ECON 594: Applied Economics

Event Study Designs

Sam Norris
University of British Columbia

Where we are

- Last class: panel data, fixed effects, two-way fixed effects DD
- Today: event study designs
  - The dynamic version of DD
  - How effects evolve over time
  - How (sort of) to test parallel trends
  - Why the standard TWFE specification can give you the wrong answer in staggered designs, and what to do about it

From static to dynamic

- Last class's TWFE estimates a single coefficient δ:

$$Y_{st} = \alpha_s + \gamma_t + \delta \cdot D_{st} + e_{st}$$

- This is a static effect: average effect of being treated, pooled across all post-treatment time

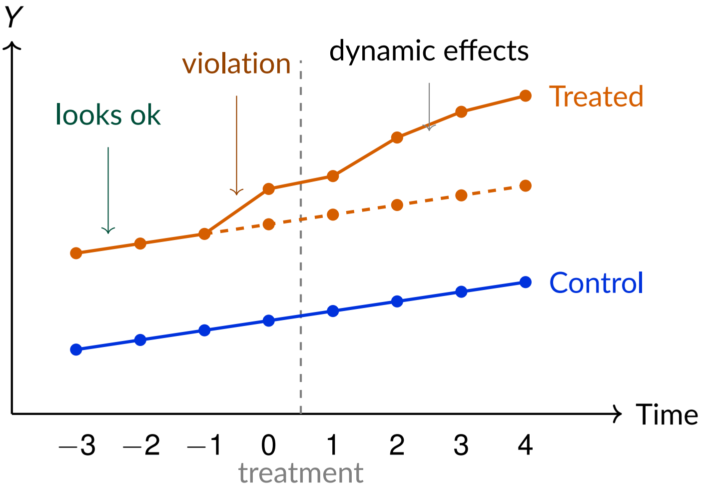
- Often we want to know how the effect evolves:
  - Does it kick in immediately, or build up?
  - Does it fade out, or persist?
  - Are there anticipation effects before the policy?

- And: are pre-treatment trends parallel?

Two motivations for event studies

- 1. Dynamic treatment effects
  - Effect of UI extensions on unemployment duration: small at first, builds as the new max binds
  - Effect of a criminal conviction: scarring that compounds
- 2. Testing parallel trends
  - Parallel trends is fundamentally untestable in the post-period
  - But we can check whether trends were parallel before treatment
  - If yes, more credible (not proof) that they would have remained parallel
- Both come out of the same regression

Many periods let us see more



Multiple pre-periods let us check parallel trends; multiple post-periods let us trace out how effects evolve.

From treatment dummy to relative-time dummies

- Replace the single treatment dummy with a set of dummies for time relative to treatment

- Define k_{it} as periods relative to unit i's last pre-treatment year:
  - $k = 0$: last period before treatment (reference)
  - $k = 1$: first period of treatment
  - $k = -1$: two periods before treatment
  - $k = 5$: fifth period of treatment

- Run:

$$Y_{it} = \alpha_i + \gamma_t + \sum_{k \neq 0} \beta_k \cdot \mathbb{1}[k_{it} = k] + e_{it}$$

- Each β_k is the average difference in Y between units in relative period k and the omitted reference period

Choices to make

- Reference period: omit one k (otherwise collinear with FEs)
  - Convention: omit $k = 0$ (last period before treatment)
  - All $\hat{\beta}_k$ are then changes relative to period 0
- Endpoint binning: pre-treatment periods $K \leq \underline{k}$ and post-treatment $K \geq \overline{k}$ get binned into one dummy each
  - Otherwise units observed at extreme values dominate the endpoint estimates
- Window: pick $\underline{k}$ and $\overline{k}$ based on what's identified across most units

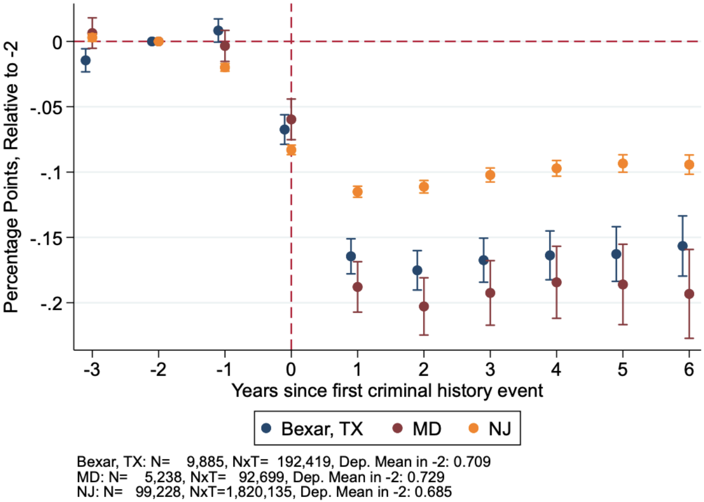
Event studies in Stata

```
* Single treatment cohort: all treated units treated in 1975.
* treat_year = 1975 if treated, missing if never-treated.

* Event time: k=0 = 1974 (last pre), k=1 = 1975 (first post)
* Never-treated stay at ktime = 0, so all event-time dummies will be 0 for them
gen ktime = 0
replace ktime = year - treat_year + 1 if !missing(treat_year)

* Bin endpoints to a window of -3 to +3
replace ktime = -3 if ktime < -3
replace ktime =  3 if ktime >  3

* Manually create dummies; k = 0 omitted as reference
gen pre3  = (ktime == -3)
gen pre2  = (ktime == -2)
gen pre1  = (ktime == -1)
gen post1 = (ktime ==  1)
gen post2 = (ktime ==  2)
gen post3 = (ktime ==  3)

reghdfe y pre3 pre2 pre1 post1 post2 post3, absorb(id year) cluster(id)

coefplot, keep(pre* post*) vertical yline(0)
```

What to look for

- Pre-period ($k < 0$): coefficients should be near zero
  - Tests parallel trends in the pre-period
  - If they trend up or down, parallel trends in the post-period is suspect
- Post-period ($k \geq 1$): the dynamic treatment effects
  - Build-up, decay, persistence visible directly
- Always plot the coefficients with confidence bands
  - One of the most informative figures you can put in a paper

Example: effect of criminal conviction on employment



Agan, Garin, Koustas, Mas, Yang (2024, AER:Insights), “The Labor Market Impacts of Reducing

Pre-trends look bad. Now what?

- If pre-period coefficients trend, parallel trends is in trouble. Options:
  - Find a better control group (most credible fix)
  - Add controls that absorb the trend (covariates, cohort interactions)
  - Use one of the modern estimators (more in a few slides)
  - Honest DiD (Rambachan & Roth 2023): how much do conclusions change under bounded violations of parallel trends?
- Don't: ignore the pre-trend, hope the referee doesn't notice, or fit a linear trend and "detrend" away the problem

TWFE with staggered treatment

- Last class: TWFE pools many “natural experiments” into one estimate

- Question we glossed over: what is each natural experiment comparing?
  - Newly-treated units to never-treated units?
    Good.
  - Newly-treated units to not-yet-treated units?
    Good (assuming parallel trends).
  - Newly-treated units to already-treated units?
    Bad.

- TWFE silently uses all three

Why already-treated controls are a problem

- Imagine treatment effects grow over time
- Compare a state treated this year to one treated three years ago
  - Newly-treated unit: small effect (just started)
  - Already-treated unit: large effect (mature)
  - DD: newly-treated minus already-treated $=$ negative
- Even though the true treatment effect is positive everywhere
- TWFE can put a negative weight on some treatment effects
- Pathological case: average treatment effect is positive, $\hat{\delta}$ is negative

Goodman-Bacon (2021)

- TWFE $=$ weighted average of all possible 2×2 DDs in the data
- Weights depend on group sizes and timing variance
- Some weights can be negative when treatment effects vary over time
- This is a property of the OLS regression, not the data
- In a clean staggered design with no never-treated units, the problem can be severe
- Decomposition (`bacondecomp` in Stata) shows you which 2×2s drive your estimate

When does this matter?

- Matters most when:
  - Treatment is staggered (units treated at different times)
  - Few or no never-treated units
  - Treatment effects are heterogeneous across cohorts or grow over time
- Matters less when:
  - Single treatment date for all treated units (classic 2×2)
  - Large never-treated control group
  - Treatment effects roughly constant over time and cohorts
- For most thesis projects: at least check whether it matters

Before you reach for the fancy estimators

- You can fix most of the problem with three changes you can do yourself:
  1. Restrict controls to never-treated units
  2. Saturate the regression with cohort-by-event-time interactions
  3. Build a stacked dataset of clean sub-experiments
- Each is a small change to the regression you'd run anyway
- Often gives nearly identical results to the modern estimators

Fix 1: restrict controls to never-treated

- The cleanest controls are units that are never treated in the sample
  - Drop already-treated units so they can't serve as controls
  - Run TWFE on treated units (with their event-time dummies) + never-treated

```
gen ever_treated = !missing(treat_year)
gen ktime = 0
replace ktime = year - treat_year + 1 if ever_treated

* Keep all never-treated, plus treated within event-time window
keep if !ever_treated | inrange(ktime, -3, 3)

* Build the dummies as before, omit k = 0 as reference, then run
reghdfe y pre3 pre2 pre1 post1 post2 post3, absorb(id year) cluster(id)
```

- Costs you statistical power, buys you a clean comparison
- Requires a meaningful never-treated group to exist

Fix 2: saturate with cohort × event time

- Let E_i = unit i's treatment cohort (year of first treatment)

- Let each cohort c have its own dynamic treatment effects:

$$Y_{it} = \alpha_i + \gamma_t + \sum_{c} \sum_{k \neq 0} \beta_{c,k} \cdot \mathbb{1}[E_i = c] \cdot \mathbb{1}[k_{it} = k] + e_{it}$$

- Then aggregate the $\hat{\beta}_{c,k}$ across cohorts to get an average dynamic effect
    - Weight by cohort size

- This is the Sun–Abraham (2021) estimator, almost

- Stata: `eventstudyinteract`

- Tradeoff: many parameters, less precision

Fix 3: stacked DiD

- Cengiz, Dube, Lindner, Zipperer (QJE 2019):
  1. For each treated cohort c, build a sub-experiment:
     - Treated units in cohort c
     - Never-treated control units (or not-yet-treated within a window)
     - Time window $[c - K_{pre},\ c + K_{post}]$
  2. Define event time within each sub-experiment
  3. Stack all sub-experiments into one long dataset
  4. Run TWFE with sub-experiment $\times$ unit FE and sub-experiment $\times$ event-time FE

- Each comparison is clean (no already-treated controls)

- Pre-periods aligned across cohorts $\Rightarrow$ event study coefficients have one interpretation

- Stata: `stackedev`, or roll your own with a loop over cohorts and `append`

If you want the off-the-shelf version

- Several papers proposed estimators that handle staggered treatment + heterogeneous effects properly

- All do something similar to the DIY fixes, but cleaner and with better standard errors

- The four most cited:
  - Callaway & Sant'Anna (2021)
  - Sun & Abraham (2021)
  - Borusyak, Jaravel, Spiess (2024)
  - de Chaisemartin & D'Haultfœuille (2020)

- For a thesis: pick one, use it as a robustness check, cite the others

The big four

Estimator	Idea	Stata
Callaway–Sant'Anna	Compute group-time ATTs $ATT(g,t)$ for each cohort/period using clean controls; aggregate	`csdid`
Sun–Abraham	Saturate with cohort $\times$ event-time interactions; aggregate by cohort weights	`eventstudyinteract`
Borusyak–Jaravel–Spiess	Impute the never-treated counterfactual for treated obs; difference	`did_imputation`
de Chaisemartin–D'Haultfœuille	Use only switchers vs. stayers in adjacent periods	`did_multiplegt`

- All require `ssc install`; check the help file before running

Useful surveys

- Roth, Sant'Anna, Bilinski, Poe (2023, J. Econometrics): "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature"
  - Decision tree for which estimator to use
  - Honest treatment of the limitations of each
- de Chaisemartin & D'Haultfœuille (2023): survey of the same literature from one of its authors
- Goodman-Bacon (2021, J. Econometrics): the original decomposition paper
- Roth (2024): "Interpreting Event-Studies from Recent Difference-in-Differences Methods"

For your thesis

- Always:
  - Plot your event study with confidence bands
  - Use $k = 0$ as reference, bin the endpoints, cluster at the unit level
  - Show your TWFE result alongside at least one robust estimator
- If staggered treatment + few never-treated:
  - Run `bacondecomp` to see where your estimate comes from
  - Restrict to clean comparisons or use a modern estimator
- If pre-trends look bad:
  - Don't ignore them
  - Better control group is the first lever to pull

Recap

- Event studies are dynamic DD: coefficient for each period relative to treatment
- Pre-period coefficients test parallel trends; post-period are dynamic effects
- Naive TWFE on staggered data can be badly biased when effects vary across cohorts or over time
- Three DIY fixes: restrict to never-treated, saturate with cohort $\times$ event-time, stack
- Several modern estimators do this for you, with proper standard errors
- Plot it, cluster it, robustness-check it

Coming up

- No class Monday (Victoria Day)
- Tuesday: Tables and Figures
- Wednesday: Discrete Choice Models
- Reminder: progress report due Wednesday June 3rd