9: Instrumentvariable

Videregående kvantitative metoder i studiet af politisk adfærd

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

Institut for Statskundskab
Københavns Universitet

14. november 2016

1. Formalia
2. Opsamling fra sidst
3. Instrumentvariable
4. Implementering i R
5. Case: Arunachalam & Watson
6. Kig fremad
7. Hvad f**k skete der i tirsdags?

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	Gerber et al. (2008)
8	31/10	Eksperimenter II	GG kap 3+4+5	Gerber & Green (2000)
9	14/11	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	

Eksamen

- Format: *seminaropgave*
- Frist: 20/12
- Omfang: min. 10, max. 20 ns.
- Rammebeskrivelse for seminaropgaven uploades 20/11
- Mini-workshop om databehandling primo december (dato/lok. tbd)

Spørgsmål?

- Clustered assignment
- Brug af pre-treatment mål
- Brug af andre kovariater
- Blocking
- Noncompliance
- Case: Gerber & Green (2000)

Fra holdtime 8: for hvert subjekt i defineres

$$ITT_{i,D} \equiv d_i(1) - d_i(0) \tag{1}$$

$$ITT_{i,Y} \equiv Y_i(1) - Y_i(0) \tag{2}$$

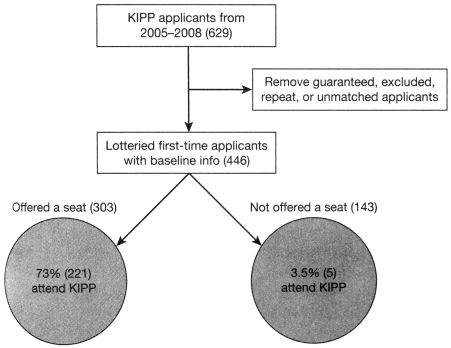
CACE er forholdet mellem $\overline{ITT_{i,Y}}$ og $\overline{ITT_{i,D}}$:

$$CACE = \frac{ITT}{ITT_D} \tag{3}$$

Spørgsmål?

Motiverende eksempel: KIPP charter schools

FIGURE 3.1
Application and enrollment data from KIPP Lynn lotteries



Note: Numbers of Knowledge Is Power Program (KIPP) applicants are shown in parentheses.

KIPP tilbud opfylder tre kriterier:

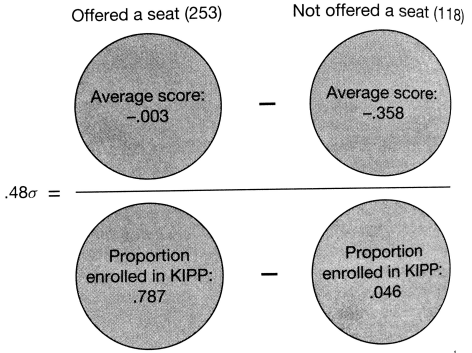
1. first stage effekt: instrumentet påvirker treatment (*relevance criterion*)
2. instrumentet er ukorreleret med evt. omitted variables (*independence assumption*)
3. instrumentet påvirker alene outcome gennem treatment (*exclusion criterion*)

Med kriterierne opfyldt identificerer forholdet ml. reduced form og first stage LATE

$$LATE = \frac{reduced\ form}{first\ stage} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = \lambda = \frac{\rho}{\phi} \tag{4}$$

IV-estimat for KIPP-pladstilbud

FIGURE 3.2
IV in school: the effect of KIPP attendance on math scores



Note: The effect of Knowledge Is Power Program (KIPP) enrollment described by this figure is $.48\sigma = .355\sigma/.741$.

Fire typer compliance:

1. compliers
2. never-takers
3. always-takers
4. defiers

Antagelse om *monotonicitet*, dvs. ingen defiers $\rightarrow$

$$\lambda = LATE = E[Y_1 - Y_0|C = 1] = CACE \tag{5}$$

Alternativt effektbegreb: treatment effect on the treated (TOT)

$$TOT = E[Y_1 - Y_0 | D = 1] \tag{6}$$

Gruppen med $D = 1$ omfatter compliers med $Z = 1$ + always-takers
$\rightarrow$ i fravær af always-takers er TOT = LATE

IV-modeller kan estimeres med funktionen `ivreg()` i pakken `AER`:

```
ivreg(<yvar> ~ <treatment> , ~ <instrument>, data=<data>)
```

Udgangspunkt: korr. m. højde og stemmeadfærd

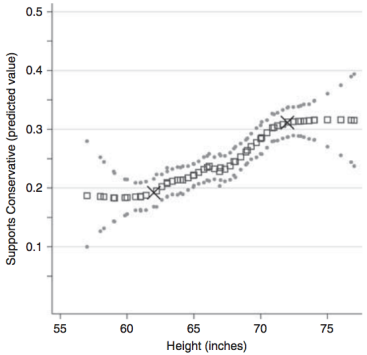


Fig. 1. Taller people support Conservatives
Note: Running line smooth of 'supports Conservative party' on height, adjusted for age and gender. The Xs mark the 10^{th} and 90^{th} percentile of the height distribution; 95 per cent pointwise bootstrapped confidence intervals displayed.

First stage

TABLE 4 *Instrumental Variables: First Stage*

	Whole	Whole	Female	Female	Male	Male
	(1)	(2)	(3)	(4)	(5)	(6)
First Stage:						
Height (inches)	0.352***	0.211***	0.247***	0.162***	0.491***	0.239***
	(0.049)	(0.047)	(0.053)	(0.051)	(0.081)	(0.078)
Controls:						
Age, region	X	X	X	X	X	X
Sex	X	X				
Extended		X		X		X
F-Stat excl. instrument	47.678	16.684	23.413	10.413	30.049	7.009
N	11,303	11,001	6,145	6,004	5,158	4,997

Note: dependent variable is 'Real Income ('000s of pounds)'. First stage of 2SLS regression corresponding to Table 1. Extended controls include: married, white, years of schooling, religion. Full models reported in the Appendix. Heteroskedasticity-robust standard errors, clustered by household. Statistical significance: *10%; **5%; ***1%.

Second stage

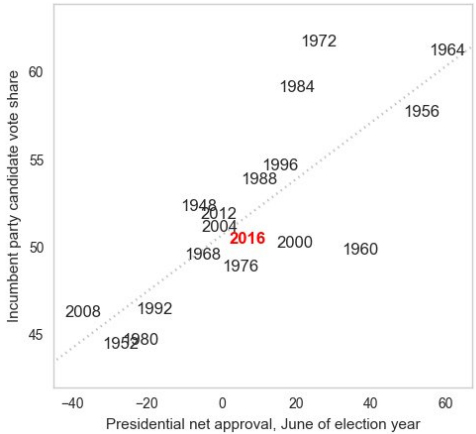
TABLE 5 *Support for Conservatives: Second-Stage IV and OLS*

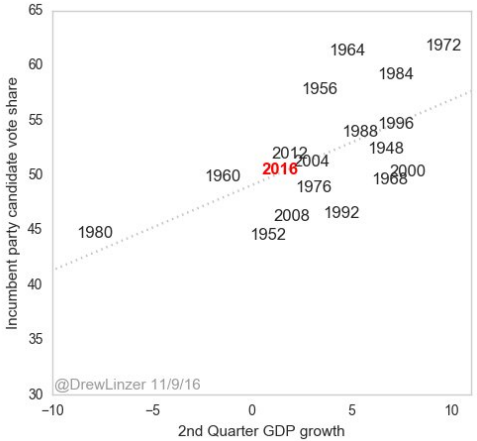
	Whole	Whole	Prime	Prime	Cog	Fam	Fam	F	F	M	M
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
IV Second Stage:											
Real Income (000s)	0.024***	0.032***	0.020***	0.027***	0.048***	0.029***	0.019***	0.022***	0.026*	0.023***	0.037***
	(0.005)	(0.011)	(0.005)	(0.009)	(0.023)	(0.012)	(0.010)	(0.009)	(0.015)	(0.006)	(0.016)
OLS:											
Real Income (000s)	0.003***	0.002***	0.003***	0.003***	0.002***	0.002***	0.002***	0.001	0.001	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Controls:											
Age, region	X	X	X	X	X	X	X	X	X	X	X
Sex	X	X	X	X	X	X	X				
Extended		X		X	X	X	X	X	X	X	X
Cognitive ability					X						
Parents' schooling						X					
Father's HGS							X				
Prime age only			X	X							
F-stat	32.783	24.524	19.514	15.390	15.682	18.365	28.267	28.389	23.914	16.273	9.337
A-R Conf. interval	[0.014, 0.036]	[0.015, 0.065]	[0.012, 0.033]	[0.013, 0.057]	[0.019, 0.138]	[0.011, 0.069]	[0.002, 0.047]	[0.007, 0.045]	[0.002, 0.079]	[0.013, 0.037]	[0.016, 0.1]
N	9,616	9,377	5,477	5,419	9,341	7,917	7,085	5,104	4,994	4,512	4,383

Note: dependent variable is 'supports Conservative Party'. The first row reports coefficients on income from second stage of 2SLS regressions; the second row reports coefficients on income from OLS regressions. Extended controls include: married, white, years of schooling, religion. Full models reported in the Appendix. Heteroskedasticity-robust standard errors, clustered by household. Statistical significance: *10%; **5%; ***1%.

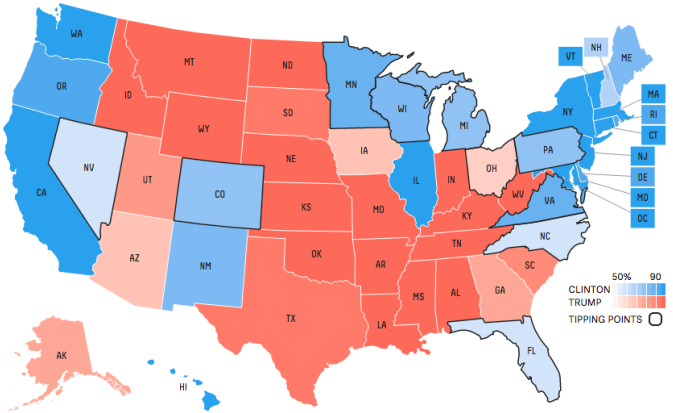
Næste gang: RDD

‘Fundamentals’ indikerede uao. et tæt resultat

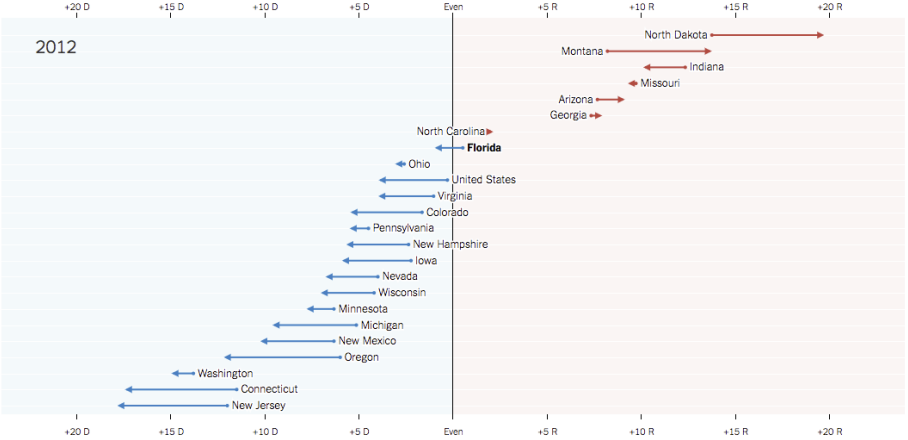




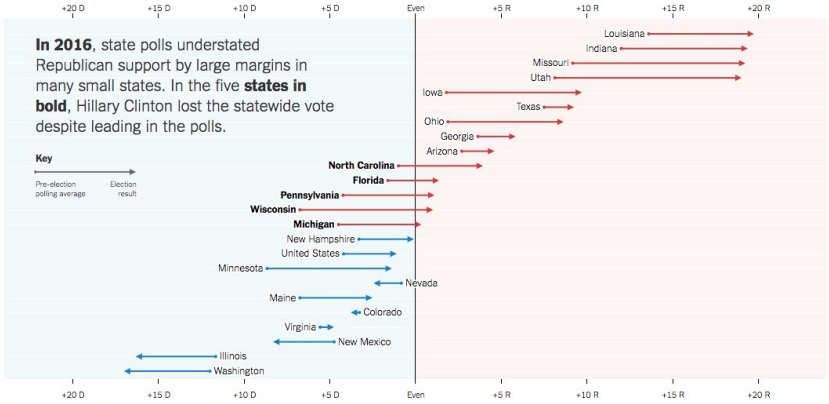
538's 'polls-only' på valgaftenen



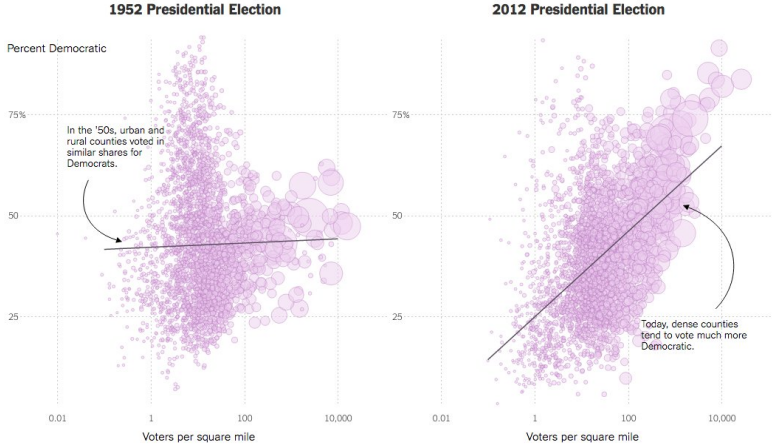
2012-fejl i Obamas favør → inferens om 'ground game'-effekt



Polling-fejl var uens fordelt

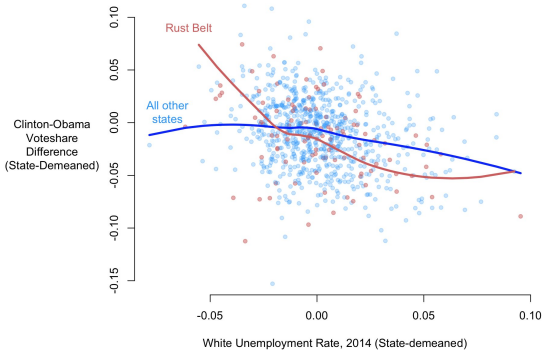


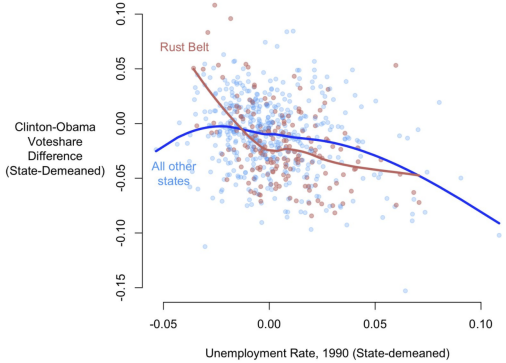
Tendens til stærkere land/by skel fortsætter



NYT om land ctr. by 2012/2016

Tegn på smh. m. svage lokale økonomier...





... men potentielt confoundet af uddannelse

```
Call:
lm(formula = presidential_swing ~ acs2011_median_hh_income +
    acs2011_pct_educ_bachelors, data = dat)

Residuals:
      Min        1Q    Median        3Q       Max
-0.254728 -0.028513  0.001691  0.030636  0.315298

Coefficients:
                                 Estimate     Std. Error t value Pr(>|t|)
(Intercept)                -0.13374866236  0.00332848484 -40.183   <2e-16 ***
acs2011_median_hh_income   -0.00000007226  0.00000009733  -0.742    0.458
acs2011_pct_educ_bachelors  0.34485944007  0.01396562309  24.693   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04366 on 2719 degrees of freedom
Multiple R-squared:  0.3079,    Adjusted R-squared:  0.3074
F-statistic: 604.8 on 2 and 2719 DF,  p-value: < 2.2e-16
```

Tak for i dag!