Hvad er der los med difference-in-differences - Del 1

Hans H. Sievertsen (h.h.sievertsen@bristol.ac.uk)

31. marts 2022

Introduction

Difference-in-Differences (DiD) is very popular

- 26 out of the 100 most cited articles in the American Economic Review 2015-2019 use DiD

Growing recent research on issues with commonly applied DiD methods

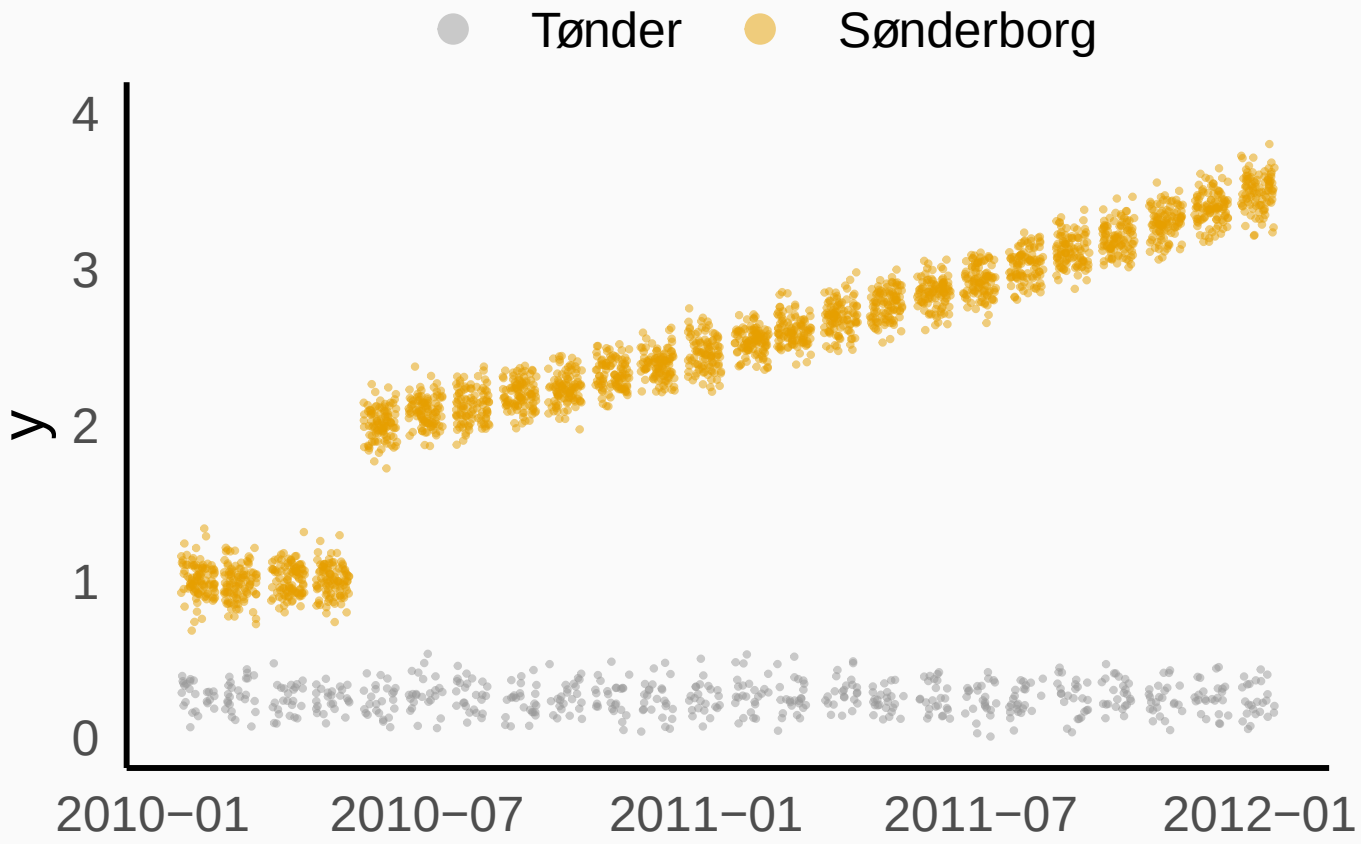
1. *DiD Estimates are oddly weighted averages of treatment effects* (even negative weighted)
   - Chaisemartin & D'Haultfoeuille (2020), Goodman-Bacon (2021), Chaisemartin & D'Haultfoeuille (2021), Callaway & Sant'Anna (2021), Sun & Abraham (2021), Borusyak, Jaravel, & Spiess (2021) [and more]
2. *Tests for // trends are under powered and lead to sample selection*
   - Roth (2021), Rambachan & Roth (2021)

Goal for today

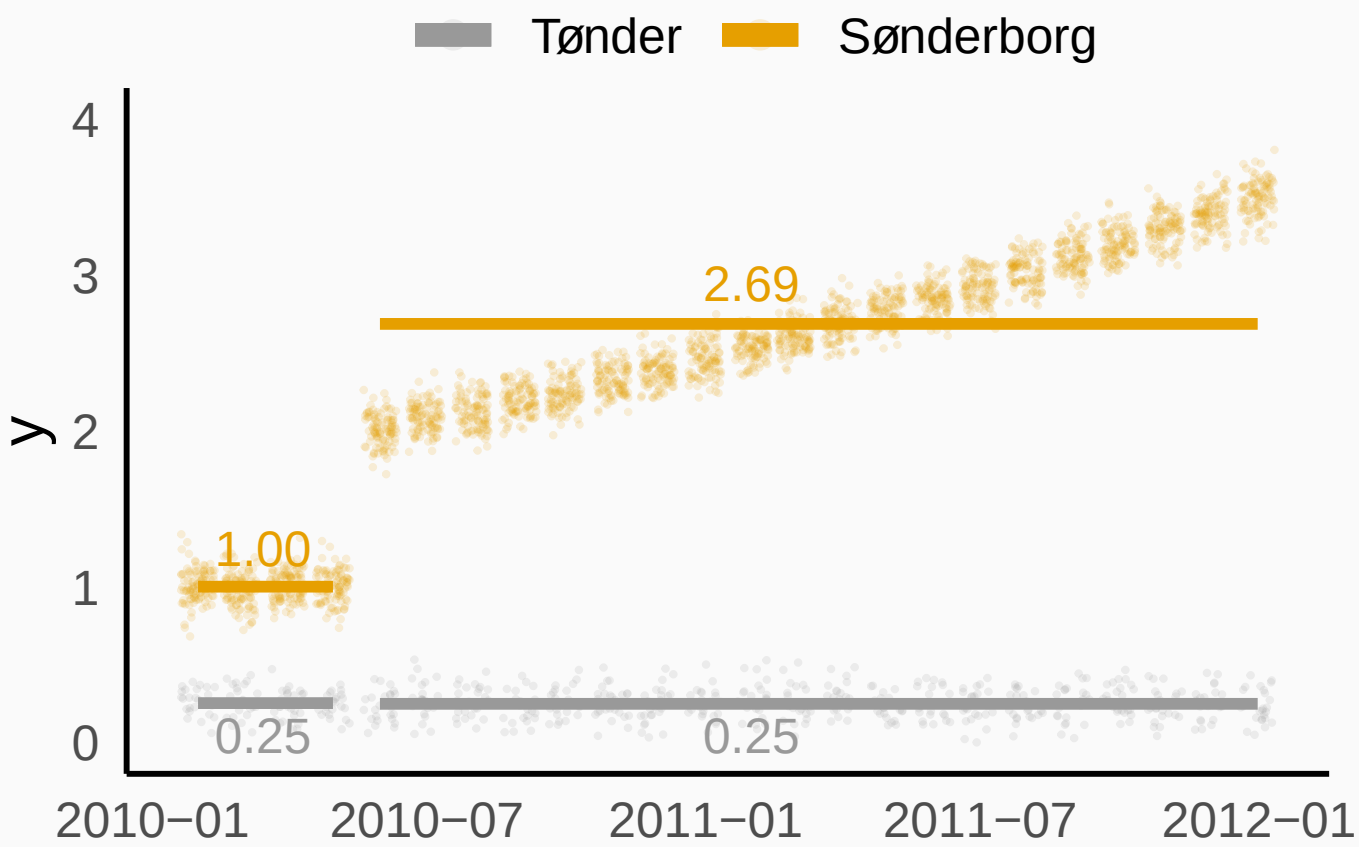
- Introduction to the issues with *1. DiD Estimates are...*
- A teaser on solutions.

Our hypothetical working example

- Sønderborg Kommune is treated with an intervention in May 2010 and onwards ($D=1$).
- Tønder Kommune is never treated ($D=0$)
- Outcome of interest y
- Data for individuals in Sønderborg (N=100) and Tønder (N=25) for the period January 2010 to December 2011.



The DiD



- Difference Sønderborg: 2.69-1.00=1.69
- Difference Tønder: 0.25-0.25=0.00
- Difference-in-Differences=1.69-0=1.69

The regression DiD

We use OLS to estimate

$y = \beta_0 + \beta_1 treated + \beta_2 after + \beta_D iDafter \times treated + u$

- *after* 1 if May 2010 or later, 0 otherwise.
- *treated* 1 if Sønderborg, 0 otherwise.

```
feols(y~treated+after+afterXtreated,data=analysisdata)
```

```
## OLS estimation, Dep. Var.: y
## Observations: 3,000
## Standard-errors: IID
##                 Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)     0.253682   0.038169  6.646306 3.5569e-11 ***
## treated         0.747120   0.042674 17.507594  < 2.2e-16 ***
## after          -0.004521   0.041812 -0.108116 9.1391e-01
## afterXtreated   1.690295   0.046747 36.158308  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.381434   Adj. R2: 0.879117
```

The more flexible regression DiD

We use OLS to estimate the Two-Way Fixed Effects (TWFE) model

$y = \alpha + \beta_{TWFE} D + \tau' t + \mu' m + u$

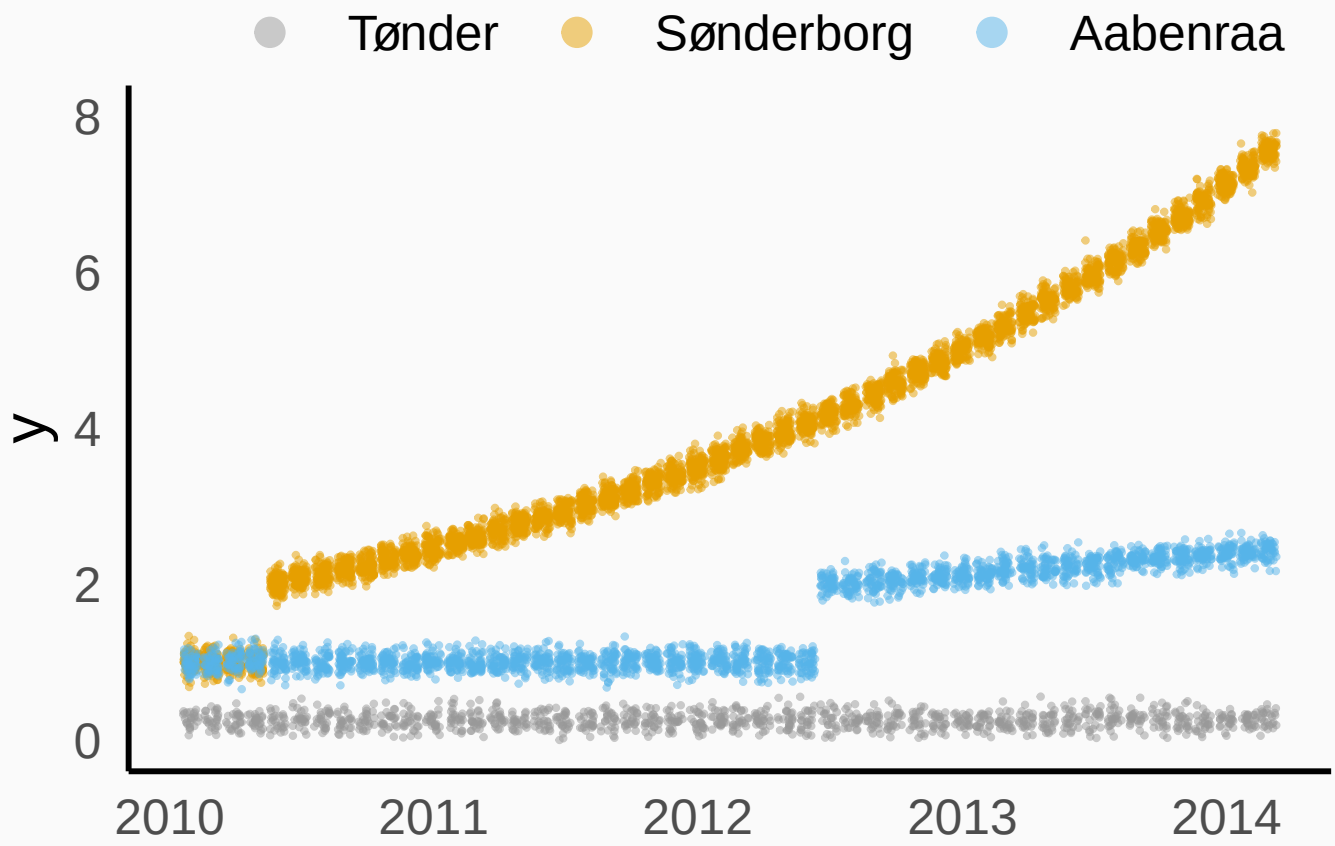
- t a vector of month dummies
- m a vector of municipality dummies
- D =1 if treated, 0 otherwise

```
feols(y~D|t+m,data=analysisdata)
```

```
## OLS estimation, Dep. Var.: y
## Observations: 3,000
## Fixed-effects: t: 24,  m: 2
## Standard-errors: Clustered (t)
##   Estimate Std. Error t value   Pr(>|t|)
## D   1.6903   0.104023 16.2493 4.2321e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.192473     Adj. R2: 0.968992
##                  Within R2: 0.63152
```

More data - and a new treated group!

- Aabenraa Kommune is treated with an intervention in June 2012 (*D=1*).
- Data for individuals in Sønderborg, Tønder, and Aabenraa (N=50) for the period January 2010 to February 2014.



Let's reestimate the DiD

We again use OLS to estimate the Two-Way Fixed Effects (TWFE) model

$y = \alpha + \beta_{TWFE} D + \tau' t + \mu' m + u$

- t a vector of month dummies
- m a vector of municipality dummies
- D =1 if in Sønderborg May 2010 or later, 1 if in Aabenraa in June 2012 or later, 0 otherwise

```
feols(y~D|t+m,data=analysisdata_update)
```

```
## OLS estimation, Dep. Var.: y
## Observations: 8,750
## Fixed-effects: t: 50,  m: 3
## Standard-errors: Clustered (t)
##    Estimate Std. Error   t value Pr(>|t|)
## D -0.188677   0.363058 -0.519688  0.60562
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.737278     Adj. R2: 0.865463
##                  Within R2: 0.003704
```


WHAT IS GOING ON

This approach is used in practice

Sievertsen & Wüst (2017)

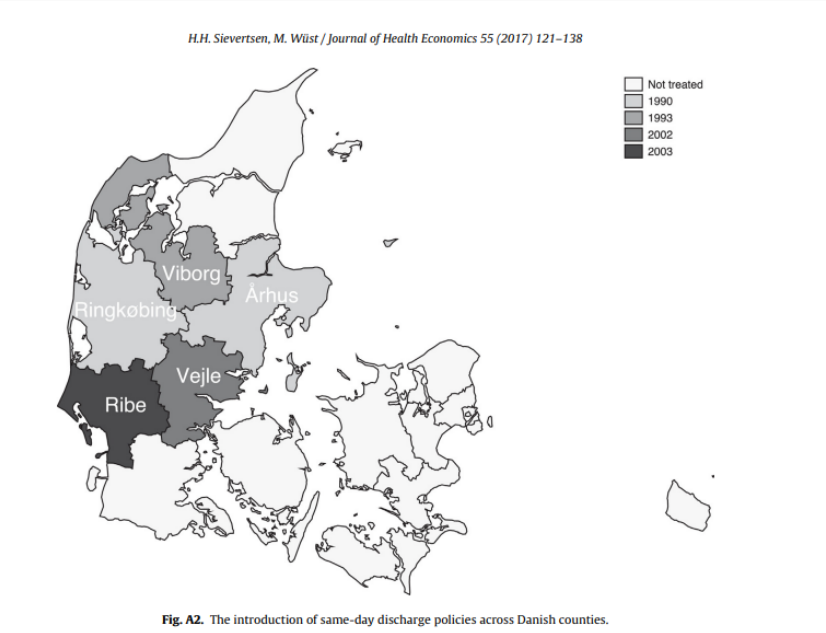


Fig. A2. The introduction of same-day discharge policies across Danish counties.

The sudden increase in the share of same-day discharge in the hospitals of five counties gives rise to exogenous variation in the probability of being discharged from hospital on the day of childbirth. Thus we compare the differences in outcomes of multiparous mothers who give birth before and after the implementation of new administrative policies in treated hospitals to the differences in outcomes of multiparous mothers who give birth in the same years in non-implementing hospitals. Our reduced form relationship is:

$$y_{ith} = \gamma_0 + \gamma_1 Post_t \times Treated_h + \boldsymbol{\alpha}'_1 \boldsymbol{X}_{ith} + \boldsymbol{\alpha}'_2 \boldsymbol{\lambda}_h + \boldsymbol{\alpha}'_3 \boldsymbol{\theta}_t + \boldsymbol{\alpha}'_4 \boldsymbol{\omega}_h \times f(\text{year}) + \epsilon_{\text{ith}} \quad (1)$$

where $\boldsymbol{X}_{ith}$ is a vector of child and mother covariates and ϵ_{ith} is a random error. As hospitals vary systematically in their population of mothers, we include $\boldsymbol{\lambda}_h$, a set of hospital indicators accounting for time-invariant differences across hospitals. Hospitals are nested in counties, the level of the policy changes. $\boldsymbol{\theta}_t$ is a set of year indicators, accounting for macro shocks. As mothers' and children's outcomes may develop differentially across hospitals (e.g. due to the faster adoption of new technologies in some hospitals or differential

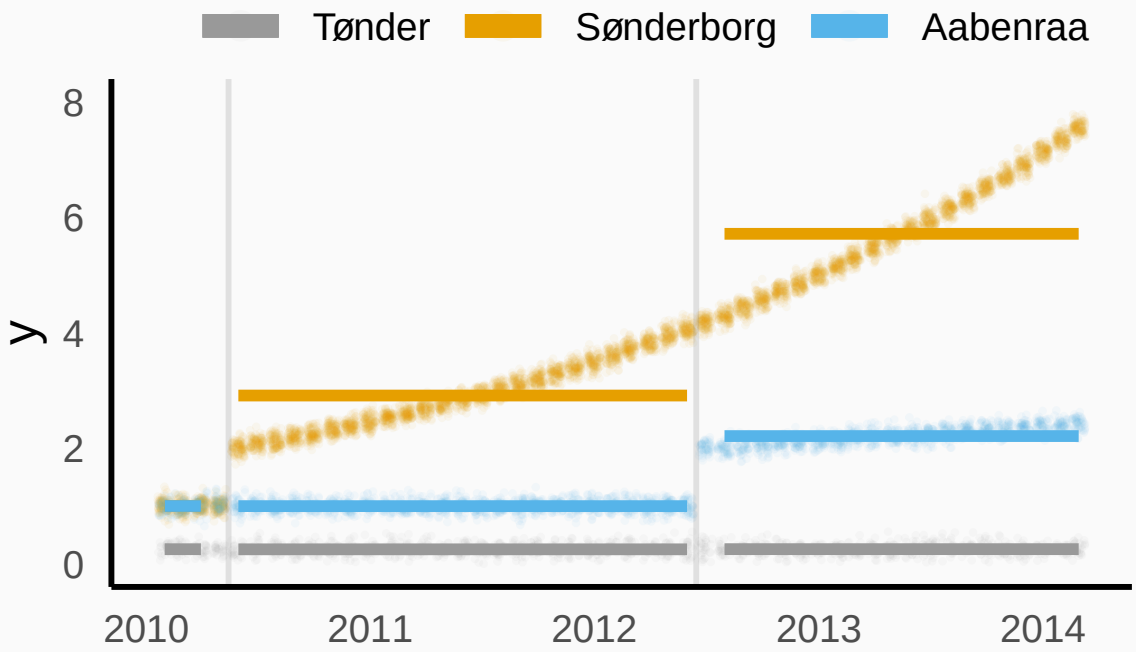
How does β_{TWFE} relate to β_{DiD}?

Big question: What is β_{TWFE} actually capturing?

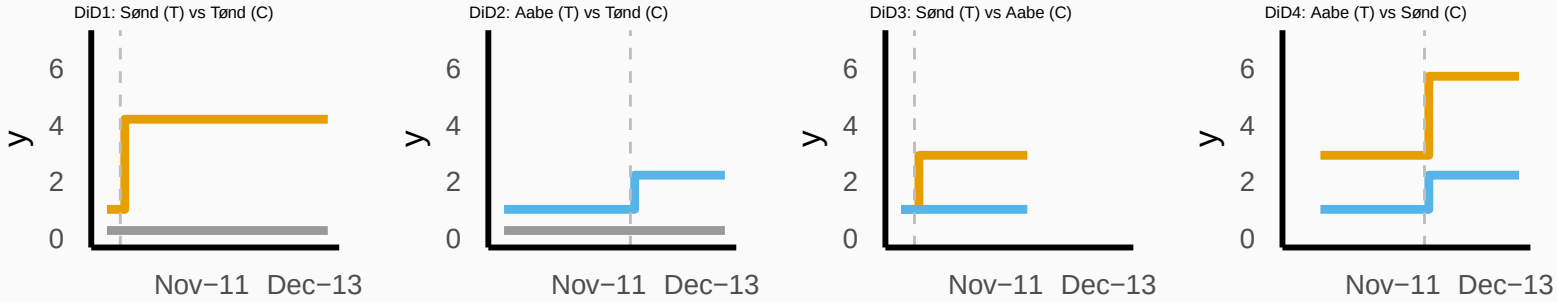
- *Goodman-Bacon (GB): "Difference-in-Differences with Variation in Treatment Timing" (2021, JoE)*
  - **Decomposition of $\hat{\beta}_{TWFE}$ in weighted $\hat{\beta}_{DiD}$s**
  - Applicable to staggered adoption designs.
- *Chaisemartin & D'Haultfoeuille (CD): "Two-way fixed effects estimators with heterogeneous treatment effects" (2020, AER)*
  - **Decomposition of $\hat{\beta}_{TWFE}$ in weighted TEs across (t, m) cells.**
  - Applicable to 2-way (e.g., group & time) fixed effects approaches.
- (There are other papers.)

The GB approach

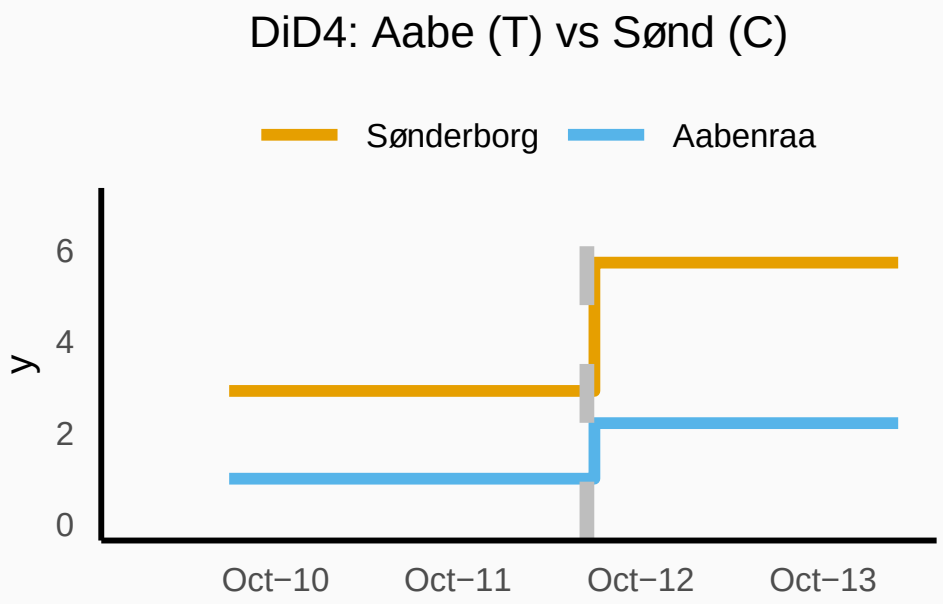
- GB: we can decompose



- into four 2X2 DiDs:



An aside: DiD4 the bad guy!



- Difference Aabenraa (Ab), After (Af) minus Before (Be)

$$Y_{Ab,Af} - Y_{Ab,Be} = (TE_{Ab,Af} + Y0_{Ab,Af}) - (Y0_{Ab,Be})$$

- Difference Sønderborg (Sø)

$$Y_{S\emptyset,Af} - Y_{S\emptyset,Be} = (TE_{S\emptyset,Af} + Y0_{S\emptyset,Af}) - (TE_{S\emptyset,Be} + Y0_{S\emptyset,Be})$$

- Difference Aabenraa minus Difference Sønderborg (under // trends assumption)
  $E[DiD] = E[Y_{Ab,Af} - Y_{Ab,Be}] = E[TE_{Ab,Af} - TE_{S\emptyset,Af} + TE_{S\emptyset,Be})]$
- Alert: $TE_{S\emptyset,Af}$ gets negative weight!

GB β_{TWFE} decomposition

β_{TWFE} ***is the weighted average across these 4 DiDs***:

$$\hat{\beta}_{TWFE} = w_1\hat{\beta}_{DiD1} + w_2\hat{\beta}_{DiD2} + w_3\hat{\beta}_{DiD3} + w_4\hat{\beta}_{DiD4}$$

- Key insight: $w \neq$ population shares, but also **depends on when a group gets treated**.
- See Theorem 1 in GB for the general expression of (3) (equation 10a in his paper) and the definition of the weights.
- GB then shows that:

$$\beta^{TWFE} = VWATT + VWCT - \Lambda ATT \tag{1}$$

where:

- $VWATT$: variance weighted ATE.
- $VWCT$: variance weighted common trends.
- ΔATT time varying treatment effect.

GB ugly weights

Theorem 1. Difference-in-Differences Decomposition Theorem

Assume that the data contain $k = 1, \ldots, K$ groups of units ordered by the time when they receive a binary treatment, $t_k^* \in (1, T]$. There may be one group, U, that includes unites that never receives treatment or are always treated. The OLS estimate, $\widehat{\beta}^{DD}$, in a two-way fixed-effects regression (2) is a weighted average of all possible two-by-two DD estimators.

$$\widehat{\beta}^{DD} = \sum_{k \neq U} s_{kU} \widehat{\beta}_{kU}^{2x2} + \sum_{k \neq U} \sum_{\ell > k} \left[s_{k\ell}^{k} \widehat{\beta}_{k\ell}^{2x2,k} + s_{k\ell}^{\ell} \widehat{\beta}_{k\ell}^{2x2,\ell} \right]. \tag{10a}$$

Where the 2x2 DD estimators are:

$$\widehat{\beta}_{kU}^{2x2} \equiv \left(\overline{y}_k^{POST(k)} - \overline{y}_k^{PRE(k)} \right) - \left(\overline{y}_U^{POST(k)} - \overline{y}_U^{PRE(k)} \right), \tag{10b}$$

$$\widehat{\beta}_{k\ell}^{2x2,k} \equiv \left(\overline{y}_k^{MID(k,\ell)} - \overline{y}_k^{PRE(k)} \right) - \left(\overline{y}_\ell^{MID(k,\ell)} - \overline{y}_\ell^{PRE(k)} \right), \tag{10c}$$

$$\widehat{\beta}_{k\ell}^{2x2,\ell} \equiv \left(\overline{y}_\ell^{POST(\ell)} - \overline{y}_\ell^{MID(k,\ell)} \right) - \left(\overline{y}_k^{POST(\ell)} - \overline{y}_k^{MID(k,\ell)} \right). \tag{10d}$$

The weights are:

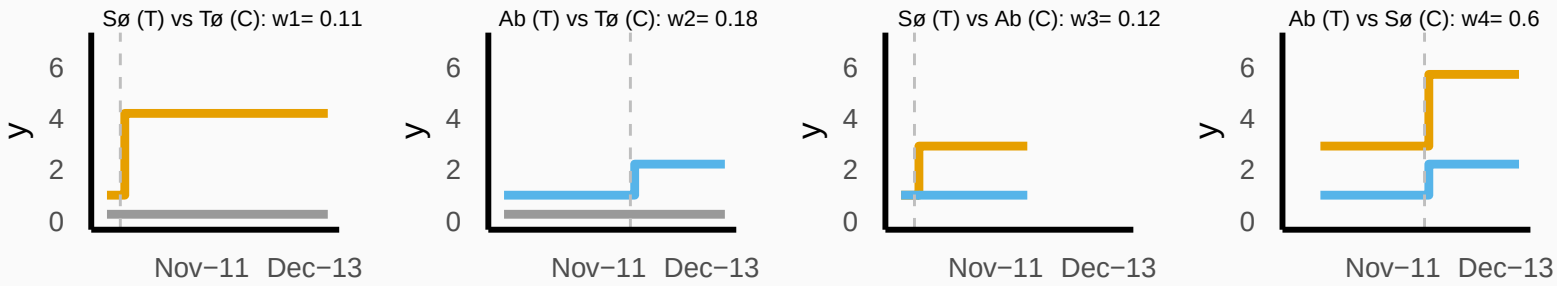
$$s_{kU} = \frac{(n_k + n_U)^2 \overbrace{n_{kU}(1 - n_{kU}) \overline{D}_k (1 - \overline{D}_k)}^{\hat{V}_{kU}^D}}{\hat{V}^D}, \tag{10e}$$

$$s_{k\ell}^{k} = \frac{\left((n_k + n_\ell)(1 - \overline{D}_\ell)\right)^2 \overbrace{n_{k\ell}(1 - n_{k\ell}) \frac{\overline{D}_k - \overline{D}_\ell}{1 - \overline{D}_\ell} \frac{1 - \overline{D}_k}{1 - \overline{D}_\ell}}^{\hat{V}_{k\ell}^{D,k}}}{\hat{V}^D}, \tag{10f}$$

$$s_{k\ell}^{\ell} = \frac{\left((n_k + n_\ell)\overline{D}_k\right)^2 \overbrace{n_{k\ell}(1 - n_{k\ell}) \frac{\overline{D}_\ell}{\overline{D}_k} \frac{\overline{D}_k - \overline{D}_\ell}{\overline{D}_k}}^{\hat{V}_{k\ell}^{D,\ell}}}{\hat{V}^D}. \tag{10g}$$

and $\sum_{k \neq U} s_{kU} + \sum_{k \neq U} \sum_{\ell > k} \left[s_{k\ell}^{k} + s_{k\ell}^{\ell} \right] = 1.$

Weights in our example



Decomposing the TWFE

$\hat{\beta}_{TWFE} = 0.106 \times 3.195 + 0.175 \times 1.21 + 0.115 \times 1.917 + 0.604 \times -1.589 = -0.19$

- Weight Sønderborg:0.11+0.12=0.23. $N_{Sø} = 100$
- Weight Aabenraa:0.18+0.60+0.78. $N_{Ab} = 50$
- Aabenraa gets a larger weight because it is treated more in the middle.
- Not necessarily a problem if you know that it is not population weighted!

Can we get just remove the bad guy?

- We know the true population weighted ATE!
  - $ATE_{S\varnothing} = 3.2$
  - $ATE_{Ab} = 1.2$
  - $ATE_{popw} = 2.83$
- Let us set $w_4 = 0$ (and adjust the weights) (this is not right way of doing this!):

$$\hat{\beta}_{TWFE} = \frac{w_1}{w_1 + w_2 + w_3}\hat{\beta}_{DiD1} + \frac{w_2}{w_1 + w_2 + w_3}\hat{\beta}_{DiD2} + \frac{w_3}{w_1 + w_2 + w_3}\hat{\beta}_{DiD3}$$

gives and overall ATE:

$$\hat{\beta}_{TWFE} = 1.95$$

- Setting also $w2 = 0$ (and adjusting weights) gives an ATE for Sønderborg

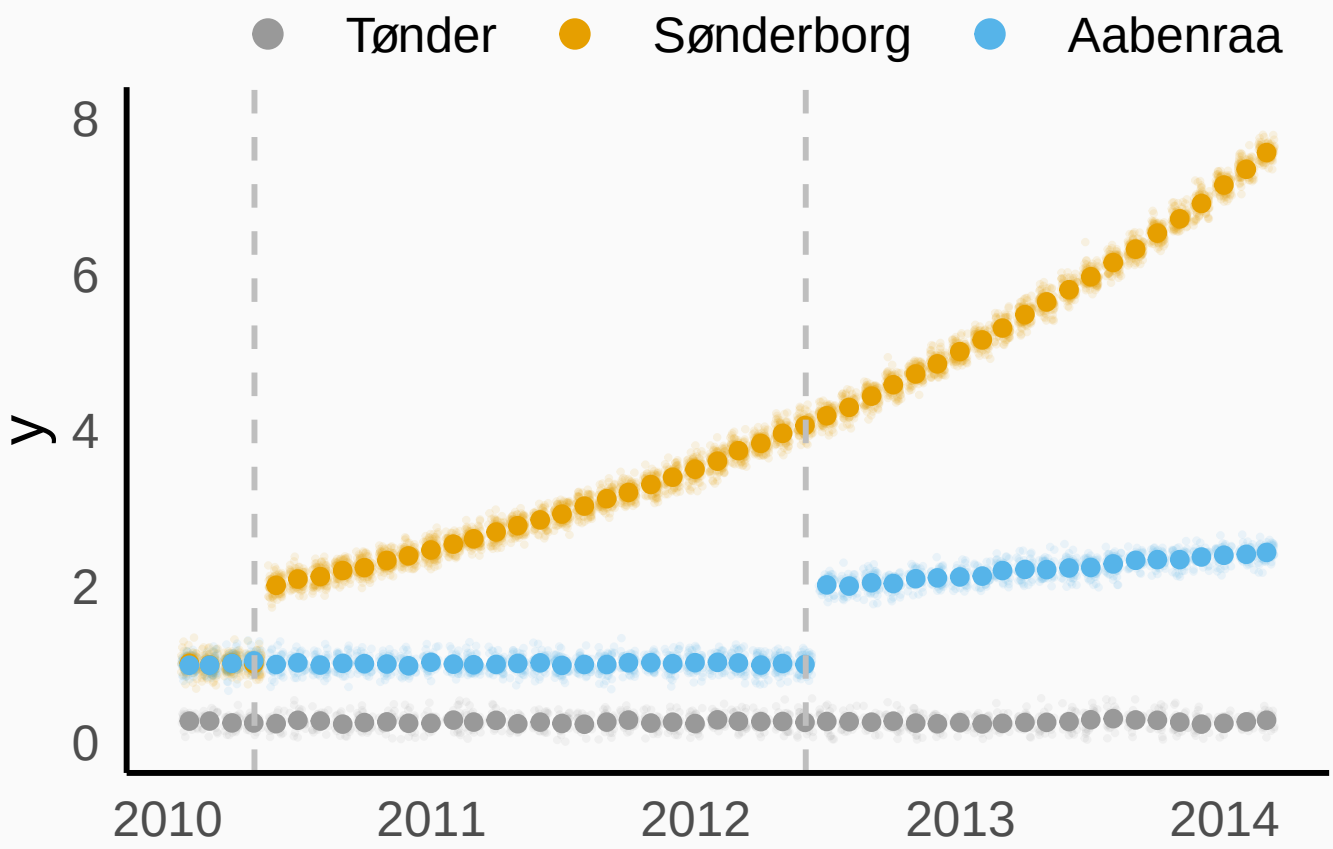
$$\hat{\beta}_{TWFE,S\varnothing} = 2.5$$

- Setting only $w2 \neq 0$ (and adjusting weights) gives an ATE for Aabenraa

$$\hat{\beta}_{TWFE,Ab} = 1.21$$

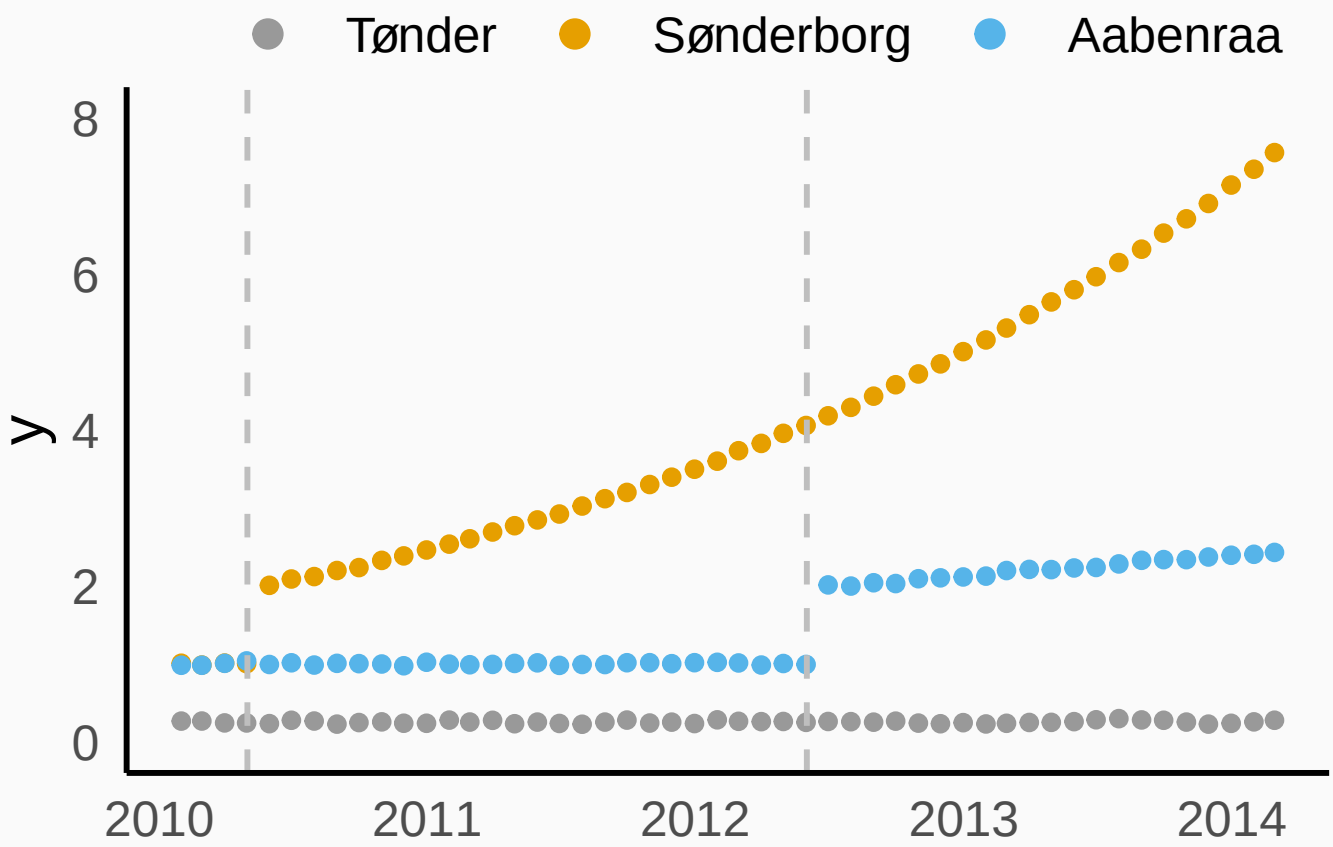
The CD decomposition

Decompose β_{TWFE} in weighted TEs across $(m, t) : D_{m,t} = 1$ cells



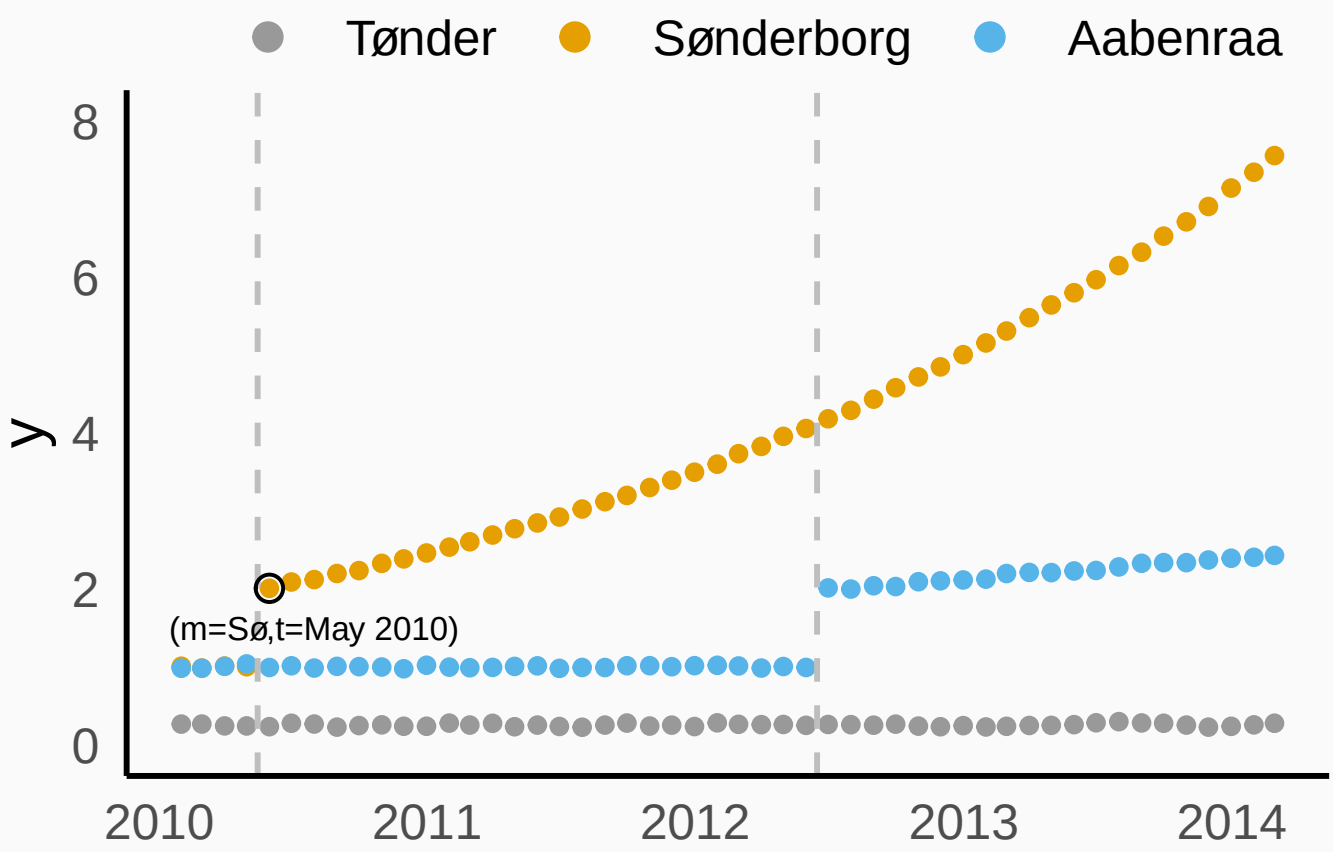
The CD decomposition

Decompose β_{TWFE} in weighted TEs across $(m, t) : D_{m,t} = 1$ cells



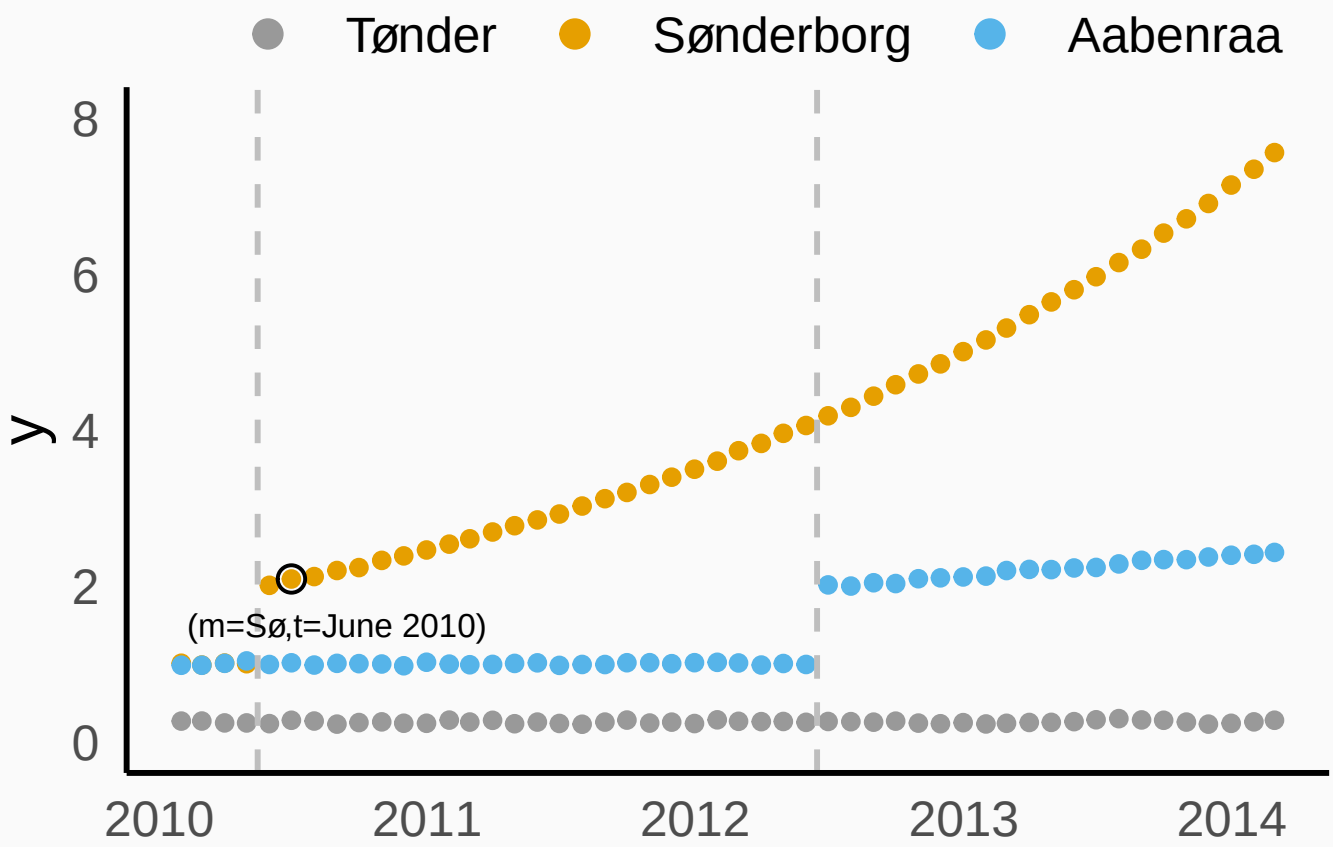
The CD decomposition

Decompose β_{TWFE} in weighted TEs across $(m, t) : D_{m,t} = 1$ cells



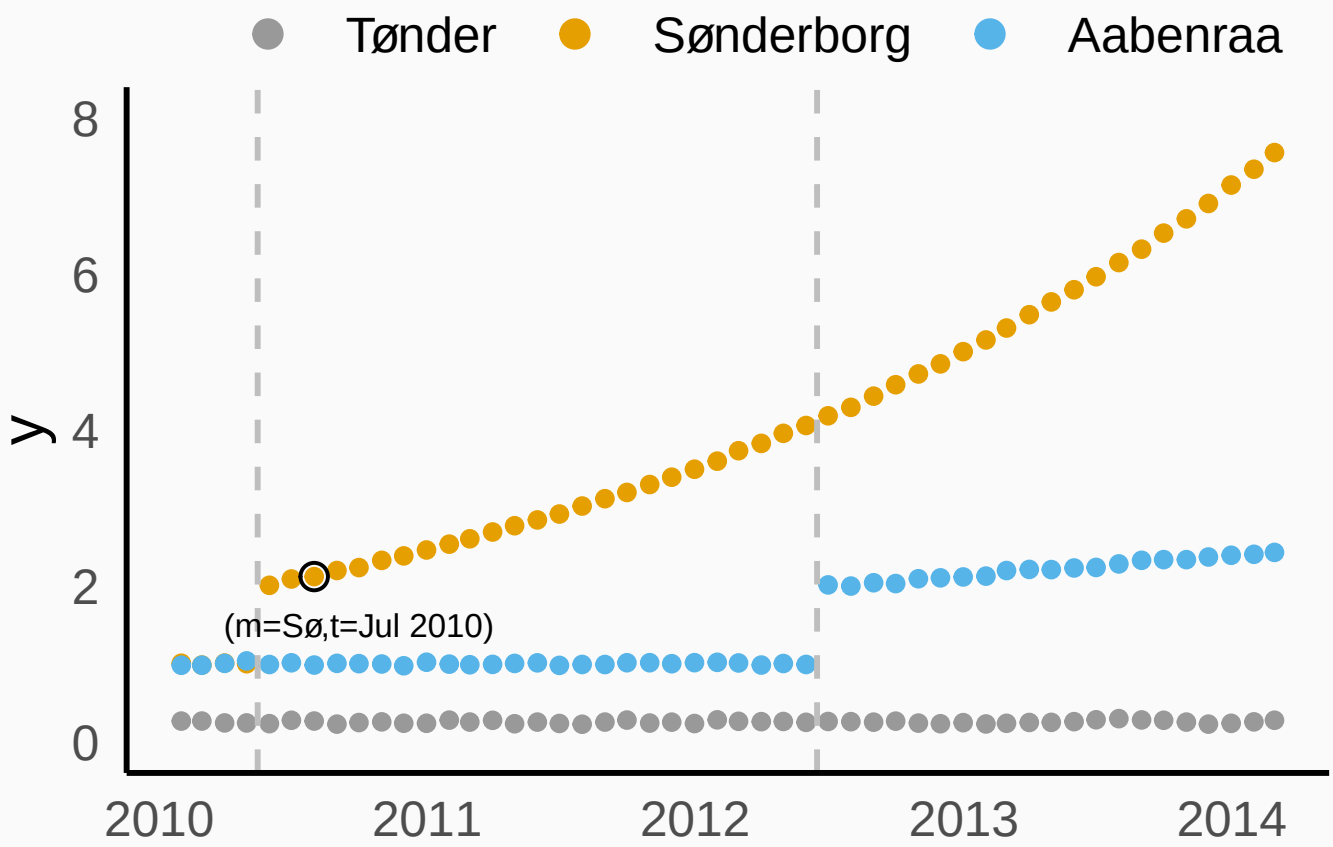
The CD decomposition

Decompose β_{TWFE} in weighted TEs across $(m, t) : D_{m,t} = 1$ cells



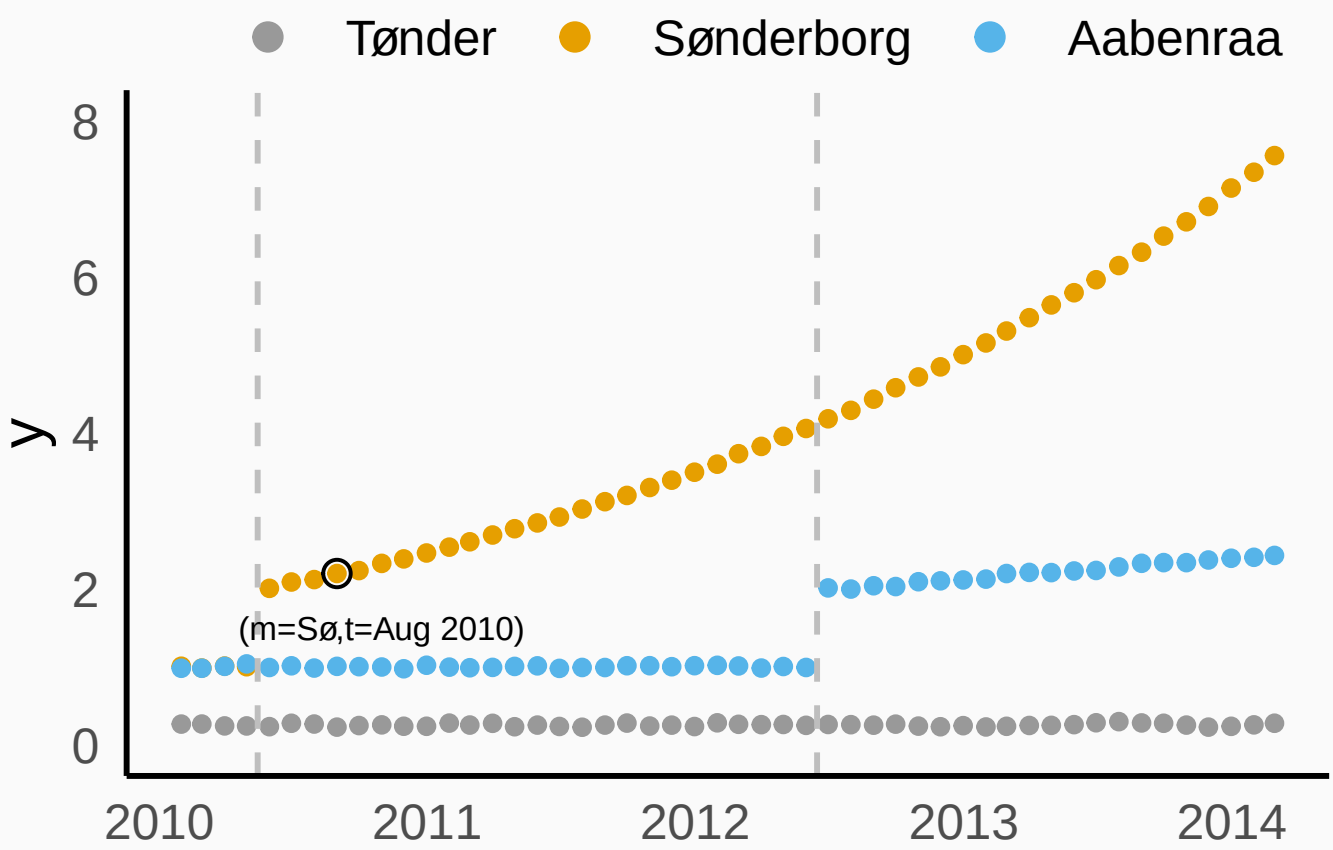
The CD decomposition

Decompose β_{TWFE} in weighted TEs across $(m, t) : D_{m,t} = 1$ cells



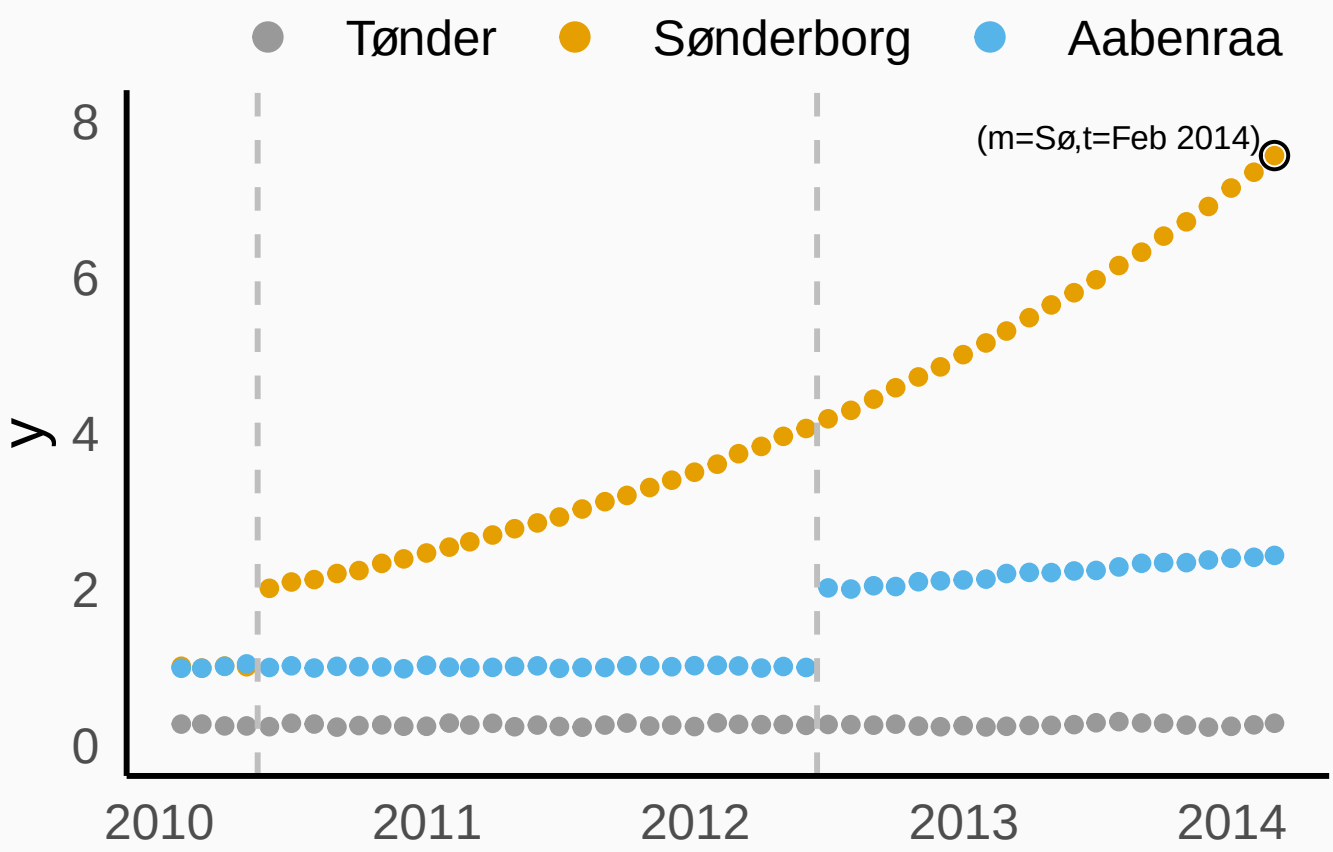
The CD decomposition

Decompose β_{TWFE} in weighted TEs across $(m, t) : D_{m,t} = 1$ cells



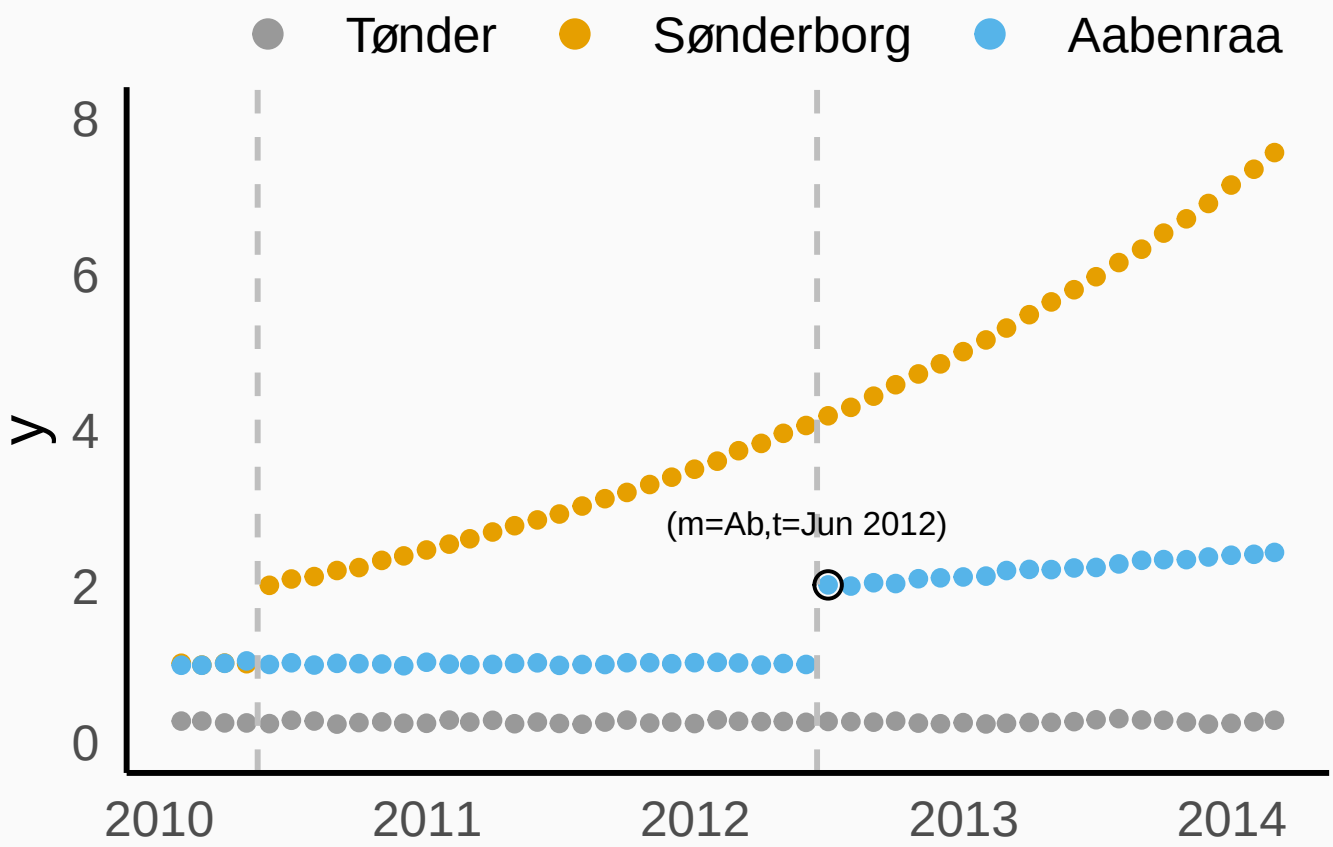
The CD decomposition

Decompose β_{TWFE} in weighted TEs across $(m, t) : D_{m,t} = 1$ cells



The CD decomposition

Decompose β_{TWFE} in weighted TEs across $(m, t) : D_{m,t} = 1$ cells



The CD decomposition

Decompose β_{TWFE} in weighted TEs across $(m, t) : D_{m,t} = 1$ cells

- β_{TWFE} is then given by

$$\beta_{TWFE} = E\left[\sum_{(m,t):D_{m,t}=1} \frac{N_{m,t}}{N_{D=1}} w_{m,t}\Lambda_{m,t}\right]$$

where

- $w_{m,t}$ is the weight on group m in period t
  - $\Lambda_{m,t}$ is the TE in group m in period t
  - $N_{D=1}$ is the number of treated units.
  - $N_{m,t}$ is the number of observations in cell m, t.

(See Theorem 1 in CD)

- How are the weights defined?

Calculating the CD weights

1. Run a regression of $D_{m,t}$ on a constant, m and t fixed effects.
2. Save the residual, $e_{g,t}$
3. Create the weight of that cell as

$$w_{m_t} = \frac{e_{m,t}}{\sum_{(m,t):D_{m,t}=1} \frac{N_{m,t}}{N_{D=1}} e_{m,t}}$$

Intuition

- Consider cell (m, t) from a group that is treated for almost the entire period and at a time where almost all cells are treated.
- We would predict that this cell is treated (because it is from a group that is mostly treated at a time that is mostly treated).
- $\Rightarrow$ small or even negative residual
- $\Rightarrow$ small or even negative weight.
- Cells that are treated when no one else is treated + for a group that is rarely treated => large weight!

An example from CD

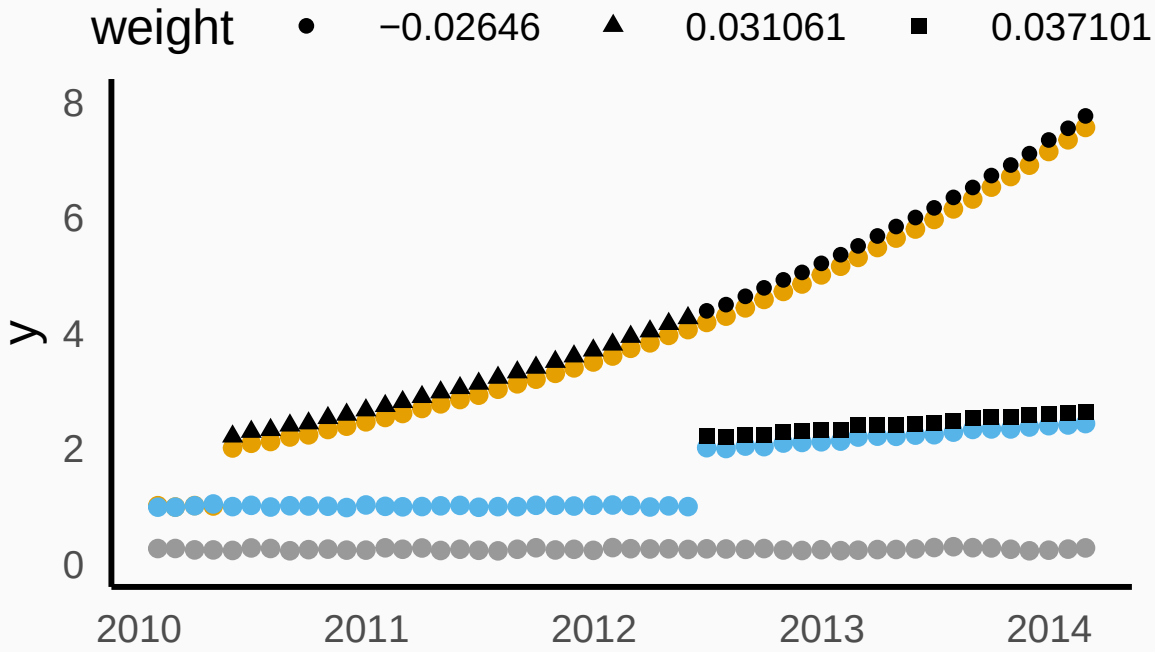
2 equally sized groups, 3 periods

- group 1 is only treated in period 3
- group 2 is treated in period 2 and 3
- The weights are then:
  - $e_{fe,1,3} = 1/6$
  - $e_{fe,2,2} = 2/6$
  - $e_{fe,2,3} = -1/6$

What do we estimate?

- So that $\beta_{TWFE} = 1/2 \times E(\Lambda_{1,3}) + 1 \times E(\Lambda_{2,2}) - 1/2 \times E(\Lambda_{2,3})$
  - If for example $E(\Lambda_{2,3}) = 4$ & $E(\Lambda_{1,3}) = E(\Lambda_{2,2}) = 1$, then $\beta_{TWFE} = -1/2$
  - Estimate is negative, although all cell's treatment effects are positive!

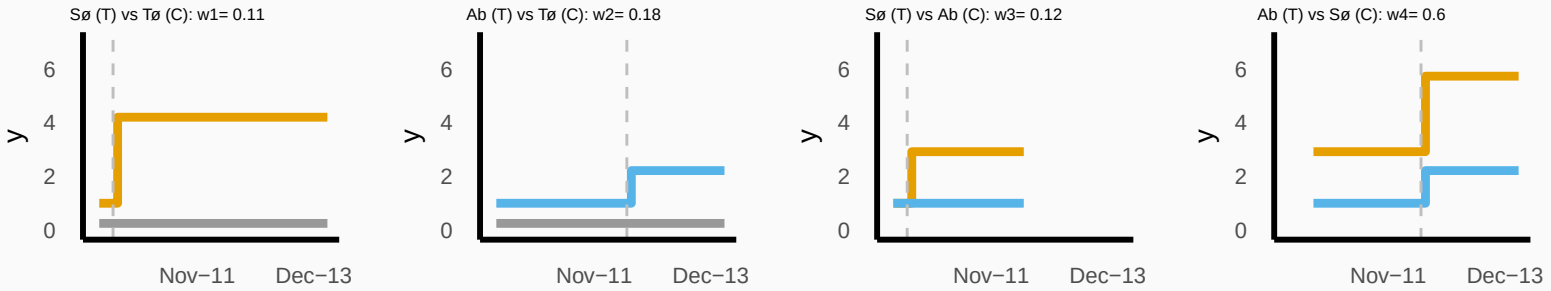
The CD weights in our example



- Sønderborg cells get negative weight after $t = 30$ because they are from a group that is mostly treated and at a time that that is mostly treated.
- Note that in contrast to GB, we cannot decompose β_{TWFE} because we don't know the TEs in the m, t cells!

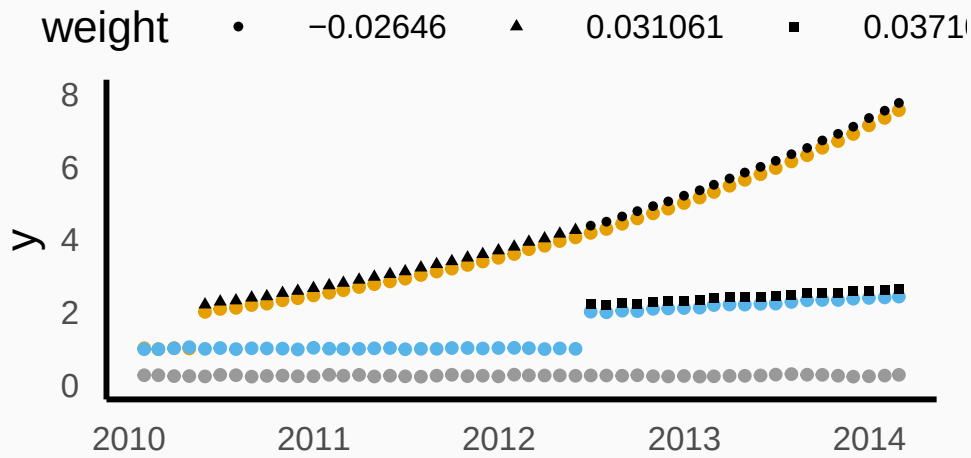
Linking GB and CD

GB weights



- Weight Sønderborg:0.11+0.12=0.23.
- Weight Aabenraa:0.18+0.60+0.78.

CD weights



- Weight Sønderborg:-0.026*21+0.031*25 =0.23.
- Weight Aabenraa:0.037*21=0.78.

So what should you do?

1. Use CD to assess whether you have negative cells!
2. If you have negative weights: calculate $\underline{\sigma}_{TWFE}$ (See Corollary 1 in CD paper)

- tells us the amount of treatment heterogeneity that is required to for the pop weighted average treatment effect to have the opposite sign of $\hat{\beta}_{TWFE}$.
- So if $\underline{\sigma}_{TWFE}$ is very small you should be worried!
- Are treatment effects likely to be correlated with weights? Check if weights are correlated with covariates.

3. If things still look bad: use one of the new estimators.

- For example by CD:

$$DID_M = \sigma_{t=2}^{T} \left(\overbrace{\frac{N_{1,0,t}}{N_S} DID_{+,t}}^{\text{"joiners"}} + \overbrace{\frac{N_{0,1,t}}{N_S} DID_{-,t}}^{\text{"leavers"}} \right)$$

- Or one of the many other new estimators!
  - Callaway & Sant'Anna (2021)/Sun & Abraham (2021)
  - General idea: Aggregate groups, choose your weights, and use good comparisons only (more about that next time).

Teaser 1: the CD estimator on our data

```
did_multiplegt(df=df, "y", "G", "t", "D")
```

```
## $effect
## treatment
##  1.029583
##
## $N_effect
## [1] 250
##
## $N_switchers_effect
## [1] 150
```

Also available in Stata!

Teaser 2: a real example

Valente, Sievertsen, & Puri (2021)

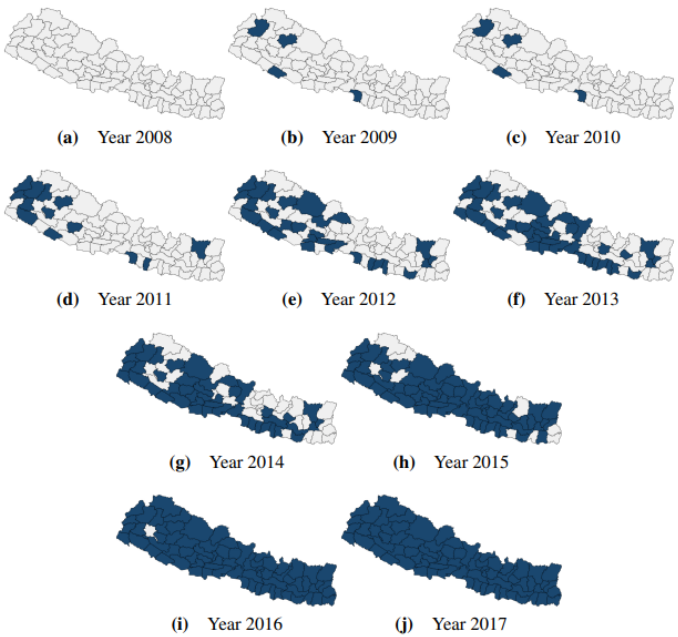


Figure 1: CHX cord application roll-out across districts over time (adopted CHX=blue).

Table 4: Assessing the role of negative weights in the two-way fixed effects estimator.

	Coefficient (1)	N_w (2)	$N_{w<0}$ (3)	$\sum_{w<0} w$ (4)	σ_{FE} (5)
Baseline	-0.028**	862	140	-0.053974	0.0307
	(0.011)				
Iteration 1	-0.029***	662	46	-0.008862	0.0394
	(0.010)				
Iteration 2	-0.028***	600	12	-0.000816	0.0419
	(0.010)				
Iteration 3	-0.027**	582	8	-0.000241	0.0425
	(0.011)				
Iteration 4	-0.026**	570	4	-0.000034	0.0415
	(0.011)				
Iteration 5	-0.026**	565	0	0.000000	NA
	(0.011)				

Notes: All models are estimated on the sample: P(home birth)>0.5, with the full set of demographic, SES, and program controls as well as district and month of birth fixed effects. Demographic controls include birth order (three indicators), five year maternal age group indicators, and gender. SES controls include education (three indicators), wealth (four indicators), rural indicator, altitude quintile indicators, and a Dalit ethnicity indicator. Program controls include controls for the CB-NCP and CB-IMNCI health programs. The place of delivery is predicted using the linear probability shown in Appendix Table A.2. Bootstrapped standard errors based on 200 iterations and clustered at the district level in parentheses. Asterisks indicate significance at the following levels * p<0.1, ** p<0.05, and *** p<0.01.

21

Summary

- It is common to generalize DiD with TWFE (include treatment dummy and control for time and region)
- The estimated coefficient is a weighted average of TEs
- The weights are not proportional to population sizes, but also depend on treatment timing!
- This is only an issue with heterogeneous treatment effects (but who believes in homogeneous treatment effects?)
- Worrying: if early adopters benefit more and treatment effects grow over time $=>$ underestimate effects!
- CD decomposition is more general than GB, but GB gives both weights and TEs.

Next time: Stata and R solutions