Recent Advances in Treatment Effect Estimation with Staggered Entry

Jeff Wooldridge
Department of Economics
Michigan State University

Bank of Italy
October 21, 2022

1. Introduction
2. Staggered Interventions, Parameters, and Assumptions
3. Limitations of the Basic TWFE Estimator
4. A Flexible, Regression-Based Approach
5. All Units Eventually Treated
6. Imputation Methods
7. Rolling and Long-Differencing Methods
8. A Small Simulation Study
9. A Strategy with Exit
10. Testing and Relaxing Parallel Trends
11. Other Directions

1. Introduction

• Exploiting different timing of interventions can be powerful for determining causality.

• Still fairly common to use a constant effect model estimated by TWFE.

  ▸ Sometimes time-varying covariates are included – have to be careful with this.

• Constant effect model can be too restrictive.

  ▸ Resulting estimates can consistently estimate uninteresting weighted averages of “treatment” effects.

- At least two reactions to the limitations of basic TWFE:

1. Try to characterize the nature of the TWFE estimates.
2. Use more flexible models/estimation methods that allow more heterogeneity.
   - Possible with or without time-constant controls.

2. Staggered Interventions and Assumptions

- T time periods with no units treated in $t = 1$.
- First unit is treated at $t = q < T$.
  - Already covered $q = T$ (no staggering).
- No reversibility: once a unit is subjected to the intervention, it stays in place.
- Treated units are added up through $t = T$.
- Is there a never treated group?
  - Determines whether certain ATTs are identified.

- How to define the potential outcomes that lead to simple definitions of ATTs?
- For $g \in \{q, \ldots, T\}$, $Y_t(g)$ is the outcome if the unit is first subjected to the intervention at time g.
  - In $Y_t(g)$, the number of treated periods decreases with g.
  - $Y_t(T)$ is the outcome treated in only the final period.
  - Adopt Athey and Imbens (2021, JofE): $Y_t(\infty)$ is the outcome if a unit is never treated in $\{q, \ldots, T\}$.

- Treatment effects of primary focus:

$$TE_{gt} = Y_t(g) - Y_t(\infty), g = q, \ldots, T; \; t = g, \ldots, T$$

  - ▸ For any t, $Y_t(\infty)$ is the outcome in the control state.

- Exhaustive and mutually exclusive dummy variables:

$$D_g = 1 \text{ if unit is first subjected to intervention at } g \in \{q, \ldots, T\}$$

$$D_\infty = 1 - (D_q + D_{q+1} + \cdots + D_T)$$

  - ▸ $D_{i\infty} = 1$ means unit i is never treated (up through T).
  - ▸ $D_{i\infty} = 0$ for all i means that all units are treated by time T.

- We hope to estimate

$$\tau_{gt} \equiv E[Y_t(g) - Y_t(\infty)|D_g = 1], t = g, g+1, \ldots, T$$

  - Sometimes the focus is on the instantaneous effects, τ_{gg}.
  - τ_{gt}, $t > g$ allows us to estimate persistence.
  - If $D_\infty \equiv 0$ then we cannot estimate τ_{gT} (final period treatment effects).
  - Later: We will be able to estimate τ_{gt} for $t < T$.
- For simplicity, assume a stronger form of no anticipation.

Assumption NA (**No Anticipation**): All pre-intervention treatment effects are zero:

$$Y_t(g) = Y_t(\infty), t \in \{1, 2, \ldots, g-1\}, g \in \{q, \ldots, T\}. \square$$

- Implies $\tau_{gt} = 0, t < g$.

- Let W_{it} be the time-varying treatment indicator.
  - $W_{i,t-1} = 1 \Rightarrow W_{it} = 1$.
- Define dummies

$$pg_t = 1 \text{ if } t \geq g$$
$$pg_t = fg_t + \cdots + fT_t$$

- Then

$$W_t = D_q \bullet pq_t + D_{q+1} \bullet p(q+1)_t + \cdots + D_T \bullet pT_t$$

- Consider a simple case with $T = 3$.
  - In $t = 1$, no units treated.
  - Some units are treated in $t = 2$.
  - Some more units are treated in $t = 3$.
- Potential outcomes:
  - Outcome in the never treated state: $Y_t(\infty)$.
  - Outcome if unit first treated at $t = 2$: $Y_t(2)$.
  - Outcome if unit first treated at $t = 3$: $Y_t(3)$.

• No anticipation:

$$Y_1(2) = Y_1(3) = Y_1(\infty);\ Y_2(3) = Y_2(\infty)$$

▸ The observed outcome in the first period is the potential outcome in all three states.

▸ In $t = 2$, the $g = 3$ and $g = \infty$ cohorts are not treated so they are both control units.

• Treatment effects of primary interest:

$$Y_2(2) - Y_2(\infty)$$
$$Y_3(2) - Y_3(\infty)$$
$$Y_3(3) - Y_3(\infty)$$

- The ATTs are

$$\tau_{22} = E[Y_2(2) - Y_2(\infty)|D_2 = 1]$$
$$\tau_{23} = E[Y_3(2) - Y_3(\infty)|D_2 = 1]$$
$$\tau_{33} = E[Y_3(3) - Y_3(\infty)|D_3 = 1]$$

- With a never treated group and parallel trends, a suitable conditional mean identifies all three.

Assumption PT (**Parallel Trends**): For $t = 2, \ldots, T$,

$$E[Y_t(\infty) - Y_1(\infty)|D_q, \ldots, D_T] = E[Y_t(\infty) - Y_1(\infty)] \square$$

or

• Interpretation is the same as before but we have more treated groups.

Assumption CPT (**Conditional Parallel Trends**): For $t = 2, \ldots, T$ and controls $\mathbf{X}$,

$$E[Y_t(\infty) - Y_1(\infty)|D_q, \ldots, D_T, \mathbf{X}] = E[Y_t(\infty) - Y_1(\infty)|\mathbf{X}]. \ \square$$

- We observe $\{D_\infty, D_q, D_{q+1}, \ldots, D_T\}$ and

$$Y_t = D_\infty \cdot Y_t(\infty) + D_q \cdot Y_t(q) + \cdots + D_T \cdot Y_t(T)$$

▸ If $D_{ig} = 0$ for all i for some $q < g \leq T$, simply drop that dummy: no units in that cohort.

3. Shortcomings of Standard TWFE

- The TWFE estimator in the staggered case (or any case) is motivated by the equation

$$Y_{it} = \beta W_{it} + \theta_2 f2_t + \cdots + \theta_T fT_t + C_i + U_{it}$$

- Within transformation removes C_i.
- Put dummies in for time effects.
- This is what `xtdidreg` does.
- Same as including cross-sectional and time period dummies in a long regression.

- The estimator, $\hat{\beta}_{TWFE}$, gives a single treatment effect.
- Wooldridge (2021, WP) shows $\hat{\beta}_{TWFE} = \hat{\beta}_{POLS}$ from

$$Y_{it} \text{ on } W_{it}, 1, D_{iq}, ..., D_{iT}, fq_t, ..., fT_t, t = 1, \ldots, T;\ i = 1, \ldots, N$$

▸ Controlling for the cohort indicators and post-intervention time dummies delivers $\hat{\beta}_{TWFE}$.

▸ Adding time-constant controls $\mathbf{X}_i$ does not change $\hat{\beta}_{POLS}$.

- Goodman-Bacon (2021, Journal of Econometrics): Generally, $\hat{\beta}_{TWFE}$ is a weighted average of many 2×2 DID estimators.
- Generally difficult to interpret with staggered interventions and heterogenous/time-varying effects.
  - Weights depend on sample size and variability of treatment within comparison groups.
- Some of the 2×2 DIDs have a sensible (causal) interpretations.

- de Chaisemartin and D’Haultfoeuille (2020, AER) derive a different weighted average representation of TWFE where the estimates have causal interpretation.
  - Weights sum to unity, but some can be negative.

- Sun and Abraham (2021, JofE): TWFE of event study estimators, which put leads and lags of treatment indicator.
  - Like TWFE estimation of constant effect, difficult to interpret estimates.
  - Evidence of “pre-trends” can be due to TE heterogeneity.
  - Propose TWFE with heterogeneous effects (very similar to below).

4. A Flexible, Regression-Based Approach

- Problem with basic TWFE is that it estimates a single treatment effect.
  - ▸ But it is simple to allow for a full set of ATTs.
- Apply TWFE, POLS estimators to more flexible models.
- Underlying this is random sampling across i.
- Also, assignment at the unit level.
  - ▸ Identification results are more general.
  - ▸ Standard error calculations generally change.

No Covariates

- Wooldridge (2021, WP) shows that, under Assumptions NA and CT,

$$
\begin{aligned}
E(Y_{it}|D_{iq}, D_{i,q+1}, \ldots, D_{iT}) &= \eta + \lambda_q D_{iq} + \cdots + \lambda_T D_{iT} \\
&\quad + \sum_{s=2}^{T} \theta_s fs_t + \sum_{g=q}^{T} \sum_{s=g}^{T} \tau_{gs}(D_{ig} \cdot fs_t)
\end{aligned}
$$

- With some entry in every period and a never treated group, all τ_{gt} are identified.
  - Include all interactions that makes sense.

- Intuitively makes sense: With sufficient variation, we can identify an ATT for cohort g in periods in which they are subjected to the intervention, $t \in \{g, \ldots, T\}$.
  - But need a suitable control group.
- If all units are treated by $t = T$, the treatment effects for the last period, τ_{gT}, are not identified.
  - Some other treatment effects are if we strengthen the CT assumption.

• Equivalent POLS regressions:

$$Y_{it} \text{ on } 1, D_{iq}, \ldots, D_{iT}, f2_t, \ldots, fT_t,$$
$$D_{iq} \cdot fq_t, \ldots, D_{iq} \cdot fT_t, \ \ldots, D_{iT} \cdot fT_t$$

or

$$Y_{it} \text{ on } 1, D_{iq}, \ldots, D_{iT}, f2_t, \ldots, fT_t,$$
$$W_{it} \cdot D_{iq} \cdot fq_t, \ldots, W_{it} \cdot D_{iq} \cdot fT_t, \ \ldots, W_{it} \cdot D_{iT} \cdot fT_t$$

• The time-varying treatment effect, W_{it}, is interacted with cohort and time dummies.

▸ Formulation is useful later for obtaining standard errors.

- Equivalently [Wooldridge (2021, WP)], use TWFE with all of the treatment interactions,

$$D_{iq} \cdot fq_t,\ \ldots,\ D_{iq} \cdot fT_t,\ \ldots,\ D_{iT} \cdot fT_t$$

  - An extended version of TWFE proposed by Sun and Abraham (2021, J of E).
  - Consistency of ETWFE holds under the same assumptions as other approaches (NA and CT) – even weaker in some cases.
- New commands `jwdid` in Stata, `{etwfe}` in R.
  - Simple to use standard Stata commands (`regress, xtreg, margins`).

Aggregation and Imposing Restrictions

- Even with a moderate number of treatment periods and treatment cohorts, the number of treatment effects can be large.
- Suppose $T = 10$, $q = 6$: Five treatment periods.
- With entry in every period and a never treated group, number of ATTs is

$$5(5 + 1)/2 = 15$$

  - Sometime there will be fewer entry cohorts.
  - Sometimes all units are eventually treated.

- Especially without a lot of data, may want to impose restrictions.
  - Basic TWFE says all ATTs are the same.
  - East to test (cluster robust) with POLS or extended TWFE.
- Could test null of common effects within a treatment cohort (so no time variation).
- Could test null of common effects by calendar time.
- Could assume only intensity (amount of time subjected to treatment) matters.

- Approaches in Stata:

1. Estimate the unrestricted equation; use `lincom` to aggregate.
   - Any restrictions can be tested.
2. Impose the restrictions directly before estimating the model.
   - For example, create dummies for different intensities of treatment.

EXAMPLE: Effects of Enterprise Zones on Unemployment Claims [Papke (1994, Journal of Public Economics)]

- Two entry times, 1984 and 1985

```
. use ezunem, clear
.
. xtset city year

Panel variable: city (strongly balanced)
 Time variable: year, 1980 to 1988
         Delta: 1 unit
.
. egen totez = sum(ez), by(city)
.
. * Create the cohort treatment dummies (entry period):
.
. gen d4 = totez == 5

. gen d5 = totez == 4
.
. gen post84 = year >= 1984

. gen post85 = year >= 1985
```

```
. * Constant treatment effect, common trends:
.
. xtreg luclms ez i.year, fe cluster(city)

Fixed-effects (within) regression               Number of obs     =        198
Group variable: city                            Number of groups  =         22

                                  (Std. err. adjusted for 22 clusters in city)
------------------------------------------------------------------------------
             |               Robust
      luclms | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          ez |  -.1044148   .0728066    -1.43   0.166    -.2558245    .0469948
             |
        year |
       1981  |  -.3216319   .0472151    -6.81   0.000     -.419821   -.2234428
       1982  |   .1354957   .0570696     2.37   0.027     .0168129    .2541785
       1983  |  -.2192554   .0846508    -2.59   0.017    -.3952964   -.0432144
       1984  |  -.5791517   .0987913    -5.86   0.000    -.7845994   -.3737041
       1985  |  -.5917868   .0908661    -6.51   0.000    -.7807533   -.4028204
       1986  |  -.6212648   .0782737    -7.94   0.000    -.7840438   -.4584858
       1987  |  -.8889486   .0836945   -10.62   0.000    -1.063001   -.7148963
       1988  |  -1.227633   .0958602   -12.81   0.000    -1.426985   -1.028281
             |
       _cons |   11.69439   .0593721   196.97   0.000     11.57092    11.81786
-------------+----------------------------------------------------------------
```

```
. * Let the effect differ by cohort and year:
.
. xtreg luclms c.ez#c.d4#c.d84 c.ez#c.d4#c.d85 c.ez#c.d4#c.d86 c.ez#c.d4#c.d87 c.ez#c.d4#
>   c.ez#c.d5#c.d85 c.ez#c.d5#c.d86 c.ez#c.d5#c.d87 c.ez#c.d5#c.d88 ///
>   i.year, fe vce(cluster city)

Fixed-effects (within) regression               Number of obs     =        198
Group variable: city                            Number of groups  =         22

                                 (Std. err. adjusted for 22 clusters in city)
------------------------------------------------------------------------------
                |               Robust
         luclms | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
----------------+-------------------------------------------------------------
c.ez#c.d4#c.d84 |  -.0597296   .1209858    -0.49   0.627    -.3113334    .1918742
                |
c.ez#c.d4#c.d85 |  -.0442902   .1071004    -0.41   0.683    -.2670178    .1784373
                |
c.ez#c.d4#c.d86 |  -.1062723   .0992368    -1.07   0.296    -.3126466     .100102
                |
c.ez#c.d4#c.d87 |  -.0349116   .1140226    -0.31   0.762    -.2720347    .2022114
                |
c.ez#c.d4#c.d88 |   .0893613   .1147503     0.78   0.445     -.149275    .3279975
                |
c.ez#c.d5#c.d85 |  -.0228812   .1095033    -0.21   0.836    -.2506057    .2048434
                |
c.ez#c.d5#c.d86 |  -.1223598   .0859244    -1.42   0.169    -.3010494    .0563299
                |
c.ez#c.d5#c.d87 |  -.3380748   .0676865    -4.99   0.000    -.4788366   -.1973131
                |
c.ez#c.d5#c.d88 |  -.3371358   .1551199    -2.17   0.041    -.6597253   -.0145463
```

```
        year |
        1981  |  -.3216319   .0482529    -6.67   0.000    -.4219793   -.2212845
        1982  |   .1354957   .0583241     2.32   0.030     .0142042    .2567872
        1983  |  -.2192554   .0865115    -2.53   0.019    -.3991659   -.0393449
        1984  |  -.5913386     .10804    -5.47   0.000      -.81602   -.3666572
        1985  |  -.6230087   .1060464    -5.87   0.000    -.8435442   -.4024733
        1986  |  -.6174955   .0819437    -7.54   0.000    -.7879068   -.4470842
        1987  |  -.8654204   .0797543   -10.85   0.000    -1.031279   -.6995621
        1988  |  -1.238168     .11074   -11.18   0.000    -1.468465   -1.007872
              |
        _cons |   11.69439   .0592368   197.42   0.000      11.5712    11.81758
--------------+----------------------------------------------------------------
      sigma_u |   .5508089
      sigma_e |  .19687755
          rho |   .8867145   (fraction of variance due to u_i)
-------------------------------------------------------------------------------
```

```
. * Average effects by treatment cohort:
.
. lincom (c.ez#c.d4#c.d84 + c.ez#c.d4#c.d85 + c.ez#c.d4#c.d86 + c.ez#c.d4#c.d87 ///
>    + c.ez#c.d4#c.d88)/5

 ( 1)  .2*c.ez#c.d4#c.d84 + .2*c.ez#c.d4#c.d85 + .2*c.ez#c.d4#c.d86 +
       .2*c.ez#c.d4#c.d87 + .2*c.ez#c.d4#c.d88 = 0

------------------------------------------------------------------------------
      luclms | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |  -.0311685   .0881576    -0.35   0.727    -.2145023    .1521653
------------------------------------------------------------------------------

. lincom (c.ez#c.d5#c.d85 + c.ez#c.d5#c.d86 + c.ez#c.d5#c.d87 + c.ez#c.d5#c.d88)/4

 ( 1)  .25*c.ez#c.d5#c.d85 + .25*c.ez#c.d5#c.d86 + .25*c.ez#c.d5#c.d87
       + .25*c.ez#c.d5#c.d88 = 0

------------------------------------------------------------------------------
      luclms | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |  -.2051129   .0667865    -3.07   0.006    -.3440029   -.0662229
------------------------------------------------------------------------------
```

```
. * Test null that effect changes only by cohort, not time period:
.
. qui xtreg luclms c.ez#c.d4 c.ez#c.d4#c.d85 c.ez#c.d4#c.d86 c.ez#c.d4#c.d87 ///
>   c.ez#c.d4#c.d88 c.ez#c.d5 c.ez#c.d5#c.d86 c.ez#c.d5#c.d87 c.ez#c.d5#c.d88 ///
>   i.year, fe vce(cluster city)

.
. test c.ez#c.d4#c.d85 c.ez#c.d4#c.d86 c.ez#c.d4#c.d87 c.ez#c.d4#c.d88 ///
>         c.ez#c.d5#c.d86 c.ez#c.d5#c.d87 c.ez#c.d5#c.d88

 ( 1)  c.ez#c.d4#c.d85 = 0
 ( 2)  c.ez#c.d4#c.d86 = 0
 ( 3)  c.ez#c.d4#c.d87 = 0
 ( 4)  c.ez#c.d4#c.d88 = 0
 ( 5)  c.ez#c.d5#c.d86 = 0
 ( 6)  c.ez#c.d5#c.d87 = 0
 ( 7)  c.ez#c.d5#c.d88 = 0

       F(  7,    21) =    4.33
            Prob > F =    0.0041
```

```
. * Generated intensity-of-treatment dummies:
.
. gen intens1 = d4*d84 + d5*d85

. gen intens2 = d4*d85 + d5*d86

. gen intens3 = d4*d86 + d5*d87

. gen intens4 = d4*d87 + d5*d88

. gen intens5 = d4*d88
.
. xtreg luclms intens1 intens2 intens3 intens4 ///
>         intens5 i.year, fe vce(cluster city)

Fixed-effects (within) regression               Number of obs     =        198
Group variable: city                            Number of groups  =         22

                                  (Std. err. adjusted for 22 clusters in city)
------------------------------------------------------------------------------
             |               Robust
      luclms | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     intens1 |  -.0570786   .0872997    -0.65   0.520    -.2386282     .124471
     intens2 |  -.1040937   .0763929    -1.36   0.187    -.2629615    .0547742
     intens3 |  -.2243052   .0811183    -2.77   0.012        -.393   -.0556105
     intens4 |  -.1518154   .1076929    -1.41   0.173     -.375775    .0721442
     intens5 |   .0961228   .1136402     0.85   0.407    -.1402049    .3324504
```

year						
1981	-.3216319	.0477255	-6.74	0.000	-.4208826	-.2223813
1982	.1354957	.0576866	2.35	0.029	.0155298	.2554616
1983	-.2192554	.085566	-2.56	0.018	-.3971996	-.0413112
1984	-.5920616	.1074185	-5.51	0.000	-.8154506	-.3686727
1985	-.600481	.0948249	-6.33	0.000	-.7976801	-.4032819
1986	-.5886258	.0787799	-7.47	0.000	-.7524576	-.424794
1987	-.8542229	.0831583	-10.27	0.000	-1.02716	-.6812857
1988	-1.273707	.1053944	-12.09	0.000	-1.492887	-1.054527
_cons	11.69439	.0595819	196.27	0.000	11.57049	11.8183

Adding Covariates

- Continue to maintain the no anticipation assumption:

$$Y_t(g) = Y_t(\infty),\ t < g$$

- Conditional common (parallel) trends assumption:

$$E[Y_t(\infty) - Y_1(\infty)|D_q,\ldots,D_T,\mathbf{X}]$$
$$= E[Y_t(\infty) - Y_1(\infty)|\mathbf{X}],\ t = 2,\ldots,T$$

• Act *as if* all expectations are linear in **X**:

$$E(Y_t|D_q,\ldots,D_T,\mathbf{X}) = \eta + \sum_{g=q}^{T}\lambda_g D_g + \mathbf{X}\boldsymbol{\kappa} + \sum_{g=q}^{T}(D_g \cdot \mathbf{X})\boldsymbol{\zeta}_g$$

$$+ \sum_{s=2}^{T}\theta_s fs_t + \sum_{s=2}^{T}(fs_t \cdot \mathbf{X})\boldsymbol{\pi}_t$$

$$+ \sum_{g=q}^{T}\sum_{s=g}^{T}\tau_{gs}(D_g \cdot fs_t) + \sum_{g=q}^{T}\sum_{s=g}^{T}(D_g \cdot fs_t \cdot \dot{\mathbf{X}}_g)\boldsymbol{\rho}_{gs}$$

$$\dot{\mathbf{X}}_g \equiv \mathbf{X} - E(\mathbf{X}|D_g = 1)$$

• For a random draw i, and with the treatment indicator W_{it}:

$$\begin{aligned} E(Y_{it}|D_{iq},\ldots,D_{iT},\mathbf{X}_i) &= \eta + \sum_{g=q}^{T} \lambda_g D_{ig} + \mathbf{X}_i \boldsymbol{\kappa} + \sum_{g=q}^{T} (D_{ig} \cdot \mathbf{X}_i) \boldsymbol{\zeta}_g \\ &+ \sum_{s=2}^{T} \theta_s fs_t + \sum_{s=2}^{T} (fs_t \cdot \mathbf{X}_i) \boldsymbol{\pi}_t \\ &+ \sum_{g=q}^{T} \sum_{s=g}^{T} \tau_{gs} (W_{it} \cdot D_{ig} \cdot fs_t) \\ &+ \sum_{g=q}^{T} \sum_{s=g}^{T} (W_{it} \cdot D_{ig} \cdot fs_t \cdot \dot{\mathbf{X}}_{ig}) \boldsymbol{\rho}_{gs} \end{aligned}$$

- Lots of heterogeneity.
  - ▸ The ATTs τ_{gt} are unrestricted.
  - ▸ Effects of covariates can change with treatment cohort and calendar time.
  - ▸ Treament effects can change with $\mathbf{X}$: observed heterogeneity.
- To operationalize, use subsample averages:

$$\dot{\mathbf{X}}_{ig} = \mathbf{X}_i - \bar{\mathbf{X}}_g$$

- POLS (with variables rearranged) is

$$Y_{it} \text{ on } W_{it} \cdot D_{iq} \cdot fq_t, \ldots, W_{it} \cdot D_{iq} \cdot fT_t, \ldots,$$

$$W_{it} \cdot D_{i,q+1} \cdot f(q+1)_t, \ldots, W_{it} \cdot D_{i,q+1} \cdot fT_t, \ldots,$$

$$W_{it} \cdot D_{iT} \cdot fT_t,$$

$$W_{it} \cdot D_{iq} \cdot fq_t \cdot \dot{\mathbf{X}}_{iq}, \ldots, W_{it} \cdot D_{iq} \cdot fT_t \cdot \dot{\mathbf{X}}_{iq},$$

$$W_{it} \cdot D_{i,q+1} \cdot f(q+1)_t \cdot \dot{\mathbf{X}}_{i,q+1}, \ldots, W_{it} \cdot D_{i,q+1} \cdot fT_t \cdot \dot{\mathbf{X}}_{i,q+1}, \ldots,$$

$$W_{it} \cdot D_{iT} \cdot fT_t \cdot \dot{\mathbf{X}}_{iT},$$

$$1, D_{iq}, \ldots, D_{iT}, \mathbf{X}_i, D_{ig} \cdot \mathbf{X}_i, \ldots, D_{iT} \cdot \mathbf{X}_i,$$

$$f2_t, \ldots, fT_t, \; f2_t \cdot \mathbf{X}_i, \ldots, fT_t \cdot \mathbf{X}_i,$$

  - Can drop W_{it} everywhere to get the $\hat{\tau}_{gt}$.

- Can adjust the standard errors for sampling variation in

$$\bar{\mathbf{X}}_g = N_g^{-1} \sum_{i=1}^{N} D_{ig} \mathbf{X}_i$$

  - Can trick the `margins` command to do it.

- In addition to restricting the τ_{gt}, might want to impose constant coefficients on the interactions with $\dot{\mathbf{X}}_{ig}$.

- See Wooldridge (2021, WP) for proof of BLUE under a set of classical assumptions.

  - Homoskedasticity, no serial correlation in idiosyncratic errors.
  - Asymptotically efficient under the same assumptions.

```
. use did_staggered_6, clear

. list id year y w dinf d4 d5 d6 in 31/60, sep(6)

     +-------------------------------------------------+
     | id   year          y   w   dinf   d4   d5   d6 |
     |-------------------------------------------------|
 31. |  6   2001   15.69385   0      0    1    0    0 |
 32. |  6   2002   19.80592   0      0    1    0    0 |
 33. |  6   2003   18.36681   0      0    1    0    0 |
 34. |  6   2004   18.23068   1      0    1    0    0 |
 35. |  6   2005   27.81216   1      0    1    0    0 |
 36. |  6   2006   22.54216   1      0    1    0    0 |
     |-------------------------------------------------|
 37. |  7   2001   21.20615   0      0    0    0    1 |
 38. |  7   2002    18.0428   0      0    0    0    1 |
 39. |  7   2003   25.29553   0      0    0    0    1 |
 40. |  7   2004    24.0236   0      0    0    0    1 |
 41. |  7   2005   21.16335   0      0    0    0    1 |
 42. |  7   2006   30.44825   1      0    0    0    1 |
     |-------------------------------------------------|
 43. |  8   2001   22.04953   0      1    0    0    0 |
 44. |  8   2002   19.01489   0      1    0    0    0 |
 45. |  8   2003   21.26308   0      1    0    0    0 |
 46. |  8   2004   23.58504   0      1    0    0    0 |
 47. |  8   2005   21.93984   0      1    0    0    0 |
 48. |  8   2006   22.75257   0      1    0    0    0 |
     |-------------------------------------------------|
```

49.	9	2001	21.30811	0	0	0	0	1
50.	9	2002	17.02841	0	0	0	0	1
51.	9	2003	18.96646	0	0	0	0	1
52.	9	2004	15.77348	0	0	0	0	1
53.	9	2005	20.06797	0	0	0	0	1
54.	9	2006	25.63022	1	0	0	0	1
55.	10	2001	17.60084	0	0	0	1	0
56.	10	2002	15.69303	0	0	0	1	0
57.	10	2003	18.4134	0	0	0	1	0
58.	10	2004	18.6501	0	0	0	1	0
59.	10	2005	20.35908	1	0	0	1	0
60.	10	2006	26.95766	1	0	0	1	0

```
. tab dinf if f01

       dinf |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        238       47.60       47.60
          1 |        262       52.40      100.00
------------+-----------------------------------
      Total |        500      100.00

. tab d4 if f01

         d4 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        381       76.20       76.20
          1 |        119       23.80      100.00
------------+-----------------------------------
      Total |        500      100.00
```

```
. tab d5 if f01

         d5 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        417       83.40       83.40
          1 |         83       16.60      100.00
------------+-----------------------------------
      Total |        500      100.00

. tab d6 if f01

         d6 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        464       92.80       92.80
          1 |         36        7.20      100.00
------------+-----------------------------------
      Total |        500      100.00
```

```
. sum te4 if d4 & f04

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         te4 |        119    3.757094     2.02284  -1.457056   8.437983

. sum te4 if d4 & f05

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         te4 |        119    4.017567     1.92296   .1231937   8.042234

. sum te4 if d4 & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         te4 |        119    4.578953    1.909807   .1030998   9.104858
```

```
. sum te5 if d5 & f05

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         te5 |         83    2.737543    2.034552  -2.907673   8.347282

. sum te5 if d5 & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         te5 |         83    3.516055     1.99553  -1.392517   7.539985

. sum te6 if d6 & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         te6 |         36    1.848862    2.032388  -2.511634   6.099405

. xtset id year

Panel variable: id (strongly balanced)
 Time variable: year, 2001 to 2006
         Delta: 1 unit
```

```
. xtreg y c.w#c.d4#c.f04 c.w#c.d4#c.f05 c.w#c.d4#c.f06 ///
>         c.w#c.d5#c.f05 c.w#c.d5#c.f06 ///
>         c.w#c.d6#c.f06 ///
>         c.w#c.d4#c.f04#c.x_dm4 c.w#c.d4#c.f05#c.x_dm4 c.w#c.d4#c.f06#c.x_dm4 ///
>         c.w#c.d5#c.f05#c.x_dm5 c.w#c.d5#c.f06#c.x_dm5 ///
>         c.w#c.d6#c.f06#c.x_dm6 ///
>         i.year i.year#c.x, fe vce(cluster id)
note: 2006.year#c.x omitted because of collinearity.

Fixed-effects (within) regression               Number of obs     =      3,000
Group variable: id                              Number of groups  =        500

                                     (Std. err. adjusted for 500 clusters in id)
------------------------------------------------------------------------------------------
                       |               Robust
                     y | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-----------------------+------------------------------------------------------------------
        c.w#c.d4#c.f04 |   3.509903    .302774    11.59   0.000     2.915034    4.104773
                       |
        c.w#c.d4#c.f05 |   4.240069     .32866    12.90   0.000     3.594341    4.885797
                       |
        c.w#c.d4#c.f06 |   4.209541   .3302325    12.75   0.000     3.560724    4.858359
                       |
        c.w#c.d5#c.f05 |    3.00813   .3405944     8.83   0.000     2.338955    3.677306
                       |
        c.w#c.d5#c.f06 |   3.701545   .3676242    10.07   0.000     2.979263    4.423827
                       |
        c.w#c.d6#c.f06 |   2.016938   .5702174     3.54   0.000     .8966146     3.13726
                       |
```

c.w#c.d4#c.f04#c.x_dm4	-.3730017	1.044654	-0.36	0.721	-2.425463	1.67946
c.w#c.d4#c.f05#c.x_dm4	.5322265	1.044033	0.51	0.610	-1.519015	2.583468
c.w#c.d4#c.f06#c.x_dm4	2.587535	1.063089	2.43	0.015	.4988523	4.676218
c.w#c.d5#c.f05#c.x_dm5	.9587673	.6319748	1.52	0.130	-.2828921	2.200427
c.w#c.d5#c.f06#c.x_dm5	.1569361	.8613957	0.18	0.856	-1.535473	1.849346
c.w#c.d6#c.f06#c.x_dm6	.7239738	1.170036	0.62	0.536	-1.574831	3.022779

```
        year |
        2002 |  -.1800947   .3434733    -0.52   0.600    -.8549268    .4947375
        2003 |  -.7864622   .3321942    -2.37   0.018    -1.439134   -.1337906
        2004 |   .2930123    .369072     0.79   0.428    -.4321143    1.018139
        2005 |   .0521642   .4233275     0.12   0.902    -.7795599    .8838882
        2006 |   .3885514   .4335716     0.90   0.371    -.4632994    1.240402
             |
    year#c.x |
        2001 |  -.3602849   .4210227    -0.86   0.393     -1.18748    .4669107
        2002 |   .0356095   .4010666     0.09   0.929    -.7523778    .8235968
        2003 |   .6152694   .4202289     1.46   0.144    -.2103667    1.440905
        2004 |  -.0296922   .4088543    -0.07   0.942    -.8329803    .7735959
        2005 |  -.1214381   .4492112    -0.27   0.787    -1.004017    .7611404
        2006 |          0  (omitted)
             |
       _cons |   20.16448   .4184447    48.19   0.000     19.34235    20.98661
-------------+----------------------------------------------------------------
```

```
. reg y c.w#c.d4#c.f04 c.w#c.d4#c.f05 c.w#c.d4#c.f06 ///
>         c.w#c.d5#c.f05 c.w#c.d5#c.f06 ///
>         c.w#c.d6#c.f06 ///
>         c.w#c.d4#c.f04#c.x_dm4 c.w#c.d4#c.f05#c.x_dm4 c.w#c.d4#c.f06#c.x_dm4 ///
>         c.w#c.d5#c.f05#c.x_dm5 c.w#c.d5#c.f06#c.x_dm5 ///
>         c.w#c.d6#c.f06#c.x_dm6 ///
>         i.year i.year#c.x ///
>         d4 d5 d6 x c.d4#c.x c.d5#c.x c.d6#c.x, vce(cluster id)

Linear regression                               Number of obs     =      3,000

                                    (Std. err. adjusted for 500 clusters in id)
-------------------------------------------------------------------------------
                       |               Robust
                     y | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-----------------------+-------------------------------------------------------
        c.w#c.d4#c.f04 |   3.509903   .3031306    11.58   0.000     2.914334    4.105473
                       |
        c.w#c.d4#c.f05 |   4.240069    .329047    12.89   0.000     3.593581    4.886558
                       |
        c.w#c.d4#c.f06 |   4.209541   .3306214    12.73   0.000     3.559959    4.859123
                       |
        c.w#c.d5#c.f05 |    3.00813   .3409955     8.82   0.000     2.338167    3.678094
                       |
        c.w#c.d5#c.f06 |   3.701545   .3680571    10.06   0.000     2.978413    4.424678
                       |
        c.w#c.d6#c.f06 |   2.016938    .570889     3.53   0.000     .8952951     3.13858
                       |
```

c.w#c.d4#c.f04#c.x_dm4	-.3730017	1.045884	-0.36	0.722	-2.42788	1.681877
c.w#c.d4#c.f05#c.x_dm4	.5322266	1.045262	0.51	0.611	-1.521431	2.585884
c.w#c.d4#c.f06#c.x_dm4	2.587535	1.064342	2.43	0.015	.4963923	4.678679
c.w#c.d5#c.f05#c.x_dm5	.9587673	.6327191	1.52	0.130	-.2843545	2.201889
c.w#c.d5#c.f06#c.x_dm5	.1569361	.8624102	0.18	0.856	-1.537467	1.851339
c.w#c.d6#c.f06#c.x_dm6	.7239738	1.171414	0.62	0.537	-1.577539	3.025486

```
        year |
        2002 |  -.1800947   .3438779    -0.52   0.601    -.8557216    .4955323
        2003 |  -.7864622   .3325854    -2.36   0.018    -1.439903   -.1330219
        2004 |   .2930123   .3695067     0.79   0.428    -.4329683    1.018993
        2005 |   .0521642   .4238261     0.12   0.902    -.7805394    .8848678
        2006 |   .3885514   .4340822     0.90   0.371    -.4643027    1.241405
             |
    year#c.x |
        2002 |   .3958943   .3275757     1.21   0.227    -.2477033    1.039492
        2003 |   .9755543   .3194763     3.05   0.002     .3478697    1.603239
        2004 |   .3305927   .3509236     0.94   0.347    -.3588773    1.020063
        2005 |   .2388468    .398165     0.60   0.549    -.5434397    1.021133
        2006 |   .3602849   .4215185     0.85   0.393    -.4678849    1.188455
             |
          d4 |   .0351479   .8603456     0.04   0.967    -1.655198    1.725494
          d5 |  -1.792846   .6808996    -2.63   0.009    -3.130629   -.4550624
          d6 |  -2.537934   .9703164    -2.62   0.009    -4.444343   -.6315248
           x |    .240952   .4526635     0.53   0.595    -.6484094    1.130313
             |
     c.d4#c.x |  -1.191253   .9341302    -1.28   0.203    -3.026566    .6440597
             |
     c.d5#c.x |   .5511332   .6069469     0.91   0.364    -.6413531     1.74362
             |
     c.d6#c.x |   .9944783   .8275699     1.20   0.230    -.6314725    2.620429
             |
        _cons |    20.1107   .4531656    44.38   0.000     19.22035    21.00105
------------------------------------------------------------------------------

. * See did_staggered_6.do for margins command to adjust standard errors.
```

5. All Units Eventually Treated

- $T = 3$, staggered entry in $t = 2$, $t = 3$.
- With a never treated group, we can estimate

$$\begin{aligned} \tau_{22} &= E[Y_2(2) - Y_2(\infty)|D_2 = 1] \\ \tau_{23} &= E[Y_3(2) - Y_3(\infty)|D_2 = 1] \\ \tau_{33} &= E[Y_3(3) - Y_3(\infty)|D_3 = 1] \end{aligned}$$

- If $D_{i,\infty} \equiv 0$ for all i, we cannot estimate τ_{23}, τ_{33}.
- Instead, we can estimate τ_{22} and

$$\tau_{2|3,3} \equiv E[Y_3(2) - Y_3(3)|D_2 = 1]$$

  - For D_2 cohort, effect of being treated one period earlier.

- Need to change the PT assumption.
- Write

$$\begin{aligned} Y_3(2) - Y_3(3) &= [Y_3(2) - Y_1(2)] - [Y_3(3) - Y_1(3)] + [Y_1(2) - Y_1(3)] \\ &= [Y_3(2) - Y_1(2)] - [Y_3(3) - Y_1(3)] \end{aligned}$$

  - ▸ $Y_1(2) - Y_1(3) = 0$ by Assumption NA.

- Condition on $D_2 = 1$:

$$\begin{aligned} E[Y_3(2) - Y_3(3)|D_2 = 1] = E[Y_3(2) - Y_1(2)|D_2 = 1] \\ - E[Y_3(3) - Y_1(3)|D_2 = 1] \end{aligned}$$

  - ▸ Assume PT on the state of being first treated at $t = 3$:

$$E[Y_3(3) - Y_1(3)|D_2 = 1] = E[Y_3(3) - Y_1(3)|D_3 = 1]$$

$$\hat{\tau}_{2|3,3} = N_2^{-1}\sum_{i=1}^{N} D_{i2}\Delta Y_{i3} - N_3^{-1}\sum_{i=1}^{N} D_{i3}\Delta Y_{i3}$$

$$= N_2^{-1}\sum_{i=1}^{N} D_{i2}\Delta Y_{i3} - (N - N_2)^{-1}\sum_{i=1}^{N}(1 - D_{i2})\Delta Y_{i3}$$

- Both $\hat{\tau}_{22}$ and $\hat{\tau}_{2|3,3}$ are obtained from the POLS regression

$$Y_{it} \text{ on } 1, D_{i2}, f2_t, f3_t, D_{i2}\cdot f2_t, D_{i2}\cdot f3_t,\ t = 1,2,3;\ i = 1,\ldots,N$$

  - ▸ $D_{i2} + D_{i3} = 1$.
  - ▸ D_{i3} and $D_{i3}f3_t$ have been dropped.

• POLS/ETWFE extends immediately to the general case with all units treated by period T.

• Generally, the TEs are now

$$Y_t(g) - Y_t(T),\ g = q, \ldots, T-1;\ t = g, \ldots, T$$

▸ The gain in period t from first being treated in the earlier period g rather than the last period.

- The identified parameters are

$$\tau_{(g:T),t} \equiv E[Y_t(g) - Y_t(T)|D_g = 1],\ g = q, \ldots, T-1;\ t = g, \ldots, T$$

- The NA and CT assumptions are stated for the potential outcome $Y_t(T)$.
- If there *could* have been a never treated group, then NA implies

$$Y_t(T) = Y_t(\infty),\ t < T$$

- Then

$$\tau_{(g:T),t} = \tau_{gt},\ g = 1, \ldots, T-1; t = g, \ldots, G-1$$

  - Same treatment effects as before for $g < T$.

- Artificially make all units treated by last period.

```
. use did_staggered_6, clear
.
. * Lose identification of some effects if all units eventually treated.
. * But can estimate other effects.

. * Artificially make all units treated in last period:
.
. replace w = 1 if year == 2006
(262 real changes made)

. xtset id year

Panel variable: id (strongly balanced)
 Time variable: year, 2001 to 2006
         Delta: 1 unit

. egen wsum = sum(w), by(id)

. drop d4 d5 d6

. gen d4 = wsum == 3

. gen d5 = wsum == 2

. gen d6 = wsum == 1
```

```
. xtreg y c.d4#c.f04 c.d4#c.f05 c.d4#c.f06 ///
>         c.d5#c.f05 c.d5#c.f06 ///
>         c.d6#c.f06 i.year, fe vce(cluster id)
note: c.d6#c.f06 omitted because of collinearity.

Fixed-effects (within) regression               Number of obs     =      3,000
Group variable: id                              Number of groups  =        500

                                   (Std. err. adjusted for 500 clusters in id)
------------------------------------------------------------------------------
             |               Robust
           y | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  c.d4#c.f04 |   3.522951   .3014332    11.69   0.000     2.930717    4.115186
             |
  c.d4#c.f05 |   4.258536   .3259157    13.07   0.000       3.6182    4.898873
             |
  c.d4#c.f06 |   3.974459   .3401851    11.68   0.000     3.306087    4.642831
             |
  c.d5#c.f05 |   2.989588   .3415854     8.75   0.000     2.318465    3.660711
             |
  c.d5#c.f06 |   3.451409   .3668947     9.41   0.000      2.73056    4.172258
             |
  c.d6#c.f06 |          0  (omitted)
             |
```

```
        year |
        2002 |   .2073711   .1267308     1.64   0.102    -.0416207    .4563628
        2003 |   .1683225   .1223818     1.38   0.170    -.0721247    .4087696
        2004 |   .6134613   .1354924     4.53   0.000     .3472555    .8796671
        2005 |   .2846088   .1483121     1.92   0.056    -.0067842    .5760019
        2006 |   .9838574   .1657924     5.93   0.000     .6581201    1.309595
             |
       _cons |   19.81186   .0828789   239.05   0.000     19.64903     19.9747
-------------+----------------------------------------------------------------
     sigma_u |  2.2381961
     sigma_e |  2.1999905
         rho |  .50860774   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. reg y c.d4#c.f04 c.d4#c.f05 c.d4#c.f06 ///
>         c.d5#c.f05 c.d5#c.f06 ///
>         c.d6#c.f06 i.year d4 d5 d6, vce(cluster id)
note: c.d6#c.f06 omitted because of collinearity.
note: d6 omitted because of collinearity.

Linear regression                               Number of obs     =      3,000

                                   (Std. err. adjusted for 500 clusters in id)
------------------------------------------------------------------------------
             |               Robust
           y | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  c.d4#c.f04 |   3.522951   .3015341    11.68   0.000     2.930518    4.115384
             |
  c.d4#c.f05 |   4.258536   .3260248    13.06   0.000     3.617986    4.899087
             |
  c.d4#c.f06 |   3.974459    .340299    11.68   0.000     3.305864    4.643055
             |
  c.d5#c.f05 |   2.989588   .3416997     8.75   0.000      2.31824    3.660935
             |
  c.d5#c.f06 |   3.451409   .3670175     9.40   0.000     2.730319    4.172499
             |
  c.d6#c.f06 |          0  (omitted)
             |
```

```
        year |
        2002  |   .2073711   .1267732     1.64   0.103     -.041704    .4564462
        2003  |   .1683225   .1224228     1.37   0.170    -.0722052    .4088501
        2004  |   .6134613   .1355377     4.53   0.000     .3471664    .8797562
        2005  |   .2846088   .1483617     1.92   0.056    -.0068818    .5760994
        2006  |   .9838574   .1658479     5.93   0.000     .6580111    1.309704
              |
          d4 |  -.9452208   .2672017    -3.54   0.000      -1.4702   -.4202418
          d5 |  -.9814092    .278305    -3.53   0.000    -1.528203   -.4346153
          d6 |          0  (omitted)
       _cons |   20.19974   .1514283   133.39   0.000     19.90223    20.49726
------------------------------------------------------------------------------
```

6. Imputation Methods

- In the cross-sectional case, pooled OLS (across control and treated) with interactions is the same as imputation using regression.
- Same is true in the staggered DiD setting.
- Now the initial regression uses $W_{it} = 0$ observations.

- Under the Conditional Common Trends Assumption:

$$E[Y_{it}(\infty)|D_{iq},\ldots,D_{iT},\mathbf{X}_i] = \eta + \sum_{g=q}^{T}\lambda_g D_{ig} + \mathbf{X}_i\boldsymbol{\kappa} + \sum_{g=q}^{T}(D_{ig}\cdot\mathbf{X}_i)\boldsymbol{\zeta}_g + \sum_{s=2}^{T}\theta_s fs_t + \sum_{s=2}^{T}(fs_t\cdot\mathbf{X}_i)\boldsymbol{\pi}_t$$

- Obtained from the conditional expectation in observables with $W_{it} = 0$.

$$
\begin{aligned}
E(Y_{it}|D_{iq},\ldots,D_{iT},\mathbf{X}_i) \;=\; & \eta + \sum_{g=q}^{T} \lambda_g D_{ig} + \mathbf{X}_i \boldsymbol{\kappa} + \sum_{g=q}^{T} (D_{ig} \bullet \mathbf{X}_i)\boldsymbol{\zeta}_g \\
& + \sum_{s=2}^{T} \theta_s fs_t + \sum_{s=2}^{T} (fs_t \bullet \mathbf{X}_i)\boldsymbol{\pi}_t \\
& + \sum_{g=q}^{T} \sum_{s=g}^{T} \tau_{gs} (W_{it} \bullet D_{ig} \bullet fs_t) \\
& + \sum_{g=q}^{T} \sum_{s=g}^{T} (W_{it} \bullet D_{ig} \bullet fs_t \bullet \dot{\mathbf{X}}_{ig}) \boldsymbol{\rho}_{gs}
\end{aligned}
$$

• Imputation Approach:

(i) Using the $W_{it} = 0$ observations, run the pooled regression,

$$Y_{it} \text{ on } 1, D_{iq}, \ ..., D_{iT}, \mathbf{X}_i, D_{iq} \cdot \mathbf{X}_i, \ ..., \ D_{iT} \cdot \mathbf{X}_i,$$
$$f2_t, \ ..., fT_t, f2_t \cdot \mathbf{X}_i, \ ..., fT_t \cdot \mathbf{X}_i$$

and obtain the $\hat{\eta}$, $\hat{\lambda}_g$, $\hat{\boldsymbol{\zeta}}_g$, $\hat{\theta}_s$, $\hat{\boldsymbol{\pi}}_s$.

(ii) For the $W_{it} = 1$ subsample, obtain imputed TEs for each (i,t) :

$$\widehat{TE}_{it} = Y_{it} - \left[\hat{\eta} + \sum_{g=q}^{T} \hat{\lambda}_g D_{ig} + \mathbf{X}_i \hat{\boldsymbol{\kappa}} + \sum_{g=q}^{T} (D_{ig} \cdot \mathbf{X}_i) \hat{\boldsymbol{\zeta}}_g \right.$$

$$\left. + \sum_{s=2}^{T} \hat{\theta}_s f s_t + \sum_{s=2}^{T} (f s_t \cdot \mathbf{X}_i) \hat{\boldsymbol{\pi}}_s \right]$$

- Aggregate by treatment cohort/time period:

$$\tilde{\tau}_{gt} = N_g^{-1} \sum_{i=1}^{N} D_{ig} \cdot \widehat{TE}_{it} \quad \text{(cohort } g\text{, period } t\text{)}$$

- Wooldridge (2021, WP) shows

$$\tilde{\tau}_{gt} = \hat{\tau}_{gt}, g = q, \ldots, T, \; t = g, \ldots, T$$

- Also, the estimates $\hat{\eta}$, $\hat{\lambda}_g$, $\hat{\boldsymbol{\zeta}}_g$, $\hat{\theta}_s$, $\hat{\boldsymbol{\pi}}_s$ from the imputation method (using the $W_{it} = 0$ observations) are the same as the POLS estimates on the full sample.
  - Proof is in the paper's appendix.
- Not the same proposal as Borusyak, Jaravel, and Spiess (2021, WP): they use fixed effects in the first step.
  - But the estimates (with time-constant covariates) are still identical.

```
. use did_staggered_6

. reg y c.w#c.d4#c.f04 c.w#c.d4#c.f05 c.w#c.d4#c.f06 ///
>         c.w#c.d5#c.f05 c.w#c.d5#c.f06 ///
>         c.w#c.d6#c.f06 ///
>         c.w#c.d4#c.f04#c.x_dm4 c.w#c.d4#c.f05#c.x_dm4 c.w#c.d4#c.f06#c.x_dm4 ///
>         c.w#c.d5#c.f05#c.x_dm5 c.w#c.d5#c.f06#c.x_dm5 ///
>         c.w#c.d6#c.f06#c.x_dm6 ///
>         i.year i.year#c.x ///
>         d4 d5 d6 x c.d4#c.x c.d5#c.x c.d6#c.x, vce(cluster id)

Linear regression                                   Number of obs     =      3,000

                                             (Std. err. adjusted for 500 clusters in id)
----------------------------------------------------------------------------------------
                       |               Robust
                     y | Coefficient  std. err.       t    P>|t|     [95% conf. interval]
-----------------------+----------------------------------------------------------------
        c.w#c.d4#c.f04 |   3.509903   .3031306    11.58   0.000     2.914334    4.105473
                       |
        c.w#c.d4#c.f05 |   4.240069    .329047    12.89   0.000     3.593581    4.886558
                       |
        c.w#c.d4#c.f06 |   4.209541   .3306214    12.73   0.000     3.559959    4.859123
                       |
        c.w#c.d5#c.f05 |    3.00813   .3409955     8.82   0.000     2.338167    3.678094
                       |
        c.w#c.d5#c.f06 |   3.701545   .3680571    10.06   0.000     2.978413    4.424678
                       |
        c.w#c.d6#c.f06 |   2.016938    .570889     3.53   0.000     .8952951     3.13858
```

c.w#c.d4#c.f04#c.x_dm4	-.3730017	1.045884	-0.36	0.722	-2.42788	1.681877
c.w#c.d4#c.f05#c.x_dm4	.5322266	1.045262	0.51	0.611	-1.521431	2.585884
c.w#c.d4#c.f06#c.x_dm4	2.587535	1.064342	2.43	0.015	.4963923	4.678679
c.w#c.d5#c.f05#c.x_dm5	.9587673	.6327191	1.52	0.130	-.2843545	2.201889
c.w#c.d5#c.f06#c.x_dm5	.1569361	.8624102	0.18	0.856	-1.537467	1.851339
c.w#c.d6#c.f06#c.x_dm6	.7239738	1.171414	0.62	0.537	-1.577539	3.025486

d4	.0351479	.8603456	0.04	0.967	-1.655198	1.725494
d5	-1.792846	.6808996	-2.63	0.009	-3.130629	-.4550624
d6	-2.537934	.9703164	-2.62	0.009	-4.444343	-.6315248
x	.240952	.4526635	0.53	0.595	-.6484094	1.130313
c.d4#c.x	-1.191253	.9341302	-1.28	0.203	-3.026566	.6440597
c.d5#c.x	.5511332	.6069469	0.91	0.364	-.6413531	1.74362
c.d6#c.x	.9944783	.8275699	1.20	0.230	-.6314725	2.620429
year						
2002	-.1800947	.3438779	-0.52	0.601	-.8557216	.4955323
2003	-.7864622	.3325854	-2.36	0.018	-1.439903	-.1330219
2004	.2930123	.3695067	0.79	0.428	-.4329683	1.018993
2005	.0521642	.4238261	0.12	0.902	-.7805394	.8848678
2006	.3885514	.4340822	0.90	0.371	-.4643027	1.241405
year#c.x						
2002	.3958943	.3275757	1.21	0.227	-.2477033	1.039492
2003	.9755543	.3194763	3.05	0.002	.3478697	1.603239
2004	.3305927	.3509236	0.94	0.347	-.3588773	1.020063
2005	.2388468	.398165	0.60	0.549	-.5434397	1.021133
2006	.3602849	.4215185	0.85	0.393	-.4678849	1.188455
_cons	20.1107	.4531656	44.38	0.000	19.22035	21.00105

```
. reg y d4 d5 d6 x c.d4#c.x c.d5#c.x c.d6#c.x ///
>         i.year i.year#c.x if ~w

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          d4 |   .0351479   .4946653     0.07   0.943    -.9348629    1.005159
          d5 |  -1.792846   .4655427    -3.85   0.000    -2.705749   -.8799429
          d6 |  -2.537934   .6663068    -3.81   0.000    -3.844524   -1.231344
           x |    .240952   .3774773     0.64   0.523    -.4992597    .9811637
             |
    c.d4#c.x |  -1.191253   .5044565    -2.36   0.018    -2.180464   -.2020427
             |
    c.d5#c.x |   .5511332   .4165168     1.32   0.186    -.2656328    1.367899
             |
    c.d6#c.x |   .9944784   .5059412     1.97   0.049     .0023562    1.986601
             |
        year |
       2002  |  -.1800947   .4807986    -0.37   0.708    -1.122914    .7627243
       2003  |  -.7864622   .4807986    -1.64   0.102    -1.729281    .1563567
       2004  |   .2930123   .5294468     0.55   0.580     -.745203    1.331228
       2005  |   .0521642    .591125     0.09   0.930    -1.106999    1.211327
       2006  |   .3885514    .643844     0.60   0.546    -.8739904    1.651093
             |
    year#c.x |
       2002  |   .3958943   .4562695     0.87   0.386    -.4988244    1.290613
       2003  |   .9755543   .4562695     2.14   0.033     .0808355    1.870273
       2004  |   .3305927   .4936956     0.67   0.503    -.6375166    1.298702
       2005  |   .2388468   .5614462     0.43   0.671    -.8621175    1.339811
       2006  |   .3602849   .6375711     0.57   0.572    -.8899561    1.610526
             |
       _cons |    20.1107   .3905693    51.49   0.000     19.34482    20.87658
------------------------------------------------------------------------------
```

```
. predict double yinfhat_pols
(option xb assumed; fitted values)

. gen tehat_pols = y - yinfhat_pols

. sum tehat_pols if d4 & f04

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
  tehat_pols |        119    3.509903    3.358716  -6.376343   15.84973

. sum tehat_pols if d4 & f05

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
  tehat_pols |        119    4.240069    3.512751  -4.825089   13.72698

. sum tehat_pols if d4 & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
  tehat_pols |        119    4.209541    3.676014  -4.689328   16.64095
```

```
. sum tehat_pols if d5 & f05

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
  tehat_pols |         83     3.00813    2.733784  -3.229156   10.83753

. sum tehat_pols if d5 & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
  tehat_pols |         83    3.701545    3.441867  -4.658365   10.77978

. sum tehat_pols if d6 & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
  tehat_pols |         36    2.016938     4.03771  -7.140621   9.950262
```

```
. * Also the same as BJS (2021), who use unit fixed effects:
.
. qui reg y i.id i.year i.year##c.x if ~w

. predict double yinfhat_bjs
(option xb assumed; fitted values)

. gen tehat_bjs = y - yinfhat_bjs

. sum tehat_bjs if d4 & f04

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   tehat_bjs |        119    3.509903    3.044101  -5.994367   9.870104

. sum tehat_bjs if d4 & f05

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   tehat_bjs |        119    4.240069    3.276538  -3.418553   13.59451

. sum tehat_bjs if d4 & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   tehat_bjs |        119    4.209541    3.380562  -5.283782   12.70758
```

```
. sum tehat_bjs if d5 & f05

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   tehat_bjs |         83     3.00813    2.915553  -3.621274   10.26595

. sum tehat_bjs if d5 & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   tehat_bjs |         83    3.701545    3.105047  -3.853438   9.730703

. sum tehat_bjs if d6 & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   tehat_bjs |         36    2.016938    3.294478  -6.050471   8.052396
```

```
. * The user-written command "did_imputation" with the "allhorizons"
. * option aggregates by number of treated periods. So all of the
. * instaneous effects are averaged.
. * Similar to imposing common effects by intensity of exposure.

. * First compute the effects "by hand":

. sum tehat_bjs if (d4 & f04) | (d5 & f05) | (d6 & f06)

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   tehat_bjs |        238    3.109089    3.069704  -6.050471   10.26595

. sum tehat_bjs if (d4 & f05) | (d5 & f06)

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   tehat_bjs |        202    4.018795    3.210354  -3.853438   13.59451

. sum tehat_bjs if (d4 & f06)

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   tehat_bjs |        119    4.209541    3.380562  -5.283782   12.70758
```

```
. gen first_treat = .
(3,000 missing values generated)

. replace first_treat = 2004 if d4
(714 real changes made)

. replace first_treat = 2005 if d5
(498 real changes made)

. replace first_treat = 2006 if d6
(216 real changes made)

. did_imputation y id year first_treat, timec(x) allhorizons

                                                      Number of obs = 3,000
------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        tau0 |   3.109089   .2120538    14.66   0.000     2.693471    3.524707
        tau1 |   4.018794   .2470403    16.27   0.000     3.534604    4.502985
        tau2 |   4.209542   .3390207    12.42   0.000     3.545074     4.87401
------------------------------------------------------------------------------
```

```
. * Default in "did_imputation" is to average across all treated observations.

. * By hand:

. sum tehat_bjs if w

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   tehat_bjs |        559    3.672084     3.22053  -6.050471   13.59451

. did_imputation y id year first_treat, timec(x)

                                                     Number of obs = 3,000
------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         tau |   3.672084   .1722172    21.32   0.000     3.334545    4.009624
------------------------------------------------------------------------------
```

7. Rolling and Long-Differencing Methods

- With rolling methods we can apply standard TE methods – as on the common timing case.
- Working paper: Lee and Wooldridge (2022).
- Long differencing due to Callaway and Sant'Anna (2021, Journal of Econometrics).

- Recall the ATTs of interest:

$$\tau_{gt} = E[Y_t(g) - Y_t(\infty)|D_g = 1], g = q, \ldots, T; t = g, \ldots, T$$

- Use algebra to write

$$Y_t(g) - Y_t(\infty) = \left[Y_t(g) - \frac{1}{(g-1)} \sum_{s=1}^{g-1} Y_s(g) \right]$$

$$- \left[Y_t(\infty) - \frac{1}{(g-1)} \sum_{s=1}^{g-1} Y_s(\infty) \right]$$

$$+ \frac{1}{(g-1)} \sum_{s=1}^{g-1} [Y_s(g) - Y_s(\infty)]$$

- Under NA, the last term is zero. Or, use the weaker assumption,

$$E[Y_s(g) - Y_s(\infty)|D_g = 1] = 0, s < g$$

- Under NA we have

$$\tau_{gt} = E[\dot{Y}_t(g)|D_g = 1] - E[\dot{Y}_t(\infty)|D_g = 1]$$

$$\dot{Y}_t(g) \equiv Y_t(g) - \frac{1}{(g-1)} \sum_{s=1}^{g-1} Y_s(g)$$

- The averaging is only over pre-intervention periods for cohort g.

$$\tau_{gt} = E[\dot{Y}_t(g)|D_g = 1] - E[\dot{Y}_t(\infty)|D_g = 1]$$

- As usual, $E[\dot{Y}_t(g)|D_g = 1]$ is easily estimated because $\dot{Y}_t = \dot{Y}_t(g)$ when $D_g = 1$:

$$\hat{E}[\dot{Y}_t(g)|D_g = 1] = N_g^{-1} \sum_{i=1}^{N} D_{ig} \dot{Y}_{igt}$$

$$\dot{Y}_{igt} \equiv Y_{it} - \frac{1}{(g-1)} \sum_{s=1}^{g-1} Y_{is}$$

- For $E[\dot{Y}_t(\infty)|D_g = 1]$, use conditional CT.
  - Can also invoke NA to use all suitable control groups.
- Under CCT,

$$E[\dot{Y}_t(\infty)|D_g = 1, \mathbf{X}] = E[\dot{Y}_t(\infty)|D_\infty = 1, \mathbf{X}]$$

$$E[\dot{Y}_t(\infty)|D_\infty = 1, \mathbf{X}] = E[\dot{Y}_t(\infty)|D_h = 1, \mathbf{X}], h = q, \ldots, T$$

- Cohorts $h = t+1,\ldots,T$ have not yet been treated, and so, by NA,

$$E[\dot{Y}_t(\infty)|D_h = 1,\mathbf{X}] = E[\dot{Y}_t(h)|D_h = 1,\mathbf{X}],\ h = t+1,\ldots,T$$

- Implies we can use the never treated group and treatment cohorts $h \in \{t+1,\ldots,T\}$ in estimating $E[\dot{Y}_t(\infty)|D_\infty = 1,\mathbf{X}]$.
  - Strategy works even without an NT group.

- For estimating $E[\dot{Y}_t(\infty)|D_g = 1, \mathbf{X}]$ we use, for $t \geq g$,

$$D_g + D_{t+1} + D_{t+2} + \cdots + D_T + D_\infty = 1$$

$$D_g = 1 \text{ is treated}$$

$$D_{t+1} + D_{t+2} + \cdots + D_T + D_\infty = 1 \text{ is control}$$

  - Cohorts already treated by period t are not used as controls.

1. For $t \in \{g, g+1, \ldots, T\}$, compute

$$\dot{Y}_{igt} \equiv Y_{it} - \frac{1}{(g-1)} \sum_{s=1}^{g-1} Y_{is}$$

2. Choose as the control group cohorts $h \in \{t+1, \ldots, T\}$.

3. Using the data with

$$D_{ig} + D_{i,t+1} + D_{i,t+2} + \cdots + D_{iT} + D_{i\infty} = 1,$$

apply standard TE methods (RA, Matching, IPWRA) to

$$\{(\dot{Y}_{igt}, D_{ig}, \mathbf{X}_i)\}$$

- Can use a subset of the control cohorts.

- Using RA on

$$\{(\dot{Y}_{igg}, D_{ig}, \mathbf{X}_i)\},$$

with all possible control units, reproduces POLS/ETWFE for the instantaneous effects,

$$\hat{\tau}_{gg}, g = q, \ldots, T.$$

- See (slight) differences with $\hat{\tau}_{gt}$ for $t > g$.

```
. use did_staggered_6

. * Rolling methods. When using RA, the instantaneous effects are the same as
. * pooled OLS/ETWFE:
.
. gen y_44dot = y - (L.y + L2.y + L3.y)/3 if f04
(2,500 missing values generated)

. gen y_45dot = y - (L2.y + L3.y + L4.y)/3 if f05
(2,500 missing values generated)

. gen y_46dot = y - (L3.y + L4.y + L5.y)/3 if f06
(2,500 missing values generated)

. gen y_55dot = y - (L.y + L2.y + L3.y + L4.y)/4 if f05
(2,500 missing values generated)

. gen y_56dot = y - (L2.y + L3.y + L4.y + L5.y)/4 if f06
(2,500 missing values generated)

. gen y_66dot = y - (L.y + L2.y + L3.y + L4.y + L5.y)/5 if f06
(2,500 missing values generated)
```

```
. * tau_44 uses all other groups as control in 2004:
. reg y_44dot d4 x c.d4#c.x_dm4 if f04

------------------------------------------------------------------------------
    y_44dot | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          d4 |   3.509903   .2596378    13.52   0.000     2.999778    4.020029
           x |  -.1265569   .3148113    -0.40   0.688     -.745085    .4919713
             |
c.d4#c.x_dm4 |  -.3730017   .7152189    -0.52   0.602    -1.778234    1.032231
             |
       _cons |   .6151979   .3399179     1.81   0.071    -.0526585    1.283054
------------------------------------------------------------------------------

. * tau_45 uses d6 and dinf as controls in 2005:
. reg y_45dot d4 x c.d4#c.x_dm4 if f05 & ~d5

------------------------------------------------------------------------------
    y_45dot | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          d4 |   4.214895    .280842    15.01   0.000     3.662837    4.766953
           x |  -.1547797   .3958496    -0.39   0.696     -.932911    .6233516
             |
c.d4#c.x_dm4 |   .4687034   .7812616     0.60   0.549    -1.067042    2.004448
             |
       _cons |   .3423441   .4161692     0.82   0.411    -.4757299    1.160418
------------------------------------------------------------------------------
```

```
. * tau_46 uses dinf as controls in 2006:
. reg y_46dot d4 x c.d4#c.x_dm4 if f06 & (d4 | dinf)

------------------------------------------------------------------------------
     y_46dot | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          d4 |   4.171439   .2921879    14.28   0.000     3.596917    4.745961
           x |  -.0490833    .481128    -0.10   0.919    -.9951139    .8969474
             |
c.d4#c.x_dm4 |   2.539754   .8417799     3.02   0.003      .884582    4.194926
             |
       _cons |   .7058289   .4792865     1.47   0.142    -.2365809    1.648239
------------------------------------------------------------------------------

. * tau_55 uses d6 and dinf as controls in 2005:
. reg y_55dot d5 x c.d5#c.x_dm5 if f05 & ~d4

------------------------------------------------------------------------------
     y_55dot | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          d5 |    3.00813   .2933499    10.25   0.000     2.431323    3.584937
           x |  -.1866635   .3609428    -0.52   0.605    -.8963768    .5230497
             |
c.d5#c.x_dm5 |   .9587673   .6624562     1.45   0.149    -.3438046    2.261339
             |
       _cons |   .2205503   .3794706     0.58   0.561    -.5255937    .9666944
------------------------------------------------------------------------------
```

```
. *tau_56 uses dinf as controls in 2006:
. reg y_56dot d5 x c.d5#c.x_dm5 if f06 & (d5 | dinf)

------------------------------------------------------------------------------
    y_56dot | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          d5 |    3.71408   .3110739    11.94   0.000     3.102214    4.325945
           x |  -.1298772   .4408746    -0.29   0.768    -.9970532    .7372989
             |
c.d5#c.x_dm5 |   .2215878   .7225668     0.31   0.759    -1.199661    1.642837
             |
       _cons |   .6142881   .4391872     1.40   0.163    -.2495689    1.478145
------------------------------------------------------------------------------

. * tau_66 uses dinf as controls in 2006:
. reg y_66dot d6 x c.d6#c.x_dm6 if f06 & (d6 | dinf)

------------------------------------------------------------------------------
    y_66dot | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          d6 |   2.016938   .4371878     4.61   0.000     1.156523    2.877352
           x |  -.0278927   .4192001    -0.07   0.947    -.8529061    .7971207
             |
c.d6#c.x_dm6 |   .7239738   .9229923     0.78   0.433    -1.092536    2.540483
             |
       _cons |   .5128275   .4175957     1.23   0.220    -.3090283    1.334683
------------------------------------------------------------------------------

. * Can apply any of the usual TE estimators to the transformed data.
```

```
. teffects psmatch (y_44dot) (d4 x) if f04, atet

------------------------------------------------------------------------------
             |              AI robust
    y_44dot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d4 |
   (1 vs 0)  |   3.321456   .3711246     8.95   0.000     2.594065    4.048847
------------------------------------------------------------------------------

. teffects psmatch (y_45dot) (d4 x) if f05 & ~d5, atet

------------------------------------------------------------------------------
             |              AI robust
    y_45dot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d4 |
   (1 vs 0)  |   3.754891   .4028564     9.32   0.000     2.965308    4.544475
------------------------------------------------------------------------------

. teffects psmatch (y_46dot) (d4 x) if f06 & (d4 | dinf), atet

------------------------------------------------------------------------------
             |              AI robust
    y_46dot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d4 |
   (1 vs 0)  |   3.957854    .444257     8.91   0.000     3.087126    4.828581
------------------------------------------------------------------------------
```

```
. teffects psmatch (y_55dot) (d5 x) if f05 & ~d4, atet

------------------------------------------------------------------------------
             |              AI robust
    y_55dot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d5 |
   (1 vs 0)  |   2.966489   .4534866     6.54   0.000     2.077671    3.855306
------------------------------------------------------------------------------

. teffects psmatch (y_56dot) (d5 x) if f06 & (d5 | dinf), atet

------------------------------------------------------------------------------
             |              AI robust
    y_56dot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d5 |
   (1 vs 0)  |   3.745565   .3971389     9.43   0.000     2.967187    4.523943
------------------------------------------------------------------------------

. teffects psmatch (y_66dot) (d6 x) if f06 & (d6 | dinf), atet

------------------------------------------------------------------------------
             |              AI robust
    y_66dot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d6 |
   (1 vs 0)  |   2.169487   .6078406     3.57   0.000     .9781414    3.360833
------------------------------------------------------------------------------
```

```
. teffects ipwra (y_44dot x) (d4 x) if f04, atet

------------------------------------------------------------------------------
             |               Robust
    y_44dot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d4 |
   (1 vs 0)  |   3.514341   .3017234    11.65   0.000     2.922974    4.105708
-------------+----------------------------------------------------------------

. teffects ipwra (y_45dot x) (d4 x) if f05 & ~d5, atet

------------------------------------------------------------------------------
             |               Robust
    y_45dot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d4 |
   (1 vs 0)  |    4.21929   .3278307    12.87   0.000     3.576754    4.861827
-------------+----------------------------------------------------------------

. teffects ipwra (y_46dot x) (d4 x) if f06 & (d4 | dinf), atet

------------------------------------------------------------------------------
             |               Robust
    y_46dot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d4 |
   (1 vs 0)  |   4.170384   .3401099    12.26   0.000     3.503781    4.836987
-------------+----------------------------------------------------------------
```

```
. teffects ipwra (y_55dot x) (d5 x) if f05 & ~d4, atet

------------------------------------------------------------------------------
             |               Robust
    y_55dot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d5 |
   (1 vs 0)  |   3.013813   .3433829     8.78   0.000     2.340795    3.686831
-------------+----------------------------------------------------------------

. teffects ipwra (y_56dot x) (d5 x) if f06 & (d5 | dinf), atet

------------------------------------------------------------------------------
             |               Robust
    y_56dot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d5 |
   (1 vs 0)  |   3.700215     .36527    10.13   0.000     2.984299    4.416131
-------------+----------------------------------------------------------------

. teffects ipwra (y_66dot x) (d6 x) if f06 & (d6 | dinf), atet

------------------------------------------------------------------------------
             |               Robust
    y_66dot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d6 |
   (1 vs 0)  |   1.879977    .563915     3.33   0.001     .7747235     2.98523
-------------+----------------------------------------------------------------
```

Long Differencing

- The “long” differencing approach is similar to removing pre-treatment averages.
- Callaway and Sant’Anna (2021, Journal of Econometrics).
- As default, CS use the never treated group as the control.
  - Previous rolling methods use all not-yet-treated units.
- In addition, the default CS uses only the most recent control period.
  - Previous rolling methods average over all previous control periods.

- CS propose RA, nonparametric IPW, and doubly robust approaches.
  - Can apply other methods, such as matching, to transformed data.
- Generally, these do not use all restrictions implied by NA and CCT.
- POLS/ETWFE/Imputation use all implications of NA and CCT.
  - CS is generally inefficient when NA and CCT hold.
  - Can be more resilient to certain violations of CCT (later).

- Still interested in

$$\tau_{gt} = E[Y_t(g) - Y_t(\infty)|D_g = 1], t = g, \ldots, T; g = q, \ldots, T$$

- For previous rolling methods:

$$\dot{Y}_{igt} \equiv Y_{it} - \frac{1}{(g-1)} \sum_{s=1}^{g-1} Y_{is}$$

• Callaway and Sant'Anna:

1. Create long differences

$$\mathring{Y}_{igt} \equiv Y_{it} - Y_{i,g-1}, t \geq g$$

▸ From period t, go back $t - g + 1$ lags.

2. Use as a control group not yet subjected to the intervention at time t:

$$D_h = 1, h \in \{t+1, \ldots, T, \infty\}$$

3. Using data $D_{ig} + D_{ih} = 1$, apply TE methods to

$$\left\{\left(\mathring{Y}_{igt}, D_{ig}, \mathbf{X}_i\right)\right\}$$

- CS use IPW and Augmented IPW (AIPW).
  - AIPW usually close to IPWRA.
- Defaults for treatment cohort g:

(i) Only uses never treated group as control ($D_{ig} + D_{i\infty} = 1$).

(ii) Only uses $g - 1$ as the control period.

- User-written Stata command `csdid` by Fernando Rios.

```
. use did_staggered_6

. * Callaway and Sant'Anna (2021) use long differences.

. * By hand:
.
. gen y_44circle = y - L.y if f04
(2,500 missing values generated)

. gen y_45circle = y - L2.y if f05
(2,500 missing values generated)

. gen y_46circle = y - L3.y if f06
(2,500 missing values generated)

. gen y_55circle = y - L.y if f05
(2,500 missing values generated)

. gen y_56circle = y - L2.y if f06
(2,500 missing values generated)

. gen y_66circle = y - L.y if f06
(2,500 missing values generated)

. * First use all potential controls in IPWRA.
. * Then use only the NT group in IPWRA.
. * See did_staggered_6_202112.do for other methods.
```

```
. teffects ipwra (y_44circle x) (d4 x) if f04, atet

------------------------------------------------------------------------------
             |               Robust
  y_44circle | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d4 |
   (1 vs 0)  |   3.451603   .3497624     9.87   0.000     2.766081    4.137124
-------------+----------------------------------------------------------------

. teffects ipwra (y_45circle x) (d4 x) if f05 & ~d5, atet

------------------------------------------------------------------------------
             |               Robust
  y_45circle | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d4 |
   (1 vs 0)  |   4.051787   .3769103    10.75   0.000     3.313056    4.790518
-------------+----------------------------------------------------------------

. teffects ipwra (y_46circle x) (d4 x) if f06 & (d4 | dinf), atet

------------------------------------------------------------------------------
             |               Robust
  y_46circle | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d4 |
   (1 vs 0)  |   4.031729   .3565145    11.31   0.000     3.332973    4.730484
-------------+----------------------------------------------------------------
```

```
. teffects ipwra (y_55circle x) (d5 x) if f05 & ~d4, atet

------------------------------------------------------------------------------
             |               Robust
  y_55circle | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d5 |
   (1 vs 0)  |   3.451348   .4021369     8.58   0.000     2.663174    4.239522
-------------+----------------------------------------------------------------

. teffects ipwra (y_56circle x) (d5 x) if f06 & (d5 | dinf), atet

------------------------------------------------------------------------------
             |               Robust
  y_56circle | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d5 |
   (1 vs 0)  |   4.211341   .4176358    10.08   0.000      3.39279    5.029892
-------------+----------------------------------------------------------------

. teffects ipwra (y_66circle x) (d6 x) if f06 & (d6 | dinf), atet

------------------------------------------------------------------------------
             |               Robust
  y_66circle | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d6 |
   (1 vs 0)  |   1.499542    .709456     2.11   0.035     .1090342     2.89005
-------------+----------------------------------------------------------------
.
```

```
. teffects ipwra (y_44circle x) (d4 x) if f04 & (d4 | dinf), atet

------------------------------------------------------------------------------
             |               Robust
  y_44circle | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d4 |
   (1 vs 0)  |    3.21494   .3602832     8.92   0.000     2.508798    3.921082
-------------+----------------------------------------------------------------

. teffects ipwra (y_45circle x) (d4 x) if f05 & (d4 | dinf), atet

------------------------------------------------------------------------------
             |               Robust
  y_45circle | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d4 |
   (1 vs 0)  |   4.068499   .3781796    10.76   0.000      3.32728    4.809717
-------------+----------------------------------------------------------------

. teffects ipwra (y_46circle x) (d4 x) if f06 & (d4 | dinf), atet

------------------------------------------------------------------------------
             |               Robust
  y_46circle | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d4 |
   (1 vs 0)  |   4.031729   .3565145    11.31   0.000     3.332973    4.730484
-------------+----------------------------------------------------------------
```

```
. teffects ipwra (y_55circle x) (d5 x) if f05 & (d5 | dinf), atet

------------------------------------------------------------------------------
             |               Robust
  y_55circle | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d5 |
   (1 vs 0)  |   3.613127   .4205019     8.59   0.000     2.788959    4.437296
-------------+----------------------------------------------------------------

. teffects ipwra (y_56circle x) (d5 x) if f06 & (d5 | dinf), atet

------------------------------------------------------------------------------
             |               Robust
  y_56circle | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d5 |
   (1 vs 0)  |   4.211341   .4176358    10.08   0.000      3.39279    5.029892
-------------+----------------------------------------------------------------

. teffects ipwra (y_66circle x) (d6 x) if f06 & (d6 | dinf), atet

------------------------------------------------------------------------------
             |               Robust
  y_66circle | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
          d6 |
   (1 vs 0)  |   1.499542    .709456     2.11   0.035     .1090342     2.89005
-------------+----------------------------------------------------------------
```

```
. replace first_treat = 0 if first_treat == .
(1,572 real changes made)

. csdid y x, ivar(id) time(year) gvar(first_treat)
...............
------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
g2004        |
 t_2001_2002 |    .061275   .3187573     0.19   0.848    -.5634778    .6860277
 t_2002_2003 |   .1773472   .3383827     0.52   0.600    -.4858706    .8405651
 t_2003_2004 |   3.214938    .360283     8.92   0.000     2.508796    3.921079
 t_2003_2005 |     4.0685   .3781798    10.76   0.000     3.327282    4.809719
 t_2003_2006 |   4.031726   .3565153    11.31   0.000     3.332969    4.730483
-------------+----------------------------------------------------------------
g2005        |
 t_2001_2002 |   .1313077   .3615115     0.36   0.716    -.5772419    .8398573
 t_2002_2003 |   .6325747   .3961499     1.60   0.110    -.1438647    1.409014
 t_2003_2004 |  -1.147224   .3773126    -3.04   0.002    -1.886743   -.4077046
 t_2004_2005 |   3.614153   .4208516     8.59   0.000     2.789299    4.439007
 t_2004_2006 |   4.211175   .4176205    10.08   0.000     3.392654    5.029696
-------------+----------------------------------------------------------------
g2006        |
 t_2001_2002 |  -.2151668   .5416258    -0.40   0.691    -1.276734    .8464003
 t_2002_2003 |  -.1344075   .5990794    -0.22   0.822    -1.308582    1.039767
 t_2003_2004 |  -.4889371   .5535533    -0.88   0.377    -1.573882    .5960075
 t_2004_2005 |   .9659142   .5876442     1.64   0.100    -.1858472    2.117676
 t_2005_2006 |   1.495818   .7083758     2.11   0.035     .1074269    2.884209
------------------------------------------------------------------------------
Control: Never Treated
```

```
. * Recall POLS/ETWFE estimates:

. reg y c.w#c.d4#c.f04 c.w#c.d4#c.f05 c.w#c.d4#c.f06 ///
>         c.w#c.d5#c.f05 c.w#c.d5#c.f06 ///
>         c.w#c.d6#c.f06 ///
>         c.w#c.d4#c.f04#c.x_dm4 c.w#c.d4#c.f05#c.x_dm4 c.w#c.d4#c.f06#c.x_dm4 ///
>         c.w#c.d5#c.f05#c.x_dm5 c.w#c.d5#c.f06#c.x_dm5 ///
>         c.w#c.d6#c.f06#c.x_dm6 ///
>         d4 d5 d6 x c.d4#c.x c.d5#c.x c.d6#c.x ///
>         i.year i.year#c.x, vce(cluster id)

                                                    (Std. err. adjusted for 500 clusters in id)
-------------------------------------------------------------------------------------------
                        |               Robust
                      y | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
------------------------+------------------------------------------------------------------
        c.w#c.d4#c.f04  |   3.509903   .3031306    11.58   0.000     2.914334    4.105473
                        |
        c.w#c.d4#c.f05  |   4.240069    .329047    12.89   0.000     3.593581    4.886558
                        |
        c.w#c.d4#c.f06  |   4.209541   .3306214    12.73   0.000     3.559959    4.859123
                        |
        c.w#c.d5#c.f05  |    3.00813   .3409955     8.82   0.000     2.338167    3.678094
                        |
        c.w#c.d5#c.f06  |   3.701545   .3680571    10.06   0.000     2.978413    4.424678
                        |
        c.w#c.d6#c.f06  |   2.016938    .570889     3.53   0.000     .8952951     3.13858
                        |
...
-------------------------------------------------------------------------------------------
```

8. A Small Simulation Study

- $T = 6$, staggered entry at $q = 4$.
- One covariate. Common trends conditional on X.
- Population $R^2 \approx 0.241$.
- Approximate cohort shares:

$$\rho_\infty = 0.358,\ \rho_4 = 0.291,\ \rho_5 = 0.225,\ \rho_6 = 0.126$$

- $N = 500$, Replications = 1,000.

	ATT	ETWFE		Rolling IPWRA		Rolling PSM		CS	
$N = 500$	Mean	Mean	SD	Mean	SD	Mean	SD	Mean	SD
τ_{44}	3.99	3.99	0.288	3.99	0.288	3.99	0.363	3.99	0.362
τ_{45}	4.19	4.19	0.289	4.19	0.290	4.18	0.366	4.20	0.367
τ_{46}	4.59	4.60	0.317	4.60	0.321	4.60	0.401	4.60	0.372
τ_{55}	3.03	3.03	0.327	3.03	0.327	3.02	0.412	3.03	0.446
τ_{56}	3.62	3.63	0.358	3.63	0.369	3.64	0.456	3.63	0.430
τ_{66}	2.05	2.04	0.475	2.07	0.527	2.08	0.634	2.04	0.644

9. A Strategy with Exit

- Suppose $T = 6$ with first entry at $q = 4$.
- Three things can happen to units subjected to the intervention at $t = 4$:

(i) Subjected to the intervention through $T = 6$ (the end).

- $D_{4,\infty} = 1$.

(ii) Subjected to the intervention in periods 4 and 5 but not in period 6.

- $D_{4,6} = 1$.

(iii) Subjected to the intervention in period 4 but "exit" in period 5.

- $D_{4,5} = 1$.

- Two things can happen to units subjected to the intervention at $t = 5$:

(i) They are subjected to the intervention through $T = 6$.

  - $D_{5,\infty} = 1$.

(ii) They are subjected to the intervention in period 5 but not in period 6.

  - $D_{5,6} = 1$.

- Only one thing can happen to units subjected to the intervention at $t = 6$:
  - $D_{6,\infty} = 1$.

- With potential exit, there are six groups or cohorts.
- Can interact $D_{g,h}$ with each of fr_t, $r \geq g$.
  - When $r \geq h$, can see whether an effect dissipates after the intervention disappears.

```
. use did_exit_6, clear

. list y w year in 133/174, sep(6)

     +---------------------+
     |        y   w   year |
     |---------------------|
133. | 17.09429   0   2001 |
134. | 17.49467   0   2002 |
135. | 22.30479   0   2003 |
136. | 20.61372   0   2004 |
137. | 21.96369   0   2005 |
138. | 17.70848   0   2006 |
     |---------------------|
139. | 21.85383   0   2001 |
140. | 23.32307   0   2002 |
141. | 16.78383   0   2003 |
142. |  19.1381   0   2004 |
143. | 23.82895   1   2005 |
144. | 28.35684   1   2006 |
     |---------------------|
145. | 17.74132   0   2001 |
146. | 17.26159   0   2002 |
147. | 22.06439   0   2003 |
148. | 18.87821   0   2004 |
149. | 17.31117   0   2005 |
150. | 22.09363   0   2006 |
     |---------------------|
```

151.	20.03997	0	2001
152.	19.22337	0	2002
153.	19.73165	0	2003
154.	23.91452	1	2004
155.	24.19254	0	2005
156.	18.05728	0	2006
157.	16.34389	0	2001
158.	21.97028	0	2002
159.	17.66189	0	2003
160.	16.84226	0	2004
161.	20.50301	1	2005
162.	26.56369	1	2006
163.	18.56555	0	2001
164.	20.59137	0	2002
165.	22.81234	0	2003
166.	20.39433	0	2004
167.	20.93196	1	2005
168.	22.51773	1	2006
169.	19.30813	0	2001
170.	18.39881	0	2002
171.	18.57726	0	2003
172.	23.80016	1	2004
173.	26.67124	1	2005
174.	23.18192	1	2006

```
. reg y w i.year d4_inf d4_5 d4_6 d5_inf d5_6 d6_inf, vce(cluster id)

Linear regression                               Number of obs     =      6,000

                                (Std. err. adjusted for 1,000 clusters in id)
------------------------------------------------------------------------------
             |               Robust
           y | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           w |   3.537775   .1143473    30.94   0.000     3.313387    3.762164
             |
        year |
       2002  |   .1782131   .0847068     2.10   0.036     .0119895    .3444368
       2003  |    .318956   .0874399     3.65   0.000     .1473691    .4905428
       2004  |   .4198873   .1023753     4.10   0.000     .2189919    .6207826
       2005  |   .9462402   .1158158     8.17   0.000     .7189701     1.17351
       2006  |   1.517124   .1243014    12.21   0.000     1.273202    1.761045
             |
      d4_inf |  -.3136822   .1883186    -1.67   0.096    -.6832277    .0558633
        d4_5 |  -.0897199    .249048    -0.36   0.719    -.5784371    .3989972
        d4_6 |   .1795457     .58421     0.31   0.759    -.9668738    1.325965
      d5_inf |  -1.146581   .1944487    -5.90   0.000    -1.528156   -.7650063
        d5_6 |   .0503113   .4667667     0.11   0.914    -.8656443     .966267
      d6_inf |  -1.158893   .2532653    -4.58   0.000    -1.655886   -.6619001
       _cons |    20.1469   .1311279   153.64   0.000     19.88958    20.40421
------------------------------------------------------------------------------
```

```
. * Heterogeneous effects:
.
. reg y c.d4_inf#c.f04 c.d4_inf#c.f05 c.d4_inf#c.f06 ///
>         c.d4_6#c.f04 c.d4_6#c.f05 c.d4_6#c.f06 ///
>         c.d4_5#c.f04 c.d4_5#c.f05 c.d4_5#c.f06 ///
>         c.d5_inf#c.f05 c.d5_inf#c.f06 c.d5_6#c.f05 c.d5_6#c.f06 c.d6_inf#c.f06 ///
>         i.year d4_inf d4_5 d4_6 d5_inf d5_6 d6_inf, vce(cluster id)

Linear regression                               Number of obs     =      6,000

                                 (Std. err. adjusted for 1,000 clusters in id)
------------------------------------------------------------------------------
             |               Robust
           y | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
c.d4_inf#c.f04 |   4.010586   .2029588    19.76   0.000     3.612311     4.40886
             |
c.d4_inf#c.f05 |   5.222066   .2128694    24.53   0.000     4.804344    5.639788
             |
c.d4_inf#c.f06 |   5.860439   .2402728    24.39   0.000     5.388942    6.331937
             |
  c.d4_6#c.f04 |   4.437284   .6673331     6.65   0.000     3.127749     5.74682
             |
  c.d4_6#c.f05 |   3.685213   .9205511     4.00   0.000     1.878778    5.491649
             |
  c.d4_6#c.f06 |   5.001325   .7163158     6.98   0.000     3.595669    6.406981
             |
  c.d4_5#c.f04 |   4.418828   .3233121    13.67   0.000     3.784379    5.053276
             |
  c.d4_5#c.f05 |   2.972421   .3052596     9.74   0.000     2.373397    3.571444
             |
  c.d4_5#c.f06 |   1.501062   .3141923     4.78   0.000     .8845095    2.117615
             |
```

c.d5_inf#c.f05	3.056998	.2369621	12.90	0.000	2.591997	3.521998
c.d5_inf#c.f06	4.048205	.2315979	17.48	0.000	3.59373	4.502679
c.d5_6#c.f05	3.10332	.6463799	4.80	0.000	1.834902	4.371738
c.d5_6#c.f06	2.661681	.6763926	3.94	0.000	1.334368	3.988994
c.d6_inf#c.f06	2.599297	.3273902	7.94	0.000	1.956845	3.241748
year						
2002	.1782131	.0847989	2.10	0.036	.0118087	.3446176
2003	.318956	.087535	3.64	0.000	.1471825	.4907294
2004	.1974369	.1070883	1.84	0.066	-.012707	.4075808
2005	.3264332	.1297033	2.52	0.012	.071911	.5809553
2006	.5965085	.1555651	3.83	0.000	.2912367	.9017803
d4_inf	-1.06031	.1905078	-5.57	0.000	-1.434151	-.6864682
d4_5	-.9821425	.2520902	-3.90	0.000	-1.47683	-.4874554
d4_6	-.8284996	.5029846	-1.65	0.100	-1.815527	.1585279
d5_inf	-1.151523	.1929001	-5.97	0.000	-1.530059	-.7729873
d5_6	-.3208929	.4314939	-0.74	0.457	-1.167631	.5258454
d6_inf	-1.00248	.2508107	-4.00	0.000	-1.494656	-.5103038
_cons	20.44071	.1309661	156.08	0.000	20.18371	20.69771

10. Testing and Relaxing Parallel Trends

- Need at least two pre-treatment periods.
- Suppose $T = 3$, intervention at $t = 3$.
- Without covariates, run the regression

$$\Delta Y_{i2} \text{ on } 1, D_i, i = 1, \ldots, N$$

  - ▸ Heteroskedasticity-robust t statistic on D_i.

- Two pooled OLS approaches yield the same statistic.

1. Run the regression

$$Y_{it} \text{ on } 1, D_i, f2_t, D_i \cdot f2_t, f3_t, D_i \cdot f3_t, t = 1, 2, 3; i = 1, \ldots, N$$

and use the cluster-robust t statistic on $D_i \cdot f2_t$.

2. Run the regression

$$y_{it} \text{ on } 1, D_i, f2_t, D_i \cdot t, f3_t, D_i \cdot f3_t, t = 1, 2, 3; i = 1, \ldots, N$$

and use the cluster-robust t statistic on $D_i \cdot t$.

- Statistics are identical.
  - Coefficients on $D_i \cdot f3_t$ can be very different.

- When including $D_i \cdot t$, the coefficient on $D_i \cdot f3_t$ is a DiDiD estimator:

$$\begin{aligned}\hat{\tau}_3 &= N_1^{-1} \sum_{i=1}^{N} D_i \cdot \Delta^2 Y_{i3} - N_0^{-1} \sum_{i=1}^{N} (1 - D_i) \cdot \Delta^2 Y_{i3} \\ &= [(\bar{Y}_{3,treat} - \bar{Y}_{2,treat}) - (\bar{Y}_{2,treat} - \bar{Y}_{1,treat})] \\ &\quad - [(\bar{Y}_{3,control} - \bar{Y}_{2,control}) - (\bar{Y}_{2,control} - \bar{Y}_{1,control})]\end{aligned}$$

  - Explicitly allows non-parallel linear trends, $D_i \cdot t$.

- Equivalent expression:

$$\hat{\tau}_3 = (\overline{\Delta Y}_{3,treat} - \overline{\Delta Y}_{3,control}) - (\overline{\Delta Y}_{2,treat} - \overline{\Delta Y}_{2,control})$$

- $\overline{\Delta Y}_{3,treat} - \overline{\Delta Y}_{3,control}$ is the DID if we only use period $t = 2$ (before) and period $t = 3$ (after).
- $\overline{\Delta Y}_{2,treat} - \overline{\Delta Y}_{2,control}$ is the DID *prior* to the intervention.
  - ▸ This term is what is used for the common trends test; a placebo test.

• Testing strategies in the general case:

1. In the full POLS regression, add interactions $D_{ig} \cdot fs_t$ for $s < g$, do joint test: the event study regression.

  ▸ Not a sensible correction.

2. In the full POLS regression, add heterogenous linear trends

$$D_{iq} \cdot t, \ldots, D_{iT} \cdot t$$

and use a joint test.

  ▸ Works as a correction if the differences in trends are linear in t.

- The imputation equivalence result holds when adding heterogeneous trends.
  - The test is identical to using only the $W_{it} = 0$ observations and doing a joint test on

$$D_{iq} \bullet t, ..., D_{iT} \bullet t$$

  - Implication: Provided heterogeneous TEs are allowed, the test for pre-trends is not contaminated by using the long regression.

Simulations

- $N = 500$, $T = 6$, staggered entry at $q = 4$.
- One covariate. Common trends holds conditional on X.
- $R^2 = 0.241$.
- Approximate cohort shares:

$$\rho_\infty = 0.358,\ \rho_4 = 0.290,\ \rho_5 = 0.225,\ \rho_6 = 0.127$$

- 1,000 replications.

	ATT	POLS		CS		Het. Trends	
$N = 500$	Mean	Mean	SD	Mean	SD	Mean	SD
τ_{44}	3.99	3.99	0.288	3.99	0.362	3.99	0.396
τ_{45}	4.19	4.19	0.289	4.20	0.367	4.20	0.513
τ_{46}	4.59	4.60	0.316	4.60	0.372	4.61	0.662
τ_{55}	3.03	3.03	0.326	3.03	0.446	3.02	0.423
τ_{56}	3.62	3.63	0.358	3.63	0.430	3.62	0.521
τ_{66}	2.05	2.04	0.474	2.04	0.644	2.05	0.546

- Rejection rate of event study test (9 df, 5% level): 0.061
- Rejection rate of common trends test (3 df, 5% level): 0.045

- $N = 500$, $T = 6$, staggered entry at $q = 4$.
- One covariate.
- Heterogeneous linear trends based on D_{ig}.
- $R^2 = 0.314$.
- Cohort shares: $\rho_\infty = 0.358$, $\rho_4 = 0.291$, $\rho_5 = 0.225$,
$\rho_6 = 0.127$.
- 1,000 replications.

	ATT	POLS		CS		Het. Trends		Event Study	
$N = 500$	Mean	Mean	SD	Mean	SD	Mean	SD	Mean	SD
τ_{44}	3.99	4.34	0.288	4.24	0.362	3.99	0.396	4.74	0.367
τ_{45}	4.19	4.82	0.289	4.70	0.367	4.20	0.513	5.19	0.366
τ_{46}	4.59	5.55	0.317	5.35	0.372	4.61	0.662	5.85	0.369
τ_{55}	3.03	3.34	0.326	3.20	0.446	3.02	0.423	3.70	0.409
τ_{56}	3.62	4.19	0.358	3.96	0.430	3.62	0.521	4.46	0.415
τ_{66}	2.05	2.43	0.474	2.16	0.644	2.05	0.546	2.69	0.551

- Rejection rate of event study test (9 df, 5% level): 0.266
- Rejection rate of common trends test (3 df, 5% level): 0.401

11. Other Directions

- Nonlinear DiD: Wooldridge (2022, WP) allows for binary, fractional, count, corner solutions.
  - Use poole quasi-MLE; same as imputation.
- Quantile DiD: Callaway and Li (2019, QE); Miller (2022, WP).