Module 4: Difference-in-Differences and Effects of Medicaid Expansion

Part 3: Understanding Difference-in-Differences

Ian McCarthy | Emory University
Econ 470 & HLTH 470

The Idea of DD

Setup

Want to estimate $E[Y_1(1) - Y_0(1)|D = 1]$

	Post-period	Pre-period
Treated	$E(Y_1(1)\|D = 1)$	$E(Y_0(0)\|D = 1)$
Control	$E(Y_0(1)\|D = 0)$	$E(Y_0(0)\|D = 0)$

Problem: We don't see $E[Y_0(1)|D = 1]$

Setup

Want to estimate $E[Y_1(1) - Y_0(1)|D = 1]$

	Post-period	Pre-period
Treated	$E(Y_1(1)|D = 1)$	$E(Y_0(0)|D = 1)$
Control	$E(Y_0(1)|D = 0)$	$E(Y_0(0)|D = 0)$

Strategy 1: Estimate $E[Y_0(1)|D = 1]$ using $E[Y_0(0)|D = 1]$ (before treatment outcome used to estimate post-treatment)

Setup

Want to estimate $E[Y_1(1) - Y_0(1)|D = 1]$

	Post-period	Pre-period
Treated	$E(Y_1(1)|D = 1)$	$E(Y_0(0)|D = 1)$
Control	$E(Y_0(1)|D = 0)$	$E(Y_0(0)|D = 0)$

Strategy 2: Estimate $E[Y_0(1)|D = 1]$ using $E[Y_0(1)|D = 0]$ (control group used to predict outcome for treatment)

Setup

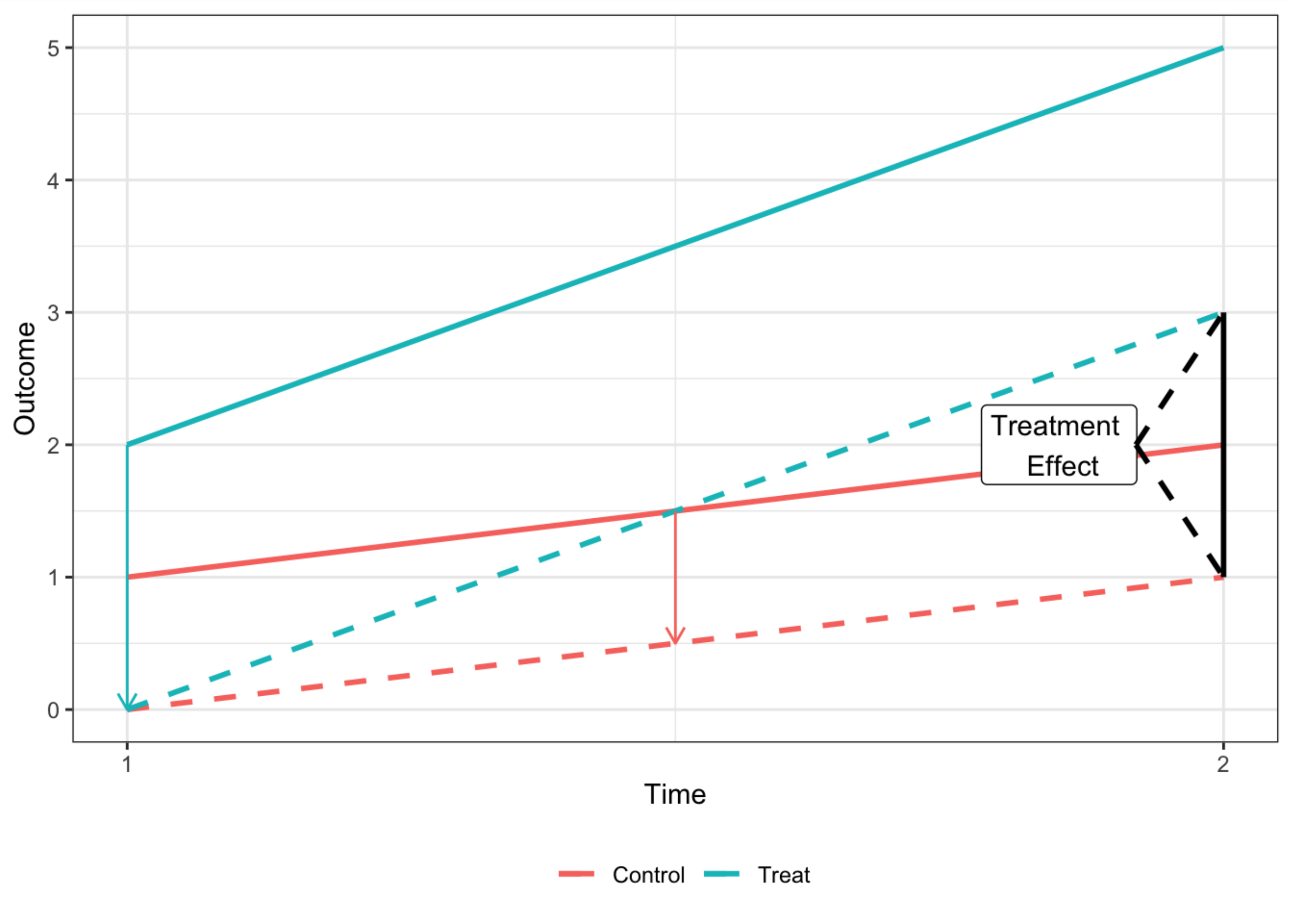
Want to estimate $E[Y_1(1) - Y_0(1)|D = 1]$

	Post-period	Pre-period
Treated	$E(Y_1(1)\|D = 1)$	$E(Y_0(0)\|D = 1)$
Control	$E(Y_0(1)\|D = 0)$	$E(Y_0(0)\|D = 0)$

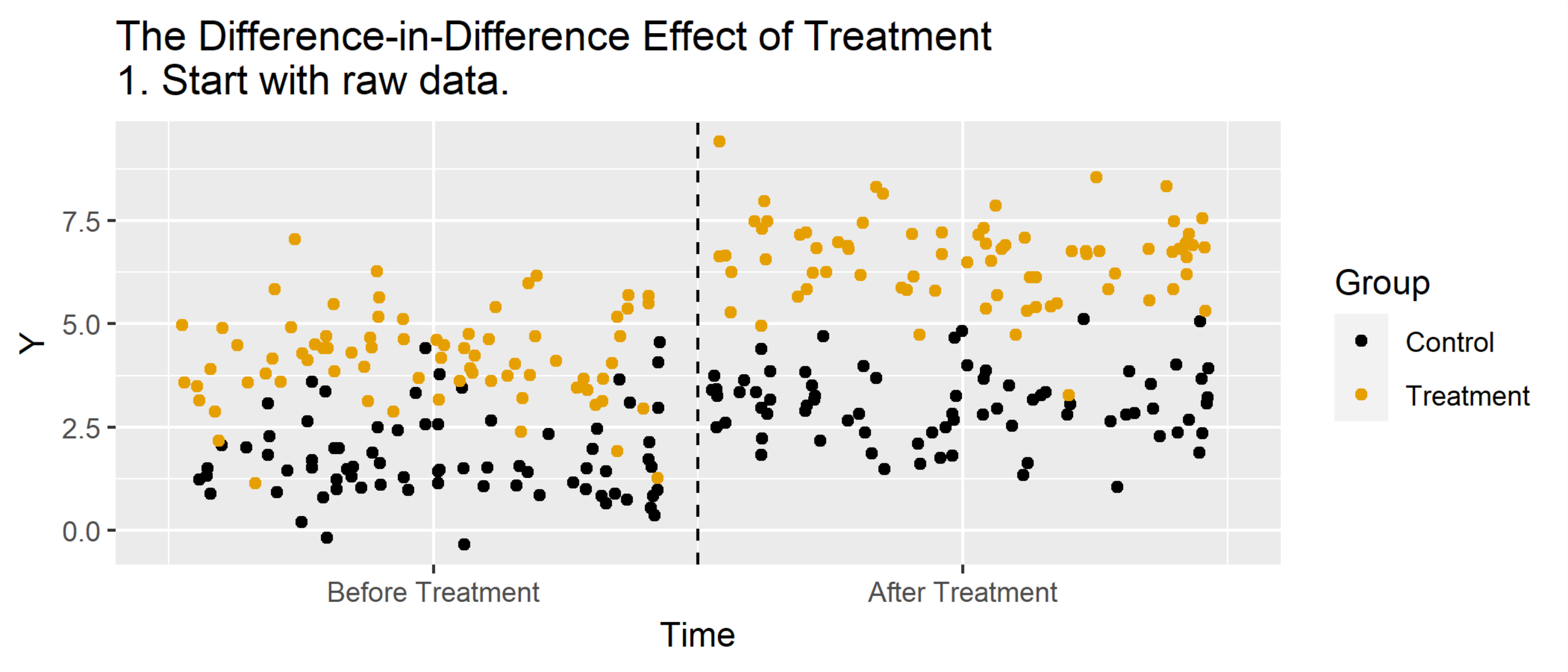
Strategy 3: DD estimate...

Estimate $E[Y_1(1)|D = 1] - E[Y_0(1)|D = 1]$ using $E[Y_0(1)|D = 0] - E[Y_0(0)|D = 0]$ (pre-post difference in control group used to predict difference for treatment group)

Graphically



Animations



The Difference-in-Difference Effect of Treatment
1. Start with raw data.
Y
7.5
5.0
2.5
0.0
Before Treatment
After Treatment
Time
Group
Control
Treatment

Average Treatment Effects with DD

Estimation

Key identifying assumption is that of *parallel trends*

$$E[Y_0(1) - Y_0(0)|D = 1] = E[Y_0(1) - Y_0(0)|D = 0]$$

Estimation

Sample means:

$$
\begin{aligned}
E[Y_1(1) - Y_0(1)|D=1] = \quad & (E[Y(1)|D=1] - E[Y(1)|D=0]) \\
& - (E[Y(0)|D=1] - E[Y(0)|D=0])
\end{aligned}
$$

Estimation

Regression:

$$y_{it} = \alpha + \beta D_i + \lambda \times Post_t + \delta \times D_i \times Post_t + \varepsilon_{it}$$

	After	**Before**	**After - Before**
Treated	$\alpha + \beta + \lambda + \delta$	$\alpha + \beta$	$\lambda + \delta$
Control	$\alpha + \lambda$	α	λ
Treated - Control	$\beta + \delta$	β	δ

Simulated data

```
N <- 5000
dd.dat <- tibble(
  d = (runif(N, 0, 1)>0.5),
  time_pre = "pre",
  time_post = "post"
)

dd.dat <- pivot_longer(dd.dat, c("time_pre","time_post"), values_to="time") %>%
  select(d, time) %>%
  mutate(t=(time=="post"),
         y.out=1.5+3*d + 1.5*t + 6*d*t + rnorm(N*2,0,1))
```

Mean differences

```
dd.means ← dd.dat %>% group_by(d, t) %>% summarize(mean_y = mean(y.out))
knitr::kable(dd.means, col.names=c("Treated","Post","Mean"), format="html")
```

Treated	Post	Mean
FALSE	FALSE	1.536235
FALSE	TRUE	3.014374
TRUE	FALSE	4.515127
TRUE	TRUE	11.970610

Mean differences

In this example:

- $E[Y(1)|D=1] - E[Y(1)|D=0]$ is 8.9562357
- $E[Y(0)|D=1] - E[Y(0)|D=0]$ is 2.9788923

So the ATT is 5.9773434

Regression estimator

```
library(modelsummary)
dd.est <- lm(y.out ~ d + t + d*t, data=dd.dat)
modelsummary(dd.est, gof_map=NA, coef_omit='Intercept')
```

	(1)
dTRUE	2.979
	(0.028)
tTRUE	1.478
	(0.028)
dTRUE × tTRUE	5.977
	(0.040)

Seeing things in action

Application

- Try out some real data on Medicaid expansion following the ACA
- **Question:** Did Medicaid expansion reduce uninsurance?

Step 1: Look at the data

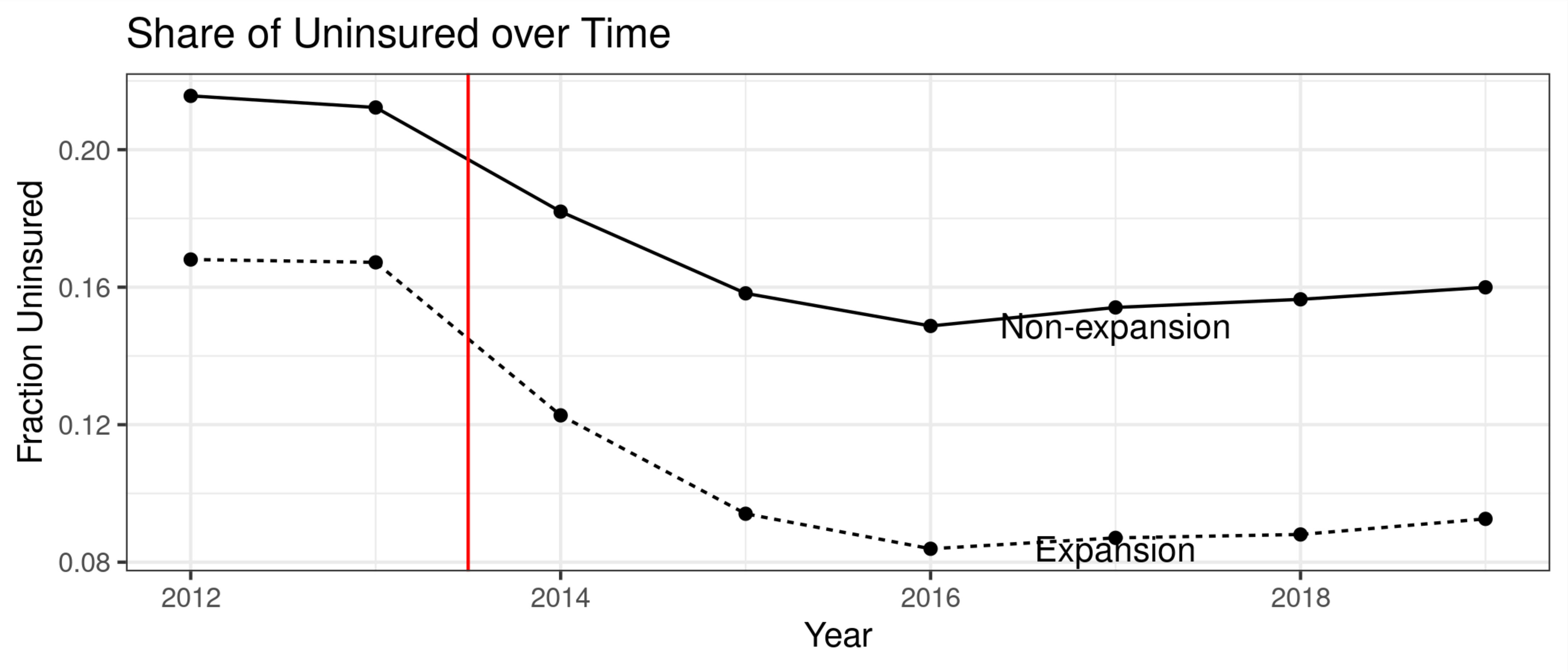
Stata

```
insheet using "data/acs_medicaid.txt", clear
gen perc_unins=uninsured/adult_pop
keep if expand_year=="2014" | expand_year=="NA"
drop if expand_ever=="NA"
collapse (mean) perc_unins, by(year expand_ever)
graph twoway (connected perc_unins year if expand_ever==
  (connected perc_unins year if expand_ever=="TRUE", col
  xline(2013.5) ///
    ytitle("Fraction Uninsured") xtitle("Year") legend(o
```

R

```
library(tidyverse)
# mcaid.data <- read_tsv("https://raw.githubusercontent.
mcaid.data <- read_tsv("../data/acs_medicaid.txt")
ins.plot.dat <- mcaid.data %>% filter(expand_year==2014
  mutate(perc_unins=uninsured/adult_pop) %>%
  group_by(expand_ever, year) %>% summarize(mean=mean(pe

ins.plot <- ggplot(data=ins.plot.dat, aes(x=year,y=mean,
  geom_line() + geom_point() + theme_bw() +
  geom_vline(xintercept=2013.5, color="red") +
  geom_text(data = ins.plot.dat %>% filter(year == 2016)
            aes(label = c("Non-expansion","Expansion"),
                x = year + 1,
                y = mean)) +
  guides(linetype="none") +
  labs(
    x="Year",
    y="Fraction Uninsured",
    title="Share of Uninsured over Time"
  )
```

Step 1: Look at the data



Step 2: Estimate effects

Interested in δ from:

$$y_{it} = \alpha + \beta \times Post_t + \lambda \times Expand_i + \delta \times Post_t \times Expand_i + \varepsilon_{it}$$

Stata

```
insheet using "data/acs_medicaid.txt", clear
gen perc_unins=uninsured/adult_pop
keep if expand_year=="2014" | expand_year=="NA"
drop if expand_ever=="NA"
gen post=(year>=2014)
gen treat=(expand_ever=="TRUE")
gen treat_post=(expand=="TRUE")

reg perc_unins treat post treat_post

*also try didregress
```

R

```
library(tidyverse)
library(modelsummary)
mcaid.data <- read_tsv("../data/acs_medicaid.txt")
reg.dat <- mcaid.data %>% filter(expand_year==2014 | is.
  mutate(perc_unins=uninsured/adult_pop,
         post = (year>=2014),
         treat=post*expand_ever)

dd.ins.reg <- lm(perc_unins ~ post + expand_ever + post*
```

Step 2: Estimate effects

	DD (2014)
postTRUE	−0.054
	(0.003)
expand_everTRUE	−0.046
	(0.016)
postTRUE × expand_everTRUE	−0.019
	(0.007)

Final DD thoughts

- Key identification assumption is **parallel trends**
- Inference: Typically want to cluster at unit-level to allow for correlation over time within units, but problems with small numbers of treated or control groups:
    - Conley-Taber CIs
    - Wild cluster bootstrap
    - Randomization inference
- "Extra" things like propensity score weighting and doubly robust estimation

DD and TWFE

What is TWFE?

- Just a shorthand for a common regression specification
- Fixed effects for each unit and each time period, γ_i and γ_t
- More general than 2x2 DD but same result

What is TWFE?

Want to estimate δ:

$$y_{it} = \alpha + \delta D_{it} + \gamma_i + \gamma_t + \varepsilon_{it},$$

where γ_i and γ_t denote a set of unit i and time period t dummy variables (or fixed effects).

TWFE in Practice

2x2 DD

```
library(tidyverse)
library(modelsummary)
mcaid.data <- read_tsv("../data/acs_medicaid.txt")
reg.dat <- mcaid.data %>% filter(expand_year==2014 | is.na(expand_year), !is.na(expand_ever)) %>%
  mutate(perc_unins=uninsured/adult_pop,
         post = (year>=2014),
         treat=post*expand_ever)
m.dd <- lm(perc_unins ~ post + expand_ever + treat, data=reg.dat)
```

TWFE

```
library(fixest)
m.twfe <- feols(perc_unins ~ treat | State + year, data=reg.dat)
```

TWFE in Practice

	DD	TWFE
postTRUE	−0.054	
	(0.003)	
expand_everTRUE	−0.046	
	(0.016)	
treat	−0.019	−0.019
	(0.007)	(0.007)

Event Studies

What is an event study?

Event study is poorly named:

- In finance, even study is just an *interrupted time series*
- In econ and other areas, we usually have a treatment/control group *and* a break in time

What is an event study?

- Allows for heterogeneous effects over time (maybe effects phase in over time or dissipate)
- Visually very appealing
- Offers easy evidence against or consistent with parallel trends assumption

What is an event study?

Estimate something akin to…

$$y_{it} = \gamma_i + \gamma_t + \sum_{\tau=-q}^{-2} \delta_\tau D_{i\tau} + \sum_{\tau=0}^{m} \delta_\tau D_{i\tau} + \beta x_{it} + \epsilon_{it},$$

where q captures the number of periods before the treatment occurs and m captures periods after treatment occurs.

How to do an event study?

1. Create all treatment/year interactions
2. Regressions with full set of interactions and group/year FEs
3. Plot coefficients and standard errors

Things to address

1. "Event time" vs calendar time
2. Define baseline period
3. Choose number of pre-treatment and post-treatment coefficients

Event time vs calendar time

Essentially two "flavors" of event studies

1. Common treatment timing
2. Differential treatment timing

Define baseline period

- Must choose an "excluded" time period (as in all cases of group dummy variables)
- Common choice is $t = -1$ (period just before treatment)
- Easy to understand with calendar time
- For event time...manually set time to $t = -1$ for all untreated units

Number of pre-treatment and post-treatment

- On event time, sometimes very few observations for large lead or lag values
- Medicaid expansion example: Late adopting states have fewer post-treatment periods
- Norm is to group final lead/lag periods together

Commont treatment timing

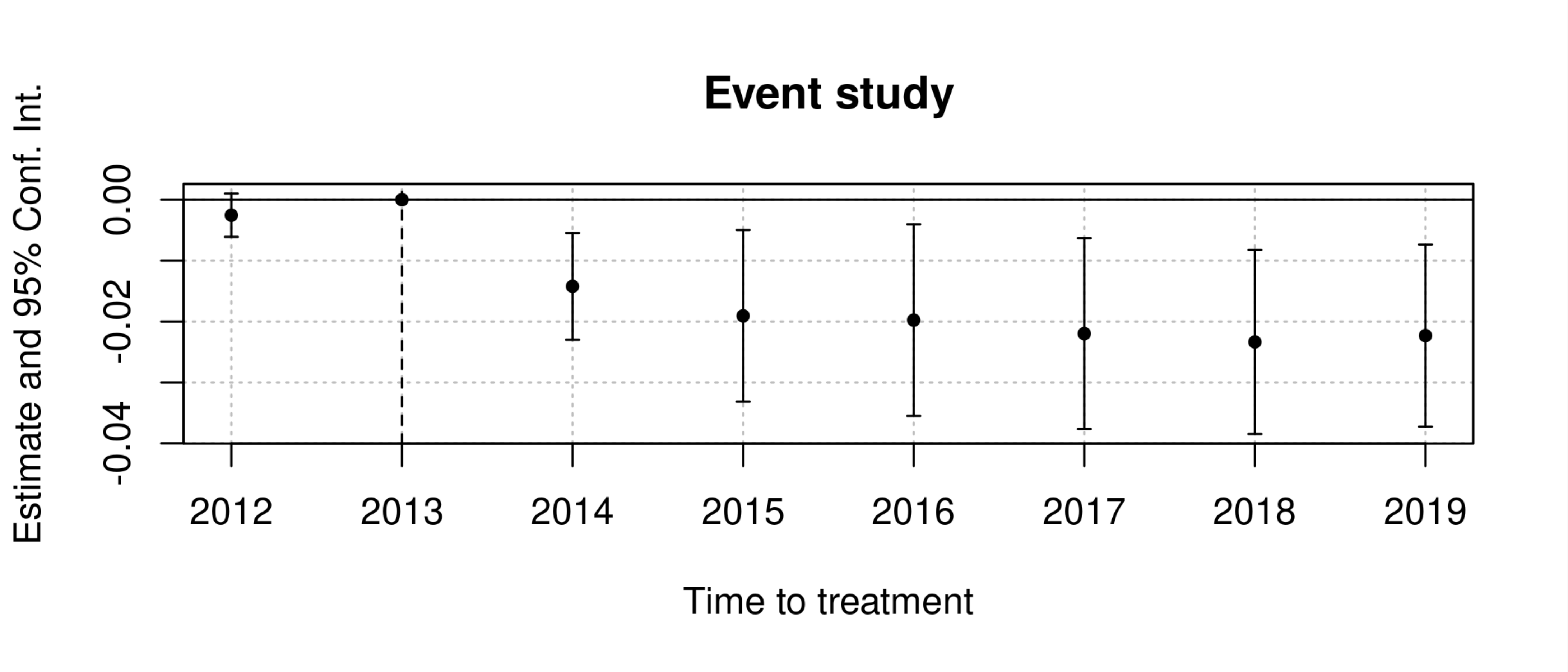
Stata

```
ssc install reghdfe

insheet using "data/acs_medicaid.txt", clear
gen perc_unins=uninsured/adult_pop
keep if expand_year=="2014" | expand_year=="NA"
drop if expand_ever=="NA"
gen post=(year>=2014)
gen treat=(expand_ever=="TRUE")
gen treat_post=(expand=="TRUE")

reghdfe perc_unins treat##ib2013.year, absorb(state)
gen coef = .
gen se = .
forvalues i = 2012(1)2018 {
    replace coef = _b[1.treat#`i'.year] if year == `i'
    replace se = _se[1.treat#`i'.year] if year == `i'
}

 Make confidence intervals
gen ci_top = coef+1.96*se
gen ci_bottom = coef - 1.96*se
```

R

```
library(tidyverse)
library(modelsummary)
library(fixest)
mcaid.data <- read_tsv("../data/acs_medicaid.txt")
reg.dat <- mcaid.data %>%
  filter(expand_year==2014 | is.na(expand_year), !is.na(
  mutate(perc_unins=uninsured/adult_pop,
         post = (year>=2014),
         treat=post*expand_ever)

mod.twfe <- feols(perc_unins~i(year, expand_ever, ref=20
                  cluster=~State,
                  data=reg.dat)
```

Common treatment timing



Differential treatment timing

- Now let's work with the full Medicaid expansion data
- Includes late adopters
- Requires putting observations on "event time"

Differential treatment timing

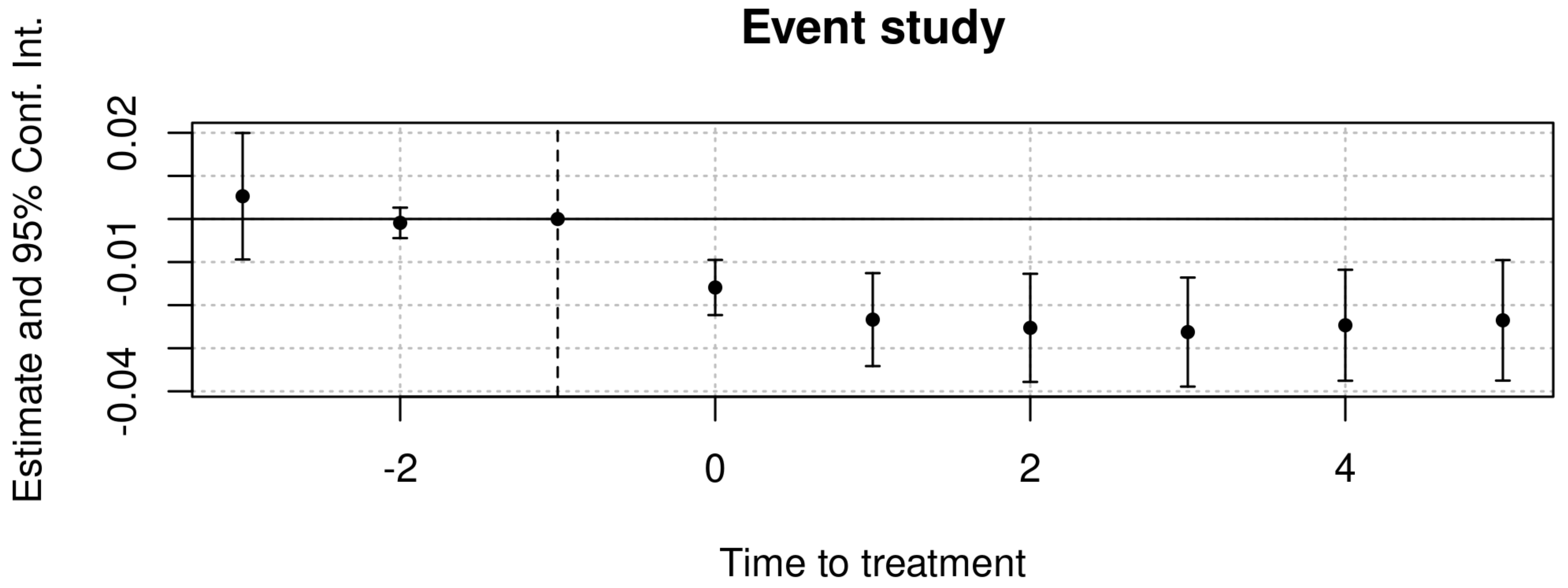
Stata

```
ssc install reghdfe

insheet using "data/acs_medicaid.txt", clear
gen perc_unins=uninsured/adult_pop
drop if expand_ever=="NA"
replace expand_year="." if expand_year=="NA"
destring expand_year, replace
gen event_time=year-expand_year
replace event_time=-1 if event_time==.

forvalues l = 0/4 {
    gen L`l'event = (event_time==`l')
}
forvalues l = 1/2 {
    gen F`l'event = (event_time==-`l')
}
gen F3event=(event_time<=-3)

reghdfe perc_unins F3event F2event L0event L1event L2eve
gen coef = .
gen se = .
```

R

```
library(tidyverse)
library(modelsummary)
library(fixest)
mcaid.data <- read_tsv("../data/acs_medicaid.txt")
reg.dat <- mcaid.data %>%
  filter(!is.na(expand_ever)) %>%
  mutate(perc_unins=uninsured/adult_pop,
         post = (year>=2014),
         treat=post*expand_ever,
         time_to_treat = ifelse(expand_ever==FALSE, 0, y
         time_to_treat = ifelse(time_to_treat < -3, -3,

mod.twfe <- feols(perc_unins~i(time_to_treat, expand_eve
                  cluster=~State,
                  data=reg.dat)
```

Differential treatment timing



What are we estimating?

Problems with TWFE

- Recall goal of estimating ATE or ATT
- TWFE and 2x2 DD identical with homogeneous effects and common treatment timing
- Otherwise...TWFE is biased and inconsistent for ATT

Intuition

- OLS is a weighted average of all 2x2 DD groups
- Weights are function of size of subsamples, size of treatment/control units, and timing of treatment
- Units treated in middle of sample receive larger weights
- Prior-treated units act as controls for late-treated units

Just the length of the panel will change the estimate!

Does it really matter?

- Definitely! But how much?
- Large treatment effects for early treated units could reverse the sign of final estimate
- Let's explore this nice Shiny app from Kyle Butts: Bacon-Decomposition Shiny App.

Note on parallel trends

Parallel trends violated, in general, if:

1. Policy endogeneity (e.g., selection into treatment due to prior outcome)
2. Compositional differences (problematic in repeated cross-sections)