ &= E[Y_i(1) \mid X_i = c] - E[Y_i(0) \mid X_i = c] \\ &= \tau_{SRD}(c).\end{aligned}$$

- First line: assignment and consistency.
- Second line: continuity of both conditional potential-outcome means.
- The fitted jump estimates these limits using observations near c.

Check: which step fails if another grant begins at exactly the same score?

A Causal DAG Clarifies the Threat; Continuity Does the Work



- Scores predict both scholarship receipt and outcomes.
- RDD isolates the rule's jump from the score's smooth relationship with Y.

What the DAG cannot establish

At a fixed score, sharp treatment is deterministic: ordinary adjustment has no treatment overlap. The local continuity argument identifies the effect; this DAG alone does not.

Roadmap

Sharp RDD: A Rule and a Local Counterfactual

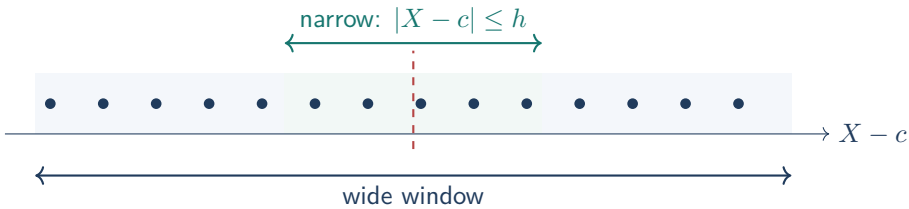
Local Estimation and Credibility Checks

Fuzzy RDD: Reconnecting the Threshold to IV

Synthetic Control: Constructing an Untreated Path

Heterogeneity, Replication, and the Final Method Map

Bandwidth: How Local Should the Comparison Be?



Narrow: closer comparisons, fewer observations, less precision.

Wide: more observations, but greater risk from fitting the wrong curve.

The Equation Behind Two Local Lines

Keep observations with $|X_i - c| \leq h$, then fit:

$$Y_i = \alpha + \tau D_i + \beta_1(X_i - c) + \beta_2 D_i(X_i - c) + u_i.$$

Just below the cutoff

$$\widehat{Y} = \widehat{\alpha} + \widehat{\beta}_1(X - c)$$

Just above the cutoff

$$\widehat{Y} = (\widehat{\alpha} + \widehat{\tau}) + (\widehat{\beta}_1 + \widehat{\beta}_2)(X - c)$$

Interpret the coefficient

At $X = c$, the estimated vertical gap is $\widehat{\tau}$. Allow separate slopes; avoid high-order global polynomials.

A Local Estimate Has Both Noise and Approximation Bias

$$\widehat{\tau}(h) - \tau_{SRD}(c) \;=\; \text{local approximation bias}(h) + \text{sampling noise}.$$

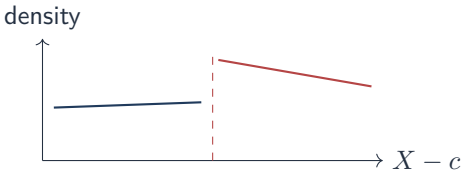
- Wider windows can reduce noise while increasing curvature bias.
- A conventional OLS interval may ignore bias that matters for coverage.
- Report bandwidth, kernel, observations on each side and robust bias-corrected intervals.

The target stays local

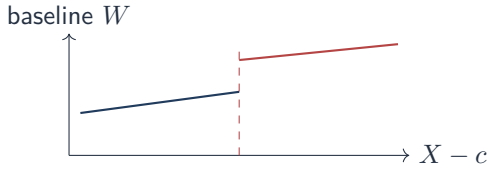
Changing h usually changes the approximation and precision, not the intended cutoff estimand. Do not choose the window for statistical significance.

Cattaneo, Idrobo and Titiunik (2020), Foundations, estimation and inference.

Diagnostics: What Would Make You Less Confident?



Bunching just above c?



A pre-treatment variable jumps?

Name a threat, then a check

Could scores be appealed? Does another benefit start at 70? Choose a density, covariate, placebo-cutoff or institutional check and explain its purpose.

Smooth covariates and density **support** the design; they do not prove continuity.

Illustrative schema; not estimates from a paper.

R Mini-Lab: Does the Local Estimate Depend on the Window?

Simulated outcome-score data: cutoff centred at zero.

```
source("r_exercises/week_07_rdd.R")
h <- c(10, 20, 35)
data.frame(bandwidth = h,
  jump = sapply(h, function(b) estimate_jump("Y", b)))
estimate_jump("W", 20)  # baseline covariate diagnostic
```

Eight minutes: compute and interpret

Does the sign or magnitude change? Is the jump in W small? Write one sentence about the local effect and one about what this diagnostic cannot establish.

These are point estimates; a full empirical analysis also needs appropriate RDD intervals.

Roadmap

Sharp RDD: A Rule and a Local Counterfactual

Local Estimation and Credibility Checks

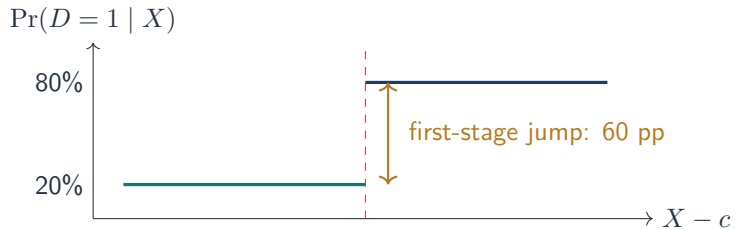
Fuzzy RDD: Reconnecting the Threshold to IV

Synthetic Control: Constructing an Untreated Path

Heterogeneity, Replication, and the Final Method Map

Eligibility Is an Offer; Receipt Can Remain a Choice

$$Z_i = \mathbf{1}\{X_i \geq c\}, \qquad D_i \neq Z_i \text{ for some pupils.}$$



The threshold indicator Z is a **candidate instrument** for receipt D.

Illustrative schema; not estimates from a paper.

A Local Wald Ratio: The Same Logic as Session 4

For any variable V, let $\Delta_V(c)$ be its jump from just below to just above c. The local Wald ratio identifies the target **under the local IV assumptions**:

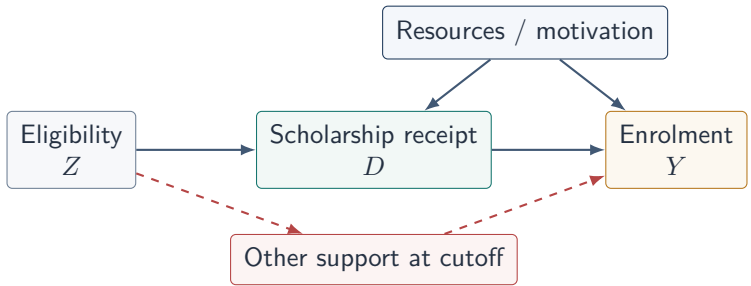
$$\underbrace{\tau_{FRD}(c)}_{\text{local complier effect}} = \frac{\overbrace{\Delta_Y(c)}^{\text{outcome jump}}}{\underbrace{\Delta_D(c)}_{\text{receipt jump}}}.$$

Two-minute calculation: illustrative scholarship numbers

Enrolment jumps by 6 percentage points; receipt jumps by 60 percentage points. Calculate the ratio. Whose effect is it, under the assumptions?

$$\widehat{\tau}_{FRD} = \frac{0.06}{0.60} = 0.10 \quad \Rightarrow \quad \text{10 pp for } \textbf{cutoff compliers}.$$

The Familiar IV DAG, Now with a Local Assignment Rule



The red bypass would violate exclusion **for the effect of scholarship receipt**. **Local qualification:** this graph suppresses smooth score-related paths; continuity around c remains essential.

What Makes the Local Ratio Causal?

Condition	Question at the cutoff
Relevance	Does crossing c change receipt probability?
Continuity	Would potential outcomes and potential receipt behaviour vary smoothly through c?
Exclusion	Can eligibility affect enrolment without changing scholarship receipt?
Monotonicity	Could becoming eligible make anyone *less* likely to receive it?

IV connection

Session 4: ITT_Y / ITT_D for invitation compliers.
Here: $\Delta_Y(c)/\Delta_D(c)$ for **eligibility compliers near the cutoff**.

Why the Ratio Describes Cutoff Compliers

Let $D_i(z)$ be receipt if eligibility were z. For binary receipt, define

$$C_i = 1\{D_i(1) > D_i(0)\}, \qquad D_i(1) \geq D_i(0).$$

Under local continuity, exclusion, monotonicity and a positive first stage:

$$\begin{aligned}\Delta_D(c) &= \Pr(C_i = 1 \mid X_i = c),\\ \Delta_Y(c) &= \Pr(C_i = 1 \mid X_i = c)\ \underbrace{E[Y_i(1) - Y_i(0) \mid C_i = 1, X_i = c]}_{\tau_{FRD}(c)}.\end{aligned}$$

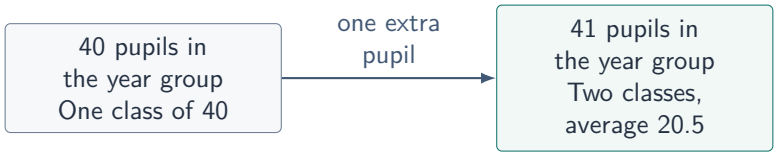
The complier share cancels. Always-takers and never-takers do not change receipt.

A small denominator is an inference problem

A weak jump can make the ratio unstable; report both jumps and their uncertainty. A strong first stage still does not establish exclusion.

Angrist and Lavy: Can Class Size Improve Achievement?

Problem: small classes may be allocated to pupils already facing difficulties. **Institution:** Israeli primary schools use Maimonides' rule, with a class-size cap of 40.

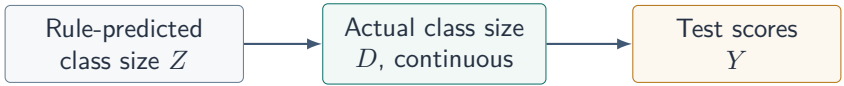


Predict before reading the result

If smaller classes improve learning, what pattern should appear in test scores when enrolment crosses this rule?

Angrist and Lavy (1999), QJE, pp. 535 and 540–542. Rule schema, not observed class sizes.

What the Class-Size Paper Teaches About IV and Locality



- The authors use rule-induced variation to instrument actual class size.
- They find gains from smaller classes for fourth and fifth graders, with little evidence for third graders.
- **Interpretation:** this is a continuous-treatment IV application; it is not literally the binary-scholarship complier example.

Ask: does the enrolment rule change other educational inputs or pupil composition?

Angrist and Lavy (1999), QJE, pp. 533–536.

Ten-Minute Break

Return at minute 105

We have a local comparison when an institution supplies a threshold.
What if a policy affects one whole region, with no lottery or useful cutoff?

Roadmap

Sharp RDD: A Rule and a Local Counterfactual

Local Estimation and Credibility Checks

Fuzzy RDD: Reconnecting the Threshold to IV

Synthetic Control: Constructing an Untreated Path

Heterogeneity, Replication, and the Final Method Map

California: Did a Tobacco-Control Programme Reduce Cigarette Sales?

Proposition 99: approved in 1988; the programme combined a cigarette-tax increase with other tobacco-control measures.

- One treated state; the post-treatment outcome is observed.
- The missing counterfactual is California's sales path **without the programme**.
- California and the rest of the US already had different pre-policy trends.

Two-minute donor-pool discussion

Which states would you exclude if they adopted a similar programme or were affected by California's policy?

Abadie, Diamond and Hainmueller (2010), JASA, pp. 498–499.

The SCM Target and Its Missing Potential Outcome

Let T_0 denote the **first treated period**. For unit 1 at $t \geq T_0$:

$$\tau_{1t} = Y_{1t}(1) - Y_{1t}(0).$$

Here treatment means the policy path beginning at T_0; state 0 means no policy.

- Observe $Y_{1t} = Y_{1t}(1)$ after adoption; $Y_{1t}(0)$ is missing.
- Untreated donors supply observed $Y_{jt}(0)$, if unaffected by the policy.
- No anticipation makes the treated unit's pre-policy outcomes untreated too.

Population and horizon

This target concerns **one aggregate unit at a particular date**.
It is not an ATE over donor states or a population of individuals.

The Equation: A Weighted Untreated California

$$\widehat{Y}_{1t}(0) = \sum_{j=2}^{J+1} \widehat{w}_j Y_{jt}, \qquad \widehat{w}_j \geq 0, \qquad \sum_{j=2}^{J+1} \widehat{w}_j = 1.$$

- Unit 1 receives the policy; units $2, \ldots, J+1$ are untreated donors.
- Choose weights using **pre-treatment** outcomes and predictors.
- Apply those fixed weights to the donors' post-treatment outcomes.

The identifying claim

Absent the programme, the weighted donor path would continue to approximate California's untreated path after treatment.

How Are the Donor Weights Chosen?

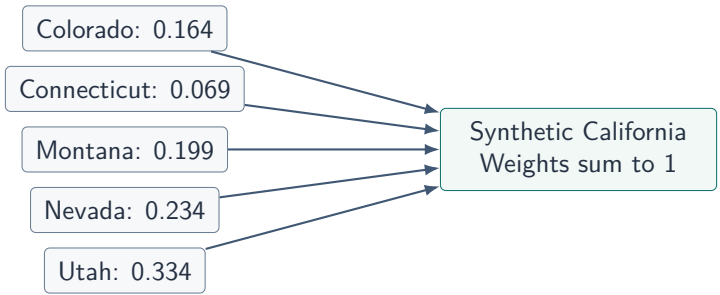
Let P_j collect selected **pre-treatment** outcomes and predictors for unit j.

$$\widehat{w} = \arg \min_{\substack{w_j \geq 0 \\ \sum_j w_j = 1}} \left(P_1 - \sum_j w_j P_j \right)' V \left(P_1 - \sum_j w_j P_j \right).$$

- V is diagonal with non-negative predictor-importance weights, selected using pre-policy information only.
- Non-negative weights summing to one restrict the synthetic predictor vector to the donors' convex hull.
- Inspect what is matched and what is not; a numerical solution need not fit well.

Freeze the design before interpreting post-policy gaps.
Use a pre-policy validation period where feasible to assess prediction beyond the fitting window.

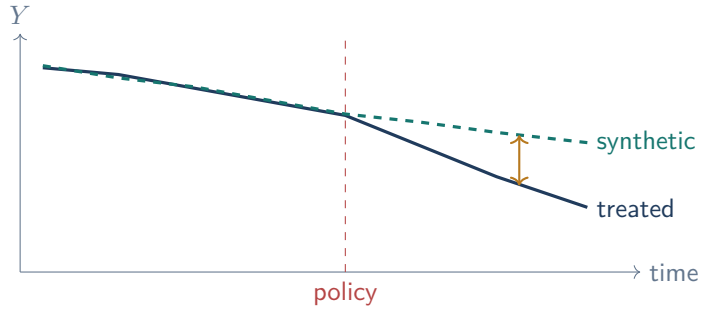
The Published Weights Make the Comparison Visible



The other eligible donor states receive zero weight. These weights construct a comparison, **not an effect**.

Abadie, Diamond and Hainmueller (2010), Table 2. This is a weighting schema, not a DAG.

Read the Paths: Fit Before the Policy, Gap Afterwards



- Good pre-treatment fit makes the comparison plausible.
- It cannot rule out a new shock affecting only the treated unit afterwards.

Illustrative schema; not estimates from a paper.

The Gap: A Treated-Unit Effect at a Particular Time

$$\widehat{\tau}_{1t} = Y_{1t} - \widehat{Y}_{1t}(0), \qquad t \geq T_0.$$

Published application: by 2000, estimated annual cigarette sales in California were about **26 packs per capita lower** than its synthetic comparison.

Three-minute interpretation

Is this a before–after change, the effect in every US state, or an estimated California counterfactual gap? Would lower recorded sales necessarily mean an identical reduction in consumption?

Abadie, Diamond and Hainmueller (2010), abstract and pp. 500–501.

A Fitted Gap Equals an Effect Plus Counterfactual Error

With no spillovers to donors and $t \geq T_0$:

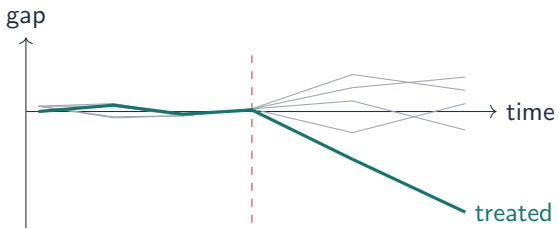
$$\begin{aligned}\widehat{\tau}_{1t} &= Y_{1t} - \sum_j \widehat{w}_j Y_{jt} \\ &= \tau_{1t} + \underbrace{\left[Y_{1t}(0) - \sum_j \widehat{w}_j Y_{jt}(0) \right]}_{\text{unobserved post-policy prediction error}} .\end{aligned}$$

What must carry forward

The donor combination must approximate the treated unit's **untreated** post-policy path. Pre-fit alone does not bound the missing prediction error.

Check: would a new factory closure unique to the treated state appear in this bracket or necessarily in the policy effect?

Placebos: Is the Treated Gap Unusual?

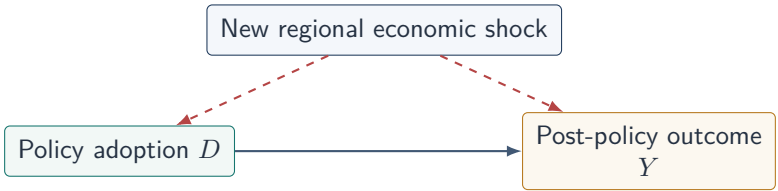


Pretend each donor was treated at the **same policy date**, and build its own synthetic control.
Compare gaps only with attention to **pre-fit quality** and comparable units.

Also remove influential donors and try an earlier placebo date. Each check should address a named threat.

Illustrative schema; not estimates from a paper.

A New Regional Shock Can Survive Perfect Pre-Fit



Evidence audit: three minutes

A factory closure influences policy adoption and the treated region's outcome after the pre-period. Would matching the earlier path remove this confounding? What information or comparison would help?

Roadmap

Sharp RDD: A Rule and a Local Counterfactual

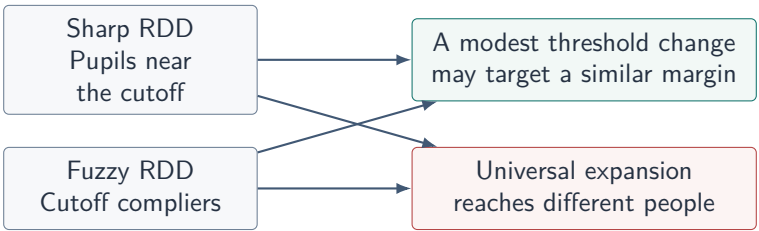
Local Estimation and Credibility Checks

Fuzzy RDD: Reconnecting the Threshold to IV

Synthetic Control: Constructing an Untreated Path

Heterogeneity, Replication, and the Final Method Map

Local Evidence: Which Policy Expansion Does It Inform?



Transport question: are the population, treatment quality, institutions and scale comparable?

Heterogeneity: Put the Difference into the Equation

Let $H_i = 1$ denote high baseline need, measured **before treatment**. Target: $\tau(h) = E[Y_i(1) - Y_i(0) \mid H_i = h]$.

$$Y_i = \alpha + \tau D_i + \theta(D_i \times H_i) + \gamma H_i + u_i.$$

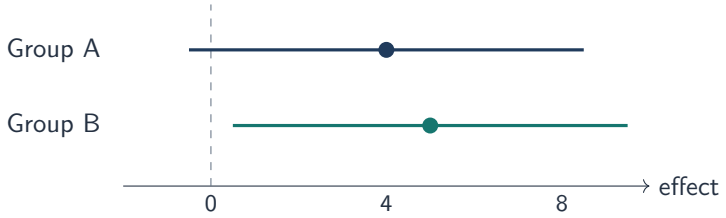
Lower need: treatment contrast τ.
Higher need: treatment contrast $\tau + \theta$.

Difference in contrasts: θ.
Test $H_0 : \theta = 0$, and report its uncertainty.

Causal qualification

The design must identify treatment effects within the relevant baseline groups. Choose a few motivated groups before seeing results; label exploratory searches.

One Significant Estimate and One Insignificant Estimate



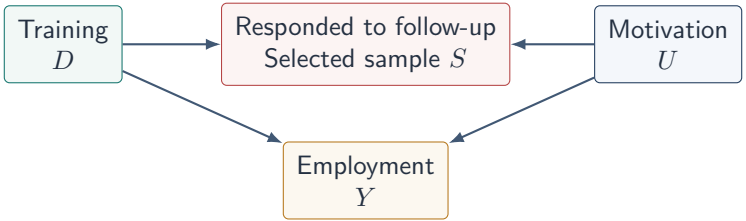
Two-minute vote

One interval includes zero and the other does not. Have we established that the effects differ?

No: estimate the difference and its confidence interval directly. Overlapping intervals alone also do not determine the test of the difference.

Illustrative schema; not estimates from a paper.

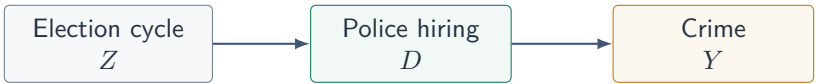
Do Not Define Subgroups Using a Post-Treatment Response



Conditioning on survey response can select on a **collider**: $D \to S \leftarrow U$. Treatment and motivation can become associated among respondents. **Prefer:** baseline need or another pre-treatment, substantively motivated dimension.

Replication Can Change What the Evidence Supports

Question: do more police reduce crime?



- Levitt's electoral-cycle IV aimed to avoid reverse causality in police hiring.
- McCrary's reanalysis identified implementation and election-dating problems; corrected pooled IV estimates shrank and lost statistical significance.
- Levitt acknowledged errors and presented additional evidence in his reply.

McCrary (2002), AER, 92(4), 1236–1243.

Levitt (2002), AER, 92(4), 1244–1250 (reply).

A Replication Audit Starts with a Specific Threat

Three-minute pair exercise

For the police study or synthetic California, name one check addressing each question:

1. Did code and data implement the claimed comparison?
2. Does the comparison identify the intended counterfactual?
3. Is the estimate precise enough for the policy claim?

A useful robustness check states: the threat, the changed choice, and whether the estimand changes.

Successful rerunning establishes reproducibility of an output. Defending the design requires more evidence.

Method Map: Assignment, Instruments, and Time

Method	Variation	Causal assumption	Target
RCT	Random assignment Z	Assignment independent of potential outcomes	Assignment effect; treatment effect with full compliance
IV	Z-induced receipt D	Independence, exclusion, relevance, monotonicity	LATE for compliers with binary Z, D
DiD	Policy timing and untreated changes	Parallel untreated trends, no anticipation	ATT for treated groups and periods

Method Map: Dynamics, Thresholds, and Donors

Method	Variation	Causal assumption	Target
Staggered DiD	Cohort timing	Appropriate untreated trends and clean comparison groups	$ATT(g,t)$, then explicit aggregation
Sharp / fuzzy RDD	Threshold rule	Continuity; fuzzy adds local IV restrictions	Cutoff effect / cutoff-complier LATE
SCM	Treated unit versus fixed weighted donors	Donors reproduce its untreated path; no contaminating shocks/spillovers	Treated-unit effect at time t

Common task: defend the observed comparison as a substitute for the missing counterfactual.

Exit Ticket: Defend a Comparison in Four Sentences

Choose **scholarships near 70**, **class-size rules**, or **synthetic California**.

Four minutes to write; two minutes to debrief

1. The identifying variation comes from _____.
2. The causal interpretation requires _____.
3. The estimate describes _____.
4. I would trust the conclusion less if _____.

Course reflex: variation → assumption → estimand.
Then assess estimation, uncertainty, and policy relevance.

Appendix: Lee (2008), Party Incumbency in Close Elections

Question: does winning today improve a party's result at the next election?

- US House election data, 1946–1998; compare barely winning with barely losing.
- Running variable: current Democratic vote margin; cutoff: zero.
- Treatment: Democratic victory; outcome: next-election Democratic vote share.
- Finding: a positive party incumbency advantage at close races.

Defend the local comparison

Why does comparing all winners with all losers mix victory with party strength?
What would precise control of the final vote margin do to the RD argument?

Lee (2008), Journal of Econometrics, 142(2), 675–697.

Optional Practice After Class

- RDD practice: `homework/session_07_homework.md`.
- R starter: `r_exercises/week_07_rdd.R`.
- Optional synthesis starter: `r_exercises/week_08_scm_hte_replication.R`.

> **One manageable deliverable**
>
> One RDD graph and bandwidth table, plus a short evidence audit naming the counterfactual, assumption, local estimand and main remaining doubt.

Paths start at the course root. The optional synthesis script extends the same reasoning to a simulated regional panel.

Appendix: Fuzzy RDD as Local 2SLS

Within a window around c, define $R_i = X_i - c$, $Z_i = \mathbf{1}\{R_i \geq 0\}$. An illustrative local-linear specification is:

$$D_i = \pi_0 + \pi_1 Z_i + \pi_2 R_i + \pi_3 Z_i R_i + \eta_i,$$

$$Y_i = \alpha + \tau \widehat{D}_i + \beta_1 R_i + \beta_2 Z_i R_i + \varepsilon_i.$$

Instrument D with Z, including R and ZR as the same controls in both stages. With the same window and weights, the just-identified coefficient is the ratio of fitted outcome and receipt jumps.

Implementation

Use an IV/RDD command for appropriate inference; naive second-stage OLS standard errors are incorrect.

Appendix: What a Donut Check Can and Cannot Do

- Exclude observations extremely close to the cutoff when heaping or appeals are a concern.
- The remaining observations require extrapolation back to the cutoff.
- Removing a suspicious region does not automatically repair selection caused by manipulation.
- Explain the changed sample and which effect the procedure is intended to recover.

Other optional extensions: multiple cutoffs, geographic boundaries, and regression kink designs.

Appendix: Donor Support and Sensitivity

- Non-negative weights summing to one keep the synthetic unit within donor support.
- Poor pre-fit can signal that no available mixture resembles the treated unit.
- Remove influential donors, inspect policy spillovers, and vary justified pre-period choices.
- Show whether a large post-gap is already preceded by large pre-period gaps.

DiD connection: both designs ask whether untreated units reproduce a treated unit's missing untreated path. SCM makes its weighted comparison explicit.

Appendix: One Model Explaining Why Pre-Fit Can Help

An illustrative untreated-outcome model is

$$Y_{jt}(0) = \delta_t + \lambda_t' \mu_j + \varepsilon_{jt}.$$

λ_t: common time-varying forces; μ_j: unit-specific exposure to them.

$$Y_{1t}(0) - \sum_j w_j Y_{jt}(0) = \lambda_t' \left(\mu_1 - \sum_j w_j \mu_j \right) + \varepsilon_{1t} - \sum_j w_j \varepsilon_{jt}.$$

- If donors reproduce μ_1, common forces cancel at later dates too.
- A long, informative pre-period can help reveal those exposures.
- Noise, weak support or a new treated-specific shock can defeat this argument.

Appendix: Compare Post-Gaps Relative to Pre-Fit

For unit j, let e_{jt} be its actual-minus-synthetic gap. One statistic is

$$R_j = \frac{\sqrt{\text{mean}_{t \geq T_0}\, e_{jt}^2}}{\sqrt{\text{mean}_{t < T_0}\, e_{jt}^2}}.$$

- Refit each placebo using the same rule and a suitable untreated donor pool.
- Declare any pre-fit exclusion rule before comparing post-treatment gaps.
- Near-zero pre-fit error can make a ratio unstable; inspect the paths as well.

Appendix: When Can a Placebo Rank Be a Test?

For M donor placebos, one upper-tail rank statistic is

$$p_{rank} = \frac{1 + \#\{j : R_j \geq R_1\}}{M + 1}.$$

- An exact randomisation interpretation needs exchangeable assignment among the comparison units, a sharp no-effect null and symmetric refitting.
- Donor eligibility and any pre-fit filtering must respect that assignment argument.
- With few placebos, ranks are coarse: the smallest value is $1/(M + 1)$.

Interpretation

Without that assignment argument, report how unusual the treated gap is; the rank is not automatically a randomised-test p-value.

Appendix: What If Scores Are Discrete or Missing Near the Cutoff?

- A continuity argument uses outcomes approaching c from both sides.
- With a few distinct scores, more pupils at each score do not fill the gaps.
- A donut exclusion also leaves a hole that must be extrapolated across.
- State the extra smoothness or local-randomisation assumptions needed.

Interpretation check

If only scores 60, 70 and 80 are possible, does a million observations make the comparison infinitesimally local?

References: Threshold and IV Applications

- Lee, D. S. (2008). "Randomized Experiments from Non-random Selection in U.S. House Elections." *Journal of Econometrics*, 142(2), 675–697.
  Author-hosted article.
- Angrist, J. D., and Lavy, V. (1999). "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *QJE*, 114(2), 533–575.
  University-hosted article.
- Cattaneo, M. D., Idrobo, N., and Titiunik, R. (2020). *A Practical Introduction to Regression Discontinuity Designs: Foundations*.

References: Synthetic Control and Replication

- Abadie, A., Diamond, A., and Hainmueller, J. (2010). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *JASA*, 105(490), 493–505.
  Author-hosted article.
- McCrary, J. (2002). "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Comment." *AER*, 92(4), 1236–1243.
  Author-hosted comment.
- Levitt, S. D. (2002). "Using Electoral Cycles in Police Hiring to Estimate the Effects of Police on Crime: Reply." *AER*, 92(4), 1244–1250.
  Author-hosted reply.