11: Regressionsdiskontinuitetsdesigns

Videregående kvantitative metoder i studiet af politisk adfærd

Frederik Hjorth
fh@ifs.ku.dk
fghjorth.github.io
@fghjorth

Institut for Statskundskab
Københavns Universitet

22. november 2017

1. Formalia
2. Opsamling fra sidst
3. Regressionsdiskontinuitetsdesigns
   - Motiverende eksempel
   - Formel definition
   - Udfordringer
   - RDD vs. Diff-in-Diff
   - Eksempler på RD-designs
4. Case: Eggers & Hainmueller
5. Kig fremad

Fagets opbygning

Blok 1

Gang	Tema	Litteratur	Case
1	Introduktion til R	Leeper (2016)	
2	R workshop + tidy data	Wickham (2014), Zhang (2017)	
3	Regression I: OLS brush-up	AP kap 3	Newman et al. (2015), Solt et al. (2017)
4	Regression II: Paneldata	AGS kap 4	Larsen et al. (2017)

Fagets opbygning

Blok 2

5	Introduktion til kausal inferens	Hariri (2012), Samii (2016)	Eckles & Bakshy (2017)
6	Matching	Justesen & Klemmensen (2014)	Nall (2015)
Efterårsferie			
7	Eksperimenter I	AP kap 1+2, GG kap 1+2	Gerber, Green & Larimer (2008)
8	Eksperimenter II	GG kap 3+4+5	Gerber & Green (2000)
9	Instrumentvariable	AP kap 4	Lundborg et al. (2017)
10	Difference-in-differences	AP kap 5	
11	Regressionsdiskontinuitetsdesigns	AP kap 6	Eggers & Hainmueller (2009)

Fagets opbygning

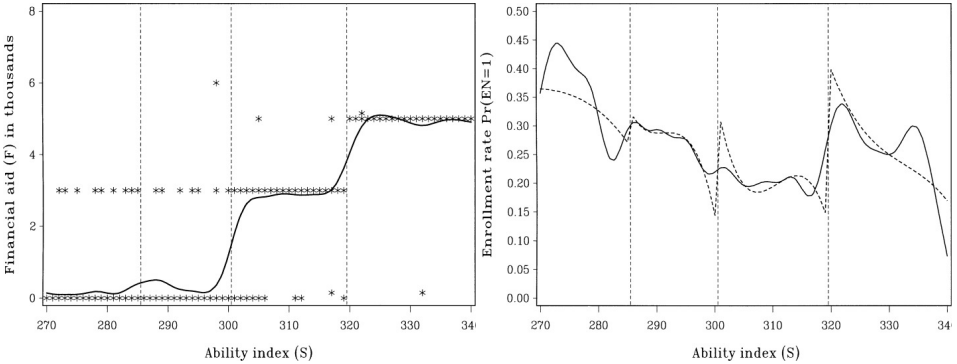
Blok 3

12	Tekst som data	Grimmer & Stewart (2013), Benoit & Nulty (2016)	Baturo & Mikhaylov (2013)
13	Scraping af data fra online-kilder	MRMN kap 9+14	Hjorth (2016)
14	'Big data' og maskinlæring	Varian (2014), Montgomery & Olivella (2017)	Theocharis et al. (2016)

Opsamling fra sidst

- eksempel: Card & Krueger om mindsteløn og beskræftigelse
- logikken i DiD
- parallel trends assumption
- DiD i regressionsform
- case: flygtningecentre og EU-afstemninger

National Merit Award Scholarships



Kilde: Van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression–discontinuity approach. *International Economic Review*, 43(4), 1249-1287.

Formelt:

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

Outcome varierer med treatment D_i, men kan også variere med running/forcing variable x_i

RD-model:

$$Y_i = \alpha + \beta x_i + \rho D_i + \eta_i \tag{2}$$

- Y_i er diskontinuert funktion af kriterierne i ligning (1)
- Y_i er samtidig kontinuert funktion af x_i
- hvis vi kan skille disse to fra hinanden kan vi estimere ρ

Klassisk RD-design i politologien: studiet af *incumbency advantage*

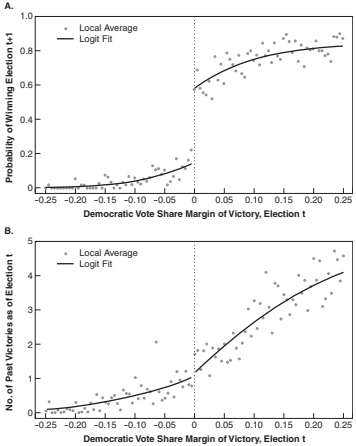


Figure 6.1.2 The probability of winning an election by past and future vote share (from Lee, 2008). (A) Candidate's probability of winning election $t+1$, by margin of victory in election t: local averages and logit polynomial fit. (B) Candidate's accumulated number of past election victories, by margin of victory in election t:

- formelt set ingen common support
- $\rightarrow$ vi skal gøre antagelser om running-variablens funktionelle form

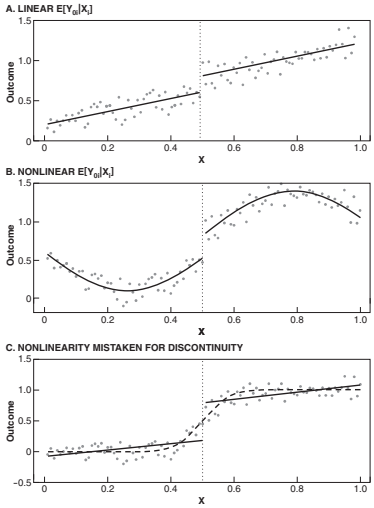


Figure 6.1.1 The sharp regression discontinuity design.

Model (1) estimeres i et smalt 'vindue' rundt om cutoff $\rightarrow$ i fastsættelsen af vinduets 'bandwidth', et klassisk tradeoff:

»[I]f the window is very narrow, there are few observations left, meaning the resulting estimates are likely to be too imprecise to be useful. Still, we should be able to trade the reduction in bias near the boundary against the increased variance suffered by throwing data away, generating some kind of optimal window size.« (*Mastering 'Metrics*, 161)

Metoder til optimal bandwidth selection, fx. Imbens, G., & Kalyanaraman, K. (2011). "Optimal bandwidth choice for the regression discontinuity estimator". *The Review of economic studies*.

Samii om RD (o.a.) -designs' 'localness':

»The LATE theorem states that under a set of basic identifying conditions, an instrumental variable identifies the average causal effect for the subpopulation of units whose treatment status is in fact moved by the instrument. (...) Similarly, regression discontinuity identifies effects local to the relevant cut points, matching with calipers identifies effects local to the region of common covariate support, experiments identify effects local to the typically nonrepresentative sample of experimental subjects, and so on.« (950)

Implementering i R: `rdd`-pakken

```
RDestimate(formula, data, cutpoint, bw)
```

Potentielt problem v. RDD: *sorting*

Figure 4. Democratic Vote Share Relative to Cutoff: Popular Elections to the House of Representatives, 1900-1990

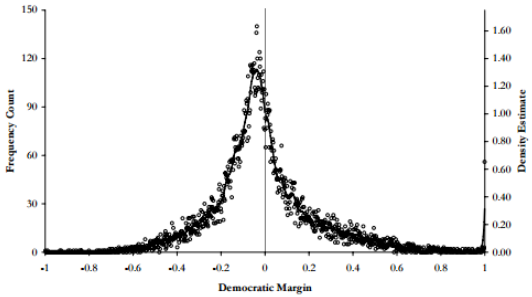
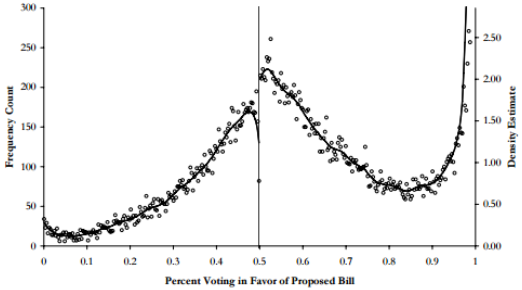


Figure 5. Percent Voting Yeay: Roll Call Votes, U.S. House of Representatives, 1857-2004



$\rightarrow$ testes i R m. `DCdensity()` $\rightarrow$ tester density omkring cutoff mod nulhypotese om ingen sorting

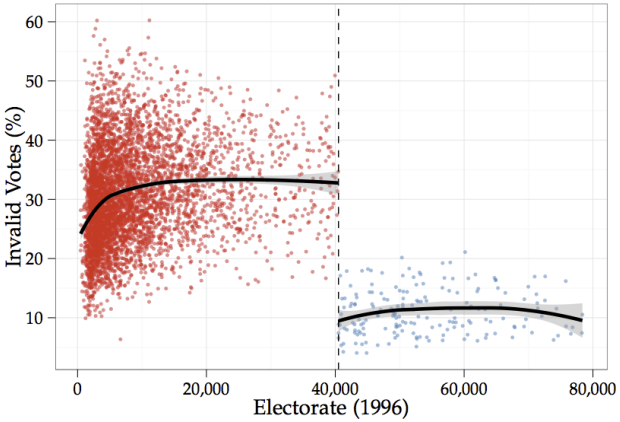
Mest oplagte case til DiD:

- to sammenlignelige (typer af) enheder
- én 'treates', én treates ikke
- målinger på outcome før og efter for begge
- flere før-målinger mhp. evaluering af parallel trends

Mest oplagte case til RD:

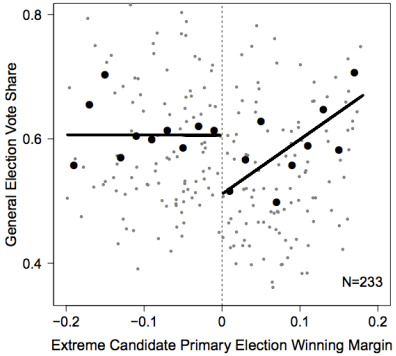
- en præcist målt, ikke-manipulerbar sorting-variabel (x_i)
- treatment (D_i) implementeres skarpt v. cutoff
- præcist estimeret kontinuert sammenhæng ml. x_i og outcome
- mange observationer tæt på cutoff (x_0)

Hidalgo (2010): elektronisk stemmeafgivning og valide stemmer

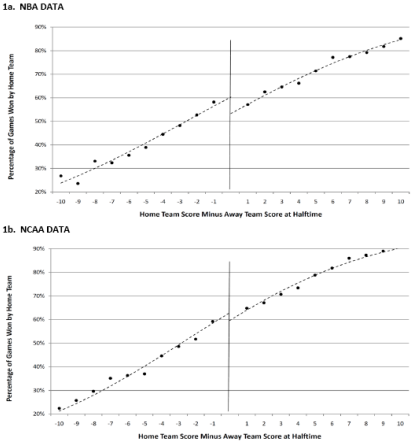


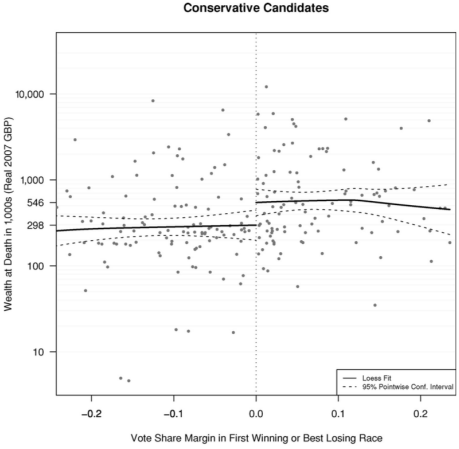
Hall (2014): ideologisk ekstreme primærvindere og valgresultater

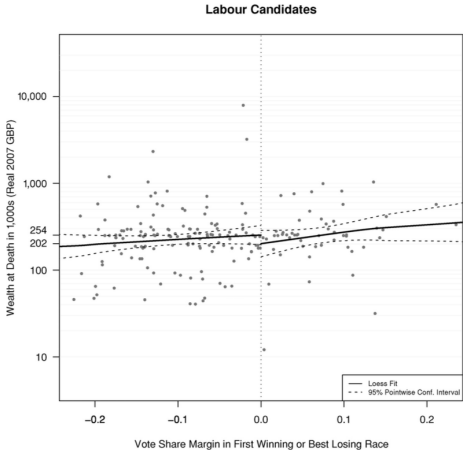
Figure 2 – General-Election Vote Share After Close Primary Elections Between Moderates and Extremists: U.S. House, 1980–2010. The close election of the more extreme primary candidate causes a decrease in general-election vote share for the party.



Berger & Pope (2011): er det bedre at være bagud ved pausen?




Conservative Candidates
Wealth at Death in 1,000s (Real 2007 GBP)
10,000
1,000
546
298
100
10
−0.2
−0.1
0.0
0.1
0.2
Vote Share Margin in First Winning or Best Losing Race
Loess Fit
95% Pointwise Conf. Interval


Labour Candidates
Wealth at Death in 1,000s (Real 2007 GBP)
10,000
1,000
254
202
100
10
−0.2
−0.1
0.0
0.1
0.2
Vote Share Margin in First Winning or Best Losing Race
Loess Fit
95% Pointwise Conf. Interval

Eggers & Hainmueller om estimatets 'localness':

»As is well known, the RD design is likely to have a very high degree of internal validity, but we pay a price in terms of decreased external validity and also efficiency. τ_{RDD} is a local average treatment effect informative only for marginal candidates close to the threshold of winning (unless additional homogeneity assumptions are introduced). This is desirable in our context, however, because the counterfactual is more reasonable for marginal compared to "unbeatable" candidates. Moreover, given that candidates in closer races attract more public scrutiny and face a higher risk of electoral defeat, rent seeking may be limited compared to candidates in safe districts (Barro 1973; Besley and Burgess 2002; Besley and Case 1995). Presumably, our estimates of the returns to office therefore provide a conservative lower bound for the average across all MPs.« (fn. 27)

Næste gang: tekst som data

- tekst som data
- Grimmer & Stewart: god oversigtstekst
- Benoit & Nulty introducerer R-pakken `quanteda` (ctr. `tm`)
- læs Baturo & Mikhaylov kursorisk

Tak for i dag!