

3: Regression II: Paneldata

Videregående kvantitative metoder i studiet af politisk adfærd

Frederik Hjorth

fh@ifs.ku.dk

fghjorth.github.io

@fghjorth

Institut for Statskundskab

Københavns Universitet

26. september 2016

- 1 Formalia
- 2 Opsamling fra sidst
- 3 Paneldata
- 4 Larsen et al. (2016): hjælp boligboblen Fogh?
- 5 Kig fremad

- husk erstatningstime på **torsdag 12-14, lokale 2.0.30**
- frivillig R-workshop mandag d. 10. oktober 13-16, lokale 2.0.30
- midterm er bestået/ikke bestået
- udleveres mandag d. 10. oktober kl. 17, frist 7 dage
- efter næste holdtime: uformel halvvejsevaluering
- nogle kommende undervisningsgange: monitorering ifm. adjunktpædagogikum

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 (2013)
3	26/9	Regression II: Panelmodeller	GH kap 11	Larsen et al. (2016)
4	29/9	Regression III: Multilevelmodeller, interaktioner	GH kap 12	Dinesen & Søndberg (2014)
5	3/10	Introduktion til kausal inferens	Hariri (2012), Samii (2016)	
6	10/10	Matching	Justesen & Klemmensen (2014)	Ladd & Lenz (2015)
	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?

- motivation: oligarki i USA?
- OLS: intuition
- formel form: ex. m. human kapital
- kilder til variation i e_i
- kilder til variation i $SE(\widehat{\beta}_k)$
- omitted variable bias

Opsamles i dag:

- subsetting
- plotting
- hvad er en regressionskoefficient?

Subsetting: isolation af dele af et datasæt

- generel notation: `data[<række(r)>,<kolonne(r)>]`
- fx.: `df[3,5]` refererer til række 3, kolonne 5 i et datasæt
- man kan subsetting flere rækker/kolonner ad gangen: fx. `df[2:3,1:5]`
- i en data frame er kolonner d.s.s. *variable*
- subsetting af en enkelt variabel kan også gøres m. dollartegn: `df$var1`
- hvis `var1` er den 5. variabel i `df` er `df[,5]` altså d.s.s. `df$var1`

Plotting: scatterplot af to variable over for hinanden

- generel notation: `plot(x1,x2)`
- ex.: data framen `df` med variablene `var1` og `var2`
- → hvordan plotter jeg de to variable i `df` mod hinanden?
- → `plot(df$var1,df$var2)`

Simuleret datasæt, simdat:

y	x
2	1
8	3
11	5

→ hvordan subsetter jeg variabelen x ?

→ hvordan subsetter jeg kun de to nederste rækker?

Formlen for en regressionskoefficient (jf. MM s. 86n):

$$b = \beta = \frac{\text{Cov}(Y_i, X_i)}{\text{Var}(X_i)} \quad (1)$$

dvs.: kovariansen ml. X og Y divideret med variansen på X , hvor

$$\text{Cov}(Y_i, X_i) = \frac{\sum_i^N (Y_i - \bar{Y})(X_i - \bar{X})}{n - 1} \quad (2)$$

og

$$\text{Var}(X_i) = \frac{\sum_i^N (X_i - \bar{X})^2}{n - 1} \quad (3)$$

Simuleret datasæt, simdat:

y	x
2	1
8	3
11	5

→ hvad er $Cov(Y_i, X_i)$?

→ hvad er $Var(X_i)$?

→ hvad er β ?

Hvad er en regressionskoefficient?

Spørgsmål?

Paneldata: data hvor enheder *observeres flere gange*

Husk formlen fra sidste gang:

$$Y_i = \alpha + \beta P_i + \gamma A_i + e_i \quad (4)$$

Antag nu at vi observerer indkomst (Y_i) og privatskoleuddannelse (P_i) over tid t :

$$Y_{it} = \alpha + \beta P_{it} + \gamma A_i + e_{it} \quad (5)$$

NB: A_i varierer her ikke med t , dvs. er *tidsinvariant*

Så længe A_i er tidsinvariant kan vi med paneldata estimere β uden bias **uden at observere A_i** :

$$Y_i = \alpha_i + \lambda_t + \beta P_i + e_i \quad (6)$$

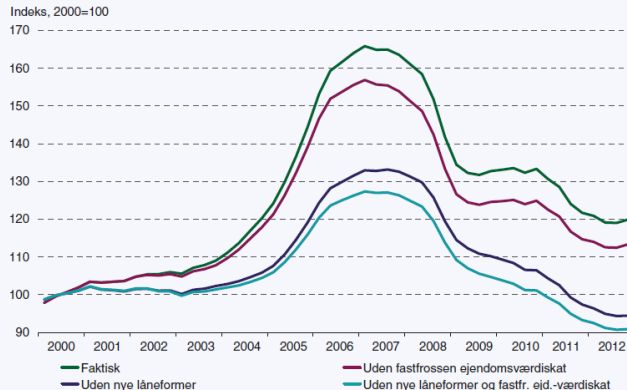
- kaldes en 'fixed effects' (FE) model
- α_i = fixed effects for enheder \rightarrow opfanger *tidsinvariant uobserveret heterogenitet ml. enheder*
- λ_t = fixed effects for tid \rightarrow opfanger *enhedsinvariant uobserveret heterogenitet ml. tidsperioder*
- tilbageværende variation kun variationen 'inden for' enheder \rightarrow FE-model kaldes også *within-estimator*

- OBS: FE-modellen tager ikke højde for at observationer ikke er uafhængige
- specifikt: observationer for samme enhed 'klumper sammen'
- konsekvensen er at standardfejlene underestimeres → det er ikke godt!
- løsningen er at inkorporere i modellen at observationer 'klumper'
- → mere om dette i næste uge

Spørgsmål?

Larsen et al. (2016): hjælp boligboblen Fogh?

Figur B Huspriser med og uden nye låneformer og fastfrosset ejendomsværdiskat



Anm.: Kontrafaktiske forløb baseret på estimeret efterspørgselsrelation. Variabelt forrentede lån antages først at slå igennem i 1. kvartal 2000. I fravær af den fastfrosne ejendomsværdiskat antages den i MONAs databank imputerede ejendomsværdiskat at blive holdt konstant som andel af boligbeholdningen opgjort til markedsværdi. Fastfrysningen af ejendomsværdiskatten er modelleret som en permanent nedsættelse af ejendomsværdiskatten med 63 pct. i 1. kvartal 2002, svarende til nutidsværdien i ændringen af det fremtidige skatteprovenu under antagelse af at skattestoppet er permanent. Den grønne linje (det faktiske forløb) angiver den samlede effekt af henholdsvis nye låneformer (rød linje) og fastfrosne ejendomsværdiskat (blå linje). Beregningerne er nærmere dokumenteret i Dam m.fl. (2011). Serierne er genberegnet og forlænget frem til udgangen af 2012.

Kilde: Danmarks Nationalbank.

»Third, the panel set-up of data enables us to **rule out time-invariant structural differences** between local contexts as explanations of any observed relationship between local house prices and support for incumbents by using only within-precinct/individual variation in local housing prices (**by means of fixed effects**). This is particularly important given the strong urban-rural gradient in local property values, which would very likely confound any observed cross-sectional relationship with support for the sitting government.«

Table 1: Estimated effects of house prices on electoral support for governing parties.

	(1)	(2)	(3)	(4)
Δ housing price	0.104** (0.008)	0.048** (0.007)	0.053** (0.008)	0.030** (0.007)
Unemployment rate				-1.904** (0.221)
Log(Median income)				-0.887** (0.064)
Year FE		✓	✓	✓
Precinct FE			✓	✓
Observations	4197	4197	4197	4177
RMSE	8.407	6.751	5.716	5.326

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$

Spørgsmål?

Næste gang:

- regression III: multilevel-modeller
- opsamling på interaktioner (I hope)
- GH kap 12:
 - partial pooling vs. complete pooling (12.2+12.3)
 - multilevel-modeller i R (12.4)
 - group-level prediktorer (12.6)
- case-tekst: *optional*

Tak for i dag!