# Experimental Analysis of Public Policies

Session 6

Event Studies and Staggered DiD

Etienne Dagorn - Universite de Lille

# Session Objectives

By the end of this session, students should be able to:

- → define event time and interpret omitted periods, leads, and lags;
- → read an event-study graph as both diagnostic evidence and effect dynamics;
- → explain why traditional TWFE event-study coefficients need qualification;
- → identify anticipation, dynamic effects, and weak support in long lags;
- → explain why staggered adoption complicates TWFE;
- → distinguish never-treated, not-yet-treated, and already-treated comparison units;
- → define $Y_{it}(g)$ and interpret $ATT(g,t)$;
- → describe why modern DiD estimators keep cohort and event time explicit.

# Roadmap

# Roadmap

From 2x2 DiD To Event Time

Event Time And Traditional TWFE

Staggered Adoption: Who Is Control?

Why Naive TWFE Can Mislead

Group-Time ATT And Aggregation

Practice And Bridge

# From DiD To Dynamics

- → Session 5: one treated group, one comparison group, one treatment date.
- → Core assumption: the comparison group reveals the treated group's untreated trend.
- → New problem: effects may appear slowly, fade, or differ across cohorts.
- → New question: who serves as control at each date?

**Central question**

What changes when treatment timing and treatment effects are heterogeneous?

# Essential Versus Optional

| | |
|---|---|
| **Essential** | event time, reference period, leads, lags, support, comparison groups |
| **Optional** | full negative-weight algebra and estimator derivations |

**Teaching choice**

Students should leave with the comparison logic even if they do not master every estimator.

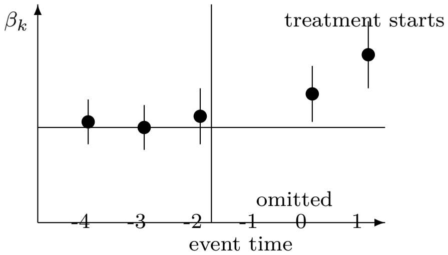
# Why Event Studies?

- → Visualise pre-treatment dynamics.
- → Estimate short-run and medium-run effects.
- → Detect possible anticipation.
- → Communicate timing in a way tables often cannot.

**Interpretation habit**

Always name the omitted period before saying what any coefficient means.

# Event-Study Graph: What To Read




- → Reference period.
- → Pre-treatment coefficients.
- → Change around treatment onset.
- → Confidence intervals and support.

# Event-Study Workshop: Read The Graph

**Prompt:** an event-study graph shows coefficients from $k = -5$ to $k = 5$, omitting $k = -1$.

1. What is the reference period?
2. Do the leads suggest pre-trends, anticipation, noise, or low power?
3. Does the effect appear immediately, gradually, or temporarily?
4. Which event times are likely based on fewer cohorts?

**Student output**

One sentence on pre-trends, one on effect timing, one on uncertainty or support.

# Roadmap

# Event Time

$$k = t - G_i$$

- $G_i$: first period in which unit $i$ receives treatment.
- $k = 0$: first treated period.
- $k = -1$: period just before treatment.
- $k = 1$: one period after treatment.
- Never-treated units have $G_i = \infty$, so event time is not defined for them.

# Event-Time Data Layout

| | 2022 | 2023 | 2024 | 2025 |
|---|---|---|---|---|
| Early adopter, $G = 2023$ | $k = -1$ | $k = 0$ | $k = 1$ | $k = 2$ |
| Late adopter, $G = 2025$ | $k = -3$ | $k = -2$ | $k = -1$ | $k = 0$ |
| Never-treated, $G = \infty$ | control | control | control | control |

**Student output**

Convert two calendar years into event time and mark which units are untreated in each year.

# Traditional TWFE Event-Study Regression

$$Y_{it} = \alpha_i + \lambda_t + \sum_{k \in \mathcal{K},\, k \neq -1} \beta_k 1[G_i < \infty,\ t - G_i = k] + u_{it}$$

- → Unit fixed effects absorb permanent unit differences.
- → Time fixed effects absorb common calendar shocks.
- → One event-time period is omitted as the reference.
- → Never-treated units have all event-time indicators equal to zero.
- → Endpoints such as $k \leq -5$ or $k \geq 5$ are often binned.

# Traditional TWFE: Immediate Caveat

**Do not overread the dots**

In staggered designs with heterogeneous effects, traditional TWFE $\beta_k$'s may not equal clean event-time ATTs.

- → Already-treated units can enter comparisons for newly treated units.
- → Dynamic effects from early cohorts can contaminate lead or lag coefficients.
- → The graph still teaches timing, but it needs a comparison-group audit.

# The Omitted Period

- → Usually omit $k = -1$, the period just before treatment.
- → Each coefficient is interpreted relative to that omitted period.
- → Changing the omitted period changes the visual reference point, not the underlying information.

**Reading cue**

If $k = -1$ is omitted, a coefficient at $k = 2$ is relative to one period before treatment.

# Leads And Pre-Trend Diagnostics

- → Lead coefficients correspond to periods before treatment.
- → In a clean comparison design, leads should be close to zero under no anticipation and parallel pre-trends.
- → In traditional staggered TWFE, leads may also reflect contamination from heterogeneous post-treatment effects.
- → They are diagnostics, not definitive tests.

**Low-power warning**

A flat pre-period with wide confidence intervals may simply mean the design cannot detect violations.

# Lags And Dynamic Effects

- → Lag coefficients correspond to periods after treatment.
- → In a clean event-study estimator, lags describe post-treatment dynamics.
- → In traditional staggered TWFE, lags need a comparison-group and weighting audit before causal interpretation.
- → Long-run lags may be imprecise if few cohorts are observed that long.

**Policy example**

A training programme may show no effect at $k = 0$, gains at $k = 2$, and fade-out by $k = 5$.

# Anticipation

- → Units may respond before official treatment starts.
- → Firms may invest before a tax change.
- → Households may move before a boundary reform.
- → Researchers may redefine the treatment date or omit anticipation periods.

**Interpretation choice**

Sometimes the relevant policy is the announcement plus implementation, not only the legal start date.

# How To Read The Graph

1. Locate the omitted period.
2. Identify the estimator and comparison group.
3. Check the pre-treatment coefficients without treating them as proof.
4. Look for a change around treatment onset.
5. Compare magnitudes to policy relevance.
6. Notice confidence intervals and sample support.

# Roadmap

# Example: Minimum Wage

- → Cities raise minimum wages in different years.
- → Employment may adjust slowly.
- → Early adopters may differ politically and economically.
- → Later adopters can be untreated controls before they adopt, but not after.

**Class question**

Should a city that raised the minimum wage three years ago serve as a control for a newly treated city?

# Staggered Adoption

- → Different units adopt the policy in different periods.
- → Some units may never adopt.
- → Some treated units may act as controls for later-treated units before they adopt.
- → Already-treated units may also enter traditional TWFE comparisons.

**Why it matters**

Control status becomes relative to calendar time and treatment cohort, not a fixed label.

# Cohort-By-Calendar Grid

| **Group** | 2022 | 2023 | 2024 | 2025 |
|---|---|---|---|---|
| Early adopters | untreated | treated | treated | treated |
| Late adopters | untreated | untreated | untreated | treated |
| Never-treated | untreated | untreated | untreated | untreated |

**Student output**

For 2024, mark who is treated, who is not-yet-treated, who is never-treated, and who should not be used as an untreated control.

# Staggered Adoption: Debrief Grid

| Comparison | Interpretation risk |
|---|---|
| Treated vs never-treated | clean if never-treated units are valid controls |
| Treated vs not-yet-treated | useful before later adoption occurs |
| Treated vs already-treated | risky when effects evolve over time |
| Early vs late adopters | may mix cohort differences and dynamics |

**Reflex**

In staggered DiD, always ask: who is serving as the control group at each date?

# Comparison Groups

- Never-treated units are often the cleanest comparison group if they are credible.
- Not-yet-treated units can be useful before they adopt.
- Already-treated units are problematic if treatment effects evolve over time.
- The comparison group determines the estimand and the external-validity claim.

**Diagnostic prompt**

For each event time, ask whether the comparison units are never-treated, not-yet-treated, or already exposed.

# Roadmap

# TWFE Weights: Intuition

- → TWFE combines many 2x2 comparisons.
- → In staggered designs, some comparisons may be undesirable.
- → If effects differ by cohort or time since treatment, the weighted average can be hard to interpret.
- → The issue is not fixed effects themselves; it is the comparison structure.

**Plain words**

A single TWFE coefficient can mix early-vs-late, late-vs-early, and treated-vs-never comparisons.

# Negative Weights

**Advanced warning**

With staggered adoption and heterogeneous effects, TWFE can place negative or unintuitive weights on some comparisons.

- → This can make the coefficient differ from a transparent average treatment effect.
- → The full algebra is optional.
- → The essential habit is to mistrust a coefficient whose comparison groups are opaque.

# Heterogeneous Effects

- → Effects may differ across cohorts.
- → Effects may differ by exposure length.
- → Early adopters may be systematically different.
- → Treatment quality may change as the policy scales.

**Example**

Early minimum-wage adopters may be high-income cities, while later adopters may face different labour-market conditions.

# Common Mistakes

**Watch for these**

- → Treating insignificant leads as proof of validity.
- → Forgetting which period is omitted.
- → Interpreting traditional TWFE leads and lags without checking comparisons.
- → Mixing already-treated units into comparisons without interpretation.
- → Reporting one TWFE coefficient when dynamics are central.
- → Ignoring support for long-run event times.

# Roadmap

# Potential Outcomes With Treatment Timing

- $Y_{it}(g)$: potential outcome at time $t$ if unit $i$ first receives treatment in period $g$.
- $Y_{it}(\infty)$: potential outcome at time $t$ if unit $i$ is never treated.
- $G_i = g$: unit $i$'s observed treatment cohort.
- This notation makes timing part of the causal object.

**Why new notation?**

Static $Y_i(1)$ and $Y_i(0)$ hide when treatment begins and how long exposure has lasted.

# Group-Time ATT

$$ATT(g,t) = E[Y_{it}(g) - Y_{it}(\infty) \mid G_i = g], \qquad t \geq g$$

- → $g$: cohort first treated in period $g$.
- → $t$: calendar period.
- → The target is local to a cohort and date.
- → Identification compares cohort $g$ to credible never-treated or not-yet-treated units.

# Cleaner Comparison Logic

- → Estimate group-time treatment effects.
- → Compare treated cohorts to never-treated or not-yet-treated units.
- → Keep already-treated units out of untreated control roles.
- → Aggregate effects transparently across cohorts and event times.

**Teaching cue**

Modern estimators are not magic; they make the comparison group and aggregation visible.

# Aggregation

- Dynamic aggregation: average by event time.
- Cohort aggregation: average by treatment cohort.
- Calendar-time aggregation: average effects in a given year.
- Overall aggregation: average across treated observations.
- Different aggregations answer different policy questions.

**Reporting rule**

Never report an aggregate without saying which event times and cohorts receive weight.

# Inference And Graph Design

- → Cluster at the unit or treatment-assignment level.
- → Report confidence intervals for event-time or group-time estimates.
- → Long leads and lags can be noisy because support falls at the extremes.
- → Graphs should show event time, treatment onset, omitted period, uncertainty, comparison group, and estimator.

**Graph warning**

Do not interpret each dot as equally precise or equally well supported.

# Code Mini-Lab

- → Create event-time indicators.
- → Estimate a simple traditional TWFE event-study regression.
- → Plot coefficients with confidence intervals.
- → Compare that representation to group-time ATT logic.
- → Make event time visible in the data before fitting the regression.

# Roadmap

From 2x2 DiD To Event Time

Event Time And Traditional TWFE

Staggered Adoption: Who Is Control?

Why Naive TWFE Can Mislead

Group-Time ATT And Aggregation

Practice And Bridge

# Practice Bridge

- → Theory sheet: `exercises/theory/session_06_theory.md`.
- → QCM: `assessments/qcm/session_06_qcm.md`.
- → Exercise: interpret an event-study graph and identify threats.
- → Exam skill: explain what a dynamic DiD coefficient compares.

**Student deliverable**

One paragraph describing leads, lags, omitted period, support, and the relevant comparison group.

# Transition To Session 7

→ IV identified local effects for compliers.

→ Staggered DiD identifies local effects by cohort and calendar time.

→ RDD will identify effects local to units around an institutional cutoff.

**Bridge intuition**

We move from "same trend without treatment" to "smooth potential outcomes at the cutoff".

# Key Takeaways

- → Event studies show timing, diagnostics, and dynamic effects.
- → Leads are useful but not definitive tests.
- → Traditional TWFE event studies need caution under staggered adoption and heterogeneous effects.
- → Cleaner estimands keep cohort and event time explicit.
- → Always report the comparison group and the aggregation.

# Appendix: Names To Recognise

- → Sun and Abraham: interaction-weighted event studies.
- → Callaway and Sant'Anna: group-time average treatment effects.
- → Borusyak, Jaravel, and Spiess: imputation approach.
- → de Chaisemartin and D'Haultfoeuille: robust DiD estimators.

**How to use this appendix**

Recognise estimator names in papers; focus class time on the comparison logic and interpretation.

# References and Further Reading


→ Sun and Abraham (2021), "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects".

→ Callaway and Sant'Anna (2021), "Difference-in-Differences with Multiple Time Periods".

→ Goodman-Bacon (2021), "Difference-in-Differences with Variation in Treatment Timing".

→ Roth, Sant'Anna, Bilinski, and Poe (2023), "What's Trending in Difference-in-Differences?".

→ Borusyak, Jaravel, and Spiess (2024), "Revisiting Event-Study Designs: Robust and Efficient Estimation".

→ de Chaisemartin and D'Haultfœuille (2020), "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects".


**Reading reflex**

When reading a modern DiD paper, identify the comparison group and aggregation before interpreting the headline coefficient.