TWFE Metrics Discussion Group: Implementing Callaway & Sant'Anna (2020) DID Estimator

Apr 29, 2021

Olivia Healy

School of Education & Social Policy

Northwestern University

NOTE: current as of 4/29/2021; information on resources and package availability likely to change

Current state of the literature

(incomplete, but a start)

Why is staggered DiD a problem?

How do we do things differently?

Why is staggered DiD a problem?

Two reasons, according to Goodman-Bacon (2021)

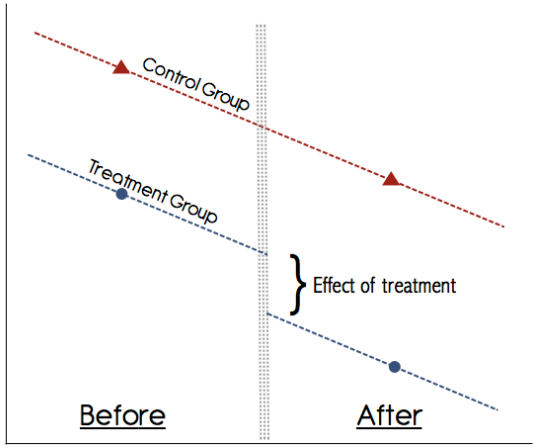
1. Variance weighting (implicit in OLS)
2. Use of past treated units as controls for later-treated units

Why is staggered DiD a problem?

Two reasons, according to Goodman-Bacon (2021)

1. Variance weighting (implicit in OLS)
2. Use of past treated units as controls for later-treated units

Goodman-Bacon (2021) *<-- DiD with constant post-period effects*



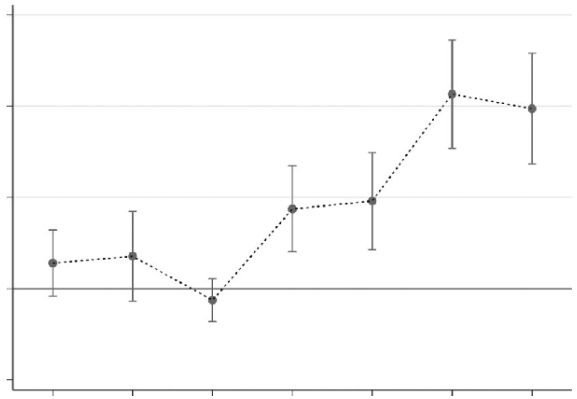
Why is staggered DiD a problem?

Two reasons, according to Goodman-Bacon (2021)

1. ~~Variance weighting (implicit in OLS)~~
2. Use of past treated units as controls for later-treated units

Sun & Abraham (2020) *<-- Focus on event-study DiD (dynamic post period effects)*

Show that event-study fixes problem #1 but not #2



Current state of the literature:

1. Why is staggered DiD a problem?

→*Goodman-Bacon (2020) provides tools to diagnose how bad of a problem staggered DiD is in your setting.*

→*That said, we know there's some degree of a problem, so let's move onto the fix.*

Current state of the literature:

1. Why is staggered DiD a problem?

→*Goodman-Bacon (2020) provides tools to diagnose how bad of a problem staggered DiD is in your setting.*

→*That said, we know there's some degree of a problem, so let's just move onto the fix.*

2. How do we do things differently?

How do we do things differently?

Current options for doing things differently (Baker, 2020)

1. Stacked DiD Estimator
2. Sun & Abraham (2020) Estimator
3. Callaway & Sant'Anna (2020) Estimator

How do we do things differently?

Current options for doing things differently (Baker, 2020)

1. Stacked DiD Estimator
2. Sun & Abraham (2020) Estimator
3. Callaway & Sant'Anna (2020) Estimator

The are the **same** in that they all:

- Rely on event-study models not pooled DiD
- Modify the set of units that act as effective controls in the estimation process

How do we do things differently?

Current options for doing things differently (Baker, 2020)

1. Stacked DiD Estimator
2. Sun & Abraham (2020) Estimator
3. Callaway & Sant'Anna (2020) Estimator

They **vary** ONLY in terms of:

- Which units are used as effective controls (for treatment cases)
- How covariates are incorporated into the analysis

How do we do things differently?

Current options for doing things differently (Baker, 2020)

1. Stacked DiD Estimator
2. Sun & Abraham (2020) Estimator
3. **Callaway & Sant'Anna (2020) Estimator**

They **vary** ONLY in terms of:

- Which units are used as effective controls (for treatment cases)
- How covariates are incorporated into the analysis

Baker (2020) suggests you implement one of these three methods to test robustness of inferences.
My take is that Callaway & Sant'Anna allows for the most flexibility + customization

Callaway & Sant'Anna Estimator:

1. Estimate group-time effects: unique estimates per cohort of units treated at the same point in time

2. Identify appropriate comparison units for each group-time treated group
   - Pairings of units in the treatment and comparison groups per group-time treatment

Implement the Callaway & Sant'Anna Estimator in R

Stata wrappers by: Jonathan Roth & Nick Huntington-Klein

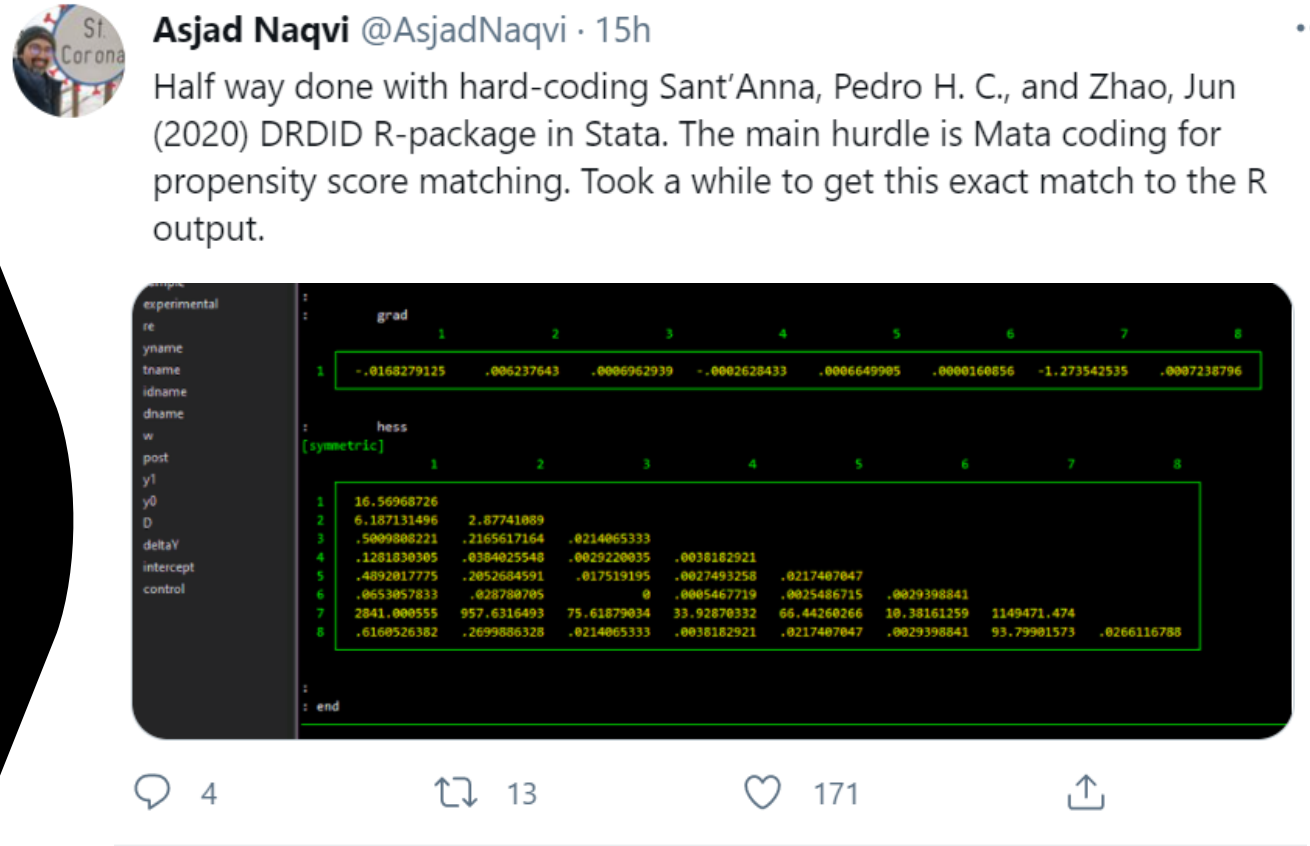
- https://github.com/jonathandroth/staggered#stata-implementation
- https://github.com/NickCH-K/did

Implement the Callaway & Sant'Anna Estimator in R

Stata wrappers by: Jonathan Roth & Nick Huntington-Klein

- https://github.com/jonathandroth/staggered#stata-implementation
- https://github.com/NickCH-K/did

Or wait for a user-written Stata package to emerge



*Note:

The DRDID R-package is different from the the Callaway & Sant'Anna estimator for multi-period fixed effects ("DID" package in R). In fact, DID builds on top of DRDID.

DRDID focuses on scenarios with only two time periods (pre-treatment and post-treatment). It essentially reweights by propensity score and controls for vars included in propensity score calculation … a.k.a. "doubly" robust ("dr")

Crash course in using R

From one novice to another

https://uvastatlab.github.io/phdplus/installR.html

A. **Download/install** R and RStudio if you've never used either before
   - You need both "R-base" and "RStudio"

B. Or, **check your RStudio version + update** to the newest if needed
   - Check for most current version on https://www.r-project.org/ → "R version 4.0.5 (Shake and Throw) has been released on 2021-03-31."

Install packages

In Stata, **ssc install**

In R, there are two necessary steps:

(1) Install

> install.packages("<package name>")

(2) "activate" the package during each analysis session you plan to use it

> library("<package name>")

https://www.datacamp.com/community/tutorials/top-ten-most-important-packages-in-r-for-data-science

Package Info: https://bcallaway11.github.io/did/

Steps

1. Install the Callaway & Sant'Anna package →

You can install **did** from CRAN with:

```
install.packages("did")
```

or get the latest version from github with:

```
# install.packages("devtools")
devtools::install_github("bcallaway11/did")
```

2. Activate the package → library(did)

3. Bring in data → maindata <- read_dta("D:/TFDW/D_R_mat_201006.dta")

4. Implement/run function…

Estimate Group-Time Average Treatment Effects

Start with the most basic approach, relying on package defaults

Source: https://cran.r-project.org/web/packages/did/vignettes/did-basics.html

Estimate Group-Time Average Treatment Effects

- Function to use in R = "**att_gt**"

```
example_attgt <- att_gt(yname = "Y",
                        tname = "period",
                        idname = "id",
                        gname = "G",
                        xformla = ~X,
                        data = dta
                        )
```

Source: https://cran.r-project.org/web/packages/did/vignettes/did-basics.html

Estimate Group-Time Average Treatment Effects

- Function to use in R = "**att_gt**"

```
example_attgt <- att_gt(yname = "Y",
                        tname = "period",
                        idname = "id",
                        gname = "G",
                        xformla = ~X,
                        data = dta
                        )
```

Calendar time or ordered units

1st period treated (constant w/in id)

Source: https://cran.r-project.org/web/packages/did/vignettes/did-basics.html

Estimate Group-Time Average Treatment Effects

- Function to use in R = "**att_gt**"

```
example_attgt <- att_gt(yname = "Y",
                        tname = "period",
                        idname = "id",
                        gname = "G",
                        xformla = ~X,
                        data = dta
                        )
```

Specify pre-treatment covariates that predict treatment or "NULL" for no covariates

Source: https://cran.r-project.org/web/packages/did/vignettes/did-basics.html

Estimate Group-Time Average Treatment Effects

- What use are covariates here?

This feature builds in a DRDID (doubly robust DID) estimator where you:

(1) reweight control units per group-time cohort using propensity score to ensure control units are similar on observables

(2) Partials out variation in outcomes due to covariates, as estimated among control units

Specify pre-treatment covariates that predict treatment or "NULL" for no covariates

```
xformla = ~X,
```

In other words, this options controls for covariates + reweights with propensity score

Source: https://cran.r-project.org/web/packages/did/vignettes/did-basics.html

Estimate Group-Time Average Treatment Effects

Example produces estimates ATE for groups 2, 3, and 4

Recall group (gname) = first period treated

```
summary(example_attgt)
#> Group-Time Average Treatment Effects:
#>  Group Time ATT(g,t) Std. Error [95% Simult.  Conf. Band]
#>      2    2   0.9615     0.0593        0.8035      1.1196 *
#>      2    3   2.0424     0.0587        1.8859      2.1989 *
#>      2    4   3.0104     0.0601        2.8503      3.1705 *
#>      3    2  -0.0392     0.0551       -0.1862      0.1077
#>      3    3   1.0877     0.0617        0.9233      1.2522 *
#>      3    4   1.9938     0.0589        1.8369      2.1506 *
#>      4    2  -0.0537     0.0585       -0.2096      0.1022
#>      4    3   0.0382     0.0563       -0.1119      0.1883
#>      4    4   0.9410     0.0578        0.7869      1.0951 *
#> ---
#> Signif. codes: `*' confidence band does not cover 0
#>
#> P-value for pre-test of parallel trends assumption:  0.79243
#> Control Group:  Never Treated,  Anticipation Periods:  0
#> Estimation Method:  Doubly Robust
```

Estimate Group-Time Average Treatment Effects

Example produces estimates ATE for groups 2, 3, and 4

Recall group (gname) = first period treated

```
summary(example_attgt)
#> Group-Time Average Treatment Effects:
#>  Group Time ATT(g,t) Std. Error [95% Simult.  Conf. Band]
#>      2    2   0.9615     0.0593        0.8035      1.1196 *
#>      2    3   2.0424     0.0587        1.8859      2.1989 *
#>      2    4   3.0104     0.0601        2.8503      3.1705 *
#>      3    2  -0.0392     0.0551       -0.1862      0.1077
#>      3    3   1.0877     0.0617        0.9233      1.2522 *
#>      3    4   1.9938     0.0589        1.8369      2.1506 *
#>      4    2  -0.0537     0.0585       -0.2096      0.1022
#>      4    3   0.0382     0.0563       -0.1119      0.1883
#>      4    4   0.9410     0.0578        0.7869      1.0951 *
#> ---
#> Signif. codes: `*' confidence band does not cover 0
#>
#> P-value for pre-test of parallel trends assumption:  0.79243
#> Control Group:  Never Treated,  Anticipation Periods:  0
#> Estimation Method:  Doubly Robust
```

Effect in 1st year treated for group 2

Estimate Group-Time Average Treatment Effects

Example produces estimates ATE for groups 2, 3, and 4

Recall group (gname) = first period treated

```
summary(example_attgt)
#> Group-Time Average Treatment Effects:
#>  Group Time ATT(g,t) Std. Error [95% Simult.  Conf. Band]
#>      2    2   0.9615     0.0593        0.8035      1.1196 *
#>      2    3   2.0424     0.0587        1.8859      2.1989 *
#>      2    4   3.0104     0.0601        2.8503      3.1705 *
#>      3    2  -0.0392     0.0551       -0.1862      0.1077
#>      3    3   1.0877     0.0617        0.9233      1.2522 *
#>      3    4   1.9938     0.0589        1.8369      2.1506 *
#>      4    2  -0.0537     0.0585       -0.2096      0.1022
#>      4    3   0.0382     0.0563       -0.1119      0.1883
#>      4    4   0.9410     0.0578        0.7869      1.0951 *
#> ---
#> Signif. codes: `*' confidence band does not cover 0
#>
#> P-value for pre-test of parallel trends assumption:  0.79243
#> Control Group:  Never Treated,  Anticipation Periods:  0
#> Estimation Method:  Doubly Robust
```

Effect in 1st year treated for group 3

Estimate Group-Time Average Treatment Effects

Example produces estimates ATE for groups 2, 3, and 4

Recall group (gname) = first period treated

```
summary(example_attgt)
#> Group-Time Average Treatment Effects:
#>  Group Time  ATT(g,t) Std. Error [95% Simult.  Conf. Band]
#>      2    2   0.9615     0.0593        0.8035      1.1196 *
#>      2    3   2.0424     0.0587        1.8859      2.1989 *
#>      2    4   3.0104     0.0601        2.8503      3.1705 *
#>      3    2  -0.0392     0.0551       -0.1862      0.1077
#>      3    3   1.0877     0.0617        0.9233      1.2522 *
#>      3    4   1.9938     0.0589        1.8369      2.1506 *
#>      4    2  -0.0537     0.0585       -0.2096      0.1022
#>      4    3   0.0382     0.0563       -0.1119      0.1883
#>      4    4   0.9410     0.0578        0.7869      1.0951 *
#> ---
#> Signif. codes: `*' confidence band does not cover 0
#>
#> P-value for pre-test of parallel trends assumption:  0.79243
#> Control Group:  Never Treated,  Anticipation Periods:  0
#> Estimation Method:  Doubly Robust
```

Effect in 1st year treated for group 4

Estimate Group-Time Average Treatment Effects

Example produces estimates ATE for groups 2, 3, and 4

Recall group (gname) = first period treated

```
summary(example_attgt)
#> Group-Time Average Treatment Effects:
#>  Group Time  ATT(g,t) Std. Error [95% Simult.  Conf. Band]
#>      2    2   0.9615     0.0593        0.8035      1.1196 *
#>      2    3   2.0424     0.0587        1.8859      2.1989 *
#>      2    4   3.0104     0.0601        2.8503      3.1705 *
#>      3    2  -0.0392     0.0551       -0.1862      0.1077
#>      3    3   1.0877     0.0617        0.9233      1.2522 *
#>      3    4   1.9938     0.0589        1.8369      2.1506 *
#>      4    2  -0.0537     0.0585       -0.2096      0.1022
#>      4    3   0.0382     0.0563       -0.1119      0.1883
#>      4    4   0.9410     0.0578        0.7869      1.0951 *
#> ---
#> Signif. codes: `*' confidence band does not cover 0
#>
#> P-value for pre-test of parallel trends assumption:  0.79243
#> Control Group:  Never Treated,  Anticipation Periods:  0
#> Estimation Method:  Doubly Robust
```

Effect in 1st year treated for group 4

Estimate Group-Time Average Treatment Effects

What are the group-time point estimates relative to?

```
summary(example_attgt)
#> Group-Time Average Treatment Effects:
#>  Group Time ATT(g,t) Std. Error [95% Simult.  Conf. Band]
#>      2    2   0.9615     0.0593        0.8035      1.1196 *
#>      2    3   2.0424     0.0587        1.8859      2.1989 *
#>      2    4   3.0104     0.0601        2.8503      3.1705 *
#>      3    2  -0.0392     0.0551       -0.1862      0.1077
#>      3    3   1.0877     0.0617        0.9233      1.2522 *
#>      3    4   1.9938     0.0589        1.8369      2.1506 *
#>      4    2  -0.0537     0.0585       -0.2096      0.1022
#>      4    3   0.0382     0.0563       -0.1119      0.1883
#>      4    4   0.9410     0.0578        0.7869      1.0951 *
#> ---
#> Signif. codes: `*' confidence band does not cover 0
#>
#> P-value for pre-test of parallel trends assumption:  0.79243
#> Control Group:  Never Treated,  Anticipation Periods:  0
#> Estimation Method:  Doubly Robust
```

Pre-treatment estimates:

- Reference is always the period before (t= t – 1)*
  *this is why period=1 is not estimated

Post-treatment estimates:

- Reference point is the period before treatment (t= -1)

Estimate Group-Time Average Treatment Effects

What are the group-time point estimates relative to?

```
summary(example_attgt)
#> Group-Time Average Treatment Effects:
#>  Group Time ATT(g,t) Std. Error [95% Simult.  Conf. Band]
#>      2    2   0.9615     0.0593        0.8035      1.1196 *
#>      2    3   2.0424     0.0587        1.8859      2.1989 *
#>      2    4   3.0104     0.0601        2.8503      3.1705 *
#>      3    2  -0.0392     0.0551       -0.1862      0.1077
#>      3    3   1.0877     0.0617        0.9233      1.2522 *
#>      3    4   1.9938     0.0589        1.8369      2.1506 *
#>      4    2  -0.0537     0.0585       -0.2096      0.1022
#>      4    3   0.0382     0.0563       -0.1119      0.1883
#>      4    4   0.9410     0.0578        0.7869      1.0951 *
#> ---
#> Signif. codes: `*' confidence band does not cover 0
#>
#> P-value for pre-test of parallel trends assumption:  0.79243
#> Control Group:  Never Treated,  Anticipation Periods:  0
#> Estimation Method:  Doubly Robust
```

Pre-treatment estimates:

- Reference is always the period before (t= t – 1)*

*this is why period=1 is not estimated

Post-treatment estimates:

- Reference point is the period before treatment (t= -1)

Estimate Group-Time Average Treatment Effects

```
summary(example_attgt)
#> Group-Time Average Treatment Effects:
#>  Group Time ATT(g,t) Std. Error [95% Simult.  Conf. Band]
#>      2    2   0.9615     0.0593        0.8035      1.1196 *
#>      2    3   2.0424     0.0587        1.8859      2.1989 *
#>      2    4   3.0104     0.0601        2.8503      3.1705 *
#>      3    2  -0.0392     0.0551       -0.1862      0.1077
#>      3    3   1.0877     0.0617        0.9233      1.2522 *
#>      3    4   1.9938     0.0589        1.8369      2.1506 *
#>      4    2  -0.0537     0.0585       -0.2096      0.1022
#>      4    3   0.0382     0.0563       -0.1119      0.1883
#>      4    4   0.9410     0.0578        0.7869      1.0951 *
#> ---
#> Signif. codes: `*' confidence band does not cover 0
#>
#> P-value for pre-test of parallel trends assumption:  0.79243
#> Control Group:  Never Treated,  Anticipation Periods:  0
#> Estimation Method:  Doubly Robust
```

Built in parallel trends test of all pre-treatment estimates

Estimate Group-Time Average Treatment Effects

Default settings shown below results

```
summary(example_attgt)
#> Group-Time Average Treatment Effects:
#>  Group Time ATT(g,t) Std. Error [95% Simult.  Conf. Band]
#>      2    2   0.9615     0.0593        0.8035      1.1196 *
#>      2    3   2.0424     0.0587        1.8859      2.1989 *
#>      2    4   3.0104     0.0601        2.8503      3.1705 *
#>      3    2  -0.0392     0.0551       -0.1862      0.1077
#>      3    3   1.0877     0.0617        0.9233      1.2522 *
#>      3    4   1.9938     0.0589        1.8369      2.1506 *
#>      4    2  -0.0537     0.0585       -0.2096      0.1022
#>      4    3   0.0382     0.0563       -0.1119      0.1883
#>      4    4   0.9410     0.0578        0.7869      1.0951 *
#> ---
#> Signif. codes: `*' confidence band does not cover 0
#>
#> P-value for pre-test of parallel trends assumption:  0.79243
#> Control Group:  Never Treated,  Anticipation Periods:  0
#> Estimation Method:  Doubly Robust
```

Estimate Group-Time Average Treatment Effects

Move beyond package defaults

Source: https://cran.r-project.org/web/packages/did/vignettes/did-basics.html

Select Alternative Control Groups

DEFAULT = never treated

- If no never treated, last treated group serves as control group for all prior treated groups (a warning will pop up)

```
example_attgt_altcontrol <- att_gt(yname = "Y",
                                   tname = "period",
                                   idname = "id",
                                   gname = "G",
                                   xformla = ~X,
                                   data = dta,
                                   control_group = "notyettreated"
                                   )
```

Includes:
(a) never treated (if any) +
(b) not yet treated

Source: https://cran.r-project.org/web/packages/did/vignettes/did-basics.html

Quick Note on Coding var for "gname"

Make sure to code your "never treated" group as "0"

```
#Replace missing values of first treatment year to 0 for never treated
dta$G[is.na(dta$G)] <- 0
```

```
example_attgt <- att_gt(yname = "Y",
                        tname = "period",
                        idname = "id",
                        gname = "G",
                        xformla = ~X,
                        data = dta
                        )
```

1st period treated

Name of dataset

Source: https://cran.r-project.org/web/packages/did/vignettes/did-basics.html

Estimate Group-Time Average Treatment Effects

Move from individual estimates of group-time ATEs by period pre/post treatment to something more useful

Source: https://cran.r-project.org/web/packages/did/vignettes/did-basics.html

Aggregating group-time average treatment effects

Recall, example produces estimates ATE for groups 2, 3, and 4, where group (gname) = first period treated

Function in R = "**aggte**"

Options for aggregation

1. Simple → Weighted average of all group time ATEs, weights proportional to group size
2. Dynamic/event-study → Average group time ATEs at each time point post treatment
3. Group-specific → Average all post-treatment time points by treatment time group
4. Calendar time → Average effect per calendar time, for groups treated at that time point

Source: https://cran.r-project.org/web/packages/did/vignettes/did-basics.html

Aggregating group-time average treatment effects

Recall, example produces estimates ATE for groups 2, 3, and 4, where group (gname) = first period treated

Function in R = "**aggte**"

Options for aggregation

1. Simple → Weighted average of all group time ATEs, weights proportional to group size
2. Dynamic/event-study → Average group time ATEs at each time point post treatment
3. Group-specific → Average all post-treatment time points by treatment time group
4. Calendar time → Average effect per calendar time, for groups treated at that time point

Source: https://cran.r-project.org/web/packages/did/vignettes/did-basics.html

Aggregating group-time average treatment effects

Recall, example produces estimates ATE for groups 2, 3, and 4, where group (gname) = first period treated

Function in R = "**aggte**"

Options for aggregation

1. Simple → Weighted average of all group time ATEs, weights proportional to group size
2. Dynamic/event-study → Average group time ATEs at each time point post treatment
3. Group-specific → Average all post-treatment time points by treatment time group
4. Calendar time → Average effect per calendar time, for groups treated at that time point

Source: https://cran.r-project.org/web/packages/did/vignettes/did-basics.html

Aggregating group-time average treatment effects

Recall, example produces estimates ATE for groups 2, 3, and 4, where group (gname) = first period treated

Function in R = "**aggte**"

Options for aggregation

1. Simple → Weighted average of all group time ATEs, weights proportional to group size
2. Dynamic/event-study → Average group time ATEs at each time point post treatment
3. Group-specific → Average all post-treatment time points by treatment time group
4. Calendar time → Average effect per calendar time, for groups treated at that time point

Source: https://cran.r-project.org/web/packages/did/vignettes/did-basics.html

Aggregating group-time average treatment effects

Recall, example produces estimates ATE for groups 2, 3, and 4, where group (gname) = first period treated

Function in R = "**aggte**"

Options for aggregation

1. Simple

```
agg.simple <- aggte(example_attgt, type = "simple")
```

2. Dynamic/event-study

```
agg.es <- aggte(example_attgt, type = "dynamic")
```

3. Group-specific

```
agg.gs <- aggte(example_attgt, type = "group")
```

4. Calendar time

```
agg.ct <- aggte(example_attgt, type = "calendar")
```

Source: https://cran.r-project.org/web/packages/did/vignettes/did-basics.html

Implementing the Callaway & Sant'Anna Estimator in R

Can specify/customize many other options!

Helpful resources include:

- https://causalinf.substack.com/p/callaway-and-santanna-dd-estimator
- https://bcallaway11.github.io/did/articles/did-basics.html

Recent paper using this method -- Gilpin, Karger, Nencka (2021):

- https://www.chicagofed.org/publications/working-papers/2021/2021-06

Thanks!