Relationships Among Recent Difference-in-Differences Estimators and How to Compute Them in Stata

30th UK Stata Conference

12-13 September 2024
London School of Economics

Jeff Wooldridge
Department of Economics
Michigan State University

1. Introduction
2. Staggered Interventions and Assumptions
3. Estimators without Controls
4. Adding Time-Constant Controls
5. Nonlinear Models
6. Extensions and Stata Wish List

1. Introduction

- Exploiting different timing of interventions can be powerful for determining causality.
- What is often (loosely) called "two-way fixed effects" (TWFE) imposes a constant effect across treatment cohort and calendar time.
- Constant effect model can be too restrictive.
  - Resulting estimates can consistently estimate uninteresting weighted averages of "treatment" effects.
  - Borusyak, Jaravel, Spiess (2024 REStud); de Chaisemartin and D'Haultfoeuille (2020, AER); Goodman-Bacon (2021, J of E).

- The “event study” (ES) or “leads and lags” estimator estimates effects for different exposure times.
  - This still can impose unwanted restrictions.
  - Sun and Abraham (2021, J of E).

- At least two reactions to the limitations of constant effect (or constant by exposure time) 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 controls.
    - Callaway and Sant'Anna (2021); BJS (2024); Sun and Abraham (2021).

- For much analysis, do not need special commands for DiD.
  - Can use existing commands in Stata, especially `regress`, `xtreg`, `teffects`.
  - `glm`, `logit`, `fracreg`, `poisson` are useful for nonlinear models.
- With many time periods, treatment cohorts, and controls, the commands become long and messy.
- Output is very busy, but you can see everything.
  - Average treatment effects; moderating effects; selection into treatment cohort; trends as a function of controls.

- Community-contributed commands: `csdid`, `jwdid` (Fernando Rios-Avila).
- Stata 17 command `xtdidregress`.
  - Assumes homogeneous effects (TWFE); want to relax this.
- Stata 18: `xthdidregress`.
  - Staggered interventions and heterogeneous TEs.

2. Staggered Interventions: Notation and Assumptions

- T time periods with no units treated in $t = 1$.
- First unit is treated at $t = q < T$.
- Initially, 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; assume so here.

- 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, Journal of Econometrics): $Y_t(\infty)$ is the outcome if a unit is never treated in $\{q, \ldots, T\}$.
  - $Y_t(0)$ is common but more confusing in this context.

- 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.

- Goal is 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.

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

$$E[Y_t(g) - Y_t(\infty)|D_q, \ldots, D_T] = 0, t \in \{1, 2, \ldots, g-1\}, g \in \{q, \ldots, T\}. \square$$

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

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

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

- Allows the D_g to be correlated with $Y_1(\infty)$; selection into treatment.

- We observe $\{D_{i\infty}, D_{iq}, D_{i,q+1}, \ldots, D_{iT}\}$ and the outcome

$$Y_{it} = D_{i\infty} \cdot Y_{it}(\infty) + D_{iq} \cdot Y_{it}(q) + \cdots + D_{iT} \cdot Y_{it}(T)$$

  - If $D_{ig} = 0$ for all i, simply drop that dummy: no units in cohort g.
- Often start with W_{it}, the time-varying treatment indicator.
  - $W_{i,t-1} = 1 \Rightarrow W_{it} = 1$.
- Define post-treatment time dummies by cohort:

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

- Then

$$W_{it} = D_{iq} \cdot pq_t + D_{i,q+1} \cdot p(q+1)_t + \cdots + D_{iT} \cdot pT_t$$

3. Estimators without Controls

- Simplest pooled OLS regression estimates a single coefficient:

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

  - $D_{iq}, ..., D_{iT}, f2_t, ..., fT_t$ act as controls.

- Algebraically the same as replacing $1, D_{iq}, ..., D_{iT}$ with unit fixed effects:

$$Y_{it} \text{ on } W_{it}, C_{i1}, ..., C_{iN}, f2_t, ..., fT_t$$

$$\text{POLS} = \text{TWFE}$$

  - Equivalence very useful to simply computation; extends to nonlinear models.

- What does $\hat{\beta}_W$ estimate?
  - Weighted average of many 2×2 DiDs.
  - Some “bad comparisons” or “forbidden contrasts.”
- `xtdidregress` imposes a constant effect.
- Reproduces `xtreg`.

• The event study estimator – leads and lags – estimates an effect for different exposure times.

▸ Chooses a base period for comparison – usually $g-1$ for each treated cohort g.

$$Y_{it} \text{ on } EXP_{it,1-T}, EXP_{it,2-T}, \ldots, EXP_{it,-2},$$
$$EXP_{it,0}, EXP_{it,1}, \ldots, EXP_{it,T-q},$$
$$1, D_{iq}, \ldots, D_{iT}, f2_t, \ldots, fT_t$$

▸ Still the same as two-way FE.

▸ Pre-treatment indicators used to detect pre-trends (failure of parallel trends).

▸ `eventdd` in Stata.

- How can we make the constant effect regression more flexible?
- Under PT, the pre-treatment indicators are redundant.

$$
\begin{aligned}
&Y_{it} \text{ on } D_{iq} \cdot fq_t, \ldots, D_{iq} \cdot fT_t, \ldots, \\
&\quad D_{i,q+1} \cdot f(q+1)_t, \ldots, D_{i,q+1} \cdot fT_t, \ldots, D_{iT} \cdot fT_t, \\
&\quad 1, D_{iq}, \ldots, D_{iT}, f2_t, \ldots, fT_t
\end{aligned}
$$

- Gives estimates by cohort-time pairs:

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

- Can aggregate these, typically weighted by cohort share:
  - Exposure time:

$$\hat{\tau}_0, \hat{\tau}_1, ..., \hat{\tau}_{T-q}$$

  - Or, a single, weighted effect.
- Avoids the "bad comparisons."
- Still the same as TWFE with treatment indicators

$$D_{iq} \cdot fq_t, ..., D_{iq} \cdot fT_t, ...,$$
$$D_{i,q+1} \cdot f(q+1)_t, ..., D_{i,q+1} \cdot fT_t, ..., D_{iT} \cdot fT_t$$

- "Extended" TWFE [Wooldridge (2021)].

- Same as a two-step imputation estimator based on cohort and time dummies: Use $W_{it} = 0$ observations to impute $Y_{it}(\infty)$.

$$\widehat{TE}_{it} = Y_{it} - \hat{Y}_{it}(\infty)$$

  - Average by (g, t) pairs.
  - Recovers the POLS = ETWFE estimates [Wooldridge (2021, 2023)].
  - Also the same as BJS imputation using unit FEs.
- POLS/TWFE makes aggregation and inference easy.

• Estimated in Stata 18:

```
xthdidregress twfe (y) (w), group(id)
estat aggregation, dynamic graph
estat aggregation
```

• For computing proper standard errors, extending to nonlinear models, useful to introduce W_{it} explicitly.

▸ Useful for emphasizing the difference with the constant coefficient estimation.

• Useful trick for obtaining standard errors that account for sampling error in weights:

$$Y_{it} \text{ on } W_{it} \cdot D_{iq} \cdot fq_t, \ldots, W_{it} \cdot D_{iq} \cdot fT_t, \ldots,$$
$$W_{it} \cdot D_{i,q+1} \cdot f(q+1)_t, \ldots, W_{it} \cdot D_{i,q+1} \cdot fT_t, \ldots, D_{iT} \cdot fT_t,$$
$$1, D_{iq}, \ldots, D_{iT}, f2_t, \ldots, fT_t$$

• By exposure time:

```
margins, dydx(w) subpop(if expj == 1)
   vce (uncond)
```

• Single effect:

```
margins, dydx(w) subpop(if w == 1)
   vce (uncond)
```

- Can add the pre-treatment indicators for a fully saturated regression:

$$
\begin{aligned}
&D_{iq} \cdot f1_t, \ldots, D_{iq} \cdot f(q-2)_t, \\
&D_{i,q+1} \cdot f1_t, \ldots, D_{i,q+1} \cdot f(q-1)_t, \\
&\ldots \\
&D_{iT} \cdot f1_t, \ldots, D_{iq} \cdot f(T-2)_t
\end{aligned}
$$

  - "Leads and lags" estimator.
  - Equivalent to TWFE: Sun and Abraham (2021).
  - User-written command is `eventstudyinteract`.

- Also the same as Callaway and Sant'Anna (2021) regression adjustment:

```
xthdidregress ra (y) (w), group(id)
estate aggregate, dynamic graph
estate aggregate
```

- Equivalent to 2×2 DiDs using the NT group as the controls.

$$\text{Treated cohort: } g$$

$$\text{Pre-treatment period: } g - 1$$

  - For the pre-treatment effects, `xthdidregress ra` does not use $g - 1$ as the reference period.

- Technically, the ES (leads and lags) only requires that PT holds starting in period $g - 1$.
- Extended TWFE (lags only) effectively averages the pre-treatment periods.
- *Might* be a tradeoff between efficiency and robustness.
  - ▸ Under the PT and the "ideal" second moment assumptions – no serial correlation or heteroskedasticity – ETWFE is more efficient.
    - ▹ ES adds redundant, collinear regressors.
  - ▸ If PT holds starting in period $g - 1$ but fails before, ES is consistent and ETWFE is inconsistent.

- However:
  - ▸ Under strong, positive serial correlation, ES can be more efficient (FD versus FE).
  - ▸ If PT is violated once the treatment begins, ES can have more bias than ETWFE.
- Ideally, the leads and lags and lags only estimators are similar.

- Using any other set of pre-treatment dummies, such as

$$D_{iq} \cdot f2_t,\ \ldots,\ D_{iq} \cdot f(q-1)_t,\ \ldots,\ D_{iT} \cdot f2_t,\ \ldots,\ D_{iT} \cdot f(T-1)_t,$$

results in the same test.

▸ Estimates on the treatment dummies will differ; they will be relative to the first time period (coefficients normalized to be zero).

4. Adding Time-Constant Controls

- Assume $\mathbf{X}_i$ not affected by the intervention (or a analyze mediating effects).
- Adding

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

does not change estimated effects.

- Also should add

$$f2_t \cdot \mathbf{X}_i, \ldots, fT_t \cdot \mathbf{X}_i \text{ (observed heterogenous trends)}$$

$$D_{ig} \cdot fs_t \cdot \dot{\mathbf{X}}_{ig} \text{ (moderating effects)}$$

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

• POLS:

$$
\begin{aligned}
&Y_{it} \text{ on } D_{iq} \cdot fq_t, \ldots, D_{iq} \cdot fT_t, \ldots, \\
&\quad D_{i,q+1} \cdot f(q+1)_t, \ldots, D_{i,q+1} \cdot fT_t, \ldots, D_{iT} \cdot fT_t, \\
&\quad D_{iq} \cdot fq_t \cdot \dot{\mathbf{X}}_{iq}, \ldots, D_{iq} \cdot fT_t \cdot \dot{\mathbf{X}}_{iT}, \\
&\quad D_{i,q+1} \cdot f(q+1)_t \cdot \dot{\mathbf{X}}_{iq}, \ldots, D_{i,q+1} \cdot fT_t \cdot \dot{\mathbf{X}}_{iT}, \ldots, D_{iT} \cdot fT_t \cdot \dot{\mathbf{X}}_{iT} \\
&\quad 1, D_{iq}, \ldots, D_{iT}, \mathbf{X}_i, D_{ig} \cdot \mathbf{X}_i, \ldots, D_{iT} \cdot \mathbf{X}_i, f2_t, \ldots, fT_t, f2_t \cdot \mathbf{X}_i, \ldots, fT_t \cdot \mathbf{X}_i
\end{aligned}
$$

▸ Same as TWFE and random effects: Wooldridge (2021).

▸ Same as cohort imputation [Wooldridge (2021)] and unit-specific imputation [BJS (2024)].

```
xthdidregress twfe (y x1 ... xK) (w), group(id)
estat aggregation, dynamic graph
estat aggregation
```

- How should one include the pre-treatment indicators?
  - As a test, probably just

$$D_{iq}\cdot f1_t,\ \ldots,\ D_{iq}\cdot f(q-2)_t,\ \ldots,\ D_{iT}\cdot f1_t,\ \ldots,\ D_{iT}\cdot f(T-2)_t$$

- For symmetry, include the interactions with $\dot{\mathbf{X}}_{ig}$.
- Fully saturated regression with $g-1$ as the base period for treatment cohort g.
  - A "very long regression."

$$Y_{it} \text{ on } \quad D_{iq} \cdot f1_t, \ ..., D_{iq} \cdot f(q-2)_t, D_{iq} \cdot fq_t, ..., D_{iq} \cdot fT_t, ...,$$

$$D_{i,q+1} \cdot f1_t, ..., D_{i,q+1} \cdot f(q-1)_t, D_{i,q+1} \cdot f(q+1)_t, ..., D_{i,q+1} \cdot fT_t,$$

$$..., D_{iT} \cdot f1_t, ..., D_{iq} \cdot f(T-2)_t, D_{iT} \cdot fT_t,$$

$$D_{iq} \cdot f1_t \cdot \dot{\mathbf{X}}_{iq}, \ ..., D_{iq} \cdot f(q-2)_t \cdot \dot{\mathbf{X}}_{iq}, D_{iq} \cdot fq_t \cdot \dot{\mathbf{X}}_{iq}, ..., D_{iq} \cdot fT_t \cdot \dot{\mathbf{X}}_{iT},$$

...

$$D_{iT} \cdot f1_t \cdot \dot{\mathbf{X}}_{iT}, ..., D_{iq} \cdot f(T-2)_t \dot{\mathbf{X}}_{iT}, D_{iT} \cdot fT_t, ..., D_{iT} \cdot fT_t \cdot \dot{\mathbf{X}}_{iT}$$

$$1, D_{iq}, ..., D_{iT}, \mathbf{X}_i, D_{ig} \cdot \mathbf{X}_i, ..., D_{iT} \cdot \mathbf{X}_i, f2_t, ..., fT_t, f2_t \cdot \mathbf{X}_i, ..., fT_t \cdot \mathbf{X}_i$$

- Still the same as RE and TWFE on flexible equation.

- Under violation of conditional PT, not necessarily to add the extra terms.
- Under CPT and ideal second moment assumptions, adding the extra terms is inefficient.
  - But with serial correlation, can be better to add the controls!
- The very long regression can be viewed as Sun and Abraham (2021) with fully flexible controls or the ES version of Wooldridge (2021).
- Gives "pre-treatment" effects and estimated ATTs:

$$\hat{\theta}_{gs}, g \in \{q, \ldots, T\}, s \in \{1, \ldots, g-2\}$$
$$\hat{\tau}_{gs}, g \in \{q, \ldots, T\}, s \in \{g, \ldots, T\}$$

  - $\hat{\theta}_{g,g-1} \equiv 0$ is the normalization.

- Typically the pre-trends test focuses on $D_{ig} \cdot fs_t$, $s = 1, \ldots, g-2$ and not these interacted with $\mathbf{X}_{ig}$.
  - Under CPT, these terms have zero population coefficients.
- For each cohort, could create an ES plot.
  - Can be noisy unless we have many units in each treated cohort.
- Can weight the $\hat{\theta}_{gs}$, $\hat{\tau}_{gs}$ by the cohort shares to create a single ES plot.

- Equivalently, define

$$NW_{it} = 1 - W_{it}$$

- Interact W_{it} with $D_{ig} \cdot fs_t$, $s \in \{g, \ldots, T\}$ (treatment).
- Interact NW_{it} with $D_{ig} \cdot fs_t$, $s \in \{1, \ldots, g-2\}$ (pre-treatment).
- Use the `subpop(expj == 1) vce(uncond)` options to account for sampling error in weights and sample averages $\bar{\mathbf{X}}_g$.

- There are other useful characterizations of this extended extended TWFE.
- For both $\hat{\tau}_{gs}$ and $\hat{\theta}_{gs}$, same estimates as using time periods $(g-1,s)$ and the flexible 2×2 DiD:

$$
\begin{aligned}
Y_{it} &= \alpha + \beta D_{ig} + \mathbf{X}_i \boldsymbol{\gamma} + (D_{ig} \cdot \mathbf{X}_i)\boldsymbol{\delta} + \gamma_s fs_t + (fs_t \cdot \mathbf{X}_i)\boldsymbol{\pi}_s \\
&\quad + \tau_{gs}(D_{ig} \cdot fs_t) + (D_{ig} \cdot fs_t \cdot \dot{\mathbf{X}}_i)\boldsymbol{\rho}_{gs} + U_{it},\ t \in (g-1,s)
\end{aligned}
$$

▸ Use only the subset $D_{ig} = 1$ or $D_{i\infty} = 1$.

- Also the same as the regression adjustment version of Callaway and Sant'Anna (2021).
  - Run separate RA on "long" differences:

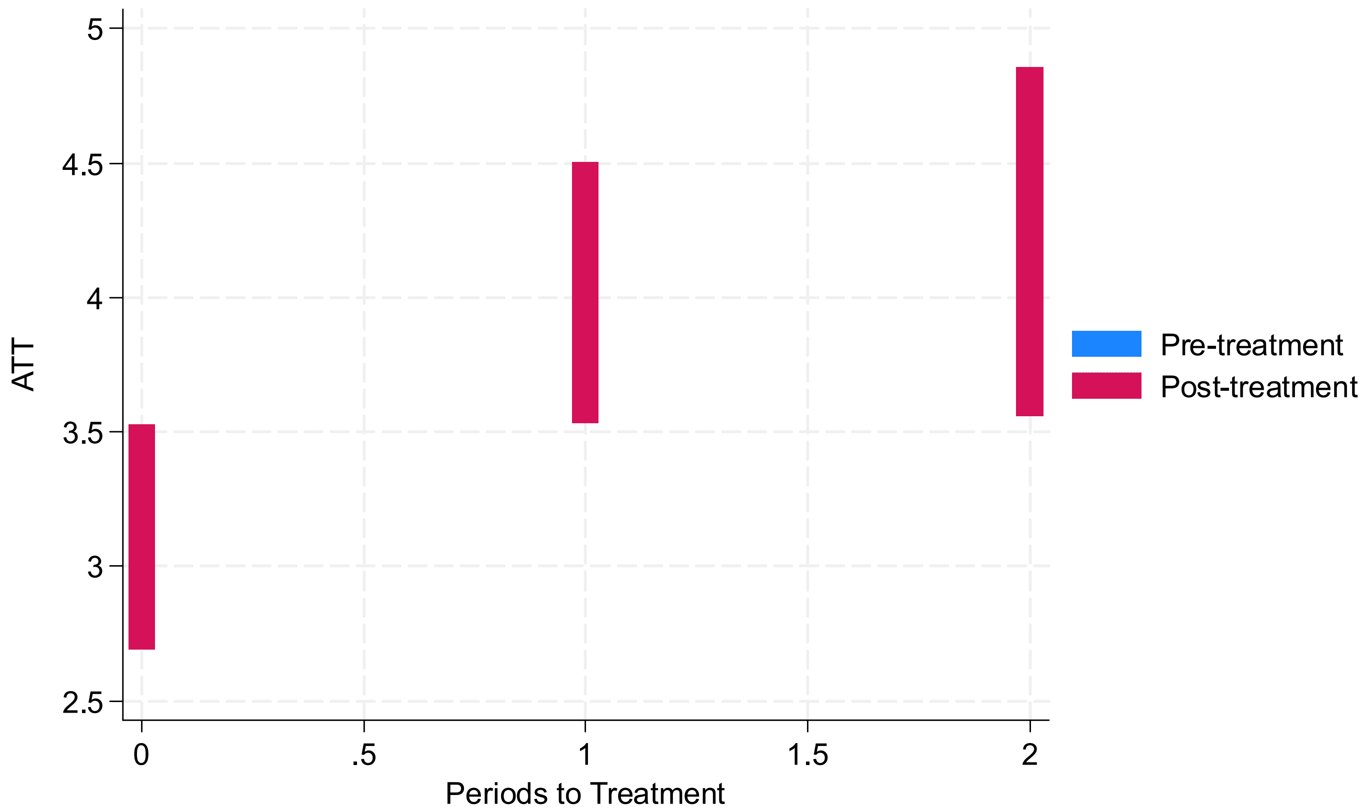
$$Y_{is} - Y_{i,g-1} \text{ on } 1, D_{ig}, \dot{\mathbf{X}}_{ig}, D_{ig} \cdot \dot{\mathbf{X}}_{ig} \text{ using } D_{ig} = 1 \text{ or } D_{i\infty} = 1$$

  - CS (2021) only tests the pre-treatment dummies; not the interactions with $\mathbf{X}_i$.
  - The coefficient on D_{ig} is $\hat{\tau}_{gs}$ ($s \geq g$) or $\hat{\theta}_{gs}$ ($s \leq g - 2$) for each cohort $g \in \{q, \ldots, T\}$.

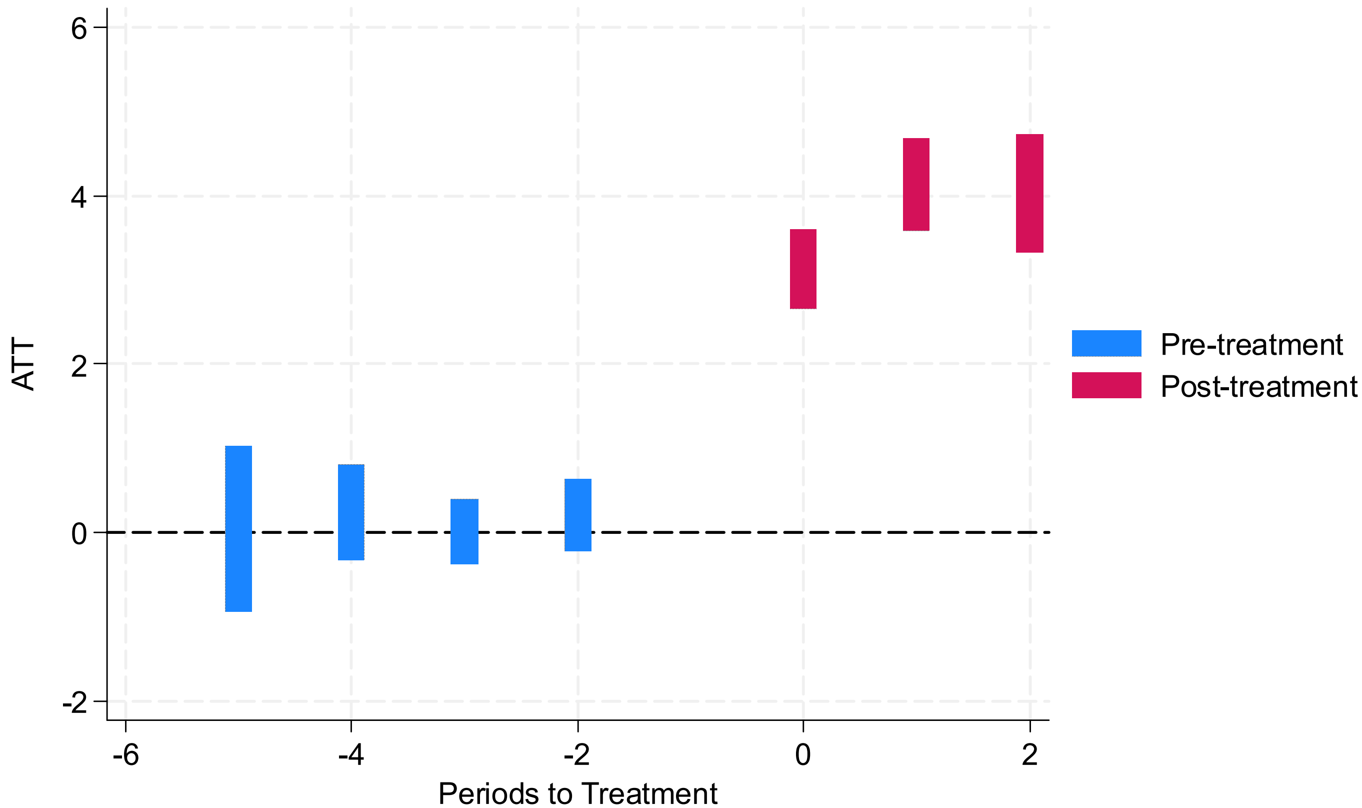
- Computed by `xthdidregress ra`, but does the original CS (2021).
  - Does not use $g - 1$ as base period.
- Computed by `csdid, method(reg) long2`.
  - `xthdidregress` does not have the "`long2`" option.
- See did_staggered_6_es.do.

```
. qui xthdidregress twfe (y x) (w), group(id)
. estat aggregation, dynamic graph

Duration of exposure ATET                              Number of obs = 3,000

                                   (Std. err. adjusted for 500 clusters in id)
------------------------------------------------------------------------------
             |               Robust
    Exposure |       ATET   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           0 |   3.109089   .2158719    14.40   0.000     2.684959    3.533219
           1 |   4.018795   .2491253    16.13   0.000     3.529331    4.508258
           2 |   4.209541    .341013    12.34   0.000     3.539543    4.879539
------------------------------------------------------------------------------
Note: Exposure is the number of periods since the first treatment time.

. estat aggregation

Overall ATET                                           Number of obs = 3,000

                                   (Std. err. adjusted for 500 clusters in id)
------------------------------------------------------------------------------
             |               Robust
           y |       ATET   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           w |
   (1 vs 0)  |   3.672084   .1752452    20.95   0.000     3.327775    4.016394
------------------------------------------------------------------------------
```

```
. qui jwdid y x, ivar(id) tvar(year) gvar(first_treat)
. estat event
. estat plot
```



```
. qui csdid y x, ivar(id) time(year) gvar(first_treat) method(reg) long2

. estat event

ATT by Periods Before and After treatment
Event Study:Dynamic effects
------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     Pre_avg |   .1255849   .2166593     0.58   0.562    -.2990596    .5502294
    Post_avg |   3.764128   .2144565    17.55   0.000     3.343801    4.184455
         Tm5 |   .0428767   .5030756     0.09   0.932    -.9431333    1.028887
         Tm4 |   .2440144    .288585     0.85   0.398    -.3216019    .8096306
         Tm3 |   .0139107   .1969577     0.07   0.944    -.3721194    .3999408
         Tm2 |   .2015377   .2202663     0.91   0.360    -.2301763    .6332518
         Tp0 |   3.129432   .2422789    12.92   0.000     2.654574     3.60429
         Tp1 |   4.129554   .2796401    14.77   0.000      3.58147    4.677639
         Tp2 |   4.033398   .3559803    11.33   0.000      3.33569    4.731107
------------------------------------------------------------------------------

. estat simple
Average Treatment Effect on Treated
------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         ATT |   3.683272   .2053372    17.94   0.000     3.280819    4.085726
------------------------------------------------------------------------------
```



6
4
2
0
-2
ATT
-6
-4
-2
0
2
Periods to Treatment
Pre-treatment
Post-treatment

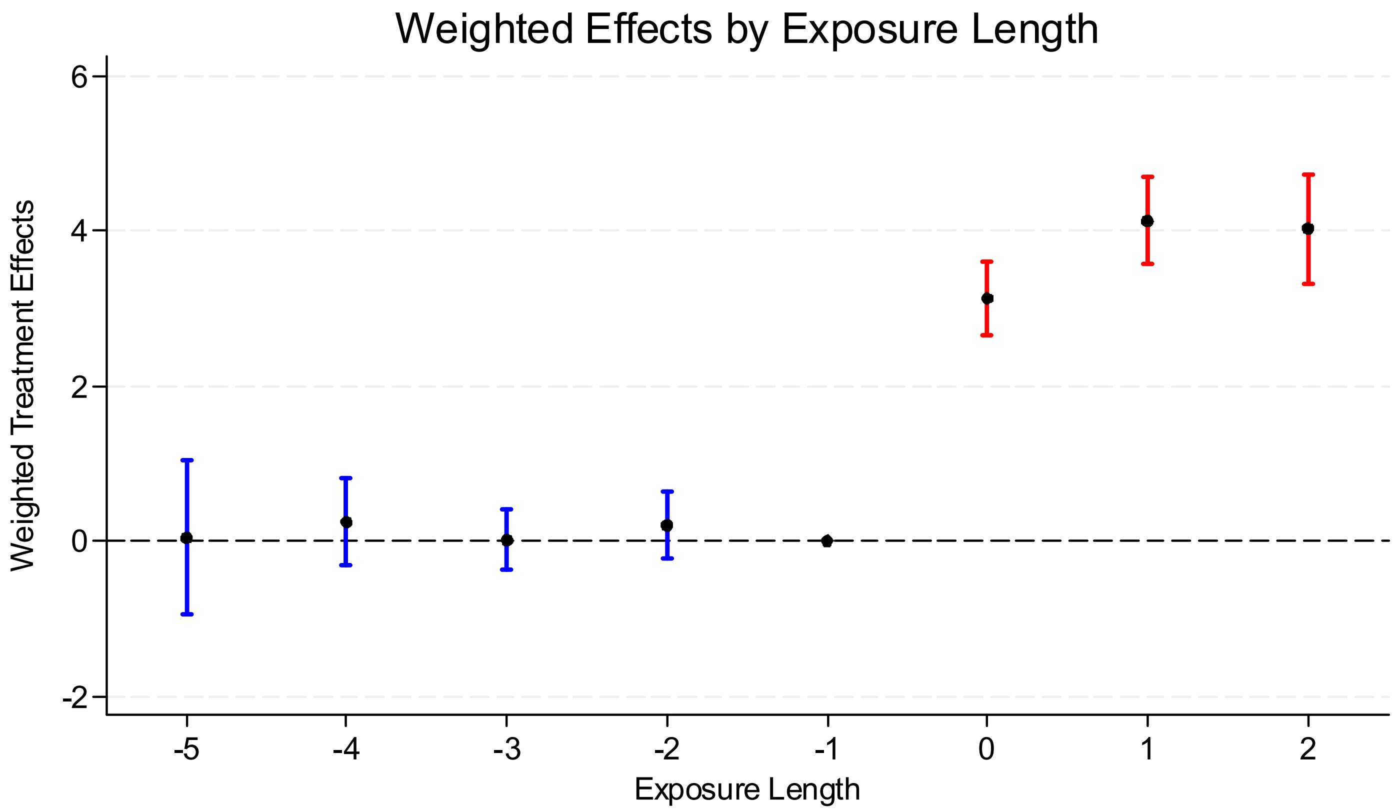
```
. * Can reproduce CS regression with the "very long regression" and
. * properly adjust standard errors:

* Pre-treatment exposure times:

gen expm1 = d4f03 + d5f04 + d6f05
gen expm2 = d4f02 + d5f03 + d6f04
gen expm3 = d4f01 + d5f02 + d6f03
gen expm4 = d5f01 + d6f02
gen expm5 = d6f01

* Treatment exposure times:

gen exp0 = d4f04 + d5f05 + d6f06
gen exp1 = d4f05 + d5f06
gen exp2 = d4f06
```

```
. qui reg y c.nw#c.d4f01 c.nw#c.d4f02 c.w#c.d4f04 c.w#c.d4f05 c.w#c.d4f06 ///
>         c.nw#c.d5f01 c.nw#c.d5f02 c.nw#c.d5f03 c.w#c.d5f05 c.w#c.d5f06 ///
>         c.nw#c.d6f01 c.nw#c.d6f02 c.nw#c.d6f03 c.nw#c.d6f04 c.w#c.d6f06 ///
>         c.nw#c.d4f01#c.x_dm4 c.nw#c.d4f02#c.x_dm4 c.w#c.d4f04#c.x_dm4 c.w#c.d4f05#c.x_dm4
>         c.nw#c.d5f01#c.x_dm5 c.nw#c.d5f02#c.x_dm5 c.nw#c.d5f03#c.x_dm5 c.w#c.d5f05#c.x_
>         c.nw#c.d6f01#c.x_dm6 c.nw#c.d6f02#c.x_dm6 c.nw#c.d6f03#c.x_dm6 c.nw#c.d6f04#c.x
>         c.d4 c.d5 c.d6 x c.d4#c.x c.d5#c.x c.d6#c.x ///
>         i.year i.year#c.x, vce(cluster id)

. margins, dydx(nw) subpop(if expm5 == 1) vce(uncond)

                                   (Std. err. adjusted for 500 clusters in id)
------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          nw |   .0428767   .5075724     0.08   0.933    -.9543658    1.040119
------------------------------------------------------------------------------

. margins, dydx(nw) subpop(if expm4 == 1) vce(uncond)

                                   (Std. err. adjusted for 500 clusters in id)
------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          nw |   .2440144   .2911646     0.84   0.402    -.3280453     .816074
------------------------------------------------------------------------------
```

```
. margins, dydx(nw) subpop(if expm3 == 1) vce(uncond)

                                      (Std. err. adjusted for 500 clusters in id)
------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          nw |   .0139107   .1987183     0.07   0.944     -.376517    .4043384
------------------------------------------------------------------------------

. margins, dydx(nw) subpop(if expm2 == 1) vce(uncond)

                                      (Std. err. adjusted for 500 clusters in id)
------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          nw |   .2015377   .2222352     0.91   0.365    -.2350943    .6381698
------------------------------------------------------------------------------
```

```
. margins, dydx(w) subpop(if exp0 == 1) vce(uncond)

                                    (Std. err. adjusted for 500 clusters in id)
------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           w |   3.129432   .2444446    12.80   0.000     2.649165    3.609699
------------------------------------------------------------------------------

. margins, dydx(w) subpop(if exp1 == 1) vce(uncond)

                                    (Std. err. adjusted for 500 clusters in id)
------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           w |   4.129554   .2821397    14.64   0.000     3.575226    4.683882
------------------------------------------------------------------------------
```

```
. margins, dydx(w) subpop(if exp2 == 1) vce(uncond)

                                  (Std. err. adjusted for 500 clusters in id)
------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           w |   4.033398   .3591623    11.23   0.000     3.327742    4.739055
------------------------------------------------------------------------------

 * Single weighted effect:
.
. margins, dydx(w) subpop(if w == 1) vce(uncond)

                                  (Std. err. adjusted for 500 clusters in id)
------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           w |   3.683272   .2071727    17.78   0.000     3.276234     4.09031
------------------------------------------------------------------------------
```


Weighted Effects by Exposure Length
Weighted Treatment Effects
Exposure Length
6
4
2
0
-2
-5
-4
-3
-2
-1
0
1
2

- See did_staggered_6_es.do for other characterizations of the estimators.
  - As 2×2 DiDs.
  - As imputation estimators.

Other Treatment Effect Estimators

- The default CS (2021) is not the linear RA estimator.
  - Doubly robust estimator based on linear RA and IPW (augmented IPW).
- Can apply *any* treatment effect estimator to the cross sections

$$\{(Y_{it} - Y_{i,g-1}, D_{ig}, \mathbf{X}_i)\}$$

- Lee and Wooldridge (2023): Replace long differences $Y_{it} - Y_{i,g-1}$ with

$$\dot{Y}_{itg} = Y_{it} - \frac{1}{(g-1)} \sum_{r=1}^{g-1} Y_{ir}$$

- Apply any TE estimator to

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

- Approaches have different sensitivities to violations of CPT.
- Lee and Wooldridge (2023): Replace $\dot{Y}_{itg}$ with unit-specific detrended outcomes.

5. Nonlinear Models

- Only rarely does adding many unit FEs not result in the incidental parameters problem.
- In linear case, equivalent to controlling for a (small) number of cohort dummies.
- Can include the cohort dummies in a variety of nonlinear models.

• Wooldridge (2023, Econometrics Journal): Use an index version of conditional PT.

$$E[Y_t(\infty)|D_q,\ldots,D_T,\mathbf{X}] = G\Bigg(\alpha + \sum_{g=q}^{T}\beta_g D_g + \mathbf{X}\boldsymbol{\kappa} + \sum_{g=q}^{T}(D_g \cdot \mathbf{X})\boldsymbol{\eta}_g + \gamma_t + \mathbf{X}\boldsymbol{\pi}_t\Bigg)$$

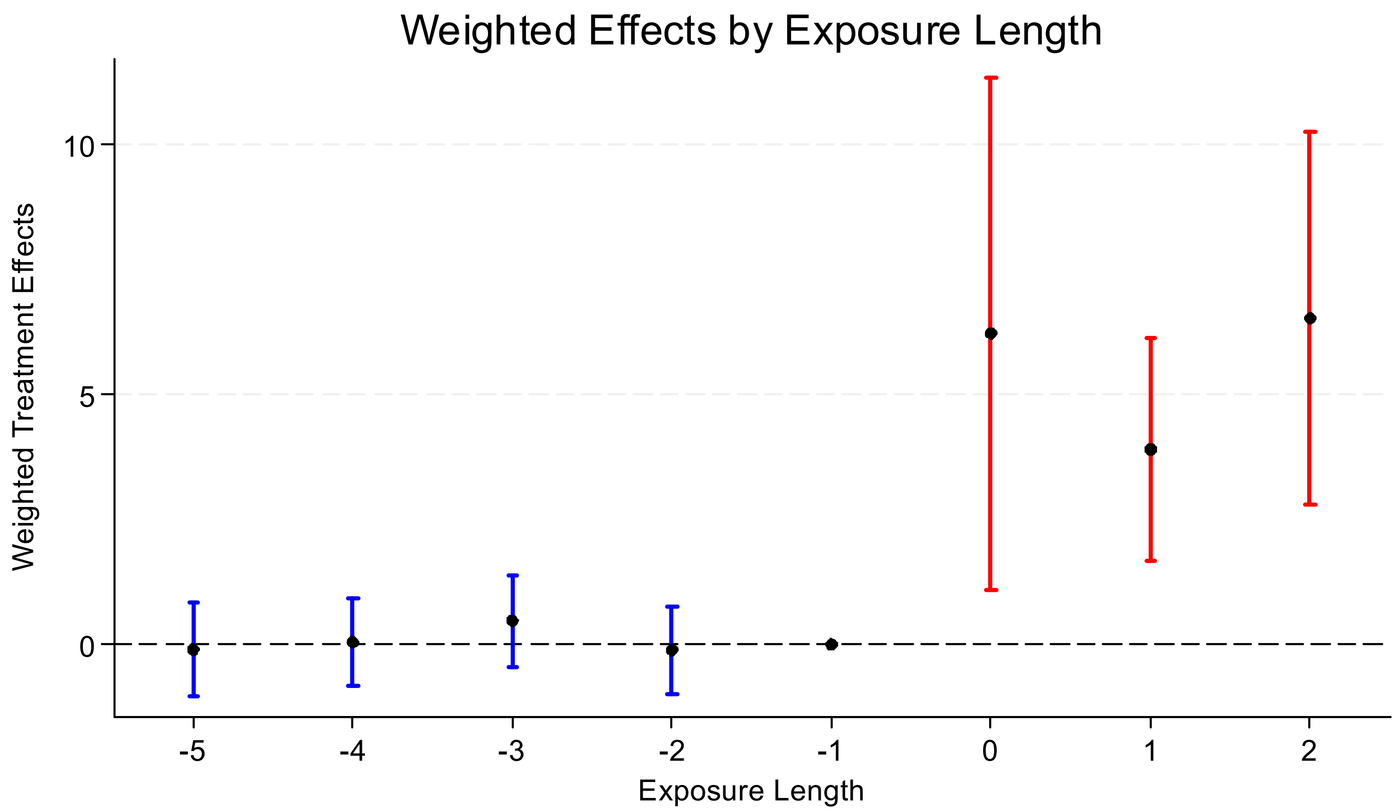
• Linear case [Wooldridge (2021)]:

$$\begin{aligned} E[Y_t(\infty)|D_q,\ldots,D_T,\mathbf{X}] &= \alpha + \sum_{g=q}^{T} \beta_g D_g + \mathbf{X}\boldsymbol{\kappa} \\ &\quad + \sum_{g=q}^{T} (D_g \cdot \mathbf{X})\boldsymbol{\eta}_g + \gamma_t + \mathbf{X}\boldsymbol{\pi}_t \end{aligned}$$

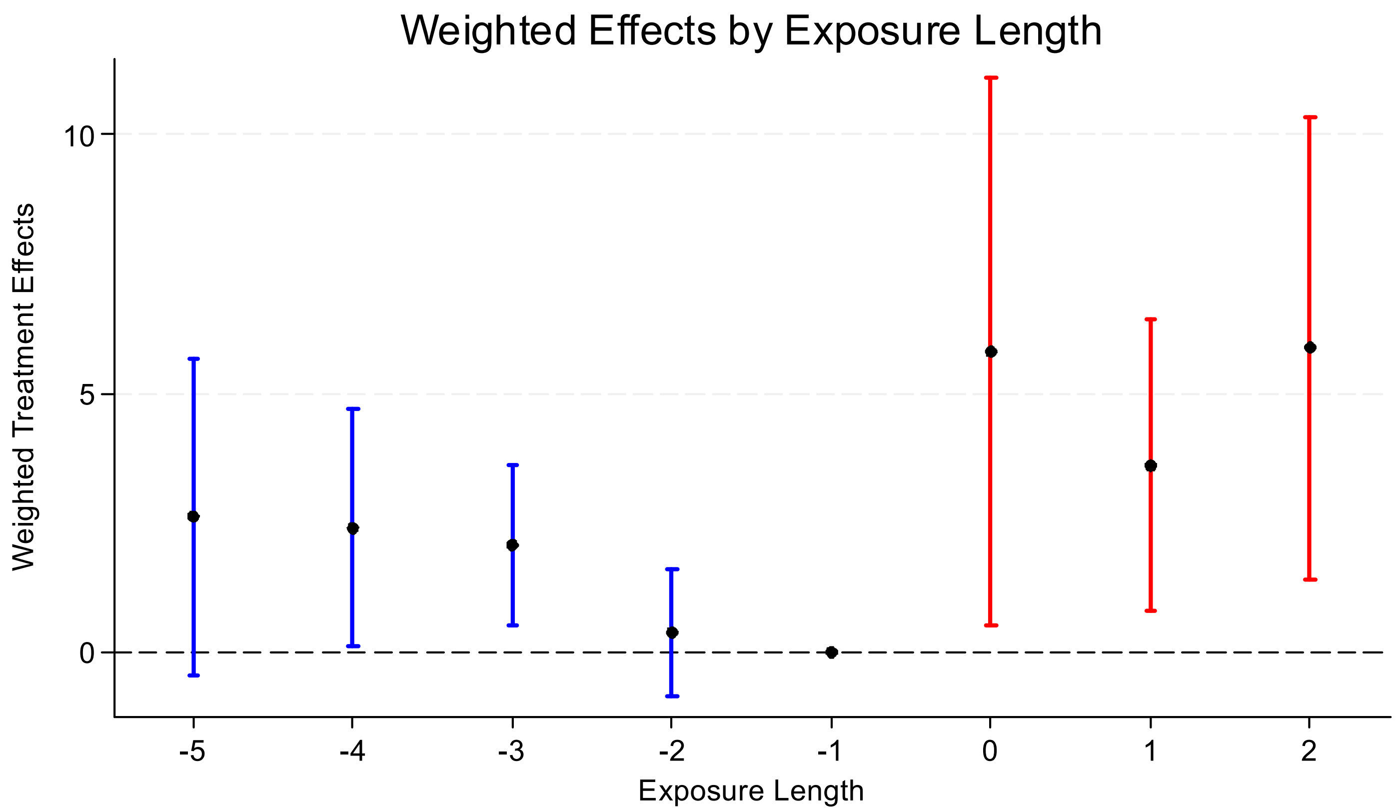
$$E[Y_t(\infty)|D_q,\ldots,D_T,\mathbf{X}] - E[Y_1(\infty)|D_q,\ldots,D_T,\mathbf{X}] = \gamma_t + \mathbf{X}\boldsymbol{\pi}_t$$

- The "lags only" estimator – extension of pooled OLS in linear case – simply replaces `regress` with `logit`, `fraclogit`, `poisson`.
- Advantage in using canonical link pairs.
  - Pooled estimation gives same ATT estimates as (theoretically justified) imputation.
- See did_common_6_logit_es.do and did_staggered_6_poisson_es.do.
- Especially in the count case, the Poisson regression is more precise and passes the pre-trends test.
  - Linear model fails conditional PT.

- Exponential, pooled Poisson:



- Linear model, pooled OLS:



6. Extensions and Stata Wish List

- Regression methods can easily allow staggered exit from treatment.
- Generally, index cohorts by entry and exit time.
  - The potential outcomes are now $Y_t(g,h)$.
  - First treated period is g; exit occurs in h.
- $D_{g,h}$ for $g < h \leq T$ are the new cohort indicators.
- With a never treated group,

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

  - $Y_r(\infty)$ is the PO in the never treated state.

- ATTs are defined even when $t \geq h$ – that is, after the intervention has been removed.
  - Can estimate persistence even after program is eliminated.
  - When $t \geq h$, can see whether an effect dissipates after the intervention disappears.
- Estimation: In place of the interactions $D_g \cdot fs_t$, $s = g, \ldots, T$, include

$$D_{g,h} \cdot fs_t, \quad g < h, s = g, \ldots, T$$

- See did_exit_6_es.do.

Extensions to `xthdidregress`?

```
xthdidregress twfe (y x), (w) group(id)
xthdidregress ra (y x), (w) group(id)
xthdidregress logit (y x), (w) group(id)
xthdidregress fraclogit (y x), (w) group(id)
xthdidregress poisson (y x), (w) group(id)
xthdidregress logit (y x), (w) group(id)
   event
xthdidregress twfe, (y x) (w) group(id)
   hetrend
```

- Easy to automatically detect exit and define cohort dummies.

```
. gen exp0 = d4inf4 + d464 + d454 + d5inf5 + d565 + d6inf6

. gen exp1 = d4inf5 + d465 + d5inf6

. gen exp2 = d4inf6

. qui reg y c.w#c.d4inf4 c.w#c.d4inf5 c.w#c.d4inf6 c.w#c.d464 c.w#c.d465 c.d466 ///
>         c.w#c.d454 c.d455 c.d456 c.w#c.d5inf5 c.w#c.d5inf6 c.w#c.d565 c.d566 c.w#c.d6inf6
>         c.w#c.d4inf4#c.x_dm4_inf c.w#c.d4inf5#c.x_dm4_inf c.w#c.d4inf6#c.x_dm4_inf c.w#
>         c.w#c.d465#c.x_dm4_6 c.d466#c.x_dm4_6 c.w#c.d454#c.x_dm4_5 c.d455#c.x_dm4_5 c.d456
>         c.w#c.d5inf5#c.x_dm5_inf c.w#c.d5inf6#c.x_dm5_inf c.w#c.d565#c.x_dm5_6 c.d566#c
>         i.year i.year#c.x ///
>         c.d4_inf c.d4_5 c.d4_6 c.d5_inf c.d5_6 c.d6_inf x ///
>         c.d4_inf#c.x c.d4_5#c.x c.d4_6#c.x c.d5_inf#c.x c.d5_6#c.x c.d6_inf#c.x, vce(cluster

. margins, dydx(w) subpop(if exp0 == 1) vce(uncond)

                                 (Std. err. adjusted for 1,000 clusters in id)
------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           w |   3.579169   .1347143    26.57   0.000     3.314814    3.843525
------------------------------------------------------------------------------
```

```
. margins, dydx(w) subpop(if exp1 == 1) vce(uncond)

                                 (Std. err. adjusted for 1,000 clusters in id)
------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           w |   4.635792   .1635907    28.34   0.000     4.314771    4.956813
------------------------------------------------------------------------------

. margins, dydx(w) subpop(if exp2 == 1) vce(uncond)

                                 (Std. err. adjusted for 1,000 clusters in id)
------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           w |    5.88682   .2415575    24.37   0.000     5.412802    6.360838
------------------------------------------------------------------------------

. margins, dydx(w) subpop(if w == 1) vce(uncond)

                                 (Std. err. adjusted for 1,000 clusters in id)
------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           w |   4.355817   .1216767    35.80   0.000     4.117046    4.594588
------------------------------------------------------------------------------
```

Exit in Event Study Estimation

- Now include $D_{gh} \cdot fs_t$ for all

$$s \neq g - 1$$

  - Again, period $g - 1$ is the comparison (base) period.

- Use `margins` with `subpop()` to obtain estimates by exposure time.

  - Gives an ES plot with pre-trends and effects within treatment.

- Can also use `margins` with `subpop()` to estimate effects of time since last exposure.
- If desired, aggregate by initial treatment time, so that effects in treatment and post treatment get averaged together.
- See did_exit_6_es.do.