# Regression discontinuity designs: an introduction

Oxford spring school in methods

27 March 2017)

Andy Eggers

“Natura non facit saltus.”
(Nature makes no jump.)
— Gottfried Leibniz

“Natura non facit saltus.”
(Nature makes no jump.)
— Gottfried Leibniz

But policies **do** make jumps:

“Natura non facit saltus.”
(Nature makes no jump.)
— Gottfried Leibniz

But policies **do** make jumps:

- Students receive a “first” if they score 70 or higher

“Natura non facit saltus.”
(Nature makes no jump.)
— Gottfried Leibniz

But policies **do** make jumps:

- Students receive a “first” if they score 70 or higher
- French municipalities use PR elections if population is 3,500 (now 1,000) or higher

“Natura non facit saltus.”
(Nature makes no jump.)
— Gottfried Leibniz

But policies **do** make jumps:

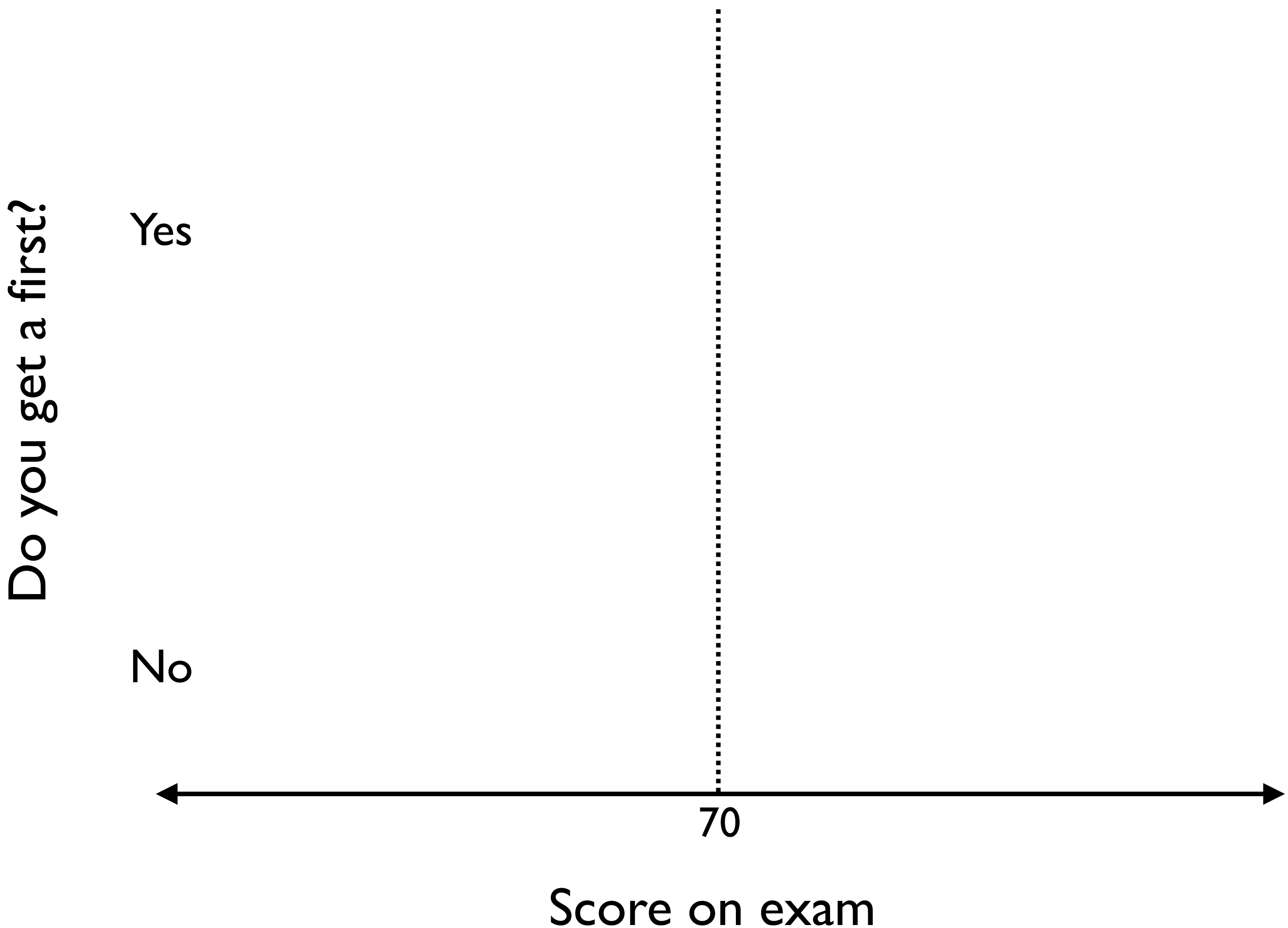
- Students receive a “first” if they score 70 or higher
- French municipalities use PR elections if population is 3,500 (now 1,000) or higher
- A candidate is elected if she receives more votes than any other candidate

"Natura non facit saltus."
(Nature makes no jump.)
— Gottfried Leibniz

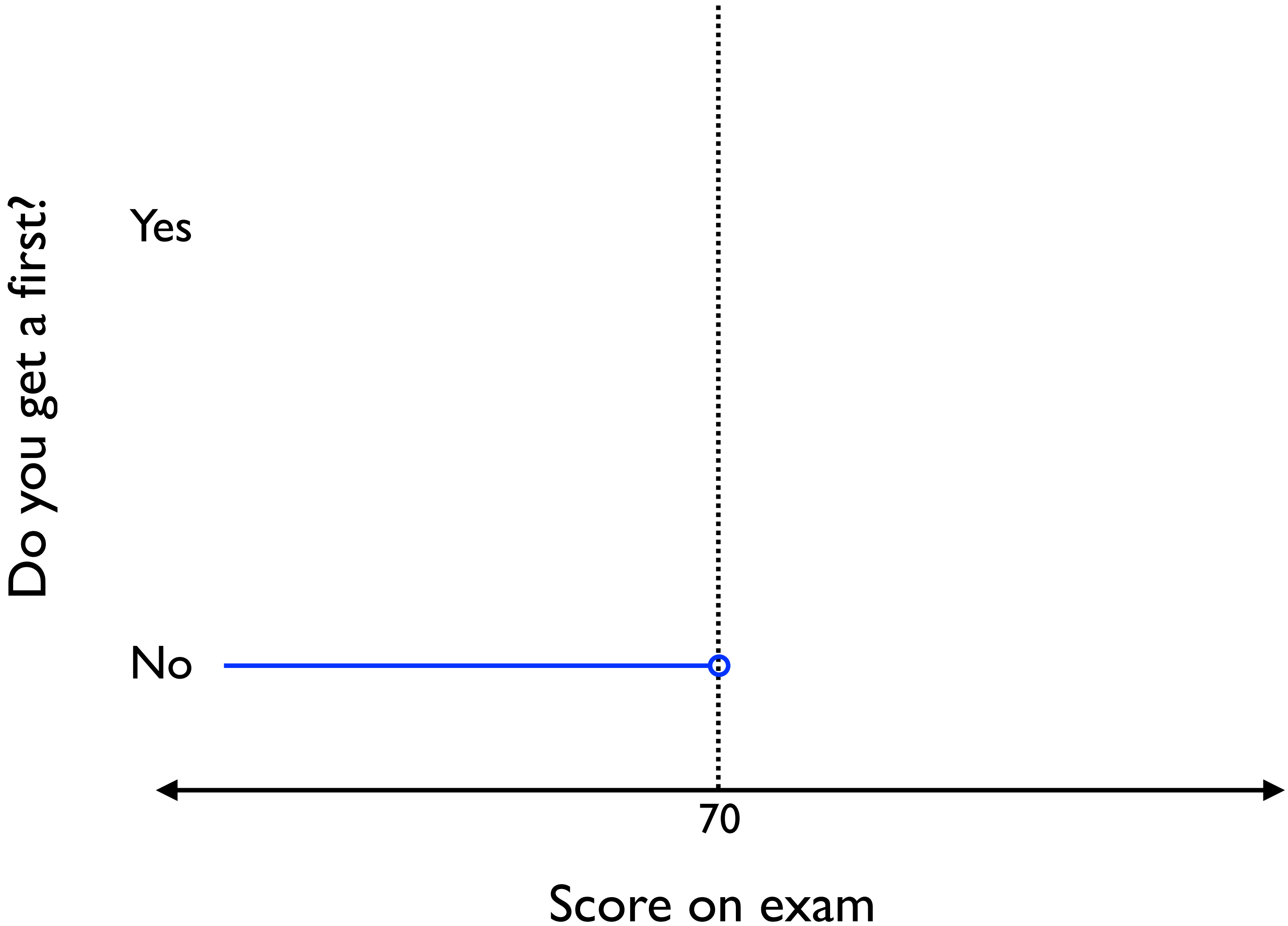
But policies **do** make jumps:

- Students receive a "first" if they score 70 or higher
- French municipalities use PR elections if population is 3,500 (now 1,000) or higher
- A candidate is elected if she receives more votes than any other candidate
- Journalists report a recession if the economy shrinks for two consecutive quarters

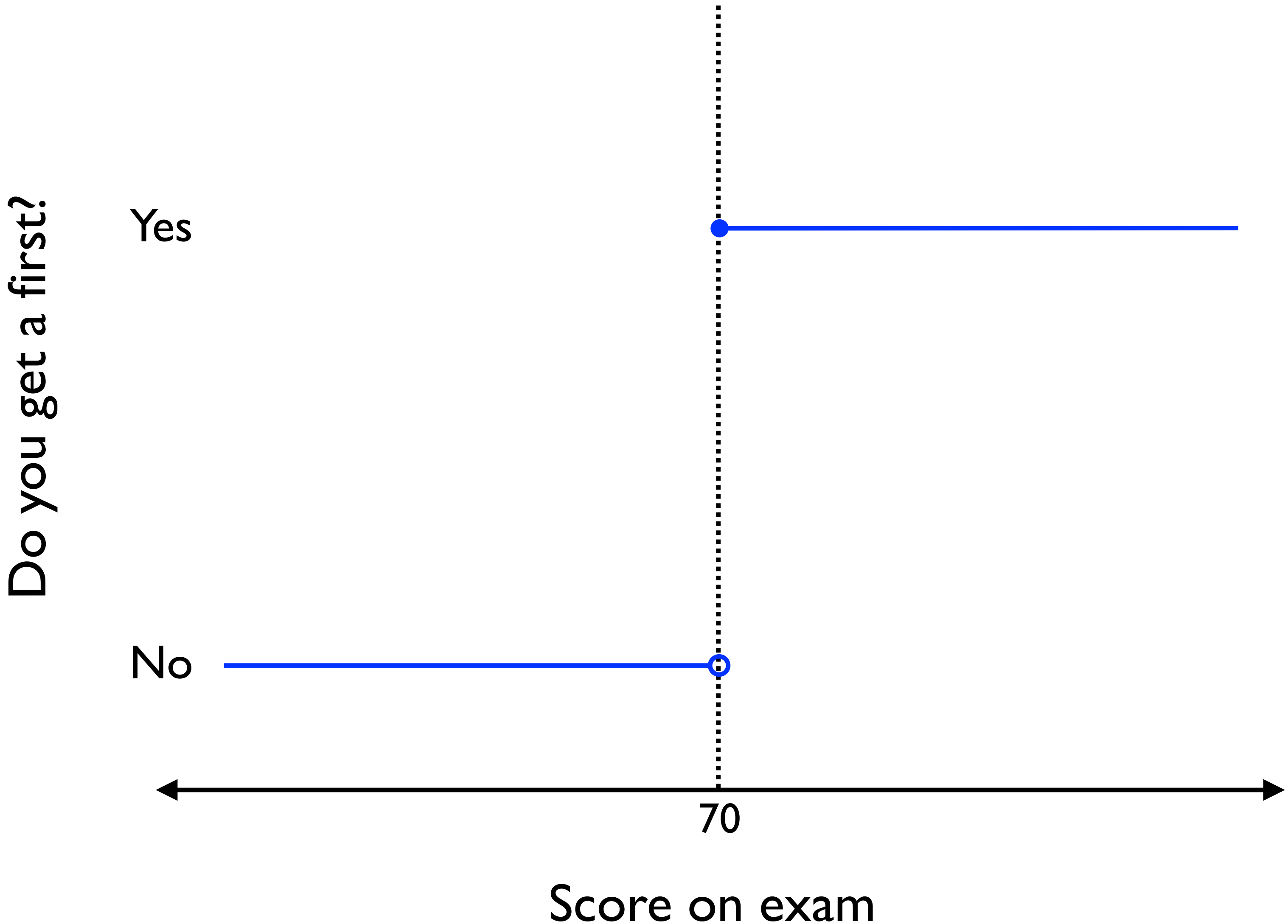
# “Policies” make jumps



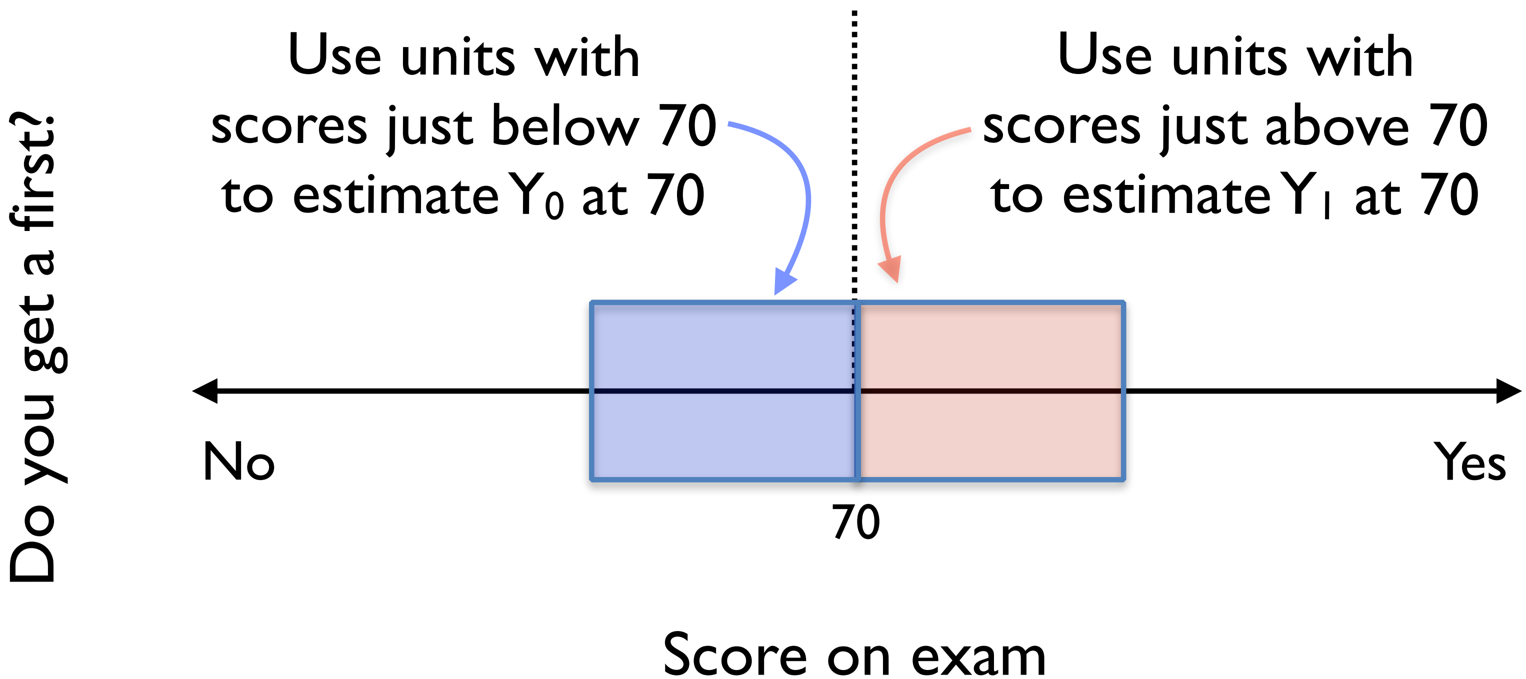
# “Policies” make jumps



# "Policies" make jumps



# If potential outcomes don't jump but treatment does, then near the threshold we have comparable groups getting different treatment.




**Goal**: estimate ATE at threshold (70), i.e. Local Average Treatment Effect (LATE), using units just above and below the threshold.

# Example 1: Eggers and Spirling on incumbency effects in UK politics

# Example 1: Eggers and Spirling on incumbency effects in UK politics

**General question:** Do incumbents have advantages in elections?

# Example 1: Eggers and Spirling on incumbency effects in UK politics

**General question:** Do incumbents have advantages in elections?

**(Party) incumbency effects:** What is the effect of party A winning a seat at time $t$-$1$ on party A's success in the election for that seat at $t$?

# Example 1: Eggers and Spirling on incumbency effects in UK politics

**General question:** Do incumbents have advantages in elections?

**(Party) incumbency effects:** What is the effect of party A winning a seat at time *t-1* on party A's success in the election for that seat at *t*?

How would you estimate that with regression? With matching?

# Example 1: Eggers and Spirling on incumbency effects in UK politics

**General question:** Do incumbents have advantages in elections?

**(Party) incumbency effects:** What is the effect of party A winning a seat at time *t-1* on party A's success in the election for that seat at *t*?

How would you estimate that with regression? With matching?

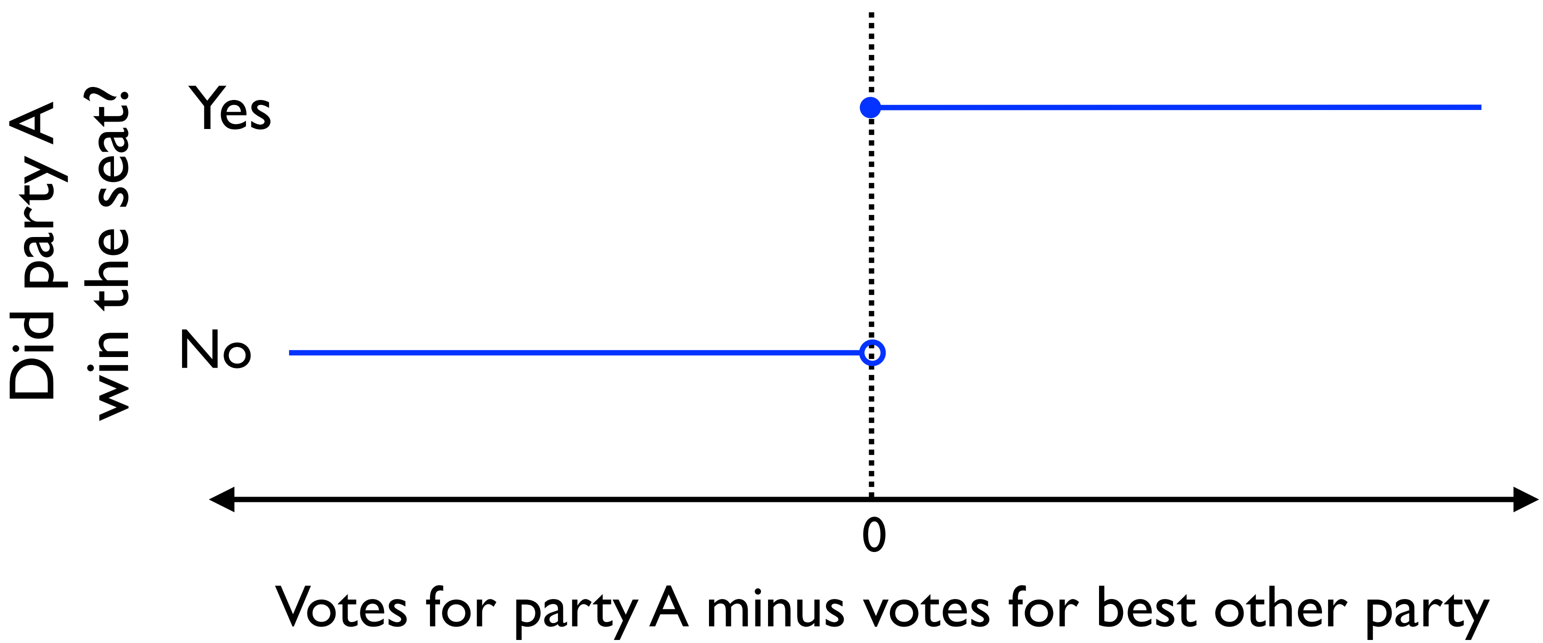
With RDD?

# Example 1: Eggers and Spirling on incumbency effects in UK politics

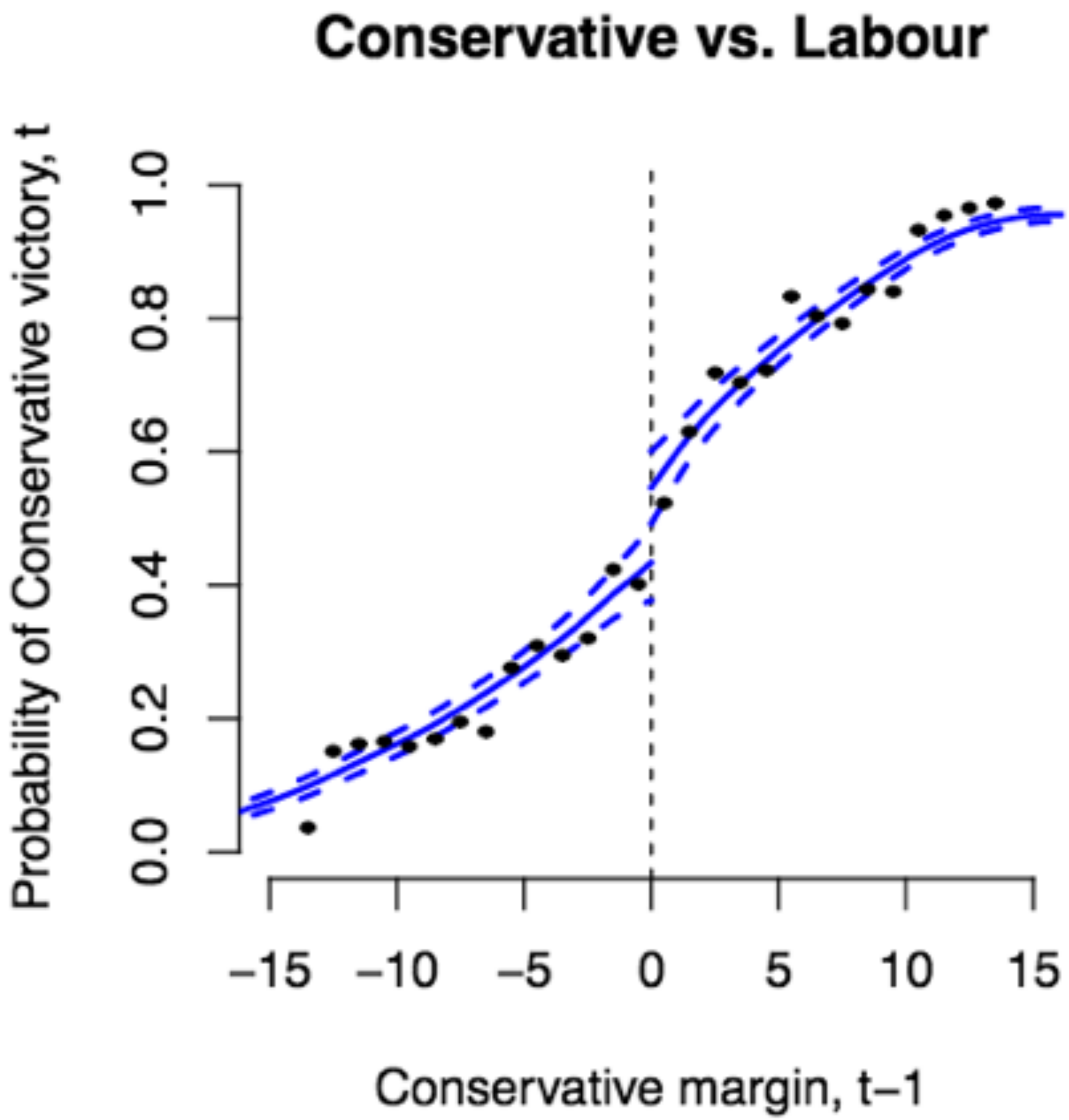
**General question:** Do incumbents have advantages in elections?

**(Party) incumbency effects:** What is the effect of party A winning a seat at time *t-1* on party A's success in the election for that seat at *t*?

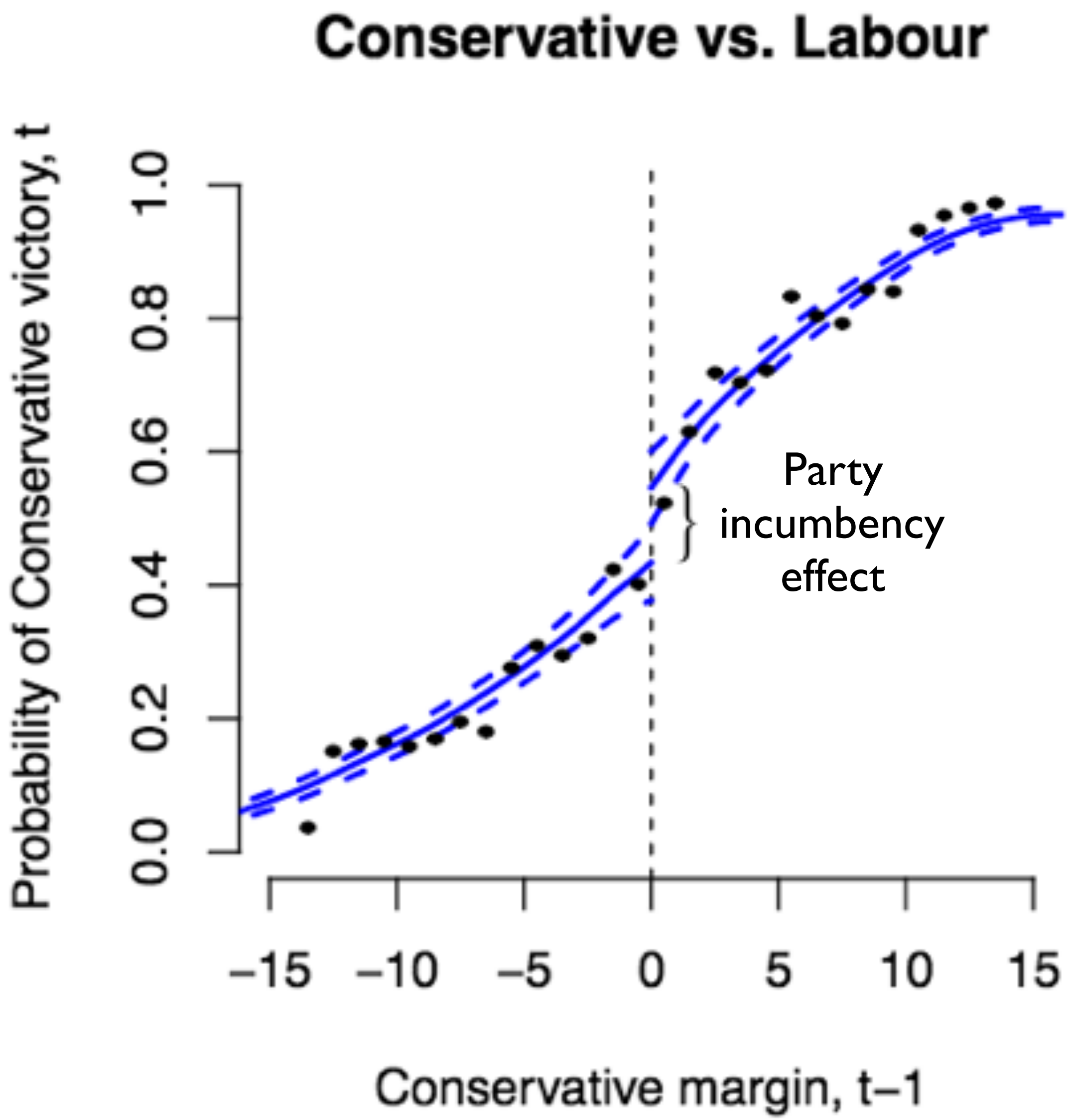
How would you estimate that with regression? With matching? With RDD?



# Example 1: Eggers and Spirling on incumbency effects in UK politics

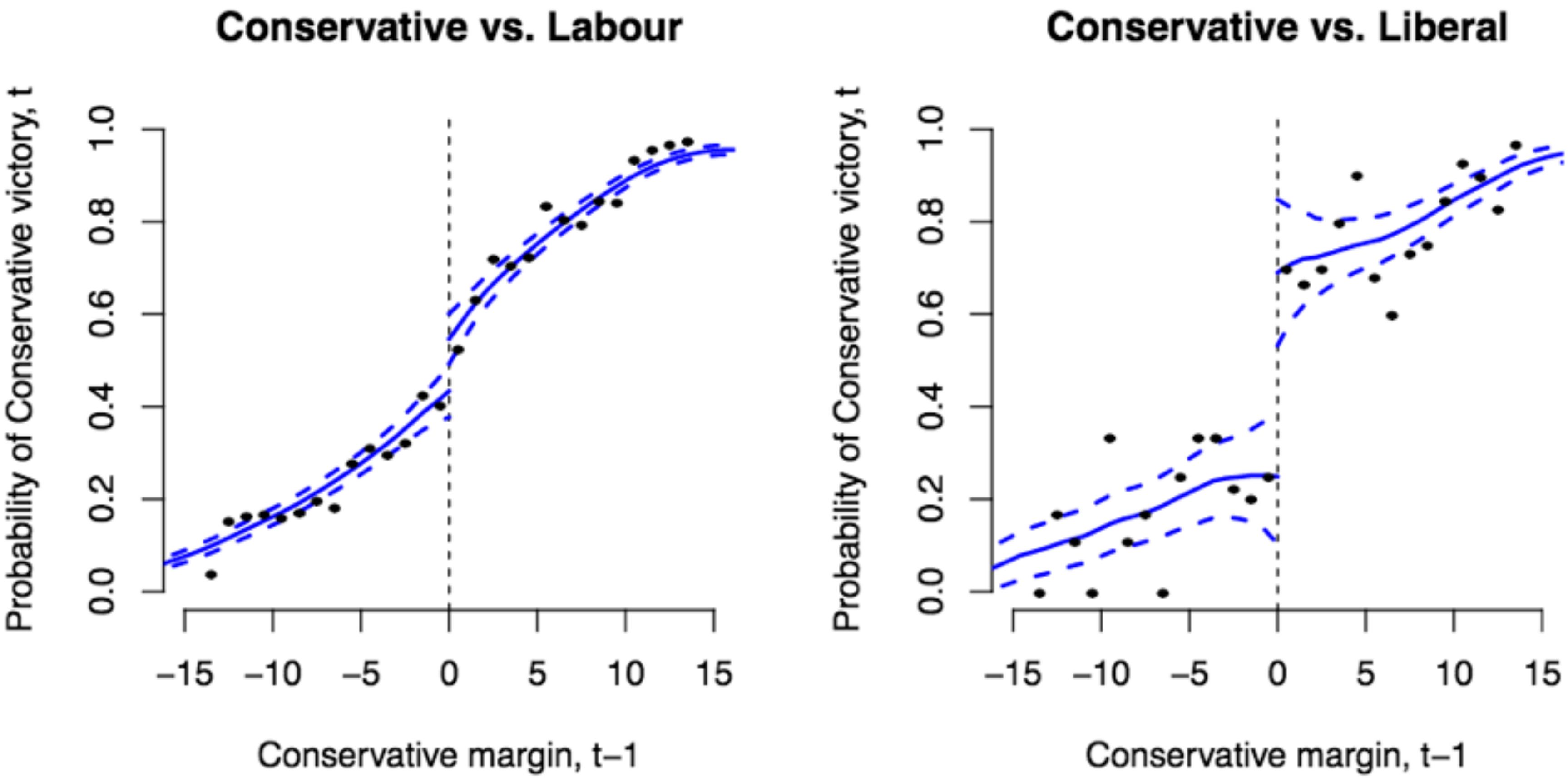


# Example 1: Eggers and Spirling on incumbency effects in UK politics

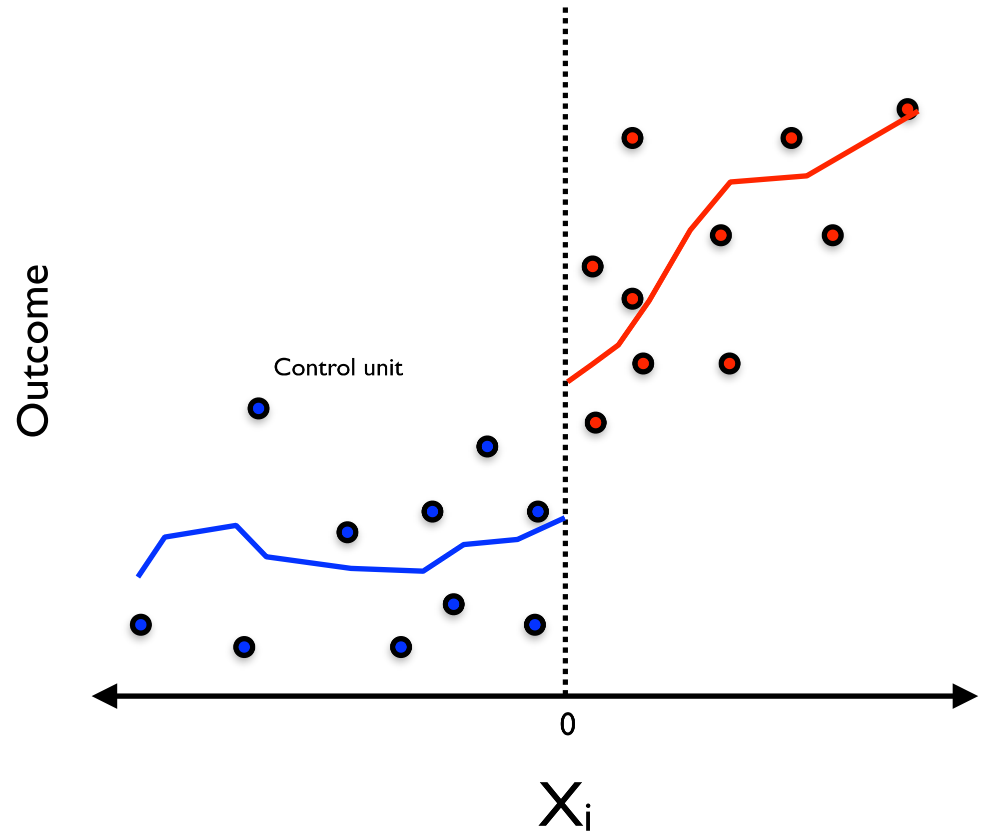


# Example 1: Eggers and Spirling on incumbency effects in UK politics

# Example 1: Eggers and Spirling on incumbency effects in UK politics



# How do we estimate it? Difference in means?

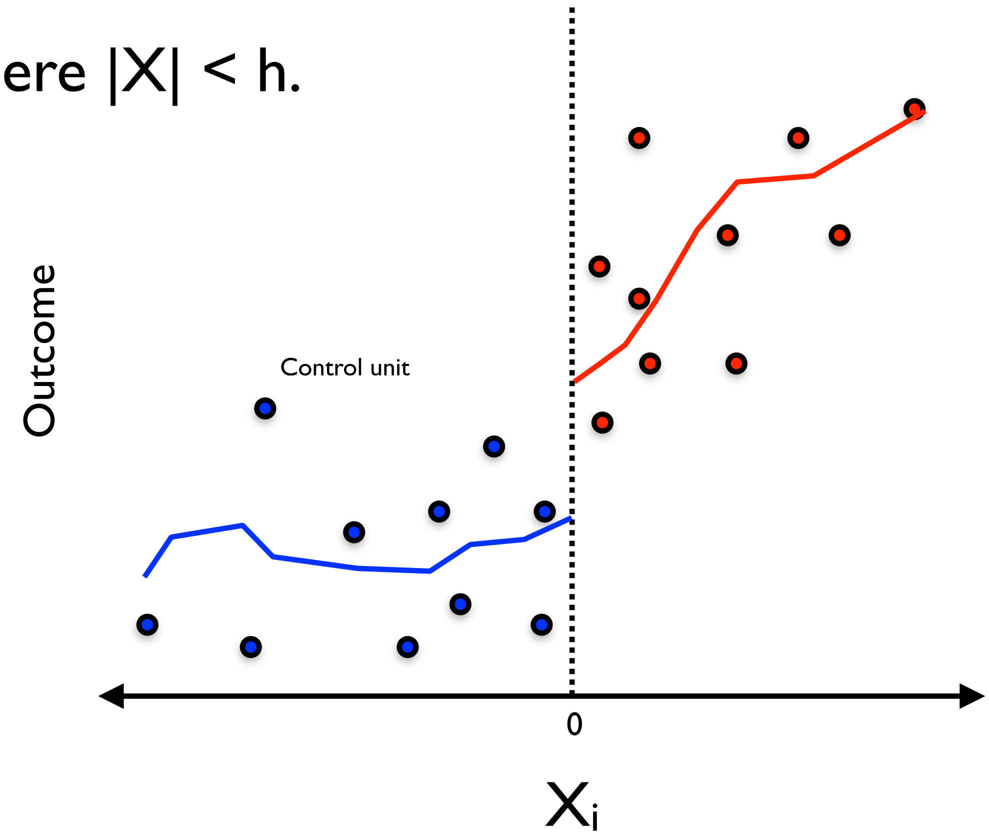


# How do we estimate it? Difference in means?

Consider running this regression:

$Y = \beta_0 + \beta_1 D$

restricting to observations where $|X| < h$.



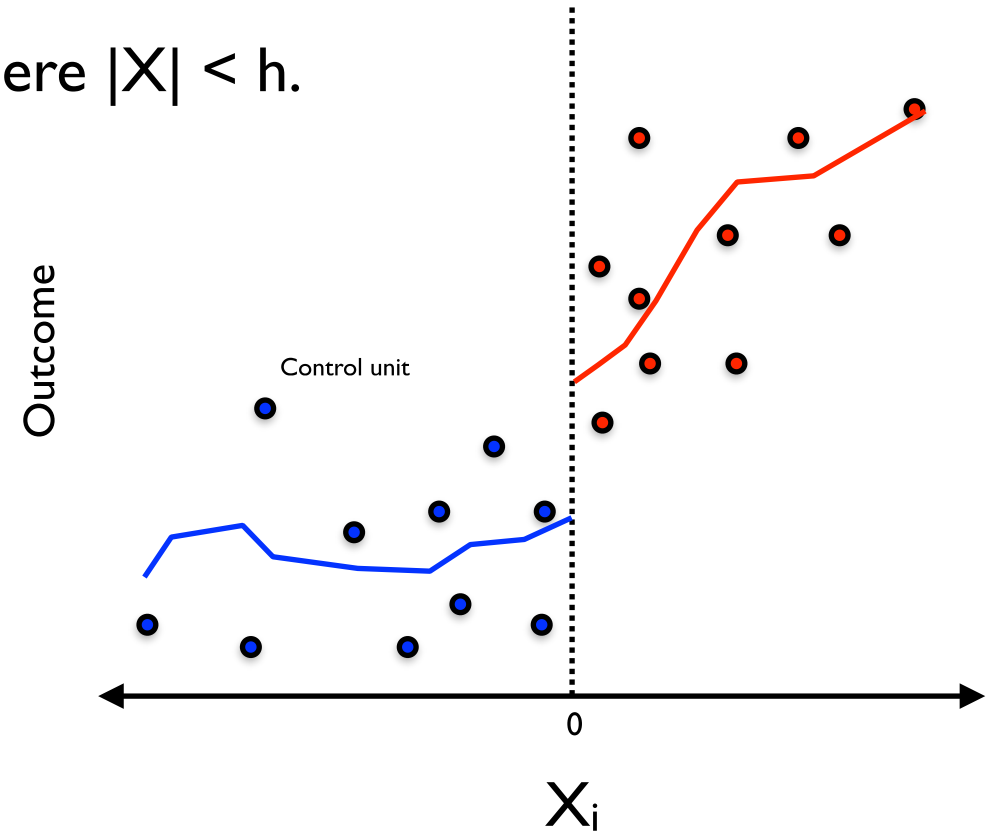
# How do we estimate it? Difference in means?

Consider running this regression:

$Y = \beta_0 + \beta_1 D$

restricting to observations where $|X| < h$.

Questions:



# How do we estimate it? Difference in means?

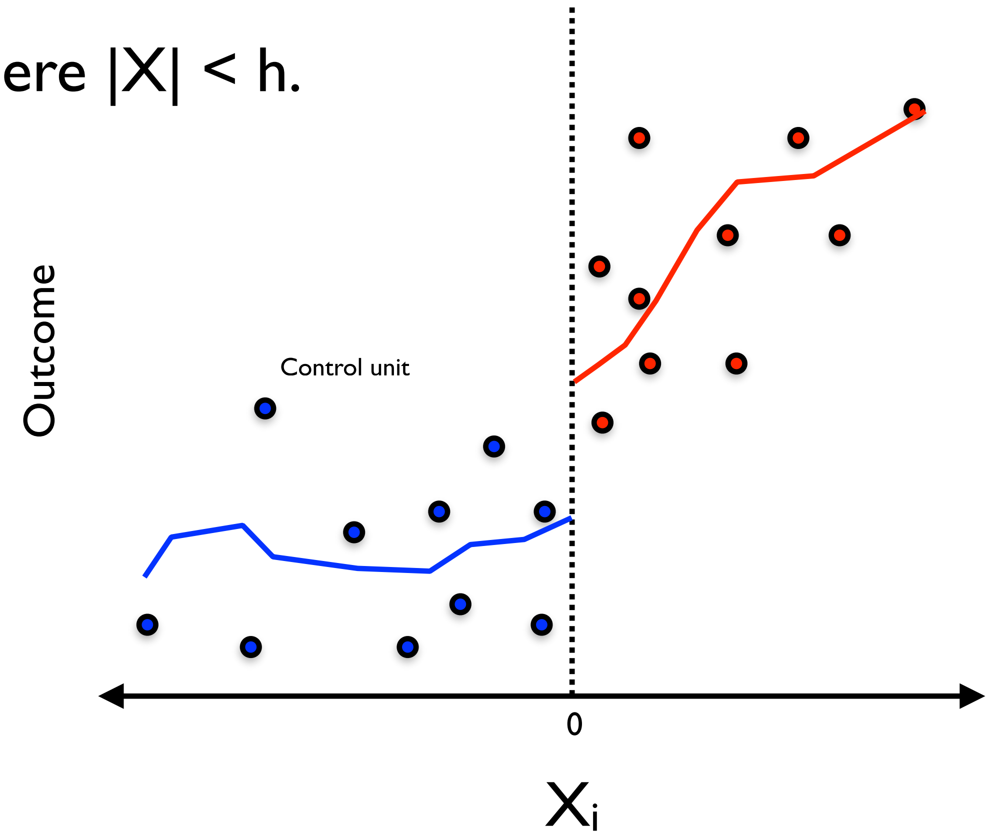
Consider running this regression:

$Y = \beta_0 + \beta_1 D$

restricting to observations where $|X| < h$.

Questions:

- What value of h (*bandwidth, window*) should you choose?



# How do we estimate it? Difference in means?

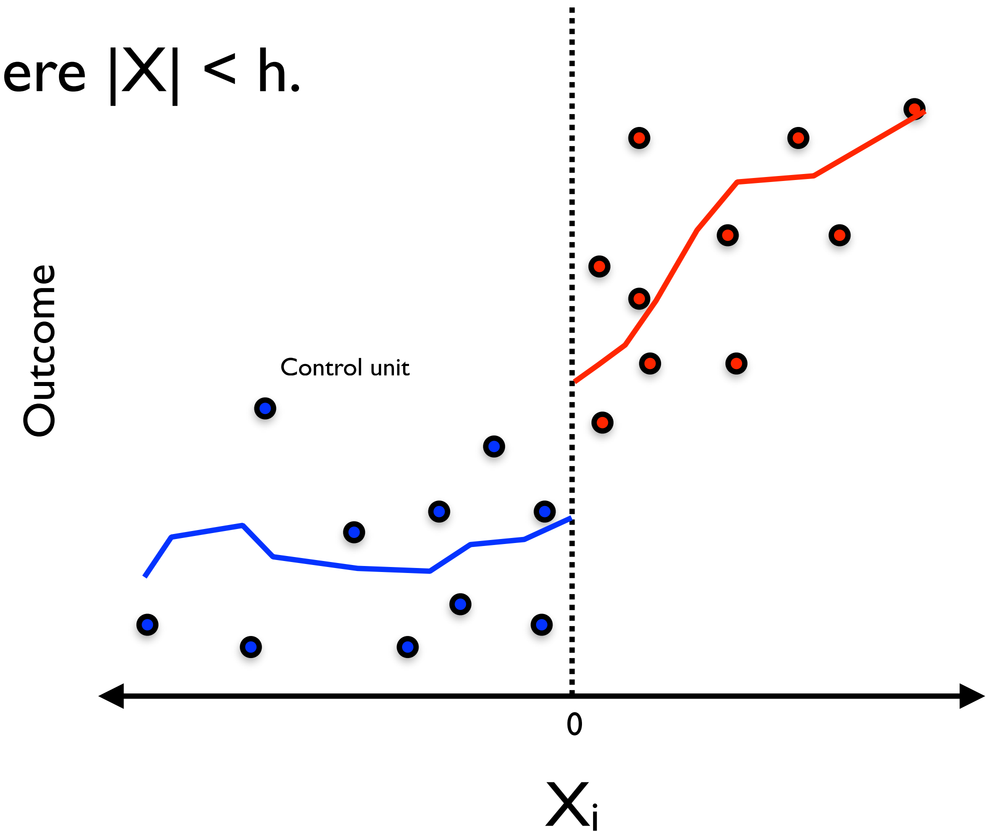
Consider running this regression:

$Y = \beta_0 + \beta_1 D$

restricting to observations where $|X| < h$.

Questions:

- What value of h (*bandwidth, window*) should you choose?
- Will your estimate be unbiased?



# How do we estimate it? Difference in means?

Consider running this regression:

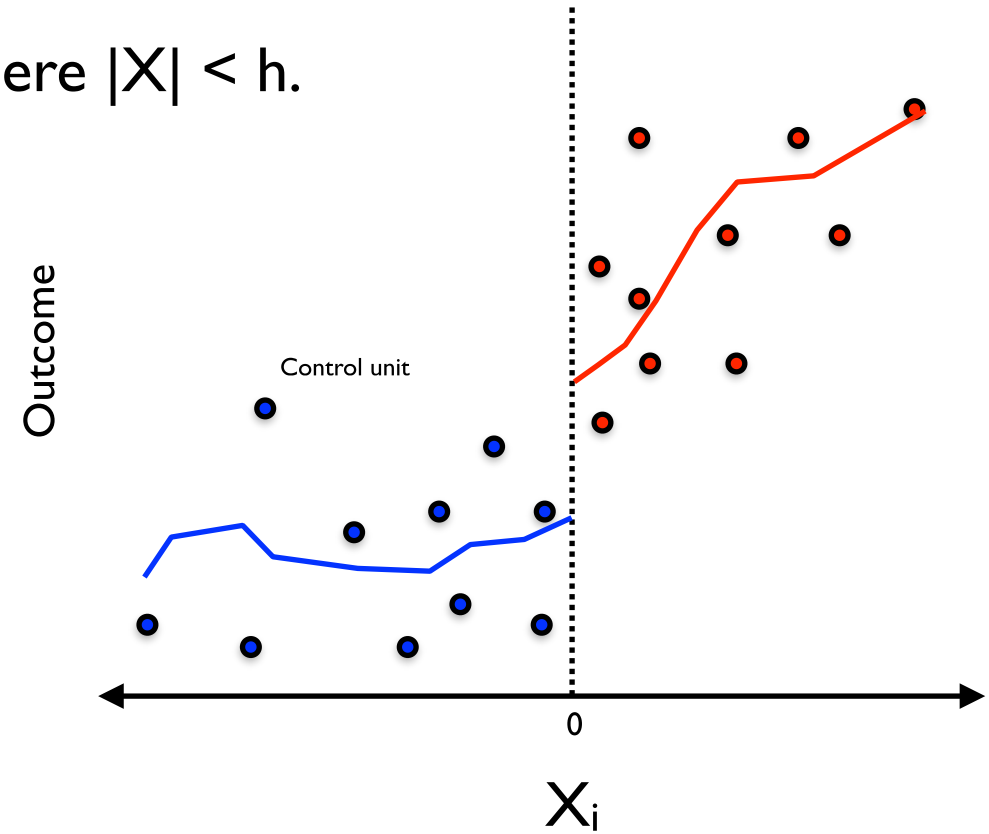
$Y = \beta_0 + \beta_1 D$

restricting to observations where $|X| < h$.

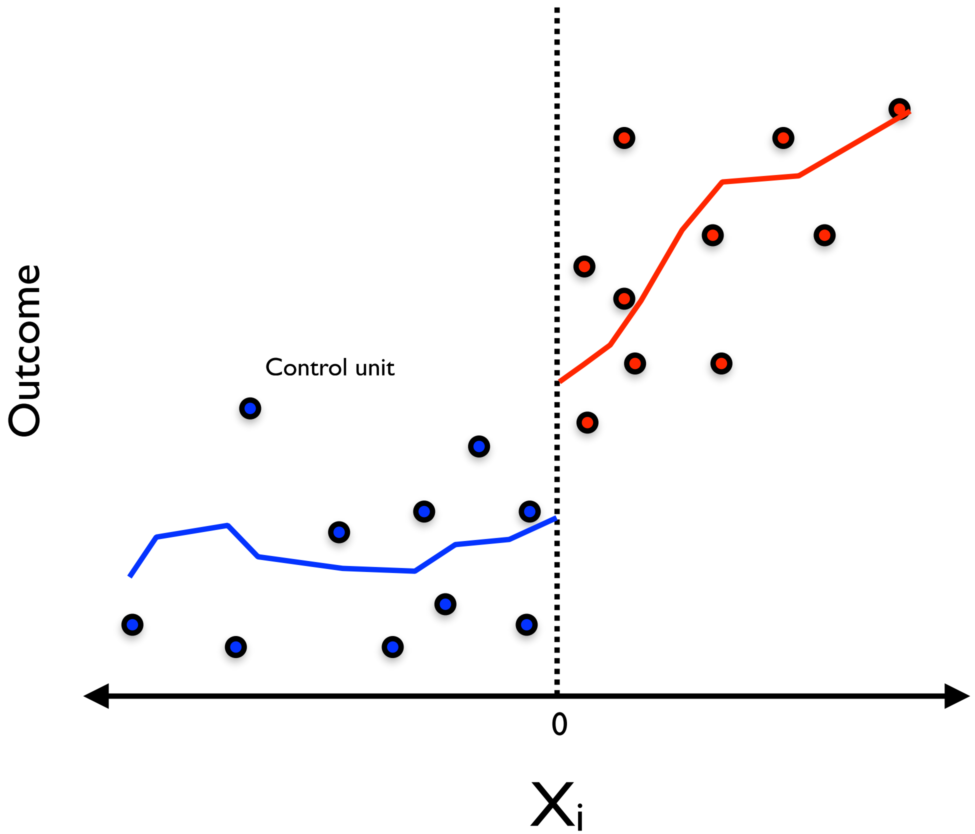
Questions:

- What value of h (*bandwidth, window*) should you choose?
- Will your estimate be unbiased?

There is a **bias-variance tradeoff**.



# How do we estimate it? Local linear regression

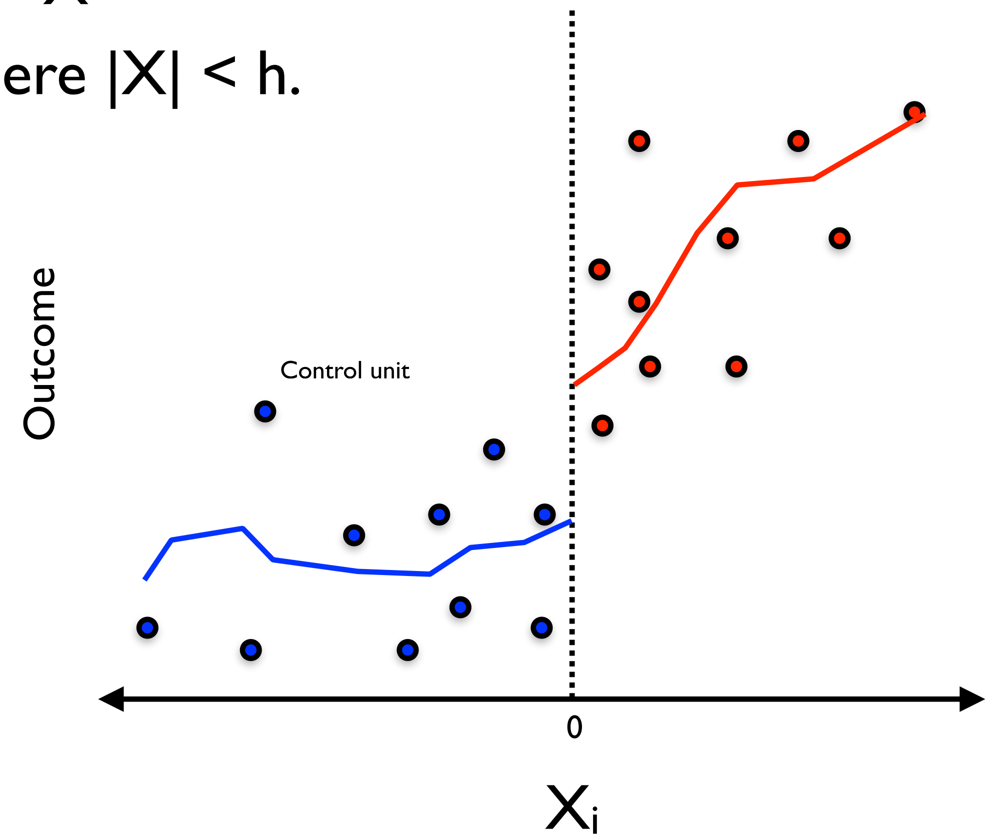


# How do we estimate it? Local linear regression

Consider running this regression:

$Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 D \times X$

restricting to observations where $|X| < h$.



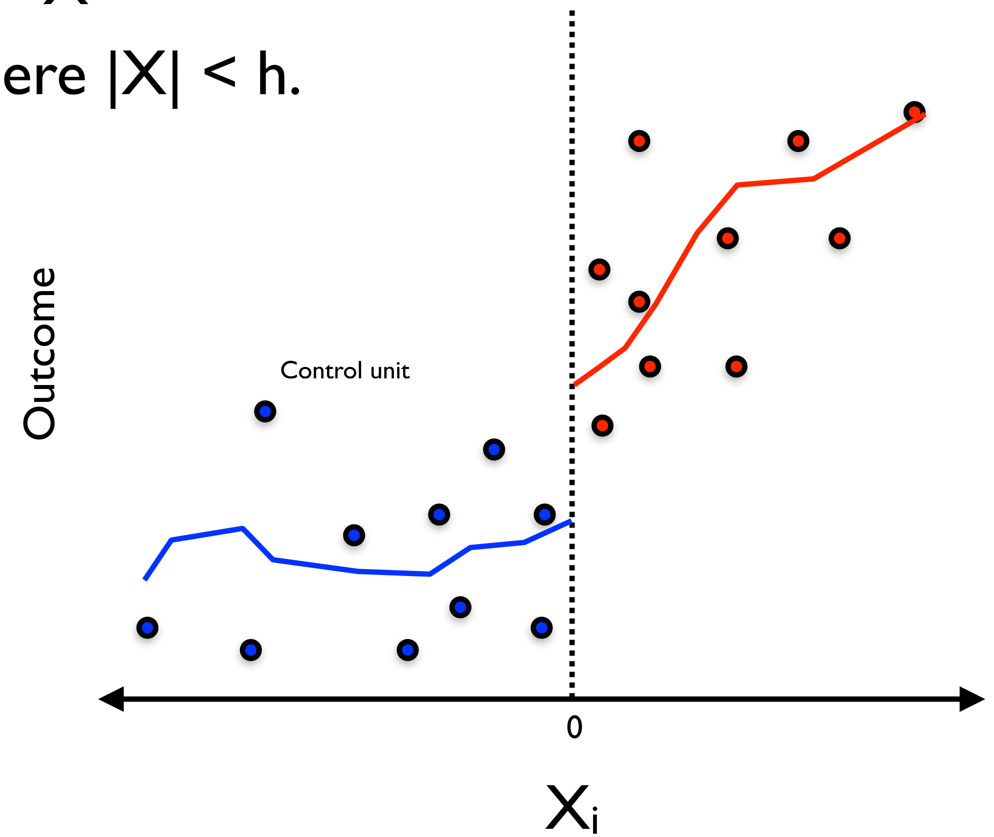
# How do we estimate it? Local linear regression

Consider running this regression:

$Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 D \times X$

restricting to observations where $|X| < h$.

Questions:



# How do we estimate it? Local linear regression

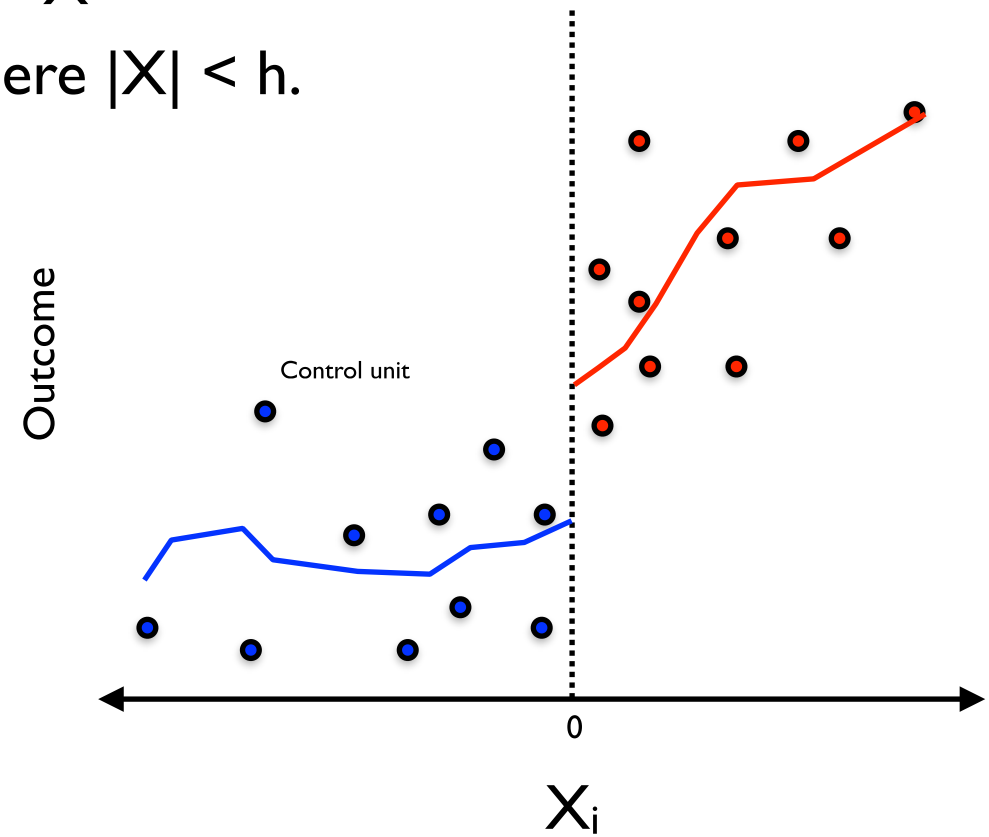
Consider running this regression:

$Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 D \times X$

restricting to observations where $|X| < h$.

Questions:

- What will the fitted equation look like?



# How do we estimate it? Local linear regression

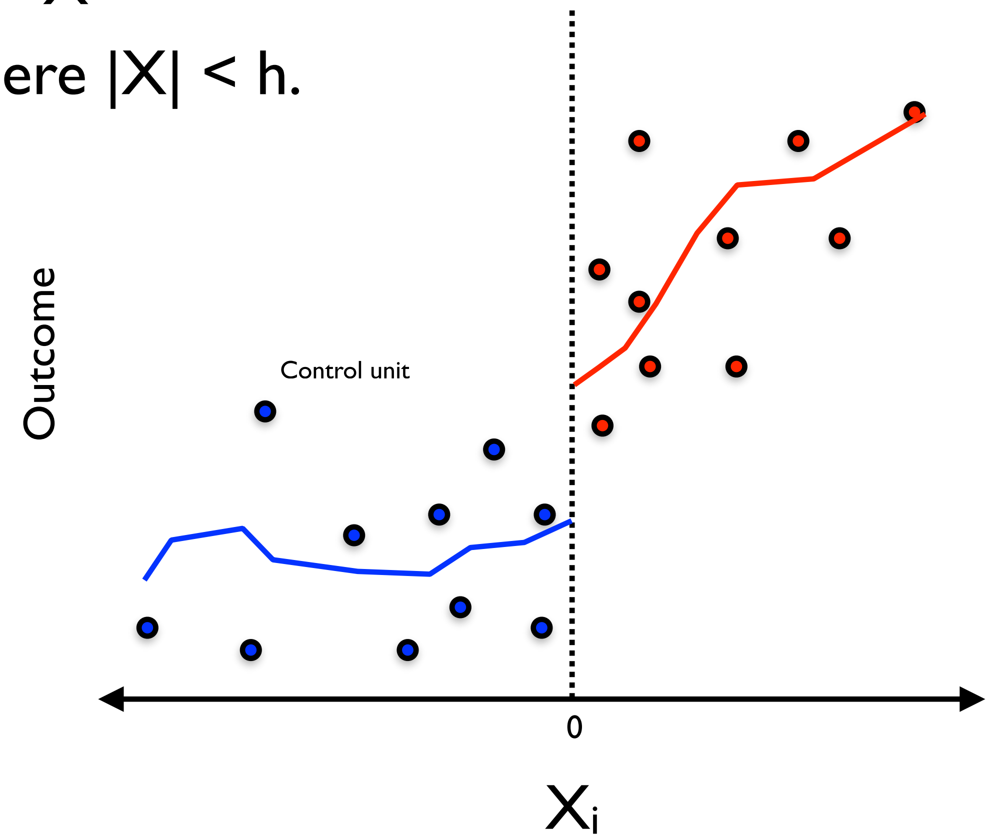
Consider running this regression:

$Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 D \times X$

restricting to observations where $|X| < h$.

Questions:

- What will the fitted equation look like?
- Which coefficient is the estimated treatment effect?



# How do we estimate it? Local linear regression

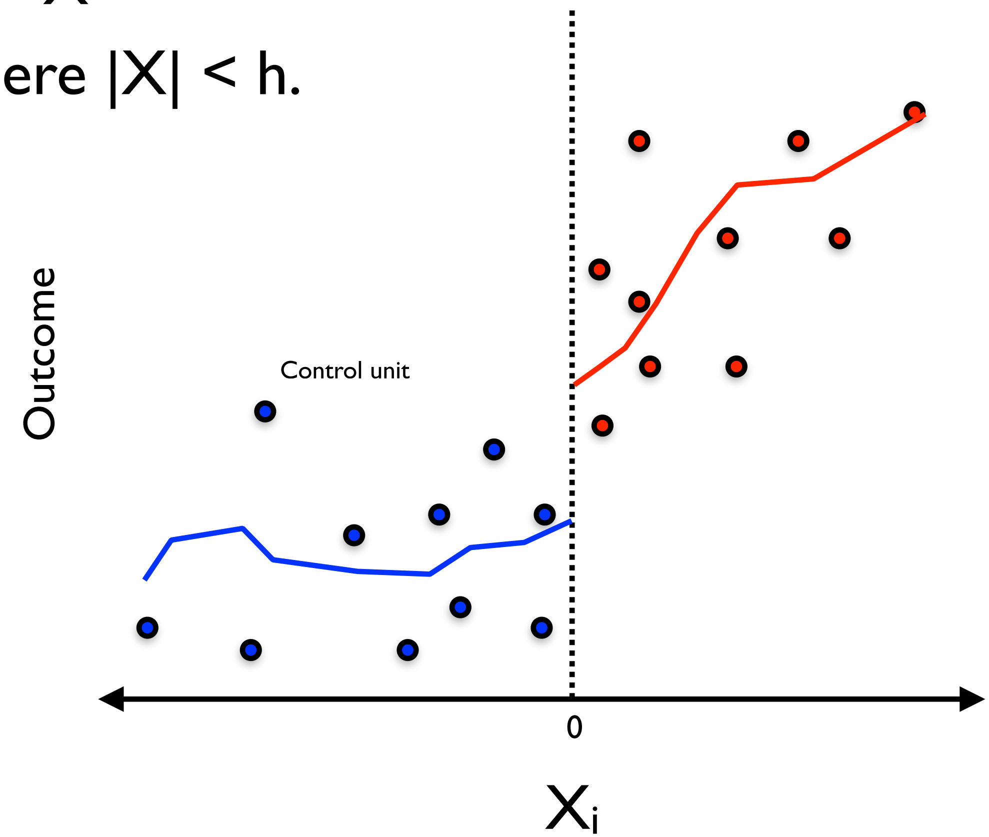
Consider running this regression:

$Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3\, D \times X$

restricting to observations where $|X| < h$.

Questions:

- What will the fitted equation look like?
- Which coefficient is the estimated treatment effect?
- What value of h (*bandwidth, window*) should you choose?



# How do we estimate it? Local linear regression

Consider running this regression:

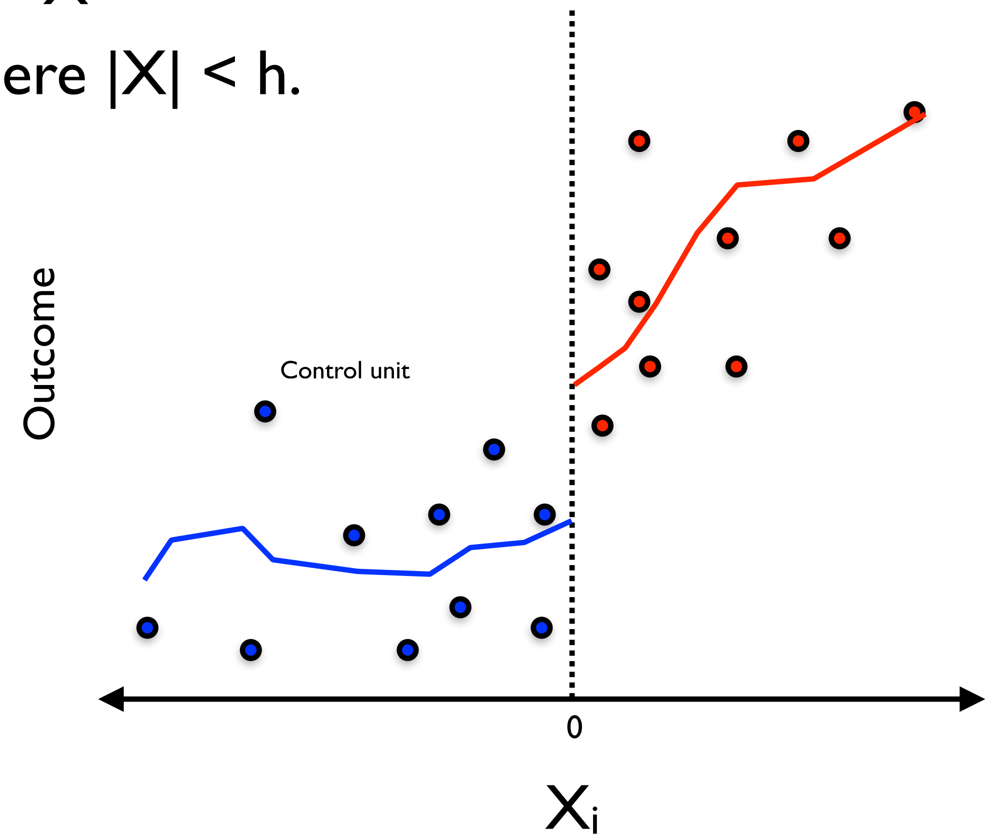
$Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 D \times X$

restricting to observations where $|X| < h$.

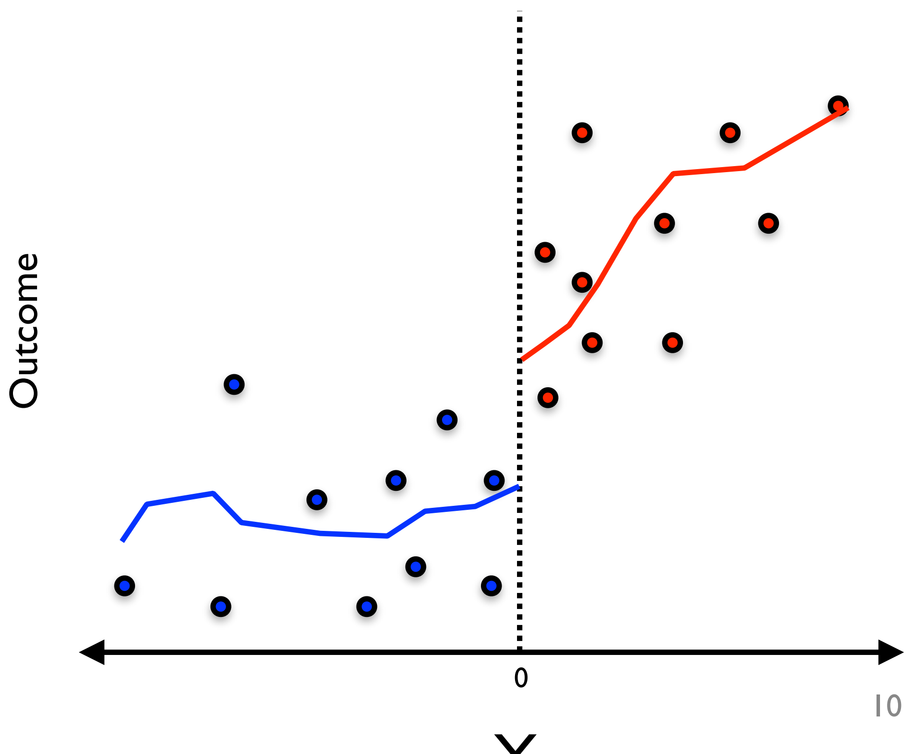
Questions:

- What will the fitted equation look like?
- Which coefficient is the estimated treatment effect?
- What value of h (*bandwidth, window*) should you choose?

There is a **bias-variance tradeoff**.

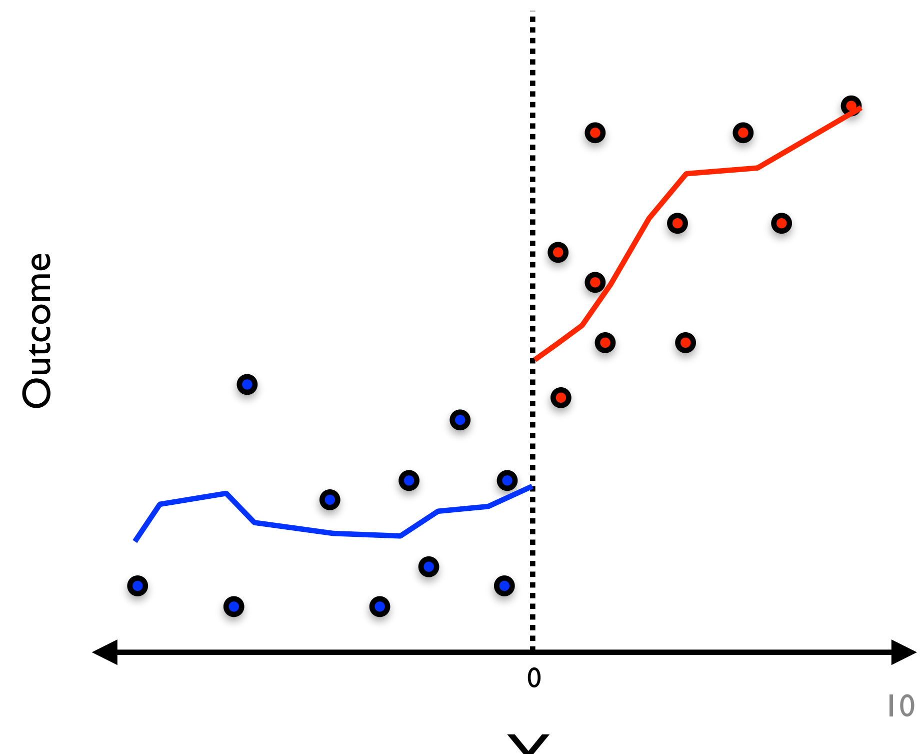


# What people do now



# What people do now

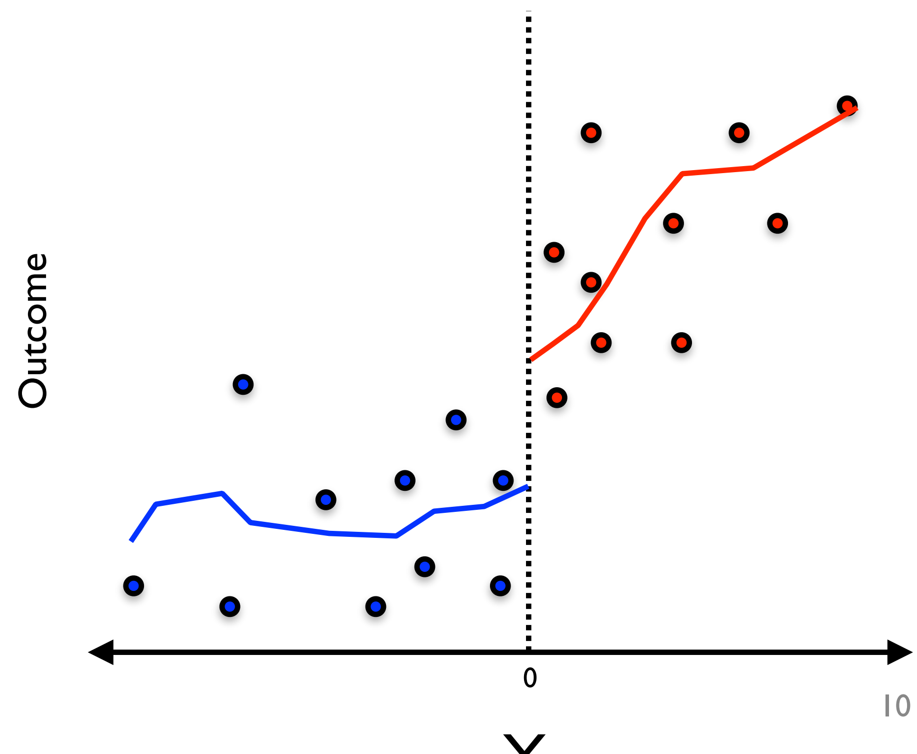
Most common approach: local regression, as on previous slide, with:



# What people do now

Most common approach: local regression, as on previous slide, with:

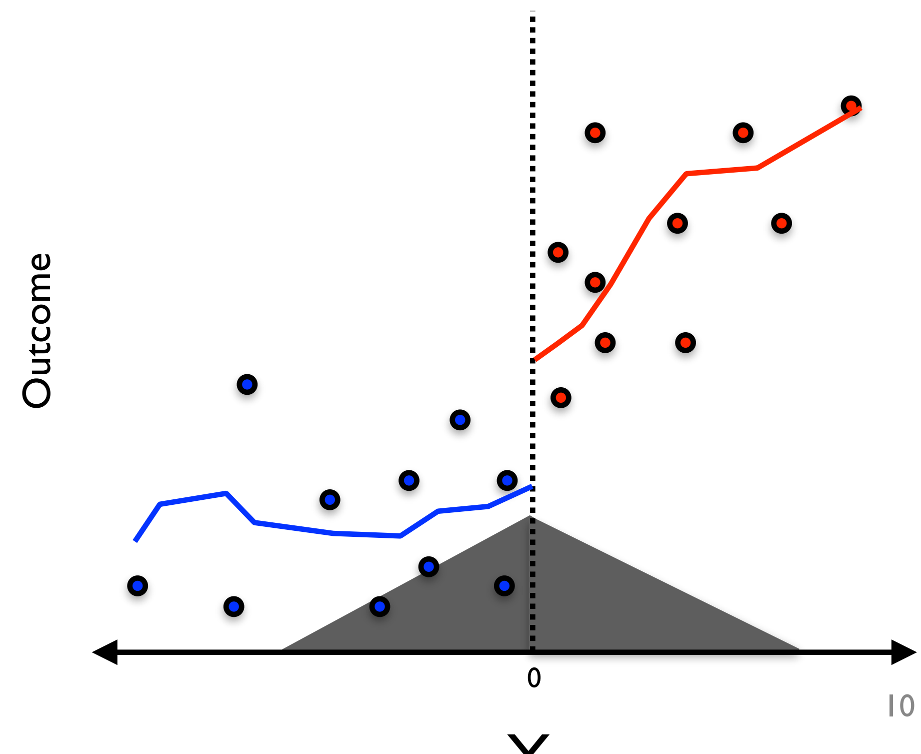
- bandwidth (h) chosen by algorithm that considers curvature of CEF and density of data



# What people do now

Most common approach: local regression, as on previous slide, with:

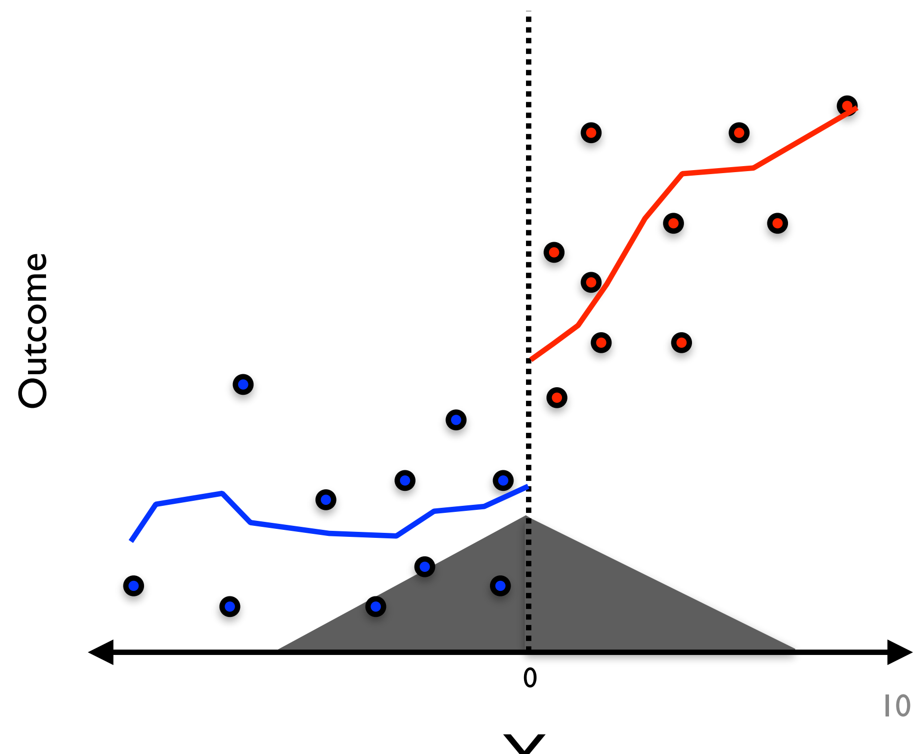
- bandwidth (h) chosen by algorithm that considers curvature of CEF and density of data
- weights that are larger closer to the threshold (triangular kernel)



# What people do now

Most common approach: local regression, as on previous slide, with:

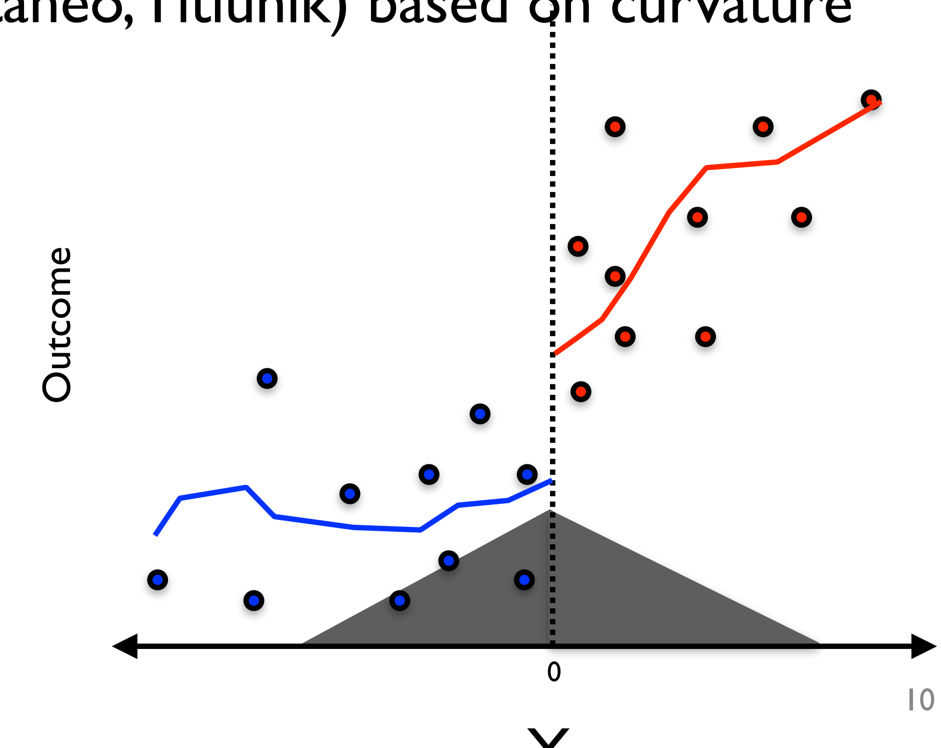
- bandwidth (h) chosen by algorithm that considers curvature of CEF and density of data
- weights that are larger closer to the threshold (triangular kernel)
- (sometimes) polynomial terms (square, cube, etc) in estimating CEF



# What people do now

Most common approach: local regression, as on previous slide, with:

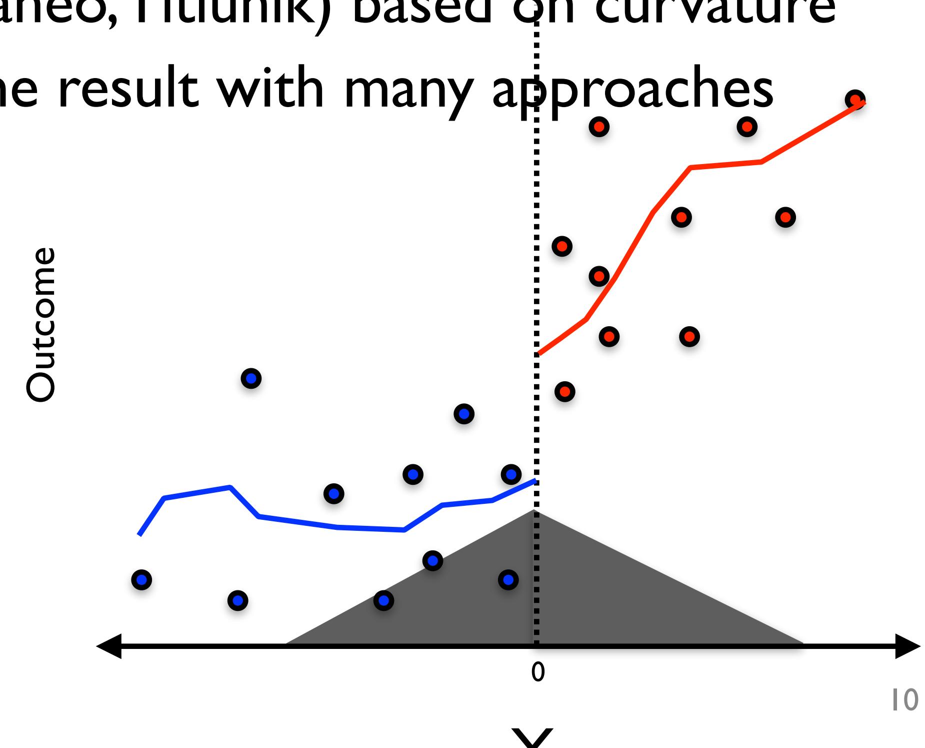
- bandwidth (h) chosen by algorithm that considers curvature of CEF and density of data
- weights that are larger closer to the threshold (triangular kernel)
- (sometimes) polynomial terms (square, cube, etc) in estimating CEF
- bias correction (Calonico, Cattaneo, Titiunik) based on curvature



# What people do now

Most common approach: local regression, as on previous slide, with:

- bandwidth (h) chosen by algorithm that considers curvature of CEF and density of data
- weights that are larger closer to the threshold (triangular kernel)
- (sometimes) polynomial terms (square, cube, etc) in estimating CEF
- bias correction (Calonico, Cattaneo, Titiunik) based on curvature
- robustness checks to show same result with many approaches



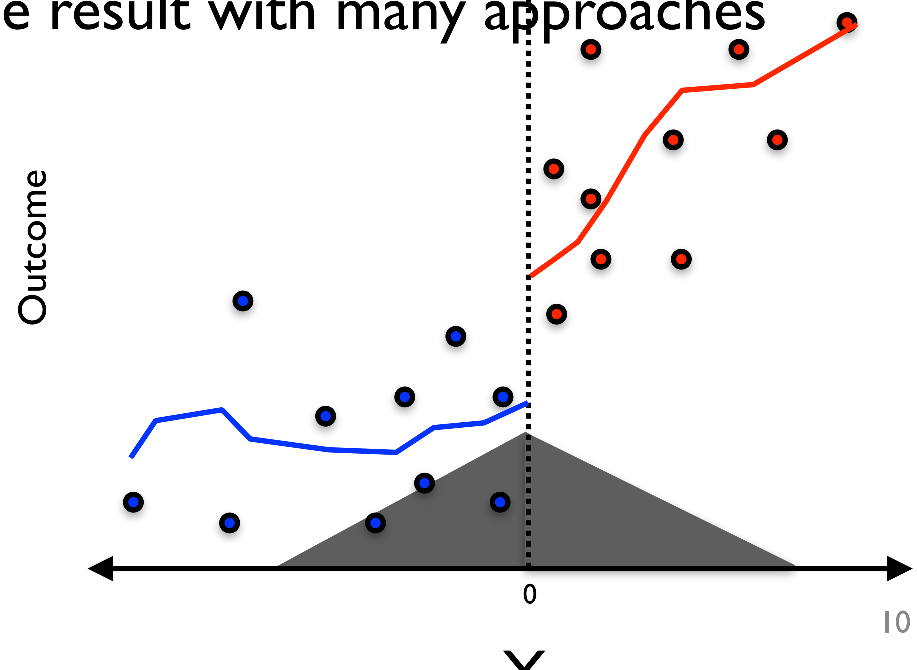
# What people do now

Most common approach: local regression, as on previous slide, with:

- bandwidth (h) chosen by algorithm that considers curvature of CEF and density of data
- weights that are larger closer to the threshold (triangular kernel)
- (sometimes) polynomial terms (square, cube, etc) in estimating CEF
- bias correction (Calonico, Cattaneo, Titiunik) based on curvature
- robustness checks to show same result with many approaches

State of the art: `rdrobust` package (Stata, R), other work by Calonico, Cattaneo, Titiunik

Rocío Titiunik



# What can go wrong?

# What can go wrong?

Remember, basic idea of RDD is to estimate LATE at threshold, where $Y_1$ is estimated using data just above threshold and $Y_0$ using data just below.

# What can go wrong?

Remember, basic idea of RDD is to estimate LATE at threshold, where $Y_1$ is estimated using data just above threshold and $Y_0$ using data just below.

What can go wrong?

# What can go wrong?

Remember, basic idea of RDD is to estimate LATE at threshold, where $Y_1$ is estimated using data just above threshold and $Y_0$ using data just below.

What can go wrong?

- Bias from **misspecification** (see previous slides — depends on algorithm, also curvature of CEF, amount of data)

# What can go wrong?

Remember, basic idea of RDD is to estimate LATE at threshold, where $Y_1$ is estimated using data just above threshold and $Y_0$ using data just below.

What can go wrong?

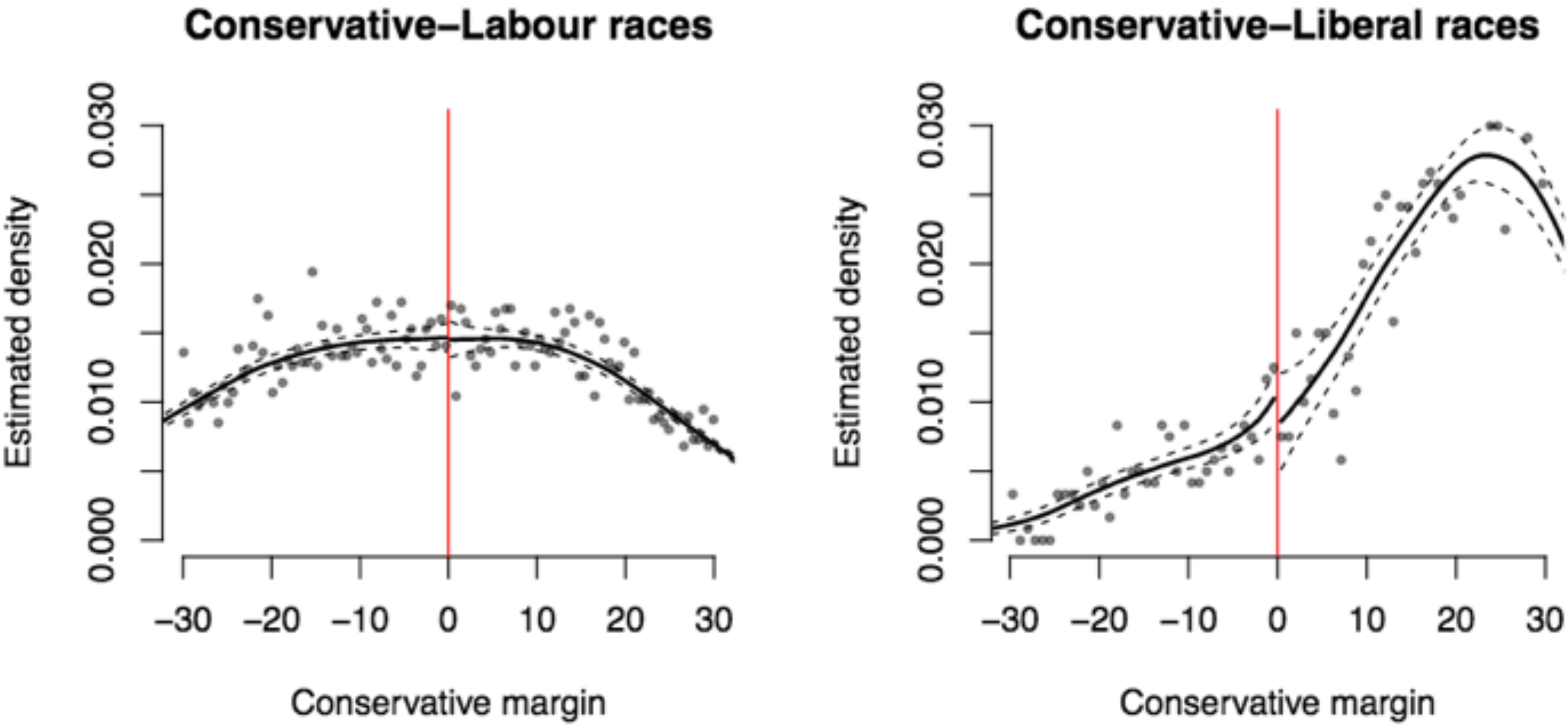
- Bias from **misspecification** (see previous slides — depends on algorithm, also curvature of CEF, amount of data)
- **Sorting**: units just above the threshold may differ substantially from those just below if *some units choose their X* (and thus treatment), and the *ability to choose depends on variables that affect outcome*.

# So how do we check for sorting?

- Is the density of the running variable continuous across the threshold? (McCrary 2008)
- Are covariates (e.g. incumbency status, mayor characteristics, lagged outcomes) continuous across the threshold?

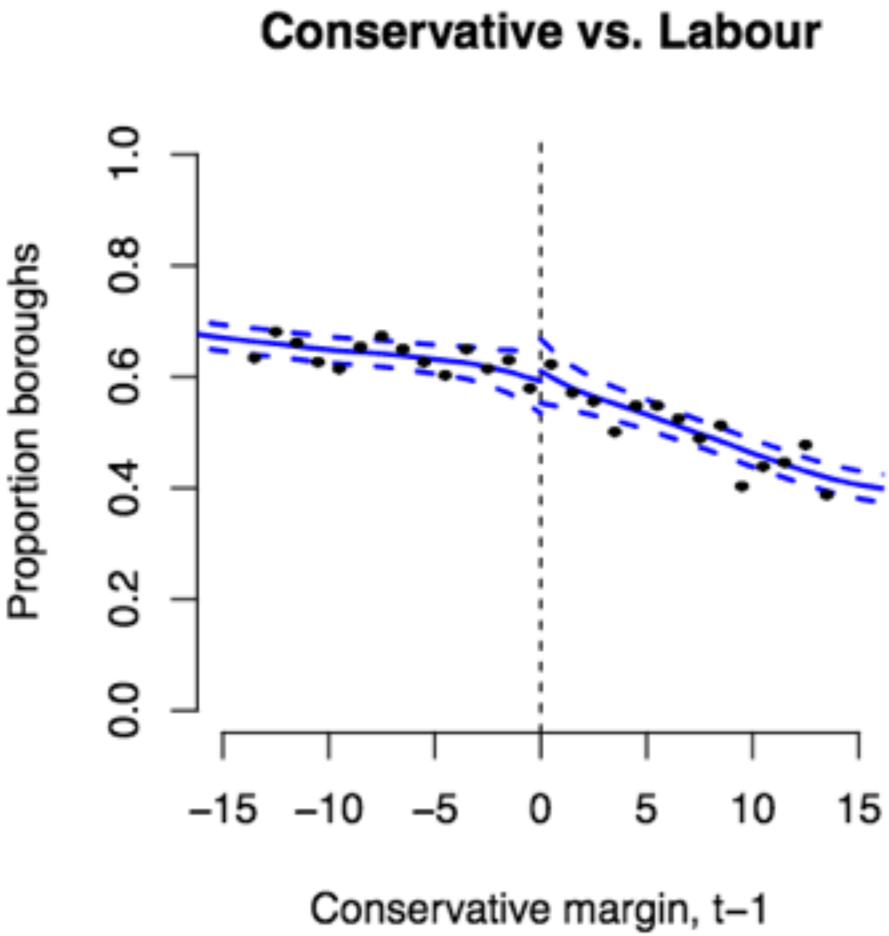
# Example 1: tests of the continuity assumption (1)

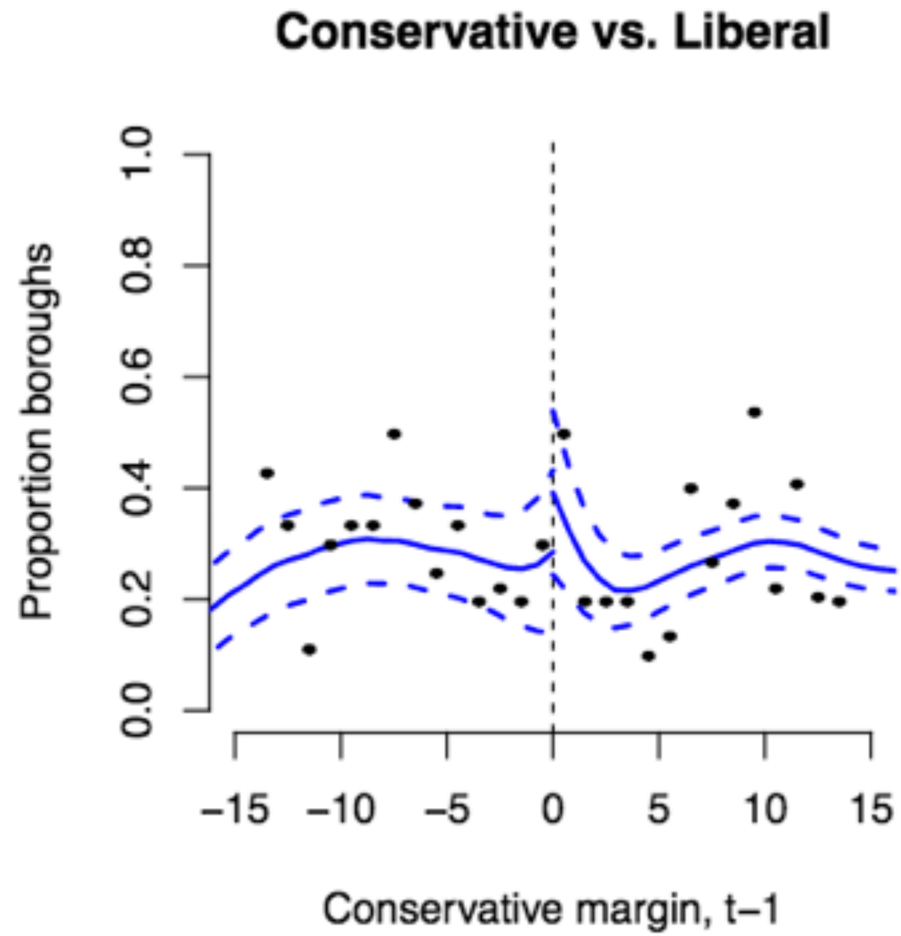
McCrary test for continuity in the density



# Example 1: tests of the continuity assumption (2)

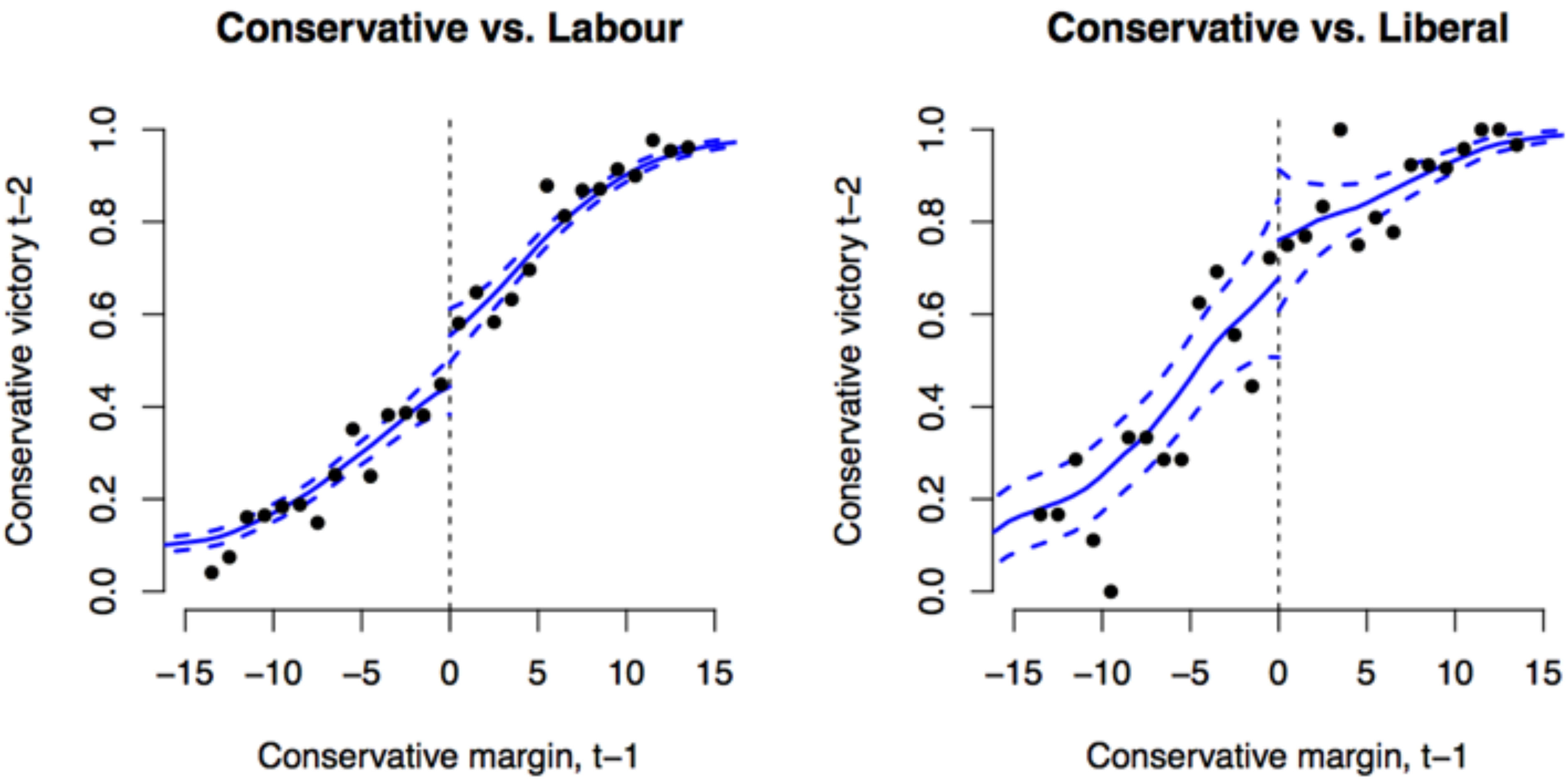
Tests for continuity in covariate: whether or not the election took place in a borough (vs county) constituency



# Example 1: tests of the continuity assumption (3)

Tests for continuity in covariate: whether or not the Conservatives won the *previous* election



# Example 2: Eggers (2015) on turnout and proportionality

# Example 2: Eggers (2015) on turnout and proportionality

**Research question:** Does a more proportional electoral system lead to more voter turnout on average?

# Example 2: Eggers (2015) on turnout and proportionality

**Research question:** Does a more proportional electoral system lead to more voter turnout on average?

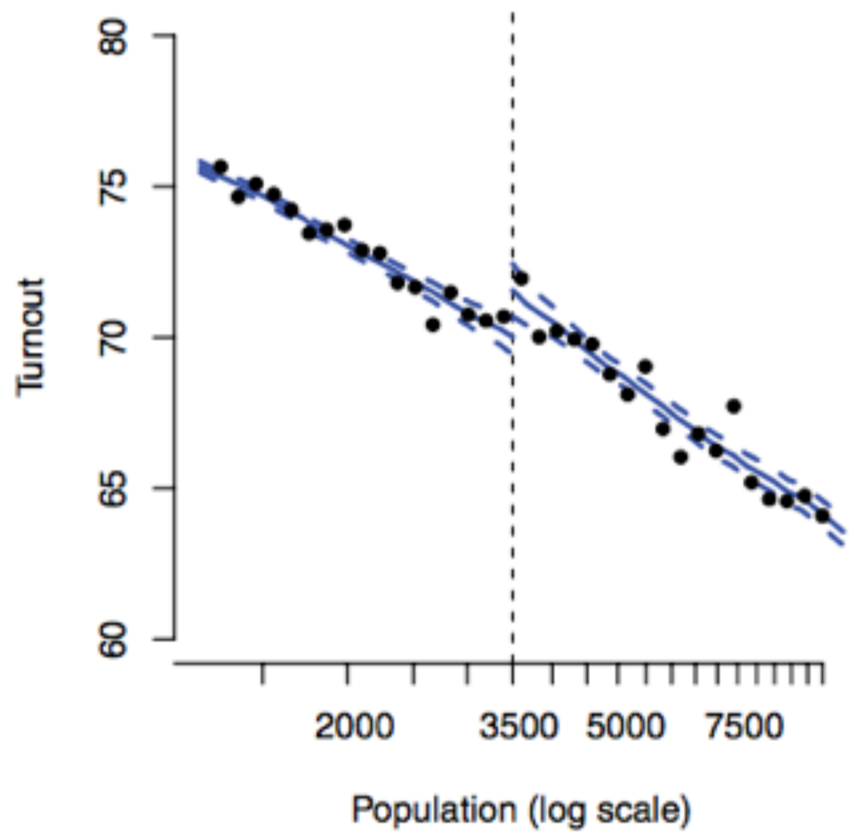
**Strategy:** Use RDD to compare French municipalities using PR and plurality systems at the 3,500 population cutoff determining electoral system

# Example 2: Eggers (2015) on turnout and proportionality

**Research question:** Does a more proportional electoral system lead to more voter turnout on average?

**Strategy:** Use RDD to compare French municipalities using PR and plurality systems at the 3,500 population cutoff determining electoral system

**Turnout in 2001 municipal elections, by municipal population**



# Example 2: testing the continuity assumption

another. One standard way of checking the validity of the RDD, due to McCrary (2008), involves testing for a jump in the density of the forcing variable at the threshold; in this case, McCrary (2008)'s test fails to reject the null ($p$ = .127). Another standard validity check is to carry out RDD analysis in which pre-treatment covariates serve as outcome variables. Table A1 in the appendix reports RDD effect estimates at varying population windows (25%, 50%, and 75%), showing that there is (as one would expect) no "effect" of crossing the 3,500 population threshold on the vast majority of placebo outcomes. These tests suggest that cities just above and below the population threshold are indeed comparable in not just observed but also unobservable features (e.g., local political culture). **(Page 144)**

# Example 2: testing the continuity assumption (2)

Estimated effect of crossing 3,500 on turnout in municipal elections and higher-level elections

| Outcome | Mean turnout | Effect (1) | (2) | (3) |
|---|---|---|---|---|
| Municipal, 2001 | 70.73 | 0.989 | 1.537** | 1.525*** |
| | | (0.778) | (0.538) | (0.433) |
| Municipal, 2008 | 69.14 | 0.763 | 0.929† | 1.476*** |
| | | (0.765) | (0.523) | (0.423) |
| Municipal, 2001 & 2008 | 69.96 | 0.878 | 1.242** | 1.502*** |
| | | (0.71) | (0.481) | (0.385) |
| Presidential, 2002 | 74.95 | −0.04 | −0.189 | −0.038 |
| | | (0.413) | (0.29) | (0.241) |
| Regional, 2004 | 63.38 | −0.448 | −0.7† | −0.241 |
| | | (0.583) | (0.414) | (0.341) |
| Presidential, 2007 | 86.33 | −0.248 | −0.439† | −0.253 |
| | | (0.326) | (0.224) | (0.185) |
| *Window:* | 25% | 25% | 50% | 75% |

# Example 2: testing the continuity assumption (3)

**Table A1.** RDD Analysis: The Effect of Crossing the 3,500 Population Threshold on Placebo (Pre-Treatment) Outcomes.

| Outcome | *Mean* | Effect estimates (1) | (2) | (3) |
|---|---|---|---|---|
| Pct. retired, 1999 | 19.07 | 0.711 | 0.313 | 0.24 |
| | | (0.667) | (0.455) | (0.361) |
| Pct. working in agriculture, 1999 | 5.34 | −0.023 | −0.068 | 0.775* |
| | | (0.638) | (0.461) | (0.38) |
| Pct. with "bac" degree, 1999 | 36.45 | −0.713 | −0.758 | −0.187 |
| | | (0.985) | (0.718) | (0.585) |
| Pct. unemployed, 2001 | 12.33 | −0.764 | −0.456 | 0.064 |
| | | (0.762) | (0.542) | (0.441) |
| Log pop., 1990 | 8.03 | 0.025† | 0.013 | 0.011 |
| | | (0.013) | (0.009) | (0.007) |
| Area (sq. km.) | 21.37 | −3.087 | −1.883 | −1.803 |
| | | (2.021) | (1.406) | (1.123) |
| Local tax revenue, 2000 (in thous. euros) | 1,142.57 | 172.027 | 125.215 | 109.848 |
| | | (106.377) | (75.952) | (66.66) |
| VAT compensation fund (FCTVA), 2000 (in thous. euros) | 101.36 | 5.422 | 4.661 | 6.406 |
| | | (8.829) | (6.544) | (5.628) |
| Turnout, 1995 pres. elections | 81.87 | −0.157 | −0.057 | −0.109 |
| | | (0.341) | (0.236) | (0.192) |
| Pct. for Chirac, 1995 pres. elections | 51.87 | 0.633 | 0.772 | 0.585 |
| | | (0.927) | (0.674) | (0.541) |
| Region: Center | 0.08 | −0.033 | −0.006 | −0.005 |
| | | (0.028) | (0.019) | (0.016) |
| Region: West | 0.22 | −0.005 | −0.076* | −0.084*** |
| | | (0.044) | (0.031) | (0.025) |
| Region: South | 0.22 | 0.018 | 0.042 | 0.052† |
| | | (0.045) | (0.032) | (0.026) |
| Region: East | 0.13 | 0.046 | 0.044 | 0.013 |
| | | (0.037) | (0.027) | (0.022) |
| Region: North | 0.16 | −0.019 | 0.004 | 0.018 |
| | | (0.039) | (0.028) | (0.023) |
| Region: SW | 0.13 | 0.013 | 0.003 | 0.015 |
| | | (0.037) | (0.026) | (0.021) |
| Region: Paris | 0.06 | −0.021 | −0.01 | −0.01 |
| | | (0.026) | (0.019) | (0.015) |
| Window | 25% | 25% | 50% | 75% |

# Does RDD work for political science applications? The case of close elections

# Does RDD work for political science applications? The case of close elections

**Caughey and Sekhon (2011):** RDD might not work for close elections — evidence that incumbents disproportionately win very close U.S. congressional elections post 1950. Imbalance (i.e. discontinuity) in incumbency, amount of money raised, predicted winner, many other pre-treatment characteristics.

# Does RDD work for political science applications? The case of close elections

**Caughey and Sekhon (2011):** RDD might not work for close elections — evidence that incumbents disproportionately win very close U.S. congressional elections post 1950. Imbalance (i.e. discontinuity) in incumbency, amount of money raised, predicted winner, many other pre-treatment characteristics.

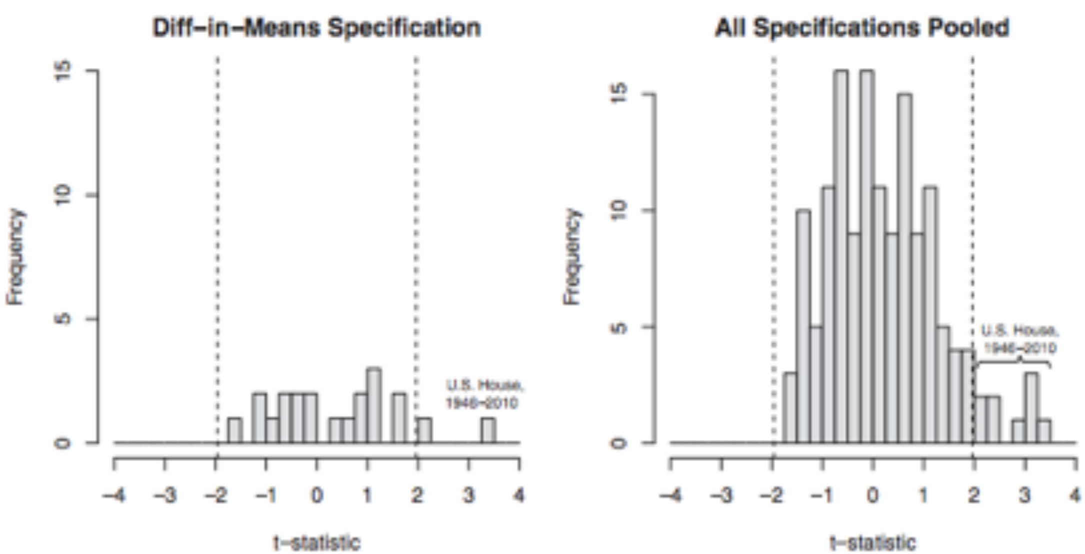
**Eggers, Fowler, Hainmueller, Hall, Snyder (2015):** Looking at other periods and legislatures in US and elsewhere, similar problems not found anywhere else. Caughey and Sekhon (2011) pattern probably a fluke.

# Does RDD work for political science applications? The case of close elections

**Caughey and Sekhon (2011):** RDD might not work for close elections — evidence that incumbents disproportionately win very close U.S. congressional elections post 1950. Imbalance (i.e. discontinuity) in incumbency, amount of money raised, predicted winner, many other pre-treatment characteristics.

**Eggers, Fowler, Hainmueller, Hall, Snyder (2015):** Looking at other periods and legislatures in US and elsewhere, similar problems not found anywhere else. Caughey and Sekhon (2011) pattern probably a fluke.

FIGURE 2 T-values for "Effect" of Party Winning at Time t on Party Winning at Time t − 1



# Does RDD work for political science applications? The case of population thresholds (1)

# Does RDD work for political science applications? The case of population thresholds (1)

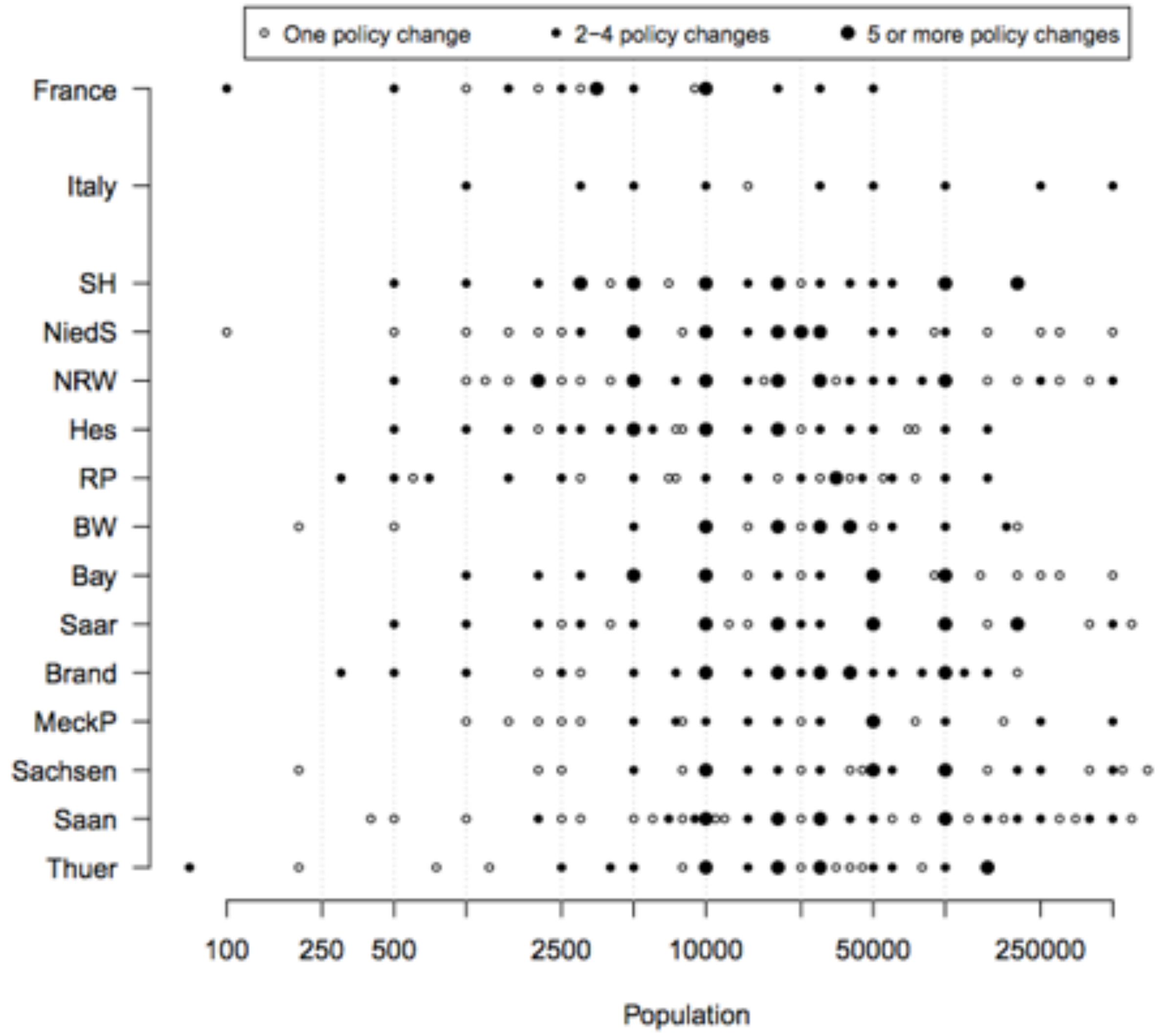
**Eggers, Freier, Grembi, and Nannicini** (forthcoming): There may be more reason to doubt RDDs based on population thresholds.

# Does RDD work for political science applications? The case of population thresholds (1)

**Eggers, Freier, Grembi, and Nannicini** (forthcoming): There may be more reason to doubt RDDs based on population thresholds.

First problem: same threshold often used to determine more than one treatment.

See Eggers, Freier, Grembi, Nannichini (forthcoming) and Eggers (2015) for ideas about handling this.



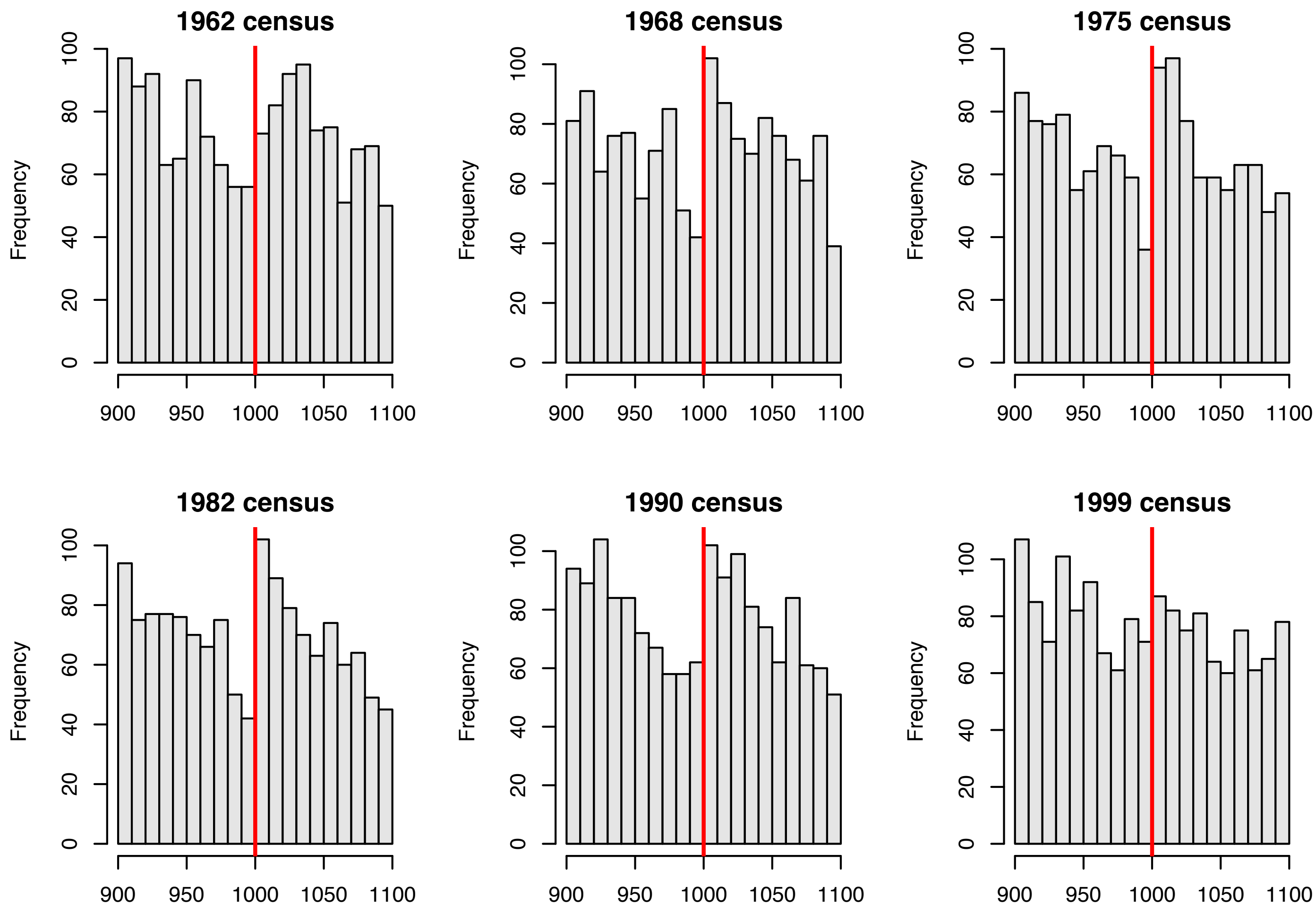
# Does RDD work for political science applications? The case of population thresholds (2)

## Does RDD work for political science applications? The case of population thresholds (2)

Second problem: **sorting**. Consider histograms of population in French villages near 1,000 in pop., where mayor's salary increases:

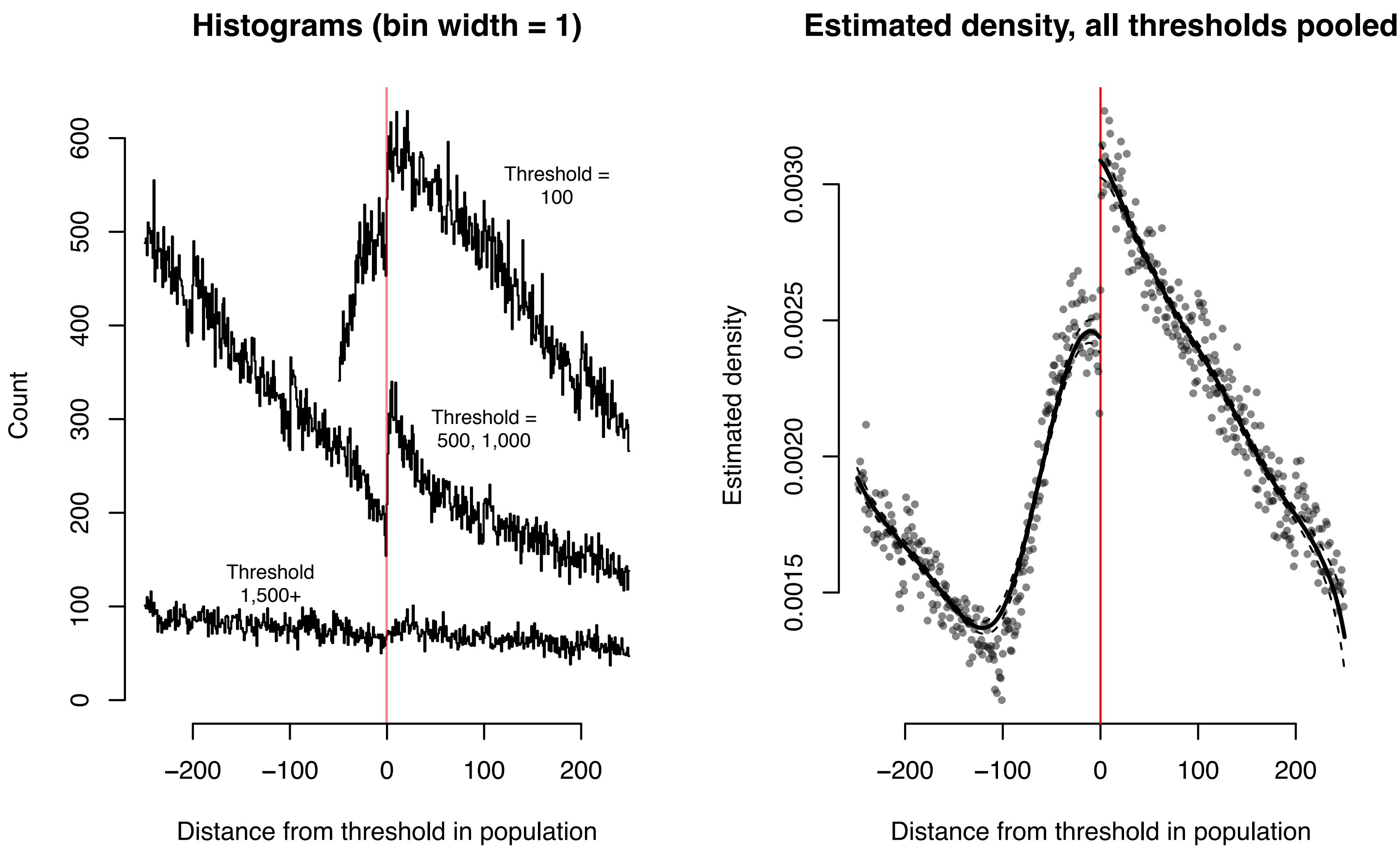
# Does RDD work for political science applications? The case of population thresholds (2)

Second problem: **sorting**. Consider histograms of population in French villages near 1,000 in pop., where mayor's salary increases:



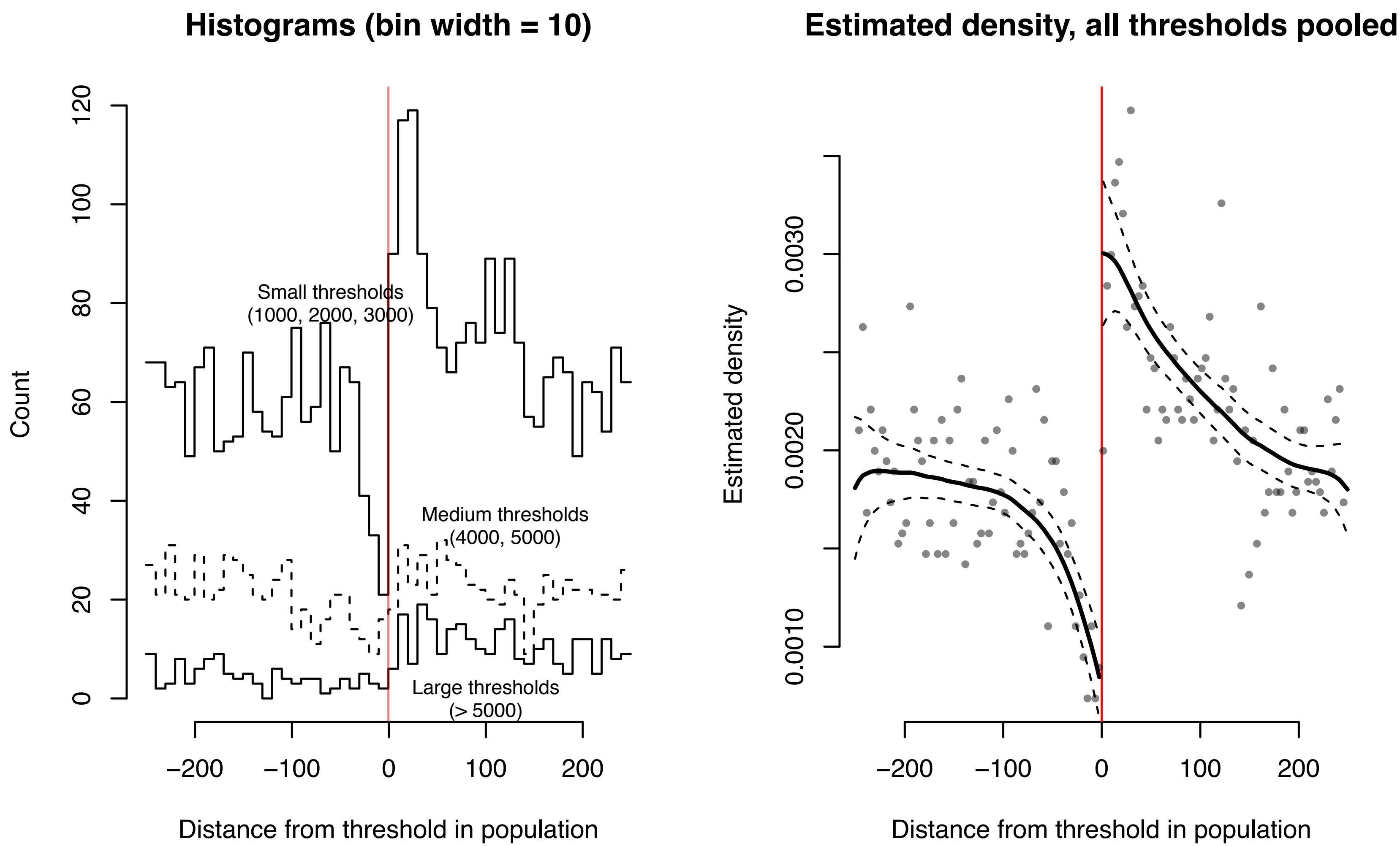
# Does RDD work for political science applications? The case of population thresholds (3)

Pooling all thresholds, censuses from France:



# Does RDD work for political science applications? The case of population thresholds (3)

Even worse in Italy:



# Does RDD work for political science applications?
# General questions about sorting

# Does RDD work for political science applications? General questions about sorting

(1) Why do you think there is sorting in the municipal population case but not (apparently) close elections?

## Does RDD work for political science applications? General questions about sorting

(1) Why do you think there is sorting in the municipal population case but not (apparently) close elections?

(2) If there is sorting, is the RDD ruined?

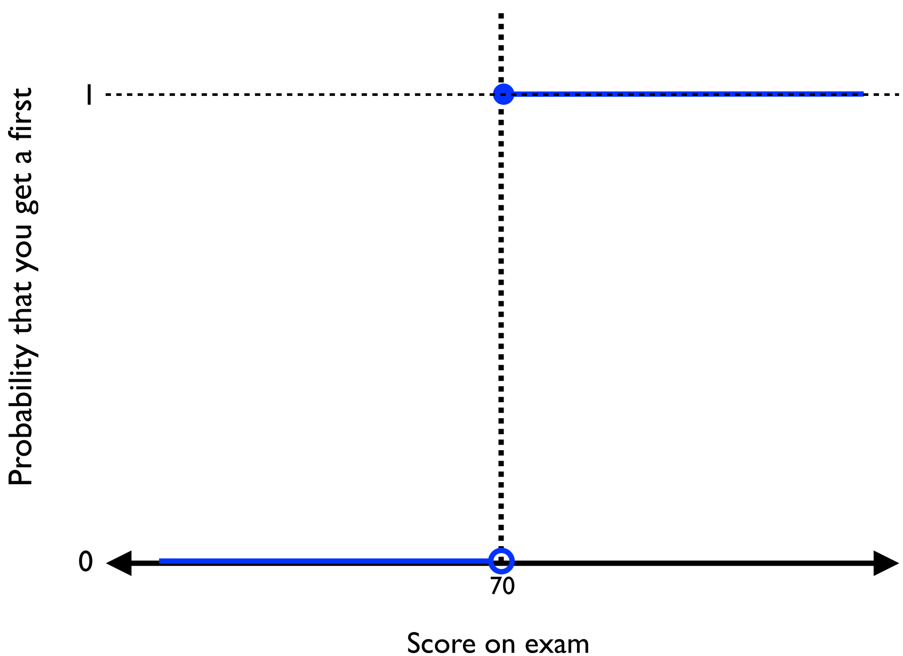
# Fuzzy RDD, briefly

# Fuzzy RDD, briefly

So far, we have only considered cases where the probability of treatment goes from 0 to 1 discontinuously at a threshold.

# Fuzzy RDD, briefly

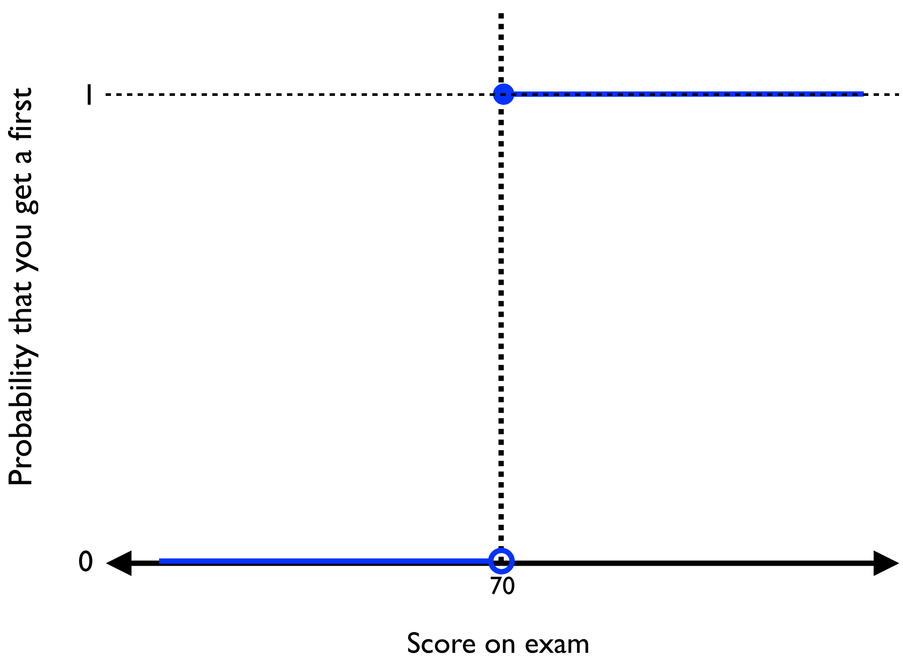
So far, we have only considered cases where the probability of treatment goes from 0 to 1 discontinuously at a threshold.



# Fuzzy RDD, briefly

So far, we have only considered cases where the probability of treatment goes from 0 to 1 discontinuously at a threshold.

**Sharp RDD.**

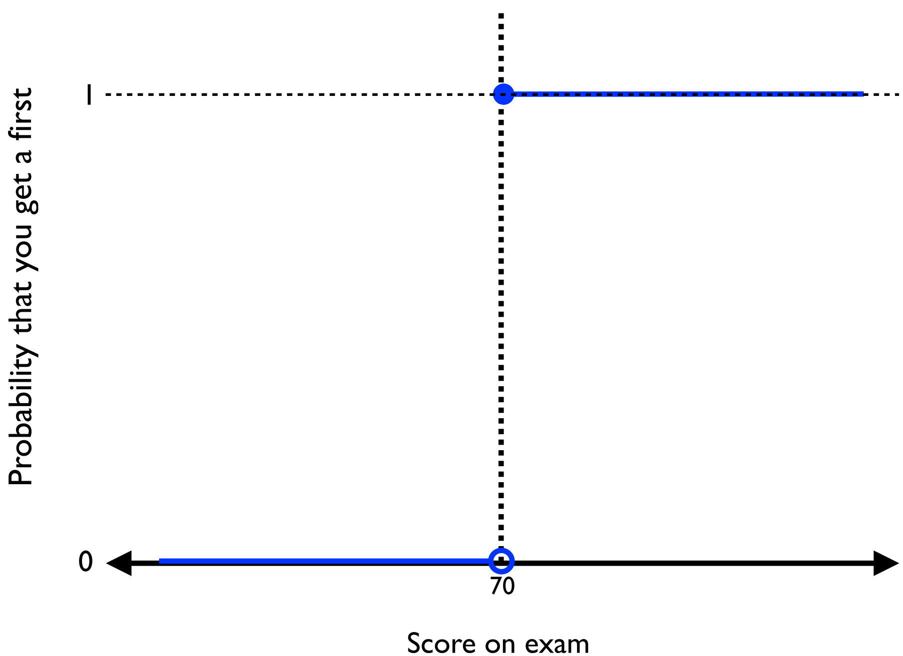


# Fuzzy RDD, briefly

So far, we have only considered cases where the probability of treatment goes from 0 to 1 discontinuously at a threshold.

In other cases, the probability of treatment jumps discontinuously, but not all the way from 0 to 1.

**Sharp RDD.**

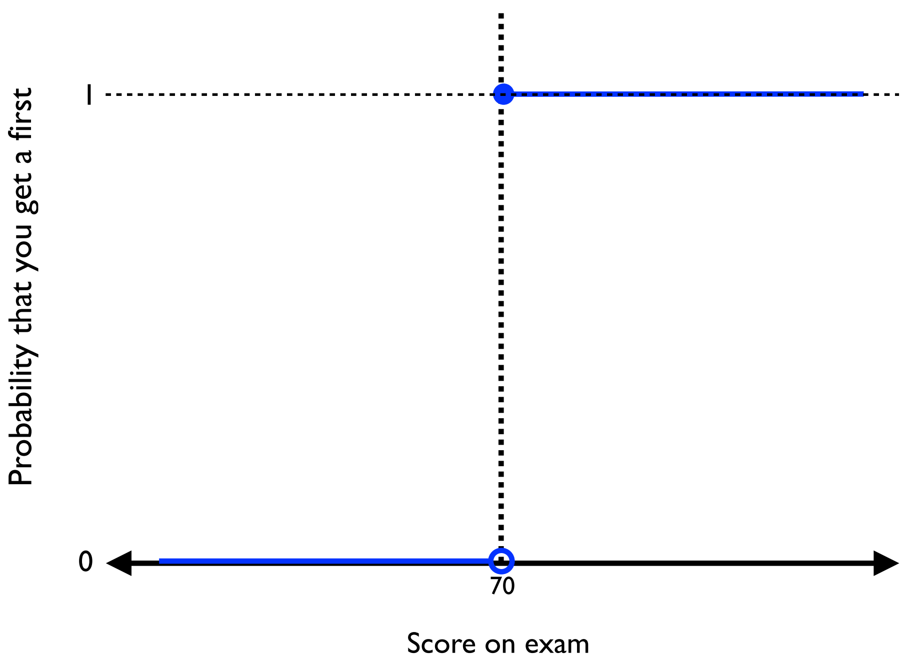


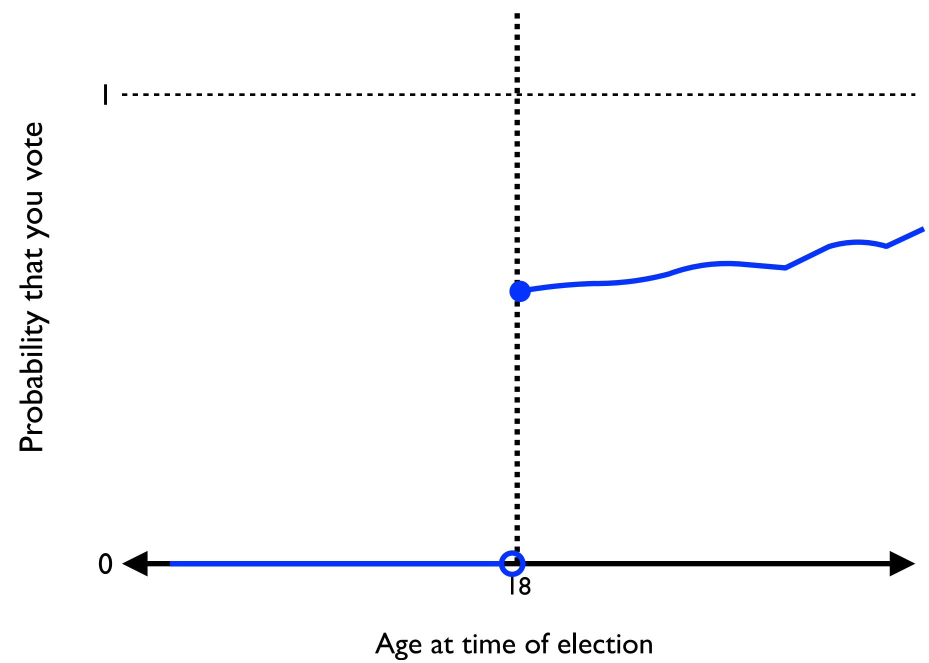
# Fuzzy RDD, briefly

So far, we have only considered cases where the probability of treatment goes from 0 to 1 discontinuously at a threshold.

In other cases, the probability of treatment jumps discontinuously, but not all the way from 0 to 1.

**Sharp RDD.**



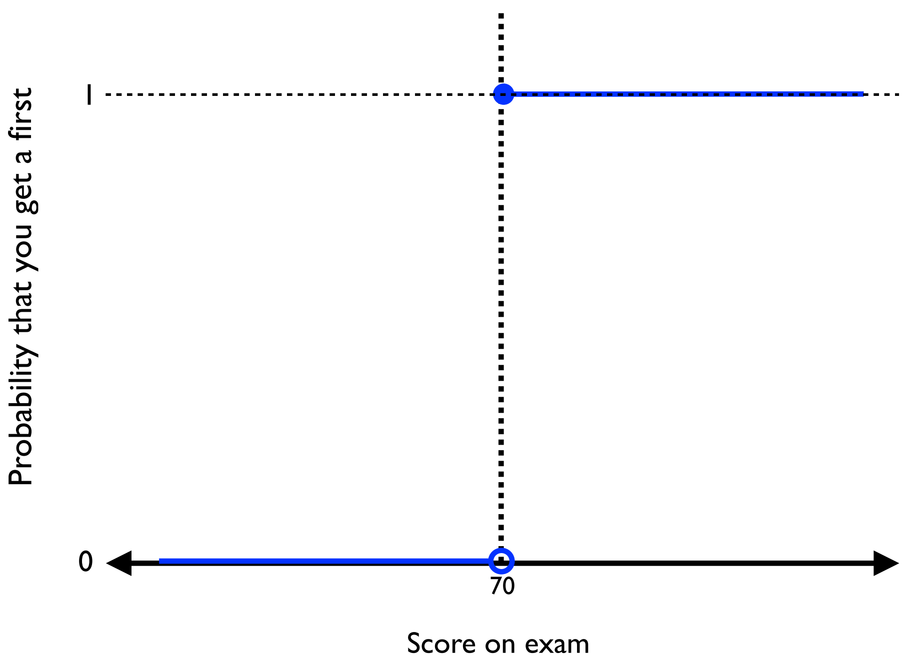



# Fuzzy RDD, briefly

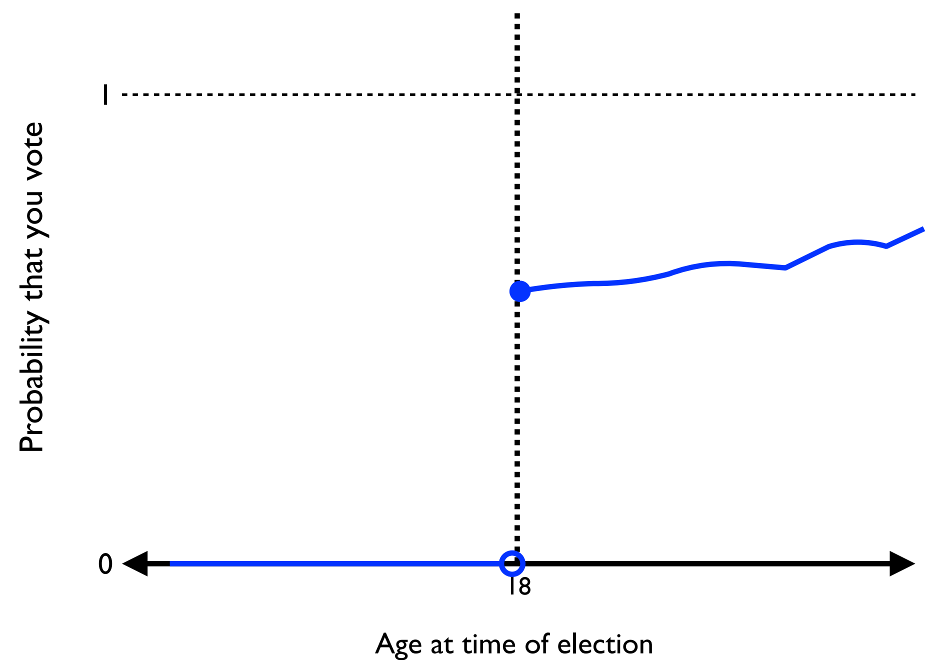
So far, we have only considered cases where the probability of treatment goes from 0 to 1 discontinuously at a threshold.

In other cases, the probability of treatment jumps discontinuously, but not all the way from 0 to 1.

## Sharp RDD.




## Fuzzy RDD.



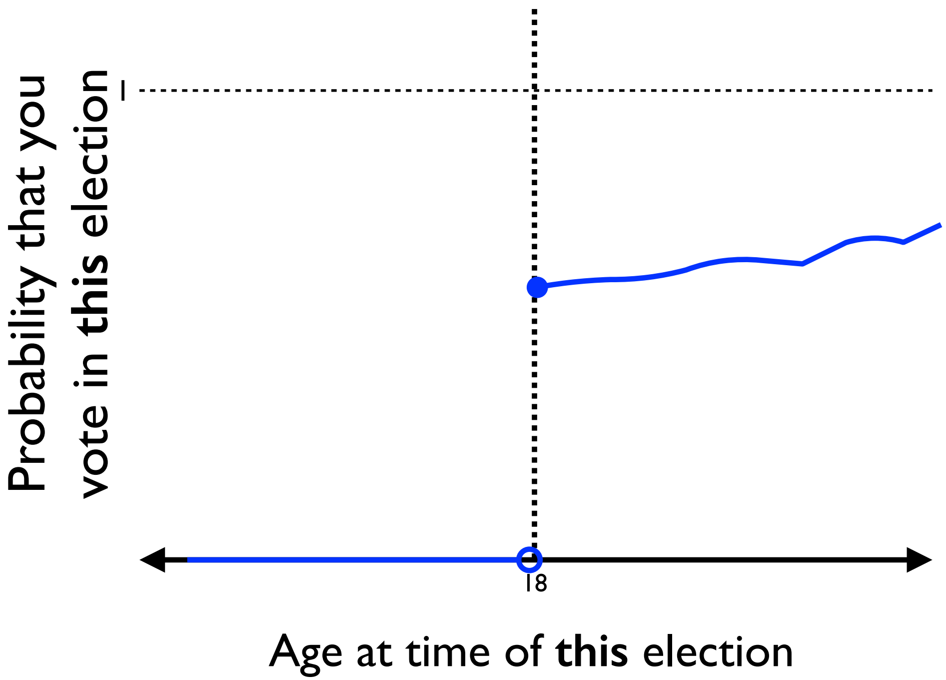
# Fuzzy RDD, briefly (2)

# Fuzzy RDD, briefly (2)

Fuzzy RDD cases can be thought of as having **non-compliance**: the threshold assigns units to treatment, but some units disobey.

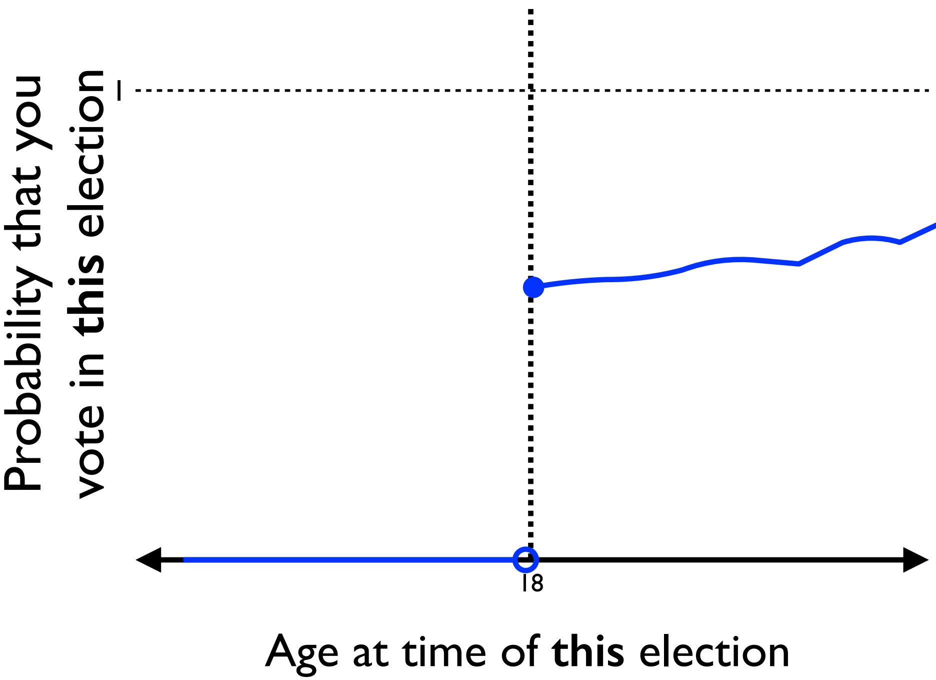
# Fuzzy RDD, briefly (2)

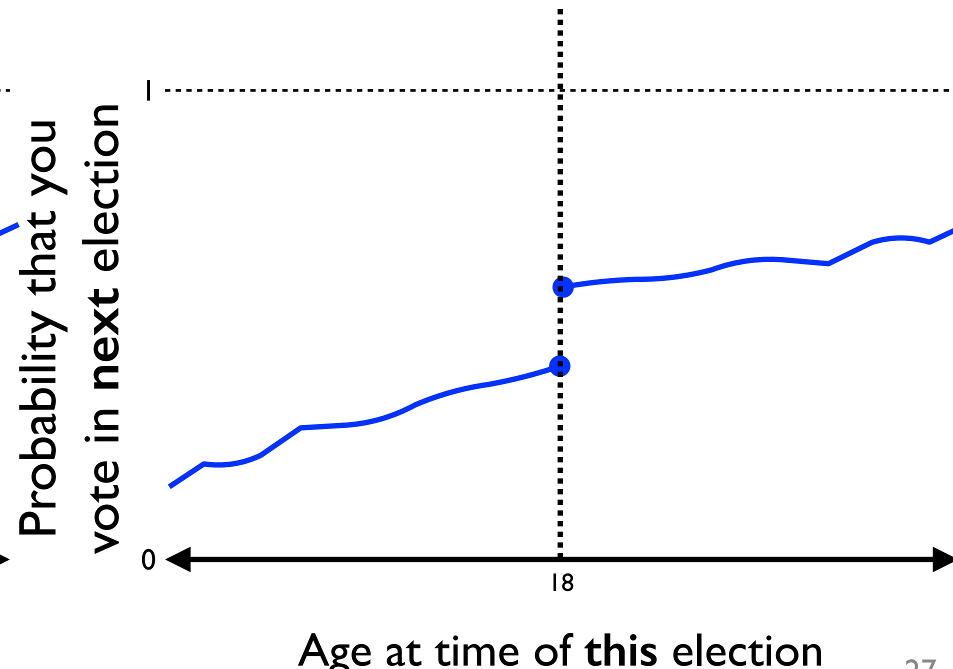
Fuzzy RDD cases can be thought of as having **non-compliance**: the threshold assigns units to treatment, but some units disobey.



# Fuzzy RDD, briefly (2)

Fuzzy RDD cases can be thought of as having **non-compliance**: the threshold assigns units to treatment, but some units disobey.

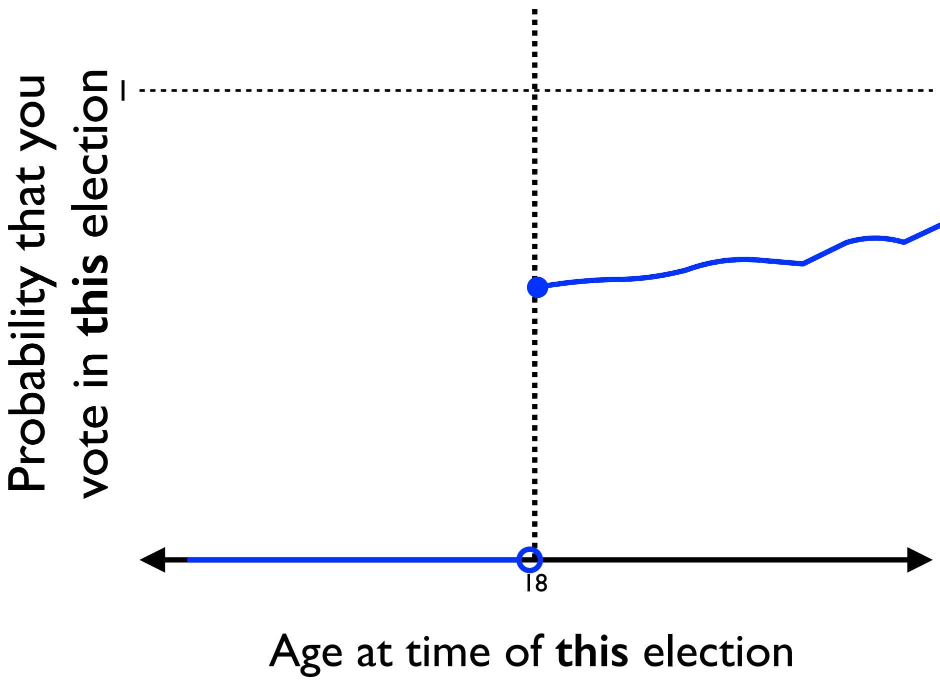


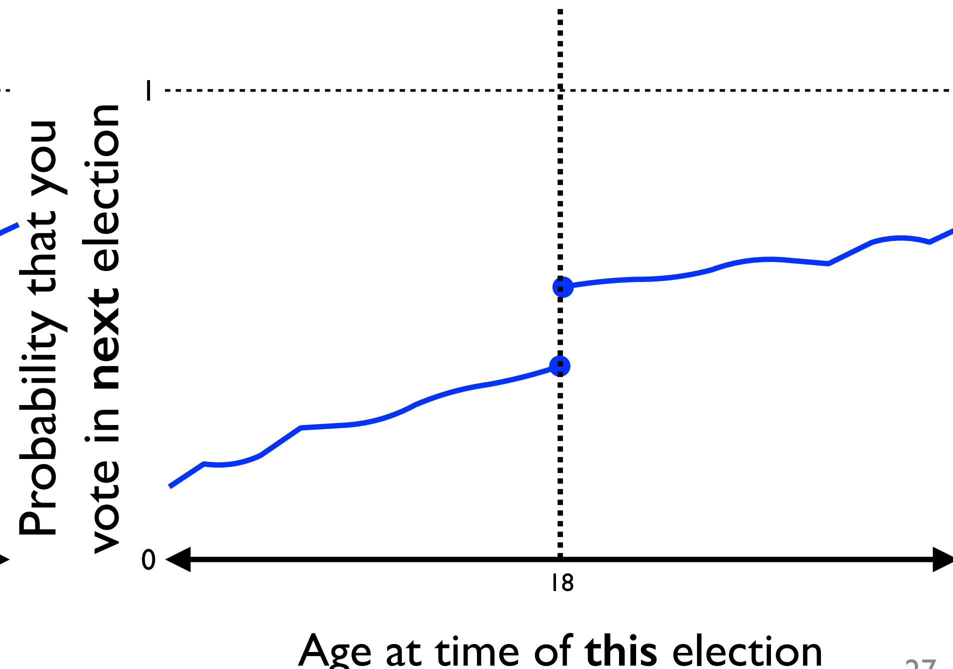



# Fuzzy RDD, briefly (2)

Fuzzy RDD cases can be thought of as having **non-compliance**: the threshold assigns units to treatment, but some units disobey.

Does this remind you of something?



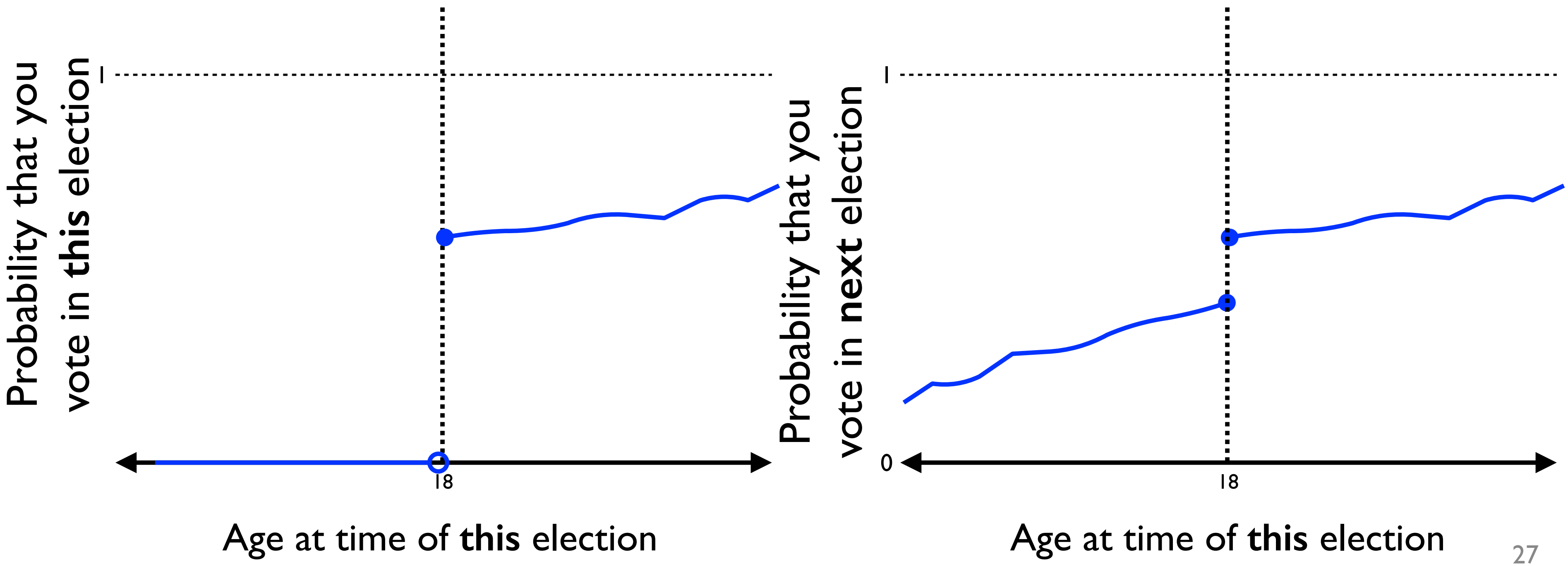



# Fuzzy RDD, briefly (2)

Fuzzy RDD cases can be thought of as having **non-compliance**: the threshold assigns units to treatment, but some units disobey.

Does this remind you of something?

Think of fuzzy RDD as IV and use the Wald estimator:

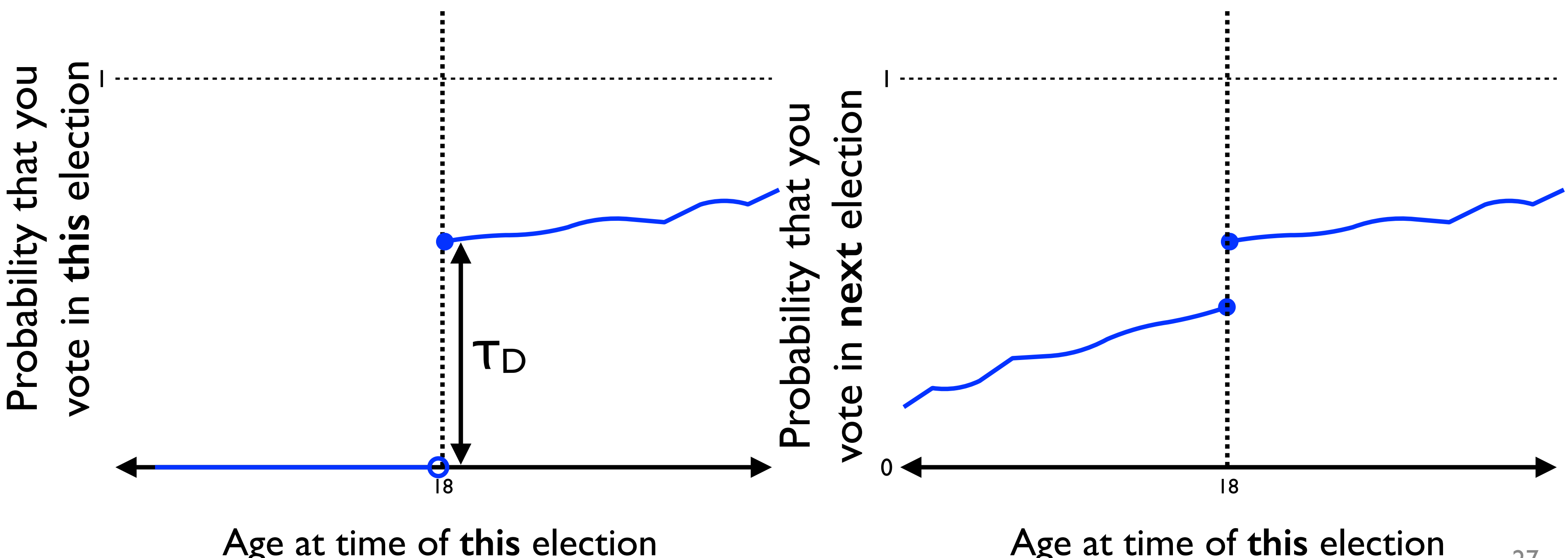


# Fuzzy RDD, briefly (2)

Fuzzy RDD cases can be thought of as having **non-compliance**: the threshold assigns units to treatment, but some units disobey.

Does this remind you of something?

Think of fuzzy RDD as IV and use the Wald estimator:

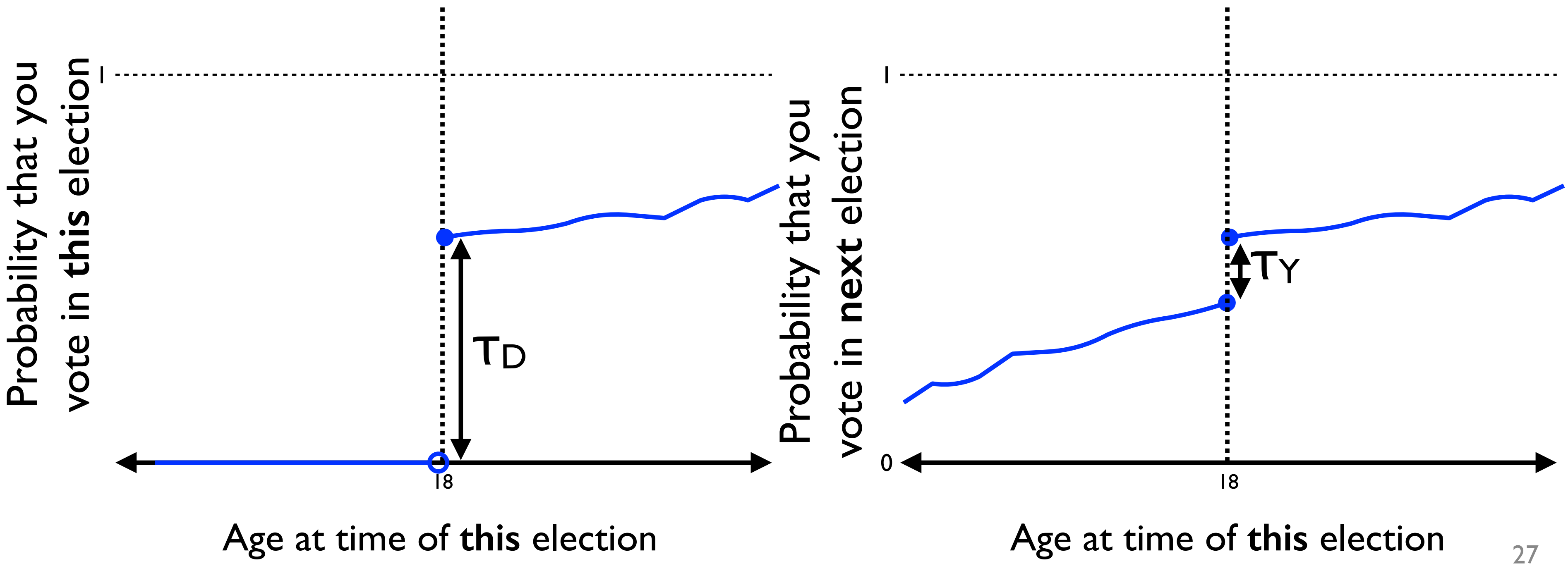


# Fuzzy RDD, briefly (2)

Fuzzy RDD cases can be thought of as having **non-compliance**: the threshold assigns units to treatment, but some units disobey.

Does this remind you of something?

Think of fuzzy RDD as IV and use the Wald estimator:



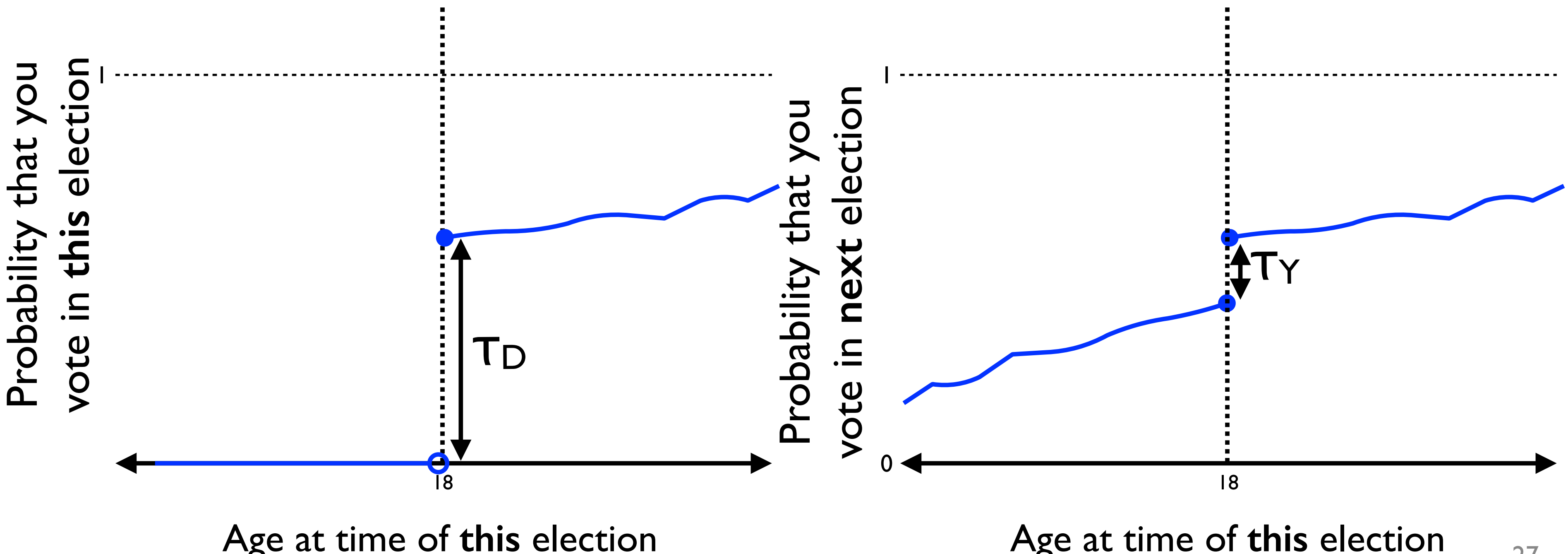
# Fuzzy RDD, briefly (2)

Fuzzy RDD cases can be thought of as having **non-compliance**: the threshold assigns units to treatment, but some units disobey.

Does this remind you of something?

Think of fuzzy RDD as IV and use the Wald estimator:

$\tau_{Fuzzy} = \tau_Y/\tau_D$



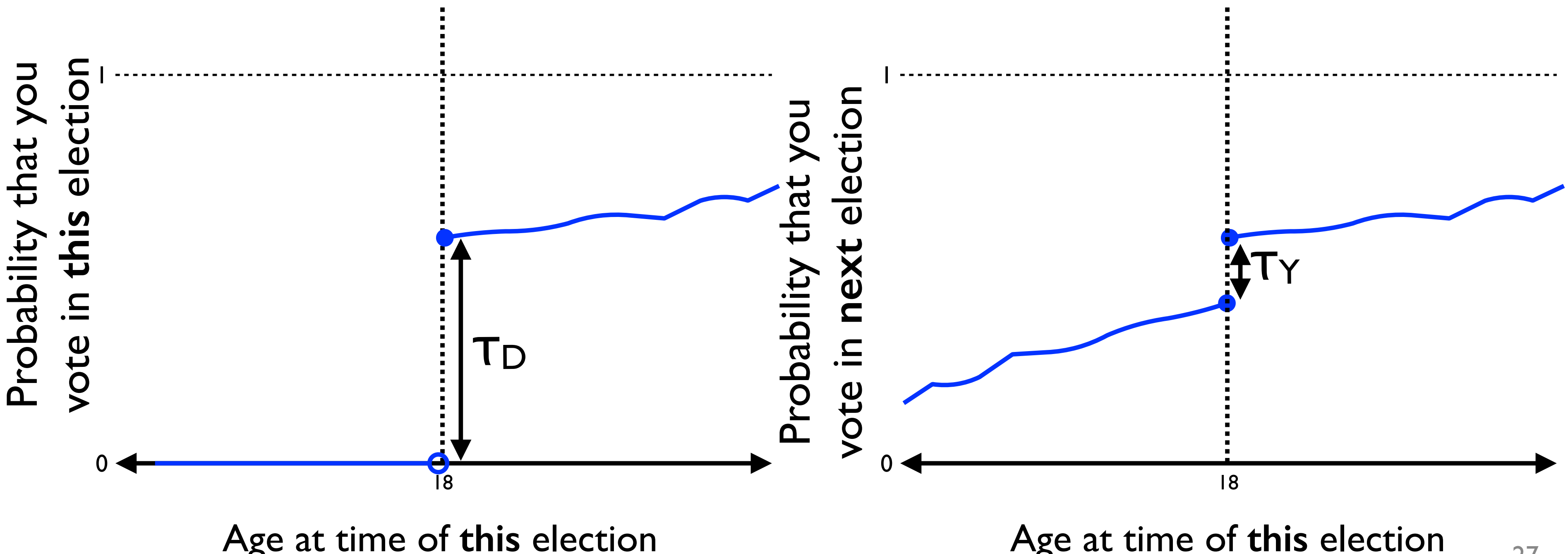
# Fuzzy RDD, briefly (2)

Fuzzy RDD cases can be thought of as having **non-compliance**: the threshold assigns units to treatment, but some units disobey.

Does this remind you of something?

Think of fuzzy RDD as IV and use the Wald estimator:

$\tau_{Fuzzy} = \tau_Y / \tau_D$



# Some bad ideas for an RDD

# Some bad ideas for an RDD

- Some political scientists define democracy as being 6 or higher in Polity score. I will study the effect of democracy on redistribution with an RDD using a 6 on Polity as the cutoff.

# Some bad ideas for an RDD

- Some political scientists define democracy as being 6 or higher in Polity score. I will study the effect of democracy on redistribution with an RDD using a 6 on Polity as the cutoff.
- Living through the stock market crash of 1929 had a major effect on people's risk tolerance and savings behavior. I will measure this effect using an RDD with "age in 1929" as the running variable and I will use a cutoff of 21 years old, as this is when people "come of age".

# Some bad ideas for an RDD

- Some political scientists define democracy as being 6 or higher in Polity score. I will study the effect of democracy on redistribution with an RDD using a 6 on Polity as the cutoff.
- Living through the stock market crash of 1929 had a major effect on people's risk tolerance and savings behavior. I will measure this effect using an RDD with "age in 1929" as the running variable and I will use a cutoff of 21 years old, as this is when people "come of age".
- Other suggestions?

# Summing up and generalizing

# Summing up and generalizing

To address selection bias, we must understand the process of selection (i.e. **assignment mechanism**).

# Summing up and generalizing

To address selection bias, we must understand the process of selection (i.e. **assignment mechanism**).

RDD is an extreme case of **selection on observables**.

# Summing up and generalizing

To address selection bias, we must understand the process of selection (i.e. **assignment mechanism**).

RDD is an extreme case of **selection on observables**.

Causal inference is most favorable when there is an assignment mechanism that

# Summing up and generalizing

To address selection bias, we must understand the process of selection (i.e. **assignment mechanism**).

RDD is an extreme case of **selection on observables**.

Causal inference is most favorable when there is an assignment mechanism that

- is transparent/well understood (at least in part)

# Summing up and generalizing

To address selection bias, we must understand the process of selection (i.e. **assignment mechanism**).

RDD is an extreme case of **selection on observables**.

Causal inference is most favorable when there is an assignment mechanism that

- is transparent/well understood (at least in part)
- treats similar units differently

# Summing up and generalizing

To address selection bias, we must understand the process of selection (i.e. **assignment mechanism**).

RDD is an extreme case of **selection on observables**.

Causal inference is most favorable when there is an assignment mechanism that

- is transparent/well understood (at least in part)
- treats similar units differently

Think about what "treating similar units differently" means in matching, IV, diff-in-diff/panel, RDD.

## Summing up and generalizing

To address selection bias, we must understand the process of selection (i.e. **assignment mechanism**).

RDD is an extreme case of **selection on observables**.

Causal inference is most favorable when there is an assignment mechanism that

- is transparent/well understood (at least in part)
- treats similar units differently

Think about what "treating similar units differently" means in matching, IV, diff-in-diff/panel, RDD.

You may find other cases where a transparent rule treats similar units differently — keep an eye out!