Econometrics of Policy Evaluation

Q&A Session

Cristian Huse

Relation between SDO and ATE

Background

The **Simple Difference in Outcomes (SDO)** is related to the **Average Treatment Effect (ATE)**, but they are only equal under specific conditions. In the general case, SDO can be decomposed into **ATE** plus a selection bias term.

$$SDO = ATE + \text{Selection Bias Term } (+ \text{ HTE Term})$$

where:

- **SDO** $= E[Y \mid D = 1] - E[Y \mid D = 0]$ (**observed difference in means**),
- **ATE** $= E[Y(1)] - E[Y(0)]$ (**true causal effect of treatment**),
- **Selection Bias Term** $= E[Y(0) \mid D = 1] - E[Y(0) \mid D = 0]$, which captures the **pre-existing differences** between treated and control groups.
- **HTE Term** assumed zero for simplicity

Relation between SDO and ATE cont'd

- The **Simple Difference in Outcomes (SDO)** method is the most basic way to estimate the **causal effect** of a treatment.
- It compares the **average outcome** between the **treated and untreated groups**:

$$SDO = E[Y|D = 1] - E[Y|D = 0]$$

where:

- Y = Outcome variable (e.g., income, test scores).
- D = Treatment indicator ($D = 1$ if treated, $D = 0$ if not).
- $E[Y|D = 1]$ = Average outcome for the **treated** group.
- $E[Y|D = 0]$ = Average outcome for the **control** group.

How SDO Relates to ATE (Average Treatment Effect)

The **Average Treatment Effect (ATE)** is the true causal effect:

$$ATE = E[Y(1)] - E[Y(0)]$$

where:

- $Y(1)$ = Outcome **if treated**.
- $Y(0)$ = Outcome **if untreated**.

For **SDO to be a valid estimator of ATE**, we need that the **treated and untreated groups must be identical in expectation** — i.e., treatment assignment must be **as good as random**.

When is SDO Biased?

SDO is **biased** if treatment assignment is **not random**. This happens when:

1. **Selection Bias (Confounding)**
   - Individuals who choose treatment are **not the same as those who don't**.
   - Example: People who attend a **health program** may already care more about their health.
2. **Reverse Causality**
   - The **outcome may affect treatment**.
   - Example: Higher-income individuals may **be more likely to afford** private education.
3. **Omitted Variables**
   - There may be **unobserved factors** affecting both treatment and outcome.
   - Example: If **ability** influences both **education choices and wages**, then the effect of education on wages is **overstated**.

When is SDO Valid?

When Treatment is Randomly Assigned (RCTs)

If treatment is **randomly assigned**, then:

$$E[Y(0)|D = 1] = E[Y(0)|D = 0]$$

This means the **control group is a valid counterfactual** for the treated group. In this case:

$$SDO = ATE$$

Example:

- A clinical trial randomly assigns patients to a **new drug (D=1)** or **placebo (D=0)**.
- Since treatment is **randomly assigned**, both groups have **identical expected outcomes** before treatment.
- The SDO correctly estimates the **causal effect** of the drug.

Conclusion: SDO is unbiased **when treatment is random**.

When Does SDO Lead to Biased Estimates?

SDO **fails** when treatment assignment is **not random**, leading to **selection bias**.

Selection Bias: Treated and Control Groups Differ in Pre-Treatment Characteristics

If individuals **self-select into treatment** (e.g., based on motivation, wealth, education), then:

$$E[Y(0)|D = 1] \neq E[Y(0)|D = 0]$$

Example: Job Training Program

- Suppose people **choose** to attend a job training program.
- Those who **enroll** are likely **more motivated** or already have better skills.
- This means their expected earnings **would have been higher even without training**.

When Does SDO Lead to Biased Estimates? cont'd

Bias in SDO

The difference in outcomes now captures both:

$$SDO = ATE + E[Y(0)|D = 1] - E[Y(0)|D = 0]$$

- The second term is **selection bias** (pre-existing differences).
- If motivated people **self-select**, then $E[Y(0)|D = 1] > E[Y(0)|D = 0]$, so SDO **overestimates the treatment effect**.

Solution: Use **Matching, Difference-in-Differences (DiD), Instrumental Variables (IV), or RCTs** to account for selection bias.

Reverse Causality

If the **outcome affects treatment**, then SDO is **misleading**.

Example: Health Insurance and Health

- Suppose we compare **health outcomes** of people **with vs. without health insurance**.
- SDO may suggest insurance **causes better health**.
- But what if **sick people are more likely to buy insurance**? (Reverse causality)
- Then, **health problems influence treatment**, rather than treatment influencing health.

Solution: Use **IV or Panel Data Methods** to correct for reverse causality.

Omitted Variable Bias (Confounding)

If treatment and outcome are correlated due to a **third variable (confounder)**, then:

$$E[Y|D] \neq E[Y|D, X]$$

Example: Education and Earnings

- Suppose we compare the earnings of **college graduates (D=1)** vs. **non-graduates (D=0)**.
- But **family background, intelligence, or social networks** affect both **education and earnings**.
- The difference in earnings reflects both **education effects and pre-existing differences**.

Bias in SDO

SDO **confounds the true treatment effect with the omitted variable's effect**.

Solution: Use **Regression, Fixed Effects, or IV to control for confounders**.

Take-aways

1. SDO = ATE only if treatment is random.
2. Selection bias and omitted variables create endogeneity, making SDO misleading.
3. Solutions include Matching, DiD, Instrumental Variables (IV), and Randomized Control Trials (RCTs).

Summary

Issue	**Is SDO Unbiased?**	**Is SDO Biased?**	**Solution**
Random Assign-ment	Yes (RCTs)	No	No correction needed
Selection Bias	No	If treated and control differ before treatment	Matching, DiD, IV
Omitted Variables	No	If confounders affect treatment & outcome	Regression, Fixed Effects, IV
Reverse Causality	No	If outcome influences treatment	IV, Panel Data

Understanding Full Compliance in Different Methods

- **Full compliance** means that everyone assigned to the treatment group actually receives the treatment, and everyone in the control group does not receive it.

Randomized Assignment (RCT)

Assumes full compliance (in the basic setup)

- **How it works:** Individuals are **randomly assigned** to treatment ($D = 1$) or control ($D = 0$).
- If compliance is **perfect**, the **Average Treatment Effect (ATE)** can be estimated simply as:

$$E[Y(1) - Y(0)]$$

- **With non-compliance:**
  - If some assigned to treatment do **not take it**, or if some in the control group **take the treatment**, then the simple difference is **biased**.
  - **Solution:** Use **Intention-to-Treat (ITT)** or **Instrumental Variables (IV)** to estimate **Local Average Treatment Effect (LATE)**.

Randomized Assignment (RCT)

Example:

- **Full compliance:** All patients assigned to a new drug actually take it, and no control patient takes the drug.
- **Non-compliance:** Some assigned to the new drug **don't take it**, and some control patients **obtain the drug anyway**.

Regression Discontinuity Design (RDD)

Full compliance is not necessary

- **RDD does not assume full compliance** because some individuals near the cutoff may **not follow the assignment rule**.
- **How it works:** Individuals receive treatment **if they pass a threshold** (e.g., a test score of 50+ leads to a scholarship).
- **With full compliance**, we estimate the **treatment effect for everyone above the cutoff**:

$$E[Y(1)|X \geq c] - E[Y(0)|X < c]$$

- **With non-compliance**, some below the cutoff **still get treated**, or some above the cutoff **refuse treatment**.
  - **Solution:** Use **Fuzzy RDD**, which relies on **Instrumental Variables (IV)** to estimate **LATE**.

Regression Discontinuity Design (RDD)

Example:

- **Full compliance:** Only students scoring 50+ receive the scholarship, and all below 50 do not.
- **Non-compliance:** Some students scoring 48 still get a scholarship, or some scoring 52 reject it.

Difference-in-Differences (DiD)

Assumes full compliance for standard DiD, but can be adapted for non-compliance

- **How it works:** Compares **before-after differences** in **treated and untreated** groups.
- **With full compliance**, the DiD estimate is:

$$(E[Y_t^T] - E[Y_{t-1}^T]) - (E[Y_t^C] - E[Y_{t-1}^C])$$

  - The effect is valid if **treated units actually receive the treatment** and control units do not.
- **With non-compliance**, some in the treatment group **don't receive treatment**, or some in the control group **somehow get treated**.
  - **Solution:** Use **Instrumental Variables (IV)** or **treatment heterogeneity models** to correct for non-compliance.

Difference-in-Differences (DiD)

Example:

- **Full compliance:** A city increases the minimum wage in 2020, and all firms in the city follow it.
- **Non-compliance:** Some firms **do not increase wages**, or firms outside the city **voluntarily raise wages**.

Matching (Propensity Score Matching - PSM)

Requires full compliance for valid estimates

- **How it works:** Matches treated and control individuals with similar **observable characteristics**.
- **Full compliance is required** because matching assumes that:
  1. **Treated individuals actually received treatment.**
  2. **Control individuals never received treatment.**
- If compliance is **imperfect**, then **treatment effects are biased** because the treatment and control groups are no longer correctly matched.

Matching (Propensity Score Matching - PSM)

Example:

- **Full compliance:** Job training participants complete the full training program, while non-participants receive no training.
- **Non-compliance:** Some assigned to training **drop out**, or some in the control group **attend private training programs**.

Take-aways

1. **Randomized Assignment**: Works well with **full compliance**, but if not, use **ITT or IV**.
2. **RDD**: Allows for **non-compliance**, but use **Fuzzy RDD** when treatment assignment is imperfect.
3. **DiD**: Works **best with full compliance**, but **IV methods** can correct for switching.
4. **Matching**: **Full compliance is necessary**; otherwise, treated and control groups are no longer comparable.

Summary

Method	**Full Compliance Needed?**	**What Happens with Non-Compliance?**
Randomized Assignment (RCT)	Yes (for simple ATE estimation)	Need **ITT** or **IV (LATE)**
Regression Discontinuity (RDD)	No (can handle non-compliance)	Use **Fuzzy RDD (IV approach)**
Difference-in-Differences (DiD)	Yes (for basic DiD)	Use **IV or alternative models**
Matching (PSM)	Yes	If non-compliance exists, becomes biased

Identifying Non-compliance

Question: "I read in an article that we cannot differentiate non-compliance by simply looking at the data. Is this only true before testing the treatment? Can't we still detect Never-Takers in the treatment group and Always-Takers in the comparison group after observing the treatment outcomes?"

Identifying Non-compliance

Answer: it depends. . .

- The **key issue** is that individuals — **Never-Takers, Always-Takers, Compliers, and Defiers** — are defined by their **potential outcomes**. These are **unobservable** for each individual.
- This is a direct consequence of the **Fundamental Problem of Causal Inference** (we never observe both treated and untreated outcomes for the same individual).

Intuition:

- **"Can we ever see the same person in both treatment and control groups?"**
  - Answer: **No**, which is why we cannot classify individuals into compliance types.
- **"Can we estimate the fraction of Compliers?"**
  - Answer: **Yes**, but only in aggregate using statistical methods.

To the details. . .

Can We Identify Non-Compliance Before Observing Treatment?

Yes, sometimes, but only if we have extra information.

- Before treatment is assigned, we **do not** observe actual treatment take-up, so we cannot distinguish between **Compliers**, **Never-Takers**, **Always-Takers**, and **Defiers**.
- However, if we have survey data on individuals' preferences or eligibility rules, we might make some educated guesses.

(How realistic?)

Can We Identify Non-Compliance After Observing Treatment?

Partially, but never fully for individuals.

Only in aggregate through statistical methods.

After treatment assignment, we observe:

- Who was **assigned to treatment** (but some may refuse).
- Who was **assigned to control** (but some may still find a way to get treated).
- The **actual treatment status** (who received treatment and who didn't).
- The **outcomes** (e.g., test scores, earnings, health improvement).

Can We Identify Non-Compliance After Observing Treatment? cont'd

However, the **fundamental problem** is:

- We never see the same person **both treated and untreated**, so we do not observe their **counterfactual behavior** (i.e., what they would have done in the alternative assignment).
- This means we **cannot perfectly classify individuals** into compliance types.

Can We Detect Never-Takers and Always-Takers?

In some cases, yes, but only for some groups.

- **Never-Takers in the treatment group:**
  - If someone was **assigned to treatment** but **did not take it**, we can **definitely** classify them as a Never-Taker.
- **Always-Takers in the control group:**
  - If someone was **assigned to control** but **still got treated**, we can **definitely** classify them as an Always-Taker.

The problem:
For everyone else, we cannot tell!

- A **treated person in the treatment group** could be a **Complier or an Always-Taker**.
- An **untreated person in the control group** could be a **Complier or a Never-Taker**.
- We can never directly observe Defiers (those who do the **opposite** of their assignment).

How Do We Detect Non-Compliance in Practice?

Since we cannot identify **Compliers and Defiers** directly, we use **Instrumental Variables (IV) methods** (in particular 2SLS) to estimate the **Local Average Treatment Effect (LATE)**, which captures the effect of treatment **only on Compliers**.

Step 1: First-Stage Regression (Predict Treatment)

$$D_i = \alpha + \gamma Z_i + \epsilon_i$$

where:

- D_i = actual treatment status
- Z_i = treatment assignment (instrument)
- γ tells us **how strongly assignment affects actual treatment take-up**.

How Do We Detect Non-Compliance in Practice? cont'd

Step 2: Estimate Treatment Effect for Compliers

$$Y_i = \beta + \tau \hat{D}_i + \nu_i$$

where:

- τ is the **LATE** (treatment effect on Compliers).

Some words on γ

The coefficient γ tells us **how strongly treatment assignment** Z **influences actual treatment** D, or the **fraction of Compliers in the population**, because $\gamma = P(\text{Compliers})$.

- **If** $\gamma = 1$**:**
  - **Perfect compliance** $\rightarrow$ Everyone assigned to treatment receives treatment, and no one assigned to control gets treated, i.e., the instrument is a **perfect predictor** of treatment.
- **If** $\gamma < 1$**:**
  - **Imperfect compliance** $\rightarrow$ Some people **assigned to treatment refuse it (Never-Takers)**, or some **assigned to control still get treated (Always-Takers)**.
  - The instrument is still correlated with treatment, but there are **Compliers, Never-Takers, and Always-Takers**.
- **If** $\gamma = 0$**:**
  - The instrument **does not affect treatment at all**, meaning assignment Z has **no predictive power** for actual treatment.
  - **IV is invalid because there's no first-stage relationship**.

Some words on γ cont'd

- **If $0 < \gamma < 1$, but small:**
  - The instrument **weakly predicts** treatment.
  - This is a **weak instrument problem**, meaning the IV estimates in the second-stage may be unreliable.

Example.

- If $\gamma = 0.6$, then **only 60% of those assigned to treatment actually take it**.
- If γ is **too small** (say, $\gamma = 0.1$), we have a **weak instrument problem**, making standard errors large and estimates unreliable.

Conclusion.

- A **higher** γ means the instrument is a **strong predictor** of treatment, leading to **more precise** estimates in the second stage.

Take-aways

1. **We can detect Never-Takers in the treatment group**:
   - If a person was **assigned to treatment ($Z = 1$) but refused treatment ($D = 0$)**, we can classify them as a **Never-Taker**.
2. **We can detect Always-Takers in the control group**:
   - If a person was **assigned to control ($Z = 0$) but still got treated ($D = 1$)**, we can classify them as an **Always-Taker**.
3. **We cannot directly detect Compliers and Defiers**:
   - A person in the treatment group who **took treatment** could be either a **Complier or an Always-Taker**.
   - A person in the control group who **did not take treatment** could be either a **Complier or a Never-Taker**.
   - We never see what each individual **would have done in the opposite assignment**.

Summary

Question	**Answer**
Can we detect non-compliance before treatment?	No, unless we have external information.
Can we detect Never-Takers in the treatment group?	Yes, they are assigned to treatment but refuse it.
Can we detect Always-Takers in the control group?	Yes, they are assigned to control but still get treated.
Can we detect Compliers individually?	No, only in aggregate using IV.
Can we detect Defiers?	No, because we never observe counterfactual compliance.
How do we estimate compliance rates?	Using **first-stage IV regression** and compliance proportions.

Relation between Fuzzy RD and 2SLS

- Fuzzy RDD is similar to IV (2SLS) because both use an instrumental variable to isolate exogenous variation in treatment. In Fuzzy RDD, the cutoff acts as an instrument because it influences treatment probability but does not directly affect the outcome. We estimate Fuzzy RDD using two-stage least squares, just like IV.
- **In short:** Fuzzy RDD is just IV with a cutoff as the instrument!
- In what follows, we will discuss why fuzzy RDD is like IV, how the cutoff acts as an instrument, and how we estimate the treatment effect.

1. Why is Fuzzy RDD Similar to 2SLS?

Both **Fuzzy RDD** and **IV (2SLS)** deal with the **same problem**:

We do not have full compliance, meaning that treatment is **not perfectly determined** by assignment.

- In **2SLS (IV regression)**, we use an instrument **Z** to deal with **endogenous treatment D**.
- In **Fuzzy RDD**, we use the **cutoff in the running variable** as an **instrument** for actual treatment.

Both use two stages to estimate the causal effect, separating the **part of treatment that is truly exogenous**.

2. How Does the Cutoff in Fuzzy RDD Act Like an Instrument?

Think of the **running variable** X (e.g., test score, income level) and the **cutoff** c (e.g., 50 points required for a scholarship).

- People **above the cutoff** ($X \geq c$) are **more likely** to receive treatment.
- People **below the cutoff** ($X < c$) are **less likely** to receive treatment.
- **But not everyone follows the rule** $\rightarrow$ some people **above the cutoff refuse treatment**, and some **below the cutoff still receive it**. This creates **non-compliance**.

2. How Does the Cutoff in Fuzzy RDD Act Like an Instrument? cont'd

Key Idea:

The cutoff **affects the probability of receiving treatment but does not directly affect the outcome**, so we can use it as an **instrument**.

We use the cutoff as an IV for treatment.

- The **running variable cutoff** is like a **randomized encouragement** to take the treatment.
- It **only** affects the outcome **through treatment**.

3. How is Fuzzy RDD Estimated Using 2SLS?

Since not everyone follows the cutoff rule, we estimate Fuzzy RDD using **Two-Stage Least Squares (2SLS):**

Step 1: First-Stage Regression (Predicting Treatment)

We estimate **how much the cutoff influences treatment**:

$$D_i = \alpha + \gamma \cdot 1(X_i \geq c) + f(X_i) + \epsilon_i$$

where:

- D_i = actual treatment received.
- $1(X_i \geq c)$ = **instrument** (binary variable: 1 if above cutoff, 0 if below).
- $f(X_i)$ = smooth function of the running variable.
- γ tells us **how strongly assignment to treatment depends on the cutoff**.

This is just like the first-stage in 2SLS, where the instrument predicts the endogenous variable.

3. How is Fuzzy RDD Estimated Using 2SLS? cont'd

Step 2: Second-Stage Regression (Estimating Treatment Effect)

Now, we use the predicted treatment **from the first stage** to estimate the treatment effect on Y:

$$Y_i = \beta + \tau \hat{D}_i + f(X_i) + \nu_i$$

where:

- Y_i = outcome (e.g., earnings, test scores).
- $\hat{D}_i$ = **predicted treatment** from the first-stage equation.
- τ = **treatment effect (LATE, Local Average Treatment Effect)**.

This is exactly like the second-stage in 2SLS, where we replace the endogenous variable with its instrumented version.

4. Why Does This Work?

Fuzzy RDD **identifies the causal effect only for Compliers**—people whose treatment status **changes due to the cutoff**.

- **Never-Takers**: Stay untreated no matter what → cutoff doesn't affect them.
- **Always-Takers**: Always get treatment → cutoff doesn't affect them.
- **Compliers**: Take treatment **only if assigned** → cutoff affects them.
- **Defiers**: Take the opposite action → assumed to be rare (monotonicity assumption – advanced topic).

Key Insight:

The cutoff **only** affects treatment probability but does **not directly affect Y (the outcome)** except through treatment → which is exactly how a valid instrument works in IV.

Summary

Concept	IV (2SLS)	Fuzzy RDD
Endogenous Variable	Treatment D	Treatment D
Instrument	External Z (e.g., lottery, encouragement)	Running variable **cutoff** $1(X \geq c)$
First-Stage Regression	$D_i = \alpha + \gamma Z_i + \epsilon_i$	$D_i = \alpha + \gamma 1(X \geq c) + f(X) + \epsilon_i$
Second-Stage Regression	$Y_i = \beta + \tau \hat{D}_i + \nu_i$	$Y_i = \beta + \tau \hat{D}_i + f(X) + \nu_i$
Estimated Effect	LATE (Local Average Treatment Effect)	LATE (Local Average Treatment Effect)
Handles Non-Compliance?	Yes	Yes