150C Causal Inference

Regression Discontinuity Design

Jonathan Mummolo

Regression Discontinuity Design (RDD)

- In the RDD the assignment to treatment is not random, but determined at least partly by the value of an observed covariate lying on either side of a fixed threshold
- Widely applicable in a rule-based world (e.g. programs with fixed eligibility criteria, sharp rules used to allocate resources, rule changes that occur based on thresholds, etc.)
- RDD is a fairly old idea (Thistlethwaite and Campbell 1960), but this design experienced a renaissance in recent years
- High internal validity: Several correspondence tests have shown that RDDs are remarkably effective at replicating results from randomized experiments while other observational study designs are not
  - Buddelmeyer and Skoufias 2004; Cook et al 2008; Berk et al. 2010; Shadish et al. 2011

Outline

Outline

Illustration Example: Sharp RDD

- Thistlethwaite and Campbell (1960) study the effects of college scholarships on later students' achievements
- Scholarships are granted based on whether a student's test score exceeds some threshold c
- Consider the following variables:
  - Binary treatment D is receipt of scholarship
  - Covariate X is SAT score with threshold c
  - Outcome Y is subsequent earnings
  - Y_0 denotes potential earnings without the scholarship
  - Y_1 denotes potential earnings with the scholarship

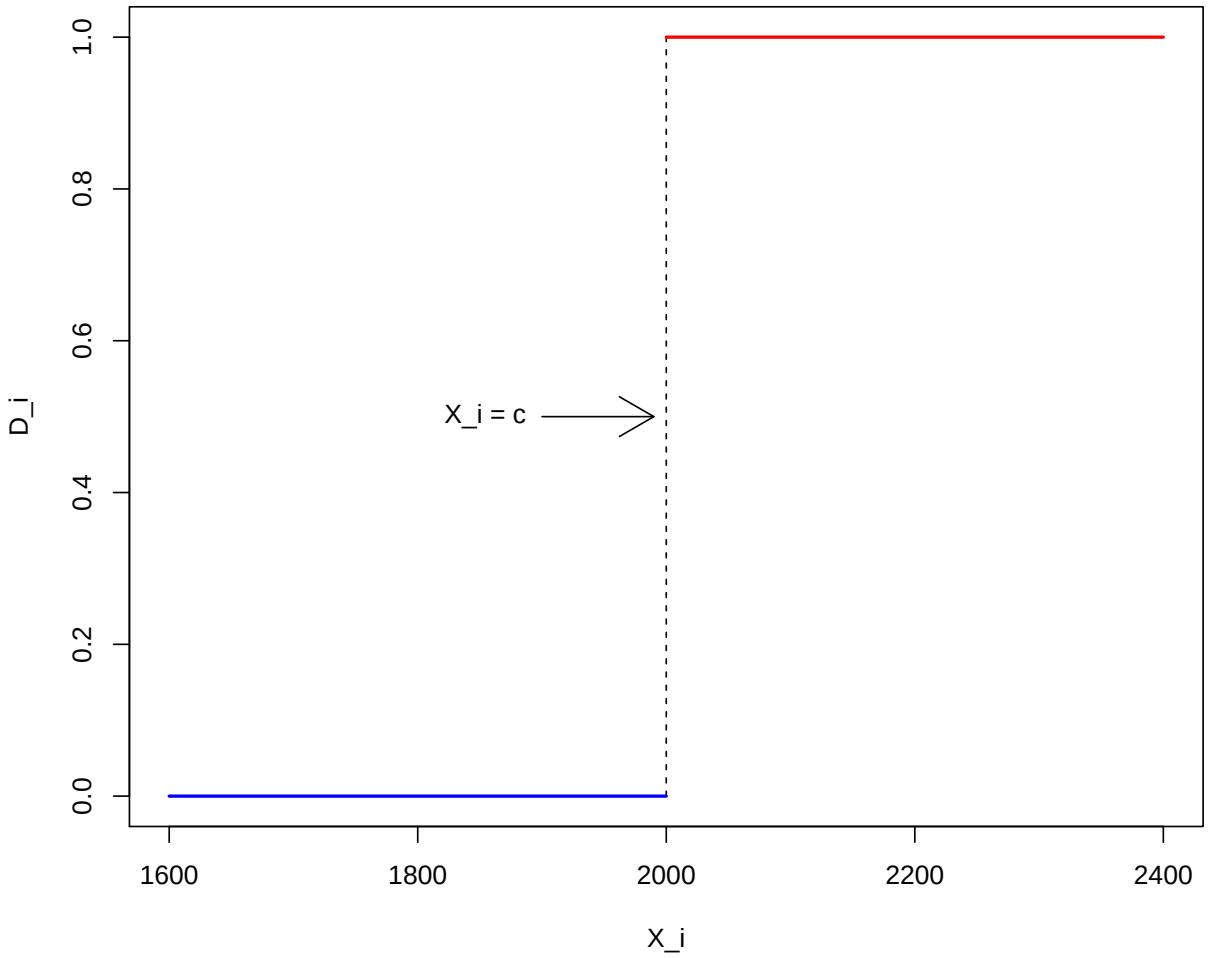
Sharp Regression Discontinuity Design

- Assignment to the scholarship treatment D_i is completely determined by the value of the SAT score X_i being on either side of the threshold c:

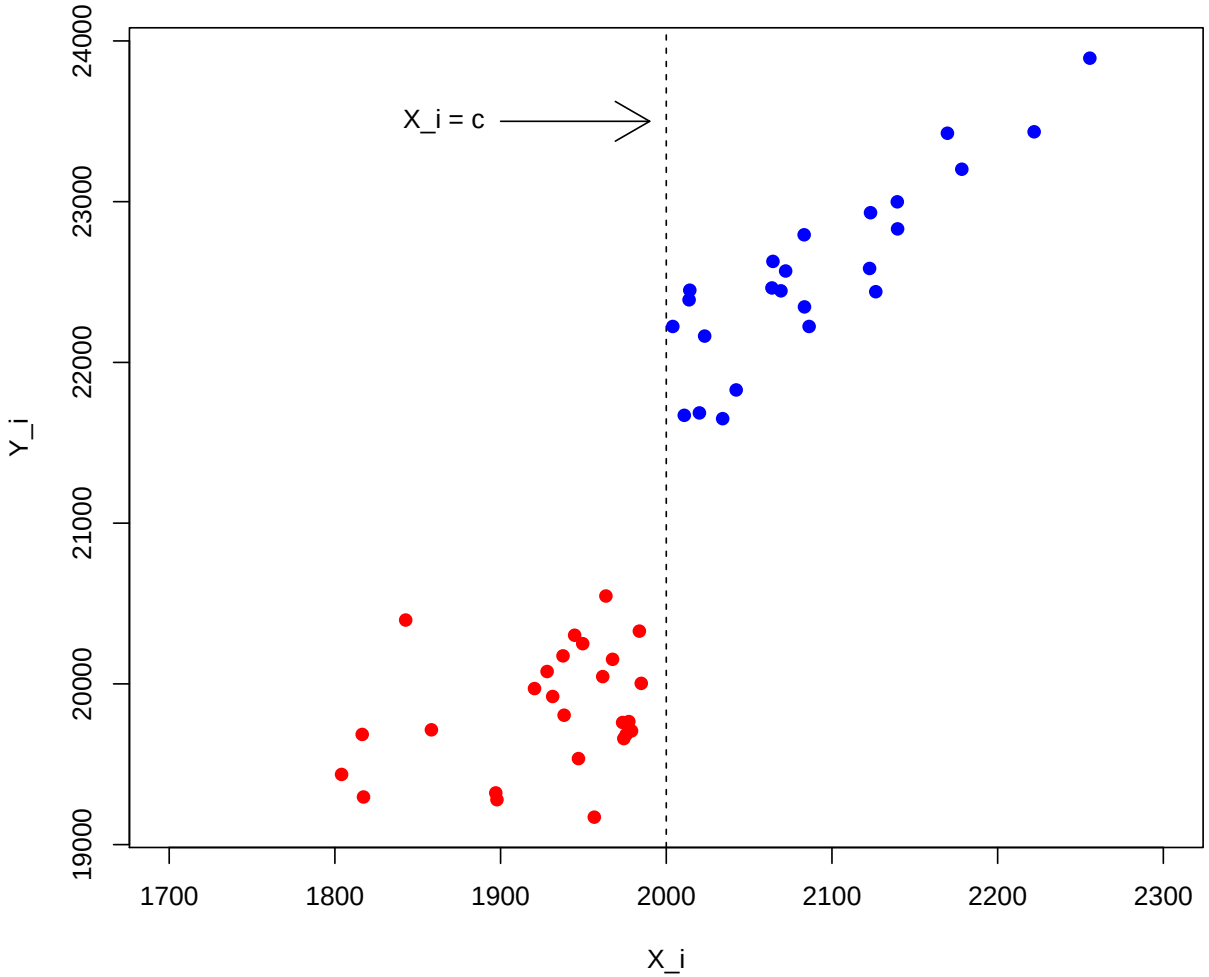
$$D_i = 1\{X_i \geq c\} \text{ so } D_i = \begin{cases} D_i = 1 & \text{if } X_i \geq c \\ D_i = 0 & \text{if } X_i < c \end{cases}$$

- X is called the forcing variable, because it "forces" units from control into treatment once X_i exceeds c
- X may be correlated with Y_1 and Y_0 so comparing treated and untreated units does not provide causal estimates (e.g. students with higher SAT scores obtain higher earnings even without the scholarship)
- If the relationship between X and the potential outcomes Y_1 and Y_0 is "smooth" around the threshold c, we can use the discontinuity created by the treatment to estimate the effect of D on Y at the threshold

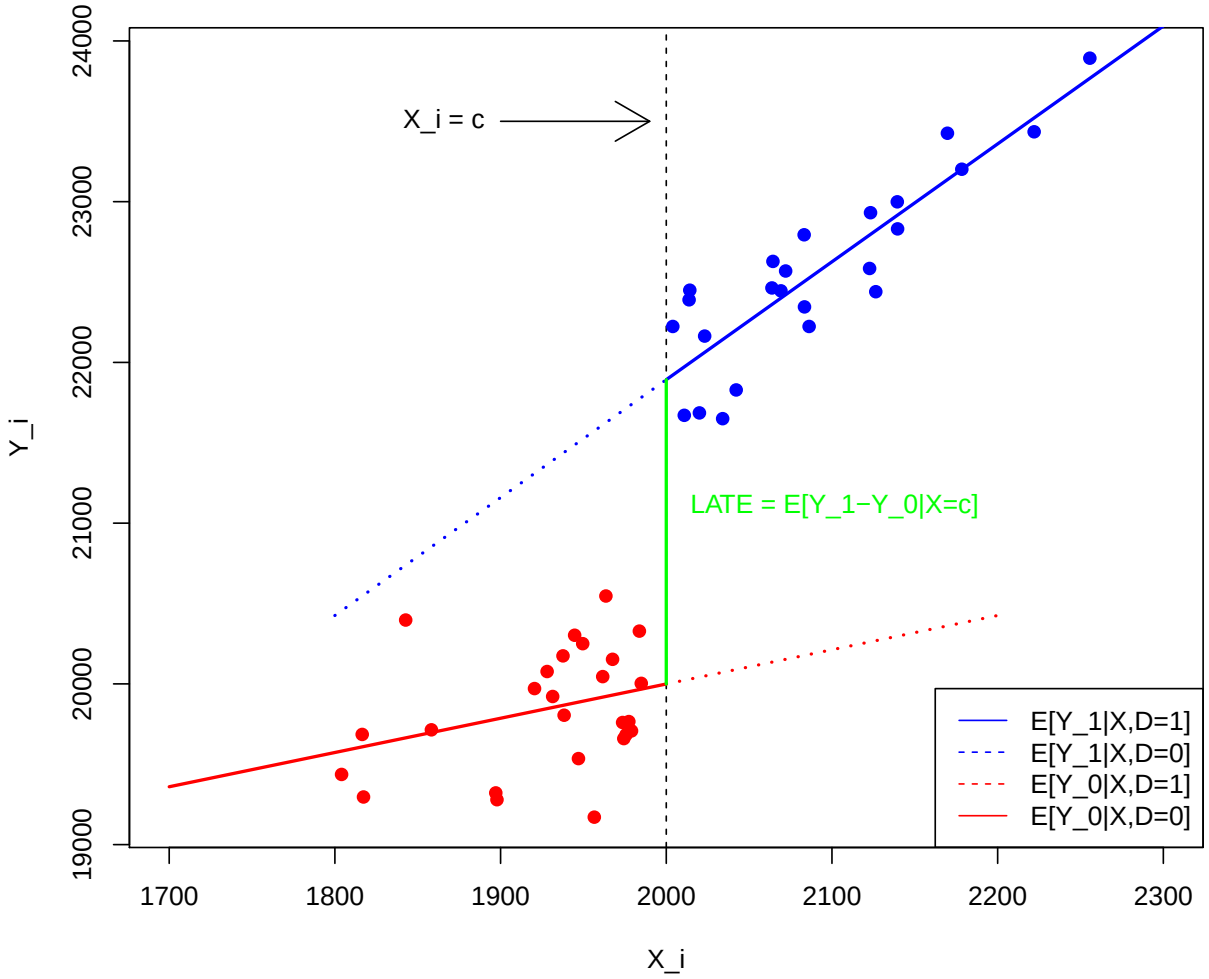
Treatment Assignment



Observed Outcomes



Potential Outcomes



Identification

Identification Assumption

1. $Y_1, Y_0 \perp\!\!\!\perp D|X$ *(trivially met)*
2. $0 < P(D = 1|X = x) < 1$ *(always violated in Sharp RDD)*
3. $E[Y_1|X, D]$ *and* $E[Y_0|X, D]$ *are continuous in* X *around the threshold* $X = c$ *(individuals have imprecise control over* X *around the threshold)*

Identification Result

The treatment effect is identified at the threshold as:

$$\begin{aligned} \alpha_{SRDD} &= E[Y_1 - Y_0|X = c] \\ &= E[Y_1|X = c] - E[Y_0|X = c] \\ &= \lim_{x \downarrow c} E[Y_1|X = c] - \lim_{x \uparrow c} E[Y_0|X = c] \end{aligned}$$

Without further assumptions α_{SRDD} *is only identified at the threshold.*

Outline

Party Incumbency Advantage

- What is the effect of incumbency status on vote shares? (Lee 2006)
- Let i indicate congressional districts, j indicate parties, and t indicate elections, d indicate incumbency status
- V_{ditj} is the vote share of j in i at t as incumbent $d = 1$ or non-incumbent $d = 0$
- Party Incumbency Effect: $V_{1itj} - V_{0itj}$
- Forcing variable: Margin of Victory for party j:

$$Z_{itj} = V_{itj} - V_{itk}$$

where k indicates the strongest opposition party.

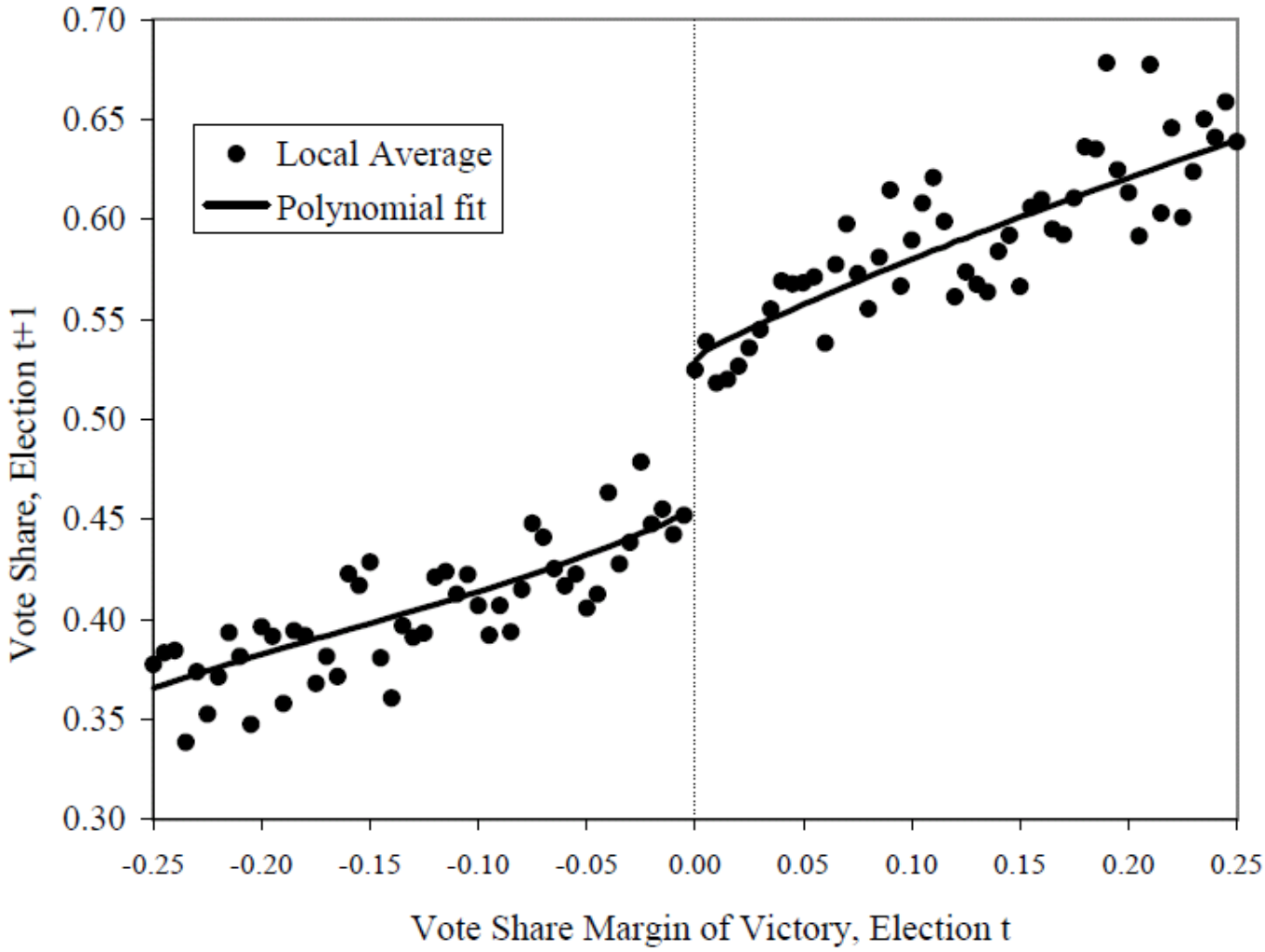
- Party Incumbency status is then assigned as:

$$D_{ij,t+1} = 1\{Z_{itj} > 0\} \text{ so } D_i = \begin{cases} D_{ij,t+1} = 1 & \text{if } Z_{itj} > 0 \\ D_{ij,t+1} = 0 & \text{if } Z_{itj} < 0 \end{cases}$$

- With only two parties we can also use $Z = V - c$ with $c = .5$

Incumbency Advantage

Figure IVa: Democrat Party's Vote Share in Election t+1, by Margin of Victory in Election t: local averages and parametric fit



Electronic Voting

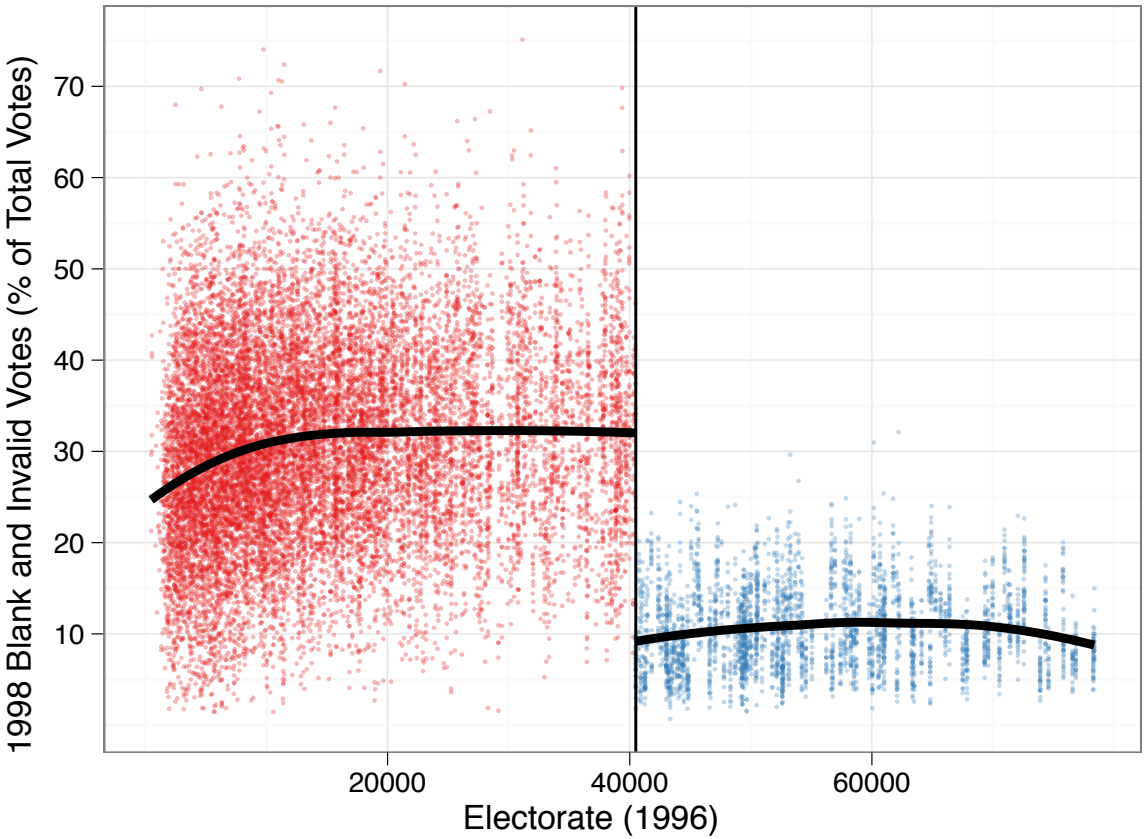


Figure 6: The effect of electronic voting on the percent of null and blank votes. Each dot is a polling station. Polling stations to the left of the vertical black line used paper ballots and polling stations to the right used electronic voting. The black horizontal line is the conditional mean of the outcome estimated with a loess regression.

Hidalgo 2012

Other Recent RDD Examples

- class size on student achievement
  - Angrist and Lavy 1999
- wage increase on performance of mayors

  Ferraz and Finan 2011; Gagliarducci and Nannicini 2013
- colonial institutions on development outcomes

  Dell 2009
- length of postpartum hospital stays on mother and infant mortality

  Almond and Doyle 2009
- naturalization on political integration of immigrants

  Hainmueller and Hangartner 2015
- financial aid offers on college enrollment

  Van der Klaauw 2002
- access to Angel funding on growth of start-ups

  Kerr, Lerner and Schoar 2010
- RDD that exploits "close" elections is workhorse model for electoral research:

  Lee, Moretti and Butler 2004, DiNardo and Lee 2004, Hainmueller and Kern 2008, Leigh 2008, Pettersson-Lidbom 2008, Broockman 2009, Butler 2009, Dal Bó, Dal Bó and Snyder 2009, Eggers and Hainmueller 2009, Ferreira and Gyourko 2009, Uppal 2009, 2010, Cellini, Ferreira and Rothstein 2010, Gerber and Hopkins 2011, Trounstine 2011, Boas and Hidalgo 2011, Folke and Snyder Jr. 2012, and Gagliarducci and Paserman 2012

Outline

Estimate $\alpha_{SRDD} = E[Y_1|X = c] - E[Y_0|X = c]$

1. Trim the sample to a reasonable window around the cutpoint c (discontinuity sample):
   - $c - h \leq X_i \leq c + h$, were h is some positive value that determines the size of the window
   - h may be determined by cross-validation
2. Code the margin $\tilde{X}$ which measures the distance to the threshold:

$$\tilde{X} = X - c \text{ so } \tilde{X}_i = \begin{cases} \tilde{X} = 0 & \text{if } X = c \\ \tilde{X} > 0 & \text{if } X > c \text{ and thus D=1} \\ \tilde{X} < 0 & \text{if } X < c \text{ and thus D=0} \end{cases}$$

3. Decide on a model for $E[Y|X]$
   - linear, same slope for $E[Y_0|X]$ and $E[Y_1|X]$
   - linear, different slopes
   - non-linear
   - always start with an visual inspection to check which model is appropriate (e.g. scatter plot with kernel/lowess)

Linear with Same Slope

- $E[Y_0|X]$ is linear and treatment effect, α, does not depend on X:

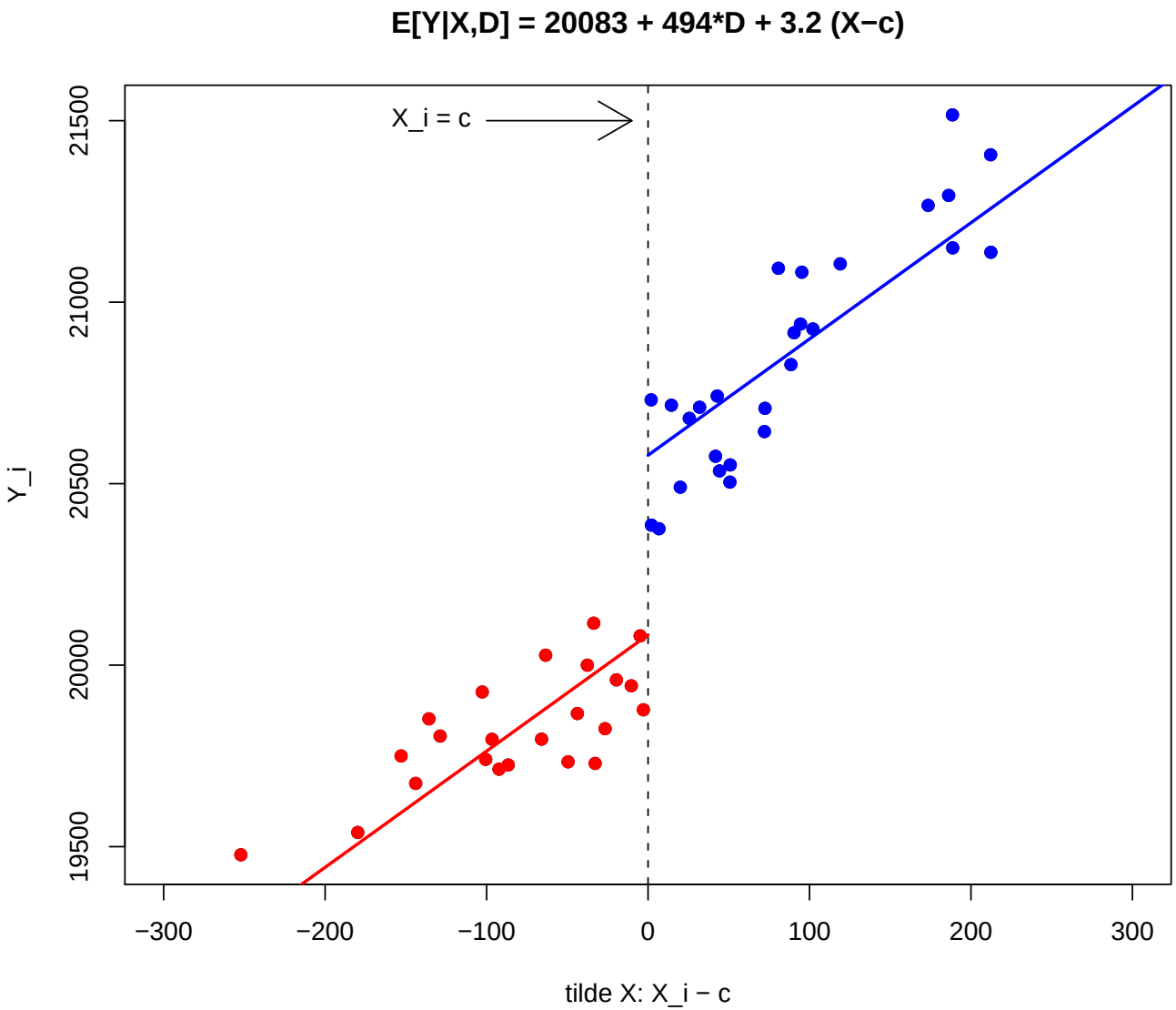
$$E[Y_0|X] = \mu + \beta X\,, \qquad E[Y_1 - Y_0|X] = \alpha$$

- Therefore $E[Y_1|X] = \alpha + E[Y_0|X] = \alpha + \mu + \beta X$
- Since D is determined given X, we have that:

$$\begin{aligned} E[Y|X,D] &= D \cdot E[Y_1|X] + (1-D) \cdot E[Y_0|X] \\ &= \mu + \alpha D + \beta X \\ &= (\mu - \beta c) + \alpha D + \beta(X - c) \\ &= \gamma + \alpha D + \beta \tilde{X} \end{aligned}$$

- So we just run a regression of Y on D and the margin $\tilde{X} = X - c$.

Linear with Same Slope



Linear with Different Slopes

- $E[Y_0|X]$ and $E[Y_1|X]$ are distinct linear functions of X, so the average effect of the treatment $E[Y_1 - Y_0|X]$ varies with X:

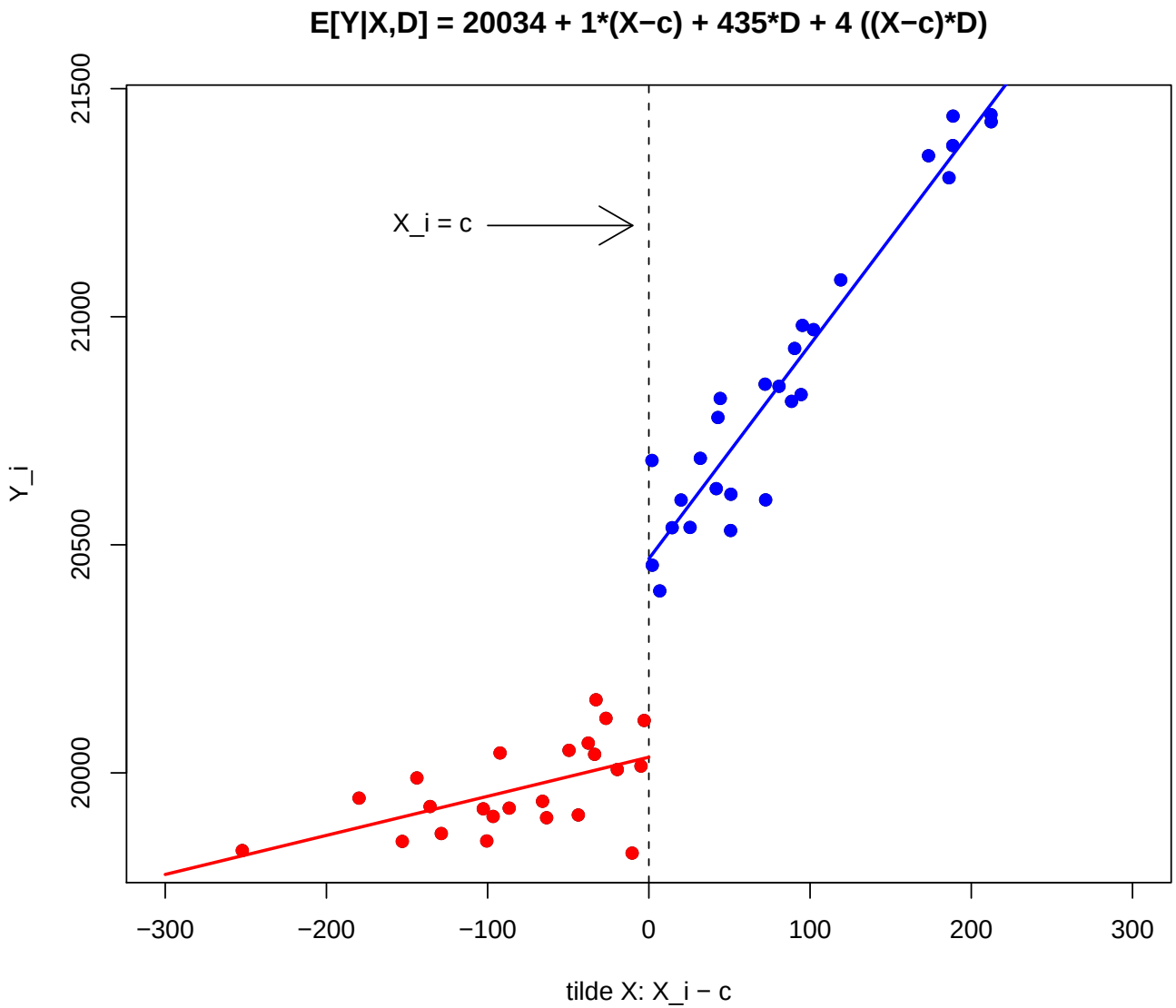
$$E[Y_0|X] = \mu_0 + \beta_0 X, \qquad E[Y_1|X] = \mu_1 + \beta_1 X$$

- So $\alpha(X) = E[Y_1 - Y_0|X] = (\mu_1 - \mu_0) + (\beta_1 - \beta_0)X$ we have

$$\begin{aligned} E[Y|X,D] &= D \cdot E[Y_1|X] + (1-D) \cdot E[Y_0|X] \\ &= \mu_1 D + \beta_1 (X \cdot D) + \mu_0 (1-D) + \beta_0 (X \cdot (1-D)) \\ &= \gamma + \beta_0 (X - c) + \alpha D + \beta_1 ((X-c) \cdot D) \\ &= \gamma + \beta_0 \tilde{X} + \alpha D + \beta_1 (\tilde{X} \cdot D) \end{aligned}$$

- Regress Y on the margin $\tilde{X}$, treatment D, and the interaction $\tilde{X} \cdot D$; the coefficient α on D identifies the local average treatment effect at $X = c$

Linear with Different Slope



Non-Linear Case

- $E[Y_0|X]$ and $E[Y_1|X]$ are distinct non-linear functions of X and the average effect of the treatment $E[Y_1 - Y_0|X]$ varies with X
- Include quadratic and cubic terms in $\tilde{X}$ and their interactions with D in the equation
- The specification with quadratic terms is

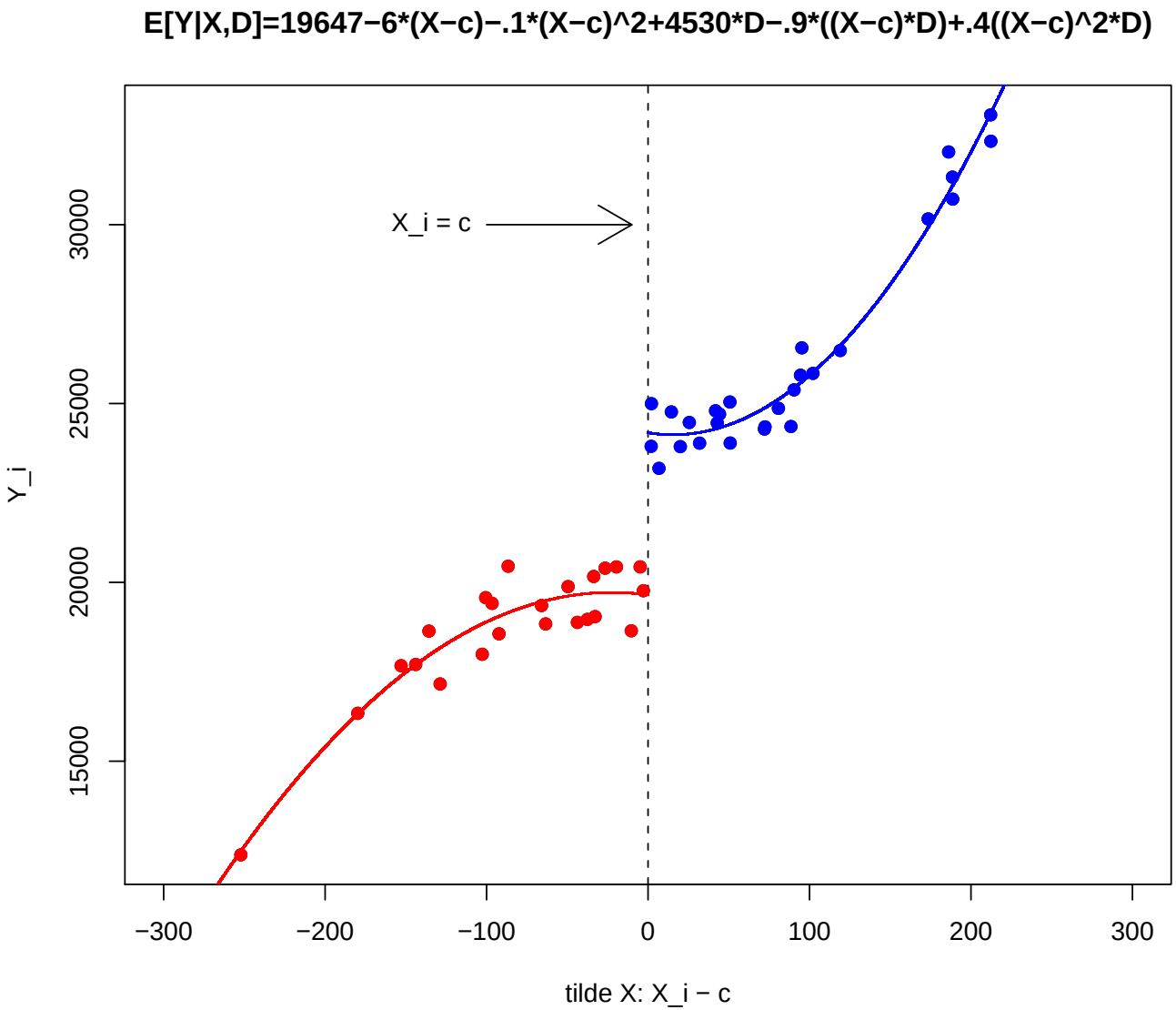
$$E[Y|X, D] = \gamma_0 + \gamma_1\tilde{X} + \gamma_2\tilde{X})^2 + \alpha_0 D + \alpha_1(\tilde{X} \cdot D) + \alpha_2(\tilde{X}^2 \cdot D)$$

  The specification with cubic terms is

$$E[Y|X, D] = \gamma_0 + \gamma_1\tilde{X} + \gamma_2\tilde{X}^2 + \gamma_3\tilde{X}^3 + \alpha_0 D + \alpha_1(\tilde{X} \cdot D) + \alpha_2(\tilde{X}^2 \cdot D) + \alpha_3(\tilde{X}^3 \cdot D)$$

- In both cases $\alpha_0 = E[Y_1 - Y_0|X = c]$

Non-Linear Case

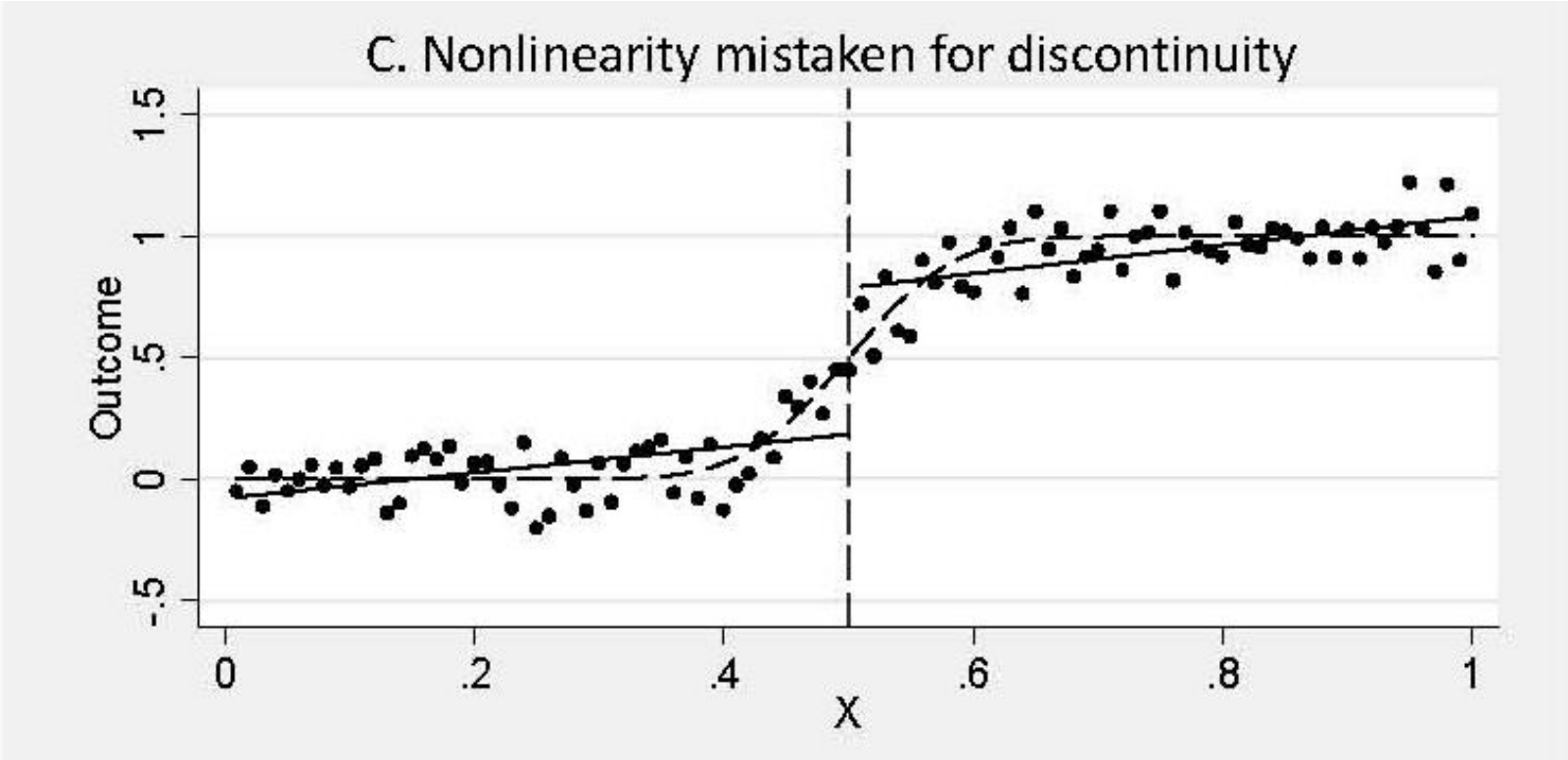


Outline

Falsification Checks

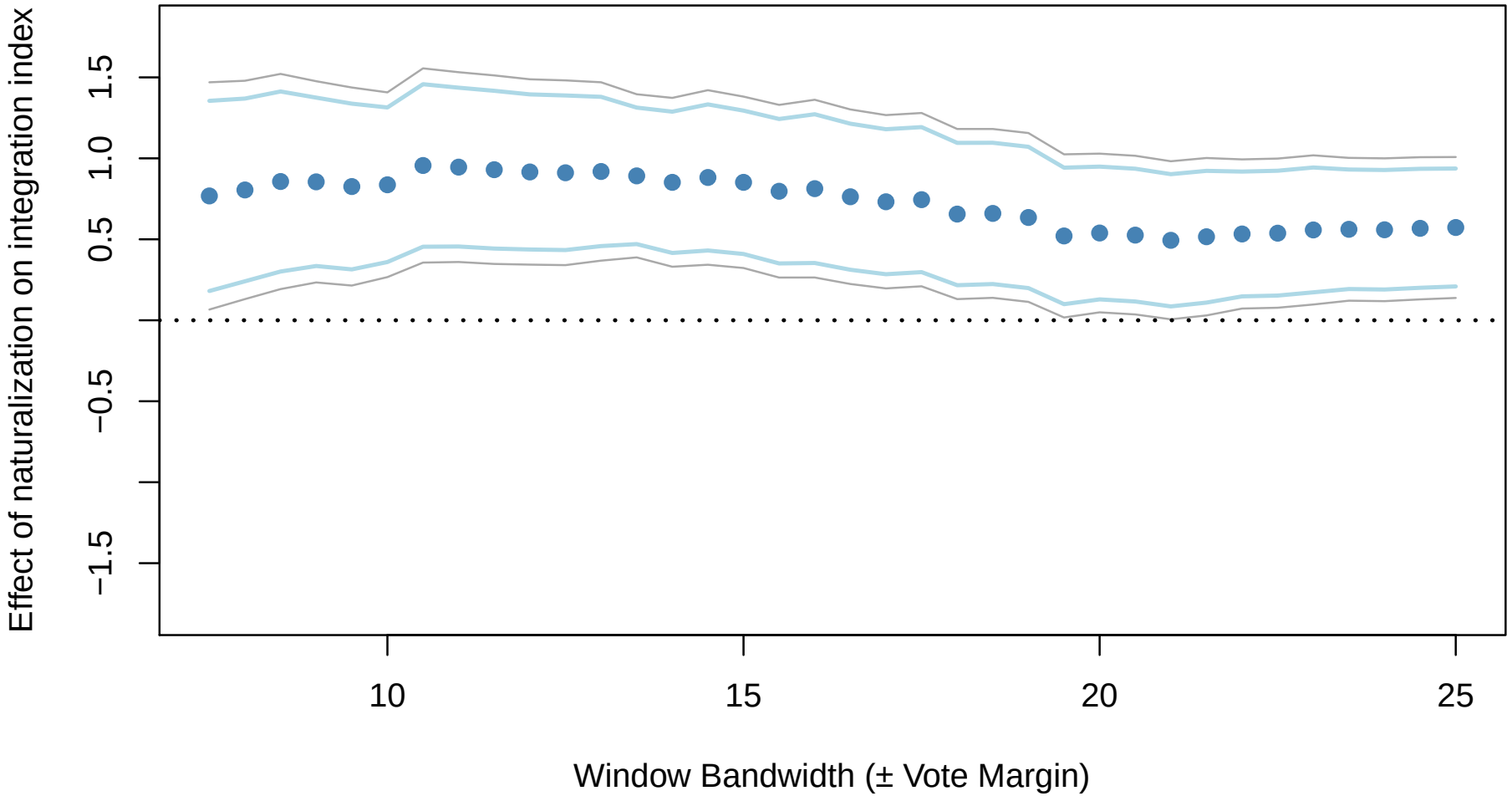
1. Sensitivity: Are results sensitive to alternative specifications?
2. Balance Checks: Do covariates jump at the threshold?
3. Check if jumps occur at placebo thresholds $c^{\star}$?
4. Sorting: Do units sort around the threshold?

Sensitivity to Specification



- $Y = f(X) + \alpha D + \varepsilon$: A miss-specified control function $f(X)$ can lead to a spurious jump: Do not confuse a nonlinear relation with a discontinuity
- More flexibility can reduce bias, but might decrease efficiency
- Check sensitivity to size of bandwidth (i.e. estimation window)

Sensitivity to Bandwidth



Hainmueller and Hangartner 2015

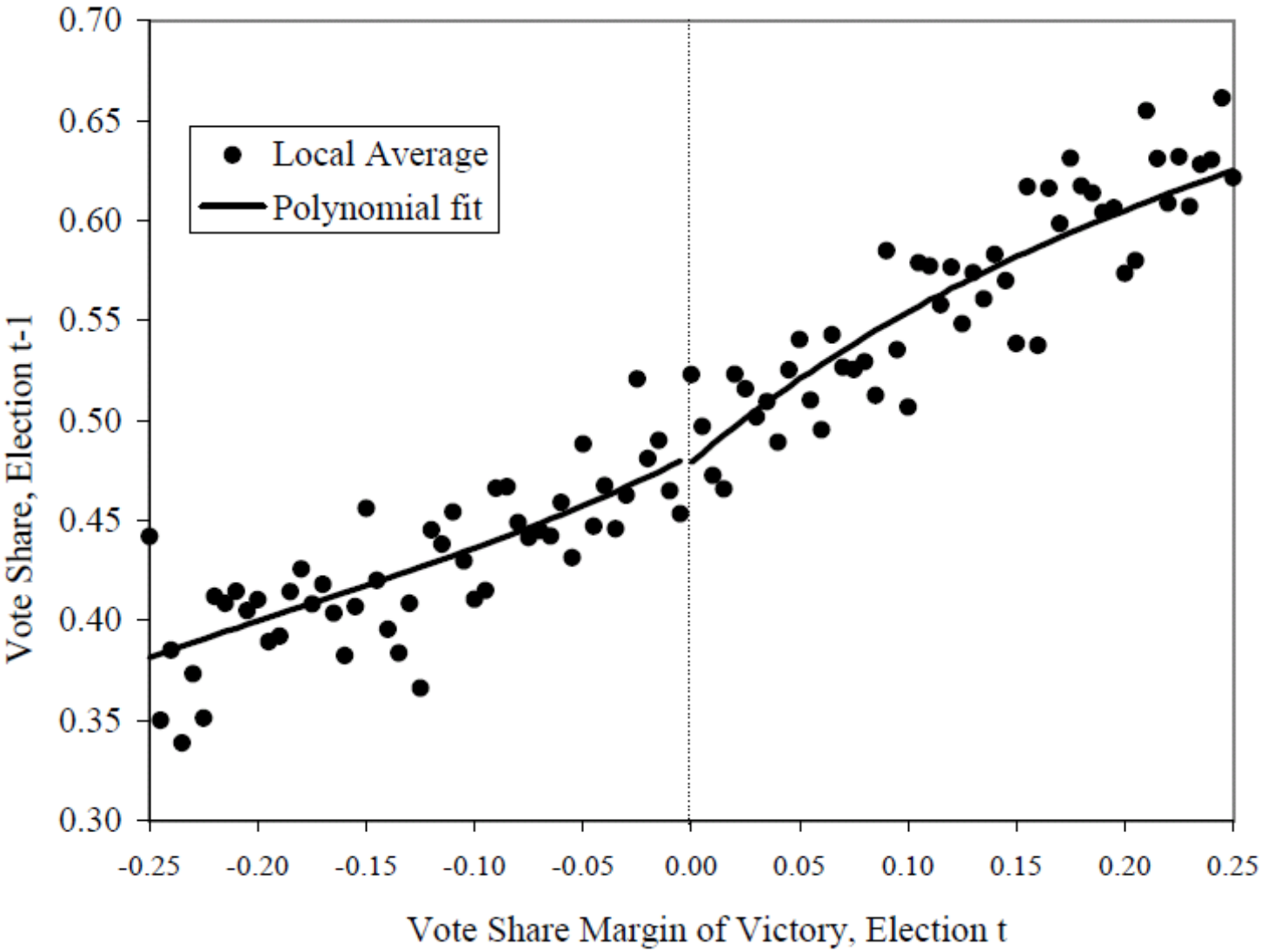
Balance Checks Test

- Test for comparability of units around the cut-off:
  - Consider a pre-treatment covariate M
  - Visual tests: Plot $E[M|X, D]$ and look for jumps, ideally the relation between covariates and treatment should be smooth around threshold
  - Run the RDD regression using M as the outcome:
    $E[M|X, D] = \beta_0 + \beta_1 \tilde{X} + \alpha_z D + \beta_3 (\tilde{X} \cdot D)$
    ideally should yield $\alpha_z = 0$ if M is balanced at the threshold

- An occasional discontinuity in $E[M|X, D]$ does not necessarily invalidate the RDD
  - Multiple testing problem
  - Can incorporate M as additional control into the RDD regression. Ideally, this should only impact statistical significance, not magnitude of treatment effect

- Balance checks address only observables, not unobservables

Falsification Test



Lee 2006

Falsification Test

TABLE 6. Effect of Serving on Placebo Outcomes

	Conservative Party			Labour Party		
Placebo Outcome	Placebo Effect	95. UB	95 LB	Placebo Effect	95. UB	95 LB
Year of birth	2.79	8.10	−2.62	2.50	8.62	−3.77
Year of death	2.08	5.97	−1.89	2.23	6.23	−1.91
Age at death	0.12	−6.32	6.56	1.41	−5.78	8.60
Female	−0.01	0.14	−0.16	−0.03	0.06	−0.12
Teacher	−0.09	0.06	−0.23	−0.23	0.01	−0.47
Barrister	0.09	0.25	−0.09	−0.07	0.05	−0.18
Solicitor	−0.13	0.07	−0.33	0.03	0.15	−0.10
Doctor	−0.00	0.12	−0.13	0.03	0.14	−0.09
Civil servant	0.04	0.10	−0.02	−0.03	0.03	−0.10
Local politician	−0.01	0.23	−0.25	0.10	0.40	−0.21
Business	−0.05	0.21	−0.31	0.00	0.13	−0.13
White collar	−0.00	0.19	−0.19	−0.00	0.15	−0.16
Union official	0.00	NA	NA	−0.04	0.12	−0.20
Journalist	−0.08	0.07	−0.22	0.05	0.29	−0.20
Miner	0.00	NA	NA	−0.02	0.02	−0.07
Schooling: Eton	0.12	0.28	−0.04	−0.04	0.02	−0.11
Schooling: public	−0.22	0.07	−0.52	0.03	0.23	−0.17
Schooling: regular	−0.15	0.12	−0.42	−0.01	0.32	−0.35
Schooling: not reported	0.25	0.46	0.03	0.02	0.33	−0.30
University: Oxbridge	0.10	0.36	−0.17	−0.04	0.21	−0.30
University: degree	−0.02	0.25	−0.30	0.10	0.42	−0.23
University: not reported	−0.08	0.21	−0.37	−0.06	0.25	−0.37
Aristocrat	0.05	0.19	−0.09	0.06	0.17	−0.06
Previous races	0.22	0.59	−0.16	0.24	0.76	−0.29
Vote margin in previous race	−0.00	0.04	−0.05	−0.05	0.01	−0.11
Size of electorate	−622	−8056	6812	−545	−7488	6397
Turnout	−0.01	−0.04	0.03	0.02	−0.02	0.05

Eggers and Hainmueller 2009

Adding Covariates

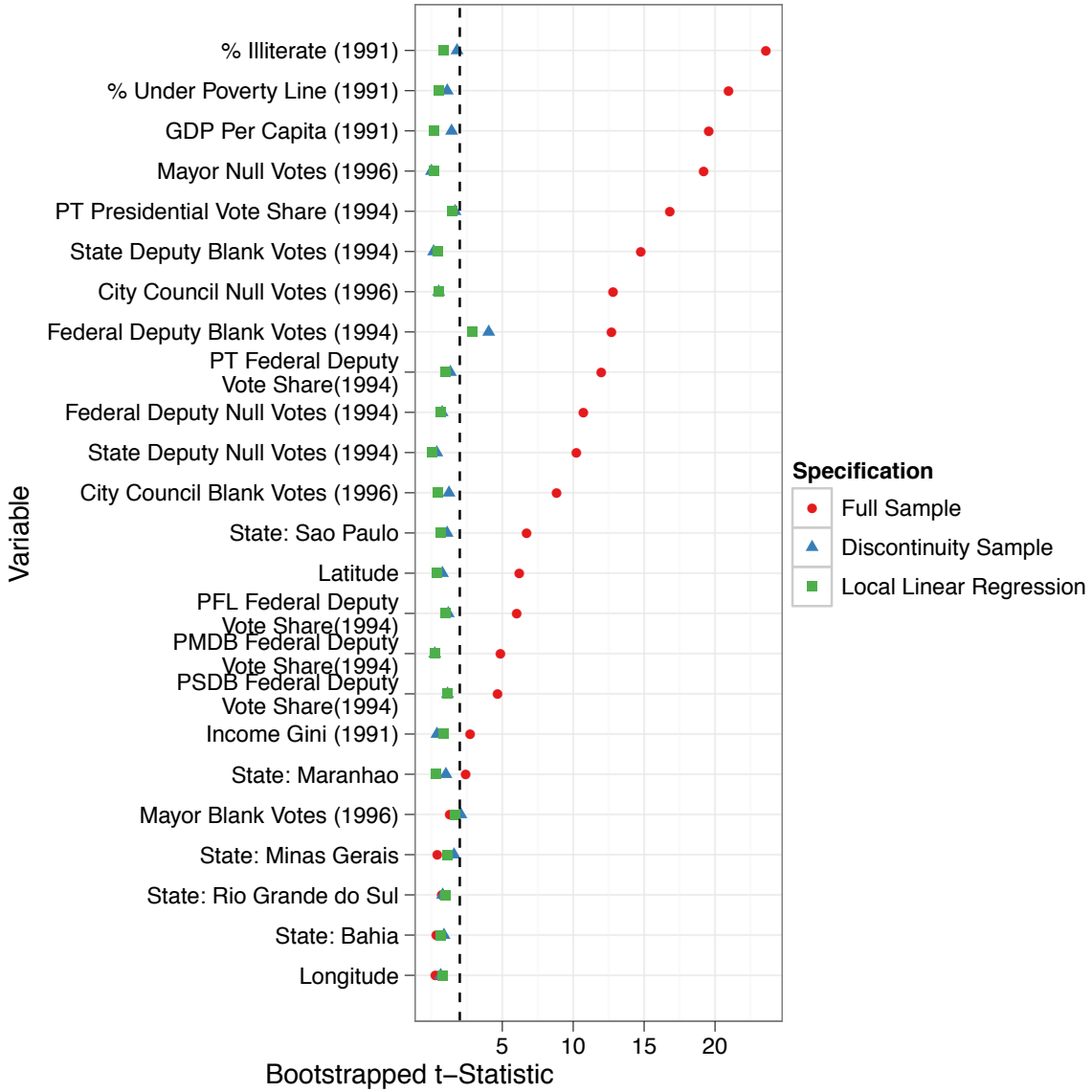
TABLE 4. Regression Discontinuity Design Results: Effect of Serving in House of Commons on (Log) Wealth at Death

	Conservative Party		Labour Party	
Effect of serving	0.61	0.66	−0.20	−0.25
Standard error	(0.27)	(0.37)	(0.26)	(.26)
Covariates		x		x
Percent wealth increase	83	94	−18	−23
95% Lower bound	8	−7	−52	−65
95% Upper bound	212	306	31	71

Note: Effect estimates at the threshold of winning $\tau_{RDD} = E[Y(1) - Y(0) \mid Z = 0]$. Estimates without covariates from local polynomial regression fit to both sides of the threshold with bootstrapped standard errors. Estimates with covariates from local linear regression with rectangular kernel (equation 2); bandwidth is 15 percentage point of vote share margin with

Eggers and Hainmueller 2009

Falsification Test



Hidalgo 2012

Placebo Threshold

- Test whether the treatment effect is zero when it should be
- Let $c^\star$ be a placebo threshold value. Run the regression of: $E[Y|X, D] = \beta_0 + \beta_1(X - c^\star) + \alpha D + \beta_3((X - c^\star) \cdot D)$ and check if α large and significant?
  - Usually we split the sample to the left and the right of the actual threshold c in order to avoid miss-specification by imposing a zero jump at c
- The existence of large placebo jumps does not necessarily invalidate the RDD, but does require an explanation
- Concern is that the relation is fundamentally discontinuous and jump at cut-off is contaminated by other factors.
- Maybe data exists in a period where there was no program or there is a time when the threshold value was changed so we can run placebo tests

Sorting Around the Threshold and Compound Treatments

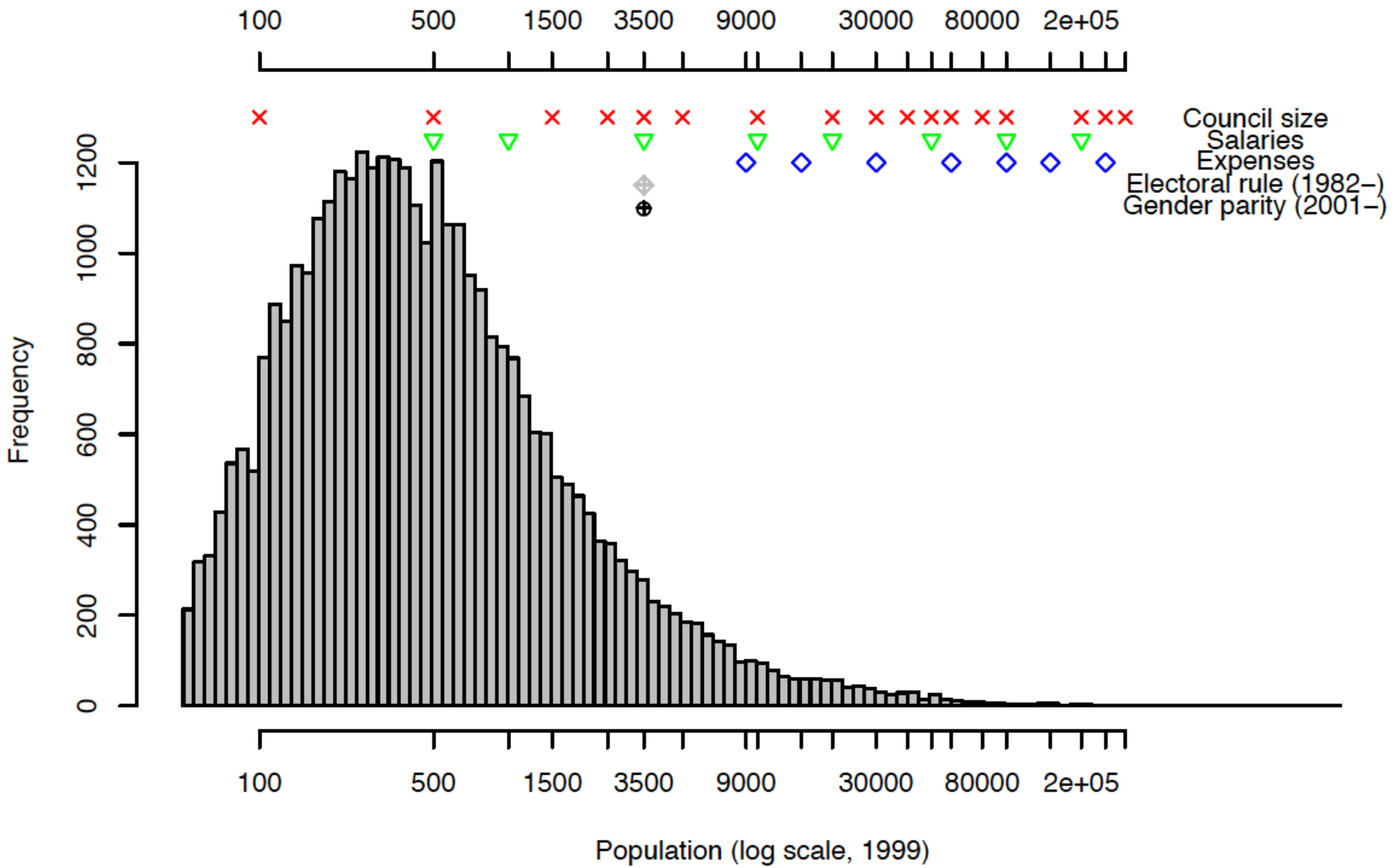
- Can sorting behavior invalidate the local continuity assumption?
  - Is it plausible that units exercise precise control over their values of the assignment variable? Theory helps!
  - Can administrators strategically choose what assignment variable to use or which cut-off point to pick?
  - Either can invalidate the comparability of subjects near the threshold because of sorting of agents around the cut-off, where those below may differ on average from those just above
  - Does not necessarily invalidate the design unless sorting is very widespread and very precise

- Is there a compound treatment? What else changes at c? Continuity can be violated in the presence of other programs that use a discontinuous assignment rule with the same assignment variable and threshold
  - Sorting and compound treatment are often a concern in spatial RDDs

Sorting Around the Threshold

- Test for discontinuity in density of forcing variable:
  - Visual Histogram Inspection:
    - Construct equal-sized non-overlapping bins of the forcing variable such that no bin includes points to both the left and right of the cut-off
    - For each bin, compute the number of observations and plot the bins to see if there is a discontinuity at the cut-off
  - Formal tests
  - DCdensity command in Stata
    - McCrary, J. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142 (2): 698–714.

Sorting Around the Threshold



Eggers 2010

Outline

Outline

Fuzzy Regression Discontinuity Design

- Threshold may not perfectly determine treatment exposure, but it creates a discontinuity in the probability of treatment exposure
- Incentives to participate in a program may change discontinuously at a threshold, but the incentives are not powerful enough to move all units from non-participation to participation
- We can use such discontinuities to produce instrumental variable estimators of the effect of the treatment (close to the discontinuity)

Fuzzy Regression Discontinuity Design

- Probability of being offered a scholarship may jump at a certain SAT threshold when applicants are given "special consideration" (but not everybody above get's it)
  - Shouldn't compare recipients with non-recipients (even close to threshold) since they differ along unobserved confounders (e.g., letters of rec, etc.)
- Administrators might offer the scholarship to everybody above the threshold, but there might be non-compliance in the take up (e.g. some of those offered the scholarship don't take it)
  - Shouldn't compare those who take it with those who do not because they differ on unobserved confounders (e.g. motivation, etc.)
- Close to the threshold we can exploit the discontinuity as an instrument to identify the LATE for the subgroup of applicants for whom scholarship receipt/uptake depends on the difference between their score and the threshold

Fuzzy Regression Discontinuity Design

- Conceptually the Fuzzy RDD is similar to the instrumental variable framework we use for encouragement design experiments
- Assume the scholarship is offered to everybody above c, but not everybody might take it
- Let $Z = 1\{X > c\}$ be a binary encourgament indicator that captures whether units are above or below the threshold c
- Let D be the binary observed treatment indicator that captures whether applicants take the scholarship or not
- Let D_z indicate *potential* treatment status given $Z = z$
  - $D_1 = 1$ encouraged to take the treatment and takes the treatment
- Observed treatment is realized as

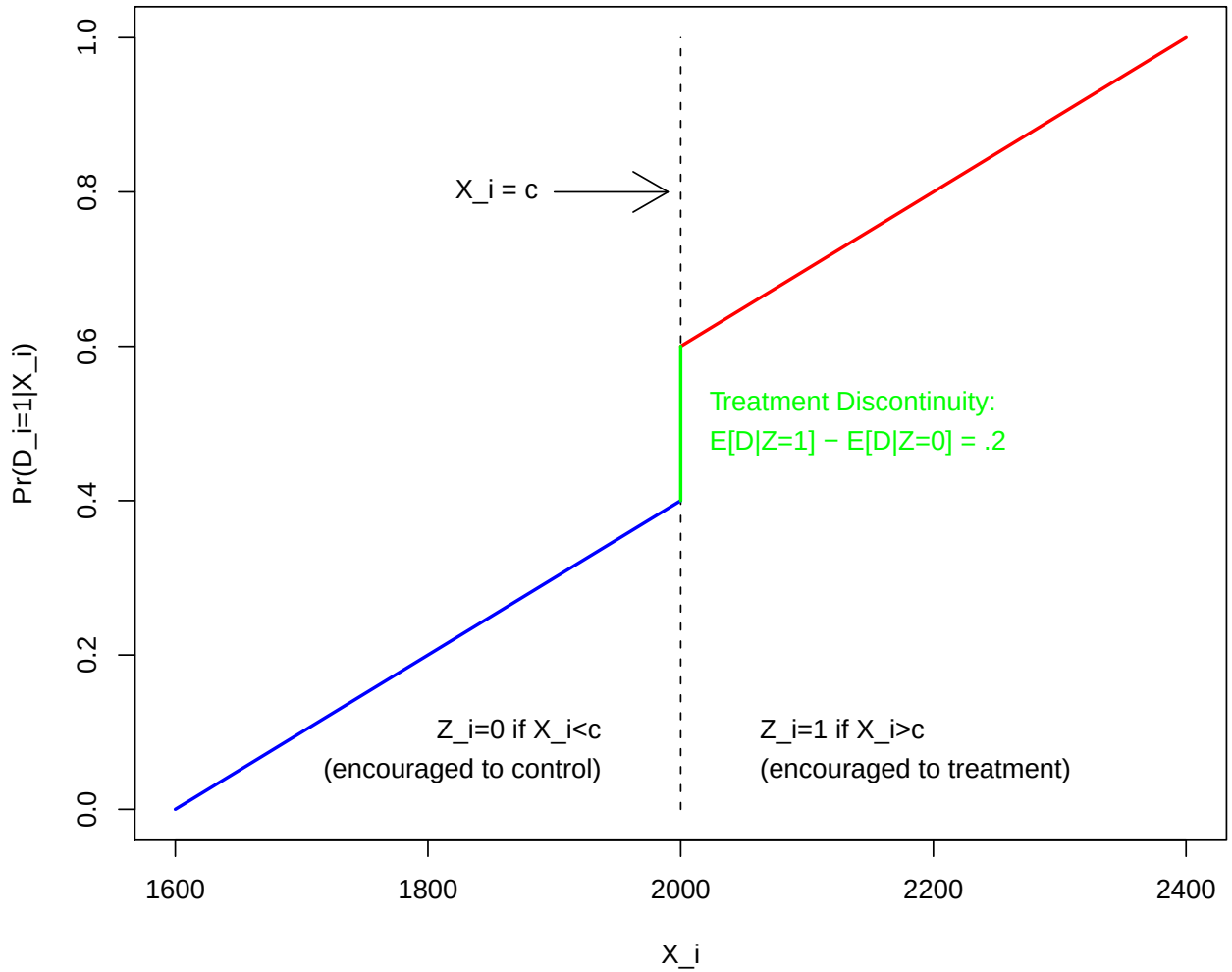
$$D = Z \cdot D_1 + (1 - Z) \cdot D_0 \text{ so } D_i = \begin{cases} D_{1i} & \text{if } Z_i = 1 \\ D_{0i} & \text{if } Z_i = 0 \end{cases}$$

Compliance Types

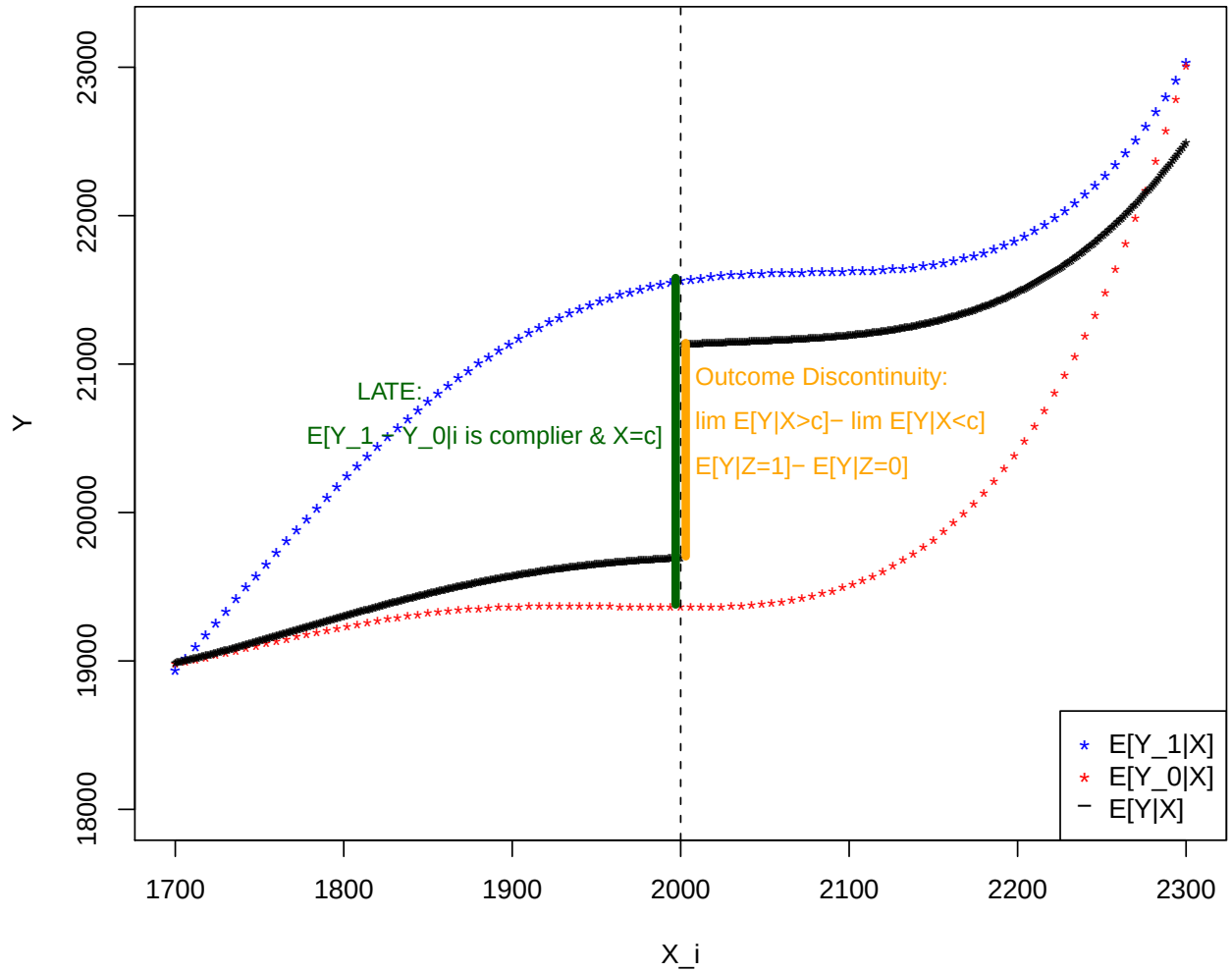
Definition

- *Compliers*:
  - $D_1 > D_0$ ($D_0 = 0$ and $D_1 = 1$)
  - Takes treatment if above threshold but not if below threshold
- *Always-takers*:
  - $D_1 = D_0 = 1$.
  - Always takes treatment, regardless if above or below threshold
- *Never-takers*:
  - $D_1 = D_0 = 0$.
  - Never takes treatment, regardless if above or below threshold
- *Defiers*:
  - $D_1 < D_0$ ($D_0 = 1$ and $D_1 = 0$).
  - Takes treatment if below threshold but not if above threshold

Discontinuity in $E[D|X]$



Discontinuity in $E[Y|X]$



Identification

Identification Assumption

- *Binary instrument* Z *with* $Z = 1\{X > c\}$
- *Restrict sample to observations close to discontinuity where* $E[Y|D, X]$ *jumps so that* $X \approx c$ *and thus* $E[X|Z = 1] - E[X|Z = 0] \approx 0$.
- *Usual IV assumptions hold (ignorability, first stage, montonicity, no defiers)*

Identification Result

$$\begin{aligned}
\alpha_{FRDD} &= E[Y_1 - Y_0 | X = c \text{ and } i \text{ is a complier}] \\
&= \frac{\lim_{x \downarrow c} E[Y|X = c] - \lim_{x \uparrow c} E[Y|X = c]}{\lim_{x \downarrow c} E[D|X = c] - \lim_{x \uparrow c} E[D|X = c]} \\
&= \frac{\textit{outcome discontinuity}}{\textit{treatment discontinuity}} \\
&\approx \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[D|Z = 1] - E[D|Z = 0]}
\end{aligned}$$

Outline

Estimation

- Cut the sample to a small window above and below the threshold (discontinuity sample)
- Code instrument: $Z = 1\{X > c\}$
  - 1 if unit is above; 0 if unit is below threshold c
- Code margin on forcing variable: $\tilde{X} = X - c$
  - distance to threshold: $+$ if above; - if below; 0 at threshold
- Fit two-stage least squares regression:

$$Y = \beta_0 + \beta_1 \tilde{X} + \beta_2 (Z \cdot \tilde{X}) + \alpha D + \varepsilon$$

  where D is instrumented with Z
- Always check whether instrument is weak (i.e. compliance ratio is too low)
- Specification can be made more flexible by adding polynomials
- Also helpful to separately plot (and or estimate) the outcome discontinuity and treatment discontinuity

Outline

Early Release Program (HDC)

- Prison system in many countries is faced with overcrowding and high recidivism rates after release.
- Early discharge of prisoners on electronic monitoring or tag has become a popular policy
- Difficult to estimate impact of early release program on future criminal behavior: best behaved inmates are usually the ones to be released early
- Marie (2008) considers Home Detention Curfew (HDC) scheme in England and Wales:
- Fuzzy RDD: Only offenders sentenced to more than three months (88 days) in prison are eligible for HDC, but not all those with longer sentences are offered HDC

Table 1: Descriptive Statistics for Prisoners Released by Length of Sentence and HDC and Non HDC Discharges

Panel A - Released Before 3 Months:			
Discharge Type	**Non HDC**	**HDC**	**Total**
Percentage Female	12.2	-	12.2
Mean Age	29.5	-	29.5
Percentage Incarcerated for Violence	17.6	-	17.6
Mean Number Previous Offences	8.8	-	8.8
Recidivism within 12 Months	52.4	-	52.4
Sample Size	42,987	0	42,987
Panel B - Released Between 3 and 6 Months:			
Discharge Type	**Non HDC**	**HDC**	**Total**
Percentage Female	8.8	8.8	8.8
Mean Age at Release	27.6	30.8	28.4
Percentage Incarcerated for Violence	20.3	18.3	19.8
Mean Number Previous Offences	10	6.5	9.1
Recidivism within 12 Months	60	30.2	52.6
Sample Size	52,091	17,222	69,313

Table 2: Descriptive Statistics for Prisoners Released by Length of Sentence and HDC and Non HDC Discharges and +/-7 Days Around Discontinuity Threshold

Panel A - Released +/- 7 Days of 3 Months (88 Days) Cut-off:			
Discharge Type	**Non HDC**	**HDC**	**Total**
Percentage Female	10.5	9.7	10.3
Mean Age at Release	28.9	30.7	29.3
Percentage Incarcerated for Violence	19.8	18.2	19.4
Mean Number Previous Offences	9.5	5.7	8.7
Recidivism within 12 Months	54.6	28.1	48.8
Sample Size	18,928	5,351	24,279
Panel B - Released +/- 7 Days of 3 Months (88 Days) Cu-off:			
Day of Release around Cut-off	**- 7 Days**	**+ 7 Days**	**Total**
Percentage Female	11	10.2	10.3
Mean Age at Release	28.8	29.4	29.3
Percentage Incarcerated for Violence	17.1	19.7	19.4
Mean Number Previous Offences	9.1	8.6	8.7
Recidivism within 12 Months	56.8	47.9	48.8
Percentage Released on HDC	0	24.4	22
Sample Size	2,333	21,946	24,279

Figure 1: Proportion Discharged on HDC by Sentence Length

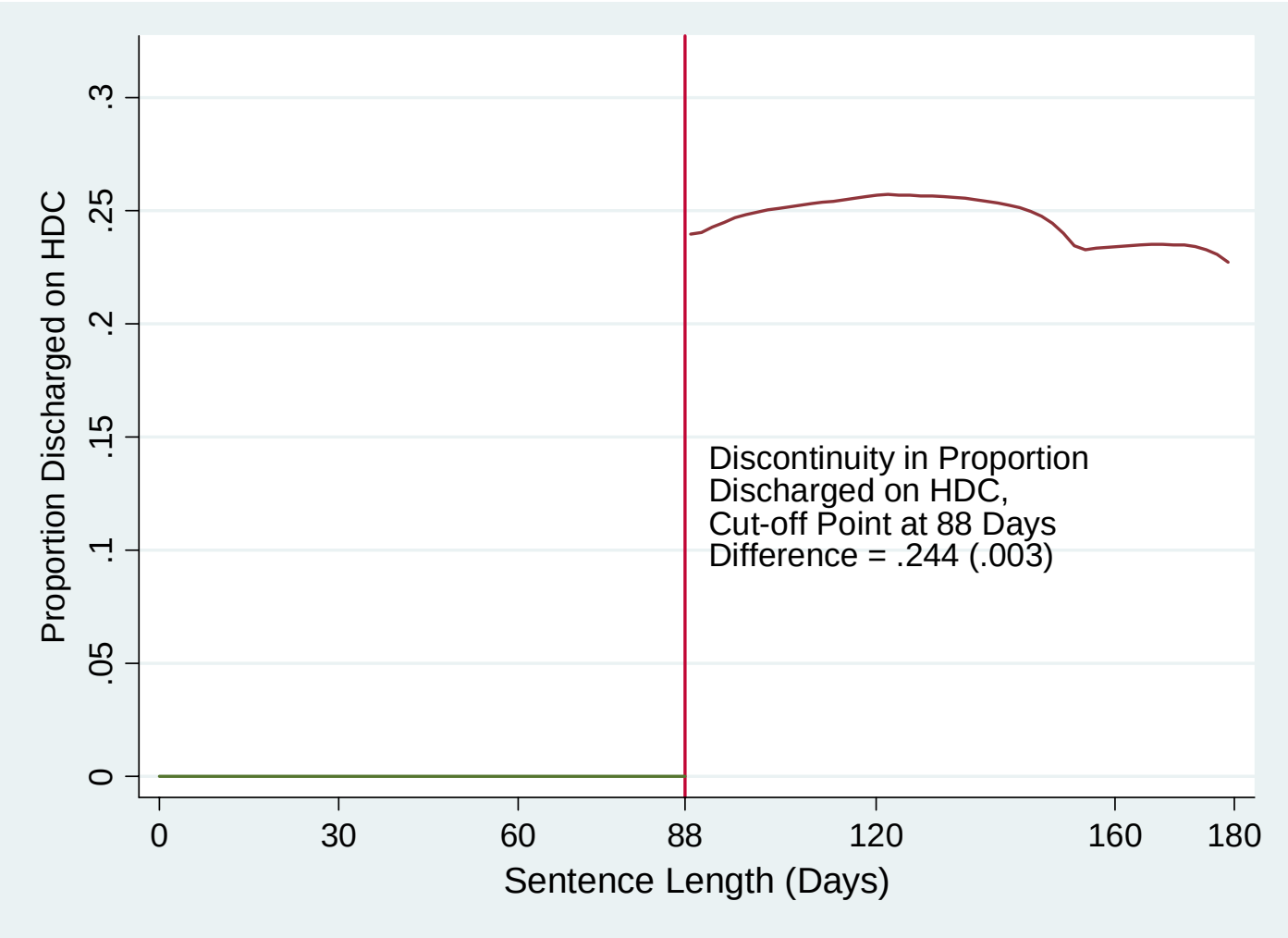


Figure 2: Mean Number of Previous Offence by Sentence Length

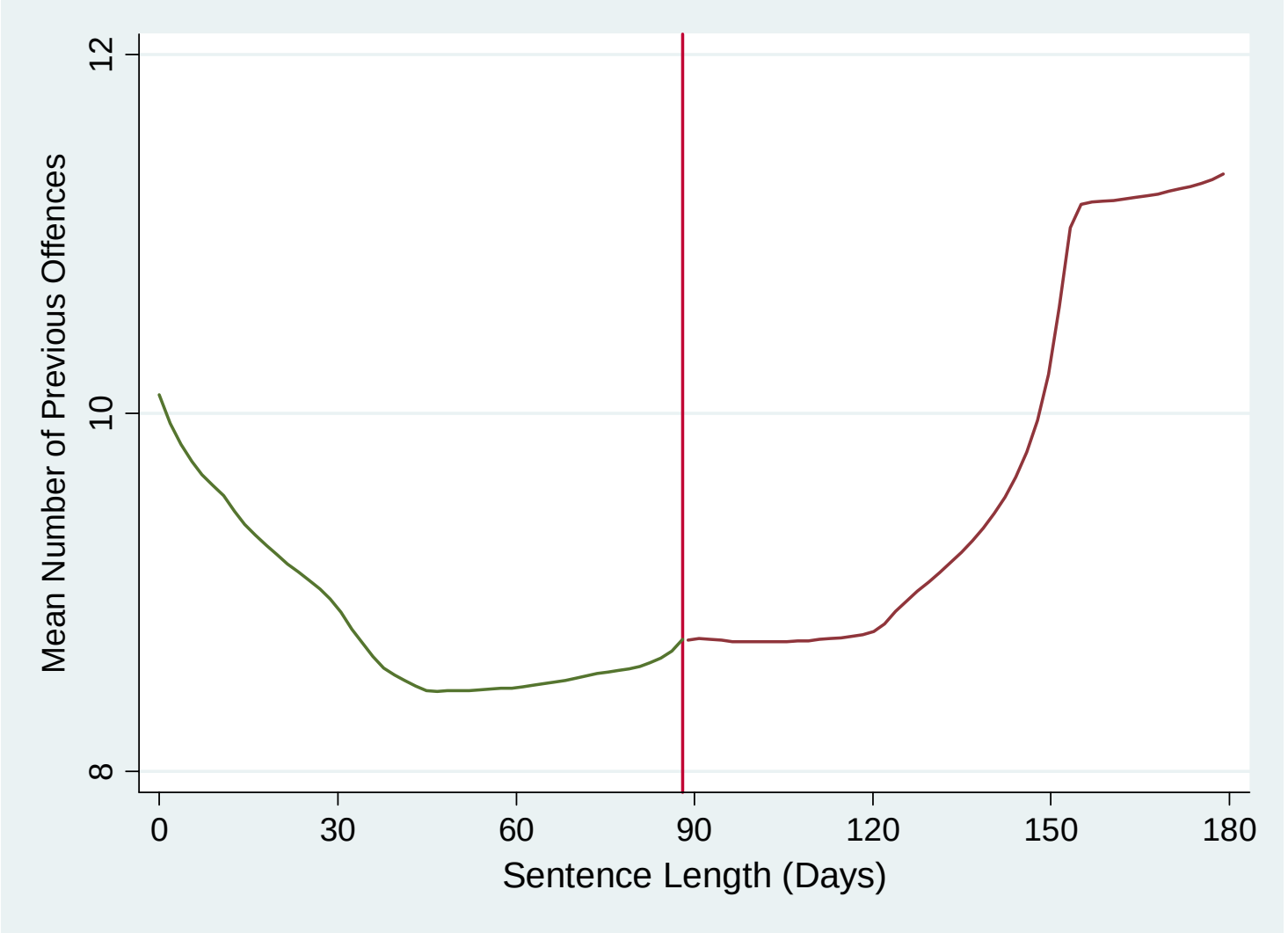


Figure 4: Recidivism within 1 Year by Sentence Length

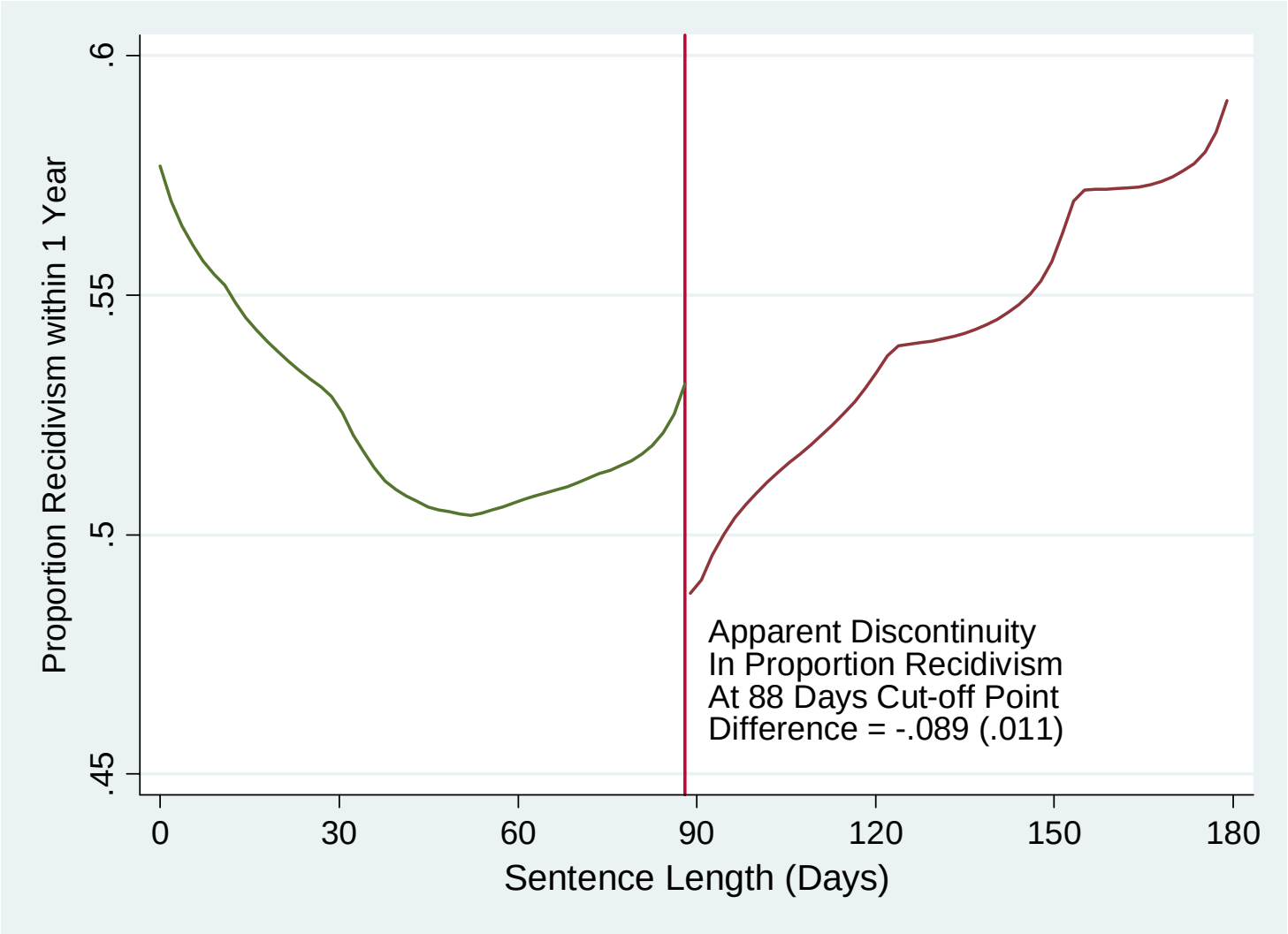


Table 4: RDD Estimates of HDC Impact on Recidivism – Around Threshold

	Dependent Variable = Recidivism Within 12 Months Estimation on Individuals Discharged +/- 7 Days of 88 Days Threshold		
	(1)	(2)	(3)
Estimated Discontinuity of HDC Participation at Threshold ($HDC^{+} - HDC^{-}$)	.243 (.009)	.223 (.009)	.243 (.003)
Estimated Difference in Recidivism Around Threshold ($Rec^{+} - Rec^{-}$)	-.089 (.011)	-.059 (.009)	-.044 (.014)
Estimated Effect of HDC on Recidivism Participation ($Rec^{+} - Rec^{-}$)/ ($HDC^{+} - HDC^{-}$)	-.366 (.044)	-.268 (.044)	-.181 (n.a.)
Controls	No	Yes	No
PSM	No	No	Yes
Sample Size	24,279	24,279	24,279

Internal and External Validity

- At best, Sharp and Fuzzy RDD estimate the average effect of the sub-population with X_i close to c
- Fuzzy RDD restricts this subpopulation even further to that of the compliers with X_i close to c
- Only with strong assumptions (e.g., homogenous treatment effects) can we estimate the overall average treatment effect
    - Some new methods to get further away from the discontinuity (Angrist and Rokkanen 2012; Hainmueller, Hall and Snyder 2015)
- RDDs have strong internal validity but may have limited external validity (although it depends...)