

# 11: Difference-in-differences

Videregående kvantitative metoder i studiet af politisk adfærd

Frederik Hjorth

[fh@ifs.ku.dk](mailto:fh@ifs.ku.dk)

[fghjorth.github.io](https://fghjorth.github.io)

@fghjorth

Institut for Statskundskab

Københavns Universitet

21. november 2016

- 1 Formalia
- 2 Opsamling fra sidst
- 3 Difference-in-differences designs
- 4 Implementering i R
  - Omkodning
  - Lang vs. bred form
  - Piping
- 5 Kig fremad

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*		

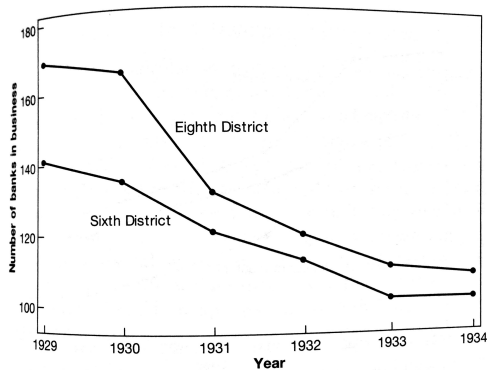
Uge	Dato	Tema	Litteratur	Case
	17/10	*Efterårsferie*		
7	24/10	Eksperimenter I	MM kap 1, GG kap 1+2	Gerber et al. (2008)
8	31/10	Eksperimenter II	GG kap 3+4+5	Gerber & Green (2000)
9	14/11	Instrumentvariable	MM kap 3	Arunachalam & Watson
10	14/11	Regressionsdiskontinuitetsdesigns	MM kap 4	Eggers & Hainmueller
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	

- RDD: cutoff og running variable
- modellering af RDD med OLS
- fastsættelse af bandwidth
- sorting
- case: Eggers & Hainmueller

Spørgsmål?

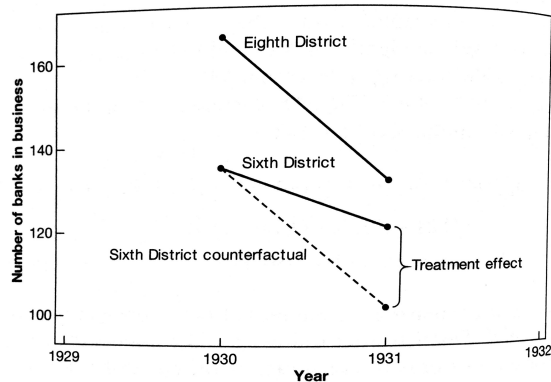
## Motiverende eksempel: pengepolitik og bankkrak

FIGURE 5.2  
Trends in bank failures in the Sixth and Eighth Federal Reserve Districts



## Motiverende eksempel: pengepolitik og bankkrak

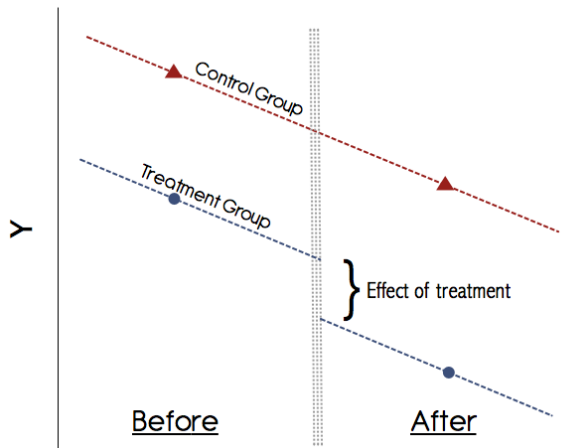
FIGURE 5.1  
Bank failures in the Sixth and Eighth Federal Reserve Districts



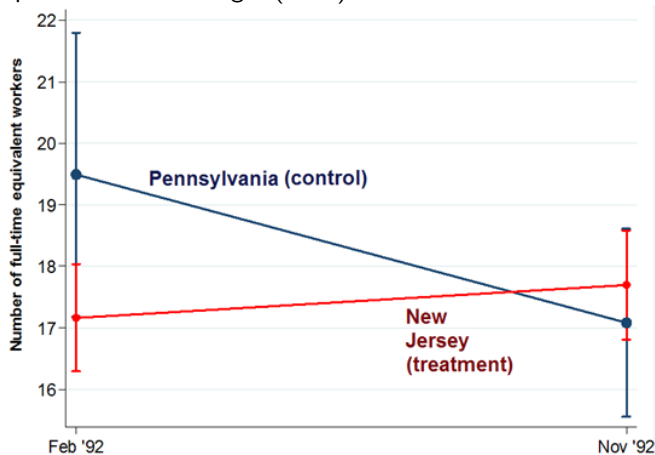
→ afgørende kilde til counterfactual: *parallel trends assumption*



## Illustration af parallel trends assumption:



## Motiverende eksempel II: Card & Krueger (1992)



Lad  $p \in \{0, 1\}$  angive om observationen er efter treatment, og  $t \in \{0, 1\}$  angive om observationen tilhører treated eller non-treated

$$\delta_{DD} = \overline{Y_{p=1,t=1}} - \overline{Y_{p=0,t=1}} - \overline{Y_{p=1,t=0}} - \overline{Y_{p=0,t=0}} \quad (1)$$

DiD i regressionsform med treatede og non-treated data 'stakket':

$$Y_{dt} = \alpha + \beta TREAT_d + \gamma POST_t + \delta_{rDD}(TREAT_d \times POST_t) + \epsilon_{dt} \quad (2)$$

En treatment-indikator i paneldata med tids- og enheds-FE har også en DiD-fortolkning:

$$Y_{st} = \alpha + \delta_{rDD} TREAT + \sum_i^N \beta_i UNIT + \sum_j^T \gamma_j TIME + \epsilon_{ij} \quad (3)$$

→ modellen kan inkludere tidsvarierende potentielle kilder til OVB

Spørgsmål?

Nyttig funktion til omkodning (fx. bestemte værdier til NA): `ifelse()`

`ifelse(<logisk betingelse>, <hvis TRUE>, <hvis FALSE>)`

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?

‘Piping’ fører et objekt gennem sekventielle funktioner forbundet med operatoren `%>%`

`%>%`  
magrittr

*Ceci n'est pas un pipe.*

→ implementeret i dplyr-pakken, kan bruges generelt vha. magrittr-pakken

Eks.: antag håbløst formateret variabel `var`

```
var<-c("1","3","2","4","1","99","3")
```

→ hvordan udfører vi mest effektivt flg.?

- ① konverter til numerisk
- ② konverter 99 til NA
- ③ tag logaritmen

Næste gang:

- 'big data' og maskinlæring
- pensum: Grimmer + Varian
- ingen case-tekst

Tak for i dag!