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

Part 3: Difference-in-Differences in Practice

Ian McCarthy | Emory University
Econ 470 & HLTH 470

Table of contents

What are Panel Data?

Nature of the Data

- Repeated observations of the same units over time

Notation

- Unit $i = 1, \dots, N$ over several periods $t = 1, \dots, T$, which we denote y_{it}
- Treatment status D_{it}
- Regression model,
  $y_{it} = \delta D_{it} + \alpha_i + \epsilon_{it}$ for $t = 1, \dots, T$ and $i = 1, \dots, N$

Benefits of Panel Data

- *May* overcome certain forms of omitted variable bias
- Allows for unobserved but time-invariant factor, α_i, that affects both treatment and outcomes

Still assumes

- No time-varying confounders
- Past outcomes do not directly affect current outcomes
- Past outcomes do not affect treatment (reverse causality)

Some textbook settings

- Unobserved "ability" when studying schooling and wages
- Unobserved "quality" when studying physicians or hospitals

Estimating Regressions with Panel Data

Regression model

$y_{it} = \alpha + \delta D_{it} + \epsilon_{it}$ for $t = 1, \ldots, T$ and $i = 1, \ldots, N$

Fixed Effects

$y_{it} = \alpha_i + \delta D_{it} + \epsilon_{it}$ for $t = 1, \ldots, T$ and $i = 1, \ldots, N$

- Allows correlation between α_i and D_{it}
- Physically estimate α_i in some cases via set of dummy variables
- More generally, "remove" α_i via:
  - "within" estimator
  - first-difference estimator

Within Estimator

$y_{it} = \alpha_i + \delta D_{it} + \epsilon_{it}$ for $t = 1, \ldots, T$ and $i = 1, \ldots, N$

- Most common approach (default in most statistical software)
- Equivalent to demeaned model,
  $y_{it} - \bar{y}_i = \delta(D_{it} - \bar{D}_i) + (\alpha_i - \bar{\alpha}_i) + (\epsilon_{it} - \bar{\epsilon}_i)$
- $\alpha_i - \bar{\alpha}_i = 0$ since α_i is time-invariant
- Requires *strict exogeneity* assumption (error is uncorrelated with D_{it} for all time periods)

First-difference

$y_{it} = \delta D_{it} + \alpha_i + \epsilon_{it}$ for $t = 1, \ldots, T$ and $i = 1, \ldots, N$

- Instead of subtracting the mean, subtract the prior period values
  $y_{it} - y_{i,t-1} = \delta(D_{it} - D_{i,t-1}) + (\alpha_i - \alpha_i) + (\epsilon_{it} - \epsilon_{i,t-1})$
- Requires exogeneity of ϵ_{it} and D_{it} only for time t and $t-1$ (weaker assumption than within estimator)
- Sometimes useful to estimate both FE and FD just as a check

Keep in mind...

- Discussion only applies to linear case or very specific nonlinear models
- Fixed effects can't solve reverse causality
- Fixed effects doesn't address unobserved, time-varying confounders
- Can't estimate effects on time-invariant variables
- May "absorb" a lot of the variation for variables that don't change much over time

Within Estimator (Default)

```r
library(readstata13)
library(fixest)
wagepan <- read.dta13("http://fmwww.bc.edu/ec-p/data/wooldridge/wagepan.dta")
feols(lwage~exper + expersq | nr, data=wagepan)
```

```
## OLS estimation, Dep. Var.: lwage
## Observations: 4,360
## Fixed-effects: nr: 545
## Standard-errors: Clustered (nr)
##          Estimate Std. Error t value Pr(>|t|))
## exper    0.122257   0.010585 11.5500 < 2.2e-16 ***
## expersq -0.004523   0.000688 -6.5742  1.15e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.329464     Adj. R2: 0.562461
##                  Within R2: 0.172696
```

Within Estimator (Manually Demean)

```r
library(readstata13)
wagepan <- read.dta13("http://fmwww.bc.edu/ec-p/data/wooldridge/wagepan.dta")
wagepan <- wagepan %>%
  group_by(nr) %>%
  mutate(demean_lwage=lwage - mean(lwage),
         demean_exper=exper - mean(exper),
         demean_expersq=expersq - mean(expersq))
summary(lm(demean_lwage~demean_exper + demean_expersq, data=wagepan))
```

```
##
## Call:
## lm(formula = demean_lwage ~ demean_exper + demean_expersq, data = wagepan)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1752 -0.1221  0.0080  0.1567  1.4875
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     6.557e-17  4.991e-03   0.000        1
## demean_exper    1.223e-01  7.661e-03  15.959  < 2e-16 ***
## demean_expersq -4.523e-03  5.637e-04  -8.024 1.31e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

First differencing

```r
library(readstata13)
wagepan <- read.dta13("http://fmwww.bc.edu/ec-p/data/wooldridge/wagepan.dta")
wagepan <- wagepan %>%
  group_by(nr) %>%
  arrange(year) %>%
  mutate(fd_lwage=lwage - lag(lwage),
         fd_exper=exper - lag(exper),
         fd_expersq=expersq - lag(expersq)) %>%
  na.omit()
summary(lm(fd_lwage~0 + fd_exper + fd_expersq, data=wagepan))
```

```
##
## Call:
## lm(formula = fd_lwage ~ 0 + fd_exper + fd_expersq, data = wagepan)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5866 -0.1454 -0.0131  0.1319  4.8341
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## fd_exper    0.120232   0.019398   6.198 6.32e-10 ***
## fd_expersq -0.004042   0.001383  -2.922   0.0035 **
## ---
```

DD with Medicaid Expansion

Key issue

What is the causal effect of Medicaid expansion?

- Clearly affects insurance markets
- but Medicaid enrollment partially crowds out private insurance

Research design

- Use pre/post and expansion/non-expansion states to identify effect of Medicaid expansion
- In a regression structure:

$$y_{it} = \alpha + \beta \times 1(Post)_t + \gamma D_i + \delta \times 1(Post)_t \times D_i + \varepsilon_{it}$$

Regression results

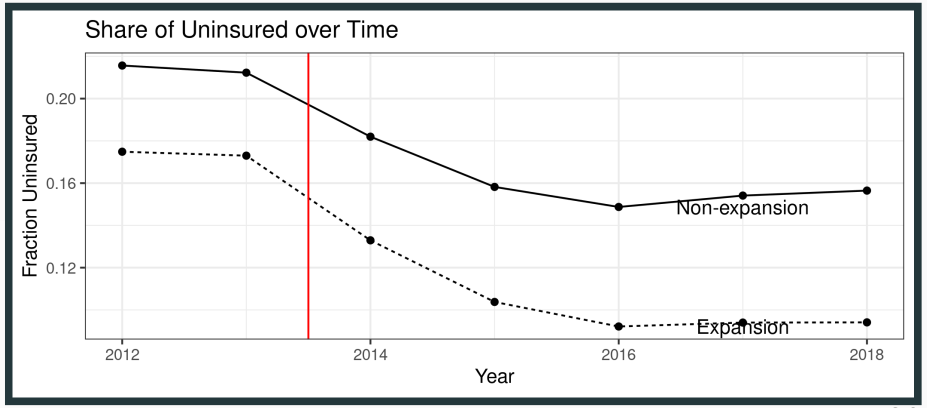
```
ins.dat.2014 <- ins.dat %>% mutate(post = (year >= 2014), treat=post*expand_ever) %>% filter(is.na(expand_year) | expand_y
dd.ins.reg <- lm(perc_unins ~ post + expand_ever + post*expand_ever, data=ins.dat.2014)
summary(dd.ins.reg)
```

```
##
## Call:
## lm(formula = perc_unins ~ post + expand_ever + post * expand_ever,
##     data = ins.dat.2014)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.115667 -0.027106 -0.006804  0.027765  0.117597
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                0.213965   0.007180  29.799  < 2e-16 ***
## postTRUE                  -0.054068   0.008496  -6.364 7.22e-10 ***
## expand_everTRUE           -0.046326   0.009166  -5.054 7.48e-07 ***
## postTRUE:expand_everTRUE -0.018403   0.010845  -1.697   0.0908 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04187 on 304 degrees of freedom
##   (7 observations deleted due to missingness)
```

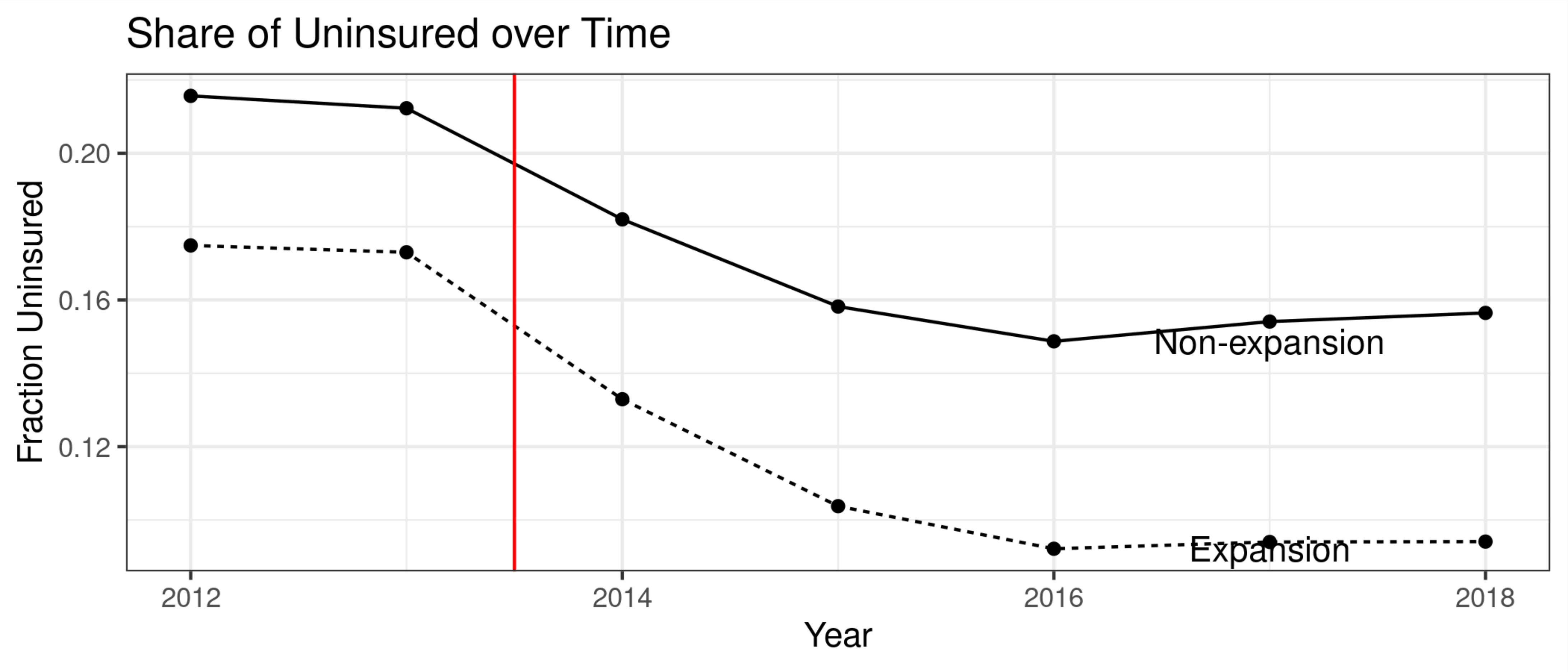
Checking pre-trends

First just plot separately by group:

```
ins.plot.dat <- ins.dat %>% filter(!is.na(expand_ever)) %>%
  group_by(expand_ever, year) %>% summarize(mean=mean(perc_unins))

ggplot(data=ins.plot.dat, aes(x=year,y=mean,group=expand_ever,linetype=expand_ever)) +
  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=FALSE) +
  labs(
    x="Year",
    y="Fraction Uninsured",
    title="Share of Uninsured over Time"
  )
```



Checking pre-trends



Some things to consider

1. Unobserved differences across units or time (TWFE)
2. Heterogeneous treatment effects (event study)

What is TWFE?

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

What is TWFE?

Want to estimate δ:

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

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

Fixed Effects?

Recall our original regression specification:

$$y_{it} = \alpha + \beta \times 1(Post)_t + \gamma D_i + \delta \times 1(Post)_t \times D_i + \varepsilon_{it}$$

This is a special case of a general fixed effects estimator:
$y_{it} = \alpha + \delta \times 1(Post)_t \times D_i + \gamma_i + \gamma_t + \varepsilon$,
where γ_i and γ_t denote a set of coefficients on state (i) and year (t) dummy variables (or fixed effects).

Fixed Effects?

In R, we can estimate the fixed effects specification using the `felm` command (among others), which is part of the `lfe` package. Intuitively, the treatment dummy is now captured by γ_i and the pre/post dummy is captured by γ_t.

- Small datasets, estimate γ_i and γ_t directly
- Large datasets, "fixed effects" estimators will "remove" those variables

Equivalence

DD is just a special case of the fixed effects approach.

```
summary(lm(perc_unins ~ post + expand_ever + post*expand
```

```
##
## Call:
## lm(formula = perc_unins ~ post + expand_ever + post * expand_ever,
##     data = ins.dat.2014)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.115667 -0.027106 -0.006804  0.027765  0.117597
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 0.213965   0.007180  29.799  < 2e-16 ***
## postTRUE                   -0.054068   0.008496  -6.364 7.22e-10 ***
## expand_everTRUE            -0.046326   0.009166  -5.054 7.48e-07 ***
## postTRUE:expand_everTRUE -0.018403   0.010845  -1.697   0.0908 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(felm(perc_unins ~ treat | factor(State) + factor
```

```
##
## Call:
## felm(formula = perc_unins ~ treat | factor(State) + factor
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.042349 -0.007307 -0.000520  0.007342  0.039814
##
## Coefficients:
##        Estimate Std. Error t value Pr(>|t|)
## treat -0.018403   0.003702  -4.971 1.22e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
##
## Residual standard error: 0.01429 on 257 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared(full model): 0.9507   Adjusted R-squared:
```

Event study

This is poorly named:

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

Event study

- Allows for different effect estimates at each time period (maybe effects phase in over time or dissipate)
- Visually very appealing
- Offers easy visual test for parallel trends assumption

Event study

Estimate something akin to…

$$y_{it} = \gamma_i + \gamma_t + \sum_{\tau=-q}^{-1} \delta_\tau D_{i\tau} + \sum_{\tau=0}^{m} \delta_\tau D_{i\tau} + x_{it} + \epsilon_{it},$$

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

Event study

First create all of the treatment/year interactions:

```
event.dat <- ins.dat.2014 %>%
  mutate(expand_2012 = expand_ever*(year==2012),
         expand_2013 = expand_ever*(year==2013),
         expand_2014 = expand_ever*(year==2014),
         expand_2015 = expand_ever*(year==2015),
         expand_2016 = expand_ever*(year==2016),
         expand_2017 = expand_ever*(year==2017),
         expand_2018 = expand_ever*(year==2018))
```

Event study

Second, run regression with full set of interactions and group/year dummies:

```
event.ins.reg <- lm(perc_unins ~ expand_2012 + expand_2014 +
                      expand_2015 + expand_2016 + expand_2017 +
                      expand_2018 + factor(year) + factor(State), data=event.dat)
point.est <- as_tibble(c(event.ins.reg$coefficients[c("expand_2012","expand_2014","expand_2015",
                                                      "expand_2016","expand_2017","expand_2018")]),
                       rownames = "term")
ci.est <- as_tibble(confint(event.ins.reg)[c("expand_2012","expand_2014","expand_2015",
                                             "expand_2016","expand_2017","expand_2018"),],
                    rownames = "term")
```

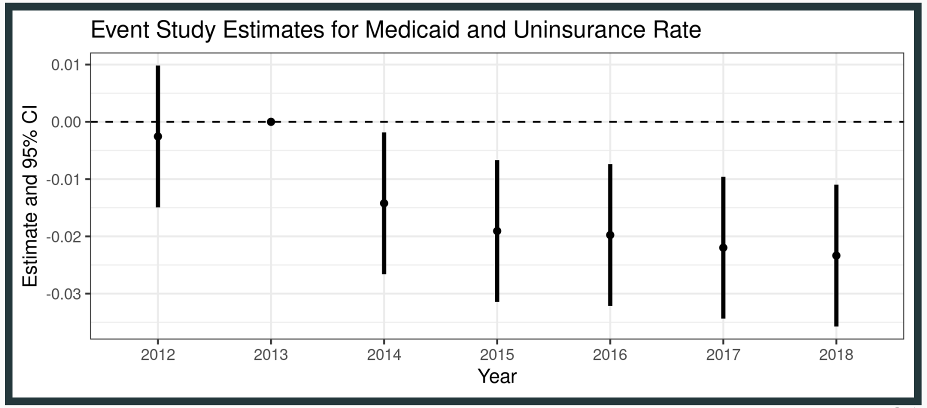
Event study

Third, organize results into a new dataset:

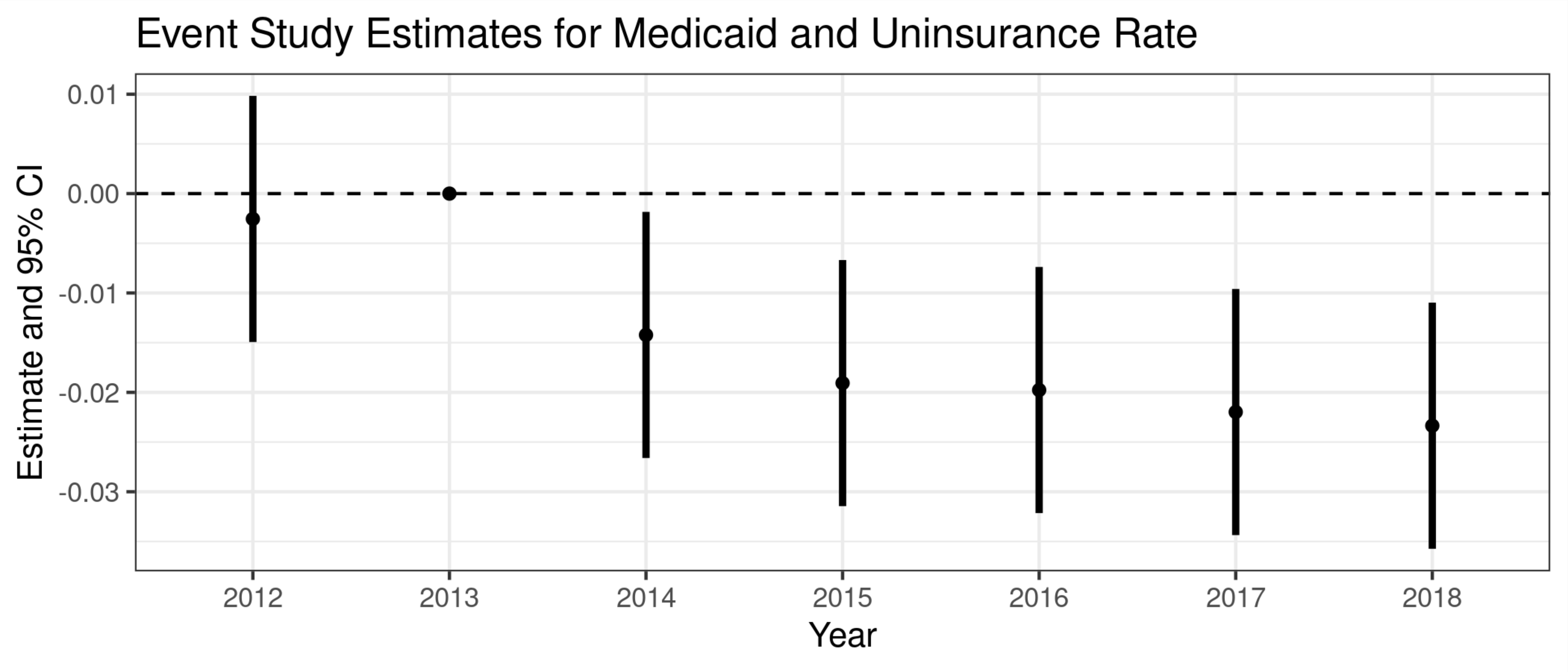
```
point.est <- point.est %>% rename(estimate = value)
ci.est <- ci.est %>% rename(conf.low = `2.5 %`, conf.high = `97.5 %`)
new.row <- tibble(
  term = "expand_2013",
  estimate = 0,
  conf.low = 0,
  conf.high = 0,
  year = 2013
)

event.plot.dat <- point.est %>%
  left_join(ci.est, by=c("term")) %>%
  mutate(year = c(2012, 2014, 2015, 2016, 2017, 2018)) %>%
  bind_rows(new.row) %>%
  arrange(year)
```

Event study

Finally, plot coefficients and confidence intervals



Event study



Event Study Estimates for Medicaid and Uninsurance Rate
0.01
0.00
-0.01
-0.02
-0.03
Estimate and 95% CI
2012
2013
2014
2015
2016
2017
2018
Year

Event study considerations

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

In practice

```
reg.dat <- ins.dat %>%
  filter(!is.na(expand_ever)) %>%
  mutate(post = (year>=2014),
         treat=post*expand_ever,
         time_to_treat = ifelse(expand_ever==FALSE, 0, year-expand_year),
         time_to_treat = ifelse(time_to_treat < -3, -3, time_to_treat))

mod.twfe <- feols(perc_unins~i(time_to_treat, expand_ever, ref=-1) | State + year,
                  cluster=~State,
                  data=reg.dat)
```

In practice

```
iplot(mod.twfe,
      xlab = 'Time to treatment',
      main = 'Event study')
```

What are we estimating?

Problems with TWFE

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

Inutition

- 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 : Bacon-Decomposition Shiny App.