Regression discontinuity designs

Intermediate Social Statistics

Week 6 (21 February 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:

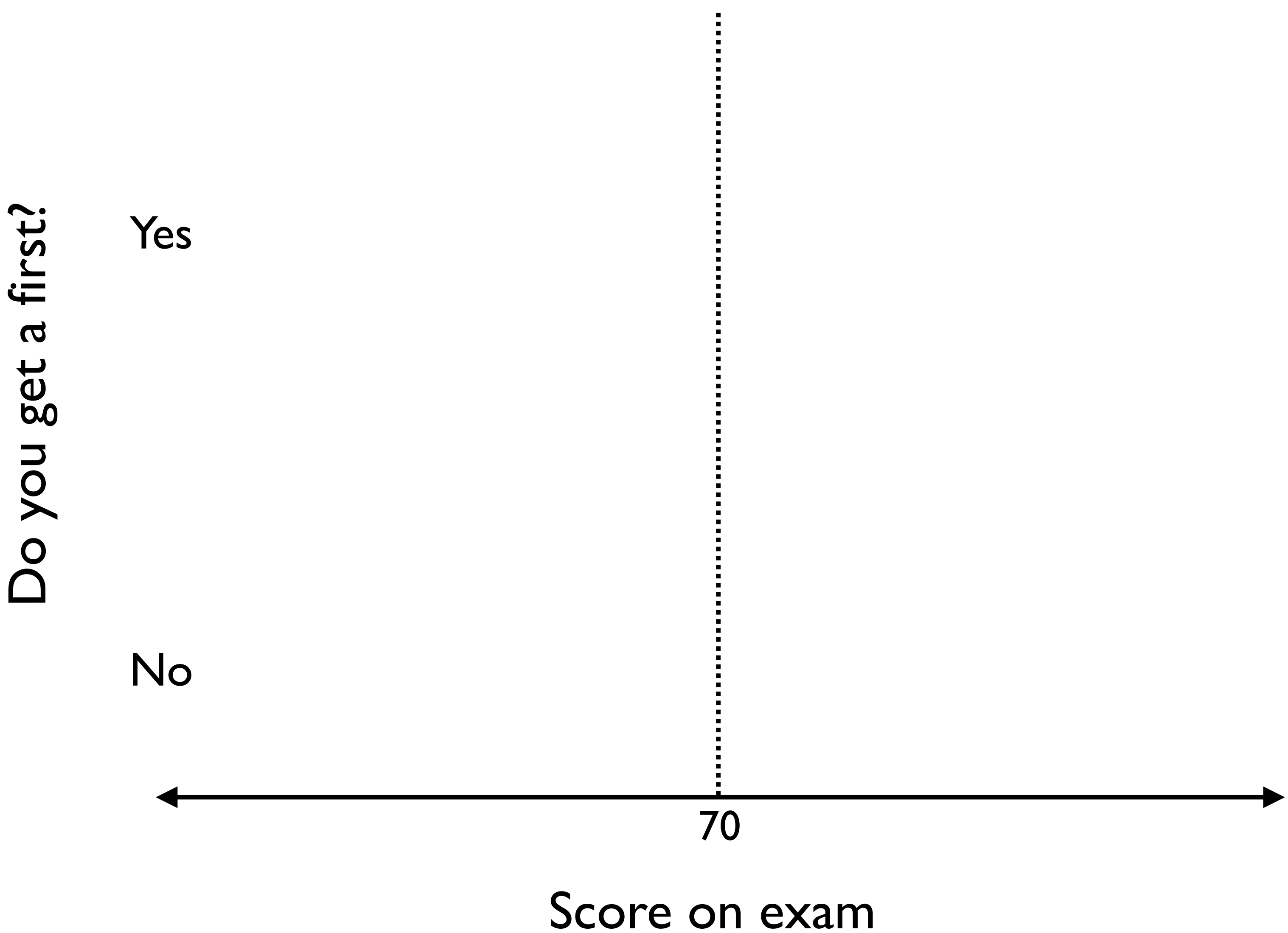
- Students receive a “first” if they score 70 or higher
- French municipalities use PR elections if population is 3,500 (now 1,000) or higher
- A candidate is elected if she receives more votes than any other candidate

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

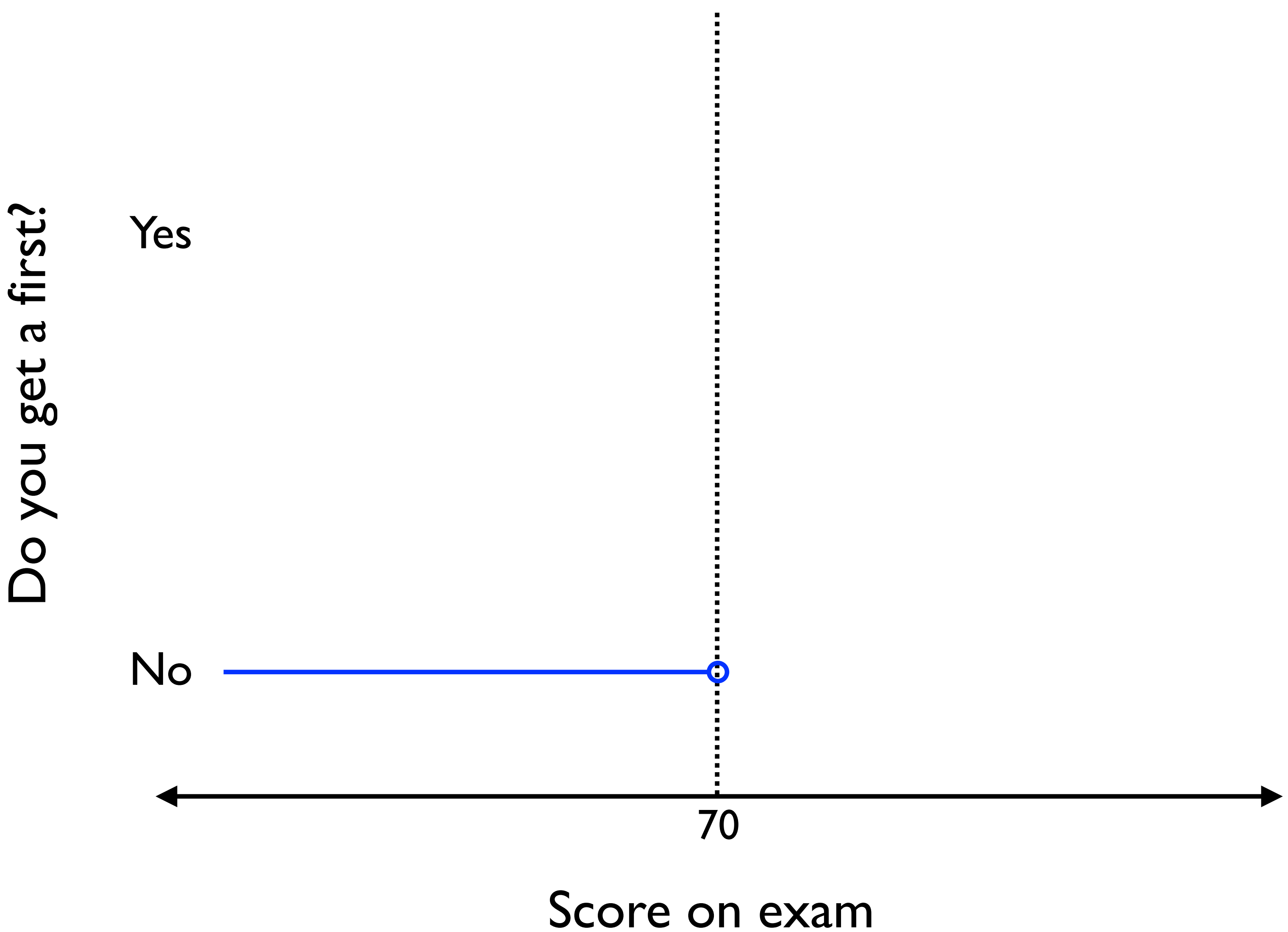
But policies **do** make jumps:

- 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

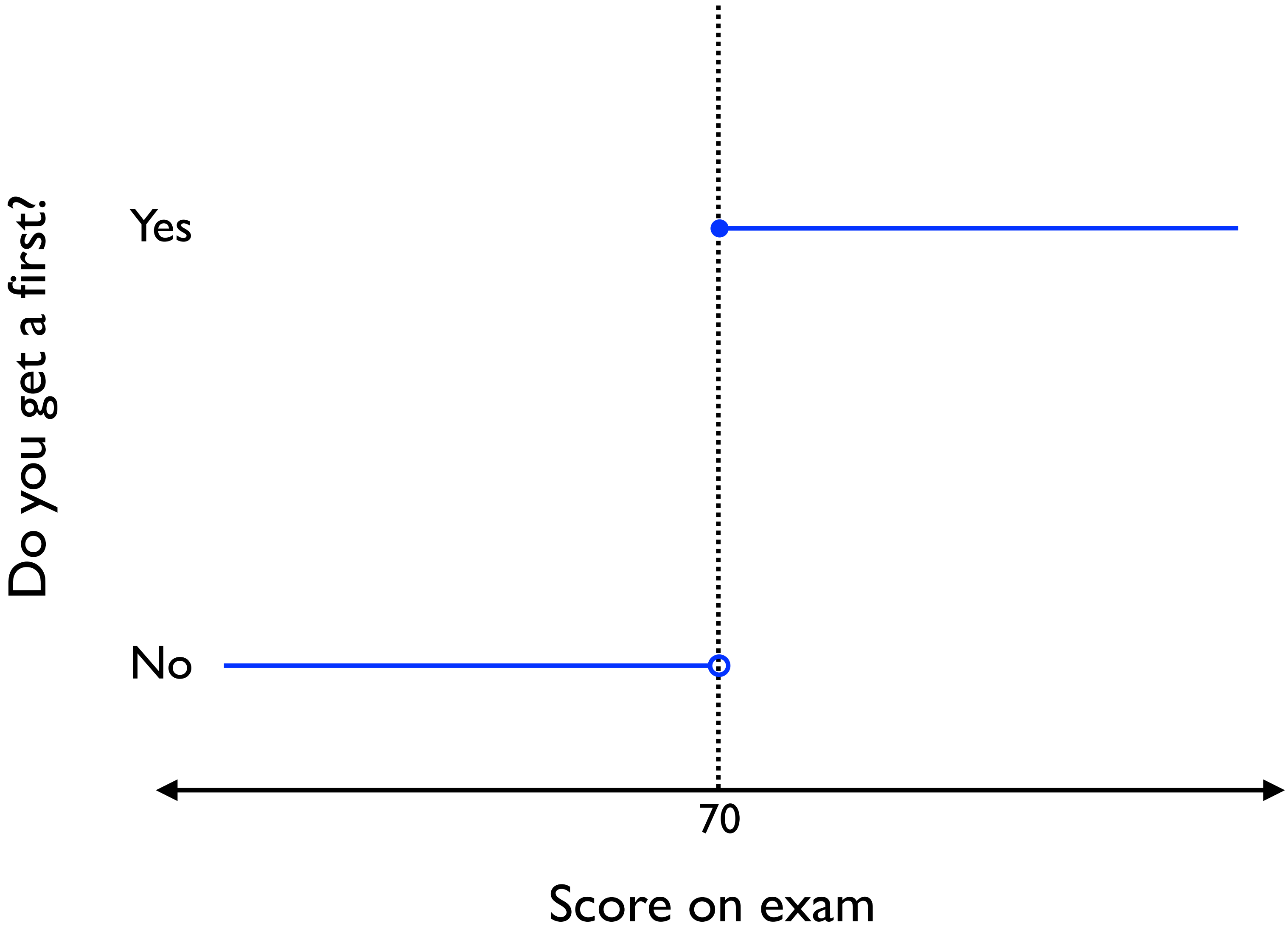
“Policies” make jumps



“Policies” make jumps



“Policies” make jumps



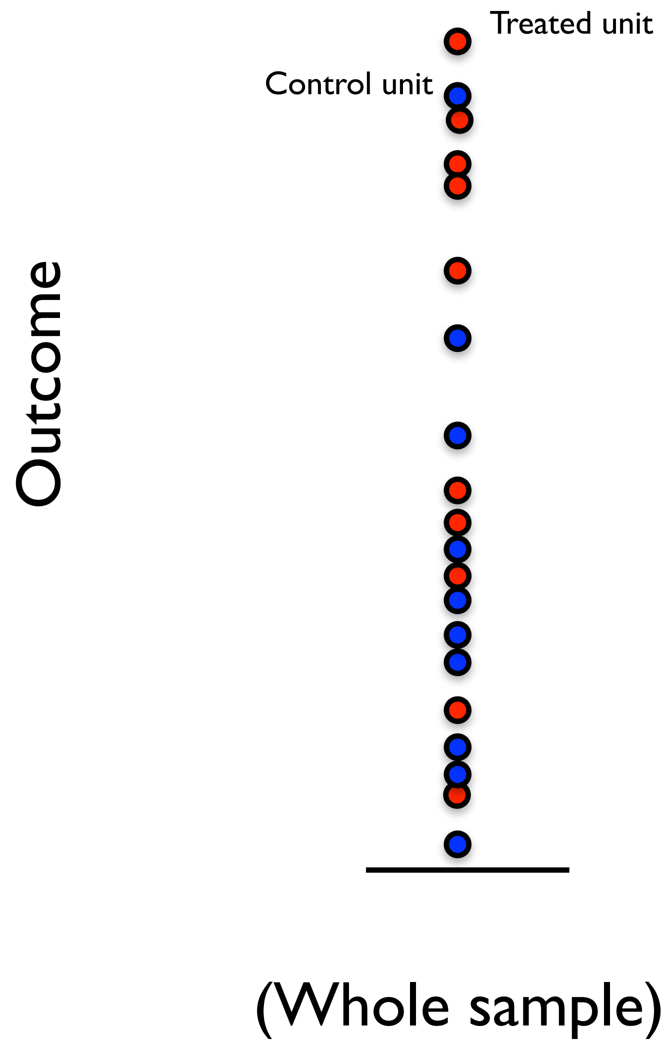
Plan

- Quick intuition (already done)
- Understanding the continuity assumption
- Examples
- Some more general evidence on RD validity
- Some guidance on estimation

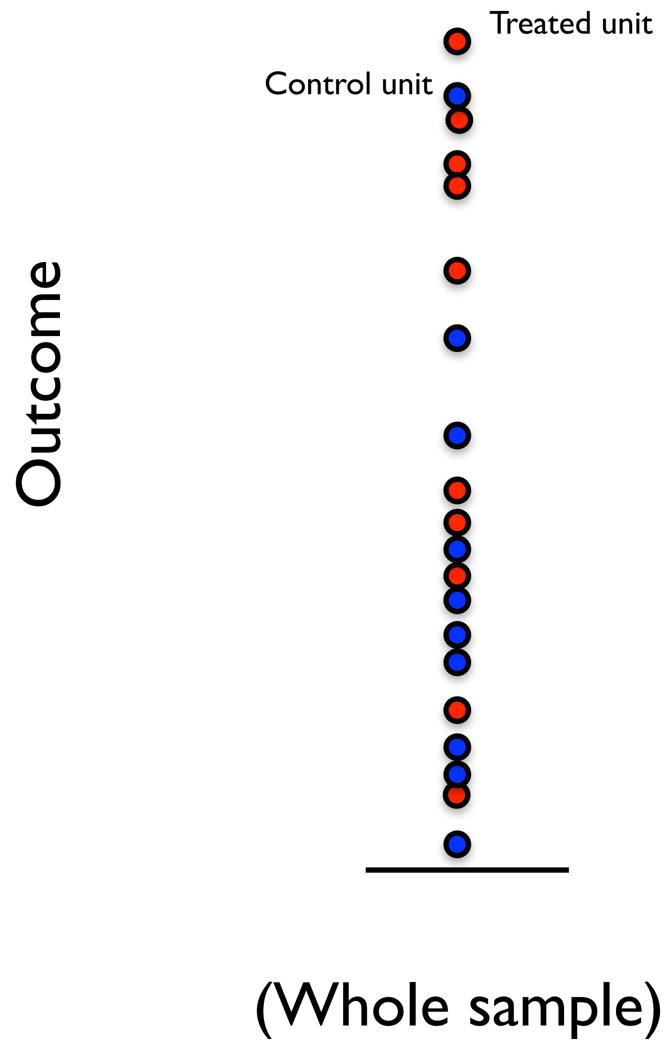
Review: causal inference as missing data problem

Unit	Y_i	D_i	Y_{1i}	Y_{0i}
1	3	1	3	?
2	1	1	1	?
3	0	0	?	0
4	1	0	?	1
…	…	…	…	…

Estimating treatment effects in the whole sample



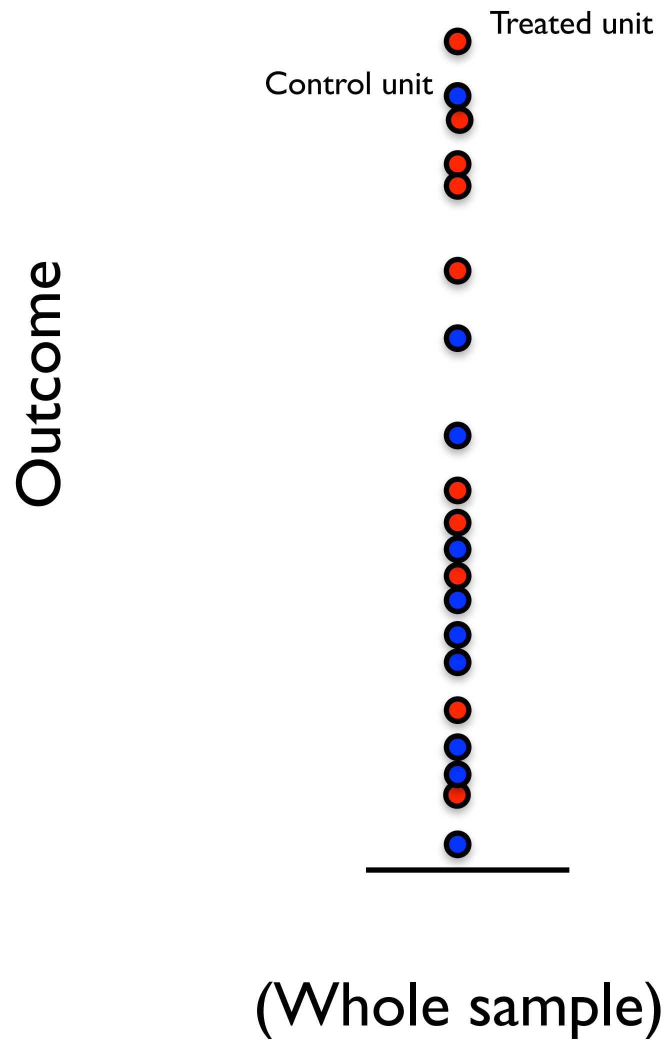
Estimating treatment effects in the whole sample



Recall ATT is

$E[Y_1|D_i=1] - E[Y_0|D_i=1]$.

Estimating treatment effects in the whole sample



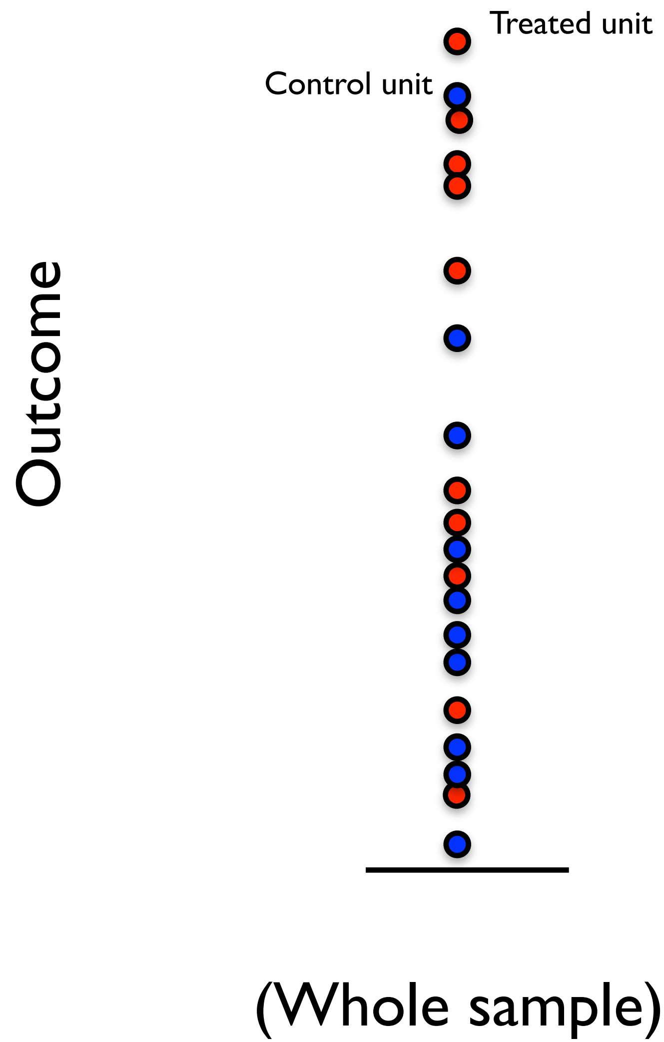
Recall ATT is

$E[Y_1|D_i=1] - E[Y_0|D_i=1]$.

Might estimate ATE with

$E[Y_1|D_i=1] - E[Y_0|D_i=0]$.

Estimating treatment effects in the whole sample



Recall ATT is

$E[Y_1|D_i=1]$ - $E[Y_0|D_i=1]$.

Might estimate ATE with

$E[Y_1|D_i=1]$ - $E[Y_0|D_i=0]$.

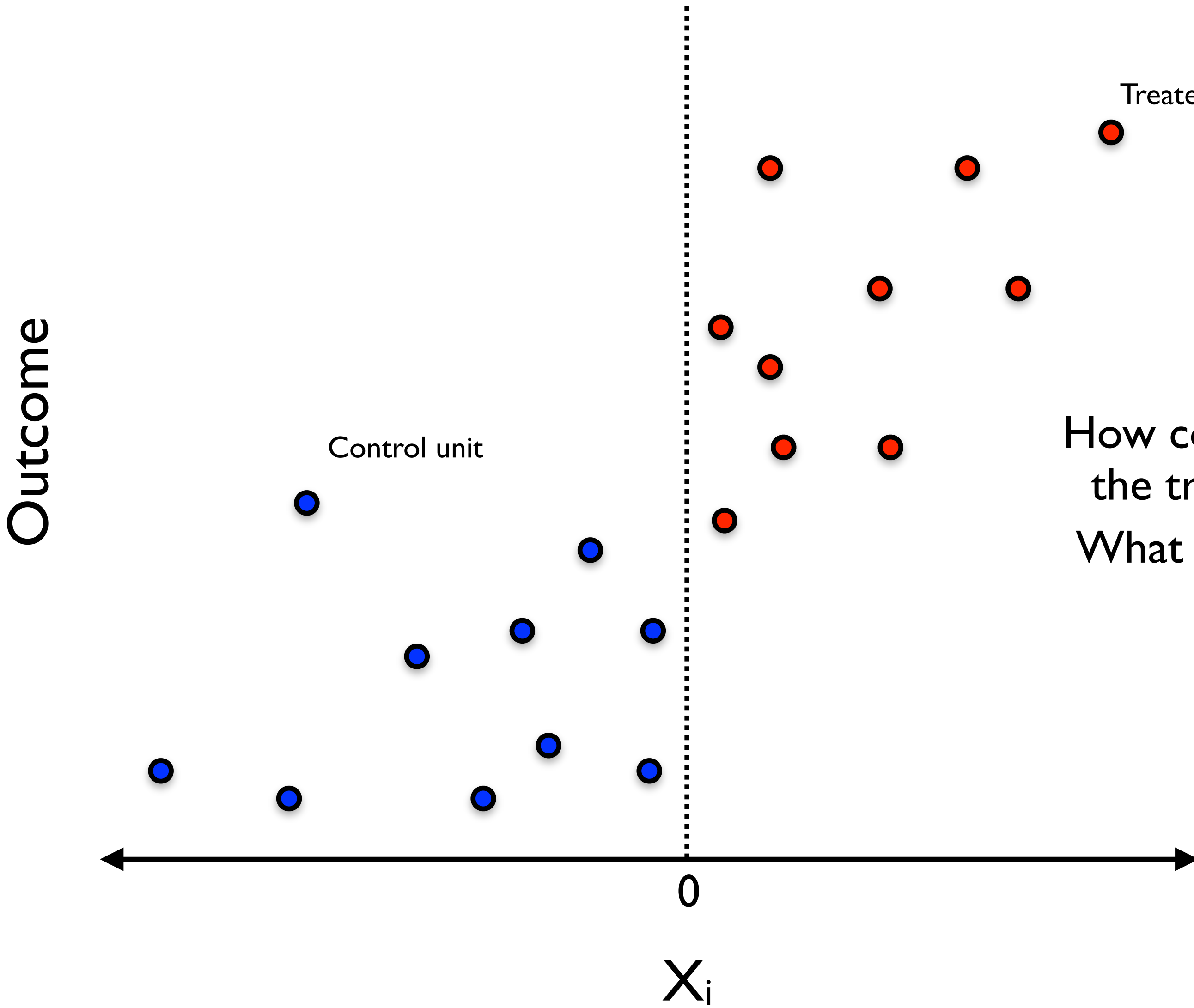
Under what conditions does this give an unbiased estimate?

What if treatment assignment depends on a cutoff value of X_i?

Assignment mechanism: $D_i = 1$ iff $X_i \geq 0$

Unit	Y_i	D_i	Y_{1i}	Y_{0i}	X_i
1	3	1	3	?	0.3
2	1	1	1	?	0.7
3	0	0	?	0	-1.2
4	1	0	?	1	-0.4
…	…	…	…	…	

Estimating treatment effects when treatment depends on a discontinuity



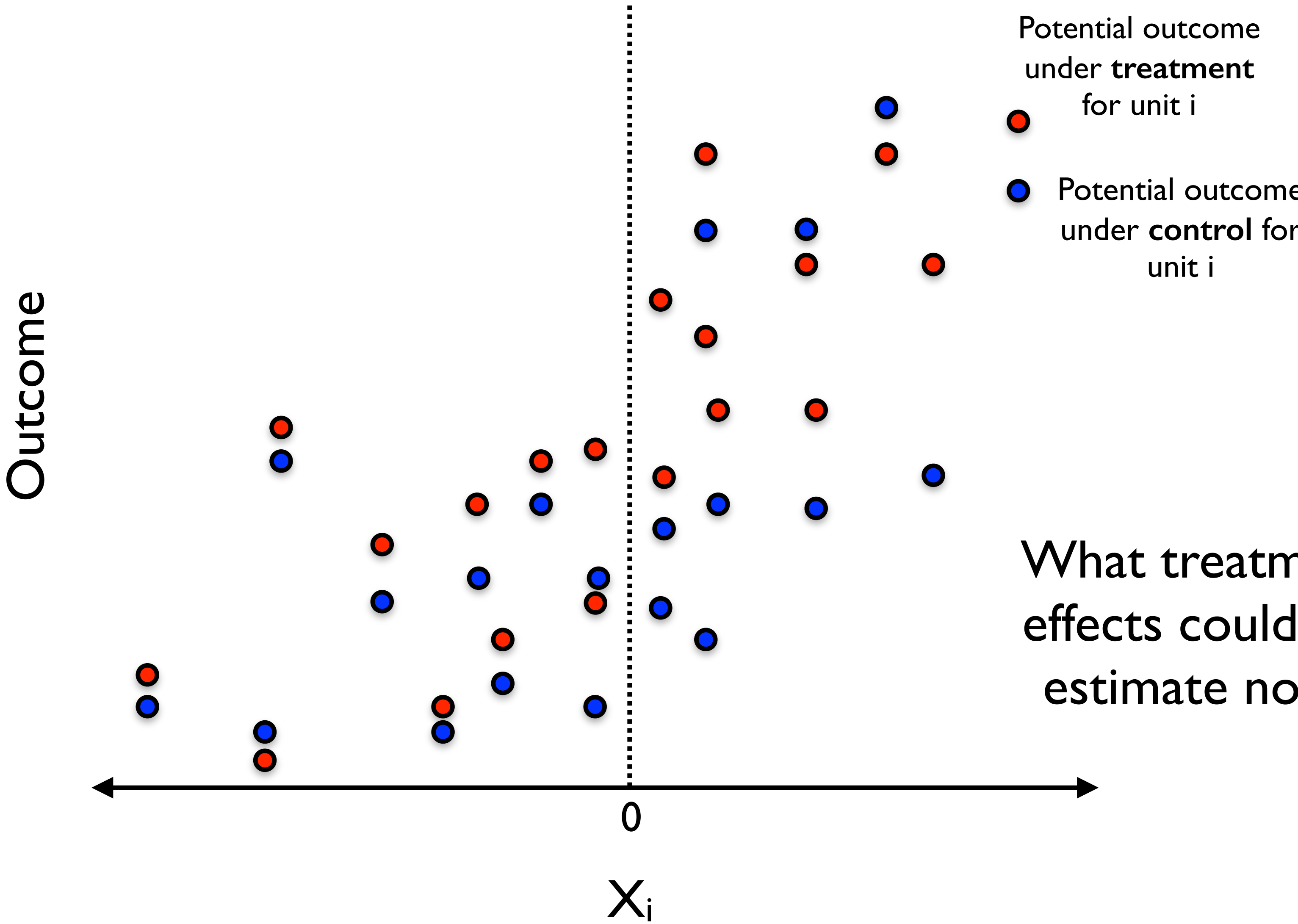
How could we estimate the treatment effect?

What assumptions are we using?

Suppose the FPOCI is overcome (rejoice!)

Unit	Y_i	D_i	Y_{1i}	Y_{0i}	X_i
1	3	1	3	1.4	0.3
2	1	1	1	0.8	0.7
3	0	0	1.3	0	-1.2
4	1	0	1.1	1	-0.4
...	...	...	...	...	

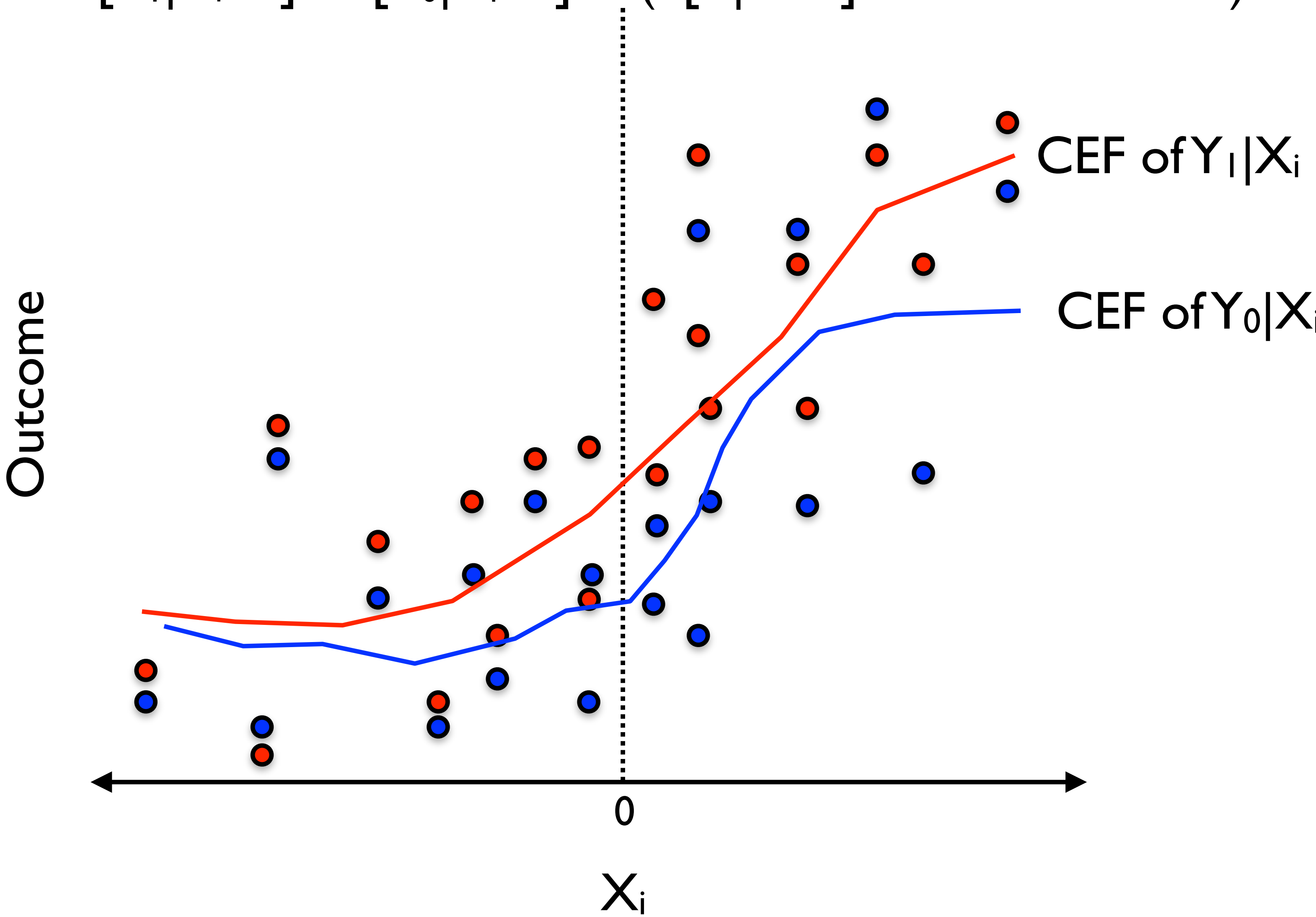
Estimating treatment effects when both potential outcomes are observed



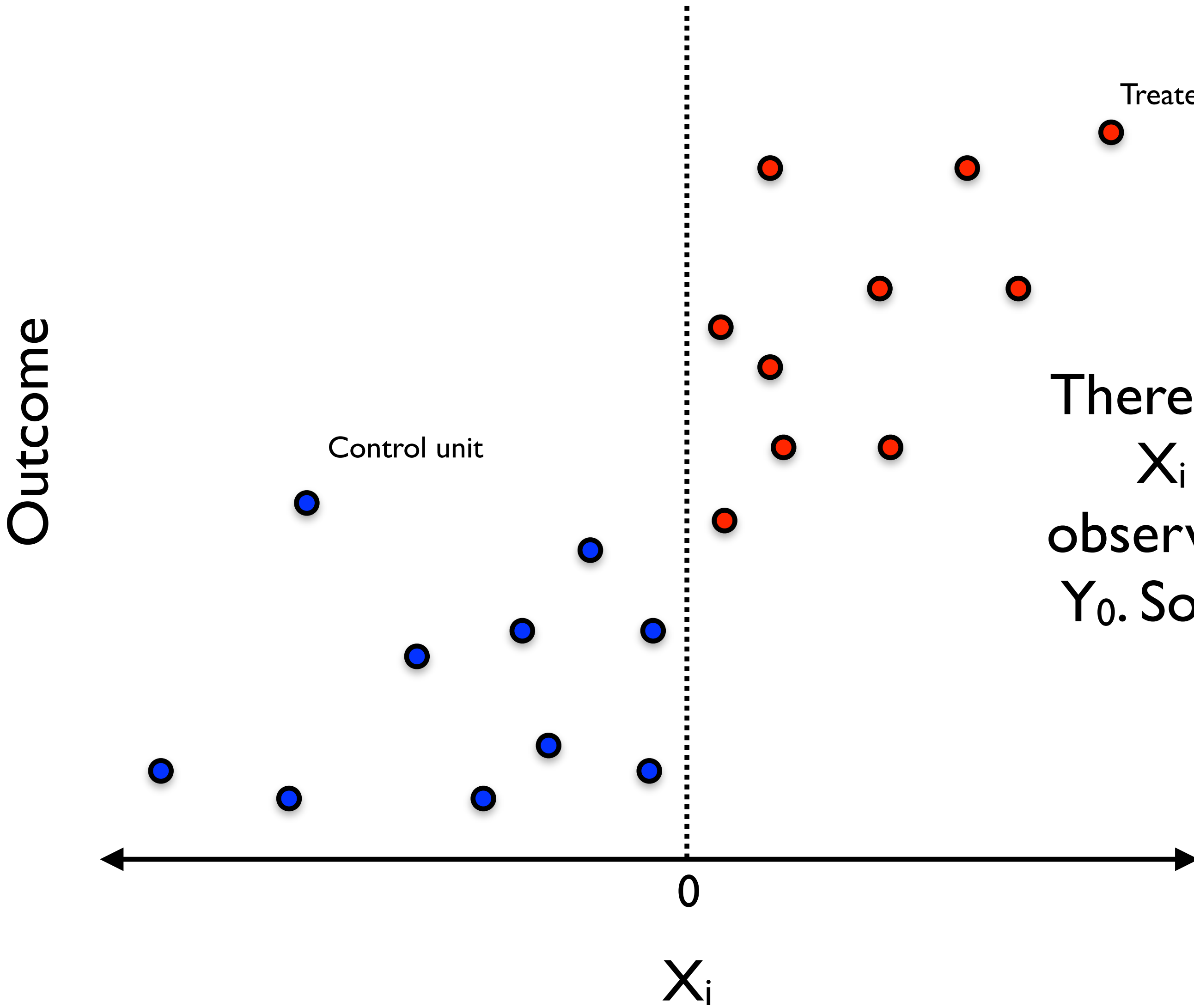
What treatment effects could we estimate now?

Here's one: Local average treatment effect (LATE)

LATE: $E[Y_1|X_i=x] - E[Y_0|X_i=x]$. ($E[Y|X=x]$ is the "CEF at x")

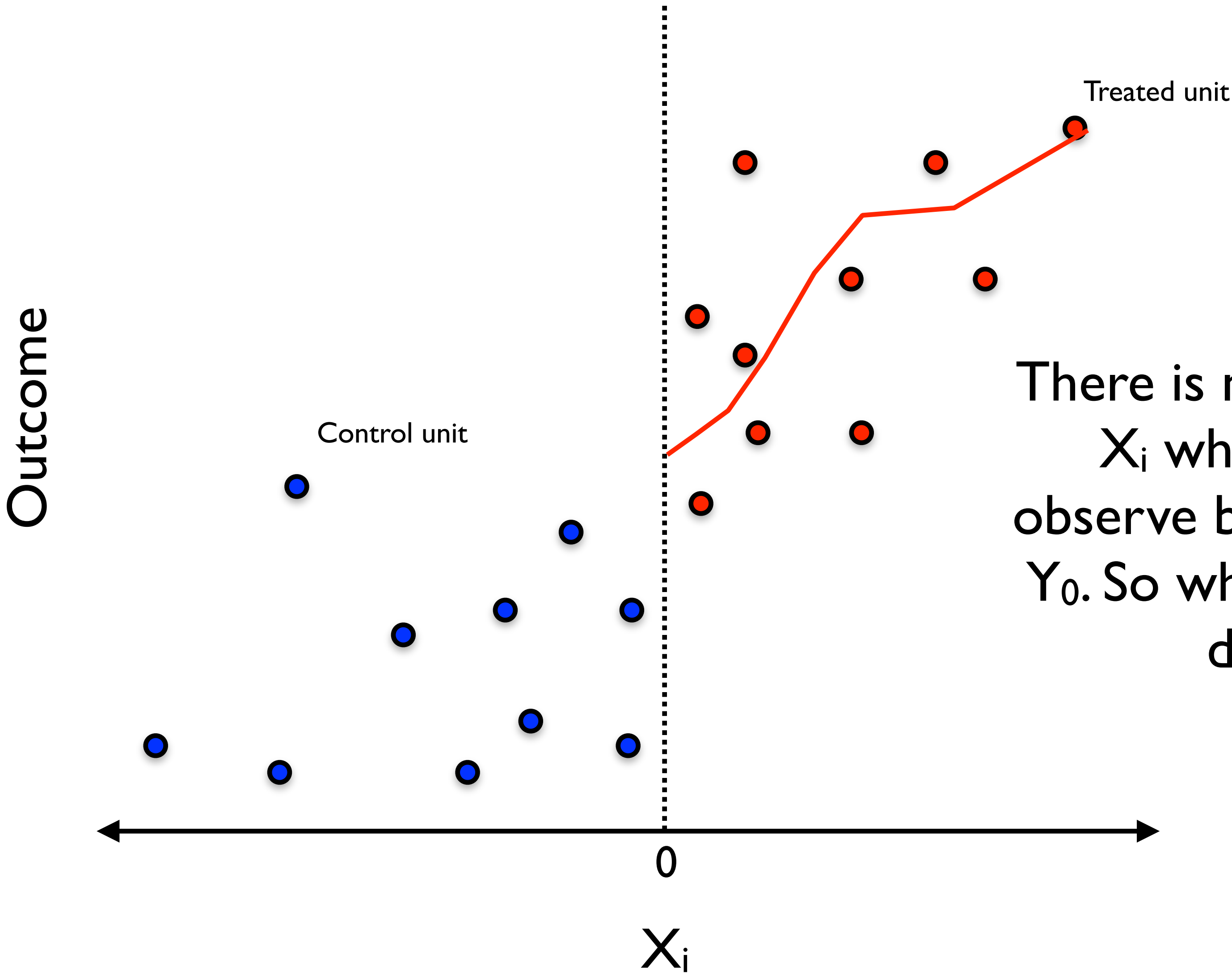


Back to reality: FPOCI



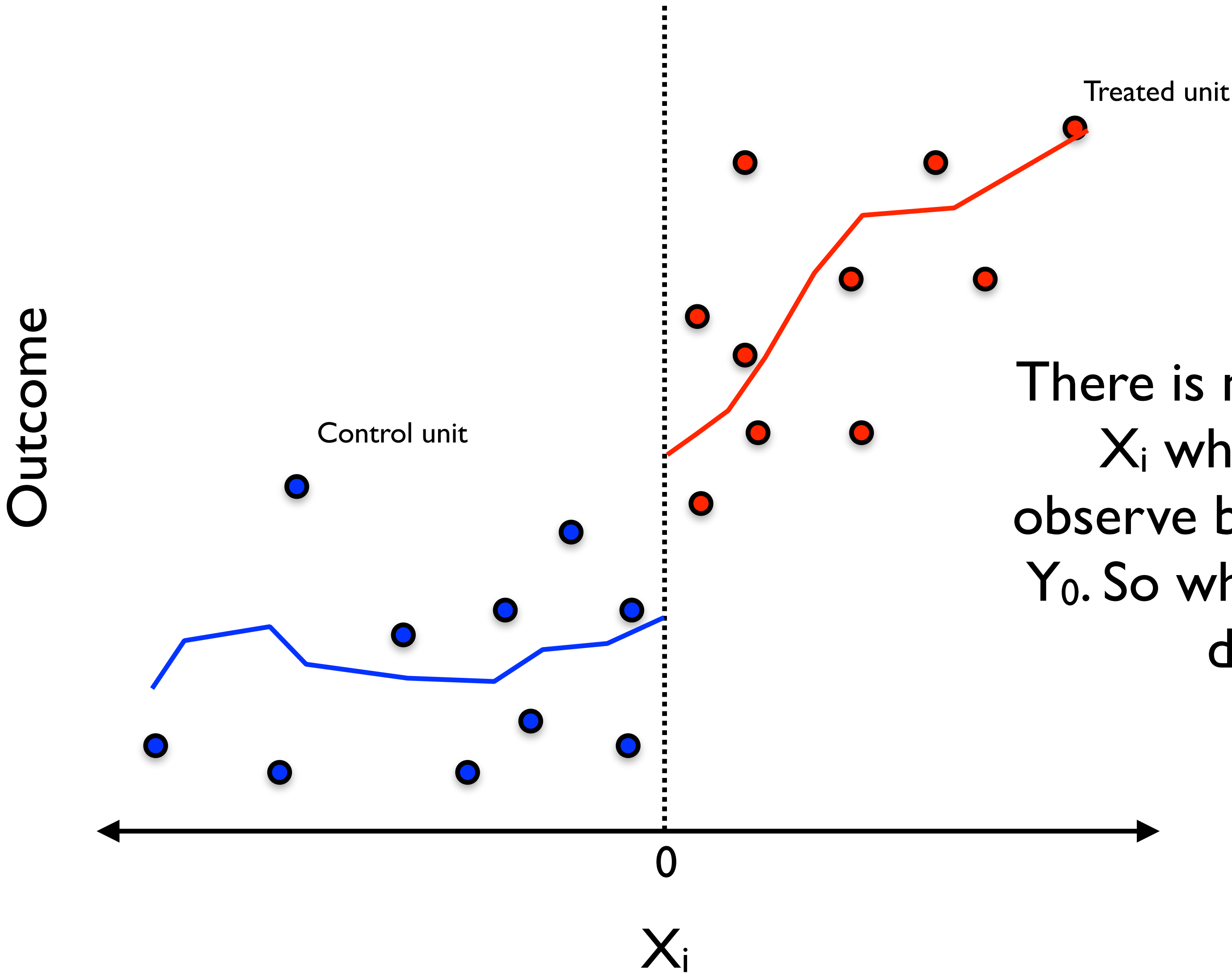
There is no value of X_i where we observe both Y_1 and Y_0. So what can we do?

Back to reality: FPOCI



There is no value of X_i where we observe both Y_1 and Y_0. So what can we do?

Back to reality: FPOCI



There is no value of X_i where we observe both Y_1 and Y_0. So what can we do?

What assumptions do we need?

The LATE at $X_i = 0$ is defined as: $E[Y_1|X_i = 0] - E[Y_0|X_i = 0]$

What assumptions do we need?

The LATE at $X_i = 0$ is defined as: $E[Y_1|X_i = 0] - E[Y_0|X_i = 0]$

If X_i is a continuous variable, no observations where $X_i = 0$.

What assumptions do we need?

The LATE at $X_i = 0$ is defined as: $E[Y_1|X_i = 0] - E[Y_0|X_i = 0]$

If X_i is a continuous variable, no observations where $X_i = 0$.

But consider these assumptions/substitutions:

$$E[Y_1|X_i = 0] = \lim_{x \to 0^+} E[Y_1|X_i = x]$$

$$E[Y_0|X_i = 0] = \lim_{x \to 0^-} E[Y_0|X_i = x]$$

What assumptions do we need?

The LATE at X_i = 0 is defined as: $E[Y_1|X_i = 0] - E[Y_0|X_i = 0]$

If X_i is a continuous variable, no observations where X_i = 0.

But consider these assumptions/substitutions:

$$E[Y_1|X_i = 0] = \lim_{x \to 0^+} E[Y_1|X_i = x]$$

"as x goes to 0 from above"

$$E[Y_0|X_i = 0] = \lim_{x \to 0^-} E[Y_0|X_i = x]$$

"as x goes to 0 from below"

What assumptions do we need?

The LATE at X_i = 0 is defined as: $E[Y_1|X_i = 0] - E[Y_0|X_i = 0]$

If X_i is a continuous variable, no observations where X_i = 0.

But consider these assumptions/substitutions:

$$E[Y_1|X_i = 0] = \lim_{x \to 0^+} E[Y_1|X_i = x]$$

"as x goes to 0 from above"

$$E[Y_0|X_i = 0] = \lim_{x \to 0^-} E[Y_0|X_i = x]$$

"as x goes to 0 from below"

In words: the CEF for treated and untreated at X_i=0 can be estimated by extrapolating each CEF to the boundary.

What assumptions do we need?

The LATE at $X_i = 0$ is defined as: $E[Y_1|X_i = 0] - E[Y_0|X_i = 0]$

If X_i is a continuous variable, no observations where $X_i = 0$.

But consider these assumptions/substitutions:

$$E[Y_1|X_i = 0] = \lim_{x \to 0^+} E[Y_1|X_i = x]$$

"as x goes to 0 from above"

$$E[Y_0|X_i = 0] = \lim_{x \to 0^-} E[Y_0|X_i = x]$$

"as x goes to 0 from below"

In words: the CEF for treated and untreated at X_i=0 can be estimated by extrapolating each CEF to the boundary.

Obviously we can estimate the LATE at $X_i = 0$ if these assumptions are valid. *(What about at other values of X_i?)*

What assumptions do we need?

The LATE at X_i = 0 is defined as: $E[Y_1|X_i = 0] - E[Y_0|X_i = 0]$

If X_i is a continuous variable, no observations where X_i = 0.

But consider these assumptions/substitutions:

$$E[Y_1|X_i = 0] = \lim_{x \to 0^+} E[Y_1|X_i = x]$$

"as x goes to 0 from above"

$$E[Y_0|X_i = 0] = \lim_{x \to 0^-} E[Y_0|X_i = x]$$

"as x goes to 0 from below"

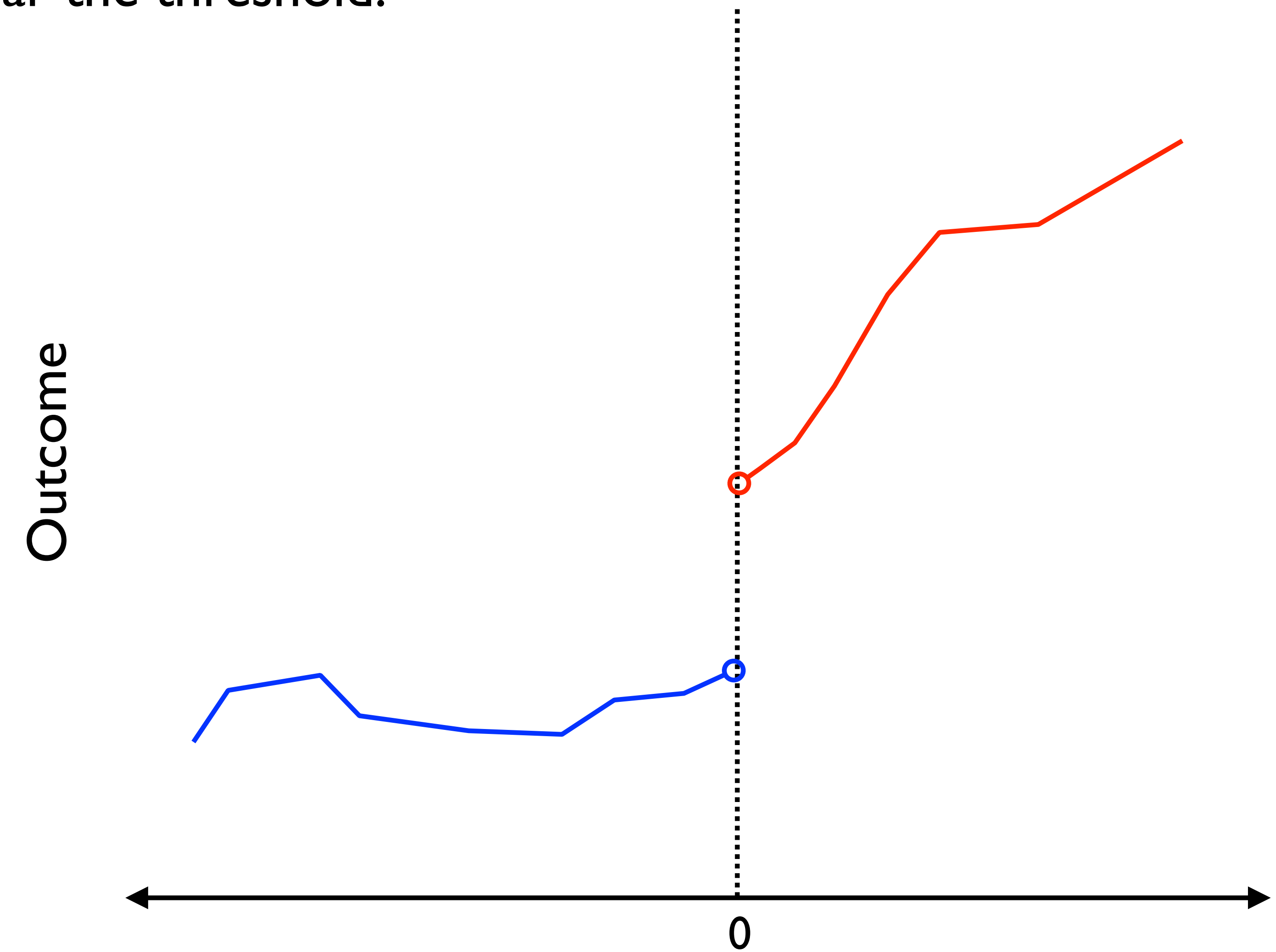
In words: the CEF for treated and untreated at X_i=0 can be estimated by extrapolating each CEF to the boundary.

Obviously we can estimate the LATE at X_i = 0 if these assumptions are valid. *(What about at other values of X_i?)*

When would these assumptions be valid?

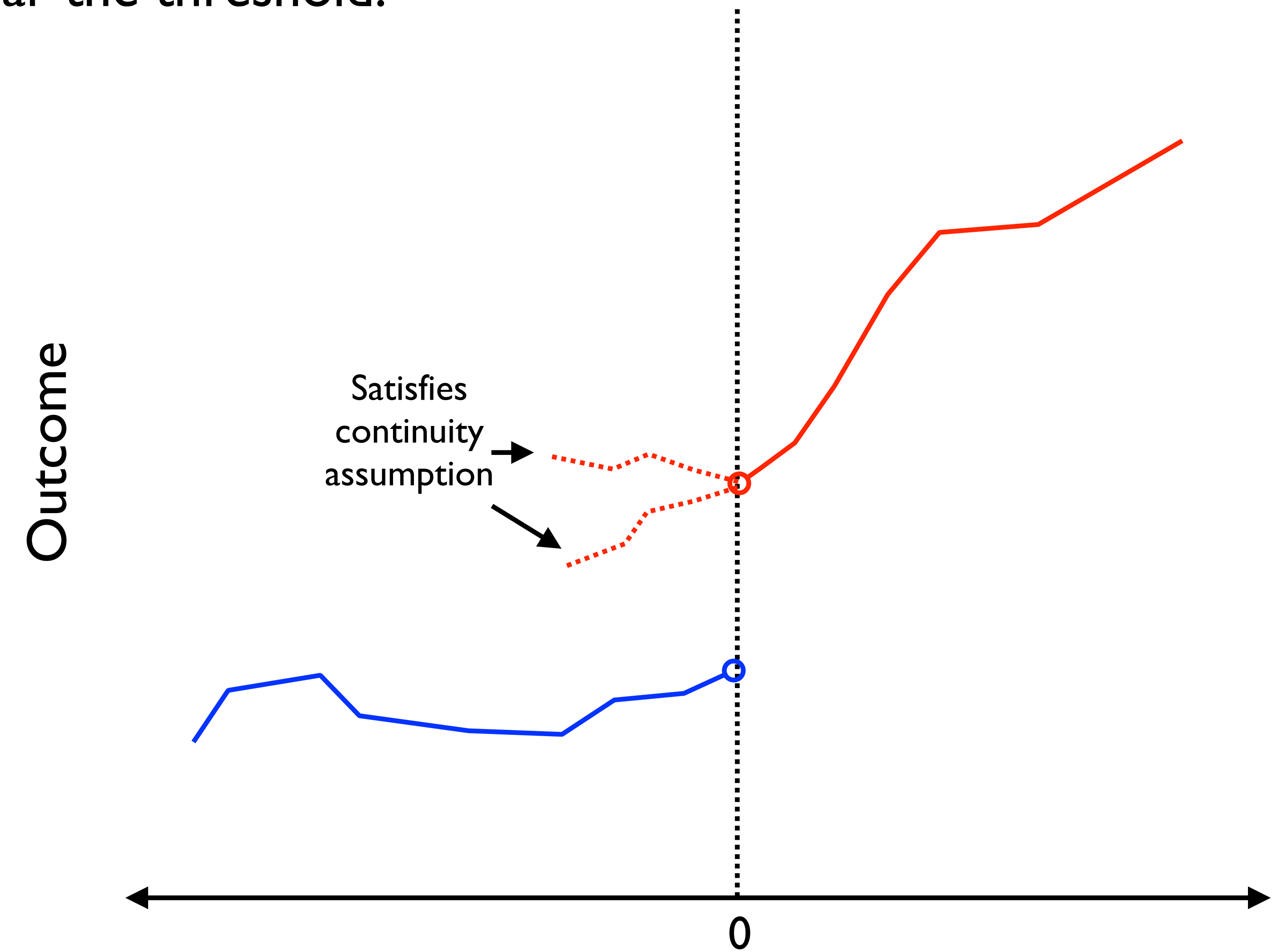
When can we extrapolate to the threshold?

Given infinite data, the only requirement is **continuity of the CEFs** near the threshold.



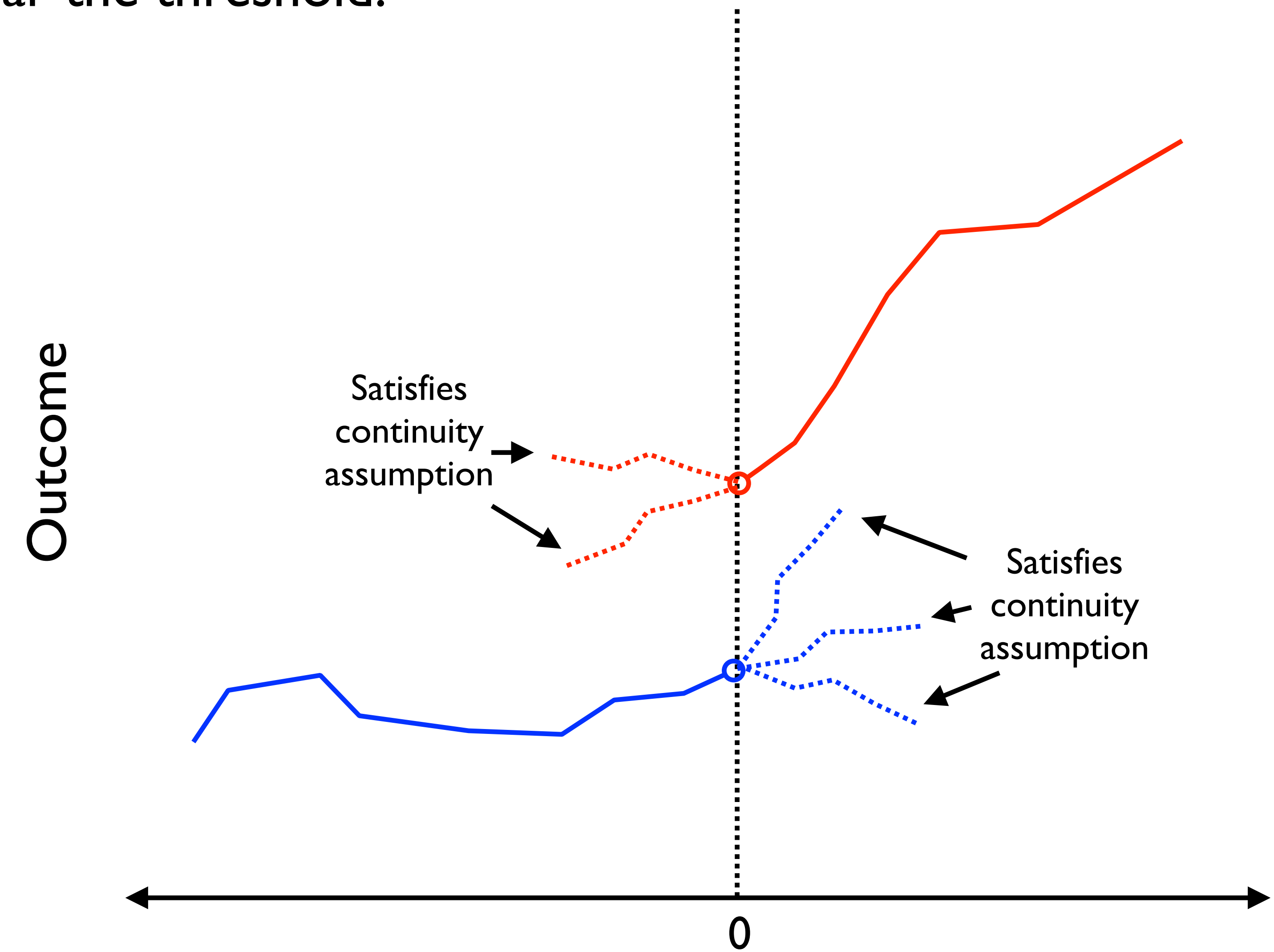
When can we extrapolate to the threshold?

Given infinite data, the only requirement is **continuity of the CEFs** near the threshold.



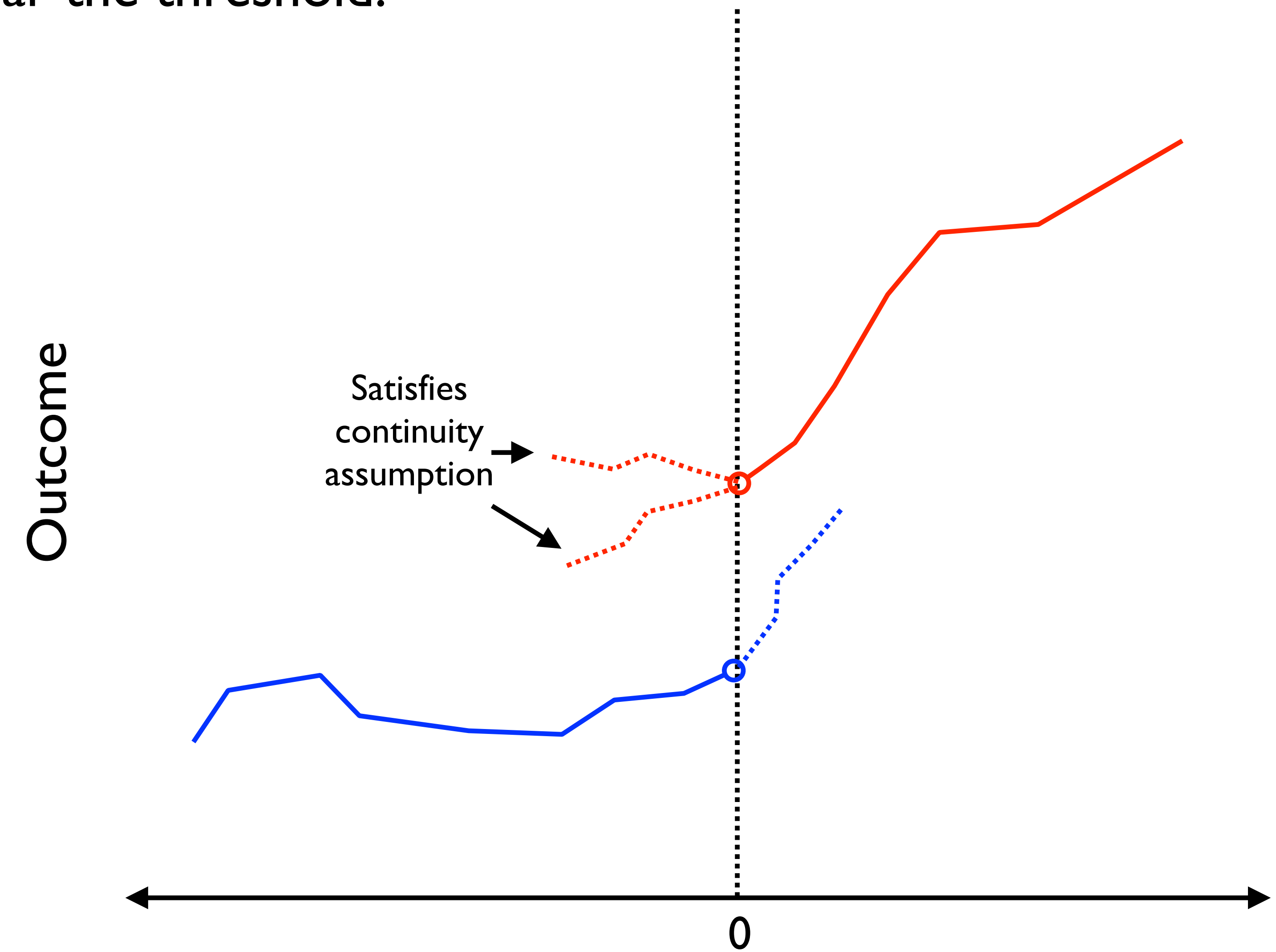
When can we extrapolate to the threshold?

Given infinite data, the only requirement is **continuity of the CEFs** near the threshold.



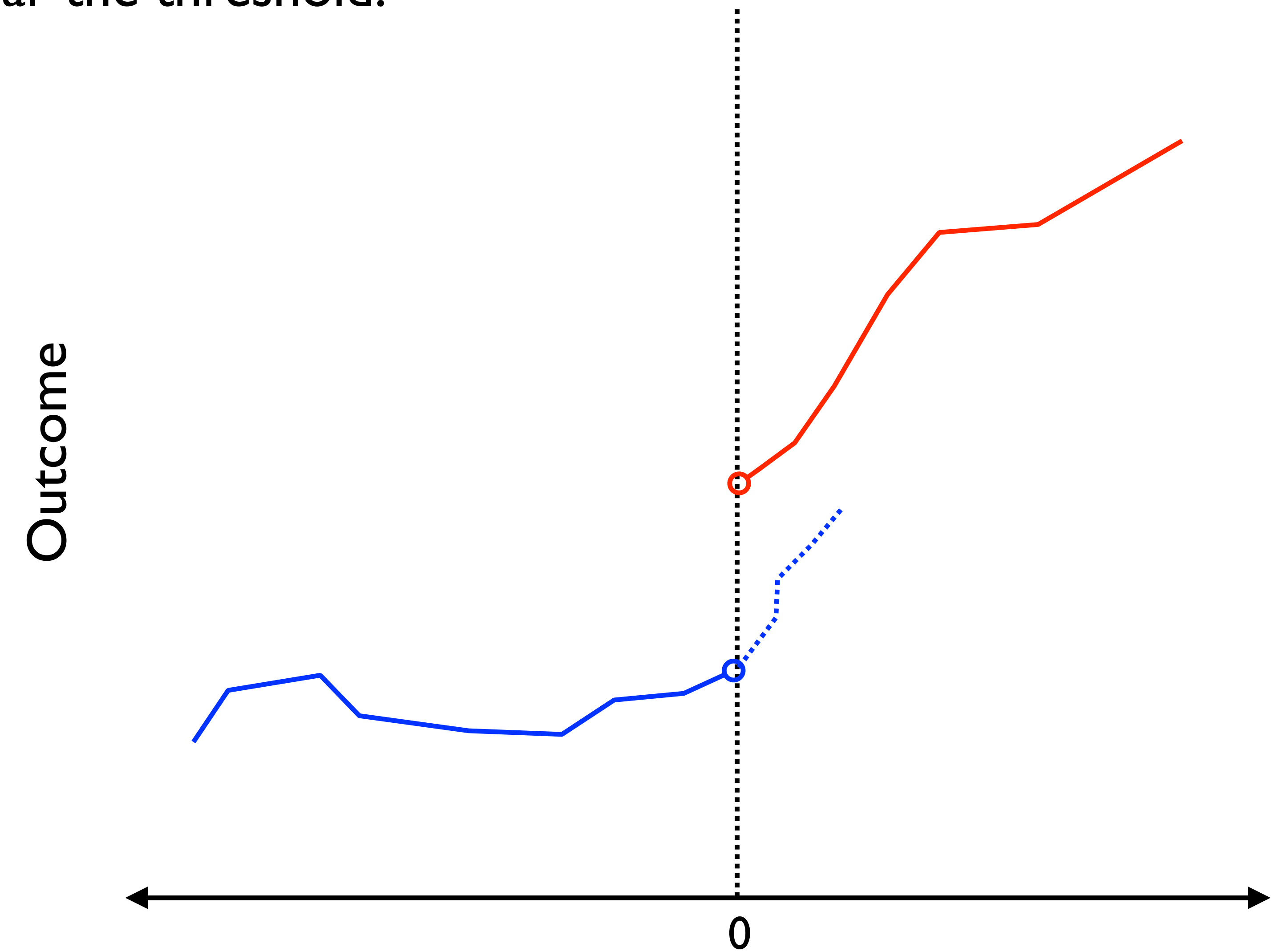
When can we extrapolate to the threshold?

Given infinite data, the only requirement is **continuity of the CEFs** near the threshold.



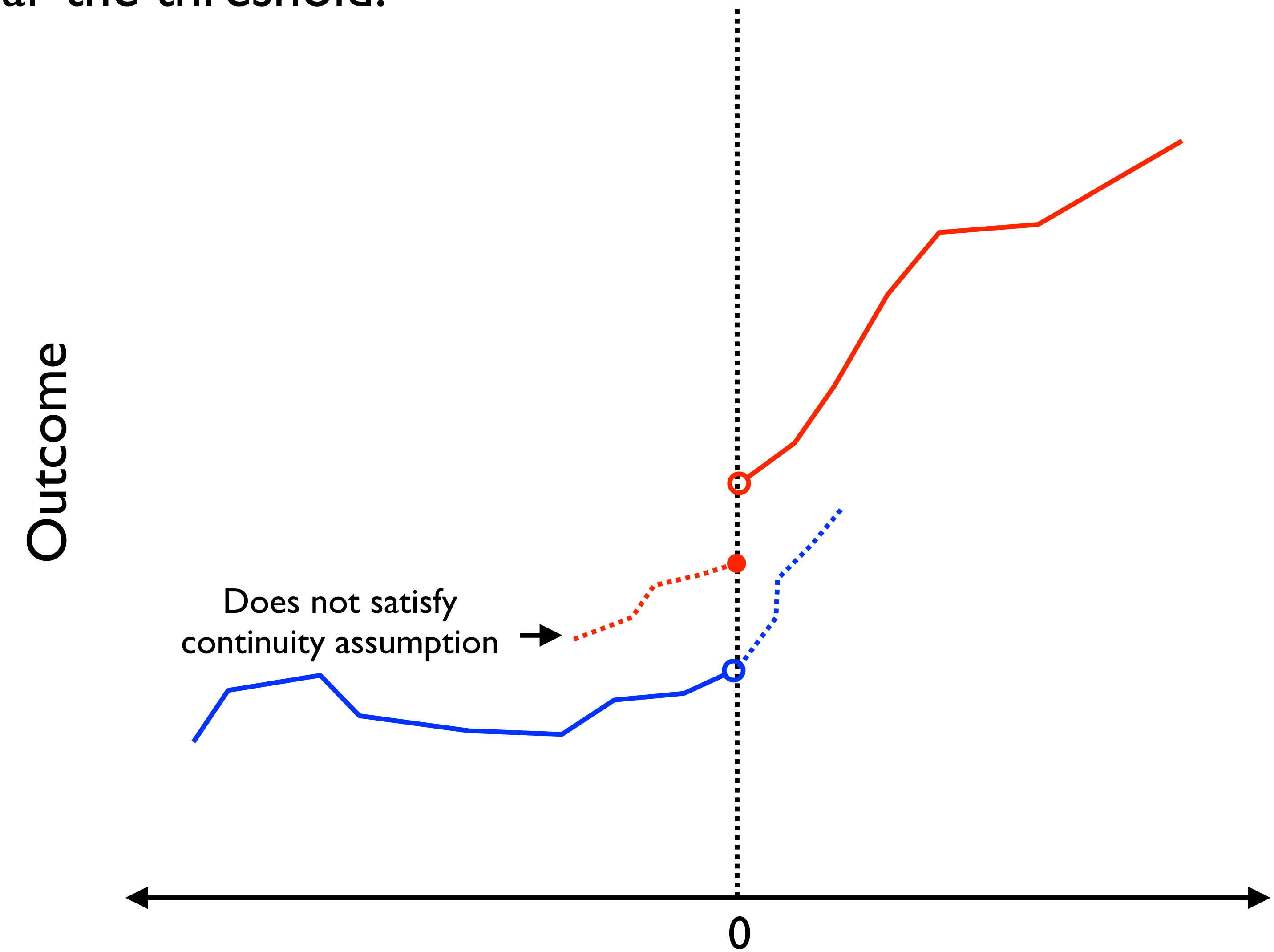
When can we extrapolate to the threshold?

Given infinite data, the only requirement is **continuity of the CEFs** near the threshold.



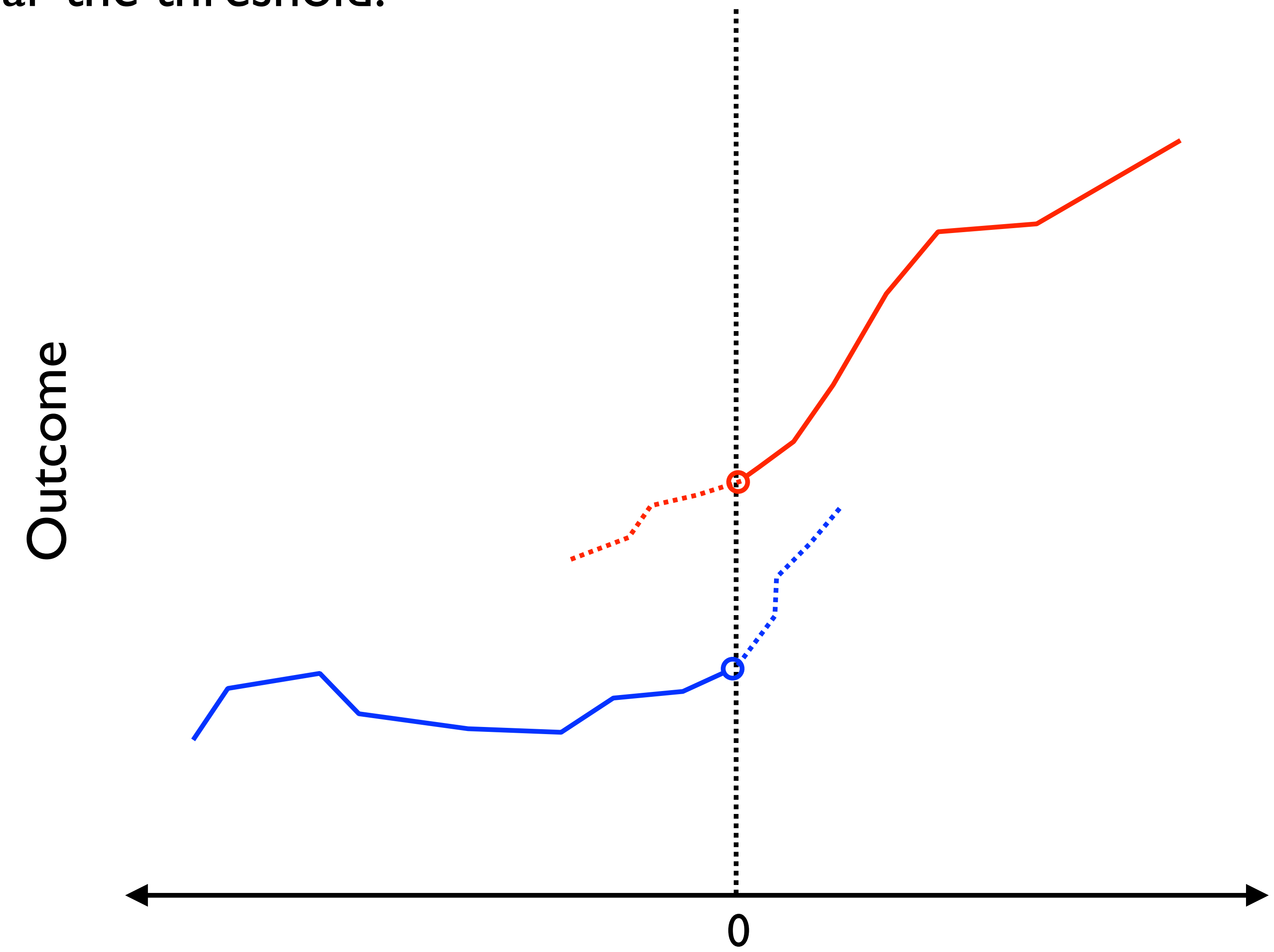
When can we extrapolate to the threshold?

Given infinite data, the only requirement is **continuity of the CEFs** near the threshold.



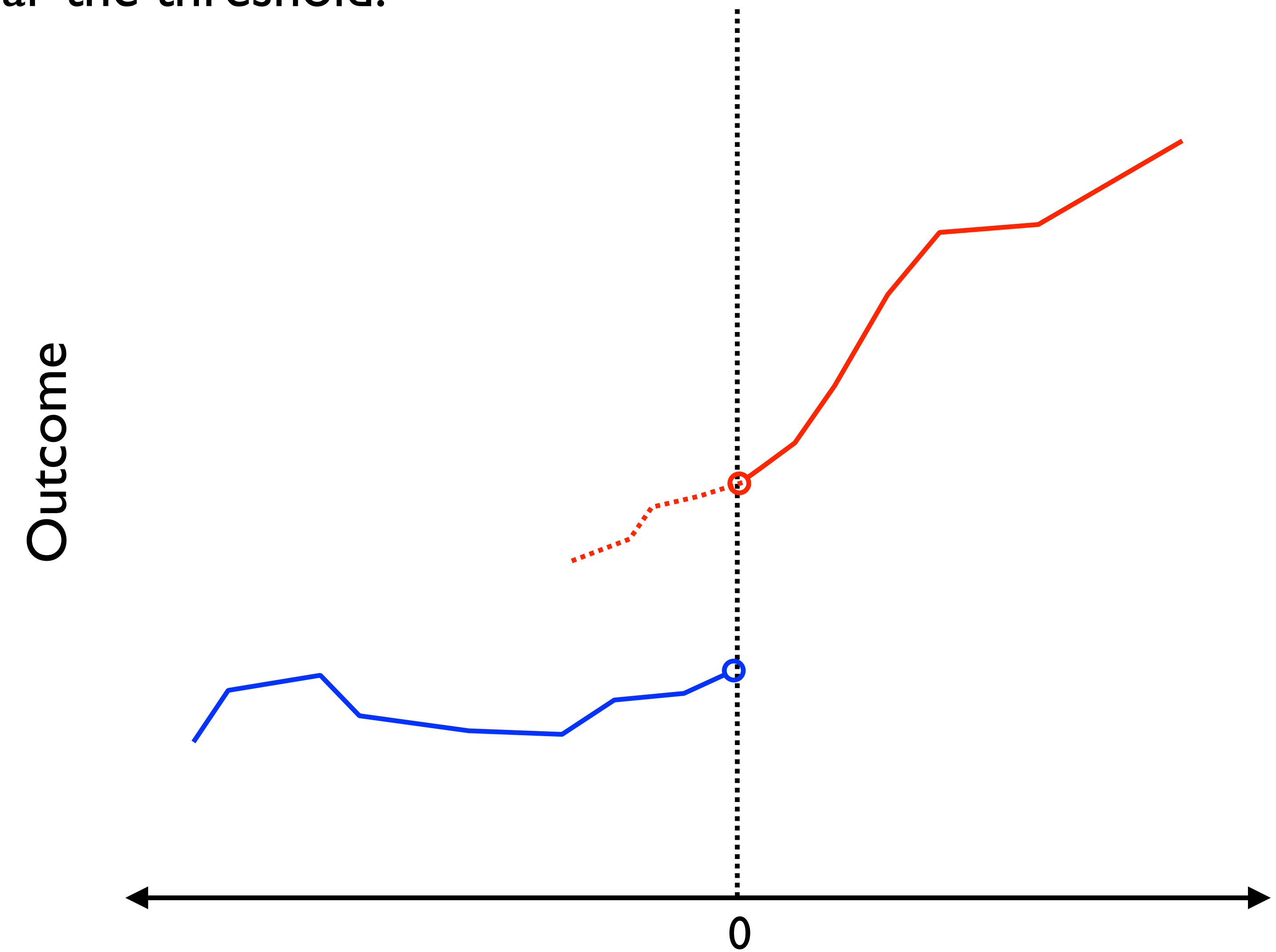
When can we extrapolate to the threshold?

Given infinite data, the only requirement is **continuity of the CEFs** near the threshold.



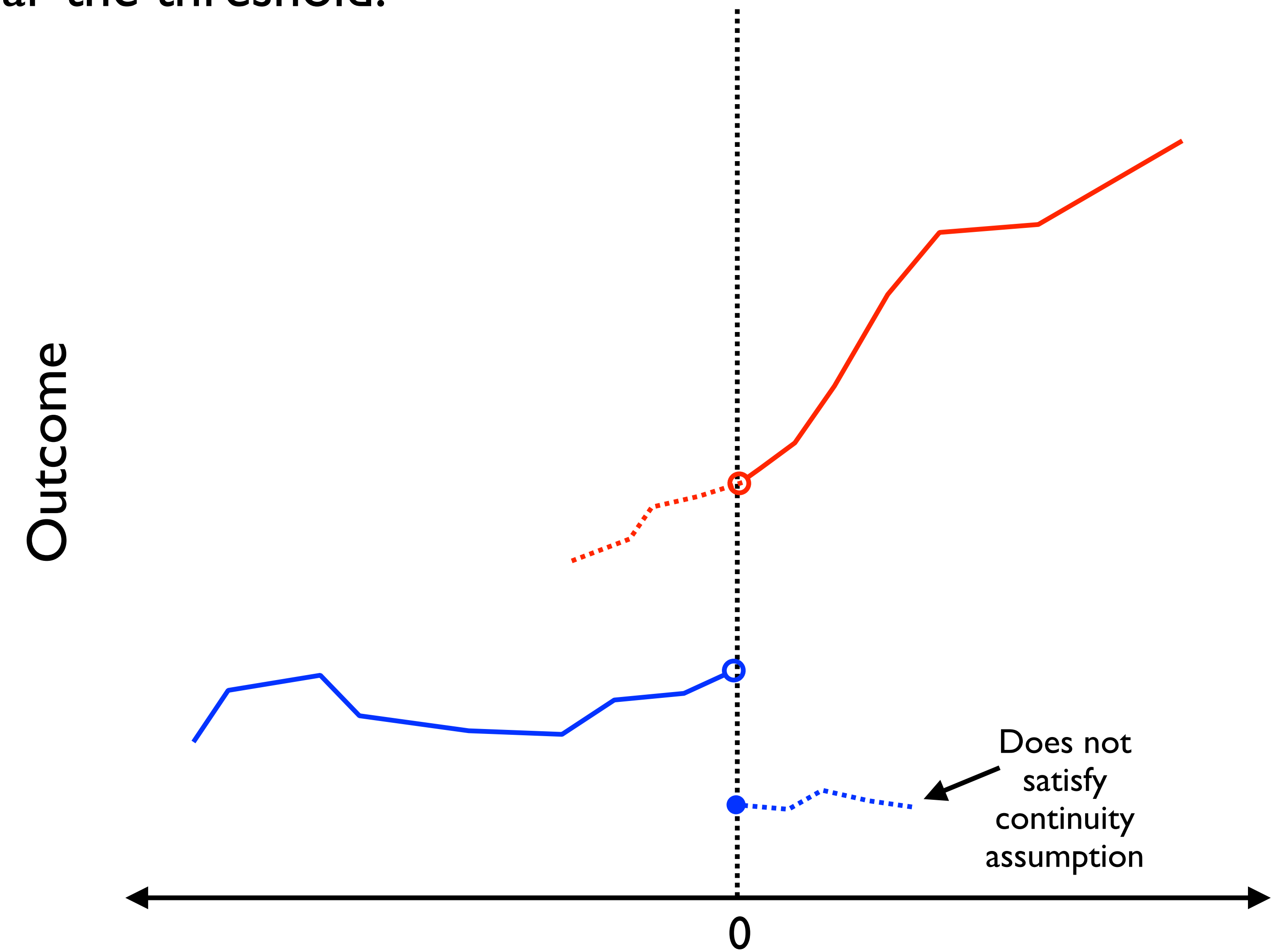
When can we extrapolate to the threshold?

Given infinite data, the only requirement is **continuity of the CEFs** near the threshold.



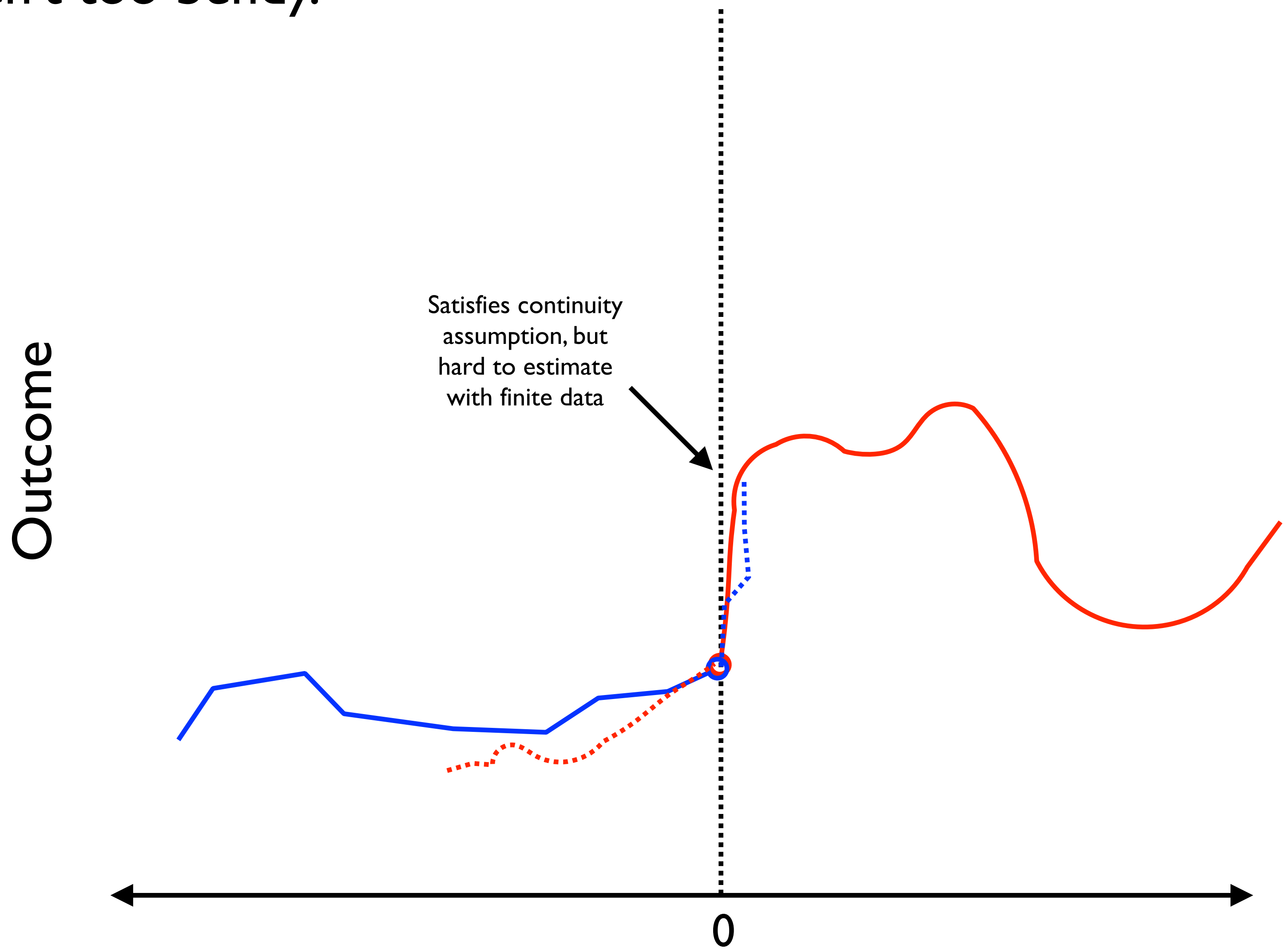
When can we extrapolate to the threshold?

Given infinite data, the only requirement is **continuity of the CEFs** near the threshold.



When can we extrapolate to the threshold? (2)

In practice, we also need either (a) a lot of data or (b) CEFs that aren't too bendy.



 15

When would CEFs be discontinuous (or really bendy) near the threshold?

When would CEFs be discontinuous (or really bendy) near the threshold?

Remember: nature makes no jumps.

When would CEFs be discontinuous (or really bendy) near the threshold?

Remember: nature makes no jumps.

Discontinuities/high bendiness comes from **sorting**: units *selecting their value of X_i* (and thus the treatment) in a way that relates to a confounder Z_i.

When would CEFs be discontinuous (or really bendy) near the threshold?

Remember: nature makes no jumps.

Discontinuities/high bendiness comes from **sorting**: units *selecting their value of X_i* (and thus the treatment) in a way that relates to a confounder $Z_{i.}$

- "In a very close election, the incumbent almost always wins (e.g. through fraud)."

When would CEFs be discontinuous (or really bendy) near the threshold?

Remember: nature makes no jumps.

Discontinuities/high bendiness comes from **sorting**: units *selecting their value of* X_i (and thus the treatment) in a way that relates to a confounder Z_i.

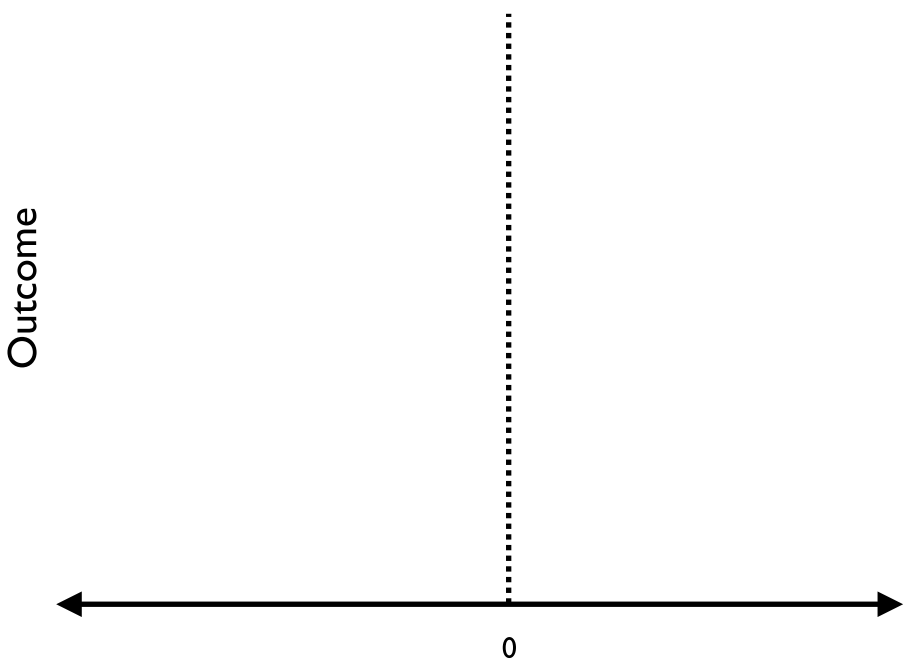
- "In a very close election, the incumbent almost always wins (e.g. through fraud)."
- "When a city is close to a consequential population threshold, savvy mayors make sure their city has enough inhabitants to cross the threshold."

When would CEFs be discontinuous (or really bendy) near the threshold?

Remember: nature makes no jumps.

Discontinuities/high bendiness comes from **sorting**: units *selecting their value of X_i* (and thus the treatment) in a way that relates to a confounder Z_i.

- "In a very close election, the incumbent almost always wins (e.g. through fraud)."
- "When a city is close to a consequential population threshold, savvy mayors make sure their city has enough inhabitants to cross the threshold."

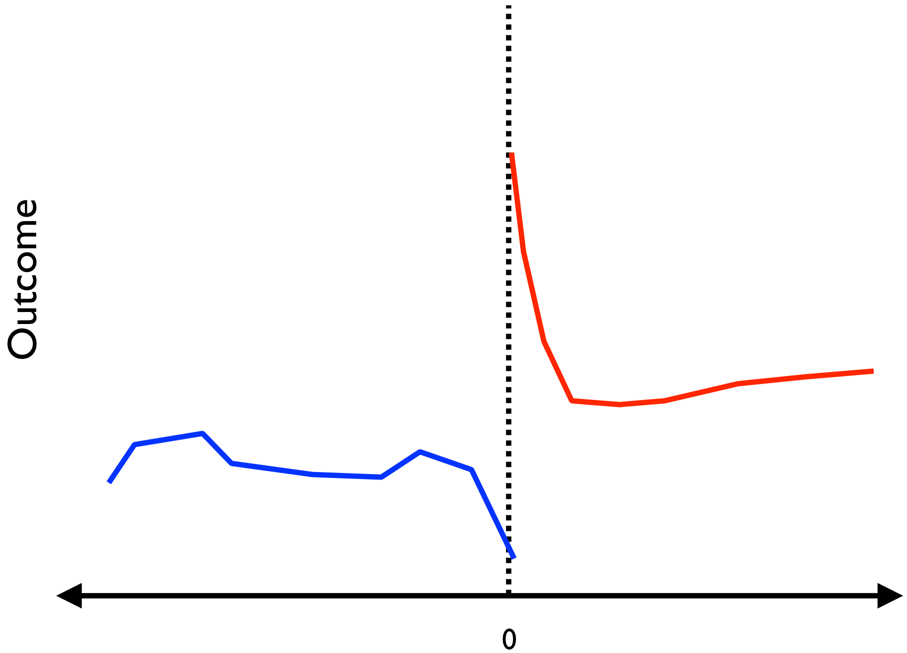


When would CEFs be discontinuous (or really bendy) near the threshold?

Remember: nature makes no jumps.

Discontinuities/high bendiness comes from **sorting**: units *selecting their value of* X_i (and thus the treatment) in a way that relates to a confounder Z_i.

- "In a very close election, the incumbent almost always wins (e.g. through fraud)."
- "When a city is close to a consequential population threshold, savvy mayors make sure their city has enough inhabitants to cross the threshold."

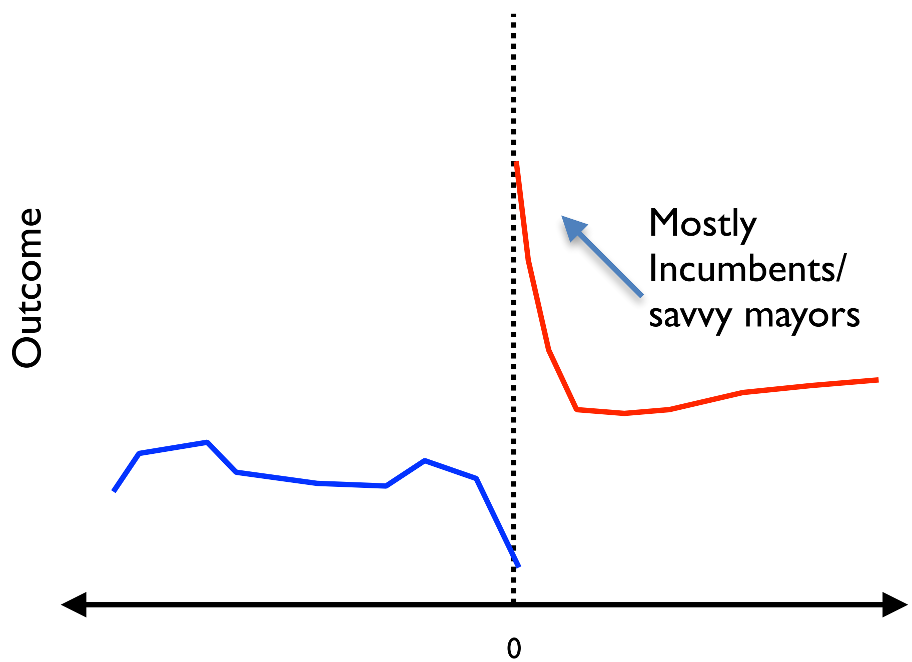


When would CEFs be discontinuous (or really bendy) near the threshold?

Remember: nature makes no jumps.

Discontinuities/high bendiness comes from **sorting**: units *selecting their value of* X_i (and thus the treatment) in a way that relates to a confounder Z_i.

- "In a very close election, the incumbent almost always wins (e.g. through fraud)."
- "When a city is close to a consequential population threshold, savvy mayors make sure their city has enough inhabitants to cross the threshold."

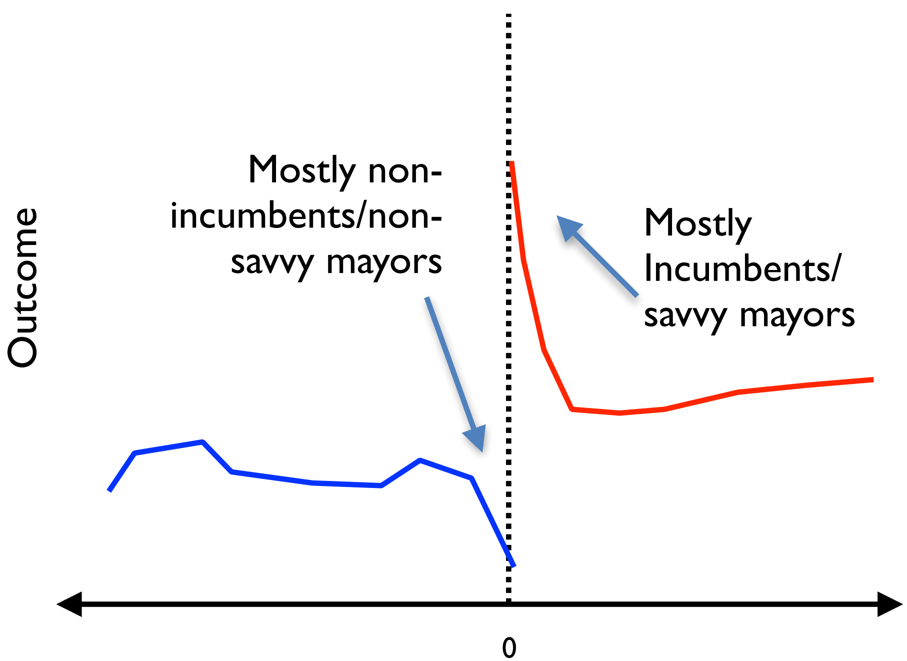


When would CEFs be discontinuous (or really bendy) near the threshold?

Remember: nature makes no jumps.

Discontinuities/high bendiness comes from **sorting**: units *selecting their value of* X_i (and thus the treatment) in a way that relates to a confounder Z_i.

- "In a very close election, the incumbent almost always wins (e.g. through fraud)."
- "When a city is close to a consequential population threshold, savvy mayors make sure their city has enough inhabitants to cross the threshold."

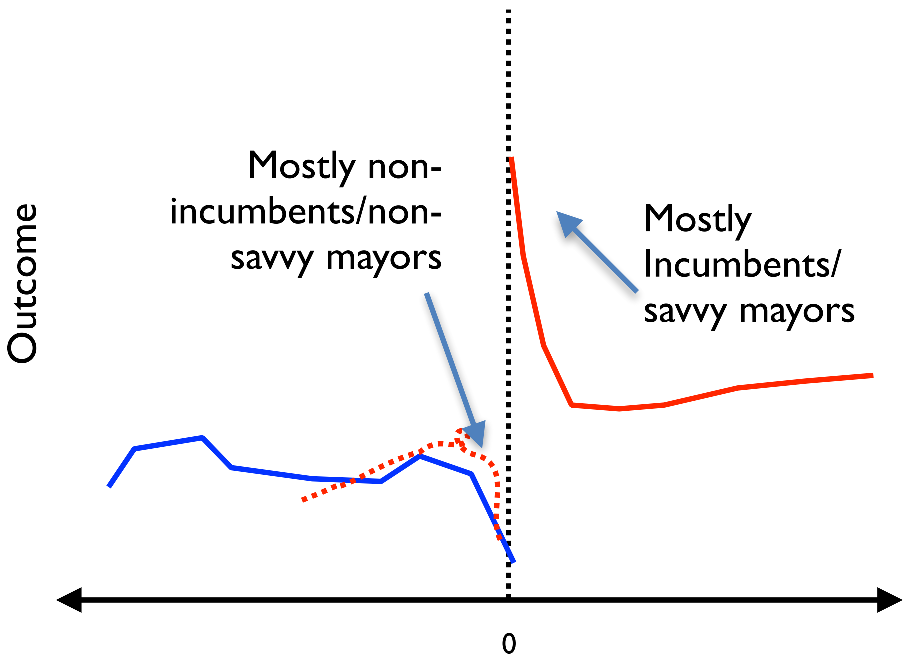


When would CEFs be discontinuous (or really bendy) near the threshold?

Remember: nature makes no jumps.

Discontinuities/high bendiness comes from **sorting**: units *selecting their value of X_i* (and thus the treatment) in a way that relates to a confounder Z_i.

- "In a very close election, the incumbent almost always wins (e.g. through fraud)."
- "When a city is close to a consequential population threshold, savvy mayors make sure their city has enough inhabitants to cross the threshold."

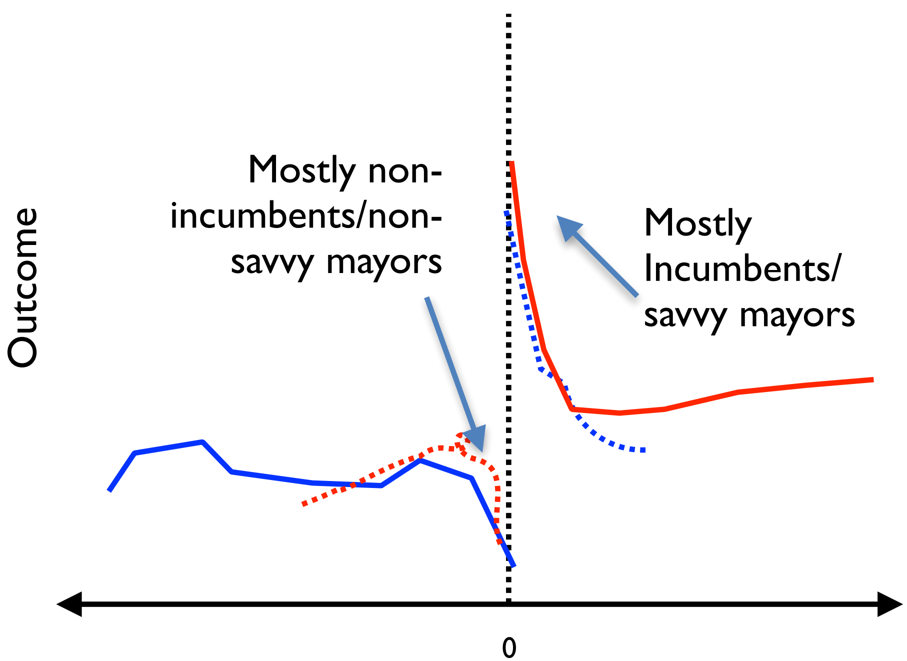


When would CEFs be discontinuous (or really bendy) near the threshold?

Remember: nature makes no jumps.

Discontinuities/high bendiness comes from **sorting**: units *selecting their value of X_i* (and thus the treatment) in a way that relates to a confounder $Z_{i.}$

- "In a very close election, the incumbent almost always wins (e.g. through fraud)."
- "When a city is close to a consequential population threshold, savvy mayors make sure their city has enough inhabitants to cross the threshold."

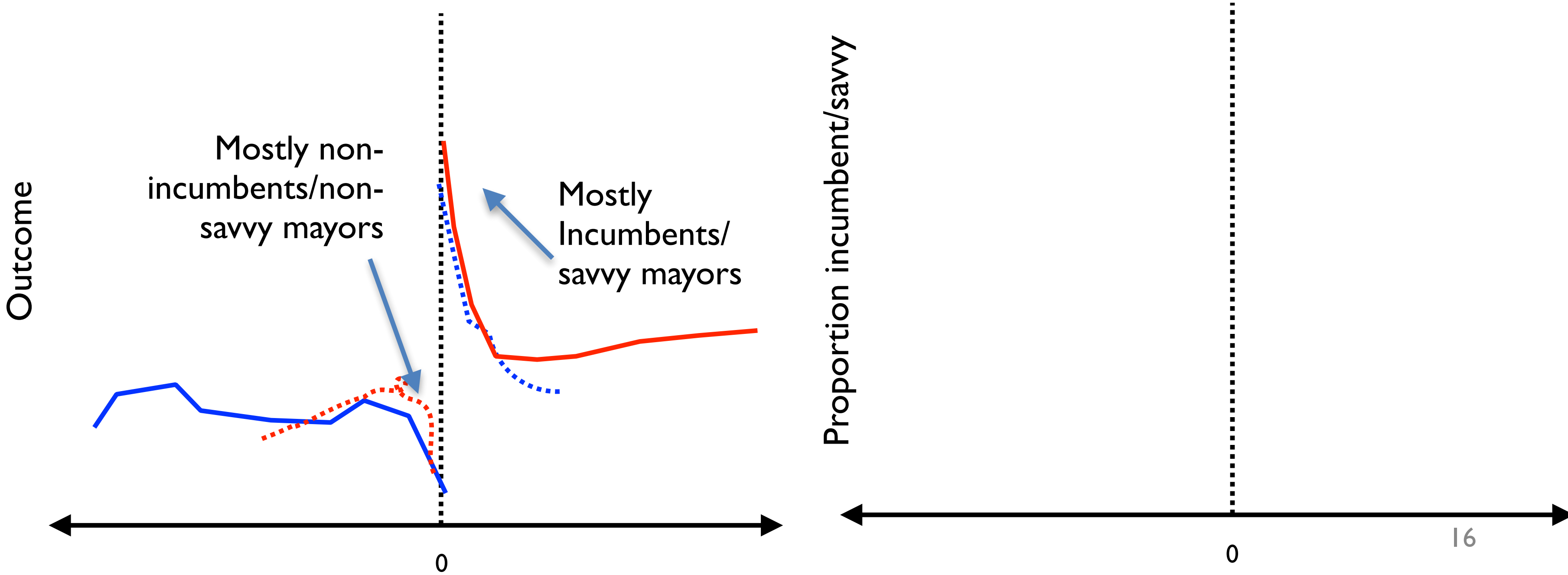


When would CEFs be discontinuous (or really bendy) near the threshold?

Remember: nature makes no jumps.

Discontinuities/high bendiness comes from **sorting**: units *selecting their value of* X_i (and thus the treatment) in a way that relates to a confounder $Z_{i.}$

- "In a very close election, the incumbent almost always wins (e.g. through fraud)."
- "When a city is close to a consequential population threshold, savvy mayors make sure their city has enough inhabitants to cross the threshold."

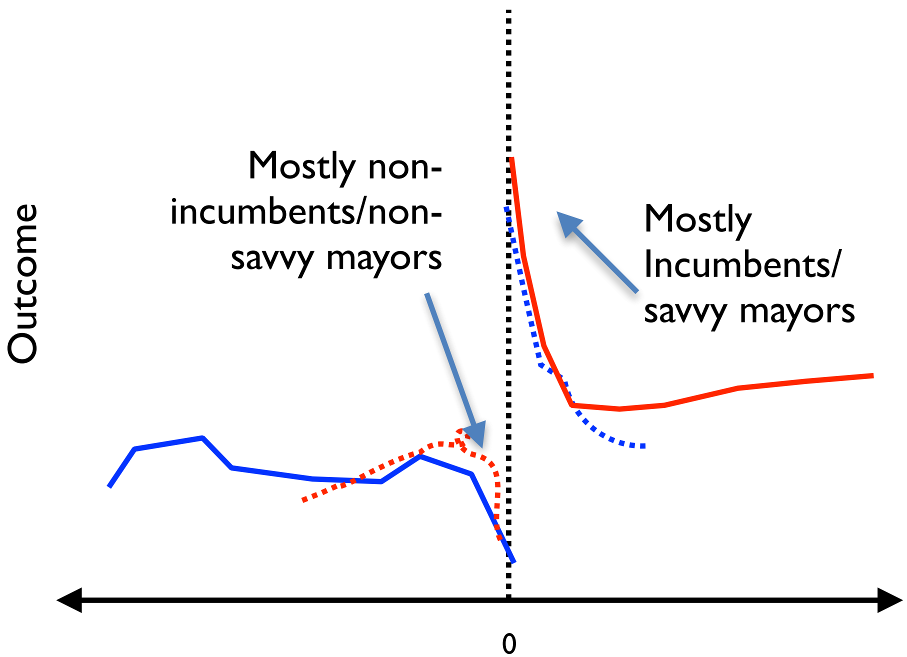


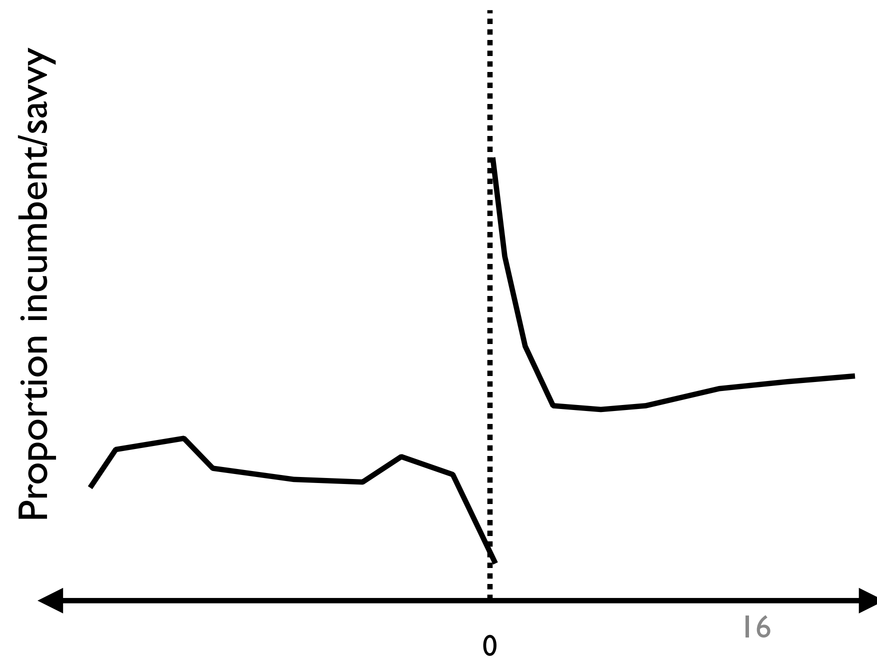
When would CEFs be discontinuous (or really bendy) near the threshold?

Remember: nature makes no jumps.

Discontinuities/high bendiness comes from **sorting**: units *selecting their value of X_i* (and thus the treatment) in a way that relates to a confounder $Z_{i.}$

- "In a very close election, the incumbent almost always wins (e.g. through fraud)."
- "When a city is close to a consequential population threshold, savvy mayors make sure their city has enough inhabitants to cross the threshold."





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: Eggers and Spirling on incumbency effects in UK politics

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

Research question: How much do voters care about candidate characteristics in British elections? Does it depend on the partisan stakes of the election?

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

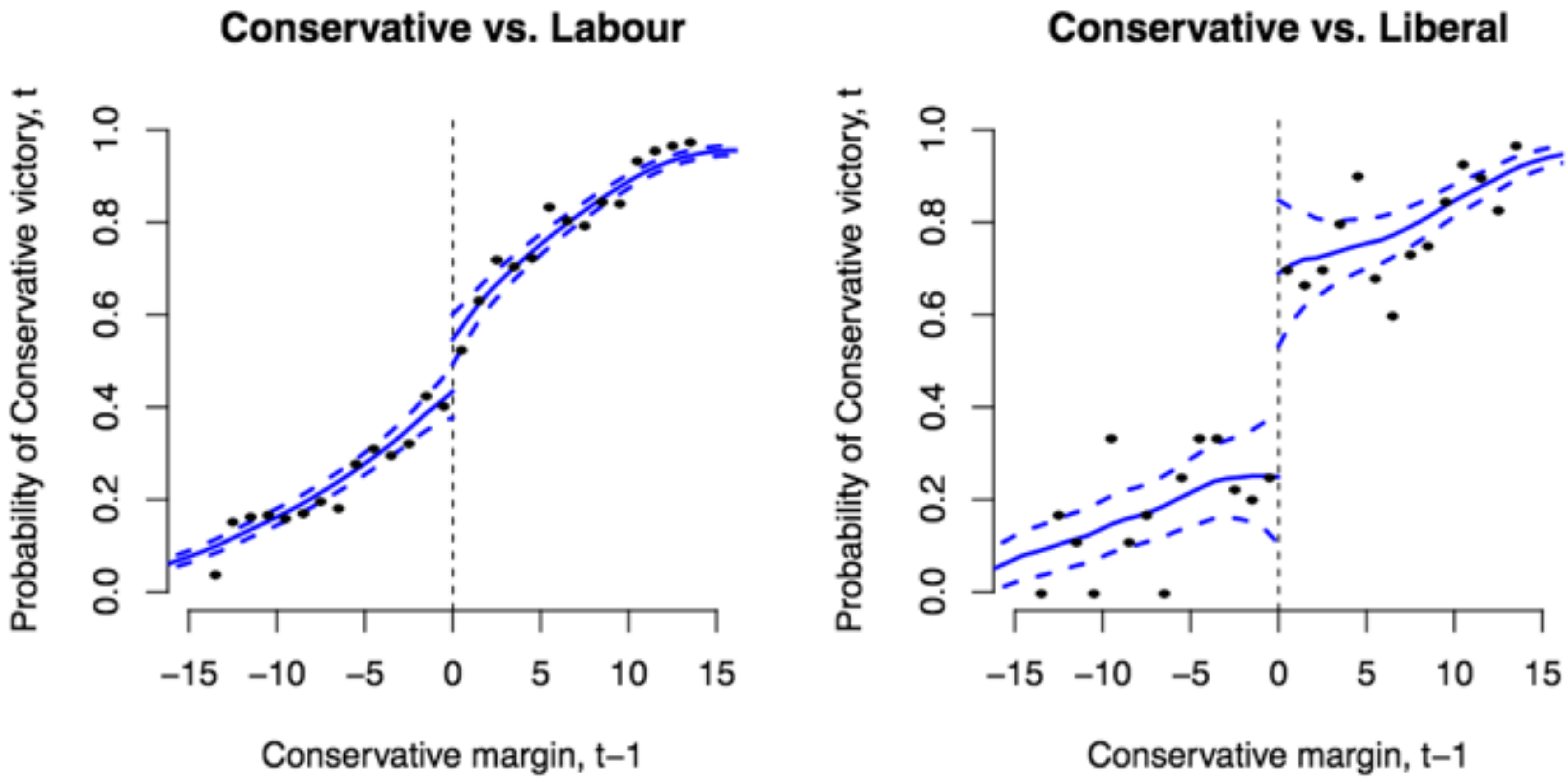
Research question: How much do voters care about candidate characteristics in British elections? Does it depend on the partisan stakes of the election?

Strategy: Use RDD to measure incumbency advantage in UK general elections; compare size of effect in Con-Lab and Con-Lib constituencies.

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

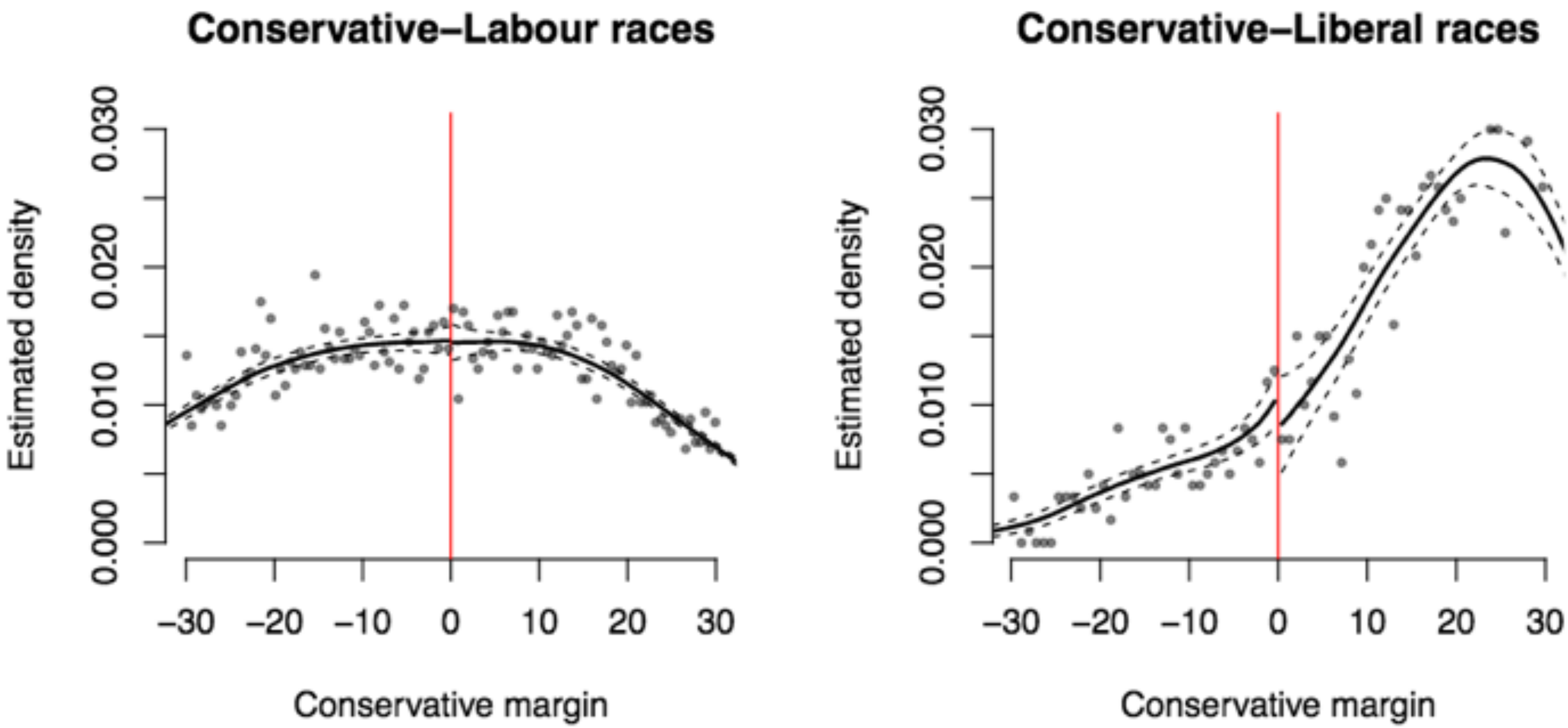
Research question: How much do voters care about candidate characteristics in British elections? Does it depend on the partisan stakes of the election?

Strategy: Use RDD to measure incumbency advantage in UK general elections; compare size of effect in Con-Lab and Con-Lib constituencies.



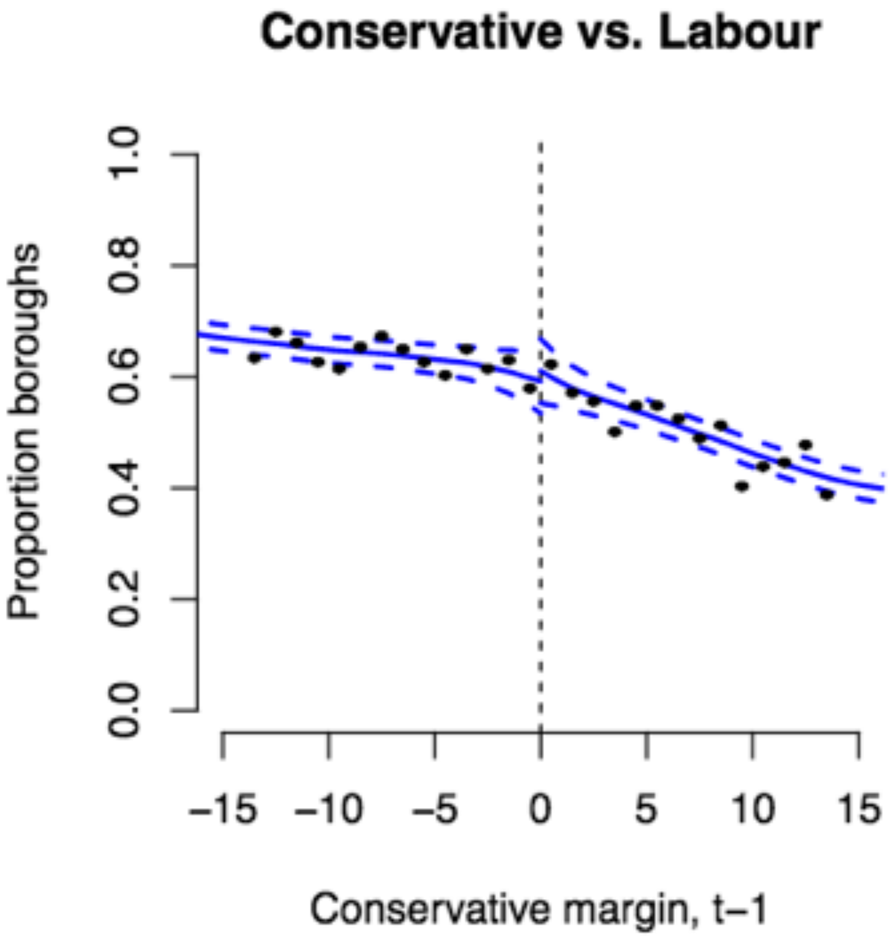
Example 1: tests of the continuity assumption (1)

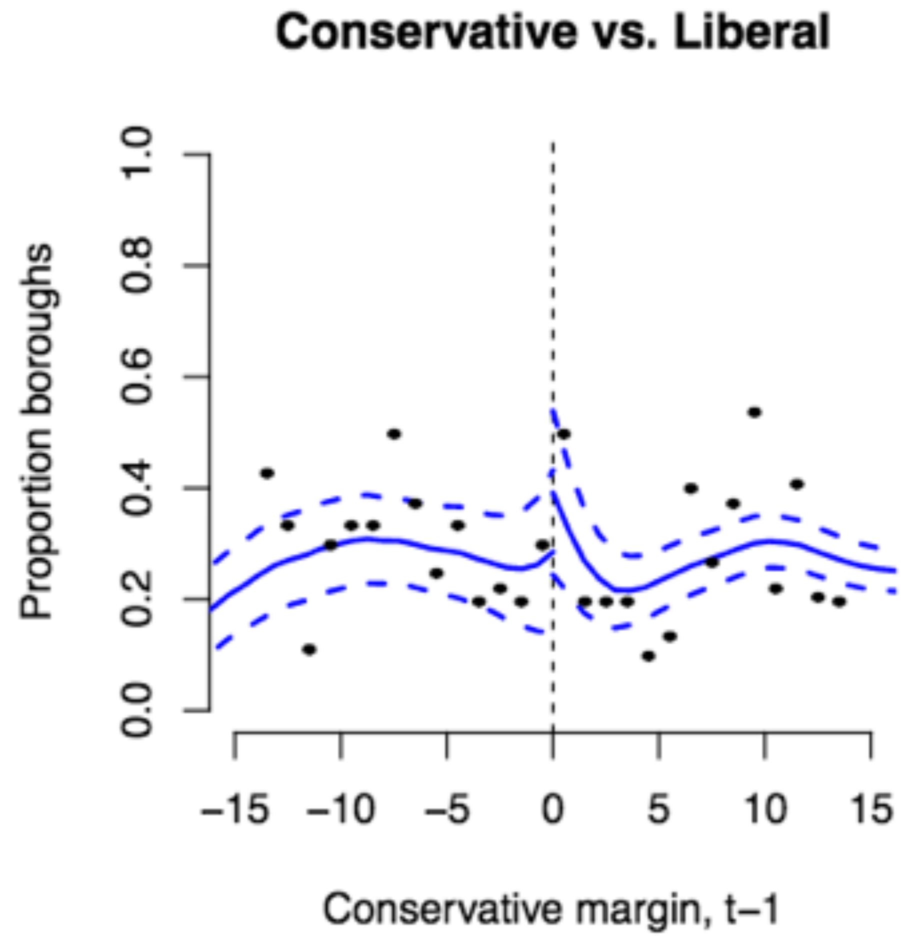
McCrary test for continuity in the density



Example 1: tests of the continuity assumption (2)

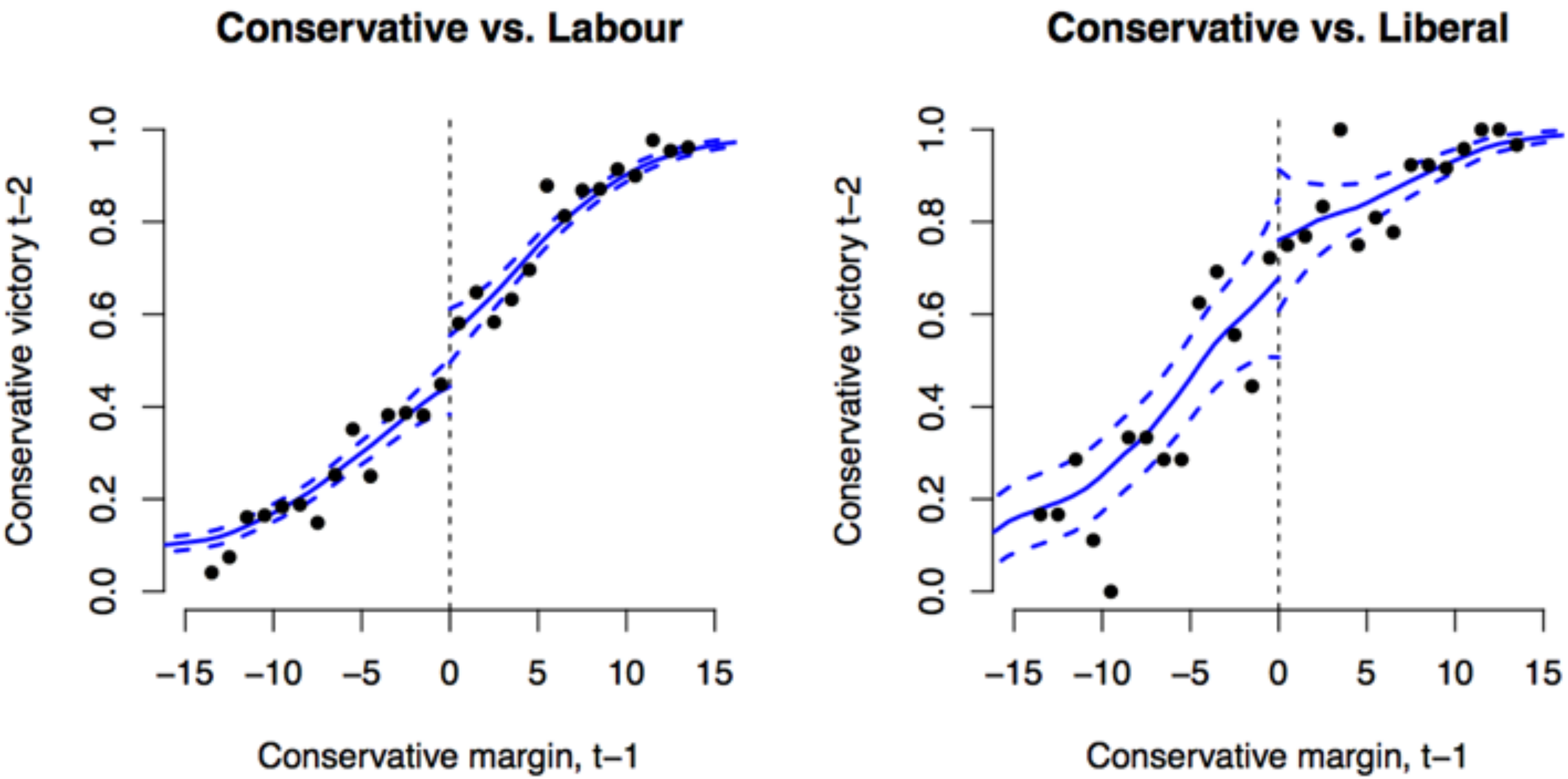
Tests for continuity in covariate: whether or not the election took place in a borough (vs county) constituency





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?

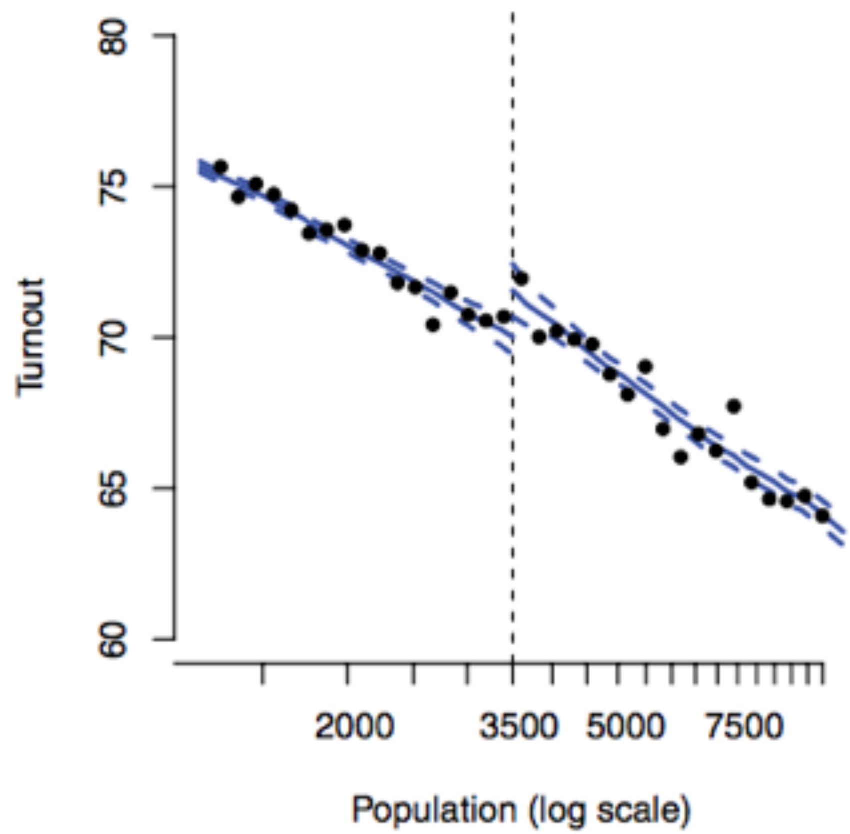
Strategy: Use RDD to compare French municipalities using PR and plurality systems at the 3,500 population cutoff determining electoral system

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)	Effect (2)	Effect (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.

		Effect estimates		
Outcome	*Mean*	(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.

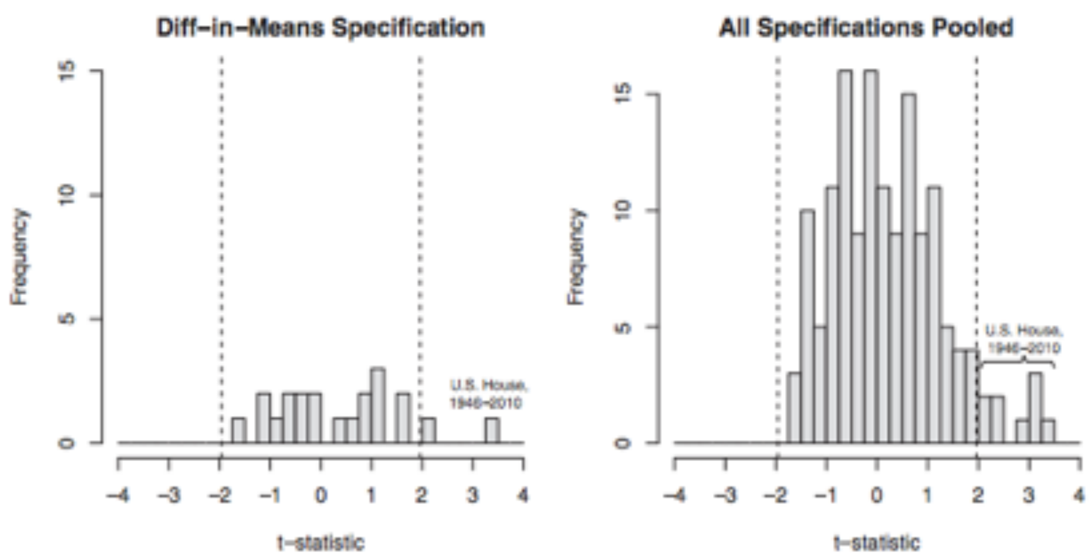
Eggers, Fowler, Hainmueller, Hall, Snyder (2015): Looking at other periods and legislatures in US and elsewhere, similar problems not found anywhere else. Caughey and Sekhon (2011) pattern probably a fluke.

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)

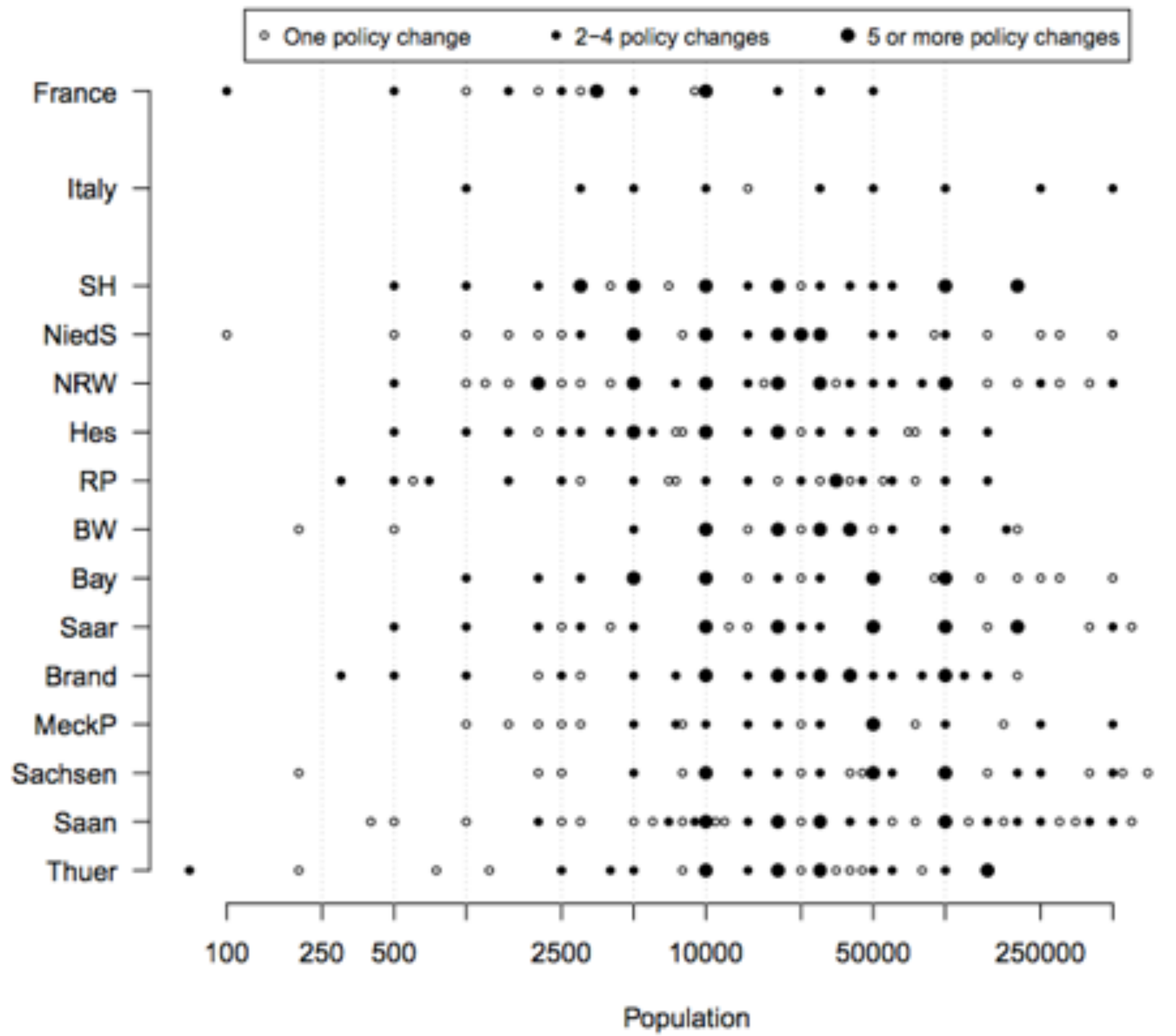
Eggers, Freier, Grembi, and Nannicini (forthcoming): There may be more reason to doubt RDDs based on population thresholds.

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.



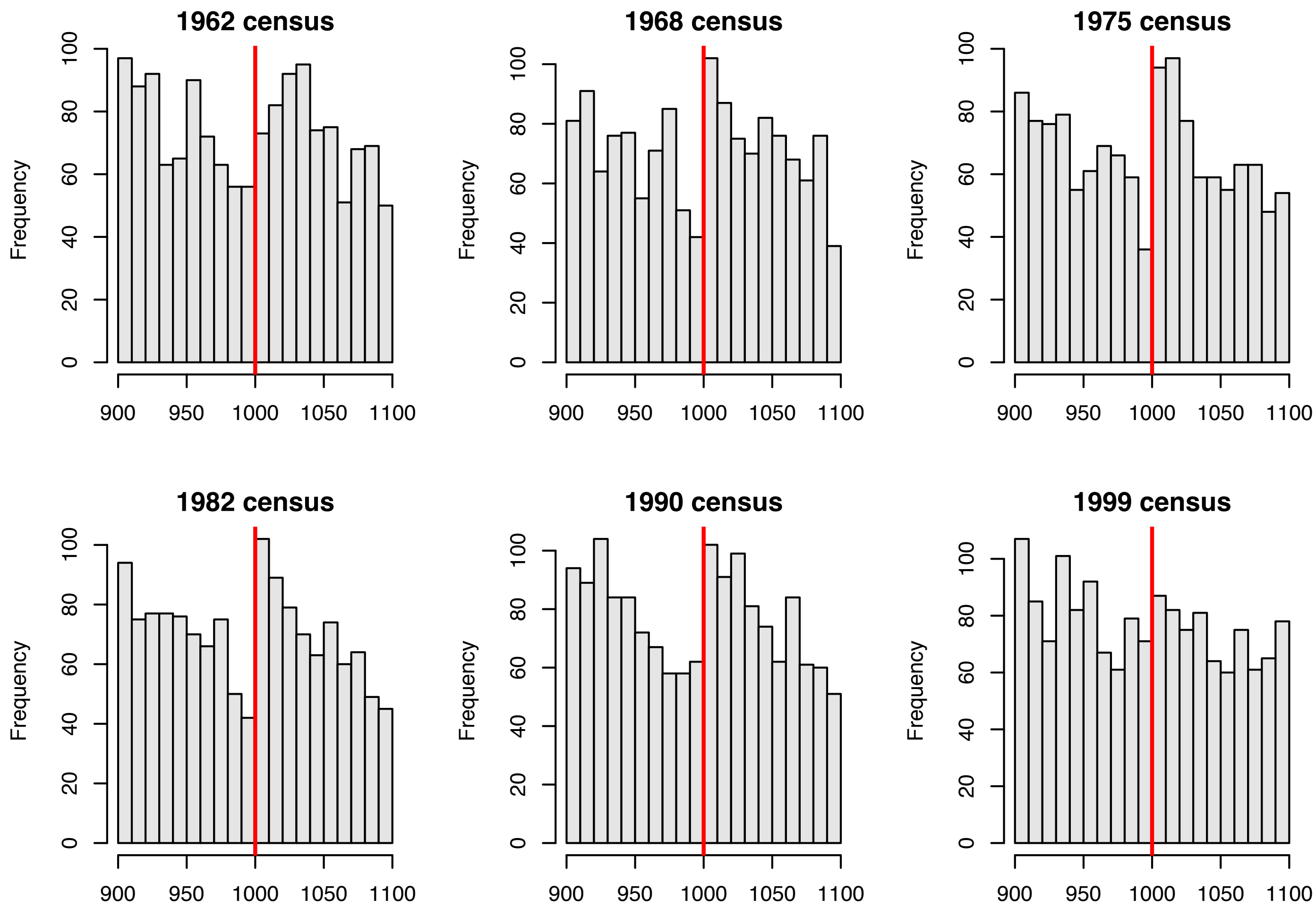
Does RDD work for political science applications? The case of population thresholds (2)

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:

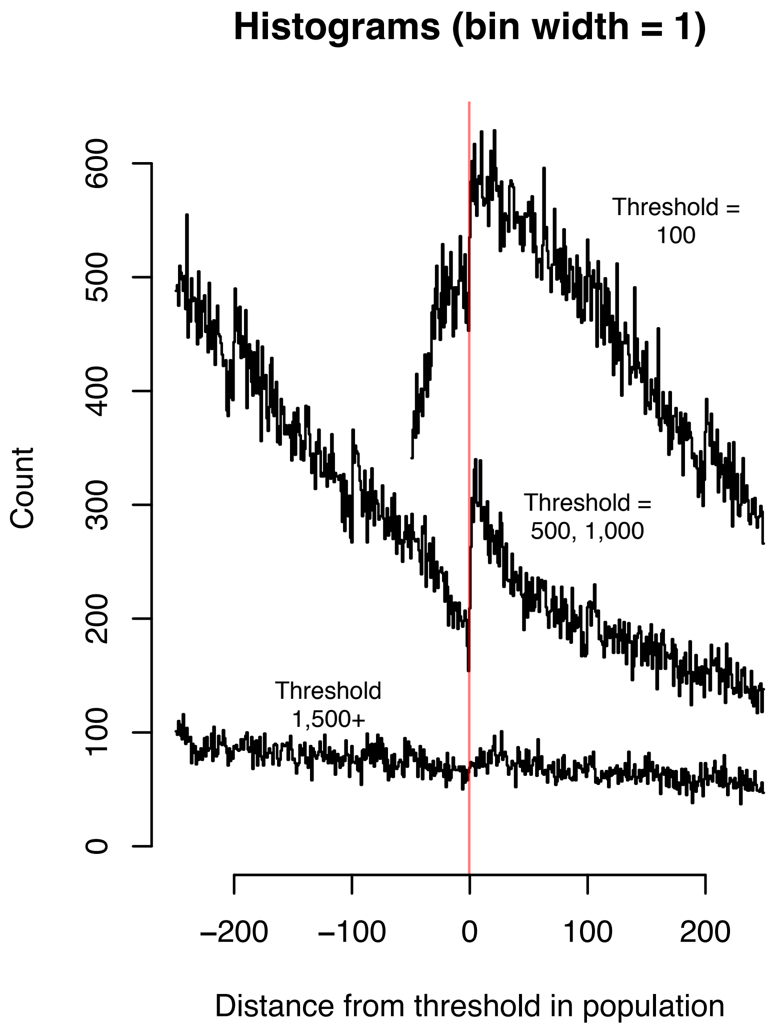
Does RDD work for political science applications? The case of population thresholds (2)

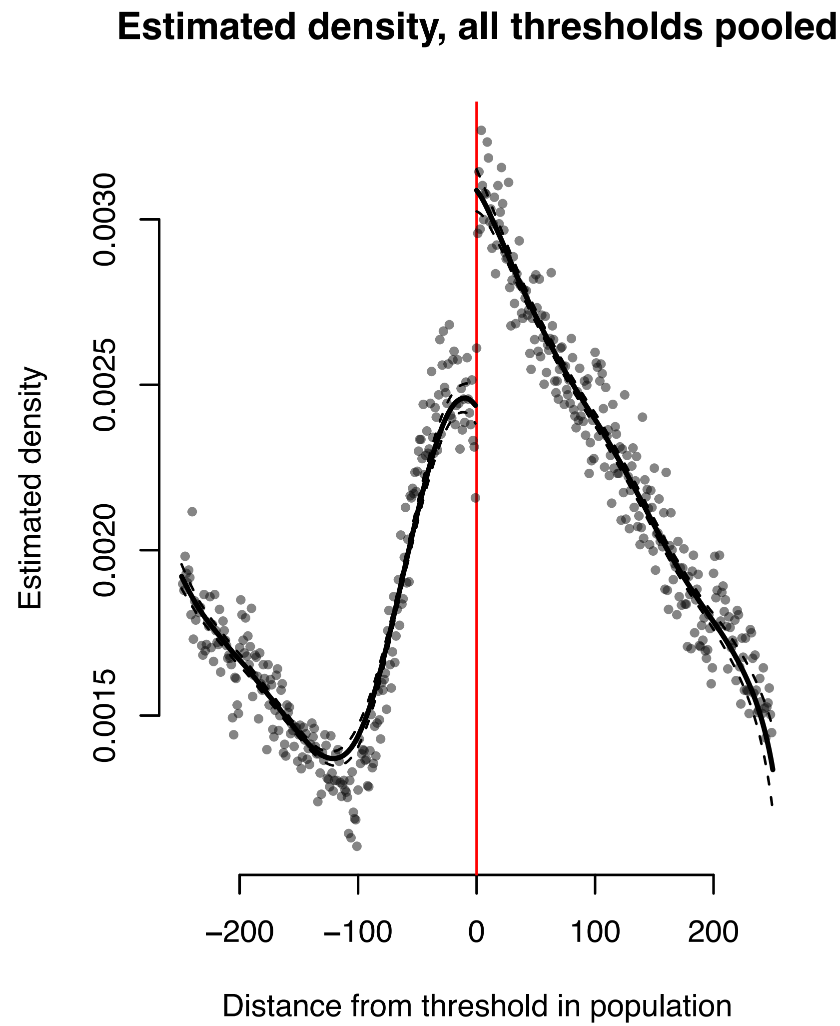
Second problem: **sorting**. Consider histograms of population in French villages near 1,000 in pop., where mayor's salary increases:



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

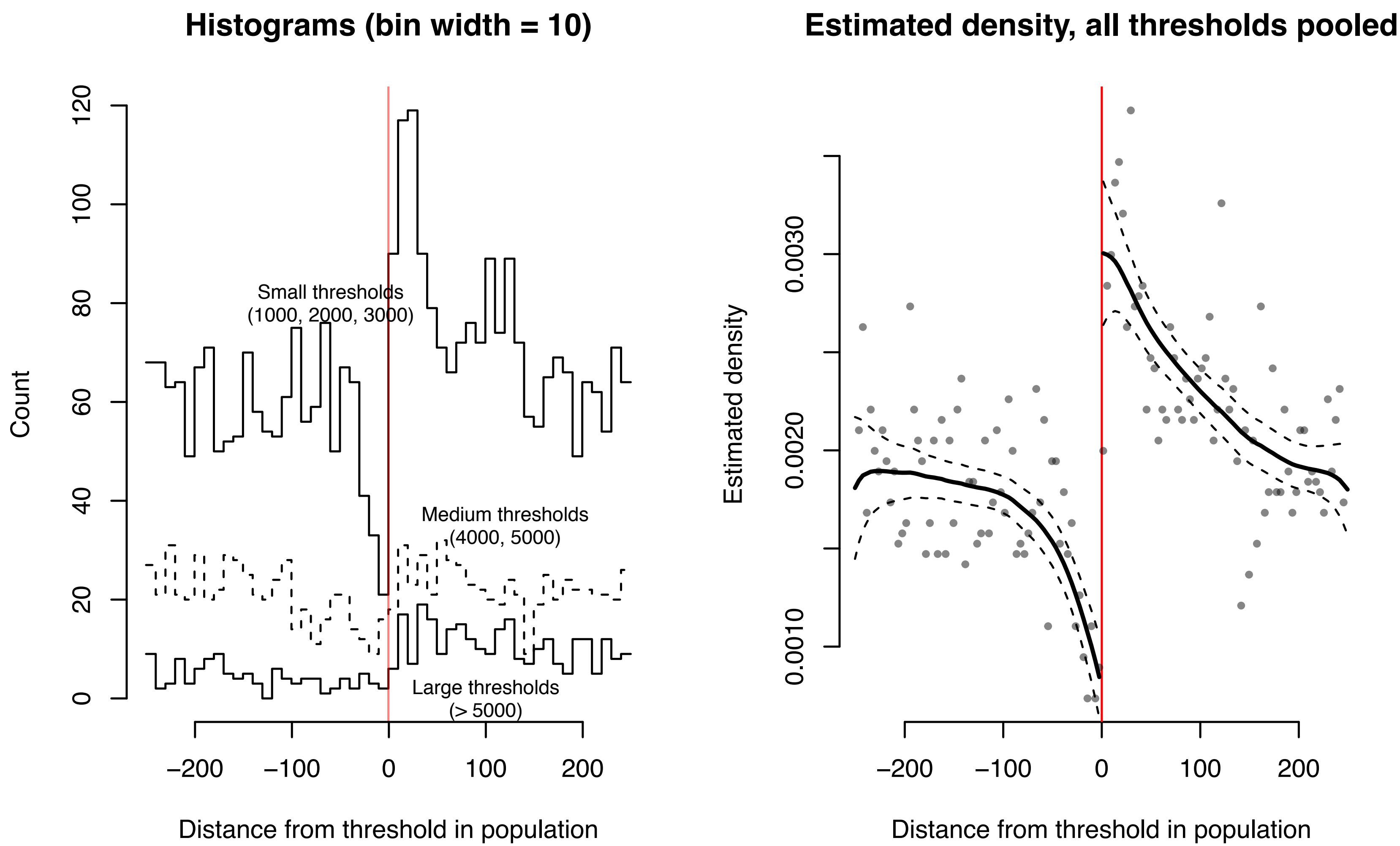
Pooling all thresholds, censuses from France:





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?

Estimation

Estimation

All we need to do is estimate two CEFs at the threshold:

Estimation

All we need to do is estimate two CEFs at the threshold:

$$E[Y_1|X_i = 0] = \lim_{x \to 0^+} E[Y_1|X_i = x]$$

$$E[Y_0|X_i = 0] = \lim_{x \to 0^-} E[Y_0|X_i = x]$$

Estimation

All we need to do is estimate two CEFs at the threshold:

$$E[Y_1|X_i = 0] = \lim_{x \to 0^+} E[Y_1|X_i = x]$$

$$E[Y_0|X_i = 0] = \lim_{x \to 0^-} E[Y_0|X_i = x]$$

So, how do we do it?

Estimation

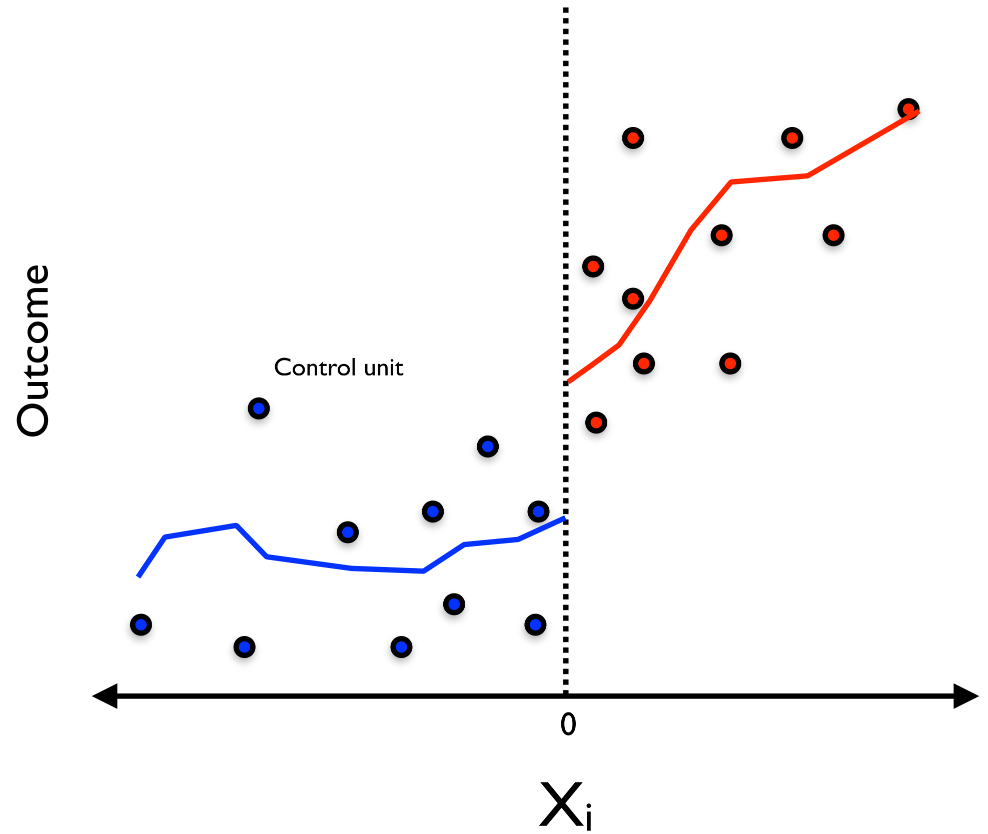
All we need to do is estimate two CEFs at the threshold:

$$E[Y_1|X_i = 0] = \lim_{x \to 0^+} E[Y_1|X_i = x]$$
$$E[Y_0|X_i = 0] = \lim_{x \to 0^-} E[Y_0|X_i = x]$$

So, how do we do it?

Over past 10 years, much variety. Something simple gets complicated!

Local linear regression approach

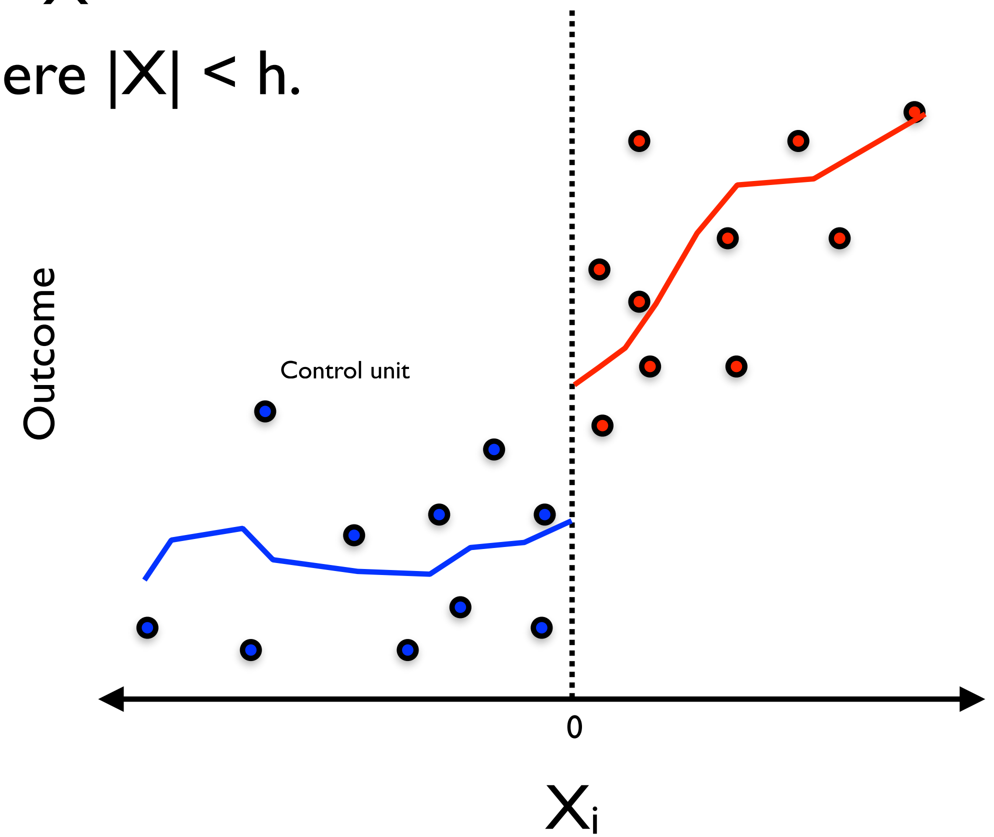


Local linear regression approach

Consider running this regression:

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

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



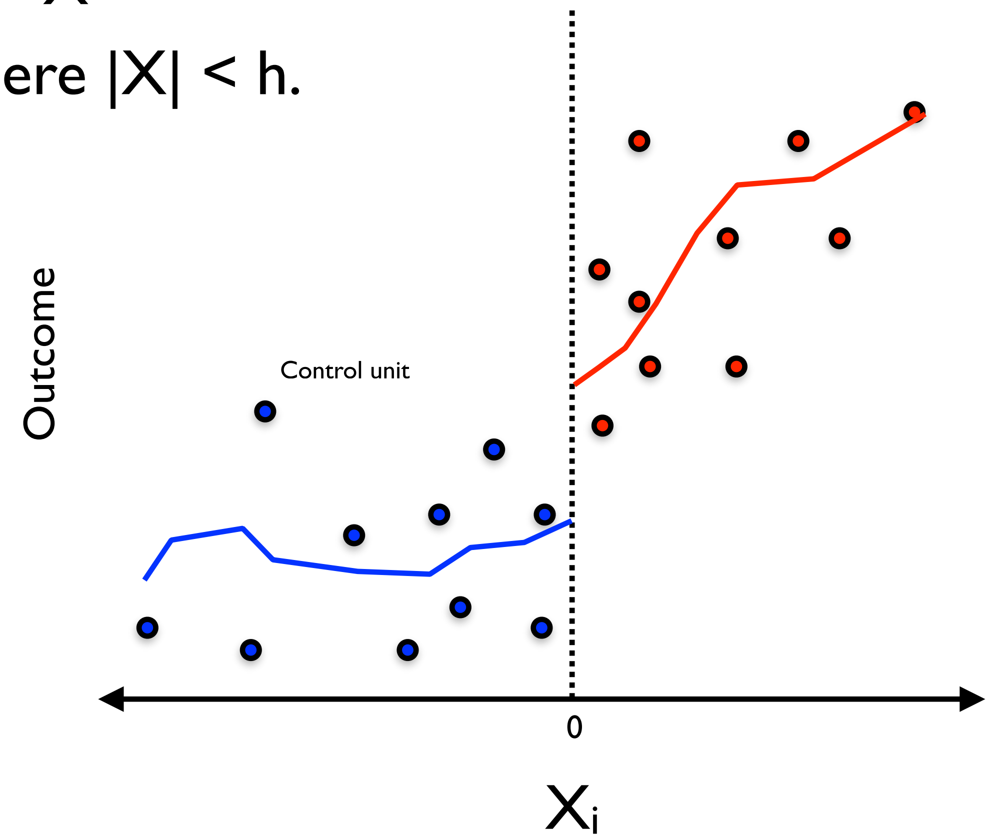
Local linear regression approach

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:



Local linear regression approach

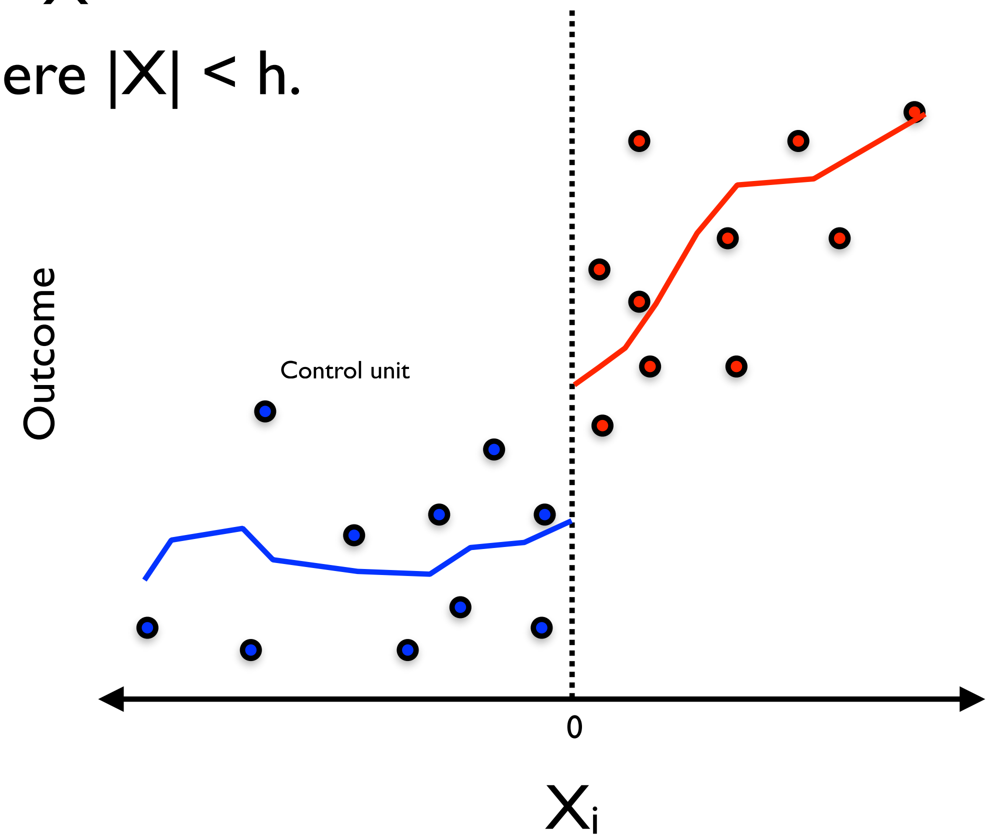
Consider running this regression:

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

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

Questions:

- What will the fitted equation look like?



Local linear regression approach

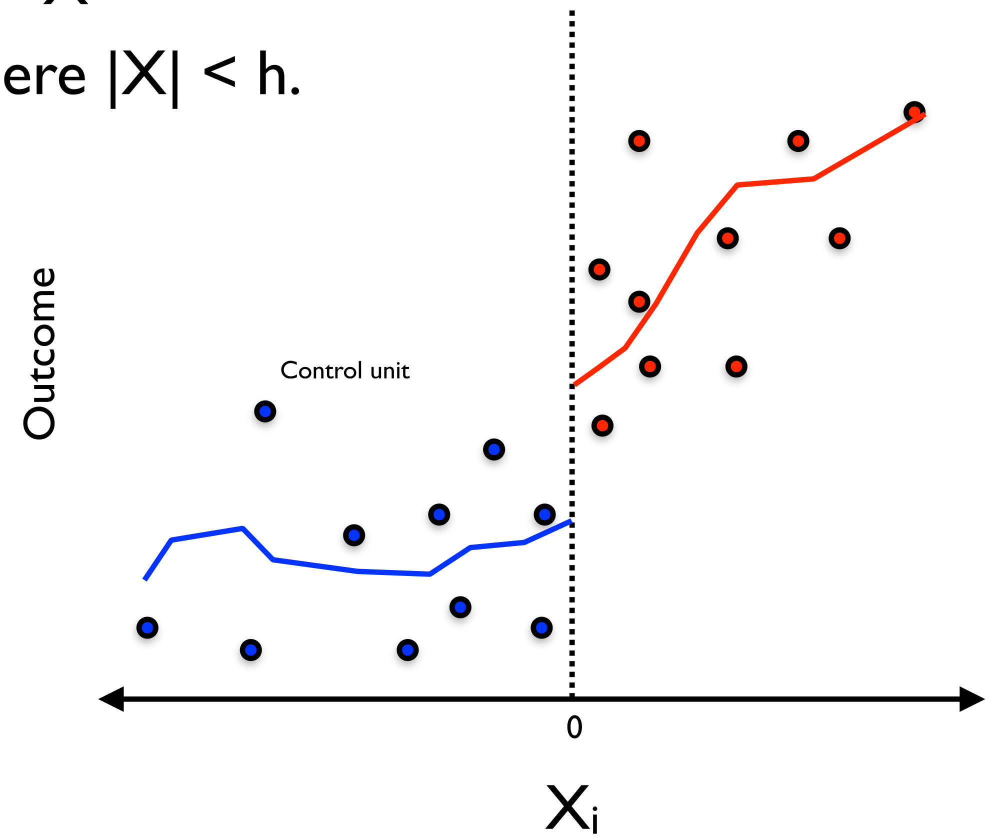
Consider running this regression:

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

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

Questions:

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



Local linear regression approach

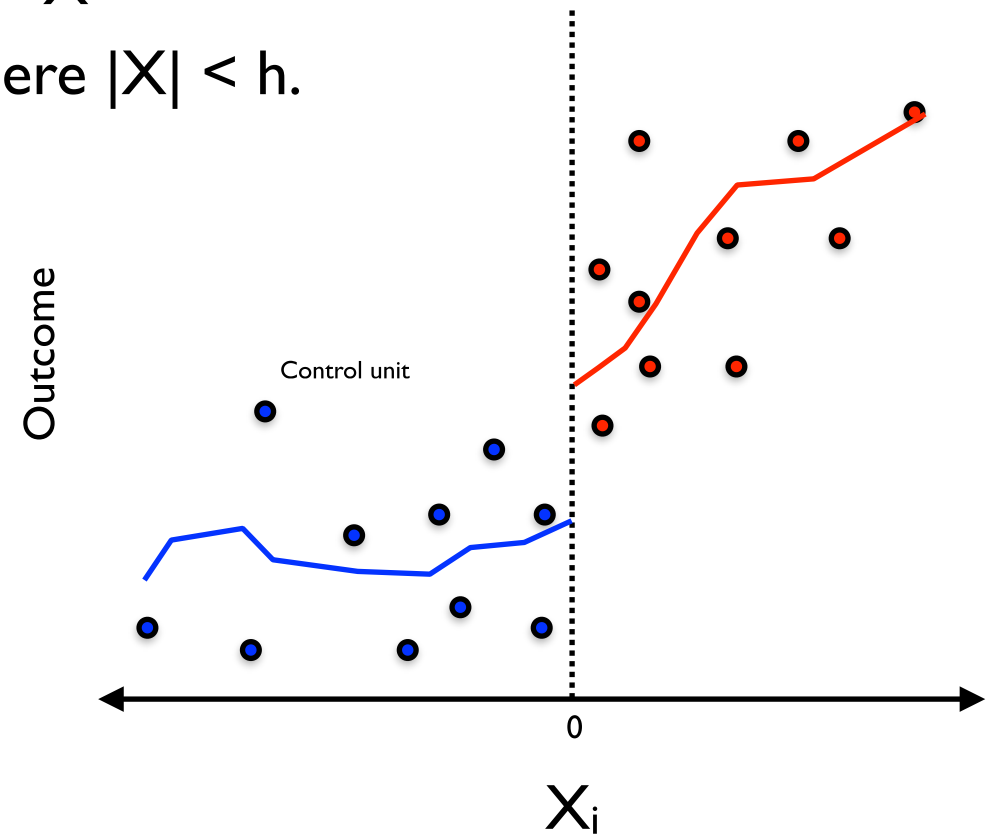
Consider running this regression:

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

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?



Local linear regression approach

Consider running this regression:

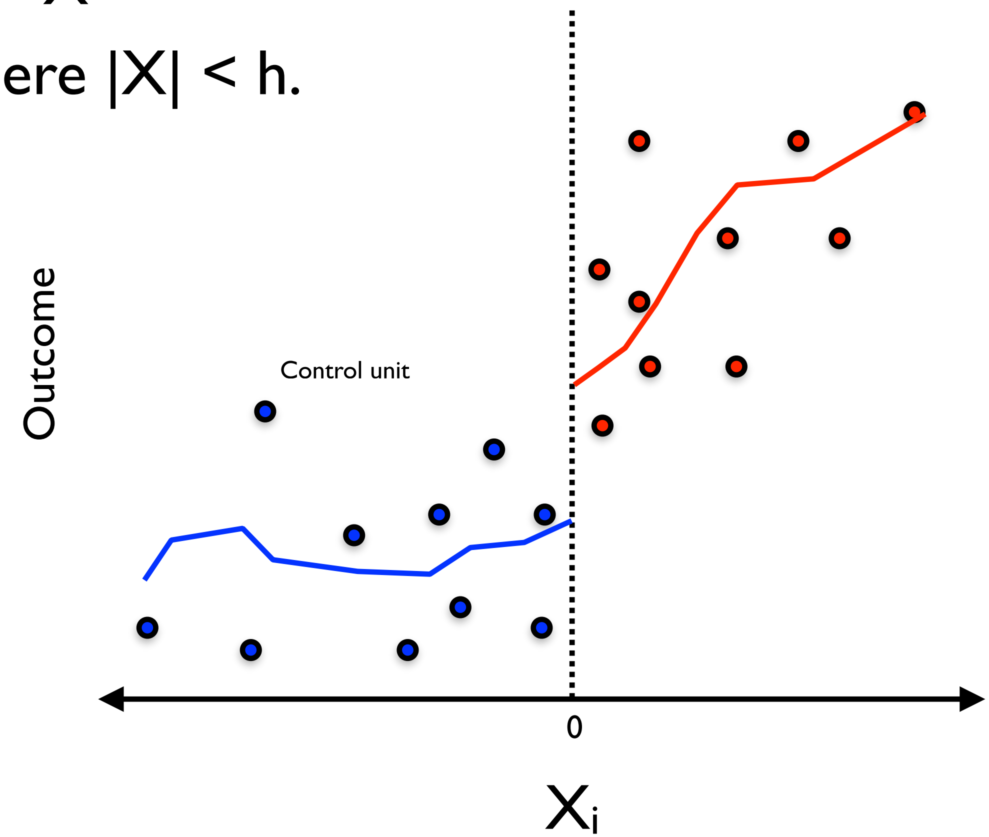
$Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 D \times X$

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

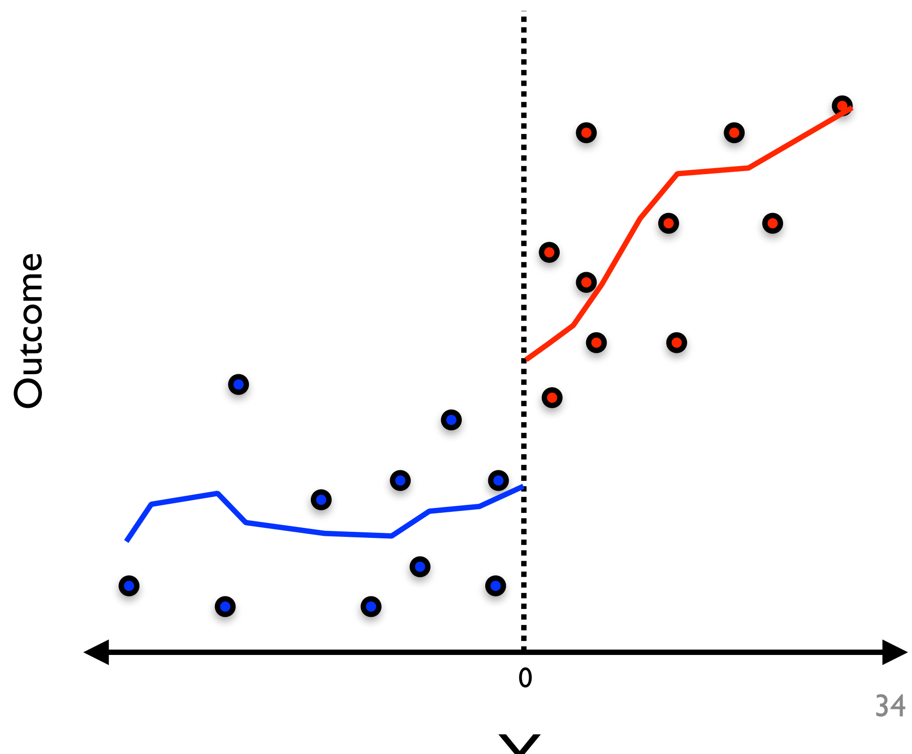
Questions:

- 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**.

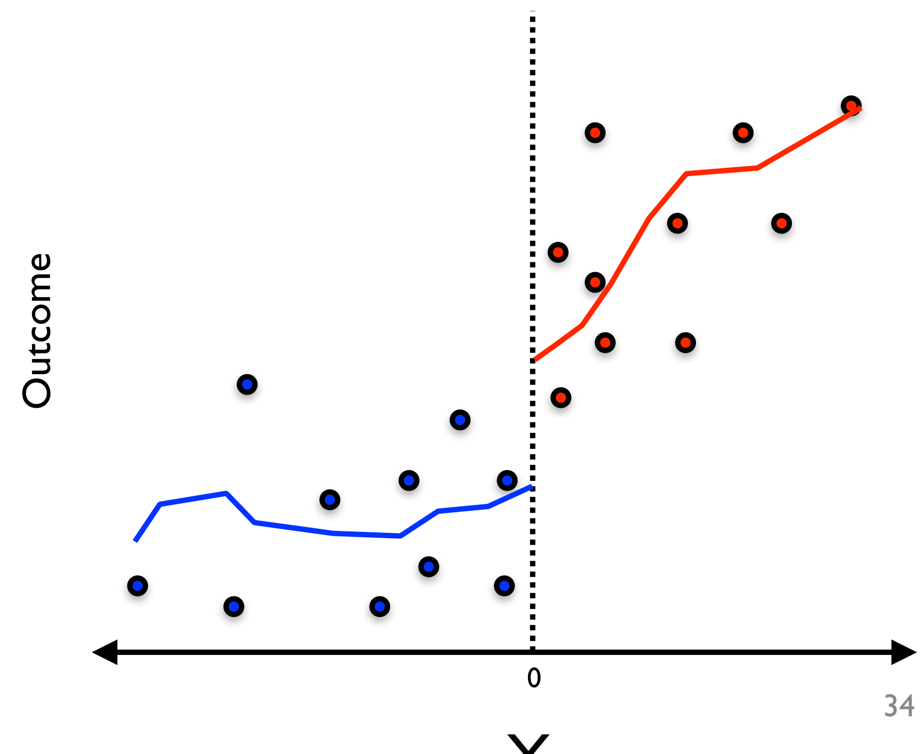


What people do now



What people do now

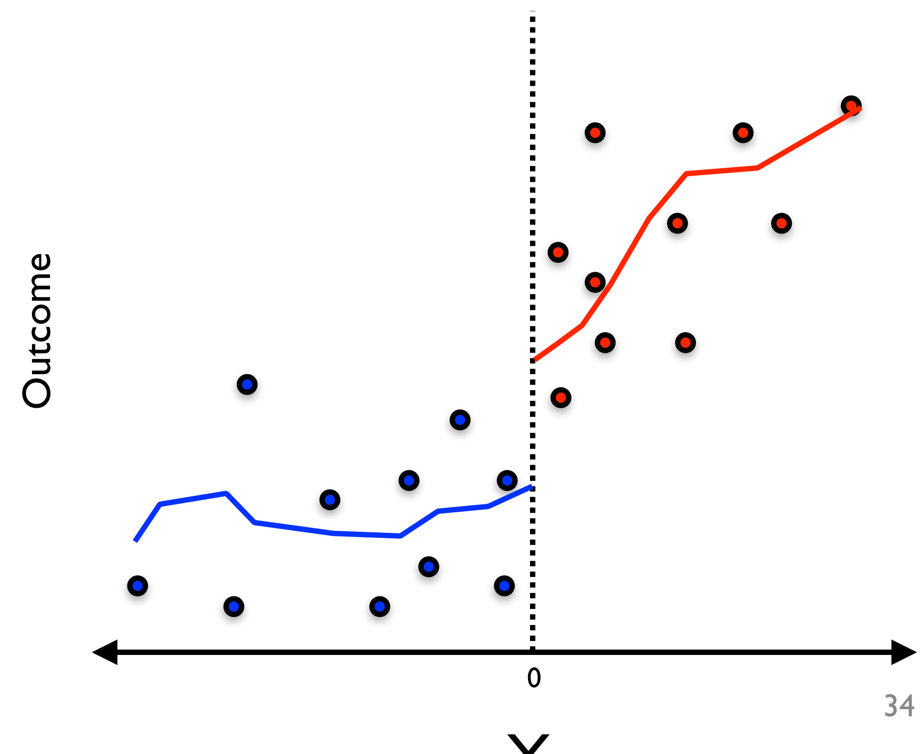
Most common approach: local regression, as on previous slide, with:



What people do now

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

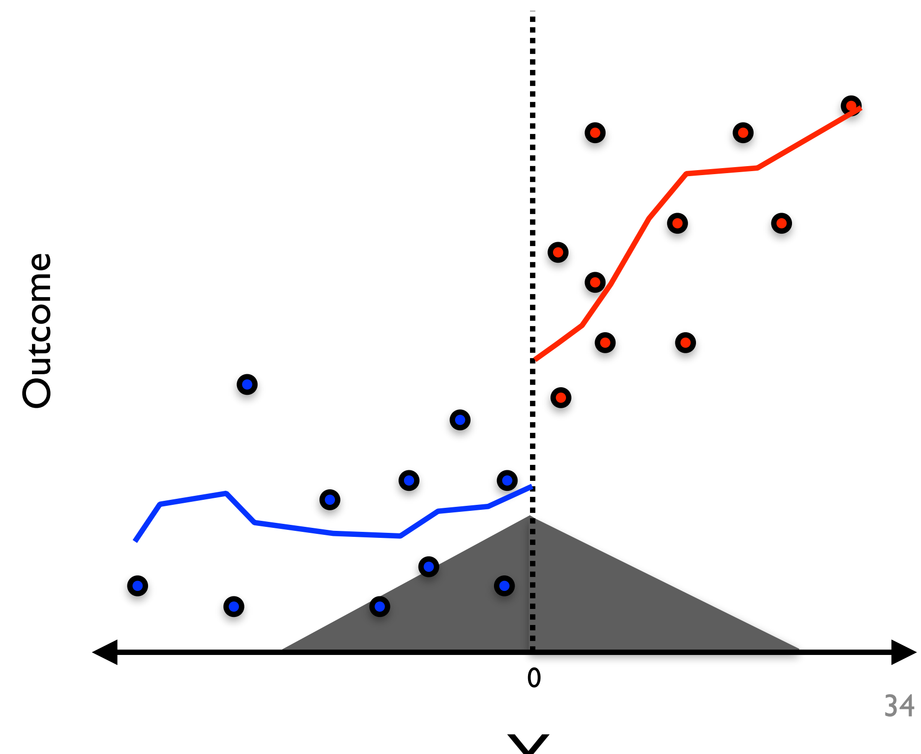
- bandwidth (h) chosen by algorithm that considers curvature of CEF and density of data



What people do now

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

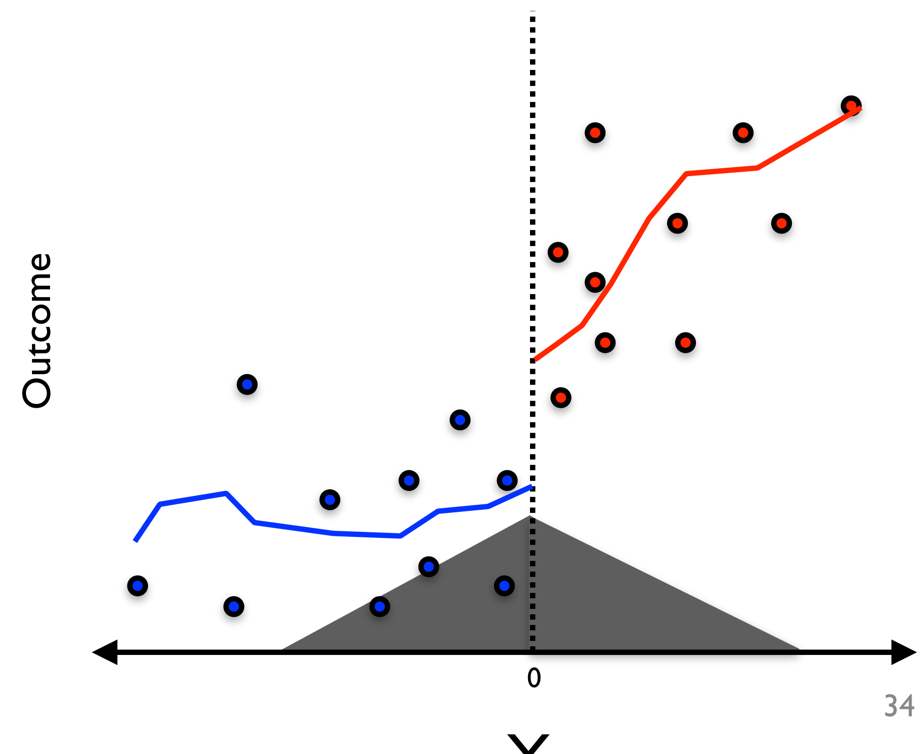
- bandwidth (h) chosen by algorithm that considers curvature of CEF and density of data
- weights that are larger closer to the threshold (triangular kernel)



What people do now

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

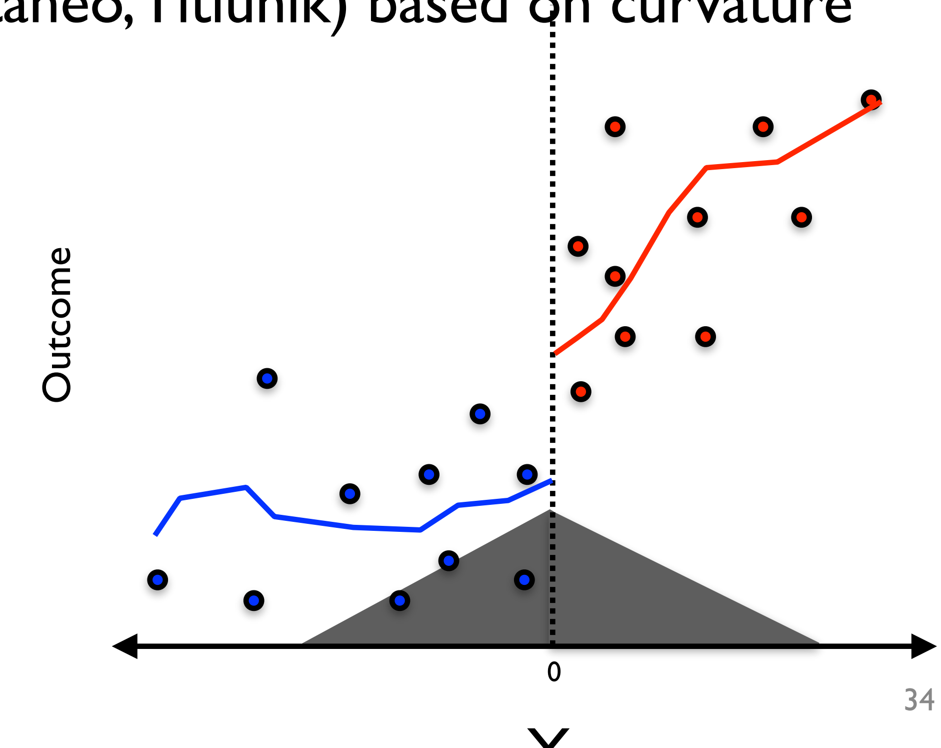
- bandwidth (h) chosen by algorithm that considers curvature of CEF and density of data
- weights that are larger closer to the threshold (triangular kernel)
- (sometimes) polynomial terms (square, cube, etc) in estimating CEF



What people do now

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

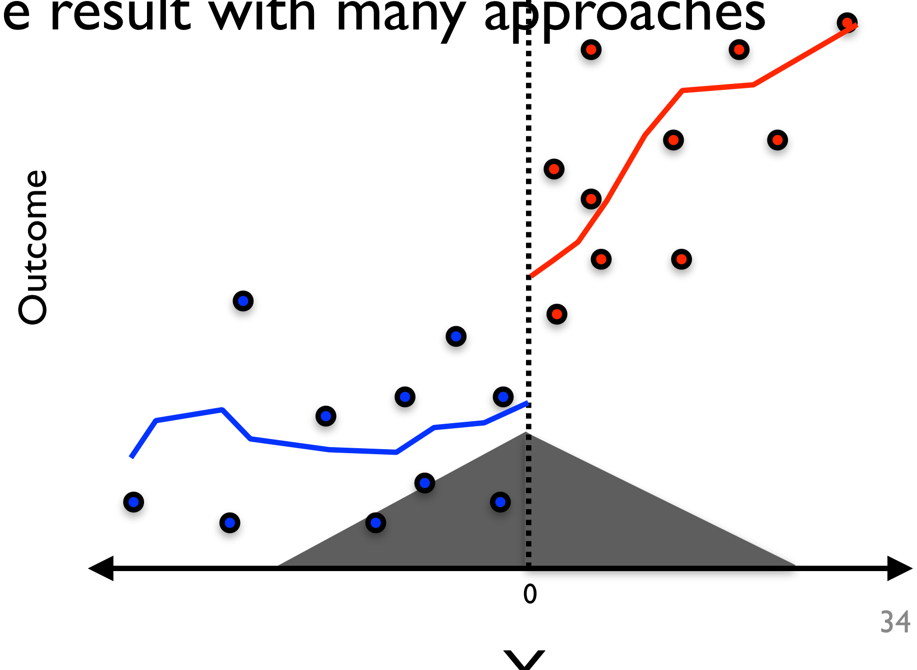
- bandwidth (h) chosen by algorithm that considers curvature of CEF and density of data
- weights that are larger closer to the threshold (triangular kernel)
- (sometimes) polynomial terms (square, cube, etc) in estimating CEF
- bias correction (Calonico, Cattaneo, Titiunik) based on curvature



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



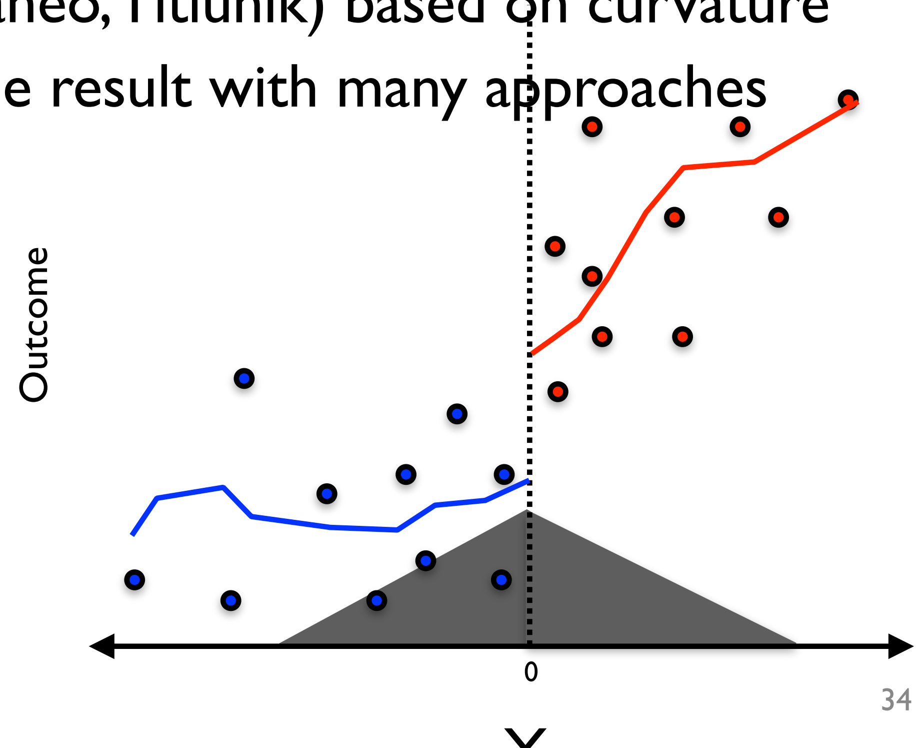
What people do now

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



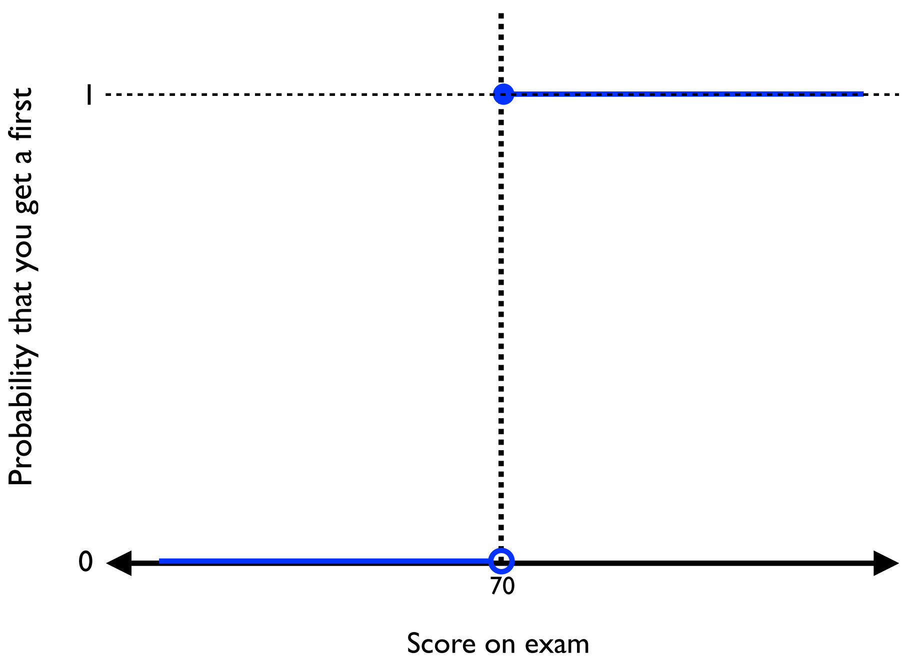
Fuzzy RDD, briefly

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

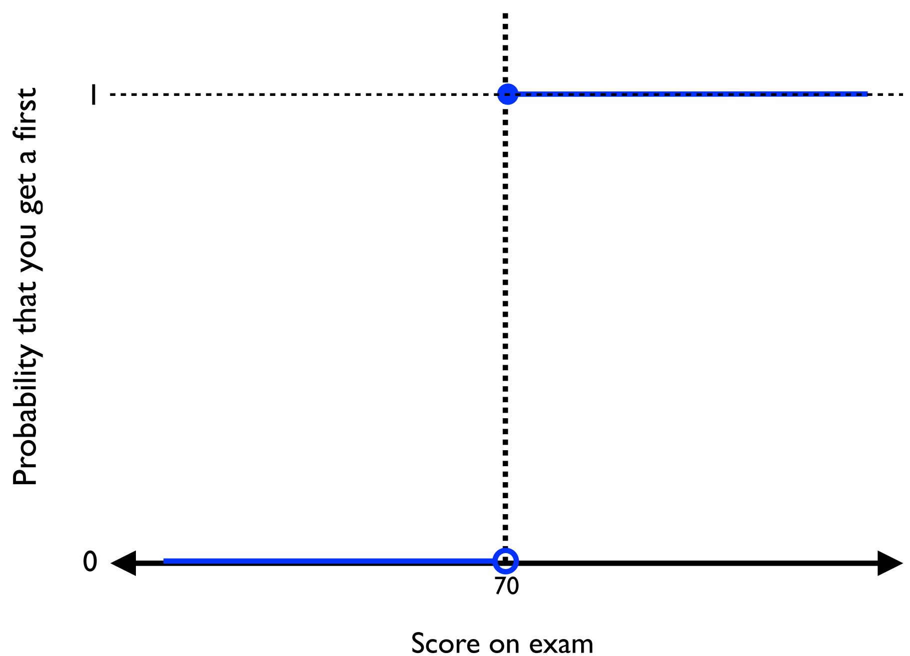
So far, we have only considered cases where the probability of treatment goes from 0 to 1 discontinuously at a threshold.



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.

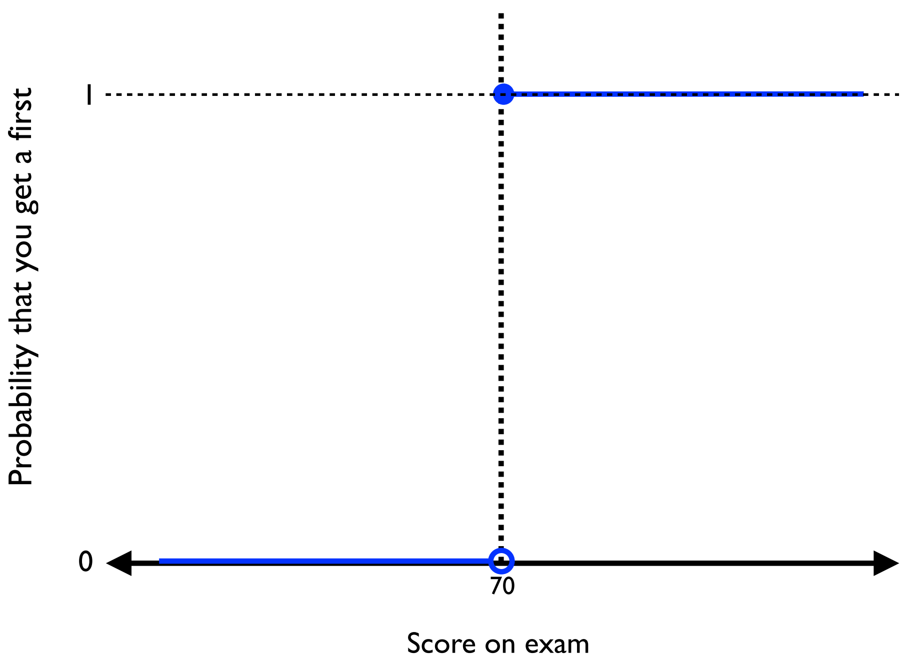


Fuzzy RDD, briefly

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.

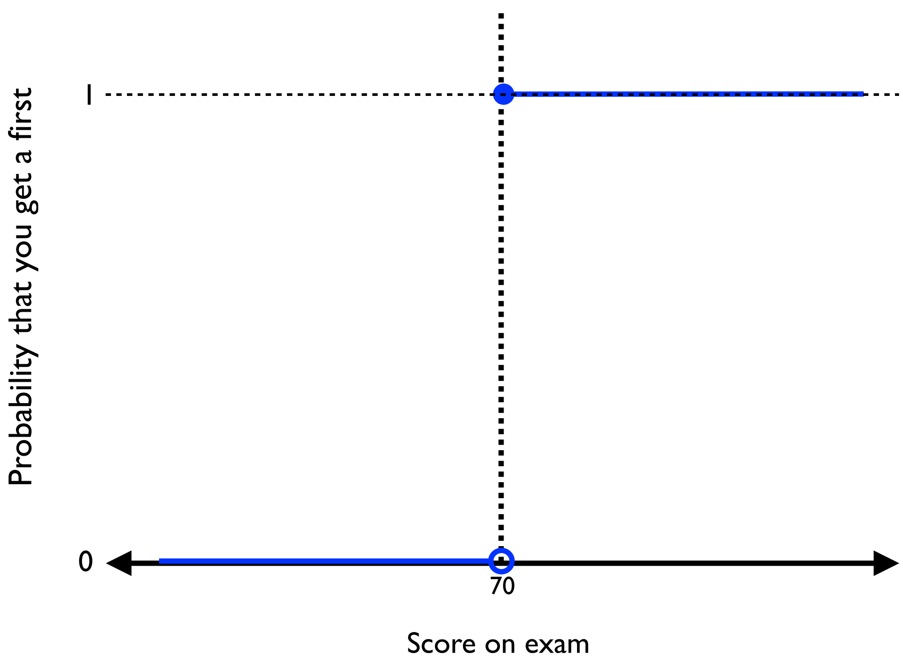


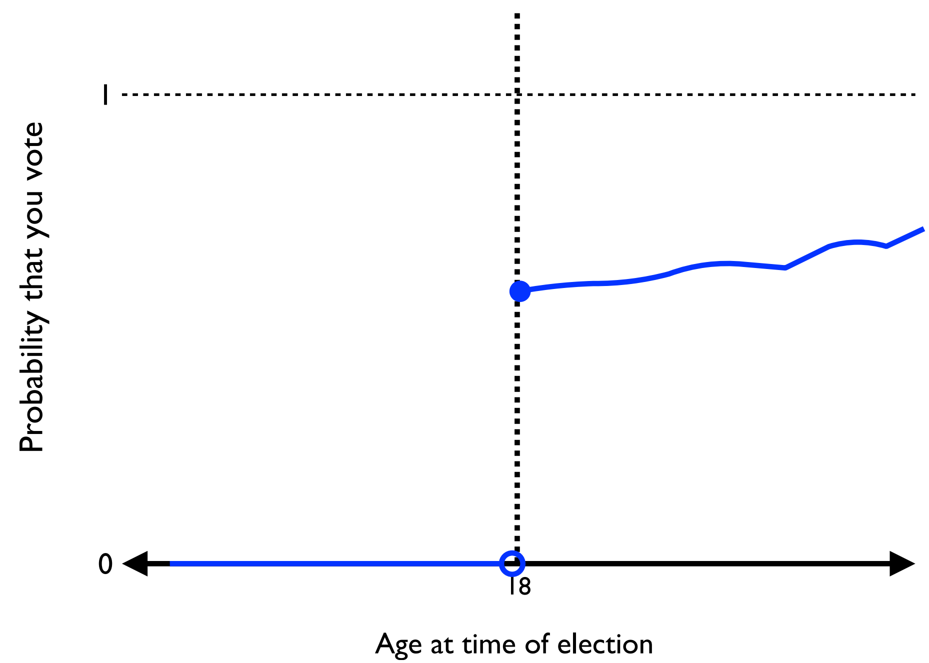
Fuzzy RDD, briefly

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.



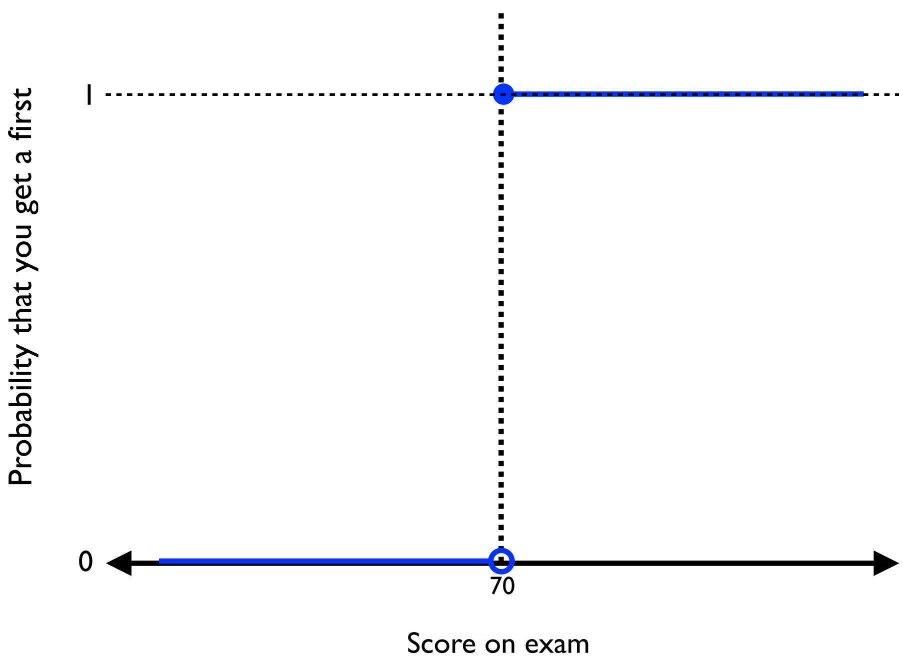


Fuzzy RDD, briefly

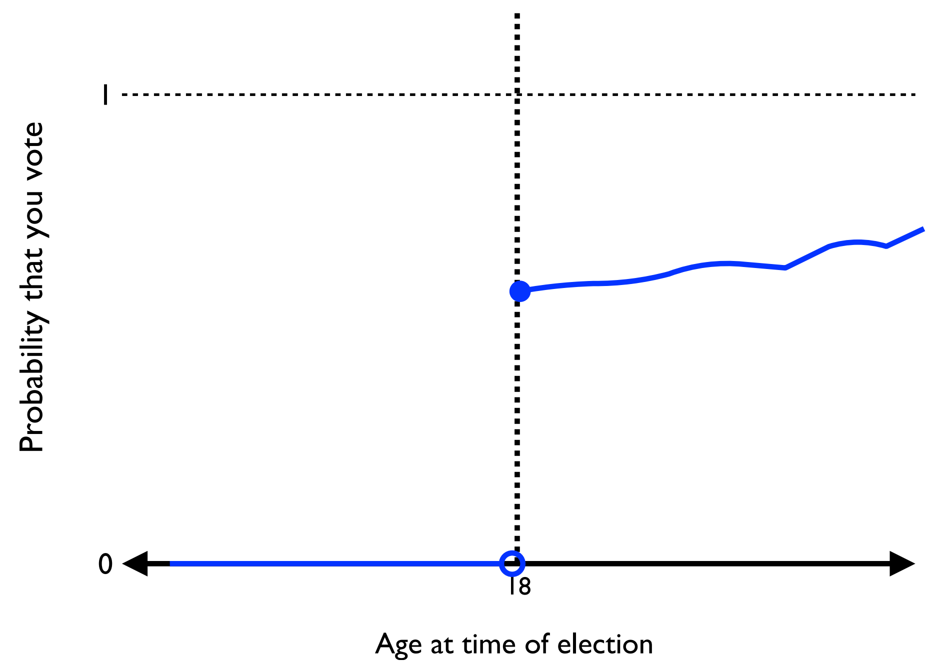
So far, we have only considered cases where the probability of treatment goes from 0 to 1 discontinuously at a threshold.

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

Sharp RDD.



Fuzzy RDD.



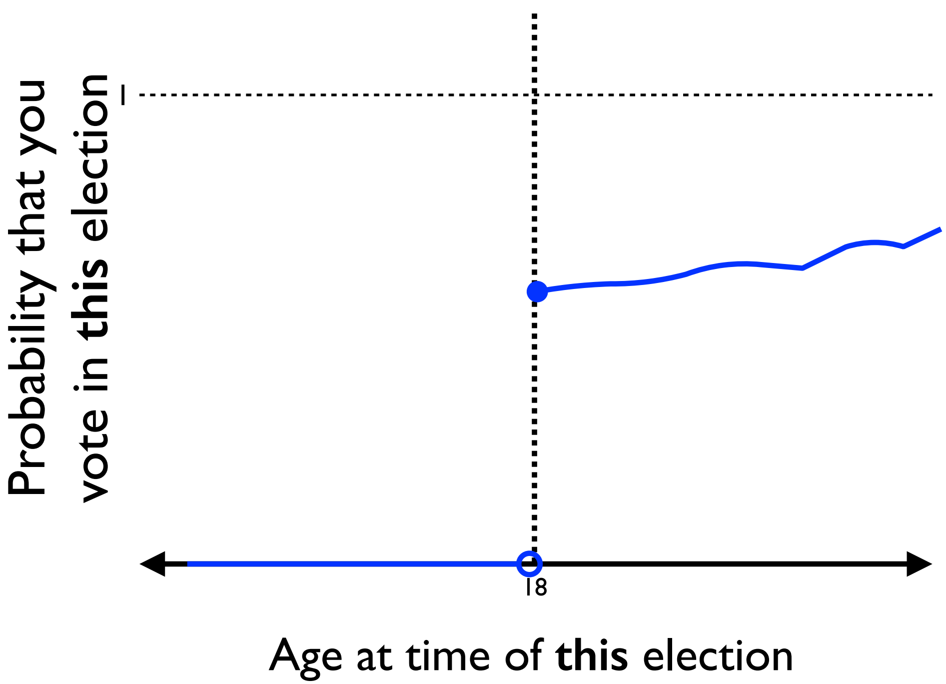
Fuzzy RDD, briefly (2)

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.

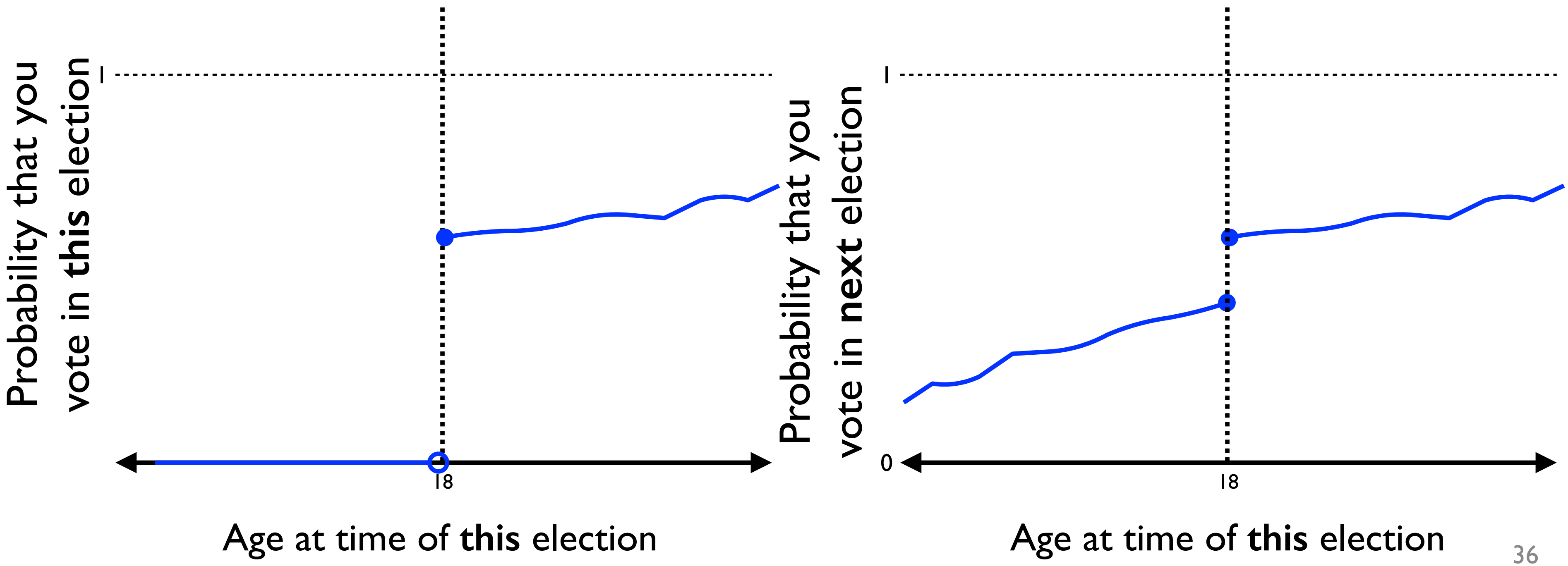
Fuzzy RDD, briefly (2)

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)

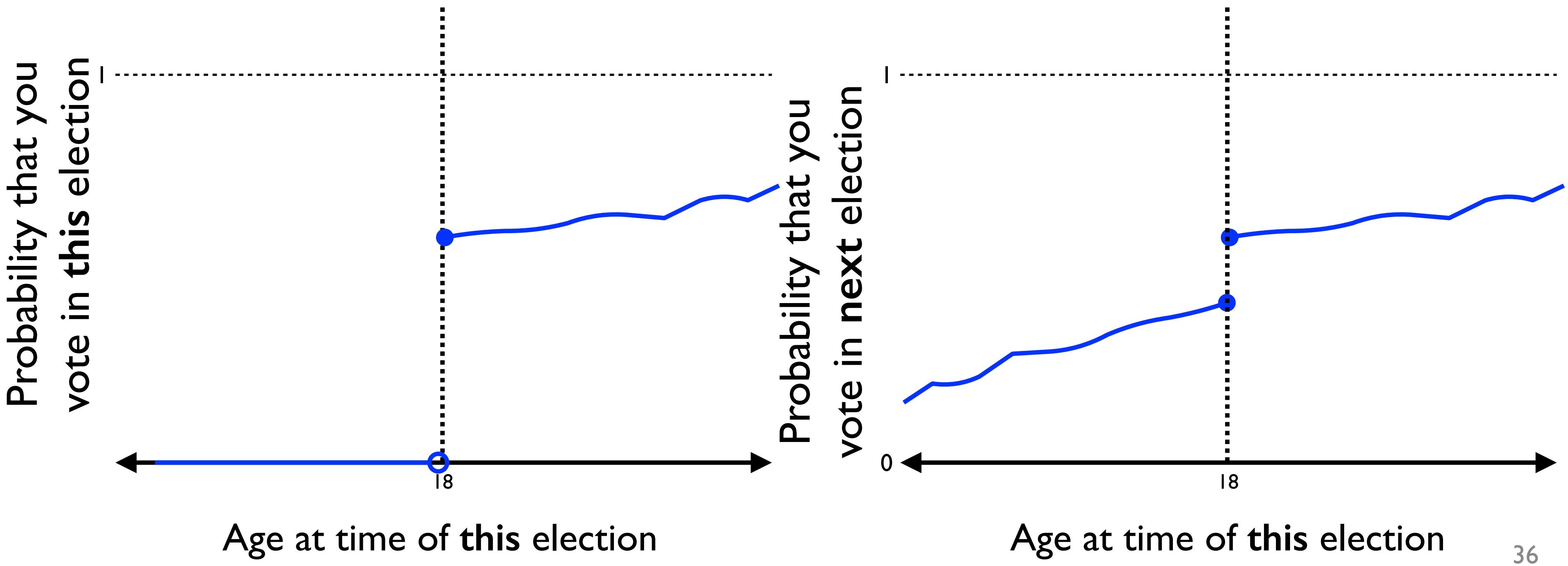
Fuzzy RDD cases can be thought of as having **non-compliance**: the threshold assigns units to treatment, but some units disobey.



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?

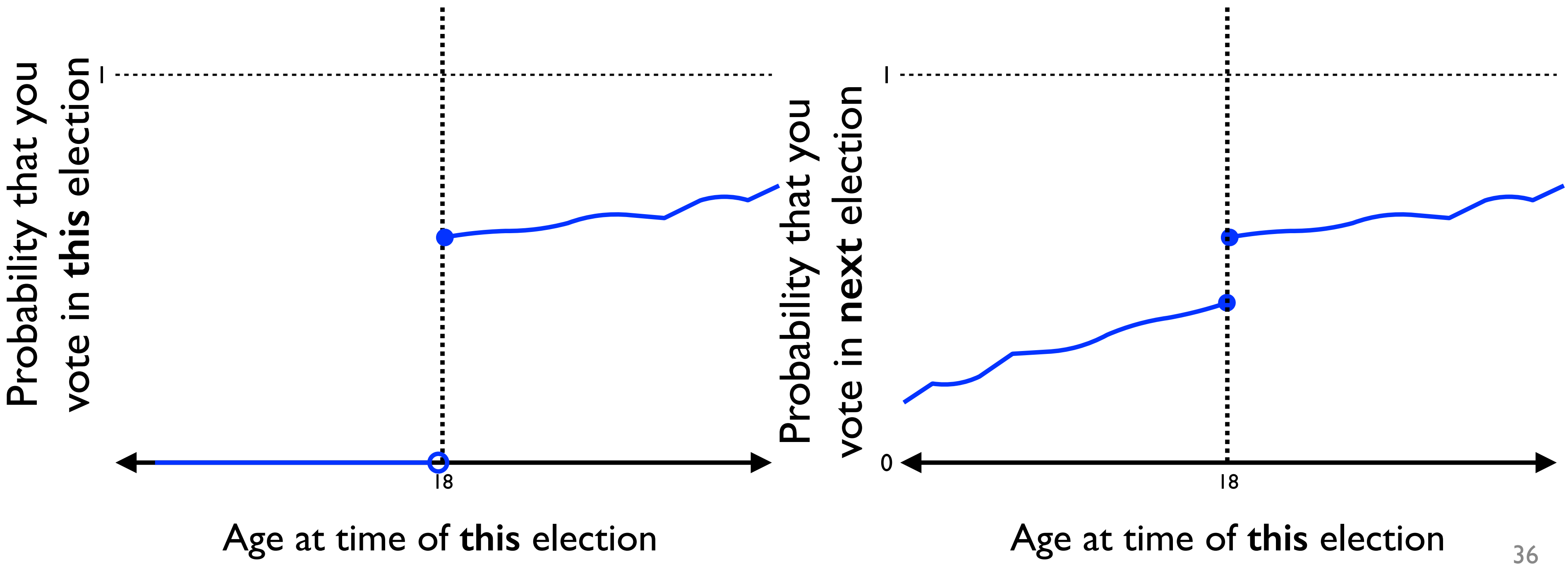


Fuzzy RDD, briefly (2)

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:

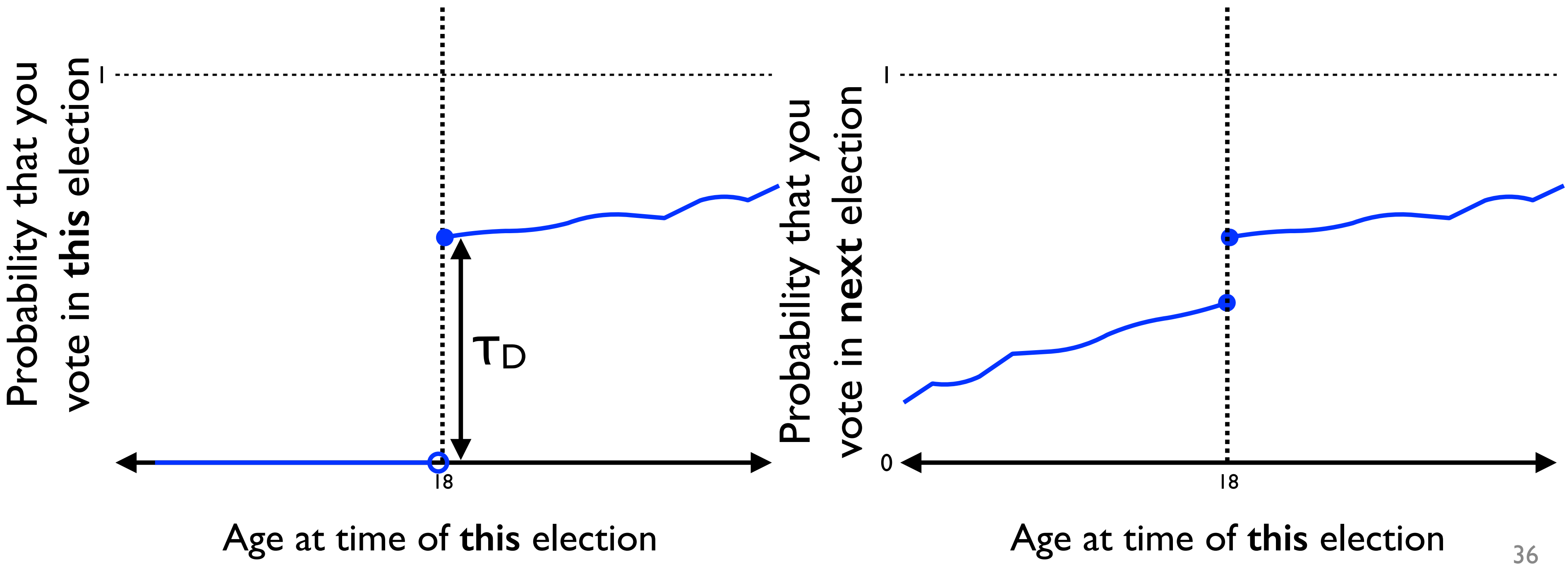


Fuzzy RDD, briefly (2)

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:

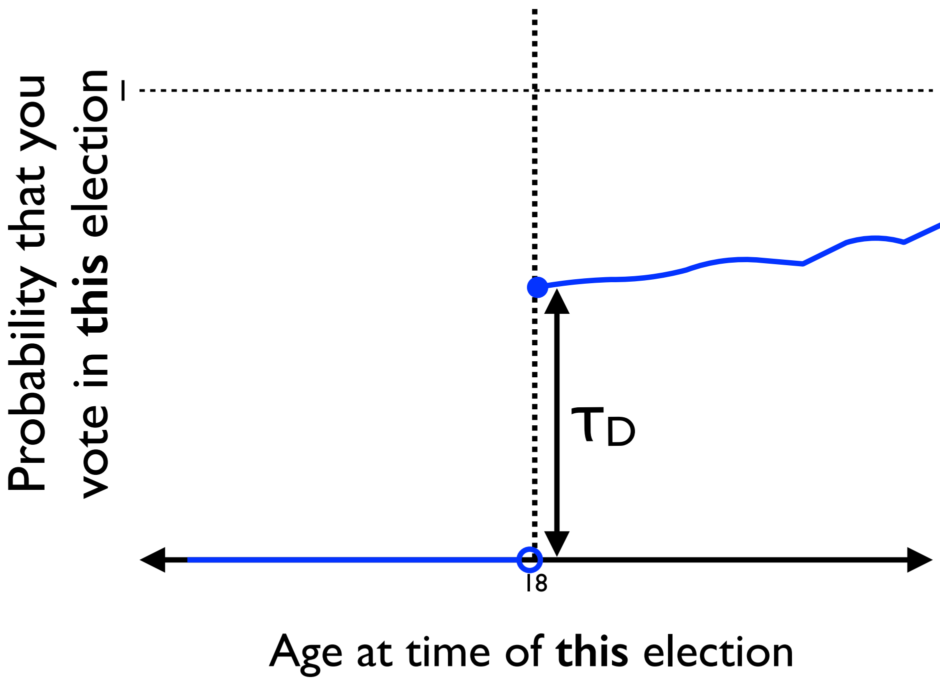


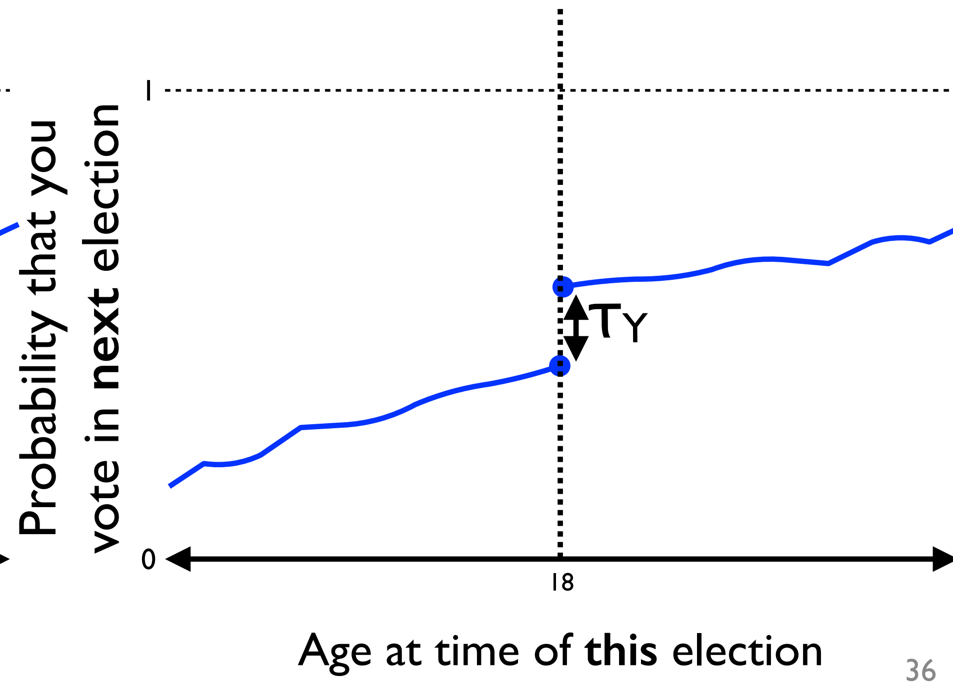
Fuzzy RDD, briefly (2)

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:





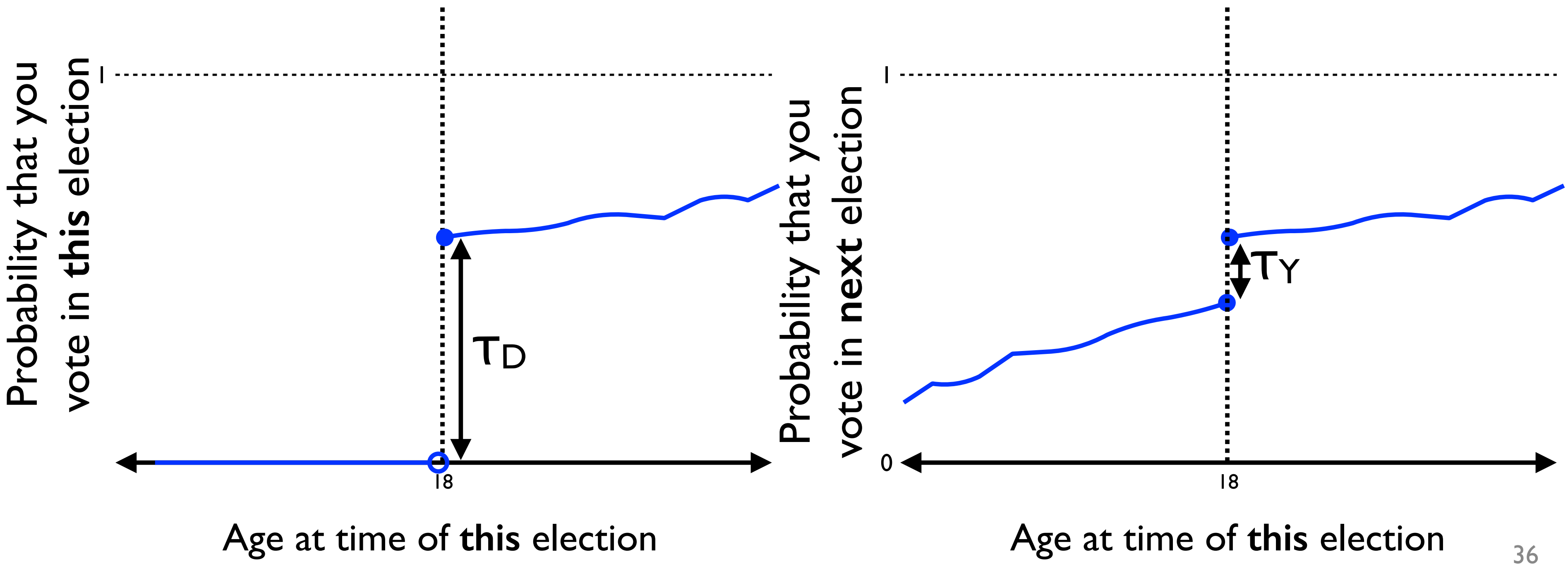
Fuzzy RDD, briefly (2)

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:

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



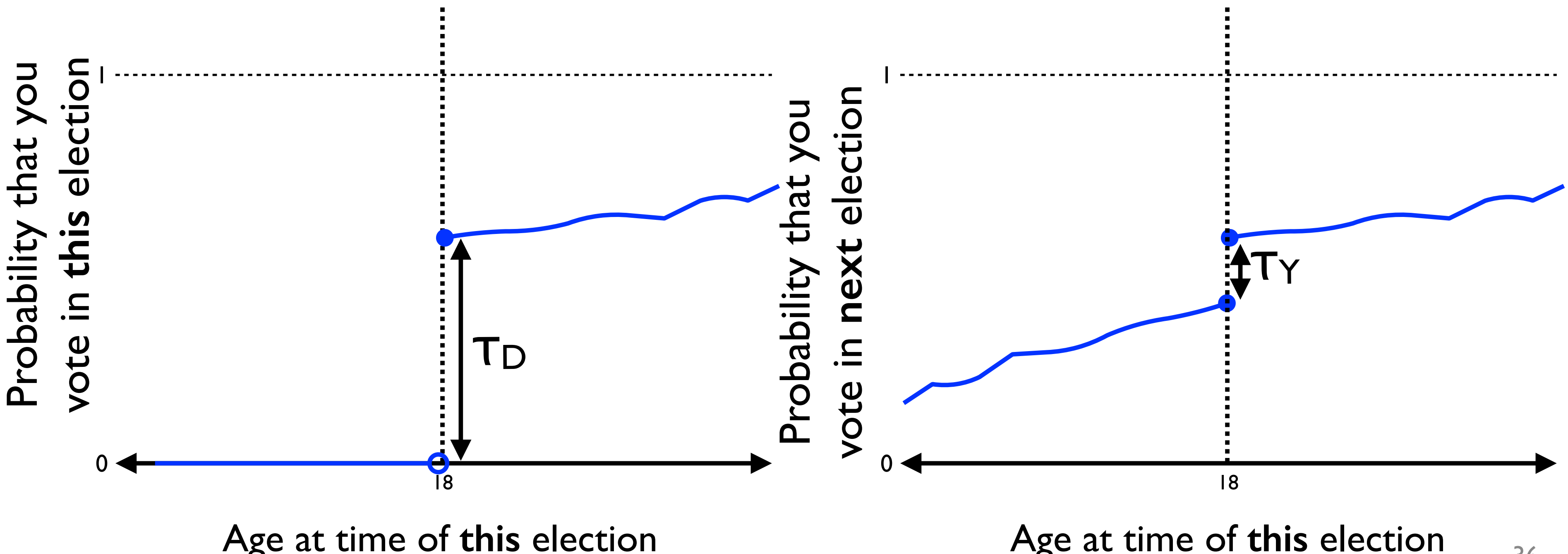
Fuzzy RDD, briefly (2)

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!