5: Introduktion til kausal inferens

Videregående kvantitative metoder i studiet af politisk adfærd

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

Institut for Statskundskab
Københavns Universitet

3. oktober 2016

- permanent lokaleændring: faste holdtimer nu i **lokale 1.0.10**
- frivillig R-workshop mandag d. 10. oktober 13-16, lokale 2.0.30
- justering: midterm udleveres på workshoppen kl. 13

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 & Watso
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?

- partial pooling
- clustering, ICC
- prædiktorer på gruppeniveau
- interaktioner
- case: Berkman & Plutzer

Opsamles i dag:

- implementering i R

Spørgsmål?

Multileveldata:

- Afhængig variabel `y`
- Uafhængige variable `x1` og `x2`
- Gruppeindikator `j`

Ren varying intercept-model:

```
lmer(y~1+(1|j),data=df)
```

Varying intercept-model med to kontinuerte variable:

```
lmer(y~x1+x2+(1|j1),data=df)
```

Varying slopes hvor `x2` varierer mellem grupper i `j`:

```
lmer(y~x1+x2+(1+x2|j1),data=df)
```

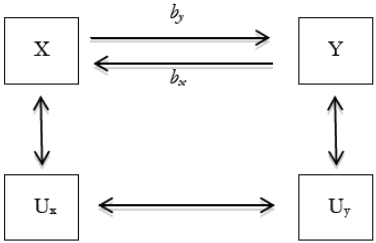
»The analysis is prone to underestimating drastically the causal impact of median-income preferences, assuming that regression coefficients even capture causality in this context: the authors' claim to causal inference is based only on the fact that they perform multiple regression.«

Bashir, O. S. (2015). Testing Inferences about American Politics: A Review of the "Oligarchy" Result. *Research & Politics*, 2(4).

Spørgsmål?

Udgangspunkt: observeret korrelation ml. x og y

Figur 1: Korrelation mellem *x* og *y*



Situationen i Fig. 1 som ligningssystem:

$$y = b_y x + a_y u_y \tag{1}$$

$$x = b_x y + a_x u_x \tag{2}$$

Nødvendige restriktioner a og b:

$$b_x = 0 \tag{a}$$

$$E[u_x u_y] = 0 \tag{b}$$

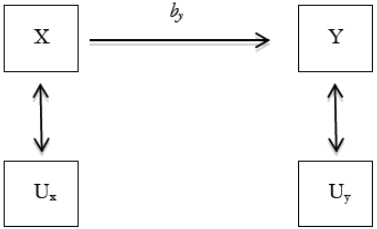
Hermed kan (1) og (2) omskrives til:

$$y = b_y x + a_y u_y \tag{1'}$$

$$x = a_x u_x \tag{2'}$$

Hermed:

Figur 2: Kausal identifikation af størrelsen af påvirkningen af x på y



Spørgsmål?

Den klassiske tilgang: masseproduktion af 'pseudo-general pseudo-facts'

»At the turn of the millennium, the modal quantitative research design was one in which researchers assembled data on theoretically interesting dependent and independent variables (...) Researchers then assessed the presumably causal relationships in these data using regressions with informally motivated sets of control variables to reduce the potential for confounding.«

Sidenhen: en 'credibility revolution' i samfundsvidenskaben

»This convention in quantitative causal research appears to be breaking down, and more quantitative causal research is moving toward causal empiricism. This (...) represents a major change in what researchers believe are credible ways of doing causal inference.«

Problemer i klassiske regressionstilgange:

1. mgl. ekstern validitet: nominel ctr. effektiv stikprøve
2. mgl. intern validitet: misspecifikation

Ad (1):

Figure 1. Nominal and effective samples from Jensen (2003), reproduced from Aronow and Samii (2016)

Ad (2):

Table 1. Replication and Auxiliary Analyses for Laitin and Fearon (2003)

	Outcome						
	Civil War Onset						Per Capita Income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Estimator	Logit	Logit	Logit	Logit	Logit	Logit	OLS
Prior war	−.95 **				−.24	−.38	
	(.31)				(.23)	(.25)	
Per capita income	−.34***			−.29***		−.29***	
	(.07)			(.07)		(.07)	
Ethnic fractionalization	.17	1.12***	1.12**	.35	1.16**	.40	−4.14***
	(.37)	(.33)	(.42)	(.39)	(.43)	(.40)	(.90)
Observations	6,327	6,610	6,610	6,373	6,610	6,373	6,373
Country-clustered SEs			Y	Y	Y	Y	Y

Spørgsmål?

Næste gang:

- matching!
- pensumtekst: Justesen & Klemmensen (2014)
- case: Ladd & Lenz (2009)

Tak for i dag!