csdid: Difference-in-Differences with Multiple Time Periods in Stata

Fernando Rios-Avila
Levy Economics Institute

Brantly Callaway
University of Georgia

Pedro H. C. Sant'Anna
Microsoft and Vanderbilt University

Stata Conference, August 2021

Big shout-out

- This project would not have reach its current point without the help and push of many.
- Special thanks goes to
    - Austin Nichols (Abt Associates)
    - Enrique Pinzón (Stata Corp)
    - Asjad Naqvi (International Institute for Applied Systems Analysis)

Big Picture

Big Picture: Problems of common practice - I

- Consider a setup with **variation in treatment timing** and **heterogeneous treatment effects**.

- Researchers routinely interpret β^{TWFE} associated with the TWFE specification

$$Y_{i,t} = \alpha_i + \alpha_t + \beta^{TWFE} D_{i,t} + \varepsilon_{i,t},$$

  as "a causal parameter of interest".

- However, β^{TWFE} is not guaranteed to recover an interpretable causal parameter (Borusyak and Jaravel, 2017; de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021).

Big Picture: Problems of common practice - II

- Researchers also routinely consider “dynamic” variations of the TWFE specification,

$$Y_{i,t} = \alpha_i + \alpha_t + \gamma_k^{-K} D_{i,t}^{<-K} + \sum_{k=-K}^{-2} \gamma_k^{lead} D_{i,t}^{k} + \sum_{k=0}^{L} \gamma_k^{lags} D_{i,t}^{k} + \gamma_k^{L+} D_{i,t}^{>L} + \varepsilon_{i,t}$$

  with the event study dummies $D_{i,t}^k = 1\{t - G_i = k\}$, where G_i indicates the period unit i is first treated (Group).

- $D_{i,t}^k$ is an indicator for unit i being k periods away from initial treatment at time t.

- Sun and Abraham (2020) demonstrated the the γ’s cannot be rigorously interpreted as reliable measures of “dynamic treatment effects”.

The heart of the drawbacks

- The heart of the these problems with these TWFE specifications is that OLS is "variational hungry".
- OLS attempts to compare all cohorts with each other, as long as there is "variation in treatment status" in that given time-window.
    - It doesn't care about "treatment" and "comparison" groups.
    - It is all about minimizing MSE.
- **Causal inference is about only exploiting the good variation, i.e., those that respect our assumptions**.

How to tackle the problems?

- With this insight in mind, it is clear what we need to do.
- We need to enforce that our estimation and inference procedure use the variations that we want it use.

How to tackle the problems?

- With this insight in mind, it is clear what we need to do.
- We need to enforce that our estimation and inference procedure use the variations that we want it use.
- Callaway and Sant'Anna (2020) propose a **transparent** way to proceed with this insight in DiD setups with multiple time periods.
- Today's talk is all about how to implement it with our Stata command, **csdid**.

Framework and Assumptions

Framework

- **csdid** accommodates both panel data and repeated cross section data.
- For simplicity, I'll focus on the panel data case.
- Consider a random sample
$$\{(Y_{i,1}, Y_{i,2}, \ldots, Y_{i,\mathcal{T}}, D_{i,1}, D_{i,2}, \ldots, D_{i,\mathcal{T}}, X_i)\}_{i=1}^{n}$$
where $D_{i,t} = 1$ if unit i is treated in period t, and 0 otherwise
- $G_{i,g} = 1$ if unit i is first treated at time g , and zero otherwise ("Treatment starting-time / Cohort dummies")
- $C = 1$ is a "never-treated" comparison group (not required, though)
- Staggered treatment adoption: $D_{i,t} = 1 \implies D_{i,t+1} = 1$, *for* $t = 1, 2, \ldots, \mathcal{T}$.

Framework (cont.)

- Limited Treatment Anticipation: There is a known $\delta \geq 0$ s.t.

$$\mathbb{E}[Y_t(g)|X, G_g = 1] = \mathbb{E}[Y_t(0)|X, G_g = 1] \; a.s..$$

for all $g \in \mathcal{G}, t \in 1, \ldots, \mathcal{T}$ such that $\underbrace{t < g - \delta}_{\text{"before effective starting date"}}$.

- For simplicity, let's take $\delta = 0$, which is arguably the norm in the literature.

- Generalized propensity score uniformly bounded away from 1:

$$p_{g,t}(X) = P\left(G_g = 1 | X, G_g + (1 - D_t)(1 - G_g) = 1\right) \leq 1 - \epsilon \; a.s.$$

Parameter of interest (or the building block of the analysis)

- Parameter of interest:

$$ATT(g,t) = \mathbb{E}\left[Y_t(g) - Y_t(0) \,|\, G_g = 1\right], \text{ for } t \geq g.$$

Average treatment effect for the group of units first treated at time period g, in calendar time t.

Parallel trend assumption based on a "never treated" group

Assumption (Conditional Parallel Trends based on a "never-treated" group)

For each $t \in \{2, \ldots, \mathcal{T}\}$, $g \in \mathcal{G}$ *such that* $t \geq g$,

$$\mathbb{E}[Y_t(0) - Y_{t-1}(0) | X, G_g = 1] = \mathbb{E}[Y_t(0) - Y_{t-1}(0) | X, C = 1] \ a.s..$$

Parallel Trends based on not-yet treated groups

Assumption (Conditional Parallel Trends based on "Not-Yet-Treated" Groups)

For each $(s, t) \in \{2, \ldots, \mathcal{T}\} \times \{2, \ldots, \mathcal{T}\}, g \in \mathcal{G}$ *such that* $t \geq g, \quad s \geq t$

$$\mathbb{E}[Y_t(0) - Y_{t-1}(0)|X, G_g = 1] = \mathbb{E}[Y_t(0) - Y_{t-1}(0)|X, D_s = 0, G_g = 0] \; a.s..$$

Recovering the ATT(g,t)'s

What if the identifying assumptions hold unconditionally?

- In the case where covariates do not play a major role into the DiD identification analysis, and one is comfortable using the "never treated" as comparison group,

$$ATT_{unc}^{nev}(g,t) = \mathbb{E}[Y_t - Y_{g-1} | G_g = 1] - \mathbb{E}[Y_t - Y_{g-1} | C = 1].$$

- If one prefers to use the "not-yet treated" as comparison groups,

$$ATT_{unc}^{ny}(g,t) = \mathbb{E}[Y_t - Y_{g-1} | G_g = 1] - \mathbb{E}[Y_t - Y_{g-1} | D_t = 0, G_g = 0].$$

What if the identifying assumptions hold unconditionally?

- In the case where covariates do not play a major role into the DiD identification analysis, and one is comfortable using the "never treated" as comparison group,

$$ATT_{unc}^{nev}(g,t) = \mathbb{E}[Y_t - Y_{g-1} | G_g = 1] - \mathbb{E}[Y_t - Y_{g-1} | C = 1].$$

- If one prefers to use the "not-yet treated" as comparison groups,

$$ATT_{unc}^{ny}(g,t) = \mathbb{E}[Y_t - Y_{g-1} | G_g = 1] - \mathbb{E}[Y_t - Y_{g-1} | D_t = 0, G_g = 0].$$

- Estimation: use the analogy principle!

- Inference: many comparisons of means!

Identification results - never treated as comparison group

- When covariates play an important role and we use the "never treated" units as comparison group, Callaway and Sant'Anna (2020) show you can use three estimation methods: OR, IPW or DR (AIPW).

- Here we show the AIPW/DR estimand:

$$ATT_{dr}^{nev}(g,t) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_g(X)C}{1-p_g(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_{g,t}^{nev}(X)\right)\right].$$

where $m_{g,t}^{nev}(X) = \mathbb{E}\left[Y_t - Y_{g-1} | X, C = 1\right]$.

- Extends Heckman, Ichimura and Todd (1997); Abadie (2005); Sant'Anna and Zhao (2020)

Identification results - never treated as comparison group

- When covariates play an important role and we use the "never treated" units as comparison group, Callaway and Sant'Anna (2020) show you can use three estimation methods: OR, IPW or DR (AIPW).

- Here we show the AIPW/DR estimand:

$$ATT_{dr}^{nev}(g,t) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_g(X)\,C}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_g(X)\,C}{1-p_g(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_{g,t}^{nev}(X)\right)\right].$$

where $m_{g,t}^{nev}(X) = \mathbb{E}\left[Y_t - Y_{g-1} | X, C = 1\right]$.

- Extends Heckman et al. (1997); Abadie (2005); Sant'Anna and Zhao (2020)

Identification results - never treated as comparison group

- When covariates play an important role and we use the "never treated" units as comparison group, Callaway and Sant'Anna (2020) show you can use three estimation methods: OR, IPW or DR (AIPW).

- Here we show the AIPW/DR estimand:

$$ATT_{dr}^{nev}(g,t) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_g(X)C}{1-p_g(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_{g,t}^{nev}(X)\right)\right].$$

where $m_{g,t}^{nev}(X) = \mathbb{E}\left[Y_t - Y_{g-1} | X, C = 1\right]$.

- Extends Heckman et al. (1997); Abadie (2005); Sant'Anna and Zhao (2020)

Identification results - never treated as comparison group

- When covariates play an important role and we use the "never treated" units as comparison group, Callaway and Sant'Anna (2020) show you can use three estimation methods: OR, IPW or DR (AIPW).

- Here we show the AIPW/DR estimand:

$$ATT_{dr}^{nev}(g,t) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_g(X)C}{1-p_g(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_{g,t}^{nev}(X)\right)\right].$$

where $m_{g,t}^{nev}(X) = \mathbb{E}\left[Y_t - Y_{g-1} | X, C=1\right]$.

- Extends Heckman et al. (1997); Abadie (2005); Sant'Anna and Zhao (2020)

Identification results - not-yet treated as comparison group

- Callaway and Sant'Anna (2020) show you can get analogous results when using "not-yet treated" units as the comparison group.
- Here we show the AIPW/DR estimand:

$$ATT_{dr}^{ny}(g,t) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_{g,t}(X)(1-D_t)}{1-p_{g,t}(X)}}{\mathbb{E}\left[\frac{p_{g,t}(X)(1-D_t)}{1-p_{g,t}(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_{g,t}^{ny}(X)\right)\right].$$

where $m_{g,t}^{ny}(X) = \mathbb{E}\left[Y_t - Y_{g-1} | X, D_t = 0, G_g = 0\right]$. .

- Extends Heckman et al. (1997); Abadie (2005); Sant'Anna and Zhao (2020), too.

Stata Implementation

Let’s get start with the csdid package in Stata

We first need to install **csdid** and its sister package, **drdid**, that implements Sant’Anna and Zhao (2020); see Rios-Avila, Naqvi and Sant’Anna (2021)

```
* Let's first install drdid
ssc install drdid, all replace

* Now let's install csdid
ssc install csdid, all replace
```

Let's get start with the csdid package in Stata

We first need to install **csdid** and its sister package, **drdid**, that implements Sant'Anna and Zhao (2020); see Rios-Avila et al. (2021)

```
* Let's first install drdid
ssc install drdid, all replace

* Now let's install csdid
ssc install csdid, all replace
```

I strongly recommend that you take a look at our help files:

```
* Help file for csdid
help csdid

* Help file for Post-estimation procedures associated with csdid
help csdid_postestimation
```

csdid syntax

csdid *depvar* [*indepvars*] [*if*] [*in*] [weight], [**ivar**(varname)] ***time***(varname) ***gvar***(varname) [options]

- *depvar*: Outcome of interest
- *indepvars*: Optional vector of covariates
- *weight*: Optional vector of (sampling) weights
- **ivar**: Cross-sectional identifier
- **time**: time-series identifier
- **gvar**: Treatment-group (cohort) identifier (0 for never-treated)
- options: where a lot of action takes place - important for choice of comparison group, estimation method and type of inference procedure

csdid syntax - some additional details inside option

- **notyet:** Use not-yet-treated units as comparison group. If not set, we will use never-treated (if any).
- **method(*method*):** Select the estimation method to be used *(only relevant if there are covariates)*. Current options are
  - **drimp (default):** Implement improved doubly robust DiD estimator based on inverse probability of tilting and weighted least squares (Sant'Anna and Zhao, 2020).
  - **dripw:** Implement doubly robust DiD estimator based on IPW and OLS. (Sant'Anna and Zhao, 2020; Callaway and Sant'Anna, 2020)
  - **reg:** Implement outcome regression DiD estimator based on OLS (Heckman et al., 1997; Callaway and Sant'Anna, 2020).
  - **ipw:** Implement (stabilized) IPW DiD estimator (Abadie, 2005; Callaway and Sant'Anna, 2020).

What about Post-Estimation?

- **csdid_plot:** Command for plotting results from csdid.
  - Need to specify the group you want to plot the effects;
  - style(styleoption): Allows you to change the style of the plot. The options are rspike (default), rarea, rcap and rbar.
- **csdid_stats pretrend** or **estat pretrend**: estimates the chi2 statistic of the null hypothesis that **all** pretreatment $ATT(g, t)$'s are equal to zero.

Illustration

Example using subset of data from CS2020

In this illustration, we will use a subset of the Callaway and Sant'Anna (2020) dataset.

This serves **purely** for syntax illustration!

Unconditional DiD with never-treated as comparison group

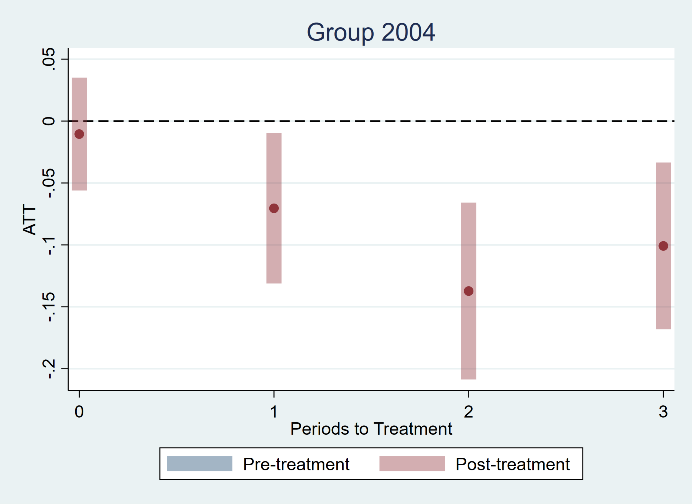
```
. * Estimation of all ATTGT's using uncondition DiD with never-treated as comparison group

.
. * Standard errors computed using analytical results

.
. csdid lemp , ivar(countyreal) time(year) gvar(first_treat)
............
Difference-in-difference with Multiple Time Periods
Outcome model  :
Treatment model:
```

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
g2004						
t_2003_2004	-.0105032	.023251	-0.45	0.651	-.0560744	.0350679
t_2003_2005	-.0704232	.0309848	-2.27	0.023	-.1311522	-.0096941
t_2003_2006	-.1372587	.0364357	-3.77	0.000	-.2086713	-.0658461
t_2003_2007	-.1008114	.0343592	-2.93	0.003	-.1681542	-.0334685
g2006						
t_2003_2004	.0065201	.0233268	0.28	0.780	-.0391996	.0522398
t_2004_2005	-.0027508	.0195586	-0.14	0.888	-.0410849	.0355833
t_2005_2006	-.0045946	.0177552	-0.26	0.796	-.0393942	.0302049
t_2005_2007	-.0412245	.0202292	-2.04	0.042	-.0808729	-.001576
g2007						
t_2003_2004	.0305067	.0150336	2.03	0.042	.0010414	.0599719
t_2004_2005	-.0027259	.0163958	-0.17	0.868	-.0348611	.0294093
t_2005_2006	-.0310871	.0178775	-1.74	0.082	-.0661264	.0039522
t_2006_2007	-.0260544	.0166554	-1.56	0.118	-.0586985	.0065896

Control: Never Treated

See Callaway and Sant'Anna (2020) for details

Unconditional DiD with never-treated as comparison group



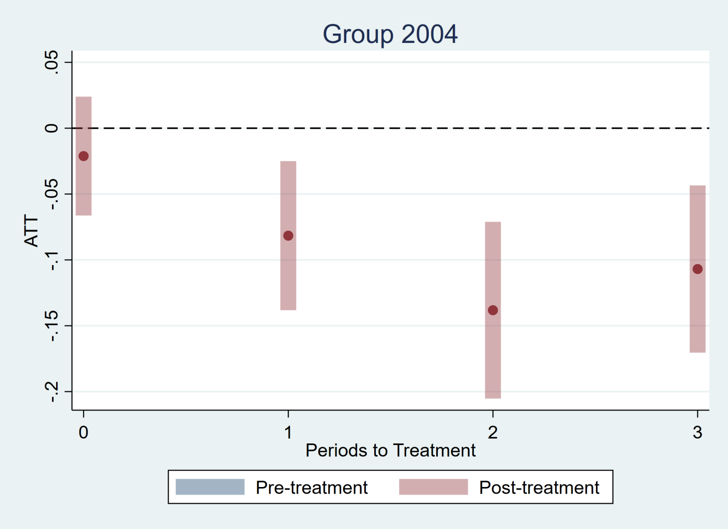
Conditional IPW-based DiD with not-yet-treated as comp. group

```
. * Estimation of all ATT(g,t)'s using IPW estimation method with not-yet-treated as comparison group

.
. * standard errors using wild-bootstrap

.
. csdid lemp lpop , ivar(countyreal) time(year) gvar(first_treat) notyet method(ipw)  wboot rseed(08052021)
............
Difference-in-difference with Multiple Time Periods
Outcome model  :
Treatment model:
```

	Coefficient	Std. err.	t	[95% conf.	interval]
g2004					
t_2003_2004	-.0211844	.0225172	-0.94	-.0663122	.0239434
t_2003_2005	-.0816065	.0288115	-2.83	-.1382072	-.0250058
t_2003_2006	-.1381948	.0339417	-4.07	-.2052931	-.0710965
t_2003_2007	-.1069341	.0311113	-3.44	-.1704361	-.0434322
g2006					
t_2003_2004	-.0075149	.0233701	-0.32	-.0530016	.0379717
t_2004_2005	-.0047093	.0189161	-0.25	-.0387104	.0292919
t_2005_2006	.0087511	.0179391	0.49	-.0237322	.0412344
t_2005_2007	-.0415457	.0203369	-2.04	-.0809737	-.0021177
g2007					
t_2003_2004	.0268608	.0144755	1.86	-.0002889	.0540106
t_2004_2005	-.004264	.0167157	-0.26	-.0351296	.0266017
t_2005_2006	-.0283679	.0184515	-1.54	-.0621979	.0054621
t_2006_2007	-.0289168	.0162066	-1.78	-.0600582	.0022246

```
Control: Not yet Treated

See Callaway and Sant'Anna (2020) for details

.
```

Conditional IPW-based DiD with not-yet-treated as comp. group



Aggregating the ATT(g,t)'s

Summarizing

- Since we have been "sub-setting the data" to get $ATT(g, t)$'s, you may be wondering: "*Are we throwing away information?*"
- Alternatively, you may be wondering how to better communicate the results, specially in setups with many groups/period.
- Aggregation of causal effects is something empiricist commonly pursue:
  - Run a TWFE "static" regression and focus on the β associated with the treatment.
  - Run a TWFE event-study regression and focus on β associated with the treatment leads and lags.
  - Collapse data into a 2 x 2 Design (average pre and post treatment periods).

Summarizing Causal Effects

- Callaway and Sant'Anna (2020) propose taking weighted averages of the $ATT(g,t)$ of the form:

$$\sum_{g=2}^{\mathcal{T}} \sum_{t=2}^{\mathcal{T}} \mathbf{1}\{g \leq t\} w_{gt} ATT(g,t)$$

- **Name-of-the-game:** we must choose "reasonable" weights such that the aggregated causal effect is easy-to-interpret.

Summarizing Causal Effects

- Callaway and Sant'Anna (2020) suggest some arguably intuitive weighting schemes, including
  - Simple weighted-average of all $ATT(g,t)$'s:

$$\theta_W^{simple} := \frac{1}{\kappa} \sum_{g=2}^{\mathcal{T}} \sum_{t=2}^{\mathcal{T}} \mathbf{1}\{g \leq t\} ATT(g,t) P(G = g | C \neq 1) \tag{1}$$

  - Average effect of participating in the treatment for the group of units that have been exposed to the treatment for exactly e time periods

$$\theta_D^{event}(e) = \sum_{g=2}^{\mathcal{T}} \mathbf{1}\{g + e \leq \mathcal{T}\} ATT(g, g+e) P(G = g | G + e \leq \mathcal{T}, C \neq 1)$$

- Implement in Stata via: **estat all** or **csdid_stats all**

Conditional DR-based DiD with never-treated as comp. group

```
. csdid lemp lpop , ivar(countyreal) time(year) gvar(first_treat) method(dripw)
............
Difference-in-difference with Multiple Time Periods
Outcome model   : least squares
Treatment model: inverse probability
```

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
g2004						
t_2003_2004	-.0145297	.0221292	-0.66	0.511	-.057902	.0288427
t_2003_2005	-.0764219	.0286713	-2.67	0.008	-.1326166	-.0202271
t_2003_2006	-.1404483	.0353782	-3.97	0.000	-.2097882	-.0711084
t_2003_2007	-.1069039	.0328865	-3.25	0.001	-.1713602	-.0424476
g2006						
t_2003_2004	-.0004721	.0222234	-0.02	0.983	-.0440293	.043085
t_2004_2005	-.0062025	.0184957	-0.34	0.737	-.0424534	.0300484
t_2005_2006	.0009606	.0194002	0.05	0.961	-.0370631	.0389843
t_2005_2007	-.0412939	.0197211	-2.09	0.036	-.0799466	-.0026411
g2007						
t_2003_2004	.0267278	.0140657	1.90	0.057	-.0008404	.054296
t_2004_2005	-.0045766	.0157178	-0.29	0.771	-.0353828	.0262297
t_2005_2006	-.0284475	.0181809	-1.56	0.118	-.0640814	.0071864
t_2006_2007	-.0287814	.016239	-1.77	0.076	-.0606091	.0030464

Control: Never Treated

See Callaway and Sant'Anna (2020) for details

Conditional DR-based DiD with never-treated as comp. group

```
.
. estat all
Pretrend Test. H0 All Pre-treatment are equal to 0
chi2(5) = 6.841824981670457
p-value       = .2326722805724239
Average Treatment Effect on Treated
```

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
ATT	-.0417518	.0115028	-3.63	0.000	-.0642969	-.0192066

ATT by group

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
G2004	-.0845759	.0245649	-3.44	0.001	-.1327222	-.0364297
G2006	-.0201666	.0174696	-1.15	0.248	-.0544065	.0140732
G2007	-.0287814	.016239	-1.77	0.076	-.0606091	.0030464

ATT by Calendar Period

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
T2004	-.0145297	.0221292	-0.66	0.511	-.057902	.0288427
T2005	-.0764219	.0286713	-2.67	0.008	-.1326166	-.0202271
T2006	-.0461757	.0212107	-2.18	0.029	-.087748	-.0046035
T2007	-.0395822	.0129299	-3.06	0.002	-.0649242	-.0142401

ATT by Periods Before and After treatment
Event Study:Dynamic effects

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
T-3	.0267278	.0140657	1.90	0.057	-.0008404	.054296
T-2	-.0036165	.0129283	-0.28	0.780	-.0289555	.0217226
T-1	-.023244	.0144851	-1.60	0.109	-.0516343	.0051463
T+0	-.0210604	.0114942	-1.83	0.067	-.0435886	.0014679
T+1	-.0530032	.0163465	-3.24	0.001	-.0850417	-.0209647
T+2	-.1404483	.0353782	-3.97	0.000	-.2097882	-.0711084
T+3	-.1069039	.0328865	-3.25	0.001	-.1713602	-.0424476

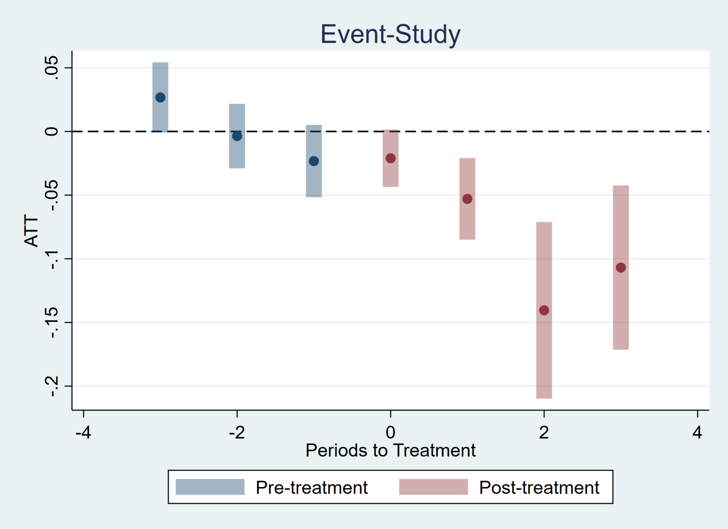
Conditional DR-based DiD with never-treated as comp. group

```
. estat event
ATT by Periods Before and After treatment
Event Study:Dynamic effects
```

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
T-3	.0267278	.0140657	1.90	0.057	-.0008404	.054296
T-2	-.0036165	.0129283	-0.28	0.780	-.0289555	.0217226
T-1	-.023244	.0144851	-1.60	0.109	-.0516343	.0051463
T+0	-.0210604	.0114942	-1.83	0.067	-.0435886	.0014679
T+1	-.0530032	.0163465	-3.24	0.001	-.0850417	-.0209647
T+2	-.1404483	.0353782	-3.97	0.000	-.2097882	-.0711084
T+3	-.1069039	.0328865	-3.25	0.001	-.1713602	-.0424476

```
.
. csdid_plot,  title("Event-Study")
```

Conditional DR-based DiD with never-treated as comp. group



Conclusion

Conclusion

- Callaway and Sant'Anna (2020) proposes semi-parametric DiD estimators when there are multiple time-periods and variation in treatment timing.
- These tools are attractive because they are transparent and avoid weighting problems associated with TWFE specifications.
- **csdid** provide a native Stata implementation of these methods.
    - Embrace TE heterogeneity in the same way as **teffects** does in cross-section setups.

References

Abadie, Alberto, “Semiparametric Difference-in-Differences Estimators,” *The Review of Economic Studies*, 2005, *72* (1), 1–19.

Borusyak, Kirill and Xavier Jaravel, “Revisiting Event Study Designs,” SSRN Scholarly Paper ID 2826228, Social Science Research Network, Rochester, NY August 2017.

Callaway, Brantly and Pedro H. C. Sant’Anna, “Difference-in-Differences with Multiple Time Periods,” *Journal of Econometrics*, 2020, (Forthcoming).

de Chaisemartin, Clément and Xavier D’Haultfœuille, “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects,” *American Economic Review*, 2020, *110* (9), 2964–2996.

Goodman-Bacon, Andrew, “Difference-in-Differences with Variation in Treatment Timing,” *Journal of Econometrics*, 2021, (Forthcoming).

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd, “Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme,” *The Review of Economic Studies*, October 1997, *64* (4), 605–654.

Rios-Avila, Fernando, Asjad Naqvi, and Pedro H. C. Sant’Anna, “drdid: Doubly robust difference-in-differences estimators in Stata,” *Working Paper*, 2021.

Sant’Anna, Pedro H. C. and Jun Zhao, “Doubly robust difference-in-differences estimators,” *Journal of Econometrics*, November 2020, *219* (1), 101–122.

Sun, Liyan and Sarah Abraham, “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects,” *Journal of Econometrics*, 2020, (Forthcoming).