Frontier Topics in Empirical Economics: Week 10
Regression Discontinuity Design

Zibin Huang [1]

[1] College of Business, Shanghai University of Finance and Economics

November 25, 2024

Introduction

- Assume that we want to examine the education quality of PKU and FDU
- The average wage for PKU graduates is 200,000 RMB/year
- The average wage for FDU graduates is 150,000 RMB/year
- Does this mean that PKU results in higher human capital growth than FDU?

Introduction

- Assume that we want to examine the education quality of PKU and FDU
- The average wage for PKU graduates is 200,000 RMB/year
- The average wage for FDU graduates is 150,000 RMB/year
- Does this mean that PKU results in higher human capital growth than FDU?

Introduction

- Assume that we want to examine the education quality of PKU and FDU
- The average wage for PKU graduates is 200,000 RMB/year
- The average wage for FDU graduates is 150,000 RMB/year
- Does this mean that PKU results in higher human capital growth than FDU?

Introduction

- Assume that we want to examine the education quality of PKU and FDU
- The average wage for PKU graduates is 200,000 RMB/year
- The average wage for FDU graduates is 150,000 RMB/year
- Does this mean that PKU results in higher human capital growth than FDU?

Introduction

- Assume that we want to examine the education quality of PKU and FDU
- The average wage for PKU graduates is 200,000 RMB/year
- The average wage for FDU graduates is 150,000 RMB/year
- Does this mean that PKU results in higher human capital growth than FDU?

Introduction

- No. Since better students select into PKU
- Self-selection is always a problem in economic research
- Is school A more efficient than school B?
- Or just because they admit students with better initial quality?
- How to deal with this issue?

Introduction

- No. Since better students select into PKU
- Self-selection is always a problem in economic research
- Is school A more efficient than school B?
- Or just because they admit students with better initial quality?
- How to deal with this issue?

Introduction

- No. Since better students select into PKU
- Self-selection is always a problem in economic research
- Is school A more efficient than school B?
- Or just because they admit students with better initial quality?
- How to deal with this issue?

Introduction

- No. Since better students select into PKU
- Self-selection is always a problem in economic research
- Is school A more efficient than school B?
- Or just because they admit students with better initial quality?
- How to deal with this issue?

Introduction

- No. Since better students select into PKU
- Self-selection is always a problem in economic research
- Is school A more efficient than school B?
- Or just because they admit students with better initial quality?
- How to deal with this issue?

Introduction

- No. Since better students select into PKU
- Self-selection is always a problem in economic research
- Is school A more efficient than school B?
- Or just because they admit students with better initial quality?
- How to deal with this issue?

Introduction

- Of course you can always construct a selection model structurally
- But there is another design-based approach:
  Regression Discontinuity Design (RDD)
- The intuition for RDD is simple
- Draw PKU students just above the PKU admission line and FDU students just below it
- They are students who enroll in PKU/FDU by chance, thus, similar in ability
- Then compare their results

Introduction

- Of course you can always construct a selection model structurally
- But there is another design-based approach:
  Regression Discontinuity Design (RDD)
- The intuition for RDD is simple
- Draw PKU students just above the PKU admission line and FDU students just below it
- They are students who enroll in PKU/FDU by chance, thus, similar in ability
- Then compare their results

Introduction

- Of course you can always construct a selection model structurally
- But there is another design-based approach:
  Regression Discontinuity Design (RDD)
- The intuition for RDD is simple
- Draw PKU students just above the PKU admission line and FDU students just below it
- They are students who enroll in PKU/FDU by chance, thus, similar in ability
- Then compare their results

Introduction

- Of course you can always construct a selection model structurally
- But there is another design-based approach: Regression Discontinuity Design (RDD)
- The intuition for RDD is simple
- Draw PKU students just above the PKU admission line and FDU students just below it
- They are students who enroll in PKU/FDU by chance, thus, similar in ability
- Then compare their results

Introduction

- Of course you can always construct a selection model structurally
- But there is another design-based approach:
  Regression Discontinuity Design (RDD)
- The intuition for RDD is simple
- Draw PKU students just above the PKU admission line and FDU students just below it
- They are students who enroll in PKU/FDU by chance, thus, similar in ability
- Then compare their results

Introduction

- Of course you can always construct a selection model structurally
- But there is another design-based approach:
  Regression Discontinuity Design (RDD)
- The intuition for RDD is simple
- Draw PKU students just above the PKU admission line and FDU students just below it
- They are students who enroll in PKU/FDU by chance, thus, similar in ability
- Then compare their results

Introduction

- Of course you can always construct a selection model structurally
- But there is another design-based approach:
  Regression Discontinuity Design (RDD)
- The intuition for RDD is simple
- Draw PKU students just above the PKU admission line and FDU students just below it
- They are students who enroll in PKU/FDU by chance, thus, similar in ability
- Then compare their results

Sharp RD

- Let's first consider a simple case: Sharp RD
- In Sharp RD, treatment rule is deterministic
- That is, you are definitely treated if you surpass the threshold
- Conversely, you are definitely not treated
- There is no uncertainty in treatment assignment

Sharp RD

- Let's first consider a simple case: Sharp RD
- In Sharp RD, treatment rule is deterministic
- That is, you are definitely treated if you surpass the threshold
- Conversely, you are definitely not treated
- There is no uncertainty in treatment assignment

Sharp RD

- Let's first consider a simple case: Sharp RD
- In Sharp RD, treatment rule is deterministic
- That is, you are definitely treated if you surpass the threshold
- Conversely, you are definitely not treated
- There is no uncertainty in treatment assignment

Sharp RD

- Let's first consider a simple case: Sharp RD
- In Sharp RD, treatment rule is deterministic
- That is, you are definitely treated if you surpass the threshold
- Conversely, you are definitely not treated
- There is no uncertainty in treatment assignment

Sharp RD

- Let's first consider a simple case: Sharp RD
- In Sharp RD, treatment rule is deterministic
- That is, you are definitely treated if you surpass the threshold
- Conversely, you are definitely not treated
- There is no uncertainty in treatment assignment

Sharp RD

- Let's first consider a simple case: Sharp RD
- In Sharp RD, treatment rule is deterministic
- That is, you are definitely treated if you surpass the threshold
- Conversely, you are definitely not treated
- There is no uncertainty in treatment assignment

Sharp RD

- Suppose that we have treatment D_i determined by some x_i

$$D_i = \mathbf{1}\{x_i \geq x_0\} = \begin{cases} 1, & \text{if } x_i \geq x_0 \\ 0, & \text{if } x_i < x_0 \end{cases}$$

- x_i is called running variable
- x_0 is a known threshold or cutoff
- D_i is a deterministic function of x_i

Sharp RD

- Suppose that we have treatment D_i determined by some x_i

$$D_i = \mathbf{1}\{x_i \geq x_0\} = \begin{cases} 1, & \text{if } x_i \geq x_0 \\ 0, & \text{if } x_i < x_0 \end{cases}$$

- x_i is called running variable
- x_0 is a known threshold or cutoff
- D_i is a deterministic function of x_i

Sharp RD

- Suppose that we have treatment D_i determined by some x_i

$$D_i = \mathbf{1}\{x_i \geq x_0\} = \begin{cases} 1, & \text{if } x_i \geq x_0 \\ 0, & \text{if } x_i < x_0 \end{cases}$$

- x_i is called running variable
- x_0 is a known threshold or cutoff
- D_i is a deterministic function of x_i

Sharp RD

- Suppose that we have treatment D_i determined by some x_i

$$D_i \quad \mathbf{1}\; x_i \; ' \; x_0 \quad \begin{cases} 1; \text{ if } x_i \; ' \; x_0 \\ 0; \text{ if } x_i \; \$ \; x_0 \end{cases}$$

- x_i is called running variable
- x_0 is a known threshold or cutoff
- D_i is a deterministic function of x_i

Sharp RD

- Suppose that we have treatment D_i determined by some x_i

$$D_i = \mathbf{1}\{x_i \geq x_0\} = \begin{cases} 1, & \text{if } x_i \geq x_0 \\ 0, & \text{if } x_i < x_0 \end{cases}$$

- x_i is called running variable
- x_0 is a known threshold or cutoff
- D_i is a deterministic function of x_i

Sharp RD

- We compare samples just above x_0 and just below x_0
- This is a special case of matching
- In conventional matching, we compare samples with identical covariates
- In RD, we compare samples within a small neighborhood at treatment threshold

Sharp RD

- We compare samples just above x_0 and just below x_0
- This is a special case of matching
- In conventional matching, we compare samples with identical covariates
- In RD, we compare samples within a small neighborhood at treatment threshold

Sharp RD

- We compare samples just above x_0 and just below x_0
- This is a special case of matching
- In conventional matching, we compare samples with identical covariates
- In RD, we compare samples within a small neighborhood at treatment threshold

Sharp RD

- We compare samples just above x_0 and just below x_0
- This is a special case of matching
- In conventional matching, we compare samples with identical covariates
- In RD, we compare samples within a small neighborhood at treatment threshold

Sharp RD

- We compare samples just above x_0 and just below x_0
- This is a special case of matching
- In conventional matching, we compare samples with identical covariates
- In RD, we compare samples within a small neighborhood at treatment threshold

Sharp RD

- We can write a simple model for this RD

$$Y_i \quad f_0 \; x_i \; \mathbf{1} \; x_i \; \$ \; x_0 \quad f_1 \; x_i \; \mathbf{1} \; x_i \; ' \; x_0 \quad D_i \quad {}_i$$

- $f_0 \; x_i$ is the smoothing function below the threshold
- $f_1 \; x_i$ is the smoothing function above the threshold
- They are used to fit the trend far away from the cutoff
- D_i is the treatment indicator, jumping at $x_i \quad x_0$

Sharp RD

- We can write a simple model for this RD

$$Y_i \quad f_0 \; x_i \; \mathbf{1} \; x_i \; \$ \; x_0 \quad f_1 \; x_i \; \mathbf{1} \; x_i \; ' \; x_0 \quad D_i \quad {}_i$$

- $f_0 \; x_i$ is the smoothing function below the threshold
- $f_1 \; x_i$ is the smoothing function above the threshold
- They are used to fit the trend far away from the cutoff
- D_i is the treatment indicator, jumping at $x_i \quad x_0$

Sharp RD

- We can write a simple model for this RD

$$Y_i \quad f_0 \; x_i \; \mathbf{1} \; x_i \; \$ \; x_0 \quad f_1 \; x_i \; \mathbf{1} \; x_i \; ' \; x_0 \quad D_i \quad i$$

- $f_0 \; x_i$ is the smoothing function below the threshold
- $f_1 \; x_i$ is the smoothing function above the threshold
- They are used to fit the trend far away from the cutoff
- D_i is the treatment indicator, jumping at $x_i \quad x_0$

Sharp RD

- We can write a simple model for this RD

$$Y_i \quad f_0 \; x_i \; \mathbf{1} \; x_i \; \$ \; x_0 \quad f_1 \; x_i \; \mathbf{1} \; x_i \; ' \; x_0 \quad D_i \quad i$$

- $f_0 \; x_i$ is the smoothing function below the threshold
- $f_1 \; x_i$ is the smoothing function above the threshold
- They are used to fit the trend far away from the cutoff
- D_i is the treatment indicator, jumping at $x_i \quad x_0$

Sharp RD

- We can write a simple model for this RD

$$Y_i \quad f_0 \; x_i \; \mathbf{1} \; x_i \; \$ \; x_0 \quad f_1 \; x_i \; \mathbf{1} \; x_i \; ' \; x_0 \quad D_i \quad {}_i$$

- $f_0 \; x_i$ is the smoothing function below the threshold
- $f_1 \; x_i$ is the smoothing function above the threshold
- They are used to fit the trend far away from the cutoff
- D_i is the treatment indicator, jumping at $x_i \quad x_0$

Sharp RD

- We can write a simple model for this RD

$$Y_i \quad f_0 \;\; x_i \;\; \mathbf{1} \;\; x_i \; \$ \; x_0 \quad f_1 \;\; x_i \;\; \mathbf{1} \;\; x_i \; ' \; x_0 \qquad D_i \qquad {}_i$$

- $f_0 \;\; x_i$ is the smoothing function below the threshold
- $f_1 \;\; x_i$ is the smoothing function above the threshold
- They are used to fit the trend far away from the cutoff
- D_i is the treatment indicator, jumping at $x_i \quad x_0$

Sharp RD

- We can choose different smoothing functions for f_0 and f_1
- The simplest ones are linear and quadratic functions

Sharp RD

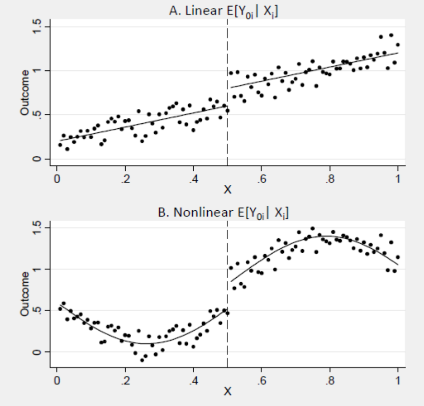
- We can choose different smoothing functions for f_0 and f_1
- The simplest ones are linear and quadratic functions

Sharp RD

- We can choose different smoothing functions for f_0 and f_1
- The simplest ones are linear and quadratic functions

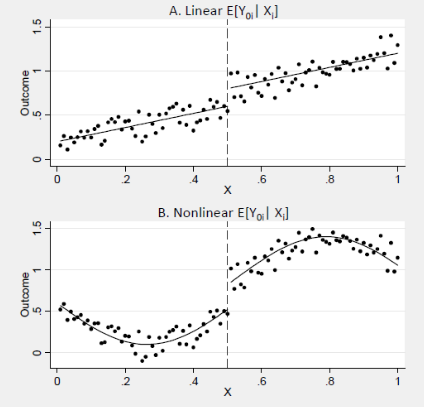
Sharp RD

- Here are two examples from Angrist and Pischke (2009), Page 255



Sharp RD

- Here are two examples from Angrist and Pischke (2009), Page 255



Sharp RD

- We can also use non-parametric and semi-parametric functions introduced in Week 2 lecture, which are more flexible
- The most recommended and commonly used one is the Local Linear/Quadratic Regression
- As we have discussed, there is a bias-variance tradeoff
- If you choose complicated smoothing function, you may lose your accuracy
- If you choose too simple smoothing function, you may get bias

Sharp RD

- We can also use non-parametric and semi-parametric functions introduced in Week 2 lecture, which are more flexible
- The most recommended and commonly used one is the Local Linear/Quadratic Regression
- As we have discussed, there is a bias-variance tradeoff
- If you choose complicated smoothing function, you may lose your accuracy
- If you choose too simple smoothing function, you may get bias

Sharp RD

- We can also use non-parametric and semi-parametric functions introduced in Week 2 lecture, which are more flexible
- The most recommended and commonly used one is the Local Linear/Quadratic Regression
- As we have discussed, there is a bias-variance tradeoff
- If you choose complicated smoothing function, you may lose your accuracy
- If you choose too simple smoothing function, you may get bias

Sharp RD

- We can also use non-parametric and semi-parametric functions introduced in Week 2 lecture, which are more flexible
- The most recommended and commonly used one is the Local Linear/Quadratic Regression
- As we have discussed, there is a bias-variance tradeoff
- If you choose complicated smoothing function, you may lose your accuracy
- If you choose too simple smoothing function, you may get bias

Sharp RD

- We can also use non-parametric and semi-parametric functions introduced in Week 2 lecture, which are more flexible
- The most recommended and commonly used one is the Local Linear/Quadratic Regression
- As we have discussed, there is a bias-variance tradeoff
- If you choose complicated smoothing function, you may lose your accuracy
- If you choose too simple smoothing function, you may get bias

Sharp RD

- We can also use non-parametric and semi-parametric functions introduced in Week 2 lecture, which are more flexible
- The most recommended and commonly used one is the Local Linear/Quadratic Regression
- As we have discussed, there is a bias-variance tradeoff
- If you choose complicated smoothing function, you may lose your accuracy
- If you choose too simple smoothing function, you may get bias

Sharp RD

- But remember, effective sample size is usually limited in RD
- You are effectively using a small neighborhood around the cutoff
- So, do not use too complicated smoothing models
- Specifically, Gelman and Imbens (2019) claim that you should avoid using high-order polynomial (over third order)
  - It leads to noisy estimates (Runge's phenomenon)
  - RDD is very sensitive to the degree of the polynomial
  - Coverage of confidence intervals is smaller than nominal

Sharp RD

- But remember, effective sample size is usually limited in RD
- You are effectively using a small neighborhood around the cutoff
- So, do not use too complicated smoothing models
- Specifically, Gelman and Imbens (2019) claim that you should avoid using high-order polynomial (over third order)
  - It leads to noisy estimates (Runge's phenomenon)
  - RDD is very sensitive to the degree of the polynomial
  - Coverage of confidence intervals is smaller than nominal

Sharp RD

- But remember, effective sample size is usually limited in RD
- You are effectively using a small neighborhood around the cutoff
- So, do not use too complicated smoothing models
- Specifically, Gelman and Imbens (2019) claim that you should avoid using high-order polynomial (over third order)
    - It leads to noisy estimates (Runge's phenomenon)
    - RDD is very sensitive to the degree of the polynomial
    - Coverage of confidence intervals is smaller than nominal

Sharp RD

- But remember, effective sample size is usually limited in RD
- You are effectively using a small neighborhood around the cutoff
- So, do not use too complicated smoothing models
- Specifically, Gelman and Imbens (2019) claim that you should avoid using high-order polynomial (over third order)
  - It leads to noisy estimates (Runge's phenomenon)
  - RDD is very sensitive to the degree of the polynomial
  - Coverage of confidence intervals is smaller than nominal

Sharp RD

- But remember, effective sample size is usually limited in RD
- You are effectively using a small neighborhood around the cutoff
- So, do not use too complicated smoothing models
- Specifically, Gelman and Imbens (2019) claim that you should avoid using high-order polynomial (over third order)
  - It leads to noisy estimates (Runge's phenomenon)
  - RDD is very sensitive to the degree of the polynomial
  - Coverage of confidence intervals is smaller than nominal

Sharp RD

- But remember, effective sample size is usually limited in RD
- You are effectively using a small neighborhood around the cutoff
- So, do not use too complicated smoothing models
- Specifically, Gelman and Imbens (2019) claim that you should avoid using high-order polynomial (over third order)
    - It leads to noisy estimates (Runge's phenomenon)
    - RDD is very sensitive to the degree of the polynomial
    - Coverage of confidence intervals is smaller than nominal

Sharp RD

- But remember, effective sample size is usually limited in RD
- You are effectively using a small neighborhood around the cutoff
- So, do not use too complicated smoothing models
- Specifically, Gelman and Imbens (2019) claim that you should avoid using high-order polynomial (over third order)
  - It leads to noisy estimates (Runge's phenomenon)
  - RDD is very sensitive to the degree of the polynomial
  - Coverage of confidence intervals is smaller than nominal

Sharp RD

- But remember, effective sample size is usually limited in RD
- You are effectively using a small neighborhood around the cutoff
- So, do not use too complicated smoothing models
- Specifically, Gelman and Imbens (2019) claim that you should avoid using high-order polynomial (over third order)
  - It leads to noisy estimates (Runge's phenomenon)
  - RDD is very sensitive to the degree of the polynomial
  - Coverage of confidence intervals is smaller than nominal

Sharp RD

- An interesting example of Sharp RD is Lee (2008)
- What is the advantage for the party incumbency on reelection?
- Hard to identify since a party may have larger group of supporters for many reasons other than incumbency
- Blue state vs. Red state vs. Swing state

Sharp RD

- An interesting example of Sharp RD is Lee (2008)
- What is the advantage for the party incumbency on reelection?
- Hard to identify since a party may have larger group of supporters for many reasons other than incumbency
- Blue state vs. Red state vs. Swing state

Sharp RD

- An interesting example of Sharp RD is Lee (2008)
- What is the advantage for the party incumbency on reelection?
- Hard to identify since a party may have larger group of supporters for many reasons other than incumbency
- Blue state vs. Red state vs. Swing state

Sharp RD

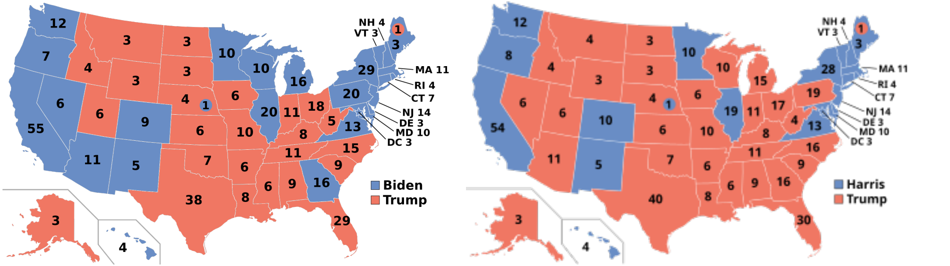
- An interesting example of Sharp RD is Lee (2008)
- What is the advantage for the party incumbency on reelection?
- Hard to identify since a party may have larger group of supporters for many reasons other than incumbency
- Blue state vs. Red state vs. Swing state

Sharp RD

- An interesting example of Sharp RD is Lee (2008)
- What is the advantage for the party incumbency on reelection?
- Hard to identify since a party may have larger group of supporters for many reasons other than incumbency
- Blue state vs. Red state vs. Swing state

Sharp RD

- Different parties are advantaged in different regions due to ideology, history, religion... reasons

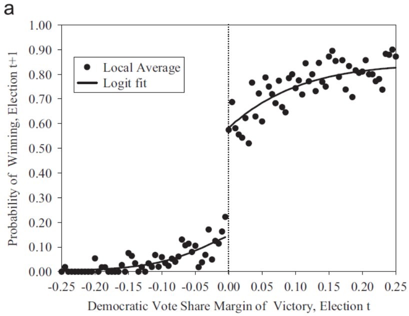


(a) 2020 (b) 2024

Figure: U.S. General Election Map

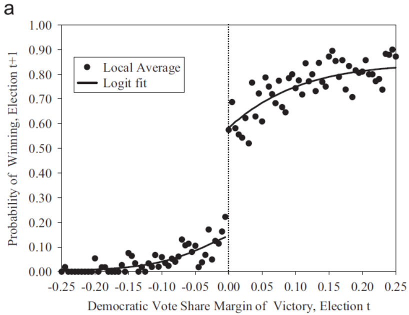
Sharp RD

- But for elections with very close results, winners and losers are similar
- Lee (2008) considers the probability of Democratic winning in regions where Democratic candidates won by small shares



Sharp RD

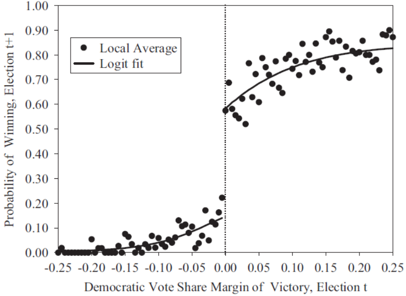
- But for elections with very close results, winners and losers are similar
- Lee (2008) considers the probability of Democratic winning in regions where Democratic candidates won by small shares



Sharp RD

- But for elections with very close results, winners and losers are similar
- Lee (2008) considers the probability of Democratic winning in regions where Democratic candidates won by small shares

a



Fuzzy RD

- A more complicated case is Fuzzy RD
- In Fuzzy RD, treatment assignment is no longer deterministic
- There is uncertainty in being treated or not
- By passing the threshold, you have larger probability to get treated

Fuzzy RD

- A more complicated case is Fuzzy RD
- In Fuzzy RD, treatment assignment is no longer deterministic
- There is uncertainty in being treated or not
- By passing the threshold, you have larger probability to get treated

Fuzzy RD

- A more complicated case is Fuzzy RD
- In Fuzzy RD, treatment assignment is no longer deterministic
- There is uncertainty in being treated or not
- By passing the threshold, you have larger probability to get treated

Fuzzy RD

- A more complicated case is Fuzzy RD
- In Fuzzy RD, treatment assignment is no longer deterministic
- There is uncertainty in being treated or not
- By passing the threshold, you have larger probability to get treated

Fuzzy RD

- A more complicated case is Fuzzy RD
- In Fuzzy RD, treatment assignment is no longer deterministic
- There is uncertainty in being treated or not
- By passing the threshold, you have larger probability to get treated

Fuzzy RD

- Discontinuity in treatment probability, but not treatment

$$P(D_i = 1 | x_i) = \begin{cases} g_1(x_i) & \text{if } x_i \geq x_0 \\ g_0(x_i) & \text{if } x_i < x_0 \end{cases}; \text{ where } g_1(x_0) \neq g_0(x_0)$$

- Let's assume that $g_1(x_0) > g_0(x_0)$ (WLOG)
- Thus, surpassing the cutoff makes treatment more likely

Fuzzy RD

- Discontinuity in treatment probability, but not treatment

$$P(D_i = 1|x_i) = \begin{cases} g_1(x_i) & \text{if } x_i \geq x_0 \\ g_0(x_i) & \text{if } x_i < x_0 \end{cases}; \text{ where } g_1(x_0) \neq g_0(x_0)$$

- Let's assume that $g_1(x_0) > g_0(x_0)$ (WLOG)
- Thus, surpassing the cutoff makes treatment more likely

Fuzzy RD

- Discontinuity in treatment probability, but not treatment

$$P(D_i = 1 \mid x_i) = \begin{cases} g_1(x_i) & \text{if } x_i \geq x_0 \\ g_0(x_i) & \text{if } x_i < x_0 \end{cases}, \text{ where } g_1(x_0) \neq g_0(x_0)$$

- Let's assume that $g_1(x_0) > g_0(x_0)$ (WLOG)
- Thus, surpassing the cutoff makes treatment more likely

Fuzzy RD

- Discontinuity in treatment probability, but not treatment

$$P(D_i = 1|x_i) = \begin{cases} g_1(x_i) & \text{if } x_i \geq x_0 \\ g_0(x_i) & \text{if } x_i < x_0 \end{cases}; \text{ where } g_1(x_0) \neq g_0(x_0)$$

- Let's assume that $g_1(x_0) > g_0(x_0)$ (WLOG)
- Thus, surpassing the cutoff makes treatment more likely

Fuzzy RD

- Denote $T_i = \mathbf{1}(x_i \geq x_0)$ as the indicator of whether passing the cutoff
- Then, we can naturally write Fuzzy RD as a 2SLS
- Treatment D_i is endogenous variable, cutoff indicator T_i is instrument
  - First stage: treatment D_i on cutoff indicator T_i
  - Second stage: outcome variable on first stage fitted value
- The smoothing function f should be included in both stages
- Very simple to implement RD in Stata: Packages such as *rdrobust*
- It helps you to implement bias-corrected CI with optimal bandwidth in Calonico, Cattaneo, and Titiunik (2014)
- You can also try optimal bandwidth in Imbens and Kalyanaraman (2012)

Fuzzy RD

- Denote $T_i = \mathbf{1}(x_i \geq x_0)$ as the indicator of whether passing the cutoff
- Then, we can naturally write Fuzzy RD as a 2SLS
- Treatment D_i is endogenous variable, cutoff indicator T_i is instrument
  - First stage: treatment D_i on cutoff indicator T_i
  - Second stage: outcome variable on first stage fitted value
- The smoothing function f should be included in both stages
- Very simple to implement RD in Stata: Packages such as *rdrobust*
- It helps you to implement bias-corrected CI with optimal bandwidth in Calonico, Cattaneo, and Titiunik (2014)
- You can also try optimal bandwidth in Imbens and Kalyanaraman (2012)

Fuzzy RD

- Denote $T_i = \mathbf{1}(x_i \geq x_0)$ as the indicator of whether passing the cutoff
- Then, we can naturally write Fuzzy RD as a 2SLS
- Treatment D_i is endogenous variable, cutoff indicator T_i is instrument
  - First stage: treatment D_i on cutoff indicator T_i
  - Second stage: outcome variable on first stage fitted value
- The smoothing function f should be included in both stages
- Very simple to implement RD in Stata: Packages such as *rdrobust*
- It helps you to implement bias-corrected CI with optimal bandwidth in Calonico, Cattaneo, and Titiunik (2014)
- You can also try optimal bandwidth in Imbens and Kalyanaraman (2012)

Fuzzy RD

- Denote $T_i = \mathbf{1}(x_i \geq x_0)$ as the indicator of whether passing the cutoff
- Then, we can naturally write Fuzzy RD as a 2SLS
- Treatment D_i is endogenous variable, cutoff indicator T_i is instrument
  - First stage: treatment D_i on cutoff indicator T_i
  - Second stage: outcome variable on first stage fitted value
- The smoothing function f should be included in both stages
- Very simple to implement RD in Stata: Packages such as *rdrobust*
- It helps you to implement bias-corrected CI with optimal bandwidth in Calonico, Cattaneo, and Titiunik (2014)
- You can also try optimal bandwidth in Imbens and Kalyanaraman (2012)

Fuzzy RD

- Denote $T_i \quad \mathbf{1} \; x_i \; ' \; x_0$ as the indicator of whether passing the cutoff
- Then, we can naturally write Fuzzy RD as a 2SLS
- Treatment D_i is endogenous variable, cutoff indicator T_i is instrument
  - First stage: treatment D_i on cutoff indicator T_i
  - Second stage: outcome variable on first stage fitted value
- The smoothing function f should be included in both stages
- Very simple to implement RD in Stata: Packages such as *rdrobust*
- It helps you to implement bias-corrected CI with optimal bandwidth in Calonico, Cattaneo, and Titiunik (2014)
- You can also try optimal bandwidth in Imbens and Kalyanaraman (2012)

Fuzzy RD

- Denote $T_i = \mathbf{1}(x_i \geq x_0)$ as the indicator of whether passing the cutoff
- Then, we can naturally write Fuzzy RD as a 2SLS
- Treatment D_i is endogenous variable, cutoff indicator T_i is instrument
  - First stage: treatment D_i on cutoff indicator T_i
  - Second stage: outcome variable on first stage fitted value
- The smoothing function f should be included in both stages
- Very simple to implement RD in Stata: Packages such as *rdrobust*
- It helps you to implement bias-corrected CI with optimal bandwidth in Calonico, Cattaneo, and Titiunik (2014)
- You can also try optimal bandwidth in Imbens and Kalyanaraman (2012)

Fuzzy RD

- Denote $T_i \quad \mathbf{1} \; x_i \; ' \; x_0$ as the indicator of whether passing the cutoff
- Then, we can naturally write Fuzzy RD as a 2SLS
- Treatment D_i is endogenous variable, cutoff indicator T_i is instrument
  - First stage: treatment D_i on cutoff indicator T_i
  - Second stage: outcome variable on first stage fitted value
- The smoothing function f should be included in both stages
- Very simple to implement RD in Stata: Packages such as *rdrobust*
- It helps you to implement bias-corrected CI with optimal bandwidth in Calonico, Cattaneo, and Titiunik (2014)
- You can also try optimal bandwidth in Imbens and Kalyanaraman (2012)

Fuzzy RD

- Denote $T_i = \mathbf{1}\{x_i \geq x_0\}$ as the indicator of whether passing the cutoff
- Then, we can naturally write Fuzzy RD as a 2SLS
- Treatment D_i is endogenous variable, cutoff indicator T_i is instrument
  - First stage: treatment D_i on cutoff indicator T_i
  - Second stage: outcome variable on first stage fitted value
- The smoothing function f should be included in both stages
- Very simple to implement RD in Stata: Packages such as *rdrobust*
- It helps you to implement bias-corrected CI with optimal bandwidth in Calonico, Cattaneo, and Titiunik (2014)
- You can also try optimal bandwidth in Imbens and Kalyanaraman (2012)

Fuzzy RD

- Denote $T_i = \mathbf{1}(x_i \geq x_0)$ as the indicator of whether passing the cutoff
- Then, we can naturally write Fuzzy RD as a 2SLS
- Treatment D_i is endogenous variable, cutoff indicator T_i is instrument
  - First stage: treatment D_i on cutoff indicator T_i
  - Second stage: outcome variable on first stage fitted value
- The smoothing function f should be included in both stages
- Very simple to implement RD in Stata: Packages such as *rdrobust*
- It helps you to implement bias-corrected CI with optimal bandwidth in Calonico, Cattaneo, and Titiunik (2014)
- You can also try optimal bandwidth in Imbens and Kalyanaraman (2012)

Fuzzy RD

- Denote $T_i \quad \mathbf{1}\ x_i\ '\ x_0$ as the indicator of whether passing the cutoff
- Then, we can naturally write Fuzzy RD as a 2SLS
- Treatment D_i is endogenous variable, cutoff indicator T_i is instrument
  - First stage: treatment D_i on cutoff indicator T_i
  - Second stage: outcome variable on first stage fitted value
- The smoothing function f should be included in both stages
- Very simple to implement RD in Stata: Packages such as *rdrobust*
- It helps you to implement bias-corrected CI with optimal bandwidth in Calonico, Cattaneo, and Titiunik (2014)
- You can also try optimal bandwidth in Imbens and Kalyanaraman (2012)

Non-parametric Identification of RD

- We have already introduced how to implement RD method
- And intuitively discussed its identification source
- But what kind of causal effect we are identifying?
- What exactly are its identification assumptions?

Non-parametric Identification of RD

- We have already introduced how to implement RD method
- And intuitively discussed its identification source
- But what kind of causal effect we are identifying?
- What exactly are its identification assumptions?

Non-parametric Identification of RD

- We have already introduced how to implement RD method
- And intuitively discussed its identification source
- But what kind of causal effect we are identifying?
- What exactly are its identification assumptions?

Non-parametric Identification of RD

- We have already introduced how to implement RD method
- And intuitively discussed its identification source
- But what kind of causal effect we are identifying?
- What exactly are its identification assumptions?

Non-parametric Identification of RD

- We have already introduced how to implement RD method
- And intuitively discussed its identification source
- But what kind of causal effect we are identifying?
- What exactly are its identification assumptions?

Non-parametric Identification of RD

- Let's go to a classic study in RDD, Hahn, Todd, and Van der Klaauw (2001)
- Do not say that you understand RDD if you never read this paper

Non-parametric Identification of RD

- Let's go to a classic study in RDD, Hahn, Todd, and Van der Klaauw (2001)
- Do not say that you understand RDD if you never read this paper

Non-parametric Identification of RD

- Let's go to a classic study in RDD, Hahn, Todd, and Van der Klaauw (2001)
- Do not say that you understand RDD if you never read this paper

Non-parametric Identification of RD

- Denote $y_{1i}; y_{0i}$ as the potential outcomes, x_i as the treatment
- We have an outcome y_i
- Thus, y_{0i}; y_{1i} y_{0i}
- Assume that we have a running variable z_i

Non-parametric Identification of RD

- Denote $y_{1i}; y_{0i}$ as the potential outcomes, x_i as the treatment
- We have an outcome y_i $_i$ x_i $_i$
- Thus, $_i$ $y_{0i};$ $_i$ y_{1i} y_{0i}
- Assume that we have a running variable z_i
  - In Sharp design, we have x_i f z_i discontinuous at z_0
  - In Fuzzy design, we have P x_i $1¶z_i$ f z_i discontinuous at z_0

Assumption (Fuzzy RD) in Hahn, Todd, and Van der Klaauw (2001)

(i) The limits x $\lim_{z \quad z_0} E$ $x_i¶z_i$ z and x $\lim_{z \quad z_0} E$ $x_i¶z_i$ z exist; (ii) x j x

Non-parametric Identification of RD

- Denote $y_{1i}; y_{0i}$ as the potential outcomes, x_i as the treatment
- We have an outcome $y_i \quad {}_i \quad x_i \quad {}_i$
- Thus, ${}_i \quad y_{0i}; \; {}_i \quad y_{1i} \quad y_{0i}$
- Assume that we have a running variable z_i
    - In Sharp design, we have $x_i \quad f \; z_i$ discontinuous at z_0
    - In Fuzzy design, we have $P \; x_i \quad 1 ¶ z_i \quad f \; z_i$ discontinuous at z_0

Assumption (Fuzzy RD) in Hahn, Todd, and Van der Klaauw (2001)

(i) The limits $x \quad \lim_{z \; z_0} E \; x_i ¶ z_i \quad z$ and $x \quad \lim_{z \; z_0} E \; x_i ¶ z_i \quad z$ exist; (ii) $x \quad j \; x$

Non-parametric Identification of RD

- Denote $y_{1i}; y_{0i}$ as the potential outcomes, x_i as the treatment
- We have an outcome $y_i \quad _i \quad x_i \quad _i$
- Thus, $_i \quad y_{0i}; \; _i \quad y_{1i} \quad y_{0i}$
- Assume that we have a running variable z_i
  - In Sharp design, we have $x_i \quad f \; z_i$ discontinuous at z_0
  - In Fuzzy design, we have $P \; x_i \quad 1 | z_i \quad f \; z_i$ discontinuous at z_0

Assumption (Fuzzy RD) in Hahn, Todd, and Van der Klaauw (2001)

(i) The limits $x \quad \lim_{z \; z_0} E \; x_i | z_i \quad z$ and $x \quad \lim_{z \; z_0} E \; x_i | z_i \quad z$ exist; (ii) $x \quad j \; x$

Non-parametric Identification of RD

- Denote $y_{1i}; y_{0i}$ as the potential outcomes, x_i as the treatment
- We have an outcome $y_i \quad {}_i \quad x_i \quad {}_i$
- Thus, ${}_i \quad y_{0i}; \; {}_i \quad y_{1i} \quad y_{0i}$
- Assume that we have a running variable z_i
  - In Sharp design, we have $x_i \quad f \; z_i$ discontinuous at z_0
  - In Fuzzy design, we have $P \; x_i \quad 1¶z_i \quad f \; z_i$ discontinuous at z_0

Assumption (Fuzzy RD) in Hahn, Todd, and Van der Klaauw (2001)

(i) The limits $x \quad \lim_{z \quad z_0} E \; x_i¶z_i \quad z$ and $x \quad \lim_{z \quad z_0} E \; x_i¶z_i \quad z$ exist; (ii) $x \quad j \; x$

Non-parametric Identification of RD

- Denote $y_{1i}; y_{0i}$ as the potential outcomes, x_i as the treatment
- We have an outcome $y_i \quad _i \quad x_i \quad _i$
- Thus, $_i \quad y_{0i}; \ _i \quad y_{1i} \quad y_{0i}$
- Assume that we have a running variable z_i
  - In Sharp design, we have $x_i \quad f \ z_i$ discontinuous at z_0
  - In Fuzzy design, we have $P \ x_i \quad 1¶z_i \quad f \ z_i$ discontinuous at z_0

Assumption (Fuzzy RD) in Hahn, Todd, and Van der Klaauw (2001)

(i) The limits $x \quad \lim_{z \quad z_0} E \ x_i¶z_i \quad z$ and $x \quad \lim_{z \quad z_0} E \ x_i¶z_i \quad z$ exist; (ii) $x \quad j \ x$

Non-parametric Identification of RD

- Denote $y_{1i}; y_{0i}$ as the potential outcomes, x_i as the treatment
- We have an outcome $y_i \quad _i \quad x_i \quad _i$
- Thus, $_i \quad y_{0i}; \ _i \quad y_{1i} \quad y_{0i}$
- Assume that we have a running variable z_i
  - In Sharp design, we have $x_i \quad f \ z_i$ discontinuous at z_0
  - In Fuzzy design, we have $P \ x_i \quad 1¶z_i \quad f \ z_i$ discontinuous at z_0

Assumption (Fuzzy RD) in Hahn, Todd, and Van der Klaauw (2001)

(i) The limits $x \quad \lim_{z \quad z_0} E \ x_i¶z_i \quad z$ and $x \quad \lim_{z \quad z_0} E \ x_i¶z_i \quad z$ exist; (ii) $x \quad j \ x$

Non-parametric Identification of RD

- Denote $y_{1i}; y_{0i}$ as the potential outcomes, x_i as the treatment
- We have an outcome $y_i \quad _i \quad x_i \quad _i$
- Thus, $_i \quad y_{0i}; \; _i \quad y_{1i} \quad y_{0i}$
- Assume that we have a running variable z_i
  - In Sharp design, we have $x_i \quad f \; z_i$ discontinuous at z_0
  - In Fuzzy design, we have $P \; x_i \quad 1¶z_i \quad f \; z_i$ discontinuous at z_0

Assumption (Fuzzy RD) in Hahn, Todd, and Van der Klaauw (2001)

(i) The limits $x \quad \lim_{z \quad z_0} E \; x_i¶z_i \quad z$ and $x \quad \lim_{z \quad z_0} E \; x_i¶z_i \quad z$ exist; (ii) $x \quad j \; x$

Non-parametric Identification of RD

- First, consider the simple case of constant treatment effects
- i across individuals
- Assume that mean untreated potential outcome is continuous at the cutoff
- That is, mean of other confounders is continuous at the cutoff

Assumption (A1) in Hahn, Todd, and Van der Klaauw (2001)

$E \quad i|z_i \quad z$ is continuous in z at z_0

Non-parametric Identification of RD

- First, consider the simple case of constant treatment effects
- $_i$ across individuals
- Assume that mean untreated potential outcome is continuous at the cutoff
- That is, mean of other confounders is continuous at the cutoff

Assumption (A1) in Hahn, Todd, and Van der Klaauw (2001)

$E \; {}_i | z_i \; z$ is continuous in z at z_0

Non-parametric Identification of RD

- First, consider the simple case of constant treatment effects
- i across individuals
- Assume that mean untreated potential outcome is continuous at the cutoff
- That is, mean of other confounders is continuous at the cutoff

Assumption (A1) in Hahn, Todd, and Van der Klaauw (2001)

E $_i$¶z_i z is continuous in z at z_0

Non-parametric Identification of RD

- First, consider the simple case of constant treatment effects
- $_i$ across individuals
- Assume that mean untreated potential outcome is continuous at the cutoff
- That is, mean of other confounders is continuous at the cutoff

Assumption (A1) in Hahn, Todd, and Van der Klaauw (2001)

$E \quad _i|z_i \quad z$ is continuous in z at z_0

Non-parametric Identification of RD

- First, consider the simple case of constant treatment effects
- i across individuals
- Assume that mean untreated potential outcome is continuous at the cutoff
- That is, mean of other confounders is continuous at the cutoff

Assumption (A1) in Hahn, Todd, and Van der Klaauw (2001)

$E[\ _i | z_i = z]$ is continuous in z at z_0

Non-parametric Identification of RD

- First, consider the simple case of constant treatment effects
- $_i$ across individuals
- Assume that mean untreated potential outcome is continuous at the cutoff
- That is, mean of other confounders is continuous at the cutoff

Assumption (A1) in Hahn, Todd, and Van der Klaauw (2001)

$E \quad {}_i \P z_i \quad z$ is continuous in z at z_0

Non-parametric Identification of RD

- We can prove that is non-parametrically identified

Theorem 1 in Hahn, Todd, and Van der Klaauw (2001)

Suppose that is fixed at . Further suppose that Assumptions (RD) and (A1) hold. We then have: $\frac{y^+ - y^-}{x^+ - x^-}$, where $y^+ = \lim_{z \to z_0^+} E[y_i | z_i = z]$ and $y^- = \lim_{z \to z_0^-} E[y_i | z_i = z]$

- Using an IV-style method, we can pin down the treatment effect

Non-parametric Identification of RD

- We can prove that is non-parametrically identified

Theorem 1 in Hahn, Todd, and Van der Klaauw (2001)

Suppose that $_i$ is fixed at . Further suppose that Assumptions (RD) and (A1) hold. We then have: $\frac{y \quad y}{x \quad x}$, where $y \quad \lim_{z \quad z_0} E \; y_i \P z_i \quad z$ and $y \quad \lim_{z \quad z_0} E \; y_i \P z_i \quad z$

- Using an IV-style method, we can pin down the treatment effect

Non-parametric Identification of RD

- We can prove that is non-parametrically identified

Theorem 1 in Hahn, Todd, and Van der Klaauw (2001)

Suppose that $_i$ is fixed at . Further suppose that Assumptions (RD) and (A1) hold. We then have: $\frac{y \quad y}{x \quad x}$, where $y \quad \lim_{z \quad z_0} E \; y_i ¶ z_i \quad z$ and $y \quad \lim_{z \quad z_0} E \; y_i ¶ z_i \quad z$

- Using an IV-style method, we can pin down the treatment effect

Non-parametric Identification of RD

- We can prove that is non-parametrically identified

Theorem 1 in Hahn, Todd, and Van der Klaauw (2001)

Suppose that $_i$ is fixed at . Further suppose that Assumptions (RD) and (A1) hold. We then have: $\frac{y \quad y}{x \quad x}$, where $y \quad \lim_{z \quad z_0} E \ y_i ¶ z_i \quad z$ and $y \quad \lim_{z \quad z_0} E \ y_i ¶ z_i \quad z$

- Using an IV-style method, we can pin down the treatment effect

Non-parametric Identification of RD

- Next, we go to more complicated heterogeneous treatment effect case
- We need one more assumption, not only is continuous at z_0, but also

Non-parametric Identification of RD

- Next, we go to more complicated heterogeneous treatment effect case
- We need one more assumption, not only is continuous at z_0, but also

Assumption (A2) in Hahn, Todd, and Van der Klaauw (2001)

$E \quad _i \P z_i \quad z$ is continuous at $z \quad z_0$

Non-parametric Identification of RD

- Next, we go to more complicated heterogeneous treatment effect case
- We need one more assumption, not only is continuous at z_0, but also

Assumption (A2) in Hahn, Todd, and Van der Klaauw (2001)

E $_i$¶z_i z is continuous at z z_0

Non-parametric Identification of RD

- Next, we go to more complicated heterogeneous treatment effect case
- We need one more assumption, not only is continuous at z_0, but also

Assumption (A2) in Hahn, Todd, and Van der Klaauw (2001)

E ${}_i$¶z_i z is continuous at z z_0

Non-parametric Identification of RD

- A1 and A2 are different
- A1 is an exogeneity assumption; A2 is a no treatment sorting assumption
- A1 says there is no systematic difference in y_{0i} around the cutoff
- A2 says there is no systematic difference in $y_{1i} - y_{0i}$
- Violation examples:
  - A1 : If students with very high ability can control their scores to be just above the cutoff line
  - A2 : If students with high return are more likely to select into treatment

Non-parametric Identification of RD

- A1 and A2 are different
- A1 is an exogeneity assumption; A2 is a no treatment sorting assumption
- A1 says there is no systematic difference in y_{0i} around the cutoff
- A2 says there is no systematic difference in $y_{1i} - y_{0i}$
- Violation examples:
  - A1: If students with very high ability can control their scores to be just above the cutoff line
  - A2: If students with high return are more likely to select into treatment

Non-parametric Identification of RD

- A1 and A2 are different
- A1 is an exogeneity assumption; A2 is a no treatment sorting assumption
- A1 says there is no systematic difference in y_{0i} around the cutoff
- A2 says there is no systematic difference in $y_{1i} - y_{0i}$
- Violation examples:
  - A1: If students with very high ability can control their scores to be just above the cutoff line
  - A2: If students with high return are more likely to select into treatment

Non-parametric Identification of RD

- A1 and A2 are different
- A1 is an exogeneity assumption; A2 is a no treatment sorting assumption
- A1 says there is no systematic difference in y_{0i} around the cutoff
- A2 says there is no systematic difference in $y_{1i} \quad y_{0i}$
- Violation examples:
  - A1: If students with very high ability can control their scores to be just above the cutoff line
  - A2: If students with high return are more likely to select into treatment

Non-parametric Identification of RD

- A1 and A2 are different
- A1 is an exogeneity assumption; A2 is a no treatment sorting assumption
- A1 says there is no systematic difference in y_{0i} around the cutoff
- A2 says there is no systematic difference in y_{1i} y_{0i}
- Violation examples:
  - A1: If students with very high ability can control their scores to be just above the cutoff line
  - A2: If students with high return are more likely to select into treatment

Non-parametric Identification of RD

- A1 and A2 are different
- A1 is an exogeneity assumption; A2 is a no treatment sorting assumption
- A1 says there is no systematic difference in y_{0i} around the cutoff
- A2 says there is no systematic difference in $y_{1i} \quad y_{0i}$
- Violation examples:
  - A1: If students with very high ability can control their scores to be just above the cutoff line
  - A2: If students with high return are more likely to select into treatment

Non-parametric Identification of RD

- A1 and A2 are different
- A1 is an exogeneity assumption; A2 is a no treatment sorting assumption
- A1 says there is no systematic difference in y_{0i} around the cutoff
- A2 says there is no systematic difference in $y_{1i} \quad y_{0i}$
- Violation examples:
    - A1: If students with very high ability can control their scores to be just above the cutoff line
    - A2: If students with high return are more likely to select into treatment

Non-parametric Identification of RD

- A1 and A2 are different
- A1 is an exogeneity assumption; A2 is a no treatment sorting assumption
- A1 says there is no systematic difference in y_{0i} around the cutoff
- A2 says there is no systematic difference in $y_{1i} \quad y_{0i}$
- Violation examples:
  - A1: If students with very high ability can control their scores to be just above the cutoff line
  - A2: If students with high return are more likely to select into treatment

Non-parametric Identification of RD

- Then we have the following result

Theorem 2 in Hahn, Todd, and Van der Klaauw (2001)

Suppose that x_i is independent of β_i conditional on z_i near z_0. Further suppose that Assumptions (RD), (A1), and (A2) hold. We then have: $E[\beta_i | z_i = z_0] = \frac{y^+ - y^-}{x^+ - x^-}$

Non-parametric Identification of RD

- Then we have the following result

Theorem 2 in Hahn, Todd, and Van der Klaauw (2001)

Suppose that x_i is independent of $_i$ conditional on z_i near z_0. Further suppose that Assumptions (RD), (A1), and (A2) hold. We then have: E $_i$¶z_i z_0 $\frac{y \quad y}{x \quad x}$

Non-parametric Identification of RD

- Then we have the following result

Theorem 2 in Hahn, Todd, and Van der Klaauw (2001)

Suppose that x_i is independent of $_i$ conditional on z_i near z_0. Further suppose that Assumptions (RD), (A1), and (A2) hold. We then have: $E \quad _i \P z_i \quad z_0 \quad \frac{y \quad y}{x \quad x}$

Non-parametric Identification of RD

- Theorem 2 tells us that under heterogeneous TE, if
  - Other confounding factors are continuous at the cutoff (A1)
  - There is no sorting over returns at the cutoff (A2)
- Then we can identify the ATT for individuals around the cutoff
- However, no sorting is a strong assumption under Fuzzy RD
- Individuals of course choose treatment based on how much they can benefit
- Just like Roy model tells us

Non-parametric Identification of RD

- Theorem 2 tells us that under heterogeneous TE, if
  - Other confounding factors are continuous at the cutoff (A1)
  - There is no sorting over returns at the cutoff (A2)
- Then we can identify the ATT for individuals around the cutoff
- However, no sorting is a strong assumption under Fuzzy RD
- Individuals of course choose treatment based on how much they can benefit
- Just like Roy model tells us

Non-parametric Identification of RD

- Theorem 2 tells us that under heterogeneous TE, if
  - Other confounding factors are continuous at the cutoff (A1)
  - There is no sorting over returns at the cutoff (A2)
- Then we can identify the ATT for individuals around the cutoff
- However, no sorting is a strong assumption under Fuzzy RD
- Individuals of course choose treatment based on how much they can benefit
- Just like Roy model tells us

Non-parametric Identification of RD

- Theorem 2 tells us that under heterogeneous TE, if
  - Other confounding factors are continuous at the cutoff (A1)
  - There is no sorting over returns at the cutoff (A2)
- Then we can identify the ATT for individuals around the cutoff
- However, no sorting is a strong assumption under Fuzzy RD
- Individuals of course choose treatment based on how much they can benefit
- Just like Roy model tells us

Non-parametric Identification of RD

- Theorem 2 tells us that under heterogeneous TE, if
  - Other confounding factors are continuous at the cutoff (A1)
  - There is no sorting over returns at the cutoff (A2)
- Then we can identify the ATT for individuals around the cutoff
- However, no sorting is a strong assumption under Fuzzy RD
- Individuals of course choose treatment based on how much they can benefit
- Just like Roy model tells us

Non-parametric Identification of RD

- Theorem 2 tells us that under heterogeneous TE, if
  - Other confounding factors are continuous at the cutoff (A1)
  - There is no sorting over returns at the cutoff (A2)
- Then we can identify the ATT for individuals around the cutoff
- However, no sorting is a strong assumption under Fuzzy RD
- Individuals of course choose treatment based on how much they can benefit
- Just like Roy model tells us

Non-parametric Identification of RD

- Theorem 2 tells us that under heterogeneous TE, if
  - Other confounding factors are continuous at the cutoff (A1)
  - There is no sorting over returns at the cutoff (A2)
- Then we can identify the ATT for individuals around the cutoff
- However, no sorting is a strong assumption under Fuzzy RD
- Individuals of course choose treatment based on how much they can benefit
- Just like Roy model tells us

Non-parametric Identification of RD

- Theorem 2 tells us that under heterogeneous TE, if
  - Other confounding factors are continuous at the cutoff (A1)
  - There is no sorting over returns at the cutoff (A2)
- Then we can identify the ATT for individuals around the cutoff
- However, no sorting is a strong assumption under Fuzzy RD
- Individuals of course choose treatment based on how much they can benefit
- Just like Roy model tells us

Non-parametric Identification of RD

- Let's see what will happen if we drop it
- We invoke a set of assumptions similar to Imbens and Angrist (1994) on LATE

Assumption (A2) — Hahn, Todd, and Van der Klaauw (2001)

(i) $(\beta_i, x_i(z))$ is jointly independent of z_i near z_0. (ii) There exists $\epsilon > 0$ such that $x_i(z_0 + e) \geq x_i(z_0 - e)$ for all $0 < e < \epsilon$

- (i) says that given choice x_i, treatment effect β_i is independent of z_i near z_0
- Running variable z can only affect y through changing treatment x
- Test scores only affect wage through changing whether you can be admitted to PKU (exclusion restriction)
- (ii) says that in a small neighborhood around the cutoff, we have monotonicity

Non-parametric Identification of RD

- Let's see what will happen if we drop it
- We invoke a set of assumptions similar to Imbens and Angrist (1994) on LATE

Assumption (A3) in Hahn, Todd, and Van der Klaauw (2001)

(i) $_i; x_i \; z$ is jointly independent of z_i near z_0. (ii) There exists % 0 such that $x_i \; z_0 \; e \; ' \; x_i \; z_0 \; e$ for all 0 $ e $

- (i) says that given choice x_i, treatment effect $_i$ is independent of z_i near z_0
- Running variable z can only affect y through changing treatment x
- Test scores only affect wage through changing whether you can be admitted to PKU (exclusion restriction)
- (ii) says that in a small neighborhood around the cutoff, we have monotonicity

Non-parametric Identification of RD

- Let's see what will happen if we drop it
- We invoke a set of assumptions similar to Imbens and Angrist (1994) on LATE

Assumption (A3) in Hahn, Todd, and Van der Klaauw (2001)

(i) ${}_i; x_i\ z$ is jointly independent of z_i near z_0. (ii) There exists % 0 such that $x_i\ z_0\ e$ ' $x_i\ z_0\ e$ for all 0 $ e $

- (i) says that given choice x_i, treatment effect ${}_i$ is independent of z_i near z_0
- Running variable z can only affect y through changing treatment x
- Test scores only affect wage through changing whether you can be admitted to PKU (exclusion restriction)
- (ii) says that in a small neighborhood around the cutoff, we have monotonicity

Non-parametric Identification of RD

- Let's see what will happen if we drop it
- We invoke a set of assumptions similar to Imbens and Angrist (1994) on LATE

Assumption (A3) in Hahn, Todd, and Van der Klaauw (2001)

(i) $_i; x_i \;\; z$ is jointly independent of z_i near z_0. (ii) There exists % 0 such that $x_i \;\; z_0 \;\; e$ ' $x_i \;\; z_0 \;\; e$ for all 0 \$ e \$

- (i) says that given choice x_i, treatment effect $_i$ is independent of z_i near z_0
- Running variable z can only affect y through changing treatment x
- Test scores only affect wage through changing whether you can be admitted to PKU (exclusion restriction)
- (ii) says that in a small neighborhood around the cutoff, we have monotonicity

Non-parametric Identification of RD

- Let's see what will happen if we drop it
- We invoke a set of assumptions similar to Imbens and Angrist (1994) on LATE

Assumption (A3) in Hahn, Todd, and Van der Klaauw (2001)

(i) $_i; x_i \; z$ is jointly independent of z_i near z_0. (ii) There exists % 0 such that $x_i \; z_0 \quad e$ ' $x_i \; z_0 \quad e$ for all 0 $ e $

- (i) says that given choice x_i, treatment effect $_i$ is independent of z_i near z_0
- Running variable z can only affect y through changing treatment x
- Test scores only affect wage through changing whether you can be admitted to PKU (exclusion restriction)
- (ii) says that in a small neighborhood around the cutoff, we have monotonicity

Non-parametric Identification of RD

- Let's see what will happen if we drop it
- We invoke a set of assumptions similar to Imbens and Angrist (1994) on LATE

Assumption (A3) in Hahn, Todd, and Van der Klaauw (2001)

(i) $_i; x_i \; z$ is jointly independent of z_i near z_0. (ii) There exists % 0 such that $x_i \; z_0 \quad e \quad ' \; x_i \; z_0 \quad e$ for all $0 \; \$ \; e \; \$$

- (i) says that given choice x_i, treatment effect $_i$ is independent of z_i near z_0
- Running variable z can only affect y through changing treatment x
- Test scores only affect wage through changing whether you can be admitted to PKU (exclusion restriction)
- (ii) says that in a small neighborhood around the cutoff, we have monotonicity

Non-parametric Identification of RD

- Let's see what will happen if we drop it
- We invoke a set of assumptions similar to Imbens and Angrist (1994) on LATE

Assumption (A3) in Hahn, Todd, and Van der Klaauw (2001)

(i) $_i; x_i \;\; z$ is jointly independent of z_i near z_0. (ii) There exists % 0 such that $x_i \;\; z_0 \;\; e$ ' $x_i \;\; z_0 \;\; e$ for all 0 \$ e \$

- (i) says that given choice x_i, treatment effect $_i$ is independent of z_i near z_0
- Running variable z can only affect y through changing treatment x
- Test scores only affect wage through changing whether you can be admitted to PKU (exclusion restriction)
- (ii) says that in a small neighborhood around the cutoff, we have monotonicity

Non-parametric Identification of RD

- Let's see what will happen if we drop it
- We invoke a set of assumptions similar to Imbens and Angrist (1994) on LATE

Assumption (A3) in Hahn, Todd, and Van der Klaauw (2001)

(i) $_i; x_i\ z$ is jointly independent of z_i near z_0. (ii) There exists % 0 such that $x_i\ z_0\ \ e$ ' $x_i\ z_0\ \ e$ for all 0 \$ e \$

- (i) says that given choice x_i, treatment effect $_i$ is independent of z_i near z_0
- Running variable z can only affect y through changing treatment x
- Test scores only affect wage through changing whether you can be admitted to PKU (exclusion restriction)
- (ii) says that in a small neighborhood around the cutoff, we have monotonicity

Non-parametric Identification of RD

- Under exclusion restriction and monotonicity, we have:

Suppose that Assumptions (RD), (A1), and (A3) hold. We then have:

- Theorem 3 says that we can identify LATE under a set of assumptions similar to Imbens and Angrist (1994)
- This LATE has two parts to be "Local"
  - Individuals who change their choice around cutoff (Compliers)
  - Individuals around the cutoff

Non-parametric Identification of RD

- Under exclusion restriction and monotonicity, we have:

Theorem 3 in Hahn, Todd, and Van der Klaauw (2001)

Suppose that Assumptions (RD), (A1), and (A3) hold. We then have:

$\lim_{e \quad 0} E \quad {}_i ¶ x_i \quad z_0 \quad e \quad x_i \quad z_0 \quad e \quad 1 \quad \frac{y \quad y}{x \quad x}$

- Theorem 3 says that we can identify LATE under a set of assumptions similar to Imbens and Angrist (1994)
- This LATE has two parts to be "Local"
  - Individuals who change their choice around cutoff (Compliers)
  - Individuals around the cutoff

Non-parametric Identification of RD

- Under exclusion restriction and monotonicity, we have:

Theorem 3 in Hahn, Todd, and Van der Klaauw (2001)

Suppose that Assumptions (RD), (A1), and (A3) hold. We then have:

$$\lim_{e\ \ 0} E \quad {}_i\P x_i \ \ z_0 \quad e \quad x_i \ \ z_0 \quad e \quad 1 \quad \frac{y \quad y}{x \quad x}$$

- Theorem 3 says that we can identify LATE under a set of assumptions similar to Imbens and Angrist (1994)
- This LATE has two parts to be "Local"
  - Individuals who change their choice around cutoff (Compliers)
  - Individuals around the cutoff

Non-parametric Identification of RD

- Under exclusion restriction and monotonicity, we have:

Theorem 3 in Hahn, Todd, and Van der Klaauw (2001)

Suppose that Assumptions (RD), (A1), and (A3) hold. We then have:

$$\lim_{e \quad 0} E \quad {}_i \P x_i \quad z_0 \quad e \quad x_i \quad z_0 \quad e \quad 1 \quad \frac{y \quad y}{x \quad x}$$

- Theorem 3 says that we can identify LATE under a set of assumptions similar to Imbens and Angrist (1994)
- This LATE has two parts to be "Local"
  - Individuals who change their choice around cutoff (Compliers)
  - Individuals around the cutoff

Non-parametric Identification of RD

- Under exclusion restriction and monotonicity, we have:

Theorem 3 in Hahn, Todd, and Van der Klaauw (2001)

Suppose that Assumptions (RD), (A1), and (A3) hold. We then have:

$$\lim_{e \quad 0} E \quad {}_i \P x_i \; z_0 \quad e \quad x_i \; z_0 \quad e \quad 1 \quad \frac{y \quad y}{x \quad x}$$

- Theorem 3 says that we can identify LATE under a set of assumptions similar to Imbens and Angrist (1994)
- This LATE has two parts to be "Local"
  - Individuals who change their choice around cutoff (Compliers)
  - Individuals around the cutoff

Non-parametric Identification of RD

- Under exclusion restriction and monotonicity, we have:

Theorem 3 in Hahn, Todd, and Van der Klaauw (2001)

Suppose that Assumptions (RD), (A1), and (A3) hold. We then have:

$$\lim_{e \quad 0} E \quad {}_i \P x_i \; z_0 \quad e \quad x_i \; z_0 \quad e \quad 1 \quad \frac{y \quad y}{x \quad x}$$

- Theorem 3 says that we can identify LATE under a set of assumptions similar to Imbens and Angrist (1994)
- This LATE has two parts to be "Local"
  - Individuals who change their choice around cutoff (Compliers)
  - Individuals around the cutoff

Non-parametric Identification of RD

- Under exclusion restriction and monotonicity, we have:

Theorem 3 in Hahn, Todd, and Van der Klaauw (2001)

Suppose that Assumptions (RD), (A1), and (A3) hold. We then have:

$$\lim_{e \quad 0} E \quad {}_i \P x_i \quad z_0 \quad e \quad x_i \quad z_0 \quad e \quad 1 \quad \frac{y \quad y}{x \quad x}$$

- Theorem 3 says that we can identify LATE under a set of assumptions similar to Imbens and Angrist (1994)
- This LATE has two parts to be "Local"
  - Individuals who change their choice around cutoff (Compliers)
  - Individuals around the cutoff

Non-parametric Identification of RD

- From this analysis of identification of RD
- We can derive what conditions we have to validate
- First, we need to check the existence of the discontinuity
- Draw the figure with x-axis as running variable, y-axis as treatment
- Draw the figure with x-axis as running variable, y-axis as outcome
- Visually detect the discontinuity

Non-parametric Identification of RD

- From this analysis of identification of RD
- We can derive what conditions we have to validate
- First, we need to check the existence of the discontinuity
- Draw the figure with x-axis as running variable, y-axis as treatment
- Draw the figure with x-axis as running variable, y-axis as outcome
- Visually detect the discontinuity

Non-parametric Identification of RD

- From this analysis of identification of RD
- We can derive what conditions we have to validate
- First, we need to check the existence of the discontinuity
- Draw the figure with x-axis as running variable, y-axis as treatment
- Draw the figure with x-axis as running variable, y-axis as outcome
- Visually detect the discontinuity

Non-parametric Identification of RD

- From this analysis of identification of RD
- We can derive what conditions we have to validate
- First, we need to check the existence of the discontinuity
- Draw the figure with x-axis as running variable, y-axis as treatment
- Draw the figure with x-axis as running variable, y-axis as outcome
- Visually detect the discontinuity

Non-parametric Identification of RD

- From this analysis of identification of RD
- We can derive what conditions we have to validate
- First, we need to check the existence of the discontinuity
- Draw the figure with x-axis as running variable, y-axis as treatment
- Draw the figure with x-axis as running variable, y-axis as outcome
- Visually detect the discontinuity

Non-parametric Identification of RD

- From this analysis of identification of RD
- We can derive what conditions we have to validate
- First, we need to check the existence of the discontinuity
- Draw the figure with x-axis as running variable, y-axis as treatment
- Draw the figure with x-axis as running variable, y-axis as outcome
- Visually detect the discontinuity

Non-parametric Identification of RD

- From this analysis of identification of RD
- We can derive what conditions we have to validate
- First, we need to check the existence of the discontinuity
- Draw the figure with x-axis as running variable, y-axis as treatment
- Draw the figure with x-axis as running variable, y-axis as outcome
- Visually detect the discontinuity

Non-parametric Identification of RD

- Second, implement balance test for samples just below and just above the cutoff
- Other variables or confounders should be similar or continuous around the cutoff
- Additionally, check the density of samples around the cutoff
- Make sure there is no bunching to either one side of it
- Good students should not control their scores to just a little above the threshold

Non-parametric Identification of RD

- Second, implement balance test for samples just below and just above the cutoff
- Other variables or confounders should be similar or continuous around the cutoff
- Additionally, check the density of samples around the cutoff
- Make sure there is no bunching to either one side of it
- Good students should not control their scores to just a little above the threshold

Non-parametric Identification of RD

- Second, implement balance test for samples just below and just above the cutoff
- Other variables or confounders should be similar or continuous around the cutoff
- Additionally, check the density of samples around the cutoff
- Make sure there is no bunching to either one side of it
- Good students should not control their scores to just a little above the threshold

Non-parametric Identification of RD

- Second, implement balance test for samples just below and just above the cutoff
- Other variables or confounders should be similar or continuous around the cutoff
- Additionally, check the density of samples around the cutoff
- Make sure there is no bunching to either one side of it
- Good students should not control their scores to just a little above the threshold

Non-parametric Identification of RD

- Second, implement balance test for samples just below and just above the cutoff
- Other variables or confounders should be similar or continuous around the cutoff
- Additionally, check the density of samples around the cutoff
- Make sure there is no bunching to either one side of it
- Good students should not control their scores to just a little above the threshold

Non-parametric Identification of RD

- Second, implement balance test for samples just below and just above the cutoff
- Other variables or confounders should be similar or continuous around the cutoff
- Additionally, check the density of samples around the cutoff
- Make sure there is no bunching to either one side of it
- Good students should not control their scores to just a little above the threshold

Application of RD: He, Wang, and Zhang (2020)

- The paper report this week is He, Wang, and Zhang (2020)
- It estimates the effect of environmental regulation on firm productivity in China
- The basic idea is very interesting
- Monitoring stations only capture emissions from upstream regions
- Thus, local gov officials enforce tighter environmental standards on firms just upstream rather than just downstream
- It gives a natural RDD setting

Application of RD: He, Wang, and Zhang (2020)

- The paper report this week is He, Wang, and Zhang (2020)
- It estimates the effect of environmental regulation on firm productivity in China
- The basic idea is very interesting
- Monitoring stations only capture emissions from upstream regions
- Thus, local gov officials enforce tighter environmental standards on firms just upstream rather than just downstream
- It gives a natural RDD setting

Application of RD: He, Wang, and Zhang (2020)

- The paper report this week is He, Wang, and Zhang (2020)
- It estimates the effect of environmental regulation on firm productivity in China
- The basic idea is very interesting
- Monitoring stations only capture emissions from upstream regions
- Thus, local gov officials enforce tighter environmental standards on firms just upstream rather than just downstream
- It gives a natural RDD setting

Application of RD: He, Wang, and Zhang (2020)

- The paper report this week is He, Wang, and Zhang (2020)
- It estimates the effect of environmental regulation on firm productivity in China
- The basic idea is very interesting
- Monitoring stations only capture emissions from upstream regions
- Thus, local gov officials enforce tighter environmental standards on firms just upstream rather than just downstream
- It gives a natural RDD setting

Application of RD: He, Wang, and Zhang (2020)

- The paper report this week is He, Wang, and Zhang (2020)
- It estimates the effect of environmental regulation on firm productivity in China
- The basic idea is very interesting
- Monitoring stations only capture emissions from upstream regions
- Thus, local gov officials enforce tighter environmental standards on firms just upstream rather than just downstream
- It gives a natural RDD setting

Application of RD: He, Wang, and Zhang (2020)

- The paper report this week is He, Wang, and Zhang (2020)
- It estimates the effect of environmental regulation on firm productivity in China
- The basic idea is very interesting
- Monitoring stations only capture emissions from upstream regions
- Thus, local gov officials enforce tighter environmental standards on firms just upstream rather than just downstream
- It gives a natural RDD setting

Application of RD: He, Wang, and Zhang (2020)

- The paper report this week is He, Wang, and Zhang (2020)
- It estimates the effect of environmental regulation on firm productivity in China
- The basic idea is very interesting
- Monitoring stations only capture emissions from upstream regions
- Thus, local gov officials enforce tighter environmental standards on firms just upstream rather than just downstream
- It gives a natural RDD setting

Extension of RDD: RKD

- An interesting extension of RDD is Regression Kink Design (RKD)
- Rather than using the discontinuity on treatment, we employ the kink on treatment
- The jump is no longer on the level, but the slope
- Or we say, the treatment probability derivative has a discontinuity (second order)

Extension of RDD: RKD

- An interesting extension of RDD is Regression Kink Design (RKD)
- Rather than using the discontinuity on treatment, we employ the kink on treatment
- The jump is no longer on the level, but the slope
- Or we say, the treatment probability derivative has a discontinuity (second order)

Extension of RDD: RKD

- An interesting extension of RDD is Regression Kink Design (RKD)
- Rather than using the discontinuity on treatment, we employ the kink on treatment
- The jump is no longer on the level, but the slope
- Or we say, the treatment probability derivative has a discontinuity (second order)

Extension of RDD: RKD

- An interesting extension of RDD is Regression Kink Design (RKD)
- Rather than using the discontinuity on treatment, we employ the kink on treatment
- The jump is no longer on the level, but the slope
- Or we say, the treatment probability derivative has a discontinuity (second order)

Extension of RDD: RKD

- An interesting extension of RDD is Regression Kink Design (RKD)
- Rather than using the discontinuity on treatment, we employ the kink on treatment
- The jump is no longer on the level, but the slope
- Or we say, the treatment probability derivative has a discontinuity (second order)

Extension of RDD: RKD

- Consider an example from Card et al. (2015) and Card et al. (2017)
- In many countries, workers can get compensation when they are unemployed
- This is called unemployment benefit (UI)
- The amount of UI depends on the wage of your last job
- If your last wage is too low, there is a minimum benefit level
- There is also a maximum value for UI (Bill Gates will not get billions once he is unemployed)

Extension of RDD: RKD

- Consider an example from Card et al. (2015) and Card et al. (2017)
- In many countries, workers can get compensation when they are unemployed
- This is called unemployment benefit (UI)
- The amount of UI depends on the wage of your last job
- If your last wage is too low, there is a minimum benefit level
- There is also a maximum value for UI (Bill Gates will not get billions once he is unemployed)

Extension of RDD: RKD

- Consider an example from Card et al. (2015) and Card et al. (2017)
- In many countries, workers can get compensation when they are unemployed
- This is called unemployment benefit (UI)
- The amount of UI depends on the wage of your last job
- If your last wage is too low, there is a minimum benefit level
- There is also a maximum value for UI (Bill Gates will not get billions once he is unemployed)

Extension of RDD: RKD

- Consider an example from Card et al. (2015) and Card et al. (2017)
- In many countries, workers can get compensation when they are unemployed
- This is called unemployment benefit (UI)
- The amount of UI depends on the wage of your last job
- If your last wage is too low, there is a minimum benefit level
- There is also a maximum value for UI (Bill Gates will not get billions once he is unemployed)

Extension of RDD: RKD

- Consider an example from Card et al. (2015) and Card et al. (2017)
- In many countries, workers can get compensation when they are unemployed
- This is called unemployment benefit (UI)
- The amount of UI depends on the wage of your last job
- If your last wage is too low, there is a minimum benefit level
- There is also a maximum value for UI (Bill Gates will not get billions once he is unemployed)

Extension of RDD: RKD

- Consider an example from Card et al. (2015) and Card et al. (2017)
- In many countries, workers can get compensation when they are unemployed
- This is called unemployment benefit (UI)
- The amount of UI depends on the wage of your last job
- If your last wage is too low, there is a minimum benefit level
- There is also a maximum value for UI (Bill Gates will not get billions once he is unemployed)

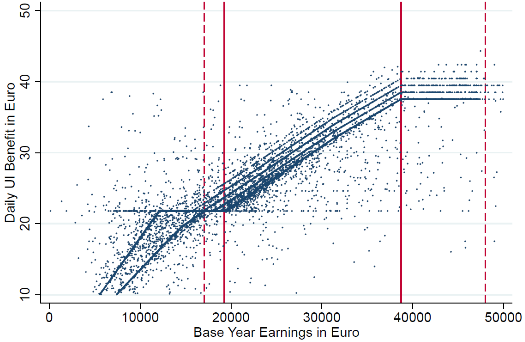
Extension of RDD: RKD

- Consider an example from Card et al. (2015) and Card et al. (2017)
- In many countries, workers can get compensation when they are unemployed
- This is called unemployment benefit (UI)
- The amount of UI depends on the wage of your last job
- If your last wage is too low, there is a minimum benefit level
- There is also a maximum value for UI (Bill Gates will not get billions once he is unemployed)

Extension of RDD: RKD

- Here is a figure for UI distribution in Austria
- Two kinks are noticeable: Minimum and Maximum

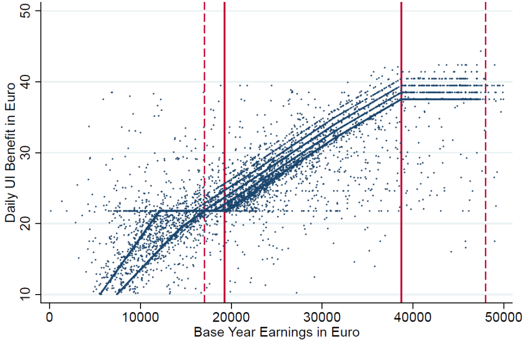
Figure 1: UI Benefits in 2004 Austria



Extension of RDD: RKD

- Here is a figure for UI distribution in Austria
- Two kinks are noticeable: Minimum and Maximum

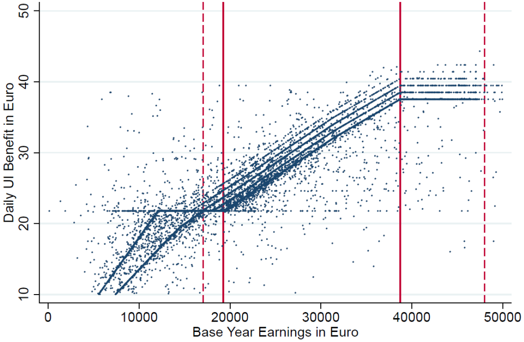
Figure 1: UI Benefits in 2004 Austria



Extension of RDD: RKD

- Here is a figure for UI distribution in Austria
- Two kinks are noticeable: Minimum and Maximum

Figure 1: UI Benefits in 2004 Austria



Extension of RDD: RKD

- A controversial issue is that too generous UI can incentive workers not to search for new jobs
- It is important to investigate the relation between UI benefit B and unemployment duration Y
- Denote V as the wage of the last job, the running variable; U as an error term
- We have $Y \quad y \; B; V; U$ as the outcome function
- In a sharp kink design, B is a deterministic function of V: $B \quad b \; V$ with a slope change at $V \quad 0$
- Here we normalize the kink to $V \quad 0$ to simplify the notation

Extension of RDD: RKD

- A controversial issue is that too generous UI can incentive workers not to search for new jobs
- It is important to investigate the relation between UI benefit B and unemployment duration Y
- Denote V as the wage of the last job, the running variable; U as an error term
- We have $Y \quad y \; B; V; U$ as the outcome function
- In a sharp kink design, B is a deterministic function of V: $B \quad b \; V$ with a slope change at $V \quad 0$
- Here we normalize the kink to $V \quad 0$ to simplify the notation

Extension of RDD: RKD

- A controversial issue is that too generous UI can incentive workers not to search for new jobs
- It is important to investigate the relation between UI benefit B and unemployment duration Y
- Denote V as the wage of the last job, the running variable; U as an error term
- We have $Y = y(B; V; U)$ as the outcome function
- In a sharp kink design, B is a deterministic function of V: $B = b(V)$ with a slope change at $V = 0$
- Here we normalize the kink to $V = 0$ to simplify the notation

Extension of RDD: RKD

- A controversial issue is that too generous UI can incentive workers not to search for new jobs
- It is important to investigate the relation between UI benefit B and unemployment duration Y
- Denote V as the wage of the last job, the running variable; U as an error term
- We have $Y \quad y \; B; V; U$ as the outcome function
- In a sharp kink design, B is a deterministic function of V: $B \quad b \; V$ with a slope change at $V \quad 0$
- Here we normalize the kink to $V \quad 0$ to simplify the notation

Extension of RDD: RKD

- A controversial issue is that too generous UI can incentive workers not to search for new jobs
- It is important to investigate the relation between UI benefit B and unemployment duration Y
- Denote V as the wage of the last job, the running variable; U as an error term
- We have $Y \quad y \;\; B; V; U$ as the outcome function
- In a sharp kink design, B is a deterministic function of V: $B \quad b \;\; V$ with a slope change at $V \quad 0$
- Here we normalize the kink to $V \quad 0$ to simplify the notation

Extension of RDD: RKD

- A controversial issue is that too generous UI can incentive workers not to search for new jobs
- It is important to investigate the relation between UI benefit B and unemployment duration Y
- Denote V as the wage of the last job, the running variable; U as an error term
- We have $Y \quad y \ B; V; U$ as the outcome function
- In a sharp kink design, B is a deterministic function of V: $B \quad b \ V$ with a slope change at $V \quad 0$
- Here we normalize the kink to $V \quad 0$ to simplify the notation

Extension of RDD: RKD

- A controversial issue is that too generous UI can incentive workers not to search for new jobs
- It is important to investigate the relation between UI benefit B and unemployment duration Y
- Denote V as the wage of the last job, the running variable; U as an error term
- We have $Y \quad y \; B; V; U$ as the outcome function
- In a sharp kink design, B is a deterministic function of V: $B \quad b \; V$ with a slope change at $V \quad 0$
- Here we normalize the kink to $V \quad 0$ to simplify the notation

Extension of RDD: RKD

- Assumption 1: (i) U is bounded; (ii) y is continuous and partially differentiable w.r.t. b and v, $y_b \; b; v; u$ is continuous (Regularity)
- Assumption 2: $y_v \; b; v; u$ is continuous around the kink $v \; 0$ (Exclusion). The kink exists only for $b \; v$, but not for the effect of v directly on y.
- Assumption 3: Treatment assignment rule $b \; v$ is known, continuous, and has a kink at $v \; 0$ (Kink existence)
- Assumption 4: Conditional density $f_{V\P U} \; v$ and its partial derivative w.r.t v are continuous around the kink $v \; 0$ (No kink for confounders)

Extension of RDD: RKD

- Assumption 1: (i) U is bounded; (ii) y is continuous and partially differentiable w.r.t. b and v, $y_b \ b; v; u$ is continuous (Regularity)
- Assumption 2: $y_v \ b; v; u$ is continuous around the kink $v \quad 0$ (Exclusion). The kink exists only for $b \ v$, but not for the effect of v directly on y.
- Assumption 3: Treatment assignment rule $b \ v$ is known, continuous, and has a kink at $v \quad 0$ (Kink existence)
- Assumption 4: Conditional density $f_{V \P U} \ v$ and its partial derivative w.r.t v are continuous around the kink $v \quad 0$ (No kink for confounders)

Extension of RDD: RKD

- Assumption 1: (i) U is bounded; (ii) y is continuous and partially differentiable w.r.t. b and v, y_b $b; v; u$ is continuous (Regularity)
- Assumption 2: y_v $b; v; u$ is continuous around the kink v 0 (Exclusion). The kink exists only for b v, but not for the effect of v directly on y.
- Assumption 3: Treatment assignment rule b v is known, continuous, and has a kink at v 0 (Kink existence)
- Assumption 4: Conditional density $f_{V¶U}$ v and its partial derivative w.r.t v are continuous around the kink v 0 (No kink for confounders)

Extension of RDD: RKD

- Assumption 1: (i) U is bounded; (ii) y is continuous and partially di erentiable w.r.t. b and v, $y_b \; b; v; u$ is continuous (Regularity)
- Assumption 2: $y_v \; b; v; u$ is continuous around the kink $v \quad 0$ (Exclusion). The kink exists only for $b \; v$, but not for the e ect of v directly on y.
- Assumption 3: Treatment assignment rule $b \; v$ is known, continuous, and has a kink at $v \quad 0$ (Kink existence)
- Assumption 4: Conditional density $f_{V \,\P U} \; v$ and its partial derivative w.r.t. v are continuous around the kink $v \quad 0$ (No kink for confounders)

Extension of RDD: RKD

- Assumption 1: (i) U is bounded; (ii) y is continuous and partially di erentiable w.r.t. b and v, y_b $b; v; u$ is continuous (Regularity)
- Assumption 2: y_v $b; v; u$ is continuous around the kink v 0 (Exclusion). The kink exists only for b v, but not for the e ect of v directly on y.
- Assumption 3: Treatment assignment rule b v is known, continuous, and has a kink at v 0 (Kink existence)
- Assumption 4: Conditional density $f_{V \P U}$ v and its partial derivative w.r.t. v are continuous around the kink v 0 (No kink for confounders)

Extension of RDD: RKD

- Then we have the non-parametric identi cation of RKD

[illegible]

In a valid Sharp RKD, that is, when Assumptions [illegible] hold:

(a) $P(U \leq u | V = v)$ [illegible] is continuously differentiable in v [illegible] in the neighborhood of the kink.

(b) $E[Y_i | B_i = b_0, U, V = 0]$ [illegible]

- Sharp RKD is dividing slope change of Y ¶V by slope change of v
- On the contrary, RDD divides level by level
- Sharp RKD identi es the ATE for individuals with $B = b_0; V = 0$

Extension of RDD: RKD

- Then we have the non-parametric identi cation of RKD

Proposition 1 in Card et al. (2015)

In a valid Sharp RKD, that is, when Assumptions 1-4 hold:
(a) P U & u¶V v is continuously di erentiable inv at v 0 ¾u " I_U, where I_U is the neighborhood of the kink.

$$\text{(b) } E\ y_b\ b_0; 0; U\ ¶V\quad 0 \quad \frac{\lim_{v\ 0} \frac{dE\ Y\ ¶V\ v}{dv}\Big|_{V\ V_0} \quad \lim_{v\ 0} \frac{dE\ Y\ ¶V\ v}{dv}\Big|_{V\ V_0}}{\lim_{v\ 0} \frac{db\ v}{dv}\Big|_{V\ V_0} \quad \lim_{v\ 0} \frac{db\ v}{dv}\Big|_{V\ V_0}}$$

- Sharp RKD is dividing slope change ofE Y ¶V by slope change of b v
- On the contrary, RDD divides level by level
- Sharp RKD identi es the ATE for individuals with B b$_0$; V 0

Extension of RDD: RKD

- Then we have the non-parametric identi cation of RKD

Proposition 1 in Card et al. (2015)

In a valid Sharp RKD, that is, when Assumptions 1-4 hold:

(a) P U & u¶V v is continuously di erentiable in v at v 0 ¾u " I_U, where I_U is the neighborhood of the kink.

(b) $E\ y_b\ \ b_0; 0; U\ ¶V\ \ 0 \quad \frac{\lim_{v\ 0} \frac{dE\ Y\ ¶V\ v}{dv}\cdot_{V\ V_0}\ \ \lim_{v\ 0} \frac{dE\ Y\ ¶V\ v}{dv}\cdot_{V\ V_0}}{\lim_{v\ 0} \frac{db\ v}{dv}\cdot_{V\ V_0}\ \ \lim_{v\ 0} \frac{db\ v}{dv}\cdot_{V\ V_0}}$

- Sharp RKD is dividing slope change of E Y ¶V by slope change of b v
- On the contrary, RDD divides level by level
- Sharp RKD identi es the ATE for individuals with B b_0; V 0

Extension of RDD: RKD

- Then we have the non-parametric identi cation of RKD

Proposition 1 in Card et al. (2015)

In a valid Sharp RKD, that is, when Assumptions 1-4 hold:
(a) P U & u¶V v is continuously di erentiable in v at v 0 ¾u " I_U, where I_U is the neighborhood of the kink.

(b) E y_b b_0; 0; U ¶V 0 $\frac{\lim_{v \; 0} \frac{dE \; Y ¶V \; v}{dv} \big|_{V \; V_0} \;\; \lim_{v \; 0} \frac{dE \; Y ¶V \; v}{dv} \big|_{V \; V_0}}{\lim_{v \; 0} \frac{db \; v}{dv} \big|_{V \; V_0} \;\; \lim_{v \; 0} \frac{db \; v}{dv} \big|_{V \; V_0}}$

- Sharp RKD is dividing slope change of E Y ¶V by slope change of b v
- On the contrary, RDD divides level by level
- Sharp RKD identi es the ATE for individuals with B b_0; V 0

Extension of RDD: RKD

- Then we have the non-parametric identi cation of RKD

Proposition 1 in Card et al. (2015)

In a valid Sharp RKD, that is, when Assumptions 1-4 hold:

(a) P U & u¶V v is continuously di erentiable in v at v 0 ¾u " I_U, where I_U is the neighborhood of the kink.

(b) $E\ y_b\ b_0; 0; U\ ¶V\ \ 0 \quad \dfrac{\lim_{v\ 0} \frac{dE\ Y ¶V\ v}{dv}\Big|_{V\ V_0} \quad \lim_{v\ 0} \frac{dE\ Y ¶V\ v}{dv}\Big|_{V\ V_0}}{\lim_{v\ 0} \frac{db\ v}{dv}\Big|_{V\ V_0} \quad \lim_{v\ 0} \frac{db\ v}{dv}\Big|_{V\ V_0}}$

- Sharp RKD is dividing slope change of E Y ¶V by slope change of b v
- On the contrary, RDD divides level by level
- Sharp RKD identi es the ATE for individuals with B b_0; V 0

Extension of RDD: RKD

- Then we have the non-parametric identification of RKD

Proposition 1 in Card et al. (2015)

In a valid Sharp RKD, that is, when Assumptions 1-4 hold:

(a) $P(U \leq u \mid V = v)$ is continuously differentiable in v at $v = 0$ $\forall u \in I_U$, where I_U is the neighborhood of the kink.

(b) $$E[y_b(b_0, 0, U) \mid V = 0] = \frac{\lim_{v \to 0^+} \frac{dE[Y \mid V = v]}{dv}\Big|_{v = v_0} - \lim_{v \to 0^-} \frac{dE[Y \mid V = v]}{dv}\Big|_{v = v_0}}{\lim_{v \to 0^+} \frac{db(v)}{dv}\Big|_{v = v_0} - \lim_{v \to 0^-} \frac{db(v)}{dv}\Big|_{v = v_0}}$$

- Sharp RKD is dividing slope change of $E[Y \mid V]$ by slope change of $b(v)$
- On the contrary, RDD divides level by level
- Sharp RKD identifies the ATE for individuals with $B = b_0, V = 0$

Extension of RDD: RKD

- The intuition here is as follows
- A change in the slope of treatment probability results in a change in the slope of average outcome
- If there is no change of slope for unobserved confounders
- We can attribute all changes in outcomes to changes of treatment

Extension of RDD: RKD

- The intuition here is as follows
- A change in the slope of treatment probability results in a change in the slope of average outcome
- If there is no change of slope for unobserved confounders
- We can attribute all changes in outcomes to changes of treatment

Extension of RDD: RKD

- The intuition here is as follows
- A change in the slope of treatment probability results in a change in the slope of average outcome
- If there is no change of slope for unobserved confounders
- We can attribute all changes in outcomes to changes of treatment

Extension of RDD: RKD

- The intuition here is as follows
- A change in the slope of treatment probability results in a change in the slope of average outcome
- If there is no change of slope for unobserved confounders
- We can attribute all changes in outcomes to changes of treatment

Extension of RDD: RKD

- The intuition here is as follows
- A change in the slope of treatment probability results in a change in the slope of average outcome
- If there is no change of slope for unobserved confounders
- We can attribute all changes in outcomes to changes of treatment

Extension of RDD: RKD

- What about Fuzzy case?
- The result is very complicated, but with no surprising intuition
- In a Fuzzy RKD, we identify a LATE for individuals who have UI slope changes at the kink
- The larger you change, the larger weight you have
- Of course, we have to invoke some monotonicity assumption

Extension of RDD: RKD

- What about Fuzzy case?
- The result is very complicated, but with no surprising intuition
- In a Fuzzy RKD, we identify a LATE for individuals who have UI slope changes at the kink
- The larger you change, the larger weight you have
- Of course, we have to invoke some monotonicity assumption

Extension of RDD: RKD

- What about Fuzzy case?
- The result is very complicated, but with no surprising intuition
- In a Fuzzy RKD, we identify a LATE for individuals who have UI slope changes at the kink
- The larger you change, the larger weight you have
- Of course, we have to invoke some monotonicity assumption

Extension of RDD: RKD

- What about Fuzzy case?
- The result is very complicated, but with no surprising intuition
- In a Fuzzy RKD, we identify a LATE for individuals who have UI slope changes at the kink
- The larger you change, the larger weight you have
- Of course, we have to invoke some monotonicity assumption

Extension of RDD: RKD

- What about Fuzzy case?
- The result is very complicated, but with no surprising intuition
- In a Fuzzy RKD, we identify a LATE for individuals who have UI slope changes at the kink
- The larger you change, the larger weight you have
- Of course, we have to invoke some monotonicity assumption

Extension of RDD: RKD

- What about Fuzzy case?
- The result is very complicated, but with no surprising intuition
- In a Fuzzy RKD, we identify a LATE for individuals who have UI slope changes at the kink
- The larger you change, the larger weight you have
- Of course, we have to invoke some monotonicity assumption

Conclusion

- When you have a discontinuity in treatment, you can use RDD
- Sharp RDD is matching
  - Using samples around the cuto
  - It identi es ATT for individuals around the cuto
- Fuzzy RDD is IV
  - Using cuto indicator as instrument
  - It identi es LATE for compliers around the cuto
- When you have a discontinuity in treatment slope, you can use RKD
- It also identi es ATT and LATE in Sharp and Fuzzy settings, respectively

Conclusion

- When you have a discontinuity in treatment, you can use RDD
- Sharp RDD is matching
  - Using samples around the cuto
  - It identi es ATT for individuals around the cuto
- Fuzzy RDD is IV
  - Using cuto indicator as instrument
  - It identi es LATE for compliers around the cuto
- When you have a discontinuity in treatment slope, you can use RKD
- It also identi es ATT and LATE in Sharp and Fuzzy settings, respectively

Conclusion

- When you have a discontinuity in treatment, you can use RDD
- Sharp RDD is matching
  - Using samples around the cuto
  - It identi es ATT for individuals around the cuto
- Fuzzy RDD is IV
  - Using cuto indicator as instrument
  - It identi es LATE for compliers around the cuto
- When you have a discontinuity in treatment slope, you can use RKD
- It also identi es ATT and LATE in Sharp and Fuzzy settings, respectively

Conclusion

- When you have a discontinuity in treatment, you can use RDD
- Sharp RDD is matching
  - Using samples around the cuto
  - It identi es ATT for individuals around the cuto
- Fuzzy RDD is IV
  - Using cuto indicator as instrument
  - It identi es LATE for compliers around the cuto
- When you have a discontinuity in treatment slope, you can use RKD
- It also identi es ATT and LATE in Sharp and Fuzzy settings, respectively

Conclusion

- When you have a discontinuity in treatment, you can use RDD
- Sharp RDD is matching
  - Using samples around the cuto
  - It identi es ATT for individuals around the cuto
- Fuzzy RDD is IV
  - Using cuto indicator as instrument
  - It identi es LATE for compliers around the cuto
- When you have a discontinuity in treatment slope, you can use RKD
- It also identi es ATT and LATE in Sharp and Fuzzy settings, respectively

Conclusion

- When you have a discontinuity in treatment, you can use RDD
- Sharp RDD is matching
  - Using samples around the cuto
  - It identi es ATT for individuals around the cuto
- Fuzzy RDD is IV
  - Using cuto indicator as instrument
  - It identi es LATE for compliers around the cuto
- When you have a discontinuity in treatment slope, you can use RKD
- It also identi es ATT and LATE in Sharp and Fuzzy settings, respectively

Conclusion

- When you have a discontinuity in treatment, you can use RDD
- Sharp RDD is matching
  - Using samples around the cuto
  - It identi es ATT for individuals around the cuto
- Fuzzy RDD is IV
  - Using cuto indicator as instrument
  - It identi es LATE for compliers around the cuto
- When you have a discontinuity in treatment slope, you can use RKD
- It also identi es ATT and LATE in Sharp and Fuzzy settings, respectively

Conclusion

- When you have a discontinuity in treatment, you can use RDD
- Sharp RDD is matching
  - Using samples around the cuto
  - It identi es ATT for individuals around the cuto
- Fuzzy RDD is IV
  - Using cuto indicator as instrument
  - It identi es LATE for compliers around the cuto
- When you have a discontinuity in treatment slope, you can use RKD
- It also identi es ATT and LATE in Sharp and Fuzzy settings, respectively

Conclusion

- When you have a discontinuity in treatment, you can use RDD
- Sharp RDD is matching
  - Using samples around the cuto
  - It identi es ATT for individuals around the cuto
- Fuzzy RDD is IV
  - Using cuto indicator as instrument
  - It identi es LATE for compliers around the cuto
- When you have a discontinuity in treatment slope, you can use RKD
- It also identi es ATT and LATE in Sharp and Fuzzy settings, respectively

Conclusion

- When you have a discontinuity in treatment, you can use RDD
- Sharp RDD is matching
  - Using samples around the cuto
  - It identi es ATT for individuals around the cuto
- Fuzzy RDD is IV
  - Using cuto indicator as instrument
  - It identi es LATE for compliers around the cuto
- When you have a discontinuity in treatment slope, you can use RKD
- It also identi es ATT and LATE in Sharp and Fuzzy settings, respectively

Conclusion

- In practice, remember the following tips:
  - Do not use high-order polynomials as smoothing functions
  - A common way is to use local linear regression
  - Using packages in Stata to give you optimal bandwidth and bias-corrected inference
  - Implement balance test both visually and statistically to validate your assumptions

Conclusion

- In practice, remember the following tips:
  - Do not use high-order polynomials as smoothing functions
  - A common way is to use local linear regression
  - Using packages in Stata to give you optimal bandwidth and bias-corrected inference
  - Implement balance test both visually and statistically to validate your assumptions

Conclusion

- In practice, remember the following tips:
  - Do not use high-order polynomials as smoothing functions
  - A common way is to use local linear regression
  - Using packages in Stata to give you optimal bandwidth and bias-corrected inference
  - Implement balance test both visually and statistically to validate your assumptions

Conclusion

- In practice, remember the following tips:
  - Do not use high-order polynomials as smoothing functions
  - A common way is to use local linear regression
  - Using packages in Stata to give you optimal bandwidth and bias-corrected inference
  - Implement balance test both visually and statistically to validate your assumptions

Conclusion

- In practice, remember the following tips:
  - Do not use high-order polynomials as smoothing functions
  - A common way is to use local linear regression
  - Using packages in Stata to give you optimal bandwidth and bias-corrected inference
  - Implement balance test both visually and statistically to validate your assumptions

Conclusion

- In practice, remember the following tips:
  - Do not use high-order polynomials as smoothing functions
  - A common way is to use local linear regression
  - Using packages in Stata to give you optimal bandwidth and bias-corrected inference
  - Implement balance test both visually and statistically to validate your assumptions

References

Angrist, Joshua D and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-discontinuity Designs." *Econometrica* 82 (6):2295–2326.

Card, David, David S Lee, Zhuan Pei, and Andrea Weber. 2015. "Inference on Causal Effects in a Generalized Regression Kink Design." *Econometrica* 83 (6):2453–2483.

———. 2017. "Regression Kink Design: Theory and Practice." In *Regression Discontinuity Designs*, vol. 38. Emerald Publishing Limited, 341–382.

Gelman, Andrew and Guido Imbens. 2019. "Why High-order Polynomials Should Not be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics* 37 (3):447–456.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-discontinuity Design." *Econometrica* 69 (1):201–209.

He, Guojun, Shaoda Wang, and Bing Zhang. 2020. "Watering Down Environmental Regulation in China." *The Quarterly Journal of Economics* 135 (4):2135–2185.

Imbens, Guido and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies* 79 (3):933–959.

Imbens, Gw and Jd Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2):467–475.

Lee, David S. 2008. "Randomized Experiments from Non-random Selection in US House Elections." *Journal of Econometrics* 142 (2):675–697.