

Modern Difference-in-Differences

PP5001 · Week 10 · Martinmas 2026

Professor David A. Jaeger

Recall: Two-Way Fixed Effects

$$Y_{it} = \alpha_i + \lambda_t + \delta D_{it} + u_{it}.$$

- α_i : time-invariant differences between units.
- λ_t : common changes in calendar time.
- D_{it} : whether unit i is treated in period t .

With different adoption dates, **which comparisons determine $\hat{\delta}$?**

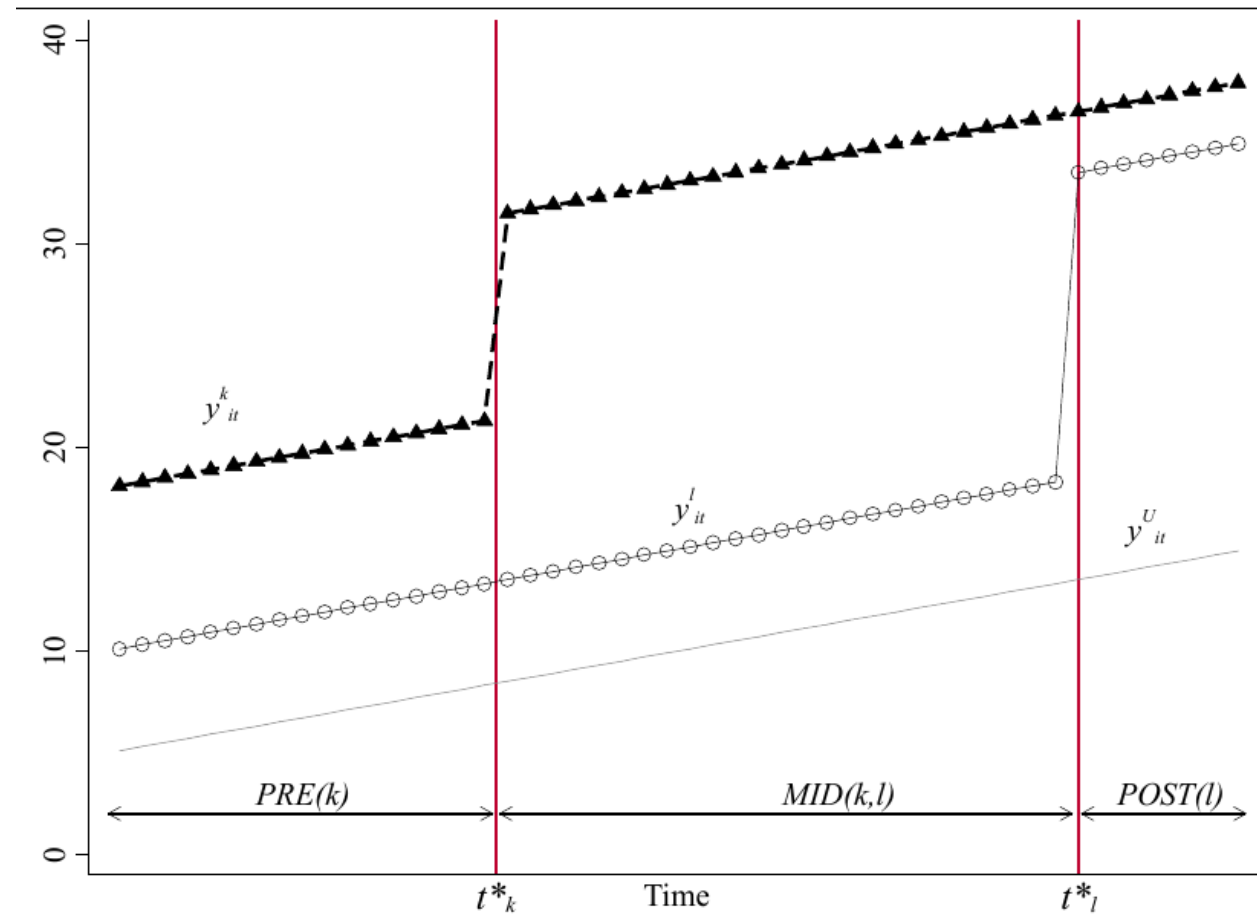
Multiple Treatment Times

Policies often begin in different places at different times:

- Minimum-wage increases.
- Health-insurance expansions.
- School reforms.
- Environmental regulations.

A **cohort** is a group of units first treated in the same calendar period.

Timing Groups



Timing groups, from the original lecture slides.

For discussion: Identify the early-treated, late-treated and never-treated groups. When can each group supply untreated observations?

Four Comparisons

With early, late and never-treated groups, TWFE combines:

1. Early treated versus never treated.
2. Late treated versus never treated.
3. Early treated versus late treated, **before the late group adopts.**
4. Late treated versus early treated, **after the early group has adopted.**

In comparison 4, the comparison group's outcomes already include treatment effects.

The Goodman-Bacon Decomposition

For the basic TWFE regression in a balanced panel,

$$\hat{\delta}^{TWFE} = \sum_j s_j \hat{\delta}_j^{2 \times 2}, \quad s_j \geq 0, \quad \sum_j s_j = 1.$$

- Each component is a two-group, two-period comparison.
- Timing and treatment variation determine the weights.
- Both not-yet-treated and already-treated groups can serve as controls.

When a Common Effect Is Adequate

Suppose untreated outcomes follow parallel trends and there is no anticipation.

If treatment has the **same constant effect** for every treated cohort and period, an already-treated group's treatment effect cancels when its outcome is differenced.

The difficulty arises when that effect changes over the comparison period.

An Illustration

Suppose every group's untreated outcome would remain at 10.

Group	Period 0	Period 1	Period 2
Early treated	10	12	16
Late treated	10	10	12
Never treated	10	10	10

Early adopters have an effect of 2 initially and 6 later. Late adopters have an effect of 2 when they adopt.

The Same Illustration Graphically

For discussion: From period 1 to period 2, whose observed outcome changes most? What would happen if we used that group as the control for late adopters?

Using an Untreated Comparison

From period 0 to period 1:

$$\hat{\delta}_{Early,Late} = (12 - 10) - (10 - 10) = 2.$$

From period 1 to period 2, using never-treated units:

$$\hat{\delta}_{Late,Never} = (12 - 10) - (10 - 10) = 2.$$

Each calculation recovers the newly treated group's effect.

Using an Already-Treated Comparison

Compare late with early adopters from period 1 to period 2:

$$\hat{\delta}_{Late,Early} = (12 - 10) - (16 - 12) = -2.$$

The late group's treatment effect is **+2**, but this comparison subtracts the **+4 change in the early group's treatment effect**.

| For discussion: Why does a negative comparison arise even though every treatment effect is positive?

What Is Being Subtracted?

Under parallel trends in untreated outcomes,

$$DiD_{Late,Early} = \text{late group's new effect} - \text{change in early group's effect.}$$

Growing effects in the early cohort reduce this component of TWFE.

Weights on Treatment Effects

The weights on the component 2×2 comparisons are positive.

But an already-treated comparison **subtracts changes in treatment effects**. When TWFE is expressed in terms of underlying cohort-period effects, some of those effects can receive negative weight.

A TWFE coefficient can consequently lie outside the range of the underlying treatment effects.

Event-Study Coefficients

In a conventional TWFE event study with staggered adoption:

- Units at different event times are observed in the same calendar period.
- Treatment effects can differ across cohorts and exposure lengths.
- An estimated lead or lag can combine effects from other event times.

Even an apparent pre-treatment coefficient can reflect post-treatment effects elsewhere.

Make the Comparisons Explicit

The approach we will use has three steps:

1. Define an untreated comparison group.
2. Estimate effects separately for each adoption cohort and calendar period.
3. Average those effects to answer a stated policy question.

Callaway and Sant'Anna (2021) organise their method around these steps.

A Group-Time Treatment Effect

Let $G_i = g$ mean that unit i first receives treatment in period g .

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

- Y_{it}^g : outcome at t if treatment first begins at g .
- Y_{it}^∞ : outcome at t if the unit remains untreated.
- $ATT(g, t)$: effect at t for the cohort first treated at g .

Reading the Two Dates

$$ATT(2004, 2006)$$

means the effect **in calendar year 2006**, for units **first treated in 2004**.

Their event time is

$$k = t - g = 2006 - 2004 = 2.$$

For discussion: Which effect would describe the 2006 cohort in its first treatment year? Which would describe that same cohort one year later?

Identify the Counterfactual

Use $g - 1$, the last pre-treatment period, as the baseline.

With never-treated units as controls, assume

$$E \left[Y_{it}^{\infty} - Y_{i,g-1}^{\infty} \mid G_i = g \right] = E \left[Y_{it}^{\infty} - Y_{i,g-1}^{\infty} \mid G_i = \infty \right] .$$

Also assume no anticipation at the baseline.

Back to a Two-by-Two Difference

Under those assumptions,

$$\begin{aligned} ATT(g, t) = & E [Y_{it} - Y_{i,g-1} \mid G_i = g] \\ & - E [Y_{it} - Y_{i,g-1} \mid G_i = \infty] . \end{aligned}$$

The sample estimator replaces each expectation with its corresponding sample mean.

We already know how to calculate this comparison.

Choosing the Comparison Group

Never treated: remain untreated throughout the observation period.

Not yet treated: remain untreated through the outcome period t .

For the latter, use units with $G_i > t$, including never-treated units where available, and maintain the corresponding parallel-trends and no-anticipation assumptions.

Conditional Parallel Trends

Sometimes parallel trends is plausible only among units with similar baseline characteristics X_i .

Estimate comparisons conditional on those characteristics, then average over the treated cohort's characteristics.

Overlap is needed: the comparison group must contain suitable counterparts for the treated cohort.

Which Average Do We Want?

Once we have $ATT(g, t)$, we can average:

- **Within cohorts:** how did each adoption cohort fare?
- **Within calendar years:** what was the effect among treated units in a particular year?
- **At a common event time:** how does the effect evolve with exposure?

These averages can differ because they combine different effects with different weights.

An Event-Time Average

At event time $k \geq 0$, combine available cohorts at $t = g + k$:

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

$\mathcal{G}_k$ contains cohorts observed at event time k with a valid comparison group. Cohort-size weights give larger cohorts more influence.

Changing Cohort Composition

Suppose data end in 2007.

First treated	Last observed event time
2004	3
2006	1
2007	0

An event-time graph can change because effects evolve **and** because the contributing cohorts change.

A balanced-cohort graph keeps the same cohorts at every displayed exposure length.

Minimum Wages Revisited

Callaway and Sant'Anna study county-level teenage employment during 2001–2007.

- The federal minimum wage stays at \$5.15 over the study window.
- Treated states raise their minimum wage above that level.
- Cohorts are defined by the first increase.
- The main comparison group is counties in states that remain at the federal minimum.

For discussion: Explain the policy, data and source of identifying variation. What differs from Card and Krueger's comparison?

Table 2: The Comparison Groups

Table 2

Summary statistics for main dataset.

Source: Quarterly Workforce Indicators and 2000 County Data Book.

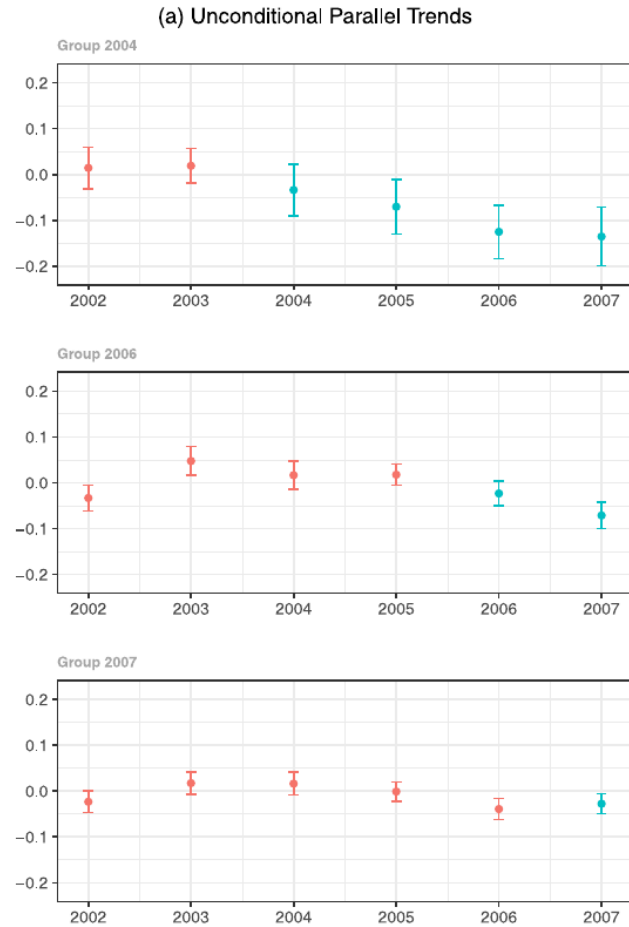
	Treated counties	Untreated counties	Diff.	P-val on Diff.
Midwest	0.59	0.34	0.25	0.00
South	0.27	0.59	−0.32	0.00
West	0.14	0.07	0.07	0.00
Population (1000s)	94.32	53.43	40.89	0.00
White	0.89	0.83	0.06	0.00
HS Graduates	0.59	0.55	0.04	0.00
Poverty Rate	0.13	0.16	−0.03	0.00
Median Inc. (1000s)	33.91	31.89	2.02	0.00

Notes: Summary statistics for counties located in states that raised their minimum wage between Q2 of 2003 and Q1 of 2007 (treated) and states whose minimum wage was effectively set at the federal minimum wage for the entire period (untreated). The sample consists of 2284 counties.

Callaway and Sant'Anna (2021), Table 2, p. 217.

For discussion: Describe the differences between treated and untreated counties. What do these differences imply for the parallel-trends argument and the choice of controls?

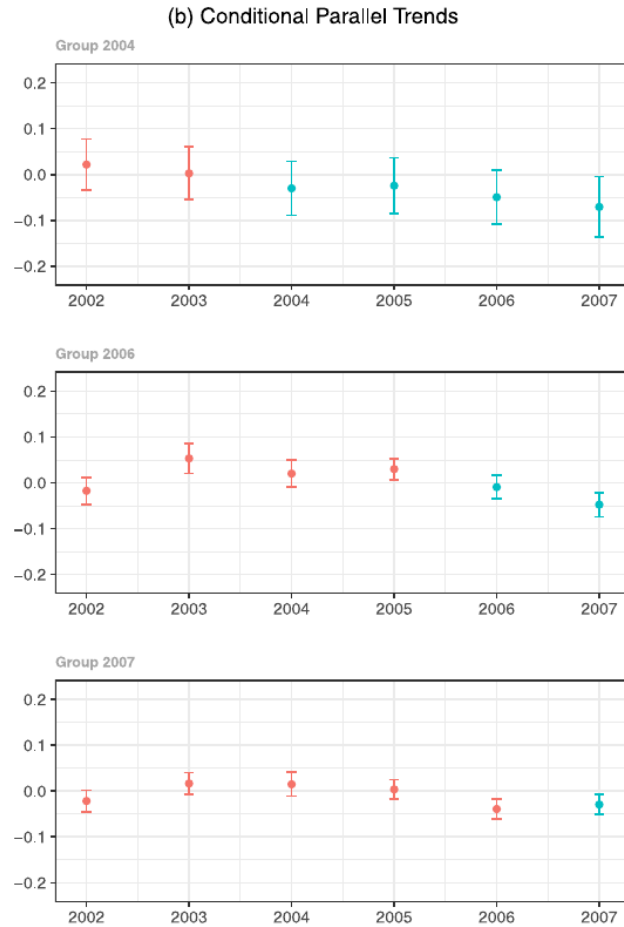
Figure 1(a): Unconditional Parallel Trends



Callaway and Sant'Anna (2021), Figure 1(a), p. 218.

For discussion: Identify the cohorts and axes. Interpret one post-treatment estimate. What concerns arise from the pre-treatment evidence?

Figure 1(b): Conditional Parallel Trends



Callaway and Sant'Anna (2021), Figure 1(b), p. 218.

For discussion: What changes after adjustment for baseline characteristics? How convincing is the identifying assumption now?

Table 3(a): Different Averages

Table 3

Minimum wage aggregated treatment effect estimates.

(a) Unconditional parallel trends

	Partially aggregated				Single parameters
TWFE					−0.037 (0.006)
Simple weighted average					−0.052 (0.006)
Group-specific effects	<u>g=2004</u> −0.091 (0.019)	<u>g=2006</u> −0.047 (0.008)	<u>g=2007</u> −0.028 (0.007)		
Event study	<u>e=0</u> −0.027 (0.006)	<u>e=1</u> −0.071 (0.009)	<u>e=2</u> −0.125 (0.021)	<u>e=3</u> −0.136 (0.023)	−0.090 (0.013)
Calendar time effects	<u>t=2004</u> −0.034 (0.019)	<u>t=2005</u> −0.071 (0.02)	<u>t=2006</u> −0.055 (0.009)	<u>t=2007</u> −0.050 (0.006)	−0.052 (0.013)
Event study w/ Balanced groups	<u>e=0</u> −0.027 (0.009)	<u>e=1</u> −0.071 (0.009)			

Callaway and Sant'Anna (2021), Table 3(a), p. 219.

For discussion: Focus on TWFE and the group-specific effects. Compare the TWFE estimate with the overall group-specific effect in the rightmost column. How similar are they?

Table 3(b): With Covariate Adjustment

(b) Conditional parallel trends					
	Partially aggregated				Single parameters
TWFE					−0.008 (0.006)
Simple weighted average					−0.033 (0.007)
Group-specific effects	<u>g=2004</u> −0.044 (0.020)	<u>g=2006</u> −0.029 (0.008)	<u>g=2007</u> −0.029 (0.008)	−0.031 (0.007)	
Event study	<u>e=0</u> −0.024 (0.006)	<u>e=1</u> −0.041 (0.009)	<u>e=2</u> −0.050 (0.022)	<u>e=3</u> −0.071 (0.026)	−0.046 (0.013)
Calendar time effects	<u>t=2004</u> −0.030 (0.022)	<u>t=2005</u> −0.025 (0.021)	<u>t=2006</u> −0.030 (0.009)	<u>t=2007</u> −0.049 (0.007)	−0.033 (0.012)
Event study w/ Balanced groups	<u>e=0</u> −0.016 (0.010)	<u>e=1</u> −0.041 (0.009)			
					−0.028 (0.008)

Callaway and Sant'Anna (2021), Table 3(b), p. 219.

For discussion: Compare TWFE with the overall group-specific effect in the rightmost column. What changes relative to panel (a) after covariate adjustment? Would your policy conclusion depend on the estimator?

Assessing the Policy Evidence

- Are untreated employment trends plausibly comparable?
- What do the pre-treatment estimates suggest?
- Minimum-wage increases differ in size. What does the binary treatment represent?

Choosing and Reporting an Estimate

1. State the policy effect and population of interest.
2. Identify the adoption cohorts and untreated comparisons.
3. Explain parallel trends, anticipation and possible spillovers.
4. Report how cohort-time effects are averaged.
5. Show uncertainty and the evidence used to assess the design.

For discussion: Which of these choices would matter most if you were advising a government about another minimum-wage increase?

What We Take Forward

The basic DiD logic remains: use an untreated group's change to construct a counterfactual.

With staggered adoption, make the comparison separately for each cohort and period, then average deliberately.

Next week: constructing a comparison for a single treated country or region using synthetic control.

References

- Callaway, B., and P. H. C. Sant'Anna (2021). [Difference-in-Differences with Multiple Time Periods](#). *Journal of Econometrics*, 225(2), 200–230.
- Goodman-Bacon, A. (2021). [Difference-in-Differences with Variation in Treatment Timing](#). *Journal of Econometrics*, 225(2), 254–277.
- Sun, L., and S. Abraham (2021). [Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects](#). *Journal of Econometrics*, 225(2), 175–199.