

6: Matching

Videregående kvantitative metoder i studiet af politisk adfærd

Frederik Hjorth

fh@ifs.ku.dk

fghjorth.github.io

@fghjorth

Institut for Statskundskab

Københavns Universitet

10. oktober 2016

- 1 Formalia
- 2 Opsamling fra sidst
- 3 Rubins kausalmodel
- 4 Matching
- 5 Ladd & Lenz (2009)
- 6 Kig fremad

- frivillig R-workshop kl. 13-16, lokale 2.0.30
- midterm løses på workshopen, frigives på github.com/fghjorth/vkme16 kl. 12.59

Centrale tendenser i evalueringen:

- flere R-opgaver
- gerne R-workshop før semesteret
- nogle pensumtekster dur ikke (især fsva. paneldata)
- lidt for meget repetition
- generel ros til underviseren

Uge	Dato	Tema	Litteratur	Case
1	5/9	Introduktion til R	Imai kap 1	
2	12/9	Regression I: OLS	GH kap 3, MM kap 2	Gilens & Page (2014)
3	26/9	Regression II: Paneldata	GH kap 11	Larsen et al. (2016)
4	29/9	Regression III: Multileveldata, interaktioner	GH kap 12	Berkman & Plutzer
5	3/10	Introduktion til kausal inferens	Hariri (2012), Samii (2016)	
6	10/10	Matching	Justesen & Klemmensen (2014)	Ladd & Lenz (2009)
	17/10	*Efterårsferie*		

Uge	Dato	Tema	Litteratur	Case
	17/10	*Efterårsferie*		
7	24/10	Eksperimenter I	MM kap 1, GG kap 1+2	Bond et al. (2012)
8	31/10	Eksperimenter II	GG kap 3+4+5	Gerber & Green (2000)
9	7/11	Instrumentvariable	MM kap 3	Arunachalam & Watson
10	14/11	Regressionsdiskontinuitetsdesigns	MM kap 4	Eggers & Hainmueller
11	21/11	Difference-in-difference designs	MM kap 5	Enos (2016)
12	28/11	'Big data' og maskinlæring	Grimmer (2015), Varian (2014)	
13	5/12	Scraping af data fra online-kilder	MRMN kap 9	
14	12/12	Tekst som data	Grimmer & Stewart (2013), Imai kap 5	

Spørgsmål?

- implementering af multilevelmodeller i R
- multikollinearitetsproblemer i Gilens & Page
- introduktion til kausal inferens: 'credibility-revolutionen'
- endogenitet
- post-treatment adjustment bias

Spørgsmål?

Lad os antage en påvirket gruppe ($T = 1$) og en upåvirket gruppe ($T = 0$).

Rubins kausalmodel:

$$\delta_i = Y_{i1} - Y_{i0} \quad (1)$$

M.a.o.: *effekten δ_i er forskellen mellem Y_i når $T = 1$ og Y_i når $T = 0$*

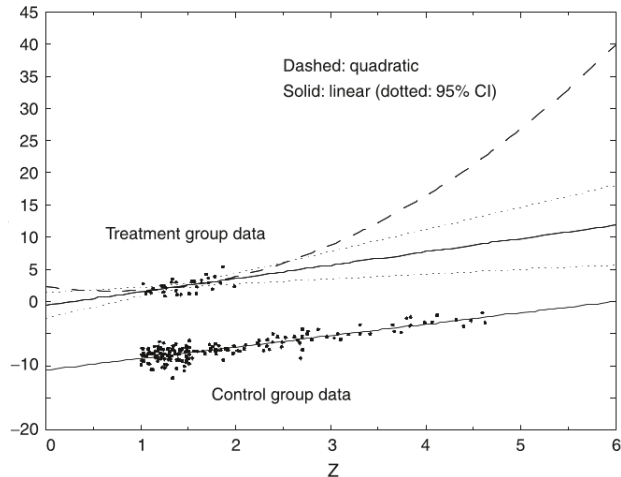
kaldes også 'the potential outcomes framework'

Problem: T er altid *enten* 1 *eller* 0 \rightarrow vi observerer altid kun Y_{i1} *eller* $Y_{i0} \rightarrow$ vi kan aldrig observere δ_i

denne uobserverbarhed kaldes **the fundamental problem of causal inference**

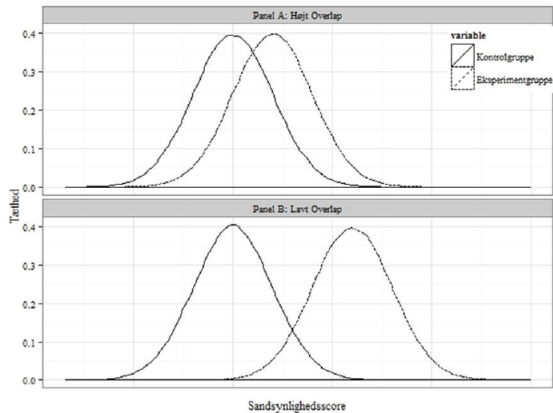
Problem i OLS-tilgange: beror (potentielt) på interpolation/ekstrapolation

→ *model dependence*

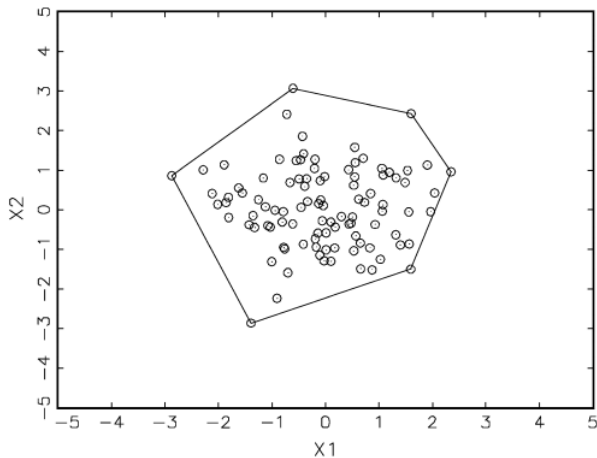


Løsning: fokuser på dele af data med *common support*

Figur 1: Intervaller i data med højt og lavt overlap



I N -dimensionelle data defineres support ud fra datas 'konvekse hylster' (*convex hull*)



Klassisk tilgang i matching: *exact matching*

Men: ofte ikke tilstrækkeligt mange eksakte matches → behov for *dimensionalitätsreduktion*

Traditionel løsning: *propensity score matching* (PSM) → hver obs. estimeres ssh for treatment

Efter PSM-estimation: fx. *nearest neighbor matching* (men: bias-variance tradeoff)

Alternativ: *radius* (caliper) matching

Sidenhen skepsis om PSM's fortræffeligheder, jf. fx. King & Nielsen (2016), "Why Propensity Scores Should Not Be Used for Matching"

Centralt problem ved PSM: misspecifikation i PS-modellen kan give bias (og er inefficiant selv ved korrekt specifikation)

Alternativ: *Coarsened Exact Matching* → variable opdeles i grove, *meningsfulde* kategorier og matches iht. kategori

I CEM måles balance med \mathcal{L}_1 :

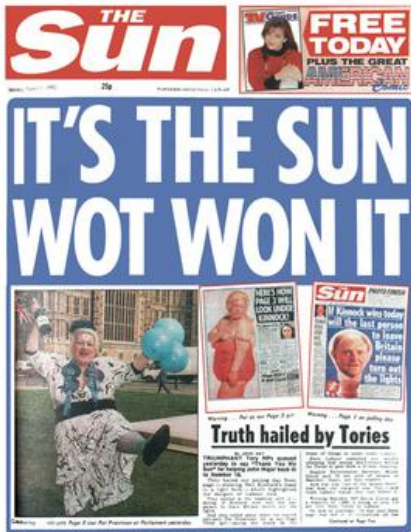
$$\mathcal{L}_1 = \frac{1}{2} \sum_{\ell_1 \dots \ell_k} |f_{\ell_1 \dots \ell_k} - g_{\ell_1 \dots \ell_k}| \quad (2)$$

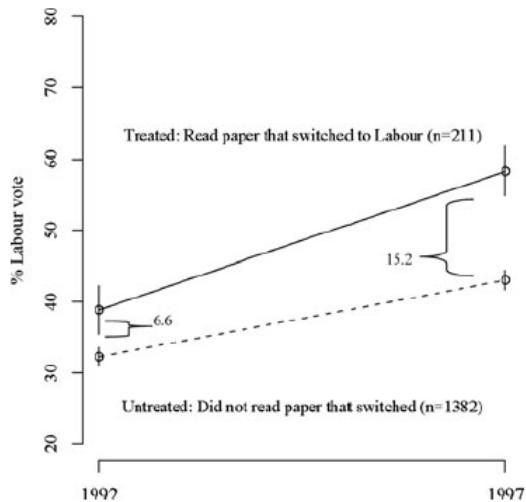
fuld common support $\rightarrow \mathcal{L}_1 = 0$; ingen common support $\rightarrow \mathcal{L}_1 = 1$

Fremgangsmåde med CEM:

- ① evaluér balance med `imbalance()` (ekskl. treatment og outcome)
- ② definér kategorier for matching-variable
- ③ kør CEM med `cem()` m. information fra (2)
- ④ evaluér balance igen
- ⑤ hvis tilfredsstillende balance, estimér effekt kun m. matched data

Spørgsmål?





Næste gang:

- efterårsferie!

Tak for i dag!