FLS 6441 - Methods III: Explanation and Causation

Week 7 - Discontinuities

Jonathan Phillips

April 2020

Classification of Research Designs

		Independence of Treatment Assignment	Researcher Controls Treatment Assignment?
Controlled Experiments	Field Experiments	✓	✓
	Survey and Lab Experiments	✓	✓
Natural Experiments	Natural Experiments	✓	
	Instrumental Variables	✓	
	Discontinuities	✓	
Observational Studies	Difference-in-Differences		
	Controlling for Confounding		
	Matching		
	Comparative Cases and Process Tracing		

Section 1

Discontinuities

Discontinuities

- ► Natural Experiments
  - ► Focused on the portion of treatment assignment which is 'as-if' random
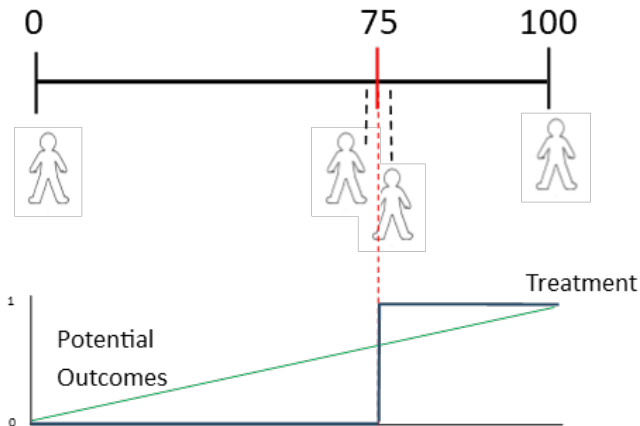
Discontinuities

- ▶ Natural Experiments
  - ▶ Focused on the portion of treatment assignment which is 'as-if' random
- ▶ Where the 'as-if' random treatment assignment comes from *discontinuities* in formal rules

Discontinuities

- ▶ Natural Experiments
  - ▶ Focused on the portion of treatment assignment which is 'as-if' random
- ▶ Where the 'as-if' random treatment assignment comes from *discontinuities* in formal rules
  - ▶ Rules that **treat very similar people very differently**

Discontinuities

- ▶ Natural Experiments
  - ▶ Focused on the portion of treatment assignment which is 'as-if' random
- ▶ Where the 'as-if' random treatment assignment comes from *discontinuities* in formal rules
  - ▶ Rules that **treat very similar people very differently**
  - ▶ Small differences on a **continuous** variable create big differences on a **binary treatment** variable

Discontinuities

Discontinuities

- ▶ Example thresholds:
  - ▶ Exam cutoffs
  - ▶ Age cutoffs
  - ▶ Policy eligibility rules
  - ▶ Close elections
  - ▶ Adminsitrative boundaries

Discontinuities

- ▶ Why do discontinuities assign treatment 'as-if' random?

Discontinuities

► Why do discontinuities assign treatment 'as-if' random?
► Maybe they don't!

Discontinuities

- ► Why do discontinuities assign treatment 'as-if' random?
- ► Maybe they don't! It depends on how much **control** people have over their 'scores'

Discontinuities

- ▶ Why do discontinuities assign treatment 'as-if' random?
- ▶ Maybe they don't! It depends on how much **control** people have over their 'scores'
    - ▶ Could you get a score of exactly 10 in naming all the Brazilian states?

Discontinuities

- ▶ Why do discontinuities assign treatment 'as-if' random?
- ▶ Maybe they don't! It depends on how much **control** people have over their 'scores'
  - ▶ Could you get a score of exactly 10 in naming all the Brazilian states?
  - ▶ Could you get a score of exactly 150 on the GRE?

Discontinuities

- ► Why do discontinuities assign treatment 'as-if' random?
- ► Maybe they don't! It depends on how much **control** people have over their 'scores'
    - ► Could you get a score of exactly 10 in naming all the Brazilian states?
    - ► Could you get a score of exactly 150 on the GRE?
- ► We need qualitative evidence that people cannot 'choose' their score perfectly

Discontinuities

- ▶ Why do discontinuities assign treatment 'as-if' random?
- ▶ Maybe they don't! It depends on how much **control** people have over their 'scores'
  - ▶ Could you get a score of exactly 10 in naming all the Brazilian states?
  - ▶ Could you get a score of exactly 150 on the GRE?
- ▶ We need qualitative evidence that people cannot 'choose' their score perfectly
- ▶ Then the factors that influence *small* changes in score should be independent of potential outcomes

Discontinuities

- ▶ Why do discontinuities assign treatment 'as-if' random?
- ▶ Maybe they don't! It depends on how much **control** people have over their 'scores'
  - ▶ Could you get a score of exactly 10 in naming all the Brazilian states?
  - ▶ Could you get a score of exactly 150 on the GRE?
- ▶ We need qualitative evidence that people cannot 'choose' their score perfectly
- ▶ Then the factors that influence *small* changes in score should be independent of potential outcomes
  - ▶ Weather
  - ▶ Chance
  - ▶ Mistakes
  - ▶ Grading you can't control

Discontinuities

- ▶ Regression Discontinuity
  - ▶ What is the Treatment Assignment Mechanism?

Discontinuities

- ▶ Regression Discontinuity
  - ▶ What is the Treatment Assignment Mechanism?

$$D_i = \begin{cases} 1 & \text{if } x_i \geq \bar{x} \\ 0 & \text{if } x_i < \bar{x} \end{cases}$$

Discontinuities

- ▶ Regression Discontinuity
  - ▶ What is the Treatment Assignment Mechanism?

$$D_i = \begin{cases} 1 & \text{if } x_i \geq \bar{x} \\ 0 & \text{if } x_i < \bar{x} \end{cases}$$

  - ▶ Intelligence/Education
  - ▶ Preparation/Effort
  - ▶ Difficulty of exam
  - ▶ Age

Discontinuities

- ▶ Regression Discontinuity
  - ▶ What is the Treatment Assignment Mechanism?

$$D_i = \begin{cases} 1 & \text{if } x_i \geq \bar{x} \\ 0 & \text{if } x_i < \bar{x} \end{cases}$$

  - ▶ Intelligence/Education
  - ▶ Preparation/Effort
  - ▶ Difficulty of exam
  - ▶ Age
  - ▶ Feeling sick on the day of the exam
  - ▶ Weather making you late
  - ▶ The questions you prepared didn't appear
  - ▶ Who graded your exam

Discontinuities

- ▶ Regression Discontinuity
  - ▶ What is the Treatment Assignment Mechanism?

$$D_i = \begin{cases} 1 & \text{if } x_i \geq \bar{x} \\ 0 & \text{if } x_i < \bar{x} \end{cases}$$

Discontinuities

- ▶ Regression Discontinuity
  - ▶ What is the Treatment Assignment Mechanism?

$$D_i = \begin{cases} 1 & \text{if } x_i \geq \bar{x} \\ 0 & \text{if } x_i < \bar{x} \end{cases}$$

  - ▶ 'As-if' random only **really close to the threshold**

Discontinuities

► Regression Discontinuity
  ► What is the Treatment Assignment Mechanism?

$$D_i = \begin{cases} 1 & \text{if } x_i \geq \bar{x} \\ 0 & \text{if } x_i < \bar{x} \end{cases}$$

  ► 'As-if' random only **really close to the threshold**
  ► For units just above and below the threshold:
    ► Their potential outcomes are almost the same

Discontinuities

- ► Regression Discontinuity
  - ► What is the Treatment Assignment Mechanism?

$$D_i = \begin{cases} 1 & \text{if } x_i \geq \bar{x} \\ 0 & \text{if } x_i < \bar{x} \end{cases}$$

  - ► 'As-if' random only **really close to the threshold**
  - ► For units just above and below the threshold:
    - ► Their potential outcomes are almost the same
    - ► Their covariates are almost the same

Discontinuities

► Regression Discontinuity
  ► What is the Treatment Assignment Mechanism?

$$D_i = \begin{cases} 1 & \text{if } x_i \geq \bar{x} \\ 0 & \text{if } x_i < \bar{x} \end{cases}$$

  ► 'As-if' random only **really close to the threshold**
  ► For units just above and below the threshold:
    ► Their potential outcomes are almost the same
    ► Their covariates are almost the same
    ► They are plausible counterfactuals for each other

Discontinuities

- ▶ Comparisons in a regression discontinuity are always *imperfect*

Discontinuities

- ▶ Comparisons in a regression discontinuity are always *imperfect*
- ▶ **Field experiment:** For every value of X we have both treated and control values
  - ▶ Overlap of covariates

Discontinuities

- ▶ Comparisons in a regression discontinuity are always *imperfect*
- ▶ **Field experiment:** For every value of X we have both treated and control values
    - ▶ Overlap of covariates
- ▶ **Discontinuity:** We **cannot** have treated and control values with the same value of the running variable x
   - ▶ 74.9 vs. 75.1

Discontinuities

- ▶ Comparisons in a regression discontinuity are always *imperfect*
- ▶ **Field experiment:** For every value of X we have both treated and control values
    - ▶ Overlap of covariates
- ▶ **Discontinuity:** We **cannot** have treated and control values with the same value of the running variable x
    - ▶ 74.9 vs. 75.1
- ▶ So we have to **extrapolate** to guess what the potential outcomes would be if unit i was treated instead of control

Discontinuities

- ▶ Comparisons in a regression discontinuity are always _imperfect_
- ▶ **Field experiment:** For every value of X we have both treated and control values
  - ▶ Overlap of covariates
- ▶ **Discontinuity:** We **cannot** have treated and control values with the same value of the running variable x
  - ▶ 74.9 vs. 75.1
- ▶ So we have to **extrapolate** to guess what the potential outcomes would be if unit i was treated instead of control
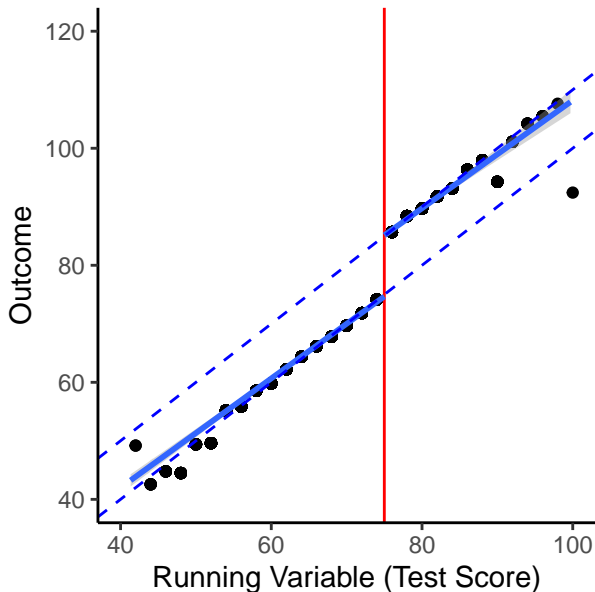- ▶ So we need more assumptions (and more N)!

Discontinuities
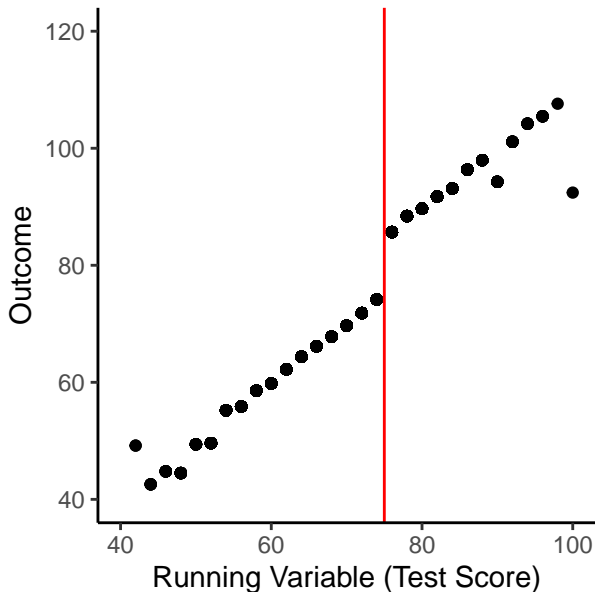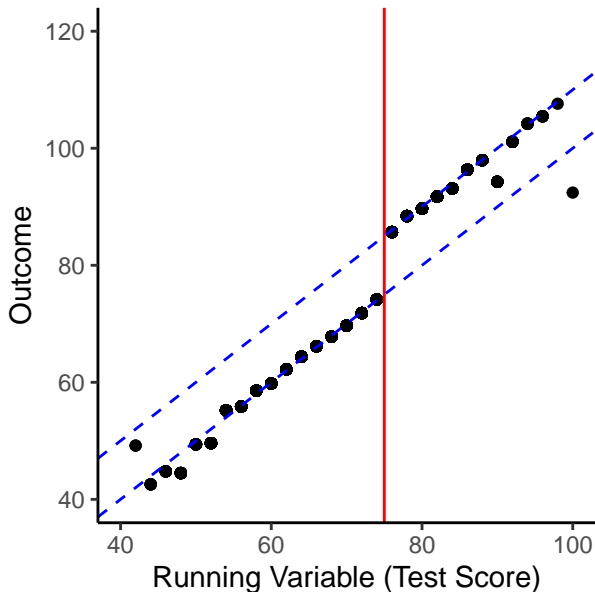
Discontinuities

Discontinuities

Discontinuities

- ▶ Regresssion Discontinuity Variables:
    - ▶ **Running Variable** x_i**:** The *continuous* variable to which the threshold/cutoff is applied, eg. exam score

Discontinuities

- ▶ Regresssion Discontinuity Variables:
    - ▶ **Running Variable** x_i**:** The *continuous* variable to which the threshold/cutoff is applied, eg. exam score
    - ▶ **Treatment** D_i**:** Binary (0/1) variable depending on whether the running variable is above or below the threshold ($x_i \geq \bar{x}$)

Discontinuities

- ▶ Regresssion Discontinuity Variables:
  - ▶ **Running Variable** x_i**:** The *continuous* variable to which the threshold/cutoff is applied, eg. exam score
  - ▶ **Treatment** D_i**:** Binary (0/1) variable depending on whether the running variable is above or below the threshold ($x_i \geq \bar{x}$)
  - ▶ **Outcome** Y_i**:** Any subsequent outcome you have measured

Discontinuities

- ▶ Regression Discontinuity Assumptions:
    1. **Potential outcomes vary continuously** (are independent of treatment) **at** the threshold

Discontinuities

► Regression Discontinuity Assumptions:
　　1. **Potential outcomes vary continuously** (are independent of treatment) **at** the threshold
　　　　1.1 **No Sorting:** Units cannot precisely control their score and sort either side of the threshold

Discontinuities

- ► Regression Discontinuity Assumptions:
  1. **Potential outcomes vary continuously** (are independent of treatment) **at** the threshold
     1.1 **No Sorting:** Units cannot precisely control their score and sort either side of the threshold
     1.2 **No Manipulation:** The threshold is not chosen strategically

Discontinuities

- ▶ Regression Discontinuity Assumptions:
  1. **Potential outcomes vary continuously** (are independent of treatment) **at** the threshold
     1.1 **No Sorting:** Units cannot precisely control their score and sort either side of the threshold
     1.2 **No Manipulation:** The threshold is not chosen strategically
  2. **No compound treatments** (additional treatments assigned by the same cutoff)

Discontinuities

- ▶ Regression Discontinuity Assumptions:
  1. **Potential outcomes vary continuously** (are independent of treatment) **at** the threshold
     1.1 **No Sorting:** Units cannot precisely control their score and sort either side of the threshold
     1.2 **No Manipulation:** The threshold is not chosen strategically
  2. **No compound treatments** (additional treatments assigned by the same cutoff)
  3. **No spillovers** (SUTVA)

Discontinuities

- ▶ The threshold is more likely to be independent of potential outcomes if:

Discontinuities

- ▶ The threshold is more likely to be independent of potential outcomes if:
  - ▶ Units are not aware of the threshold

Discontinuities

- ▶ The threshold is more likely to be independent of potential outcomes if:
  - ▶ Units are not aware of the threshold
  - ▶ The threshold is decided after units make choices

Discontinuities

- ▶ The threshold is more likely to be independent of potential outcomes if:
    - ▶ Units are not aware of the threshold
    - ▶ The threshold is decided after units make choices
    - ▶ The running variable is hard to manipulate precisely

Discontinuities

- ▶ The threshold is more likely to be independent of potential outcomes if:
  - ▶ Units are not aware of the threshold
  - ▶ The threshold is decided after units make choices
  - ▶ The running variable is hard to manipulate precisely
  - ▶ The threshold is chosen before scores are known

Discontinuities

- ▶ The threshold is more likely to be independent of potential outcomes if:
    - ▶ Units are not aware of the threshold
    - ▶ The threshold is decided after units make choices
    - ▶ The running variable is hard to manipulate precisely
    - ▶ The threshold is chosen before scores are known
- ▶ We need qualitative evidence to support these assumptions
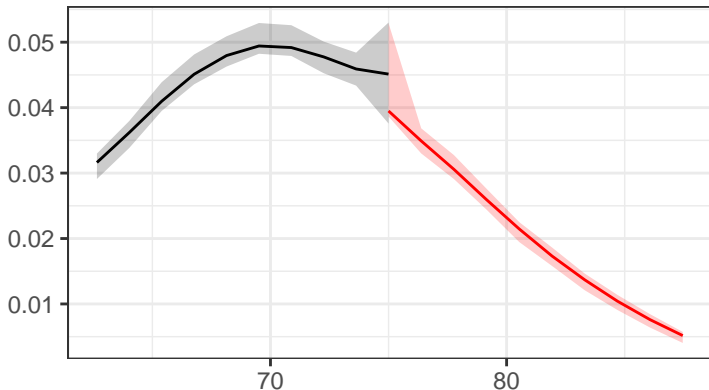
Discontinuities

- ▶ The threshold is more likely to be independent of potential outcomes if:
  - ▶ Units are not aware of the threshold
  - ▶ The threshold is decided after units make choices
  - ▶ The running variable is hard to manipulate precisely
  - ▶ The threshold is chosen before scores are known
- ▶ We need qualitative evidence to support these assumptions
- ▶ AND balance tests to show measurable pre-treatment variables are the same either side of the threshold

Discontinuities

- ▶ The threshold is more likely to be independent of potential outcomes if:
    - ▶ Units are not aware of the threshold
    - ▶ The threshold is decided after units make choices
    - ▶ The running variable is hard to manipulate precisely
    - ▶ The threshold is chosen before scores are known
- ▶ We need qualitative evidence to support these assumptions
- ▶ AND balance tests to show measurable pre-treatment variables are the same either side of the threshold
    - ▶ Simple t-test in a small window either side of the threshold

Discontinuities

- ▶ The threshold is more likely to be independent of potential outcomes if:
  - ▶ Units are not aware of the threshold
  - ▶ The threshold is decided after units make choices
  - ▶ The running variable is hard to manipulate precisely
  - ▶ The threshold is chosen before scores are known
- ▶ We need qualitative evidence to support these assumptions
- ▶ AND balance tests to show measurable pre-treatment variables are the same either side of the threshold
  - ▶ Simple t-test in a small window either side of the threshold
  - ▶ Or a 'placebo' regression discontinuity with the balance variable as the outcome

Discontinuities
- ▶ We can check for sorting with a density test
- ▶ If units are bunched just above the threshold, this suggests manipulation

Section 2

Estimating Regression Discontinuities

Estimating Discontinuities

- ► 3 Regression Discontinuity Methodologies:
  1. **Difference-in-means:** Define a small window either side of the threshold and compare average outcomes in this window
     - ► Biased since we're ignoring the omitted variable effect of the running variable on the outcome
     - ► Low power/efficiency if we have few observations near the cutoff

Estimating Discontinuities

- ► 3 Regression Discontinuity Methodologies:
 1. **Difference-in-means:** Define a small window either side of the threshold and compare average outcomes in this window
     - ► Biased since we're ignoring the omitted variable effect of the running variable on the outcome
     - ► Low power/efficiency if we have few observations near the cutoff
  2. **'Full data' regression discontinuity:** Uses _all_ the data:

$$Y_i = \alpha + \beta_1 Running\_Variable_i + \beta_2 Treatment_i + \epsilon_i$$

   - ► Controls for the continuous variation in the running variable
   - ► Raises efficiency by using all observations
   - ► BUT our estimate depends on getting the functional form correct

Estimating Discontinuities

- ▶ 3 Regression Discontinuity Methodologies:
    1. **Difference-in-means:** Define a small window either side of the threshold and compare average outcomes in this window
        - ▶ Biased since we're ignoring the omitted variable effect of the running variable on the outcome
        - ▶ Low power/efficiency if we have few observations near the cutoff
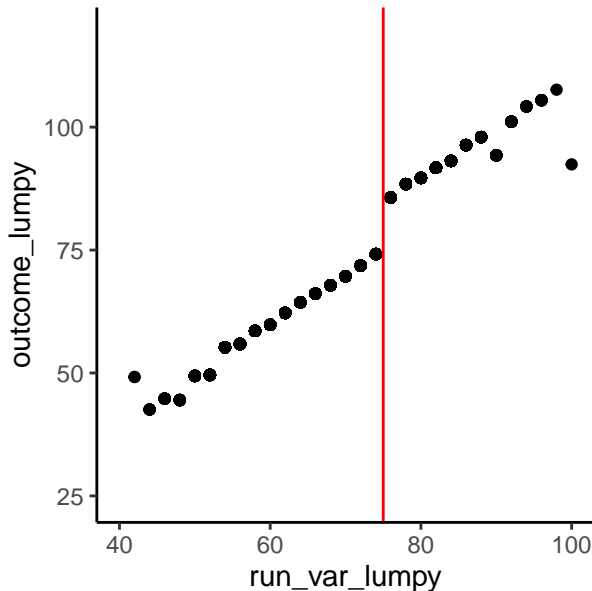   2. **'Full data' regression discontinuity:** Uses *all* the data:

        $$Y_i = \alpha + \beta_1 Running\_Variable_i + \beta_2 Treatment_i + \epsilon_i$$

        - ▶ Controls for the continuous variation in the running variable
        - ▶ Raises efficiency by using all observations
        - ▶ BUT our estimate depends on getting the functional form correct
    3. **'Limited-bandwidth' regression discontinuity:** Same regression as above bu using only data close to the threshold
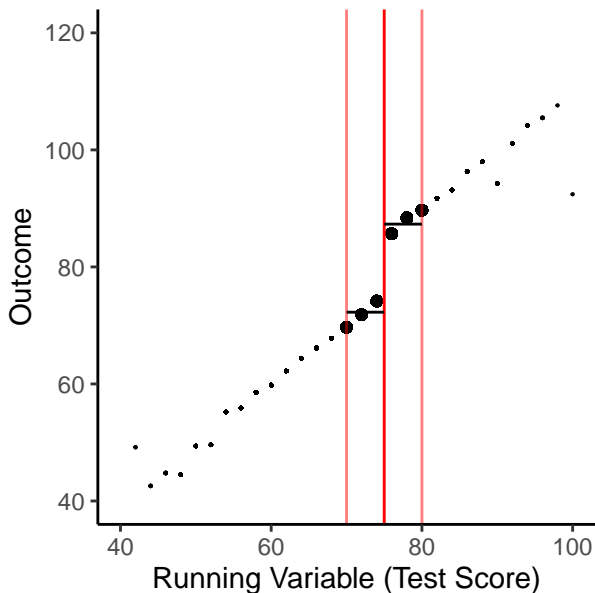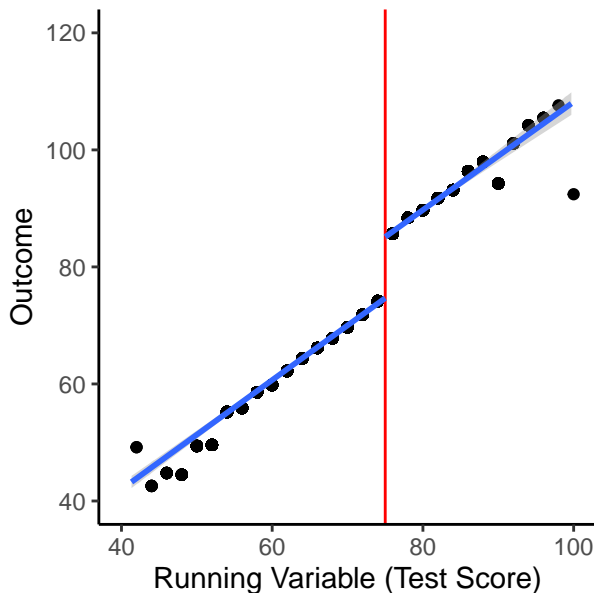   - ▶ Balancing efficiency and bias/model-dependence
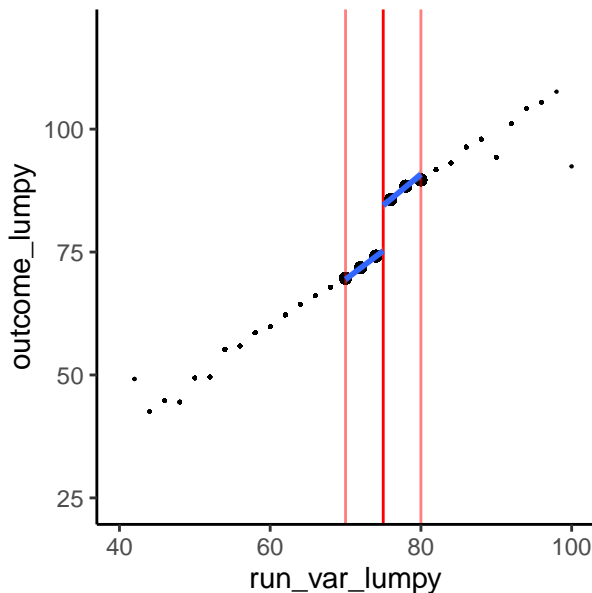
Raw Data
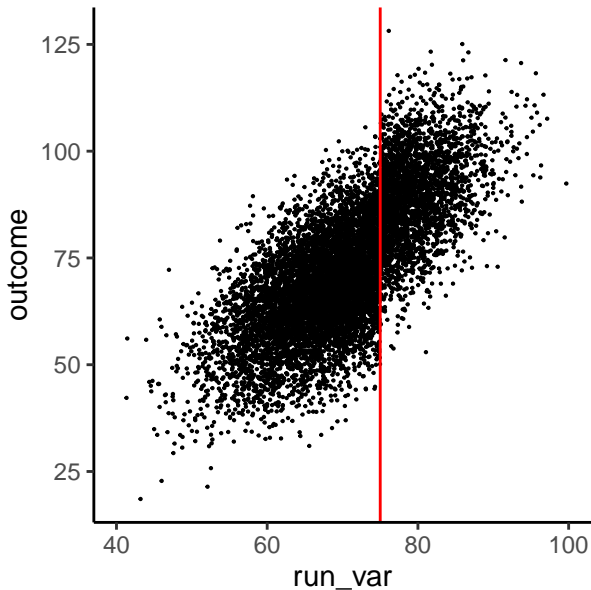
'Binned' Data

1. Difference-in-Means

2. Full Data Regression - Linear

3. Limited-bandwidth Regression - Local Linear

Estimating Discontinuities

- ► Which method?

Estimating Discontinuities

- ▶ Which method?
    - ▶ Difference-in-means is probably biased, and we can easily do better

Estimating Discontinuities

- ▶ Which method?
    - ▶ Difference-in-means is probably biased, and we can easily do better
    - ▶ The full-data approach gives more precision but depends on the right model: linear, quadratic, etc., so more risk of bias

Estimating Discontinuities

- ▶ Which method?
    - ▶ Difference-in-means is probably biased, and we can easily do better
    - ▶ The full-data approach gives more precision but depends on the right model: linear, quadratic, etc., so more risk of bias
    - ▶ The combined approach uses less data (-precision) but is less dependent on the right model (-risk of bias)

Estimating Discontinuities

- ▶ Which method?
  - ▶ Difference-in-means is probably biased, and we can easily do better
  - ▶ The full-data approach gives more precision but depends on the right model: linear, quadratic, etc., so more risk of bias
  - ▶ The combined approach uses less data (-precision) but is less dependent on the right model (-risk of bias)
- ▶ In practice, apply all three as robustness checks

Estimating Discontinuities

- Regression Discontinuity estimates a **Local Average Treatment Effect**

Estimating Discontinuities

- ▶ Regression Discontinuity estimates a **Local Average Treatment Effect**
  - ▶ Treatment assignment is only random at the threshold

Estimating Discontinuities

▶ Regression Discontinuity estimates a **Local Average Treatment Effect**
  ▶ Treatment assignment is only random at the threshold
  ▶ Our estimates only apply to units at/close to the threshold

Estimating Discontinuities

► Regression Discontinuity estimates a **Local Average Treatment Effect**
  ► Treatment assignment is only random at the threshold
  ► Our estimates only apply to units at/close to the threshold
  ► Units far from the threshold are very different for a reason, and causal effects are likely to be different

Estimating Discontinuities

- ▶ Limitations:
  - ▶ Lots of alternative specifications so no single simple test

Estimating Discontinuities

► Limitations:
  ► Lots of alternative specifications so no single simple test
  ► Less precise than a randomized trial, so we need more data

Estimating Discontinuities

► Limitations:
  ► Lots of alternative specifications so no single simple test
  ► Less precise than a randomized trial, so we need more data
  ► Risk of sorting/manipulation

Estimating Discontinuities

▶ Limitations:
  ▶ Lots of alternative specifications so no single simple test
  ▶ Less precise than a randomized trial, so we need more data
  ▶ Risk of sorting/manipulation
  ▶ Opportunistic regression discontinuities may not identify a useful causal effect or for a relevant group

Section 3

Close Elections

Close Elections

▶ Close elections are one type of regression discontinuity in which political office is 'as-if' randomized

Close Elections

- ► Close elections are one type of regression discontinuity in which political office is 'as-if' randomized
  - ► A couple of votes either way due to the weather, illness

Close Elections

- ▶ Close elections are one type of regression discontinuity in which political office is 'as-if' randomized
  - ▶ A couple of votes either way due to the weather, illness
- ▶ Useful for understanding the effects of political power

Close Elections

- ▶ Close elections are one type of regression discontinuity in which political office is 'as-if' randomized
  - ▶ A couple of votes either way due to the weather, illness
- ▶ Useful for understanding the effects of political power
  - ▶ **Running Variable:** Margin of victory
  - ▶ **Treatment:** Winning a close election
  - ▶ **Control:** Losing a close election
  - ▶ **Outcome:** Anything that happens later...

Close Elections

► How much faith should we have in 'close election' regression discontinuities?

Close Elections

- ▶ How much faith should we have in 'close election' regression discontinuities?
- ▶ Eggers et al (2013):

Close Elections

- ▶ How much faith should we have in 'close election' regression discontinuities?
- ▶ Eggers et al (2013):
  - ▶ US House of Representatives elections show sorting in very close elections (<1%)

Close Elections

- ▶ How much faith should we have in 'close election' regression discontinuities?
- ▶ Eggers et al (2013):
  - ▶ US House of Representatives elections show sorting in very close elections (<1%)
  - ▶ Politicians (incumbents, the wealthy) can control whether they win, even when it's a tight race

Close Elections

- ▶ How much faith should we have in 'close election' regression discontinuities?
- ▶ Eggers et al (2013):
  - ▶ US House of Representatives elections show sorting in very close elections (<1%)
  - ▶ Politicians (incumbents, the wealthy) can control whether they win, even when it's a tight race
  - ▶ They have extremely detailed information to predict vote results

Close Elections

- ► How much faith should we have in 'close election' regression discontinuities?
- ► Eggers et al (2013):
    - ► US House of Representatives elections show sorting in very close elections (<1%)
    - ► Politicians (incumbents, the wealthy) can control whether they win, even when it's a tight race
    - ► They have extremely detailed information to predict vote results
    - ► So potential outcomes are not balanced

Close Elections

- ► How much faith should we have in 'close election' regression discontinuities?
- ► Eggers et al (2013):
  - ► US House of Representatives elections show sorting in very close elections (<1%)
  - ► Politicians (incumbents, the wealthy) can control whether they win, even when it's a tight race
  - ► They have extremely detailed information to predict vote results
  - ► So potential outcomes are not balanced
  - ► But no other case (9 countries) has this problem

Close Elections

► Boas and Hidalgo (2011): How does incumbency affect control of the media?

Close Elections

- ▶ Boas and Hidalgo (2011): How does incumbency affect control of the media?
  - ▶ Radio licencing process depends on ability to lobby the Ministry and Congress

Close Elections

- ▶ Boas and Hidalgo (2011): How does incumbency affect control of the media?
  - ▶ Radio licencing process depends on ability to lobby the Ministry and Congress
  - ▶ Incumbents better placed to initiate exchange between Mayors and legislators

Close Elections

- ▶ Boas and Hidalgo (2011): How does incumbency affect control of the media?
    - ▶ Radio licencing process depends on ability to lobby the Ministry and Congress
    - ▶ Incumbents better placed to initiate exchange between Mayors and legislators
- ▶ What is the challenge to causal inference here?

Close Elections

- ▶ **Population:** Brazilian councillors
- ▶ **Sample:** Brazilian councillors in close elections that made radio licence applications in 2000/2004
- ▶ **Running Variable:** Vote margin
- ▶ **Treatment:** Just winning close election
- ▶ **Control:** Just losing close election
- ▶ **Treatment Assignment:** 'As-if' random in close elections
- ▶ **Outcome:** Approved radio licence application rate
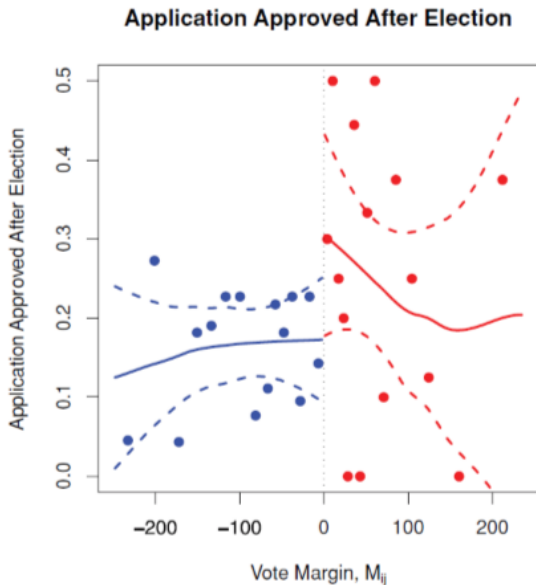
Close Elections

- ▶ Boas and Hidalgo (2011) Methodology:
    1. Local Linear regression within bandwidth of 165 votes
    2. Difference-in-Means within 10-40 vote bandwidth

Close Elections

- ▶ Results
  - ▶ Incumbent Vereadores are twice as likely (14-27 % points) to have their radio licence applications approved

Close Elections

Application Approved After Election

Vote Margin, M_{ij}

Section 4

Geographic Discontinuities

Geographic Discontinuities

▶ What is the effect of governance reform on voters' attitudes?

Geographic Discontinuities

- ▶ What is the effect of governance reform on voters' attitudes?
- ▶ Bihar is one of the poorest places on the planet and was one of the worst goverened

Geographic Discontinuities

- ▶ What is the effect of governance reform on voters' attitudes?
- ▶ Bihar is one of the poorest places on the planet and was one of the worst goverened
- ▶ **Before 2005:** 'Jungle raj': Clientelism, violence, corruption, caste bias

Geographic Discontinuities

- ▶ What is the effect of governance reform on voters' attitudes?
- ▶ Bihar is one of the poorest places on the planet and was one of the worst goverened
- ▶ **Before 2005:** 'Jungle raj': Clientelism, violence, corruption, caste bias
- ▶ **After 2005:** Bihar is a reform success case

Geographic Discontinuities

- ▶ What is the effect of governance reform on voters' attitudes?
- ▶ Bihar is one of the poorest places on the planet and was one of the worst goverened
- ▶ **Before 2005:** 'Jungle raj': Clientelism, violence, corruption, caste bias
- ▶ **After 2005:** Bihar is a reform success case

Geographic Discontinuities

Geographic Discontinuities

► People in Jharkhand are plausible counterfactuals to people in Bihar because:

Geographic Discontinuities

- ▶ People in Jharkhand are plausible counterfactuals to people in Bihar because:
  - ▶ Socioeconomic, geographic and national governance conditions are very similar at the border

Geographic Discontinuities

- ▶ People in Jharkhand are plausible counterfactuals to people in Bihar because:
  - ▶ Socioeconomic, geographic and national governance conditions are very similar at the border
  - ▶ Families have lived in their villages for decades

Geographic Discontinuities

- ▶ People in Jharkhand are plausible counterfactuals to people in Bihar because:
  - ▶ Socioeconomic, geographic and national governance conditions are very similar at the border
  - ▶ Families have lived in their villages for decades
  - ▶ The two states were only created in 2001; before that they experienced the same relationship with government

Geographic Discontinuities

▶ People in Jharkhand are plausible counterfactuals to people in Bihar because:

- ▶ Socioeconomic, geographic and national governance conditions are very similar at the border
- ▶ Families have lived in their villages for decades
- ▶ The two states were only created in 2001; before that they experienced the same relationship with government
- ▶ The border was set according to old district borders, and not politically

Geographic Discontinuities

► People in Jharkhand are plausible counterfactuals to people in Bihar because:
  ► Socioeconomic, geographic and national governance conditions are very similar at the border
  ► Families have lived in their villages for decades
  ► The two states were only created in 2001; before that they experienced the same relationship with government
  ► The border was set according to old district borders, and not politically
  ► Jharkhand did not experience the same governance improvements as Bihar

Methodology

- ▶ The 'running variable' is distance to the border, but in 2-dimensions:
- ▶ Captured by a flexible polynomial in latitude and longitude (x and y)

$$y_i = \alpha + \beta Bihar_i + x_i + y_i + x^2 + y^2 + x^3 + y^3 + x^4 + y^4 + x * y$$
$$+ x^2 * y^2 + x^3 * y^3 + x * y^2 + x * y^3 + x^2 * y + x^3 * y + \epsilon_i$$
$$(1)$$

- ▶ β is our treatment effect of interest

Methodology

- ▶ The 'running variable' is distance to the border, but in 2-dimensions:
- ▶ Captured by a flexible polynomial in latitude and longitude (x and y)

$$y_i = \alpha + \beta Bihar_i + x_i + y_i + x^2 + y^2 + x^3 + y^3 + x^4 + y^4 + x * y$$
$$+ x^2 * y^2 + x^3 * y^3 + x * y^2 + x * y^3 + x^2 * y + x^3 * y + \epsilon_i$$
$$(1)$$

- ▶ β is our treatment effect of interest

Geographic Discontinuities

- ▶ Geographic Regression Discontinuity Design
  - ▶ Exactly the same as a normal regression discontinuity, but in two dimensions (longitude and latitude)

Geographic Discontinuities

► Geographic Regression Discontinuity Design
  ► Exactly the same as a normal regression discontinuity, but in two dimensions (longitude and latitude)
  ► **The Running Variable:**

Geographic Discontinuities

- ▶ Geographic Regression Discontinuity Design
  - ▶ Exactly the same as a normal regression discontinuity, but in two dimensions (longitude and latitude)
  - ▶ **The Running Variable:** Longitude and latitude

Geographic Discontinuities

▶ Geographic Regression Discontinuity Design
  ▶ Exactly the same as a normal regression discontinuity, but in two dimensions (longitude and latitude)
  ▶ **The Running Variable:** Longitude and latitude
  ▶ **Treatment:**

Geographic Discontinuities

- ▶ Geographic Regression Discontinuity Design
  - ▶ Exactly the same as a normal regression discontinuity, but in two dimensions (longitude and latitude)
  - ▶ **The Running Variable:** Longitude and latitude
  - ▶ **Treatment:** Residents on the Bihar side of the border

Geographic Discontinuities

▶ Geographic Regression Discontinuity Design
  ▶ Exactly the same as a normal regression discontinuity, but in two dimensions (longitude and latitude)
  ▶ **The Running Variable:** Longitude and latitude
  ▶ **Treatment:** Residents on the Bihar side of the border
  ▶ **Control:**

Geographic Discontinuities

- ▶ Geographic Regression Discontinuity Design
    - ▶ Exactly the same as a normal regression discontinuity, but in two dimensions (longitude and latitude)
    - ▶ **The Running Variable:** Longitude and latitude
    - ▶ **Treatment:** Residents on the Bihar side of the border
    - ▶ **Control:** Residents on the Jharkhand side of the border
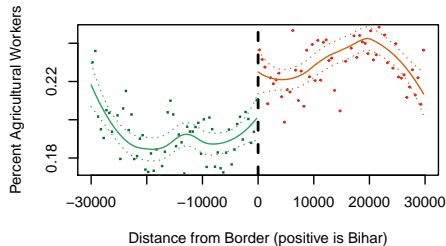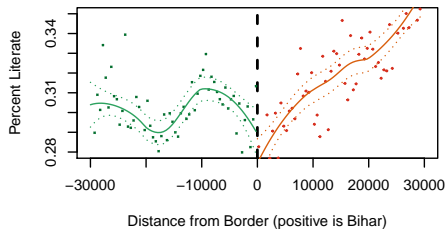
Geographic Discontinuities

- ► Geographic Regression Discontinuity Design
  - ► Exactly the same as a normal regression discontinuity, but in two dimensions (longitude and latitude)
  - ► **The Running Variable:** Longitude and latitude
  - ► **Treatment:** Residents on the Bihar side of the border
  - ► **Control:** Residents on the Jharkhand side of the border
 - ► **Treatment Assignment:**
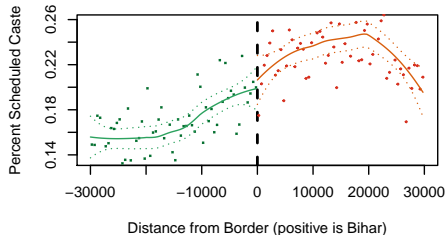
Geographic Discontinuities

- ▶ Geographic Regression Discontinuity Design
  - ▶ Exactly the same as a normal regression discontinuity, but in two dimensions (longitude and latitude)
  - ▶ **The Running Variable:** Longitude and latitude
  - ▶ **Treatment:** Residents on the Bihar side of the border
  - ▶ **Control:** Residents on the Jharkhand side of the border
  - ▶ **Treatment Assignment:** State separation in 2001, Family history, and migration
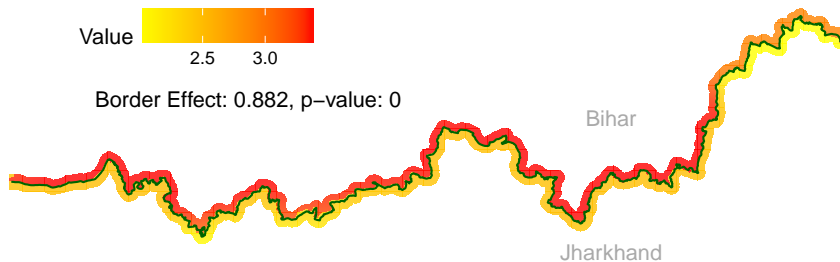
Geographic Discontinuities

- ▶ Geographic Regression Discontinuity Design
  - ▶ Exactly the same as a normal regression discontinuity, but in two dimensions (longitude and latitude)
  - ▶ **The Running Variable:** Longitude and latitude
  - ▶ **Treatment:** Residents on the Bihar side of the border
  - ▶ **Control:** Residents on the Jharkhand side of the border
  - ▶ **Treatment Assignment:** State separation in 2001, Family history, and migration
  - ▶ **Outcome:**

Geographic Discontinuities

- ▶ Geographic Regression Discontinuity Design
  - ▶ Exactly the same as a normal regression discontinuity, but in two dimensions (longitude and latitude)
  - ▶ **The Running Variable:** Longitude and latitude
  - ▶ **Treatment:** Residents on the Bihar side of the border
  - ▶ **Control:** Residents on the Jharkhand side of the border
  - ▶ **Treatment Assignment:** State separation in 2001, Family history, and migration
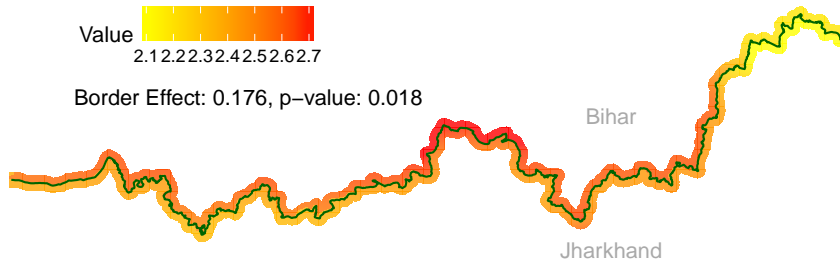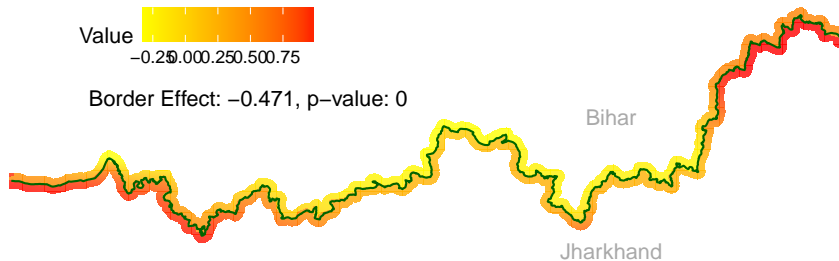 - ▶ **Outcome:** Political attitudes and behaviour

Pair Matched Villages
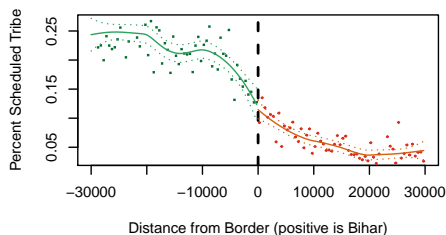
Predicted Value Plot of Likelihood of Incumbent Providing Public Goods if Reelected

Value

2.1 2.2 2.3 2.4 2.5 2.6 2.7

Border Effect: 0.176, p–value: 0.018

Bihar

Jharkhand

Predicted Value Plot of Likelihood of Corrupt Elite being Caught

Predicted Value Plot of Gram Sabha Attendance

Geographic Discontinuities

- ▶ Interpretation:
  - ▶ Governance reform has changed voters' attitudes and expectations
- ▶ But some imbalance at the border...

Geographic Discontinuities

- ▶ Interpretation:
  - ▶ Governance reform has changed voters' attitudes and expectations
- ▶ But some imbalance at the border...
- ▶ ...And compound treatment makes interpretation difficult