Regression Discontinuity

Professor Ji-Woong Chung
Korea University

Outline

Basic idea of regression discontinuity

Sharp versus fuzzy discontinuities
- Notation & sharp vs. fuzzy assumption
- Assumption about local continuity

Estimating regression discontinuity
- Sharp regression discontinuity
- Graphical analysis
- Fuzzy regression discontinuity

Checks on internal validity

Heterogeneous effects & external validity

Outline

Basic idea of regression discontinuity

Sharp versus fuzzy discontinuities
- Notation & sharp vs. fuzzy assumption
- Assumption about local continuity

Estimating regression discontinuity
- Sharp regression discontinuity
- Graphical analysis
- Fuzzy regression discontinuity

Checks on internal validity

Heterogeneous effects & external validity

Basic Idea of RDD

- Observations (e.g., firm, individual, etc.) are 'treated' based on a known cutoff rule.
  - For some observable variable, x, an observation is treated if $x \geq x'$.
  - This cutoff creates the discontinuity.
- Researchers are interested in how this treatment affects the outcome variable of interest, y.

Examples of RDD Settings

- Cutoff rules are commonplace in finance:
  - A borrower FICO score > 620 makes securitization of the loan more likely (Keys, et al., QJE 2010).
  - Accounting variable x exceeding some threshold causes loan covenant violation (Roberts & Sufi, JF 2009).

RDD and Difference-in-Differences

- Has a similar flavor to diff-in-diff natural experiment setting as you can illustrate identification with a figure.
- Plot outcome y against the independent variable x that determines treatment assignment.
- Should observe a sharp, discontinuous change in y at the cutoff value of x'.

But RDD is different. . .

- RDD has some key differences. . .
  - Assignment to treatment is NOT random; assignment is based on value of x
  - When treatment only depends on x (what I'll later call "sharp RDD", there is no overlap in treatment & controls; i.e., we never observe the same x for a treatment and a control

RDD randomization assumption

- Assignment to treatment and control isn't random, but whether an individual observation is treated is assumed to be random.
  - Researchers assume observations (e.g., firm, person, etc.) can't perfectly manipulate their x value.
  - Therefore, whether an observation's x falls immediately above or below key cutoff x' is random.

Outline

Basic idea of regression discontinuity

Sharp versus fuzzy discontinuities

- Notation & sharp vs. fuzzy assumption
- Assumption about local continuity

Estimating regression discontinuity

- Sharp regression discontinuity
- Graphical analysis
- Fuzzy regression discontinuity

Checks on internal validity

Heterogeneous effects & external validity

Outline

Basic idea of regression discontinuity

Sharp versus fuzzy discontinuities
- Notation & sharp vs. fuzzy assumption
- Assumption about local continuity

Estimating regression discontinuity
- Sharp regression discontinuity
- Graphical analysis
- Fuzzy regression discontinuity

Checks on internal validity

Heterogeneous effects & external validity

RDD Terminology

- x is called the "forcing variable"
  - Can be a single variable or multiple variables; but for simplicity, we'll work with a single variable
- x' is called the "threshold"
- $y(0)$ is outcome absent treatment
- $y(1)$ is outcome with treatment

Two Types of RDD

- Sharp RDD
  - Assignment to treatment only depends on x; i.e., if $x \geq x'$ you are treated with probability 1
- Fuzzy RDD
  - Having $x \geq x'$ only increases probability of treatment; i.e., other factors (besides x) will influence whether you are treated or not

Sharp RDD Assumption #1

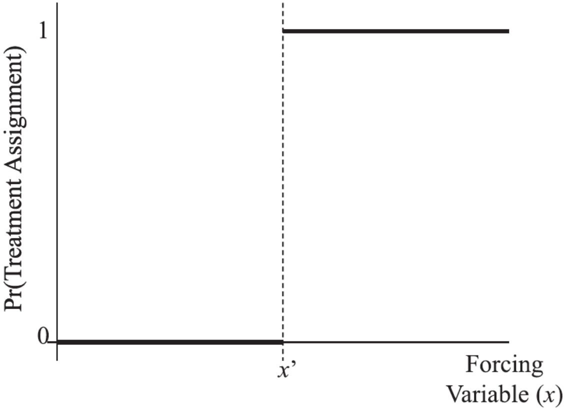
- Assignment to treatment occurs through known and deterministic decision rule:

$$\delta(x) = \begin{cases} 1 & \text{if } x \geq x' \\ 0 & \text{otherwise} \end{cases}$$

  - It is important that there exists x's around the threshold value

Sharp RDD Assumption #1 – Visually

- Probability of treatment moves from 0 to 1 around threshold value x'
- No untreated for $x > x'$ and no treated for $x < x'$
- Only x determines treatment



Sharp RDD – Examples

- Ex. #1 – PSAT score $> x'$ means student receives national merit scholarship
  - Receiving scholarship was determined solely based on PSAT scores in the past
  - Thistlewaithe and Campbell (1960) used this to study effect of scholarship on career plans

Fuzzy RDD: Assumption #1 — Stochastic Assignment

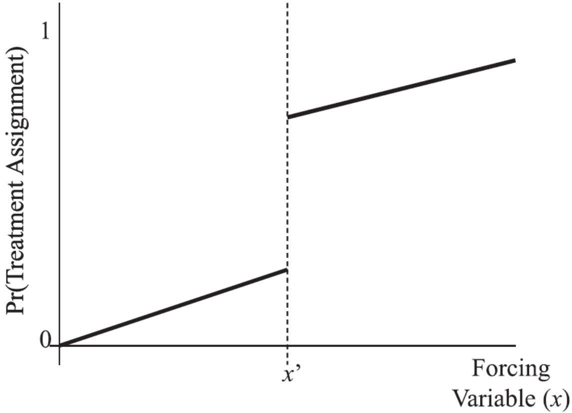
- In a **fuzzy regression discontinuity design (RDD)**, crossing the cutoff x' does not deterministically assign treatment.
- Instead, only the *probability* of receiving treatment jumps at the cutoff:

$$0 < \lim_{x \to x'^{+}} \Pr[\delta(x) = 1] \; - \; \lim_{x \to x'^{-}} \Pr[\delta(x) = 1] < 1$$

- This jump in treatment probability identifies a **local instrumental variable** at x':
    - Units just above and below x' differ only in their likelihood of treatment.
    - The direction of the jump can be upward or downward — all that matters is a discontinuity in $\Pr(\text{treatment} \mid x)$.
- Intuitively: treatment assignment around x' behaves like a randomized design.

Fuzzy RDD Assumption #1 – Visually

- Treatment probability increases at x'
- Some untreated for $x > x'$ and some treated for $x < x'$
- Treatment is not purely driven by x



Fuzzy RDD – Example

- Ex. #1 – FICO score > 620 increases likelihood of loan being securitized
  - But, extent of loan documentation, lender, etc., will matter as well. . .

Sharp versus Fuzzy RDD

- This subtle distinction affects exactly how you estimate the causal effect of treatment
  - With Sharp RDD, we will basically compare average y immediate above and below x'
  - With fuzzy RDD, the average change in y around threshold understates causal effect [Why?]
  - Answer = Comparison assumes all observations were treated, but this isn't true; if all observations had been treated, observed change in y would be even larger; we will need rescale based on change in treatment probability

Sharp vs. Fuzzy RDD

- The distinction determines **how we identify and estimate** the causal effect at the cutoff x'.
- **Sharp RDD:** treatment status changes deterministically at x'.
  - Everyone with $x \geq x'$ is treated; everyone with $x < x'$ is not.
  - Estimate the causal effect by comparing average outcomes just above and below the cutoff:

$$\tau_{\text{Sharp}} = \lim_{x \to x'+} E[y \mid x] - \lim_{x \to x'-} E[y \mid x].$$

- **Fuzzy RDD:** treatment probability (not status) jumps at x'.
  - The discontinuity in $E[y \mid x]$ reflects only a *partial compliance* effect.
  - Thus, the raw jump in y **understates** the causal effect because not everyone complies with treatment assignment.
  - The true local effect is obtained by **rescaling** the jump in y by the jump in treatment probability:

$$\tau_{\text{Fuzzy}} = \frac{\text{Jump in } E[y \mid x]}{\text{Jump in } E[\delta(x) \mid x]} = \frac{\lim_{x \to x'+} E[y \mid x] - \lim_{x \to x'-} E[y \mid x]}{\lim_{x \to x'+} E[\delta(x) \mid x] - \lim_{x \to x'-} E[\delta(x) \mid x]}.$$

- **Intuition:** The discontinuity acts as an *instrument* that encourages treatment; fuzzy RDD estimates a *local average treatment effect (LATE)* among compliers.

Example: Estimating a Fuzzy RDD

- Scholarship program for students scoring above 80 on a test.
- Not all eligible students take the scholarship $\Rightarrow$ fuzzy RDD.
- Estimate:

$$\tau_{\text{FRD}} = \frac{\text{Jump in GPA at } x'}{\text{Jump in treatment probability at } x'}.$$

- Numerator: discontinuity in outcome $\Rightarrow$ local change in GPA.
- Denominator: discontinuity in treatment probability $\Rightarrow$ partial compliance.
- Ratio $\Rightarrow$ local average treatment effect (LATE) for compliers near the cutoff.

Outline

Basic idea of regression discontinuity

Sharp versus fuzzy discontinuities
- Notation & sharp vs. fuzzy assumption
- Assumption about local continuity

Estimating regression discontinuity
- Sharp regression discontinuity
- Graphical analysis
- Fuzzy regression discontinuity

Checks on internal validity

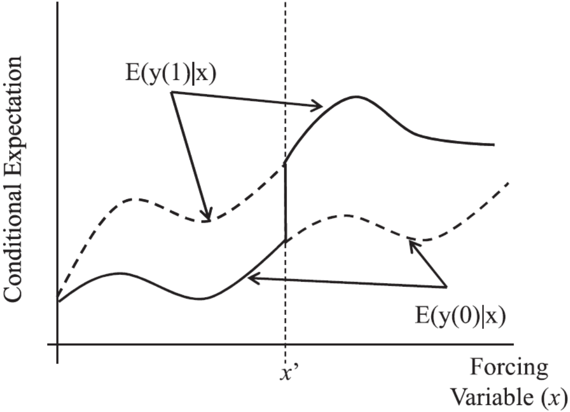
Heterogeneous effects & external validity

RDD Assumption #2

- But, both RDDs share the following assumption about **local continuity**
- Potential outcomes, $y(0)$ and $y(1)$, conditional on forcing variable, x, are continuous at threshold x'
  - In words: y would be a smooth function around threshold absent treatment; i.e., don't expect any jump in y at threshold x' absent treatment

RDD Assumption #2 – Visually

- If all obs. had been treated, y would be smooth around x'; other lines says equivalent thing for none had been treated
- Dashed lines represent unobserved counterfactuals



Outline

Basic idea of regression discontinuity

Sharp versus fuzzy discontinuities
- Notation & sharp vs. fuzzy assumption
- Assumption about local continuity

Estimating regression discontinuity
- Sharp regression discontinuity
- Graphical analysis
- Fuzzy regression discontinuity

Checks on internal validity

Heterogeneous effects & external validity

Outline

Basic idea of regression discontinuity

Sharp versus fuzzy discontinuities
- Notation & sharp vs. fuzzy assumption
- Assumption about local continuity

Estimating regression discontinuity
- Sharp regression discontinuity
- Graphical analysis
- Fuzzy regression discontinuity

Checks on internal validity

Heterogeneous effects & external validity

How *Not* to Do Sharp RDD [Part 1]

- Suppose we want to estimate the causal effect of treatment δ_i (e.g., loan securitization) on an outcome y_i (e.g., default).

$$y_i = \beta_0 + \delta_i + u_i$$

- Will this regression identify the causal effect of δ on y?
  - **No!** Treatment δ_i is typically correlated with the running variable x_i (e.g., FICO score).
  - If x_i itself affects y_i, omitting it creates bias:

$$\text{Corr}(\delta_i, u_i) \neq 0.$$

  - Example: In *Keys et al. (2010)*, borrowers with higher FICO scores are both
    1. more likely to have their loans securitized, and
    2. less likely to default.

    Thus, the simple regression above confounds the effect of securitization with the effect of borrower quality.

How *Not* to Do Sharp RDD [Part 2]

- A natural "fix" might be to control for the running variable x_i:

$$y_i = \beta_0 + \delta_i + \beta_1 x_i + u_i.$$

- Does this solve the omitted-variable problem?
  - **Not .** Two key issues remain:
    1. It imposes a linear effect of x_i on y_i across the entire range — which may be misspecified.
    2. It ignores the local randomization feature of RDD — the credible variation is only among observations *near the cutoff* x'.
  - In RDD, we exploit the fact that just above and below x', treatment assignment is "as good as random." Controlling linearly for x_i across the full sample washes out this local experiment.

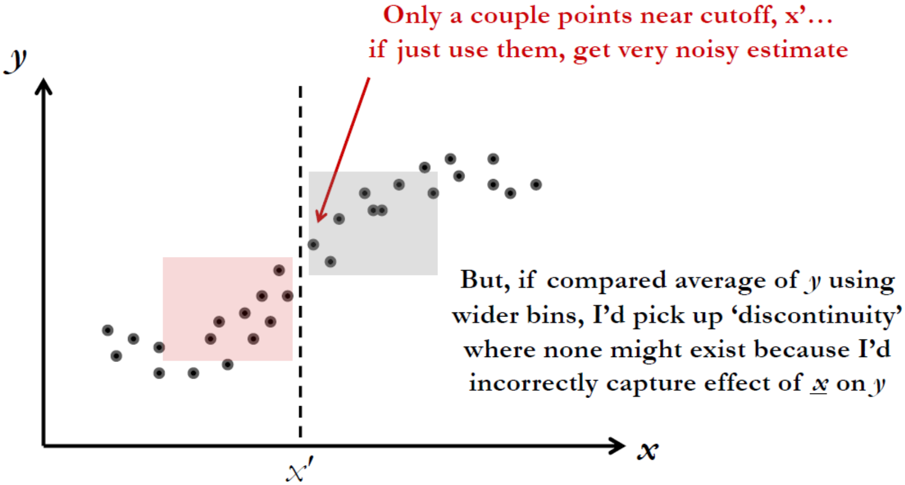
Bias–Variance Tradeoff in RDD

- In principle, we would like to compare outcomes for units with running variable x just **below** and just **above** the cutoff x'.
- In practice, we face a tradeoff when choosing how close to stay to x':
  - Using a very narrow window (small bandwidth) means the comparison is **highly local** and nearly unbiased — but based on few observations, so the estimate is **noisy**.
  - Using a wider window (large bandwidth) increases precision (less noise) — but introduces **bias** if units farther from x' differ systematically for reasons other than treatment, or if x directly affects y.
- The bandwidth thus determines the tradeoff between:

$$\text{Bias (local validity)} \quad \text{vs.} \quad \text{Variance (statistical precision)}.$$

- Modern RDD estimators (e.g., `rdrobust`) choose bandwidths that optimally balance this bias–variance tradeoff.

Bias versus Noise – Visual



Estimating Sharp RDD

- Two general strategies exist for estimating the treatment effect in a sharp RDD, each balancing bias and variance:

- **(1) Global approach:** use the full sample and flexibly control for the running variable x.
  - Fit a flexible function of x (e.g., high-order polynomial, spline).
  - Pros: uses all data, lower variance.
  - Cons: risk of misspecification away from the cutoff; may impose too much functional structure.

- **(2) Local approach:** restrict the sample to observations near the cutoff.
  - Estimate treatment effect using only data in a narrow bandwidth around x' (local linear or polynomial).
  - Pros: more credible causal identification (local randomization).
  - Cons: fewer observations $\rightarrow$ higher variance; choice of bandwidth matters.

- Modern practice: favor **local linear regression** with data close to the cutoff and bandwidth selected using data-driven procedures (e.g., `rdrobust`).

Estimating Sharp RDD: Using All Data (Global Approach)

- The **global approach** uses the entire sample but estimates separate outcome functions on each side of the cutoff x':

$$
\begin{aligned}
y_i &= \beta_b + f(x_i - x') + u_i, &\quad \text{for } x_i < x', \\
y_i &= \beta_a + g(x_i - x') + u_i, &\quad \text{for } x_i \geq x'.
\end{aligned}
$$

- Here:
  - $f(\cdot)$ and $g(\cdot)$ are smooth (continuous) functions of the running variable, normalized so that $f(0) = g(0) = 0$.
  - The treatment effect at the cutoff is the difference in the fitted intercepts:

$$\tau = \beta_a - \beta_b.$$

- This approach fits both sides of the threshold using all observations:
  - Pros: lower variance, uses all data.
  - Cons: relies heavily on correct functional form (misspecification far from x' can bias $\hat{\tau}$).
- In practice, researchers often use flexible polynomials or splines for f and g—but still test robustness using more local methods.

Interpreting the Estimates

$$y_i = \beta_b + f(x_i - x') + u_i, \qquad \text{for } x_i < x',$$
$$y_i = \beta_a + g(x_i - x') + u_i, \qquad \text{for } x_i \geq x'.$$

- **Why include** $f(\cdot)$ and $g(\cdot)$?
  - They control for the smooth relationship between the running variable x and the outcome y.
  - Without these terms, any natural slope of y in x would be incorrectly attributed to the treatment jump.

- **What do β_b and β_a represent?**
  - $\beta_b = \lim_{x \to x'-} E[y_i \mid x_i]$: the expected outcome just below the cutoff.
  - $\beta_a = \lim_{x \to x'+} E[y_i \mid x_i]$: the expected outcome just above the cutoff.
  - Their difference,

    $$\tau = \beta_a - \beta_b,$$

    captures the discontinuous jump in $E[y|x]$ at x', interpreted as the **causal treatment effect at the cutoff.**

A Simpler Way to Estimate Sharp RDD

- We can estimate both sides of the cutoff in a **single regression** using the full sample:

$$y_i = \alpha + \beta d_i + \underbrace{f(x_i - x') + d_i \times g(x_i - x')}_{\text{controls for the } x\text{–}y \text{ relationship on both sides of the cutoff}} + u_i$$

  where $d_i = \mathbf{1}(x_i \geq x')$ indicates treatment assignment.

- Here:
  - $f(\cdot)$ models the smooth relationship between x and y **below** the cutoff.
  - $g(\cdot)$ allows that relationship to differ **above** the cutoff.
  - The coefficient β captures the **jump at the cutoff:**

$$\hat{\beta} = \hat{\beta}_a - \hat{\beta}_b.$$

- **If we drop the interaction term** $d_i \times g(x_i - x')$:
  - We are assuming that the slope of y in x is the **same on both sides** of the cutoff.
  - This imposes a potentially restrictive functional form and can bias the estimated treatment effect if the trend differs above and below x'.

Choosing Functional Forms for $f(\cdot)$ and $g(\cdot)$

- In practice, we approximate the unknown functions $f(\cdot)$ and $g(\cdot)$ with **flexible polynomials** in $(x_i - x')$:

$$y_i = \alpha + \beta d_i + \sum_{s=1}^{p} \gamma_s^b (x_i - x')^s + \sum_{t=1}^{p} \delta_t^a \, d_i (x_i - x')^t + u_i,$$

  where p is the chosen polynomial order.

- **How to choose the order p?**
  - Higher-order polynomials fit the data more flexibly ($\downarrow$ bias) but can overfit and oscillate near the cutoff ($\uparrow$ variance).
  - Use model-selection tools or visual inspection to check sensitivity to different orders.
  - Modern practice often favors **local linear** or **local quadratic** fits over high-order global polynomials.

- The goal is not to perfectly fit $y(x)$ globally, but to *approximate it smoothly near the cutoff*.

Estimating Sharp RDD: Using a Window (Local Approach)

- The **local approach** estimates the RDD model using only observations near the cutoff:
$$x' - \Delta \le x_i \le x' + \Delta,$$
where $\Delta > 0$ defines the **bandwidth** or window width.

- Within this window, we fit a simple (usually linear) model on each side:
$$y_i = \alpha + \beta d_i + \underbrace{\gamma^b (x_i - x') + \gamma^a d_i (x_i - x')}_{\text{linear control for } x \text{ on each side}} + u_i.$$

- Key features:
  - Restricting the sample to a narrow window reduces bias by focusing on observations most comparable across the cutoff.
  - Lower-order controls (e.g., linear terms) avoid overfitting and reduce sensitivity to functional form.
  - The choice of Δ determines the **bias–variance tradeoff**:
    - Smaller $\Delta \to$ less bias, more noise.
    - Larger $\Delta \to$ more precision, higher bias.

- This is known as a **local linear regression**, the standard method in modern RDD estimation.

Practical Issues in Local RDD Estimation

- Two key implementation choices affect RDD estimates:
  1. The **bandwidth** (window width) Δ — how close observations must be to the cutoff.
  2. The **polynomial order** used to approximate $f(\cdot)$ and $g(\cdot)$ within that window.

- There is no single "correct" choice — both involve a tradeoff:
  - A smaller Δ yields more credible local comparisons but fewer observations ($\uparrow$ variance).
  - A larger Δ increases precision but risks bias if the relationship between x and y is nonlinear.
  - Within a small window, lower-order polynomials (typically linear or quadratic) are usually sufficient.

- **Best practice:**
  - Report results using multiple bandwidths and polynomial orders.
  - Show that conclusions are **robust** to these choices.
  - Use data-driven bandwidth selection methods (e.g., Imbens–Kalyanaraman, Calonico–Cattaneo–Titiunik).

Outline

Basic idea of regression discontinuity

Sharp versus fuzzy discontinuities
- Notation & sharp vs. fuzzy assumption
- Assumption about local continuity

Estimating regression discontinuity
- Sharp regression discontinuity
- Graphical analysis
- Fuzzy regression discontinuity

Checks on internal validity

Heterogeneous effects & external validity

Graphical Analysis of RDD

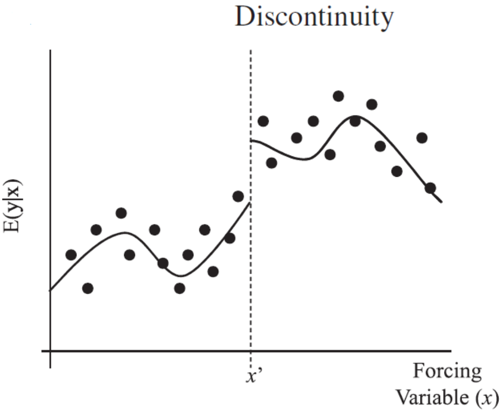
- A key diagnostic in RDD is a **binned scatter plot** of the outcome y against the running variable x.
- Purpose:
  - Visually inspect whether a **discontinuity** exists at the cutoff x'.
  - Assess whether the chosen functional form (e.g., local linear, polynomial) fits the data well.
  - Serves as a simple **sanity check**—shows the variation driving the estimate.
- Always include such a plot when presenting RDD results.

Constructing the RDD Graph

1. **Bin the data:** Divide the running variable x into intervals that do not cross the cutoff x'.
   - Example: if $x' = 4.5$, use bins like [0,0.5), [0.5,1.5), [1.5,2.5), ... ensuring one bin ends at 4.5.
2. **Compute means:** For each bin, calculate the average outcome $\bar{y}$ and plot it at the bin midpoint. These points approximate the conditional mean function $E[y|x]$.
3. **Overlay fitted lines:** Plot predicted values from your RDD estimation (e.g., local linear or polynomial fits) on each side of the cutoff.
4. **Interpret visually:** A clear jump in the fitted lines at x' indicates a treatment effect.

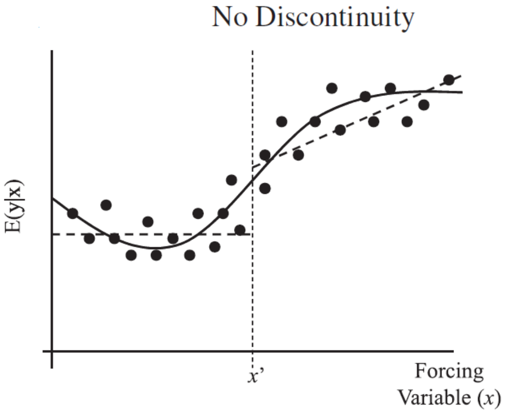
Example of Supportive Graph

- Each dot is average y for corresponding bin
- Solid line is predicted values of y from RDD regression
- Discontinuity is apparent in both estimation and nonparametric plot
- Fifth-order polynomial was needed to fit the non-parametric plot



Example of Non-Supportive Graph

- Dash lines would have been predicted values from linear RDD [i.e., polynomial of order 1]
- But, looking at nonparametric graph would make clear that a cubic version (which is plotted as solid line) would show no effect!



RDD Graphs – Diagnostic Checks

- The nonparametric (binned) RDD plot should show a clear jump in y **only at the cutoff** x'.
- If similar jumps appear elsewhere:
  - It undermines the **internal validity** of the design.
  - Suggests that the discontinuity at x' may reflect other factors unrelated to treatment (e.g., nonlinear trends, omitted variables, or manipulation in x).
- Always inspect the full scatter to confirm that the apparent discontinuity is unique to x'.

Choosing Bin Width in RDD Graphs

- The choice of bin width (or number of bins) affects how the binned RDD plot represents $E[y|x]$.
- There is no single "optimal" choice—bin width involves a **bias–variance tradeoff**:
  - **Wider bins:** include more observations per bin $\Rightarrow$ smoother, more precise averages ($\downarrow$ variance).
  - **Narrow bins:** track local variation more closely $\Rightarrow$ less bias but more noise ($\uparrow$ variance).
- Bias arises if $E[y|x]$ changes within wide bins (non-zero slope inside a bin).
- In practice:
  - Experiment with different bin widths for robustness.
  - Automated choices (e.g., evenly spaced or quantile bins) are acceptable as long as visual conclusions remain stable.

Testing for Overly Wide Graph Bins

1. Create an indicator variable for each bin of the running variable x.
2. Regress the outcome y on these bin indicators and their interactions with x:
$$y_i = \rho x_i + \sum_j \alpha_j \text{Bin}_j + \sum_j \gamma_j (\text{Bin}_j \times x_i) + u_i.$$
3. Perform a joint F-test of the interaction terms $\{\gamma_j\}$.
   - If the null is rejected, slopes differ across bins $\Rightarrow$ bins are too wide.
   - If not rejected, bin slopes are roughly constant—bin width likely acceptable.

Reference: Lee and Lemieux (2010, *JEL*) discuss this and related graphical specification checks.

Outline

Basic idea of regression discontinuity

Sharp versus fuzzy discontinuities
- Notation & sharp vs. fuzzy assumption
- Assumption about local continuity

Estimating regression discontinuity
- Sharp regression discontinuity
- Graphical analysis
- Fuzzy regression discontinuity

Checks on internal validity

Heterogeneous effects & external validity

Intuition for Fuzzy RDD

- In a **sharp RDD**, treatment status changes deterministically at the cutoff:

$$x_i \geq x' \;\Rightarrow\; D_i = 1.$$

- In a **fuzzy RDD**, crossing the cutoff only **changes the probability** of treatment:

$$\Pr(D_i = 1 \mid x_i \geq x') \;>\; \Pr(D_i = 1 \mid x_i < x').$$

- Therefore, a simple comparison of average outcomes above and below x' no longer identifies the treatment effect.
- Instead, we can use the cutoff as an **instrumental variable**:

$$\text{Indicator}(x_i \geq x') \text{ acts as an IV for actual treatment } D_i.$$

- The resulting estimate identifies a **Local Average Treatment Effect (LATE)** for units whose treatment status is influenced by crossing the cutoff.

Fuzzy RDD: Notation

- To formalize the fuzzy RDD setup, define two key variables:
  - **Treatment indicator:**
  $$D_i = \begin{cases} 1 & \text{if unit } i \text{ actually receives treatment,} \\ 0 & \text{otherwise.} \end{cases}$$
  - **Threshold (instrument) indicator:**
  $$T_i = \begin{cases} 1 & \text{if } x_i \geq x', \\ 0 & \text{if } x_i < x'. \end{cases}$$
- T_i captures assignment based on the **running variable** x_i, while D_i captures **actual treatment status.**
- In a fuzzy design, crossing the cutoff changes the **probability** of treatment:
$$E[D_i \mid T_i = 1] \neq E[D_i \mid T_i = 0].$$
- Example:
  - $D_i = 1$ if a loan is *securitized*.
  - $T_i = 1$ if the borrower's FICO score ≥ 620, which **increases** the likelihood of securitization.

Estimating Fuzzy RDD (Part 1)

- The fuzzy RDD can be estimated using a **two-stage least squares (2SLS)** framework:

  $$y_i = \alpha + \beta D_i + f(x_i - x') + u_i,$$

  where the threshold indicator $T_i = \mathbf{1}(x_i \geq x')$ serves as an instrument for D_i.

- **First stage:**

  $$D_i = \pi_0 + \pi_1 T_i + f(x_i - x') + v_i$$

- **Second stage:**

  $$y_i = \alpha + \beta \hat{D}_i + f(x_i - x') + u_i$$

- **Key IV assumptions:**
  - **Relevance:** T_i affects the probability of treatment — $\pi_1 \neq 0$.
  - **Exclusion:** T_i affects y_i only through its effect on D_i, after controlling for $f(x_i - x')$.

- Under these assumptions, $\hat{\beta}$ identifies the **Local Average Treatment Effect (LATE)** at the cutoff.

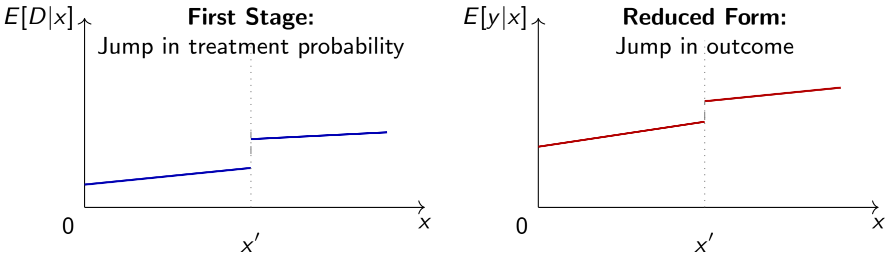
Estimating Fuzzy RDD (Part 2)

- The function $f(\cdot)$ captures the smooth relationship between the running variable x and the outcome y.
- In practice, $f(\cdot)$ is typically modeled using a low-order polynomial in $(x_i - x')$.
- Unlike in sharp RDD, allowing fully flexible functional forms above and below the cutoff is less straightforward in the 2SLS framework.
- **Practical solution:**
  - Use a **narrower bandwidth** (smaller window) around x'.
  - Within a small window, the relationship between x and y is approximately linear, making the estimate less sensitive to the choice of $f(\cdot)$.
- As always, check robustness across bandwidths and polynomial orders.

Graphical Analysis in Fuzzy RDD

- As in the sharp RDD, start by plotting the outcome variable y_i against the running variable x_i:
  - Look for a visible **discontinuity in** $E[y|x]$ at the cutoff x'.
  - Check that the chosen functional form (e.g., polynomial or local linear fit) captures the underlying trend smoothly on both sides.
- In the **fuzzy RDD**, also plot the **treatment indicator** D_i against x_i:
  - This graph should show a clear **jump in the probability of treatment** at x'.
  - That jump visually confirms the **first-stage relevance** of the instrument $T_i = \mathbf{1}(x_i \geq x')$.
- Intuitively:
  - The discontinuity in D_i reflects how crossing the cutoff changes treatment likelihood.
  - The discontinuity in y_i reflects how that change in treatment affects outcomes.

Visualizing Fuzzy RDD



(Instrument $T_i = \mathbf{1}(x_i \geq x')$)

$$\text{Estimated LATE} = \frac{\text{Jump in } E[y|x]}{\text{Jump in } E[D|x]}$$

Outline

Basic idea of regression discontinuity

Sharp versus fuzzy discontinuities
- Notation & sharp vs. fuzzy assumption
- Assumption about local continuity

Estimating regression discontinuity
- Sharp regression discontinuity
- Graphical analysis
- Fuzzy regression discontinuity

Checks on internal validity

Heterogeneous effects & external validity

Robustness Tests for Internal Validity

- Several diagnostics help verify that the estimated discontinuity truly reflects a causal effect:
  - Visual inspection — plot $E[y|x]$ around x' to confirm a single jump.
  - Check robustness to different polynomial orders $f(\cdot)$.
  - Check robustness to bandwidth choice (narrow vs. wide windows).
- Additional validity checks include:
  - Tests for manipulation of the running variable.
  - Balance tests on predetermined covariates.
  - Falsification (placebo) tests.

Check #1: Manipulation of the Running Variable

- Ask: could individuals or firms **manipulate** the running variable x to end up on one side of the cutoff?
  - If so, the **local continuity assumption** fails — treated and untreated units near x' may no longer be comparable.
- Example:
  - In Keys et al. (QJE, 2010), borrowers might inflate their FICO scores to exceed 620 and obtain better loan terms.
  - Default rates may then jump at 620 even absent the treatment, violating RDD identification.

Why Manipulation Isn't Always Fatal

- Imperfect manipulation does not necessarily invalidate RDD.
- If individuals cannot **precisely control** x, there remains random variation around x':
  - Within a small enough window, some observations fall just above and others just below the cutoff due to idiosyncratic noise.
  - This residual randomness still supports local quasi-random assignment.

Testing for Manipulation

- Look for **bunching** of observations just above or below the cutoff:
  - Excess mass on one side of x' suggests manipulation of x.
- Caveats:
  - Bunching tests assume monotonic manipulation (everyone tries to move in the same direction).
  - In some contexts, incentives may differ, so absence of bunching does not prove validity.

Check #2: Balance Tests

- The RDD assumes observations just above and below x' are comparable.
- Verify this by checking that **other covariates** do not jump at the cutoff:
  - Plot or regress each predetermined variable on x around x'.
  - If covariates are continuous, local comparability holds more credibly.
- Limitation:
  - Continuity in observables does not guarantee continuity in unobservables — these are diagnostic, not proof of validity.

Using Covariates as Controls

- You can include additional covariates in the RDD regression:

$$y_i = \alpha + \beta D_i + f(x_i - x') + \gamma Z_i + u_i,$$

  where Z_i are predetermined covariates.
- If the design is valid:
  - Adding Z_i should not change $\hat{\beta}$ materially — only improve precision.
  - Large changes in $\hat{\beta}$ imply lack of local comparability or inclusion of "bad controls."

Check #3: Falsification (Placebo) Tests

- Conduct placebo tests to ensure the estimated discontinuity is genuine:
  - Examine periods, subgroups, or outcomes where no discontinuity should exist.
  - Example: if a policy cutoff was introduced in 2010, test for jumps in pre-2010 data.
  - Or test unaffected groups (e.g., firms not covered by the law).
- A valid RDD should show **no discontinuity** in these placebo settings.

Outline

Basic idea of regression discontinuity

Sharp versus fuzzy discontinuities
- Notation & sharp vs. fuzzy assumption
- Assumption about local continuity

Estimating regression discontinuity
- Sharp regression discontinuity
- Graphical analysis
- Fuzzy regression discontinuity

Checks on internal validity

Heterogeneous effects & external validity

Heterogeneous Effects (HE) in RDD

- If treatment effects vary with the running variable x, additional assumptions are required for RDD to identify a valid **Local Average Treatment Effect (LATE)**.
- Key assumptions:
  1. **Local continuity of treatment effects:** no jump in the size of the treatment effect at x'.
  2. **Monotonicity:** crossing the threshold weakly increases the probability of treatment.
  3. **Independence:** near x', treatment assignment is unrelated to the potential size of the treatment effect.
- Assumptions (2) and (3) apply only to **fuzzy RDD**.

HE Assumption #1: Local Continuity

- Assumes that the treatment effect itself does not jump at x'.
  - Observations just above and below x' are assumed to be comparable in all relevant dimensions.
- Possible violation:
  - If the threshold x' was intentionally chosen where treatment is expected to have the largest effect.
  - Example: a regulation set at a value where policymakers knew the policy would matter most.

HE Assumption #2: Monotonicity

- Crossing the cutoff should weakly **increase** the probability of treatment:

$$\Pr(D_i = 1 | x_i \geq x') \geq \Pr(D_i = 1 | x_i < x').$$

- Intuition:
  - No group should become *less* likely to receive treatment after crossing x'.
  - This typically holds but should be checked for plausibility in your specific context.

HE Assumption #3: Independence

- Requires that treatment assignment near x' is independent of potential treatment gains.
  - Individuals for whom the treatment has a larger effect cannot systematically sort above or below the cutoff.
  - In practice, this overlaps with the "no manipulation" assumption: sorting based on expected treatment benefits would bias the RDD estimate.

How Heterogeneity Affects Interpretation

- With heterogeneous treatment effects, RDD identifies a **Local Average Treatment Effect (LATE)**:
  - The effect of treatment **around the cutoff**.
  - In a fuzzy design, this applies only to units that **change treatment status** because of crossing x' (the "compliers").
- Implication:
  - Results describe the effect for a specific local subpopulation.
  - External validity is limited—treatment effects may differ for observations far from the threshold.

External Validity and RDD (Part 1)

- RDD identifies effects using observations **close to the cutoff**.
- Treatment effects may differ for units farther from x':
  - Do not generalize beyond the local region of identification.
  - The RDD estimate is local by construction.

External Validity and RDD (Part 2)

- In a **fuzzy RDD**, identification relies on **compliers**—units whose treatment status changes because of crossing x'.
- The estimated effect applies only to this subgroup:
  - Example: studying the effect of earning a PhD on wages using a GRE cutoff.
  - If the cutoff only matters for applicants with average GPAs, the RDD identifies the effect for that subset only.
- As with IV designs, avoid overgeneralizing the LATE beyond the compliant population.

Summary of Today (Part 1)

- The Regression Discontinuity Design (RDD) provides another way to identify causal effects.
- It exploits treatment assignment rules based on a known, arbitrary cutoff rather than randomization.
- RDD is widely applicable in real-world policy and program evaluation settings—and is increasingly used in empirical research.

Summary of Today (Part 2)

- There are two main types of RDD:
  - **Sharp RDD:** treatment status changes deterministically at the cutoff ($D_i = \mathbf{1}(x_i \geq x')$).
  - **Fuzzy RDD:** probability of treatment changes discontinuously at the cutoff (stochastic assignment).
- The fuzzy design can be estimated using a 2SLS framework — it is conceptually identical to an **instrumental variable (IV)** setup.

Summary of Today (Part 3)

- **Internal validity checks:**
  - Graphical analysis of outcomes and treatment around the cutoff.
  - Robustness to bandwidth choice and functional form.
  - Tests for manipulation and covariate balance.

- When treatment effects are heterogeneous, RDD identifies the **Local Average Treatment Effect (LATE)** — the effect for units near the cutoff or, in fuzzy designs, for those who change treatment status because of the threshold.