

## 4: 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

27. september 2017

- 1 Opsamling
- 2 Motivation
- 3 Paneldata
- 4 Larsen et al. (2017): hjælp boligboblen Fogh?
- 5 Implementering i R
- 6 Kig fremad

Sidste gang:

- OLS intuition
- OLS formel form
- omitted variable bias
- læsning af OLS-output
- implementering i R

## 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. (2016)

## Fagets opbygning

### Blok 2

5	Introduktion til kausal inferens	Hariri (2012), Samii (2016)	
6	Matching	Justesen & Klemmensen (2014)	Nall (2015)
<i>Efterårsferie</i>			
7	Eksperimenter I	AP kap 1, 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	Enos (2016)
11	Regressionsdiskontinuitetsdesigns	AP kap 6	Eggers & Hainmueller (2009)

## Fagets opbygning

### Blok 3

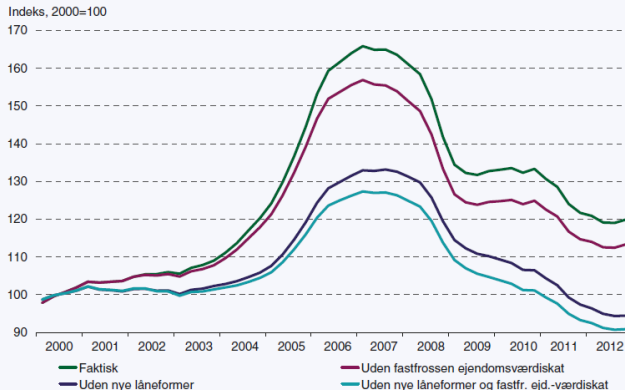
12	Tekst som data	Grimmer & Stewart (2013), Benoit & Nulty (2016)	Baturo & Mikhaylov (2013)
13	Scraping af data fra online-kilder	MRMN kap 9+14	Hjorth (2016)
14	'Big data' og maskinlæring	Varian (2014), Montgomery & Olivella (2017)	Theocharis et al. (2016)

PROFESSIONSHØJSKOLEN

METROPOL

## Larsen et al. (2017): 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.



Data på bred form kan konverteres til lang form vha. `gather()` i `dplyr`-pakken

```
gather(<data>, <key>, <value>, <...>)
```

hvor

- `key`: navn på variabel der angiver variabelnavne fra bredt format
- `value`: navn på variabel der angiver værdier fra bredt format
- `...` : intervallet af variable der skal 'stakkes', fx. `obs1992:obs1998`

Eks.: bredt data med enhederne a, b og c og outcome y observeret i t2 og t2

unit	yt1	yt2
a	1	2
b	3	6
c	6	7

→ hvordan skal data se ud på lang form?

- en normal, 'pooled' OLS 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!

Klassiske OLS-antagelser (jf. AGS boks 4.1):

- ① simpel tilfældig udvælgelse
- ② linearitet
- ③ ej perfekt multikollinearitet
- ④ fejllad ukorrelerede med  $X$ 'er
- ⑤ varianshomogenitet
- ⑥ fravær af autokorrelation
- ⑦ normalfordelte fejllad

→ hvilken trues i paneldata?

Notation i AGS (p. 127):

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + \gamma_1 z_{1i} + \dots + \gamma_j z_{ji} + u_i + e_{it} \quad (1)$$

$u_i$ : uobserverede, tidsinvariante prediktorer → med paneldata kan vi kontrollere for al confounding herfra!

Husk formlen fra sidste gang:

$$Y_i = \alpha + \rho s_i + \gamma A_i + e_i \quad (2)$$

Antag nu at vi observerer indkomst ( $Y_i$ ) og uddannelse ( $s_i$ ) over tid  $t$ :

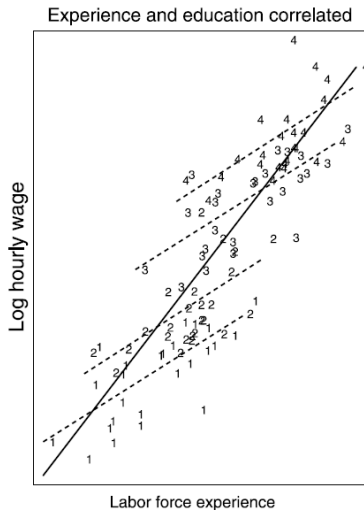
$$Y_{it} = \alpha + \rho s_{it} + \gamma A_i + e_{it} \quad (3)$$

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

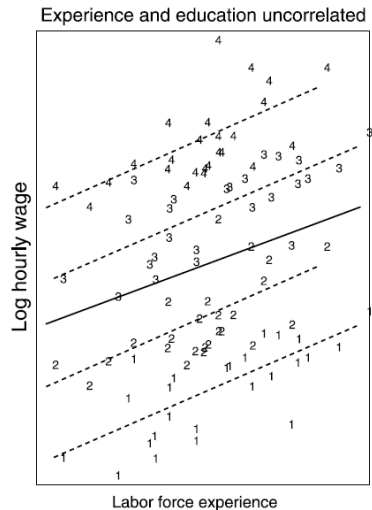
Så længe  $A_i$  er tidsinvariant kan vi med paneldata estimere  $\rho$  uden bias uden at observere  $A_i$ :

$$Y_i = \alpha_i + \lambda_t + \rho s_i + e_i \quad (4)$$

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



Source: hetbias data (see Example 3.2)



Source: nohetbias data (see Example 3.2)



»most previous studies have relied on cross-sectional data (...).While such data are often the best at hand, they come with the risk of confounding a relationship between local housing prices and support for incumbents by structural economic differences (e.g. differences in industry composition) between local contexts.This is perhaps best exemplified by the strong urban-rural gradient in local economic conditions, which would likely confound any observed cross-sectional relationship with support for the sitting government.By using panel data, we can rule out confounding by such time-invariant structural differences between local contexts by using only within-precinct/within-individual variation in local housing prices by means of fixed effects.«

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

	(1)	(2)	(3)	(4)
$\Delta$ 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	4199	4199	4199	4179
RMSE	8.405	6.749	5.715	5.325

Standard errors in parentheses

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

- LSDV FE-modeller kan implementeres med `lm()`
- clustered standardfejl kan implementeres med `multiwayvcov`-pakken

Næste gang:

- introduktion til kausal inferens
- Hariri (2012), Samii (2016)
- case-tekst: Eckles & Bakshy (kun til illustration)

Tak for i dag!