Experimental Analysis of Public Policies

Sessions 5–6: Policy timing

DiD, event studies and staggered adoption

Etienne Dagorn
University of Lille

What You Should Be Able to Defend

By the end of this class, you should be able to:

- reconstruct the treated group's missing untreated change and derive its ATT;
- separate the parallel-trends assumption from a regression and its standard errors;
- explain which comparisons contaminate staggered TWFE estimates;
- build an $ATT(g,t)$, then explain whose effects enter an event-time average.

Our three hours, including discussion

0–35: counterfactual and calculation; 35–80: identification, regression and R.
80–90: break; 90–140: event time and staggered comparisons.
140–165: cohort effects and aggregation; 165–180: synthesis and exit ticket.

Roadmap

Build the Missing Counterfactual

Assumptions, Regression and Inference

Read Effects in Event Time

Staggered Adoption: Who Is the Control?

Build and Aggregate Cohort-Specific Effects

Connect the Design to the General Method

Roadmap

Build the Missing Counterfactual

Assumptions, Regression and Inference

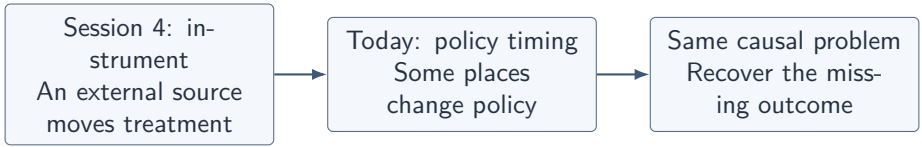
Read Effects in Event Time

Staggered Adoption: Who Is the Control?

Build and Aggregate Cohort-Specific Effects

Connect the Design to the General Method

From an Instrument to a Policy Change

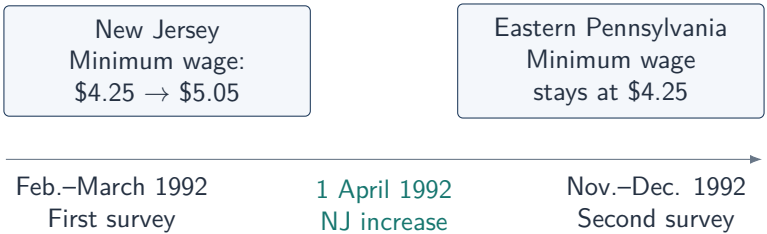


A region introduces a youth employment subsidy. Employment rises afterwards.

Opening poll

Is this evidence of a policy effect? What else could explain the increase?
What would you want to observe in a comparison region?

A Paper's Policy Question: Minimum Wages and Jobs



Question: did the increase reduce fast-food employment?
Design: survey 410 restaurants before and after; compare changes across states.

Card and Krueger (1994), AER, pp. 772–775.

Two Tempting Comparisons Leave Something Missing

Before versus after, treated only

$$\bar{Y}_{T,1} - \bar{Y}_{T,0}$$

General recovery?

Employment might rise even without the subsidy.

Treated versus control, after only

$$\bar{Y}_{T,1} - \bar{Y}_{C,1}$$

Different starting levels?

A poorer region might still look worse after a useful policy.

What we need

The treated region's outcome *after* the policy, *without* the policy.

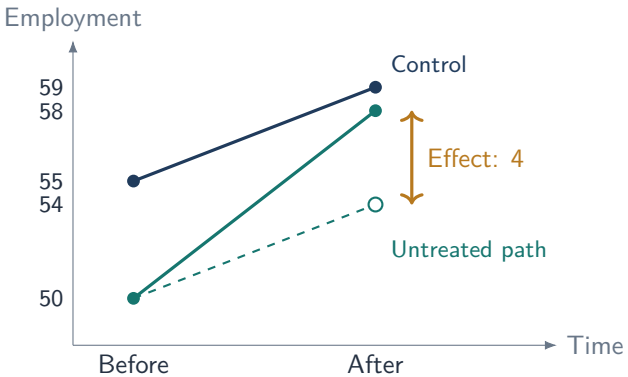
Calculate Before Naming the Estimator

Illustrative subsidy example: youth employment rate (%).

	Before	After	Change
Treated region	50	58	?
Control region	55	59	?

1. Calculate both changes. 2. Project the treated region's untreated outcome.
3. Calculate the effect. 4. State your assumption in one sentence.

Draw the Missing Untreated Path



Counterfactual:

$$50 + (59 - 55) = 54.$$

Effect estimate:

$$58 - 54 = 4.$$

The dashed path is assumed; we never observe it for treated units.

The Double Difference and Its Causal Target

$$\widehat{\tau}_{DiD} = \underbrace{(\bar{Y}_{T,1} - \bar{Y}_{T,0})}_{\text{treated change}} - \underbrace{(\bar{Y}_{C,1} - \bar{Y}_{C,0})}_{\text{proposed untreated change}} = 8 - 4 = 4.$$

Define $A_i = 1$ for the treated group and $Post_t = 1$ after implementation.
Policy exposure is $D_{it} = A_i Post_t$.

Estimand: ATT for the treated group, after treatment

$$ATT_1 = E[Y_{i1}(1) - Y_{i1}(0) \mid A_i = 1].$$

The sample double difference is an **estimate**; its causal interpretation needs assumptions.

Roadmap

Build the Missing Counterfactual

Assumptions, Regression and Inference

Read Effects in Event Time

Staggered Adoption: Who Is the Control?

Build and Aggregate Cohort-Specific Effects

Connect the Design to the General Method

Parallel Trends: The Assumption Behind the Dashed Line

In plain language

Without the policy, the treated and control groups would have experienced the same average change in outcomes.

$$E[Y_{i1}(0) - Y_{i0}(0) \mid A_i = 1] = E[Y_{i1}(0) - Y_{i0}(0) \mid A_i = 0].$$

- Groups may start at different levels.
- We need comparable *untreated changes*, including after treatment begins.
- Fixed differences and common shocks cancel; treated-specific shocks do not.

Why the Treated Change Is Not Yet an Effect

Add and subtract the missing untreated post-policy mean:

$$E[Y_{i1} - Y_{i0} \mid A_i = 1] = \underbrace{E[Y_{i1}(1) - Y_{i1}(0) \mid A_i = 1]}_{ATT_1} + \underbrace{E[Y_{i1}(0) - Y_{i0}(0) \mid A_i = 1]}_{\text{change that would occur without the policy}} .$$

- Consistency links observed and potential outcomes.
- No anticipation makes the pre-period outcome untreated.
- The last term is missing: DiD substitutes the control group's change.

Parallel Trends Removes the Remaining Bias Term

Write $\Delta Y_i(0) = Y_{i1}(0) - Y_{i0}(0)$. With untreated controls:

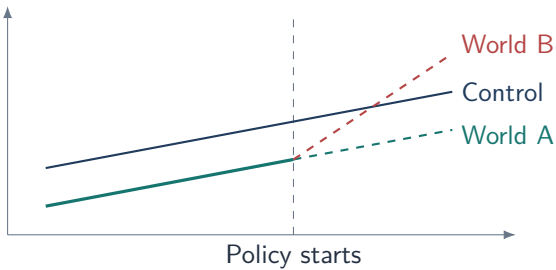
$$DiD = ATT_1 + \underbrace{E[\Delta Y_i(0) \mid A_i = 1] - E[\Delta Y_i(0) \mid A_i = 0]}_{\text{bias from different untreated trends}}.$$

Identification

Parallel trends sets the bracketed difference to zero: $DiD = ATT_1$.
In a sample, $\widehat{\tau}_{DiD}$ estimates this population contrast.

Check: if the treated region would have gained 2 extra points anyway, how does a DiD estimate of 4 relate to its true ATT?

Two Worlds Can Have the Same Pre-Trends



World A: untreated trends remain parallel.

World B: a new factory opens only in the treated region.

Both worlds fit the same pre-policy evidence.

Diagnostics do not establish identification

Flat pre-trends cannot rule out a treated-specific shock after implementation.

Three More Questions Before Calling It Causal

Condition	Subsidy example	Useful evidence
No anticipation	Firms do not react before the chosen start date	Announcement and implementation dates
No spillovers to controls	Jobs are not merely displaced from the control region	Commuting zones and neighbouring areas
Stable comparison	Migration or attrition does not change who is measured	Follow the same units; inspect composition

Design audit

Choose the most plausible problem. Propose a check that could reveal it.

The Regression Is the Same Four-Cell Comparison

$$Y_{it} = \alpha + \beta A_i + \lambda Post_t + \tau \underbrace{(A_i \times Post_t)}_{D_{it}} + u_{it}.$$

Fitted group mean	Before	After
Control	$\widehat{\alpha}$	$\widehat{\alpha} + \widehat{\lambda}$
Treated	$\widehat{\alpha} + \widehat{\beta}$	$\widehat{\alpha} + \widehat{\beta} + \widehat{\lambda} + \widehat{\tau}$

Key takeaway

$\widehat{\tau}$ equals the hand-calculated double difference.
The equation estimates the comparison; parallel trends gives it a causal interpretation.

Card and Krueger: Read the Result with Its Design

Outcome: full-time-equivalent (FTE) employment per restaurant.

Published contrast, Table 3	FTE employees per restaurant
NJ employment change minus PA employment change	$+2.76$
Reported standard error	1.36

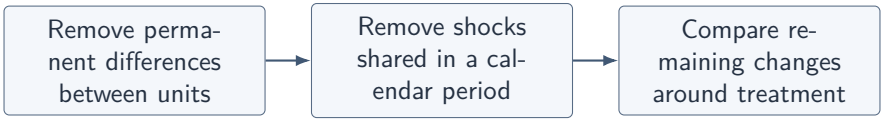
- The estimate is positive; the study does not find the predicted job loss.
- A causal reading requires comparable untreated state trends and no anticipation.
- Two survey waves cannot establish pre-trends; measurement and state shocks matter.

Scope: these restaurants and this wage increase in 1992; not every labour market.

Card and Krueger (1994), Table 3, row 3, column (iii), and p. 778.

Two-Way Fixed Effects: What Is Removed?

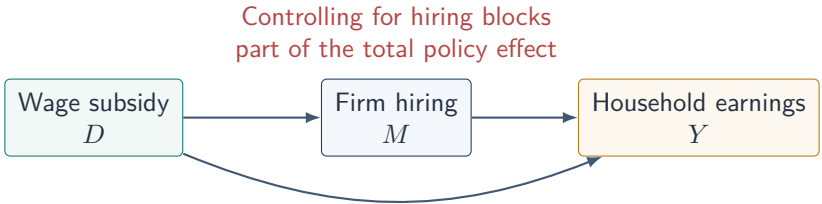
$$Y_{it} = \underbrace{\alpha_i}_{\text{unit effects}} + \underbrace{\lambda_t}_{\text{time effects}} + \tau D_{it} + u_{it}.$$



What remains

A shock affecting only treated units when the policy starts is still a threat.
With staggered dates and changing effects, we must also audit the comparisons.

Should We Control for a Policy Mechanism?



- Pre-treatment characteristics can help define a credible conditional comparison.
- A variable affected by treatment may be a mechanism, not a confounder.
- Adding controls does not establish parallel trends.

Inference: Many Observations Can Mean Few Policy Shocks

Region A: one policy decision Many workers, repeated years	Region B: one policy decision Many workers, repeated years

- Outcomes within a region and over time can be correlated.
- Standard errors should respect the assignment and dependence structure.
- Few independent clusters make inference difficult, even with many workers.

Bertrand, Duflo and Mullainathan show severe over-rejection with conventional standard errors in placebo-law exercises.

Bertrand, Duflo and Mullainathan (2004), QJE; author manuscript.

R Mini-Lab: Recover the Same Contrast Two Ways

Question: what is the subsidy effect in the simulated two-period panel?

```
source("r_exercises/week_05_did_twfe.R")
m <- with(panel, tapply(Y, list(treated, post), mean))
did <- (m["1","1"] - m["1","0"]) -
       (m["0","1"] - m["0","0"])
fit <- lm(Y ~ treated * post, data = panel)
c(by_hand = did, regression = coef(fit)["treated:post"])
```

Eight minutes: calculate, then qualify

Explain the equality. State the assumption needed for a causal interpretation.
The simulated Y is an outcome score: report **score units**, not percentage points.

Ten-Minute Break

Return at minute 90

We can defend one before–after contrast.
Next: what changes when effects evolve and units adopt in different years?

Roadmap

Build the Missing Counterfactual

Assumptions, Regression and Inference

Read Effects in Event Time

Staggered Adoption: Who Is the Control?

Build and Aggregate Cohort-Specific Effects

Connect the Design to the General Method

Align Different Calendar Dates Around Treatment

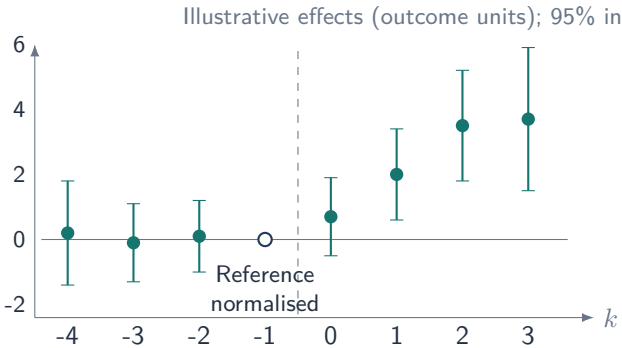
Define G_i as the first treatment date, and $k = t - G_i$ as event time.

	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

One-minute conversion

Which year is $k = 1$ for each treated cohort? Why has G replaced the earlier group indicator A?

An Event-Study Graph: Describe Before Explaining



1. Find the reference: $k = -1$.
2. Read the pre-policy evidence.
3. Describe when gains appear.
4. Inspect uncertainty and support.

The Event-Study Equation Behind the Dots

For a common adoption date g, and untreated controls:

$$Y_{it} = \alpha_i + \lambda_t + \sum_{k \in \mathcal{K},\ k \neq -1} \beta_k \underbrace{1[A_i = 1,\ t - g = k]}_{\text{treated group at event time } k} + u_{it}.$$

- Omit $k = -1$: the period just before adoption is the reference.
- Leads ($k < 0$) describe differential pre-treatment changes.
- Under the design assumptions, lags ($k \geq 0$) describe treated effects over time.

With staggered dates

Replacing g by each unit's G_i is common, but the resulting TWFE dots can mix effects. We will inspect why.

Pre-Trends, Anticipation and Uncertainty

What you see	What you can say
Pre-coefficients near zero, narrow intervals	Evidence against large pre-treatment differences; no proof about the missing post-policy trend
Pre-coefficients near zero, wide intervals	The design may be unable to detect meaningful departures
Movement before implementation	Possible anticipation, different trends, or a misspecified estimator
Long-lag intervals widen	Check how many cohorts and controls remain

Roadmap

Build the Missing Counterfactual

Assumptions, Regression and Inference

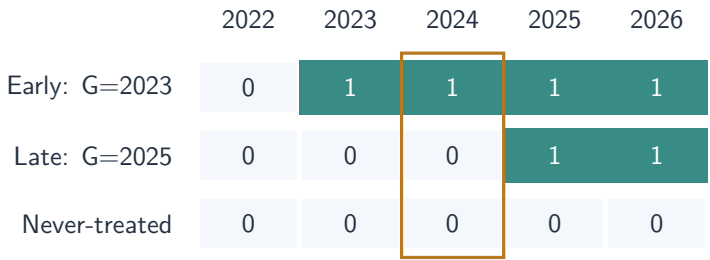
Read Effects in Event Time

Staggered Adoption: Who Is the Control?

Build and Aggregate Cohort-Specific Effects

Connect the Design to the General Method

An Adoption Heatmap Makes Control Status Visible



Inspect 2024, then 2025

Who is treated, not-yet-treated, and never-treated? Which units can supply an untreated trend?

Why an Already-Treated Control Can Subtract a Real Effect

Schematic example: no untreated trend; treatment effects grow with exposure.

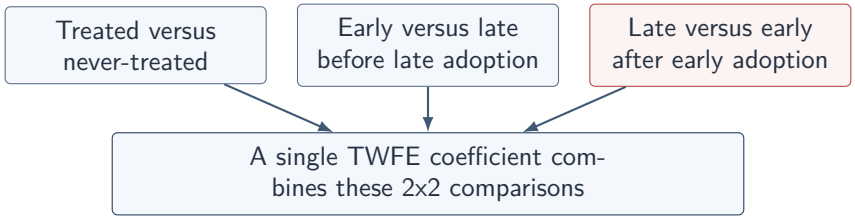
Outcome	Before late adoption	After late adoption	Change
Late adopters	0 (untreated)	2 (newly treated)	$+2$
Early adopters	2 (already treated)	6 (longer exposure)	$+4$
Never-treated	0	0	0

$$DiD_{\text{late vs early}} = 2 - 4 = -2, \qquad DiD_{\text{late vs never}} = 2 - 0 = 2.$$

The wrong counterfactual trend

Early adopters' change includes a growing treatment effect, not just an untreated trend.

Goodman-Bacon: One Coefficient, Several Comparisons



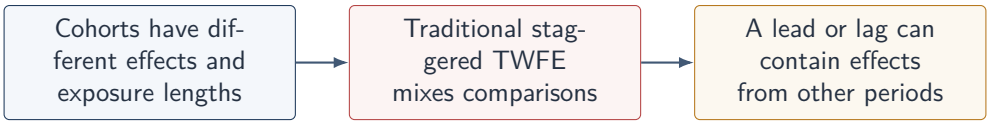
Goodman-Bacon's decomposition explains why **comparison composition** matters when treatment effects evolve.

A precise distinction

The basic decomposition has non-negative weights on the 2x2 DiDs.
Those DiDs can still subtract changing effects in already-treated units.

Goodman-Bacon (2021), Journal of Econometrics.

Sun and Abraham: Even a Lead Can Mix Post-Treatment Effects



- Apparent pre-trends can arise from treatment-effect heterogeneity.
- A good graph needs an estimator with explicit comparison groups.
- A modern estimator still requires substantive parallel trends and no anticipation.

Sun and Abraham (2021), Journal of Econometrics; author manuscript.

Roadmap

Build the Missing Counterfactual

Assumptions, Regression and Inference

Read Effects in Event Time

Staggered Adoption: Who Is the Control?

Build and Aggregate Cohort-Specific Effects

Connect the Design to the General Method

An Effect for a Cohort at a Date

- $Y_{it}(g)$: outcome at date t if treatment first starts at date g.
- $Y_{it}(\infty)$: outcome at t if treatment never starts.
- $G_i = g$: the unit belongs to the cohort first treated at g.

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

Translate the notation

What does $ATT(2023, 2025)$ describe? What is the event time?
Whose counterfactual outcome is missing?

A Clean Control Label Still Needs a Trend Assumption

For cohort g, date $t \geq g$, and baseline $b = g - 1$, choose $C_{g,t} = \{i : G_i > t\}$, or use never-treated units only.

Cohort-specific parallel untreated changes

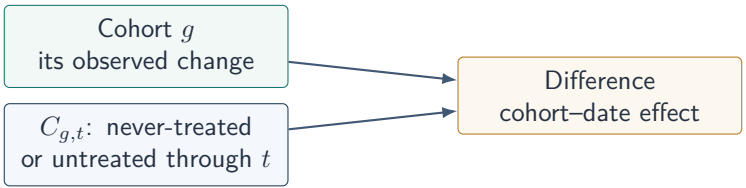
$$E[Y_{it}(\infty) - Y_{ib}(\infty) \mid G_i = g] = E[Y_{it}(\infty) - Y_{ib}(\infty) \mid i \in C_{g,t}].$$

- No anticipation: $Y_{is}(g) = Y_{is}(\infty)$ for $s < g$.
- Controls must remain unaffected at both dates, including by spillovers.
- Observe both changes in comparable populations; require a non-empty control set.

Keep the Familiar Double Difference, Choose Clean Controls

For a fixed cohort g, post date t, and base period $g-1$:

$$ATT(g,t) = E[Y_{it} - Y_{i,g-1} \mid G_i = g] \\ - E[Y_{it} - Y_{i,g-1} \mid i \in C_{g,t}].$$



Requires parallel untreated changes *for this cohort and control set*, no anticipation and no spillovers. Conditional versions also need overlap in X.

A Long-Run Effect May Describe Only Early Adopters

Suppose observation ends in 2026.

Supported event time	$k = 0$	$k = 1$	$k = 2$	$k = 3$
2023 cohort	Yes	Yes	Yes	Yes
2025 cohort	Yes	Yes	—	—
Number of cohorts	2	2	1	1

A changing population can resemble changing effects

Comparing the average at $k = 0$ with the average at $k = 3$ may change both exposure length and the cohorts represented.

Report support; consider a common cohort set when comparing dynamics.

Aggregation Is a Policy Choice, Not Just a Software Option

An average at event time k can be written:

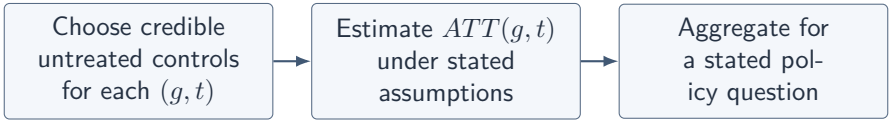
$$\theta_k = \sum_{g \in \mathcal{G}_k} w_{g,k}\, ATT(g, g+k), \qquad w_{g,k} \geq 0, \quad \sum_{g \in \mathcal{G}_k} w_{g,k} = 1.$$

- $\mathcal{G}_k$: cohorts with observed outcomes *and valid controls* at that horizon.
- Weights may reflect cohort sizes; state them explicitly.
- Other averages answer cohort, calendar-year or overall policy questions.

A one-minute calculation

At $k = 1$, one cohort has effect 2 and weight 0.75; another has effect 6 and weight 0.25. What does the average describe?

Callaway and Sant'Anna: Make the Building Blocks Visible



Empirical application: minimum wages and teen employment, 2001–2007.
The paper extends the same policy question to multiple periods and adoption cohorts.

Key takeaway

Modern DiD keeps the comparison and target visible.
It does not make an implausible control trend credible.

Callaway and Sant'Anna (2021), Journal of Econometrics; author manuscript.

Design Decision

Cities adopt a subsidy in 2023 and 2025; some never adopt. The sample ends in 2026.

Recommend an analysis in three sentences

1. For the 2025 cohort in 2026, choose the control set and base period.
2. State the trend assumption and one plausible threat.
3. Explain why an event-time-three effect cannot describe both treated cohorts.

What your recommendation must distinguish

Identification: why the comparison is causal. Estimation: which contrasts to calculate. Inference: how uncertain the estimates are.

Roadmap

Build the Missing Counterfactual

Assumptions, Regression and Inference

Read Effects in Event Time

Staggered Adoption: Who Is the Control?

Build and Aggregate Cohort-Specific Effects

Connect the Design to the General Method

From Counterfactual Paths to a Regression Method

$$\underbrace{Y_{it}(0) = \alpha_i + \lambda_t + u_{it}}_{\text{illustrative additive model for untreated outcomes}} \qquad \Longrightarrow \qquad Y_{it} = \alpha_i + \lambda_t + \tau D_{it} + u_{it}$$

for this illustration with a constant treatment effect τ.



The simple equation is useful. When effects vary by cohort and exposure, keep $ATT(g,t)$ explicit.
Untreated residual changes must remain comparable between treated and control groups.

IV and DiD: Same Questions, Different Sources of Variation

	IV (Session 4)	DiD and event studies (today)
Variation	Instrument-induced treatment changes	Policy timing relative to valid controls
Assumption	Independence, exclusion, relevance; monotonicity for LATE	Parallel untreated trends, no anticipation and appropriate comparison conditions
Target	LATE for instrument-induced recipients	ATT for treated groups, dates and exposure horizons
Estimation	Wald or 2SLS	Double differences, FE in simple designs; explicit group–time effects

Exit Ticket and the Next Policy Design

Three questions to take to every paper

1. Where does the identifying variation come from?
2. What assumption makes the comparison causal?
3. Which population, period and treatment exposure does the estimate describe?

Next class: institutional thresholds and synthesis

What if a policy is assigned by a score cutoff rather than a change over time?
RDD replaces the parallel-trend argument with a **local continuity argument**.
We will then compare what these different designs let policy-makers learn.

Appendix: Practice After This Combined Class

- DiD reasoning: `homework/session_05_homework.md`.
- Dynamics and controls: `homework/session_06_homework.md`.
- R starters: `r_exercises/week_05_did_twfe.R` and `r_exercises/week_06_event_study_staggered.R`.

Paths start at the course root.

A compact 55–70-minute route

Session 05: conceptual questions 1, 3 and 4; Session 06: conceptual question 3 and the applied cohort-grid exercise. Add the Week 05 R contrast and a five-minute review.

Appendix: Conditional Trends Require Conditional Comparisons

For baseline covariates X, assume equal untreated changes at each x:

$$E[\Delta Y_i(0) \mid A_i = 1, X_i = x] = E[\Delta Y_i(0) \mid A_i = 0, X_i = x].$$

Define $m_0(x) = E[\Delta Y_i \mid A_i = 0, X_i = x]$. Then:

$$ATT_1 = E[\Delta Y_i - m_0(X_i) \mid A_i = 1].$$

- Average the control trend over the **treated** covariate distribution.
- Need $\Pr(A_i = 0 \mid X_i = x) > 0$ on that treated support.
- A raw control mean or an arbitrary additive X term need not implement this.

Appendix: Two Different Weighting Statements

Basic Goodman-Bacon decomposition:

$$\widehat{\tau}_{TWFE} = \sum_j s_j \widehat{DiD}_j, \qquad s_j \geq 0, \quad \sum_j s_j = 1.$$

Some 2x2 DiDs use already-treated controls and subtract changing effects.

A decomposition into underlying group–time effects:
The implied weights on those causal effects can be negative.

These are different objects. Neither justifies interpreting arbitrary staggered TWFE as a transparent overall ATT.

Goodman-Bacon (2021), Journal of Econometrics.

Appendix: Cohort-Specific Event-Study Regression

A cohort-interacted representation is:

$$Y_{it} = \alpha_i + \lambda_t + \sum_g \sum_{k \neq -1} \delta_{g,k} 1[G_i = g] 1[t - g = k] + u_{it}.$$

- Choose valid untreated comparison cohorts and a supported event window.
- Estimate cohort-specific dynamic contrasts.
- Aggregate with stated non-negative cohort weights.
- Merely adding interactions does not validate the identifying assumptions.

Sun and Abraham (2021), Journal of Econometrics; author manuscript.

Appendix: A Graph of Intervals Is Not a Joint Test

- A pointwise 95% interval covers one specified horizon with its stated probability.
- A simultaneous band concerns the whole chosen set of coefficients.
- Several individual intervals crossing zero do not establish that all leads are zero.
- Cluster or resample at a level justified by policy assignment and dependence.

Before testing

Use an estimator whose comparisons identify the intended event-time effects.
Better standard errors do not remove contamination from already-treated controls.

References: Empirical Design and Inference

- Card and Krueger (1994). "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review*, 84(4), 772–793.
  Author-hosted paper.
- Bertrand, Duflo and Mullainathan (2004). "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119(1), 249–275.
  Author manuscript.

References: Why Modern DiD Became Necessary

- Goodman-Bacon (2021). "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*, 225(2), 254–277.
  Published paper.
- Sun and Abraham (2021). "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*, 225(2), 175–199.
  Author manuscript.
- Callaway and Sant'Anna (2021). "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics*, 225(2), 200–230.
  Author manuscript.