Difference-in-Differences

Econometric Modelling and Forecasting II
The SEACEN Centre

Tanisa Tawichsri

Puey Ungphakorn Institute for Economic Research (PIER)
Bank of Thailand

May 14, 2026

PIER

Overview

Quasi-Experimental Methods Are on the Rise

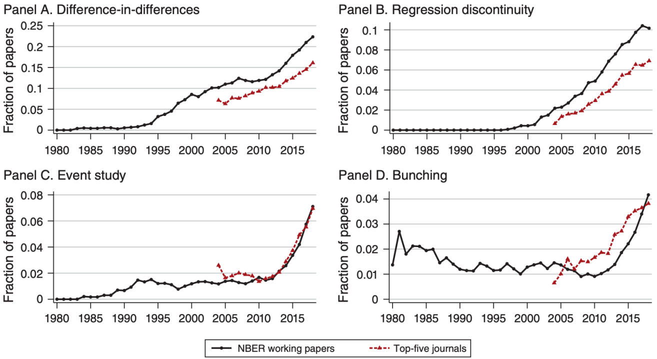


FIGURE 4. QUASI-EXPERIMENTAL METHODS

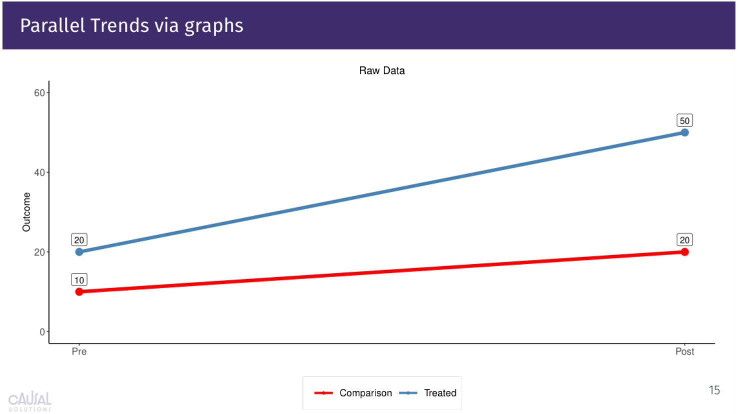
Notes: This figure shows the fraction of papers referring to each type of quasi-experimental approach. See Table A.I for a list of terms. The series show five-year moving averages.

DiD appears in >20% of NBER papers — the **dominant quasi-experimental method** in applied economics. Source: Currie, Kleven & Zwiers (2020), AEA P&P.

Recap: The Problem with Simple Comparisons

- In Lecture 1, we saw that OLS is biased when the treatment is endogenous
- The same problem arises in policy evaluation: we cannot simply compare outcomes before and after a policy change, or between countries that adopted a policy and those that did not
- **Two naive approaches — and why they fail:**
  - **Before-After (within unit):** Changes over time may reflect common trends, not the policy
    - Did NJ employment rise after the minimum wage hike — or did all US employment rise?
  - **Cross-section (between units):** Treated and untreated units may differ in ways that matter
    - Countries that adopt inflation targeting may already be different from those that don't
- **Difference-in-Differences (DiD):** Combine both comparisons to cancel out what we do not need

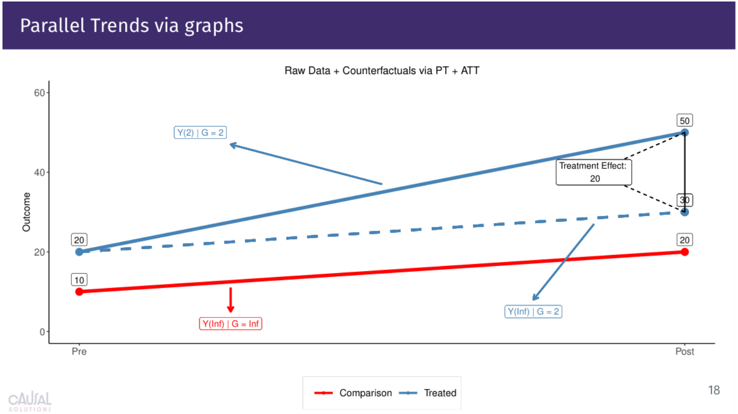
What we observe: two groups, two periods — but we cannot see the counterfactual.



Source: Sant'Anna (2025), PDHP slides, p. 15.

The DiD Idea: Constructing the Counterfactual

Use the control group's trend to impute what the treated group would have done without treatment.



Source: Sant'Anna (2025), PDHP slides, p. 18.

When DiD Applies

- DiD is appropriate when:
  1. There is a clear **treatment event** with a before and after period
  2. There is a **comparison group** that did not receive the treatment
  3. The parallel trends assumption is plausible (more on this shortly)
- **Classic applications in economics and central banking:**
  - Minimum wage changes affecting some states/countries but not others
  - Inflation targeting adoption: some countries adopt, others do not
  - Macroprudential policy interventions (LTV limits, reserve requirements)
  - Capital flow regulations applied to some markets
  - Trade policy changes for specific sectors
- DiD is a natural experiment framework when the policy creates treated/control variation and **parallel trends is credible**

DiD vs. IV: Choosing Between Designs

Both DiD and IV address selection bias — but through different mechanisms.

	Difference-in-Differences	Instrumental Variables
Key variation	Policy rollout: some units treated, others not (yet)	Exogenous shifter of the endogenous variable
What you need	Panel data; a plausible control group	A valid instrument: relevant and excludable
Main assumption	Parallel trends: groups move together absent treatment	Exclusion restriction: instrument affects outcome only via treatment
Estimand	ATT — effect on the treated	LATE — effect on compliers
Typical use case	Staggered policy rollout (IT adoption, capital rules)	Single endogenous regressor (rainfall $\to$ income $\to$ conflict)

How DiD handles it: Yes, *which* units get treated is often endogenous (e.g. states with worse labour markets adopt a minimum wage later). DiD differences this out — *as long as* that selection is time-invariant. If adoption is driven by a time-varying shock (e.g. a state raises the minimum wage *because* employment is already falling), parallel trends breaks and DiD is no longer valid.

Can be combined: IV can instrument for treatment take-up in a DiD; DiD can serve as the first stage.

Canonical DiD

The 2×2 Case

Card & Krueger: Setup and the Counterfactual

- **Setting:** NJ raises the minimum wage from \$4.25 to \$5.05 in April 1992; PA does not. **Outcome:** FTE employment at fast food restaurants (Feb 1992 pre; Nov 1992 post)
- **Two groups, two periods:**
  - Treated ($i \in \mathcal{T}$): New Jersey — exposed to the wage increase in period 2
  - Control ($i \in \mathcal{C}$): Pennsylvania — never exposed
- **The counterfactual problem:** Standard theory predicts employment falls when the MW rises above market wage — but we cannot observe what NJ employment would have been *without* the hike
- **DiD solution:** Use PA as the counterfactual — same industry structure, same national trends (1992 recession recovery)
- **Key assumption:** In the absence of treatment, NJ would have followed the same trend as PA — the **parallel trends assumption**

The DiD Formula

- Define four observable quantities:

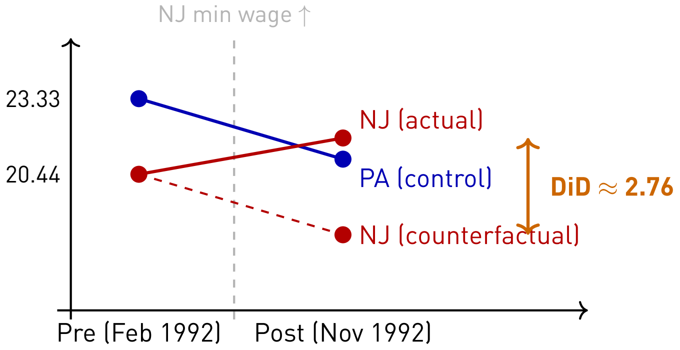
	Pre ($t = 0$)	Post ($t = 1$)
Treated (NJ)	$\bar{Y}_{T,0}$	$\bar{Y}_{T,1}$
Control (PA)	$\bar{Y}_{C,0}$	$\bar{Y}_{C,1}$

- The DiD estimator:

$$\hat{\delta}^{\text{DiD}} = \underbrace{\left(\bar{Y}_{T,1} - \bar{Y}_{T,0}\right)}_{\text{1st difference: within NJ}} - \underbrace{\left(\bar{Y}_{C,1} - \bar{Y}_{C,0}\right)}_{\text{2nd difference: within PA}}$$

- This removes:
  - Time-invariant differences between NJ and PA (e.g., state-specific wage levels)
  - Common time trends affecting both states (e.g., national employment cycle)
- What remains is the **causal effect of the treatment** (under parallel trends)

Visualising DiD: Parallel Trends



Stylised; y-axis: FTE employment per store. Source: Card & Krueger (1994).

The Parallel Trends Assumption

DiD does not require treated and control to be identical — only that they would have moved in parallel without treatment.

- Formally: $\mathbb{E}[Y^{(0)}_{T,1} - Y^{(0)}_{T,0}] = \mathbb{E}[Y^{(0)}_{C,1} - Y^{(0)}_{C,0}]$ where $Y^{(0)}$ is the potential outcome *without* treatment
- **Untestable post-treatment** — it is an assumption about the counterfactual; we can only check it indirectly via pre-trends
- **What gives it credibility:**
  - Pre-treatment trends look similar (plot the raw data)
  - Control group is contextually similar to the treated group
  - Treatment was not selected based on expected outcome trajectory
- **What undermines it:**
  - Units adopted the policy *because* they expected worse outcomes
  - Other concurrent policy changes affected only the treated group

Testing Parallel Trends in Practice

You cannot prove parallel trends — but you can make it credible or implausible.

1. **Plot raw data.** Graph outcomes for treated and control over time. Visually diverging pre-trends $\Rightarrow$ reconsider your control group.
2. **Run the event study.** Include treatment leads ($k<0$) and lags ($k\geq 0$), normalise at $k=-1$: $Y_{it} = \alpha_i + \alpha_t + \sum_{k\neq -1} \beta_k \cdot \mathbf{1}[t - t_i^* = k] + \varepsilon_{it}$. Pre-treatment $\hat{\beta}_k \approx 0$ is consistent with parallel trends.
3. **Formally test pre-trends.** Joint F-test on all $\hat{\beta}_k$ for $k < 0$. Rejection means trends were diverging before treatment.
4. **Run a placebo test.** Apply the treatment to a group theory says should be unaffected. A significant placebo effect signals a confound.
5. **Argue from context.** The statistical test supports — it does not replace — the economic argument for why parallel trends holds here.

Advanced: Passing a pre-trends test does not guarantee parallel trends holds post-treatment. For formal bounds on the ATT under plausible violations, see Rambachan & Roth (2023).

Application: Card & Krueger (1994)

- NJ raises minimum wage from \$4.25 to \$5.05 in April 1992; PA does not
- Survey of 410 fast food restaurants in NJ and PA (Feb and Nov 1992)

	Pre (Feb)	Post (Nov)	Change
New Jersey (treated)	20.44	21.03	$+0.59$
Pennsylvania (control)	23.33	21.17	-2.17
Difference-in-Differences			$+$**2.76**

FTE employment per store. Source: Card & Krueger (1994), Table 3.

- The DiD estimate is **+2.76 FTE employees per store** — employment *rose* in NJ relative to PA
- This contradicts the simple supply-and-demand prediction
- Explanation: in imperfectly competitive labour markets (monopsony), raising the minimum wage can increase both wages *and* employment

Card & Krueger: The Policy Bite

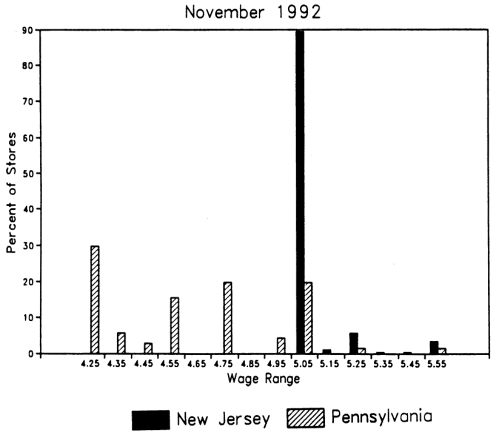


Figure 1. Distribution of Starting Wage Rates

Source: Card & Krueger (1994), Figure 1. Wage distribution, November 1992.

- By November 1992, ≈**90% of NJ stores** pay exactly $5.05 — bunched at the new floor
- PA stores remain spread across lower wages
- Treatment had real **bite**: NJ stores were genuinely forced to raise wages
- Without bite, we could not attribute employment changes to the policy

Lesson: Always verify that treatment changed the outcome mechanism — not just the treatment indicator.

DiD as a Regression

- The 2×2 DiD can be estimated as a regression:

$$Y_{it} = \alpha + \beta\, \mathbf{1}[\text{Treated}_i] + \gamma\, \mathbf{1}[\text{Post}_t] + \delta\, \underbrace{(\mathbf{1}[\text{Treated}_i] \times \mathbf{1}[\text{Post}_t])}_{\text{interaction term}} + \varepsilon_{it}$$

- **Interpreting the coefficients:**
  - α: control group outcome in the pre period (baseline)
  - β: time-invariant difference between treated and control groups
  - γ: common time trend (shared by both groups)
  - δ: **the DiD estimate** — the effect of treatment
- **Key insight:** δ identifies the treatment effect because β absorbs pre-existing level differences and γ absorbs common trends
- The regression formulation extends naturally to panel data with many units and time periods

What Each Difference Removes

DiD subtracts twice — once to remove fixed differences, once to remove common trends.

	Pre	Post	
Treated	$\alpha+\beta$	$\alpha+\beta+\gamma+\delta$	$\}\ \delta+\gamma$
Control	α	$\alpha+\gamma$	$\}\ \gamma$

$-$

DiD $= (\delta+\gamma) - \gamma = \boldsymbol{\delta}$

- **1st diff** (within treated): removes β (time-invariant gap between groups)
- **2nd diff** (subtract control change): removes γ (common trend)
- What remains: δ — the causal treatment effect

Extending to Panel Data

Two-Way Fixed Effects (TWFE)

From 2×2 to Many Units and Periods

- In practice, we have panel data: N units observed over T periods
- We need a regression that controls for:
  1. Time-invariant differences across units (e.g., NJ has a different baseline wage structure from PA)
  2. Common shocks affecting all units at the same time (e.g., national recession)
- **Solution:** include *unit fixed effects* and *time fixed effects*
- This is the **Two-Way Fixed Effects (TWFE)** estimator:

$$Y_{it} = \underbrace{\alpha_i}_{\text{unit FE}} + \underbrace{\alpha_t}_{\text{time FE}} + \delta D_{it} + \varepsilon_{it}$$

  where $D_{it} = 1$ if unit i is treated at time t, and 0 otherwise.
- δ is the DiD estimate: the average treatment effect on the treated (ATT), under parallel trends

What TWFE Controls For

- **Unit fixed effects** α_i: absorb all *time-invariant* differences between units
  - Identification comes from *within-unit* variation over time
  - Units serve as their own control: we compare each unit to its own pre-treatment baseline
- **Time fixed effects** α_t: absorb all *unit-invariant* shocks in each period
  - Common macroeconomic shocks, global trends, seasonal patterns
  - Identification comes from variation *not* shared across all units
- **What TWFE assumes:** After conditioning on unit and time FEs, parallel trends holds — the treatment indicator D_{it} is uncorrelated with ε_{it}
- **Important caveat:** TWFE also assumes the treatment effect δ is *the same* across all units and all time periods (homogeneous treatment effects). We return to why this matters shortly.

The Event Study Specification

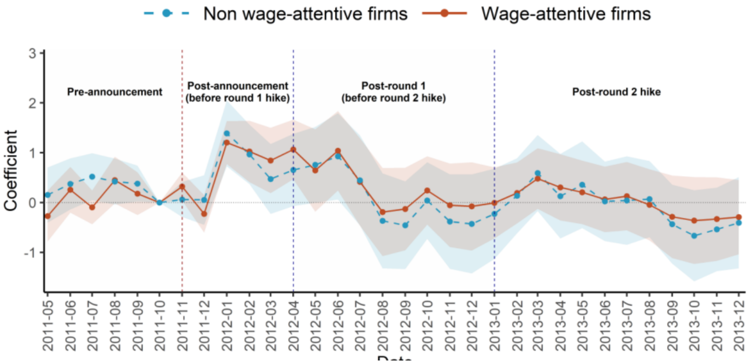
- Instead of a single treatment indicator, include **leads and lags** relative to the treatment date g_i:

$$Y_{it} = \alpha_i + \alpha_t + \sum_{k \neq -1} \beta_k \cdot \mathbf{1}[t - g_i = k] + \varepsilon_{it}$$

- $k < 0$: **pre-treatment** periods (leads); $k \geq 0$: **post-treatment** periods (lags)
- $k = -1$ is the omitted category (the period just before treatment); all effects are relative to this
- **Two things to look for in an event study plot:**
  1. **Pre-treatment coefficients** β_k for $k < 0$ should be near zero
     - If they are not, it suggests parallel trends was already violated before treatment
  2. **Post-treatment coefficients** β_k for $k \geq 0$ show the *dynamics* of the treatment effect
     - Immediate effect? Gradual build-up? Fading over time?

What Pre-Trends Testing Looks Like

Figure 7: Event-study estimates of minimum wage hike



Note: without monthly FE, treat the pattern as descriptive; with monthly FE later, effects are smaller and pre-trends flatter.

Figure 7(a), Manopimoke et al. (2026), *Firm Inflation Expectations and the Macroeconomy*, PIER WP; no monthly fixed effects.

Why TWFE Becomes Fragile

TWFE is reliable with one treatment date and constant effects — both assumptions break down with staggered adoption.

- With a *single* treatment date and homogeneous effects, TWFE works well
- **However, TWFE has two important limitations:**
  1. **Staggered adoption:** In practice, units adopt policies at different times. TWFE with staggered treatment is more complex than it appears.
  2. **Heterogeneous treatment effects:** If the effect of treatment varies across units or changes over time, TWFE does not estimate what we think it does.
- These two problems *interact*: staggered adoption combined with heterogeneous treatment effects can make TWFE estimates difficult to interpret — or even sign-wrong.
- This is the central methodological issue in the modern DiD literature. We turn to it now.

Staggered DiD

The Problem with TWFE

Real-World Policies Roll Out at Different Times

- In the Card & Krueger setting, one state changed its minimum wage at one point in time. Simple.
- In practice, policy adoption is **staggered** across units over time:
  - US states raising minimum wages in different years (1990s–2010s)
  - Countries adopting inflation targeting at different times (1990–2005)
  - Banks adopting stress-testing frameworks in different regulatory cycles
- With staggered adoption, there are **three types** of units at any point in time:
  1. **Early adopters** (already treated)
  2. **Late adopters** (not yet treated)
  3. **Never-treated** (control group)
- Researchers typically run TWFE with D_{it} as the treatment indicator. This seems reasonable. **But there is a problem.**

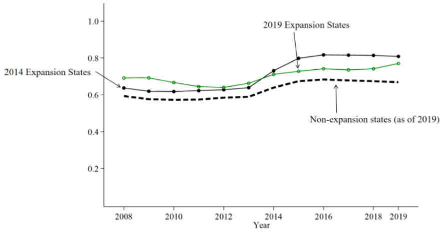
Staggered Adoption: Three Groups

Staggered policies create three types of units: early-treated, late-treated, and never-treated.

ACA Medicaid expansion:

- **Early:** 2014 expansion states
- **Late:** 2019 expansion states
- **Never:** non-expansion states
- Pre-2014 trends track closely

Figure 4: Health Insurance Rate (low-income Childless Adults Aged 25-64)



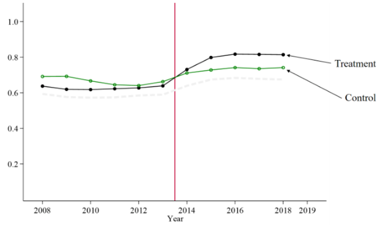
Source: Sant'Anna (2025). ACA = Affordable Care Act.

The Bad Comparison

TWFE can use already-treated groups as comparison groups for later-treated groups.

- 2014 states become the comparison group for 2019 states — but they are *already treated*
- This does *not* mean their treatment dummy turns off; it means their outcome trend is used as the benchmark
- Dashed = counterfactual TWFE cannot see
- This biases the DiD estimate

Figure 7: Health Insurance Rate (low-income Childless Adults Aged 25-64)



Source: Sant'Anna (2025).

The Decomposition Result

TWFE with staggered adoption averages clean comparisons and bad comparisons.

- Goodman-Bacon (2021): the TWFE estimator is a **weighted average** of all 2×2 DiDs:

$$\hat{\delta}^{\text{TWFE}} = \sum_{(k,\ell)} w_{k\ell}\, \hat{\delta}_{k\ell}$$

- The weights depend on group sizes and timing — not on whether comparisons are clean
- **The problem:** Some $\hat{\delta}_{k\ell}$ compare later-treated groups to already-treated groups. If the already-treated group's effect is changing, this is a **bad comparison**.
- **When does it matter most?**
  - Treatment effects vary across cohorts
  - Treatment effects change over time (build-up or fade)
  - Large weight on contaminated 2×2 comparisons

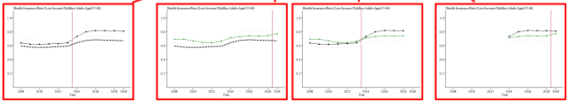
Bacon Decomposition: Visual Intuition

Bacon-Decomposition

- Main result of Goodman-Bacon (2021) is the Bacon-Decomposition:

Figure 9: Bacon-Decomposition: The 2x2 $\widehat{\beta}$

$$\hat{\beta}^{DD} = s_{kU}\boxed{\hat{\beta}^{DD}_{kU}} + s_{\ell U}\boxed{\hat{\beta}^{DD}_{\ell U}} + \left[s^{k}_{k\ell}\boxed{\hat{\beta}^{DD,k}_{k\ell}} + s^{\ell}_{k\ell}\boxed{\hat{\beta}^{DD,\ell}_{k\ell}}\right]$$



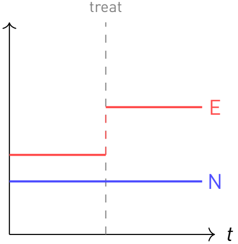
Takeaway: TWFE is not one comparison; it is a mechanical average of many 2×2 DiDs. The danger is that some components use already-treated groups as controls.

Source: Sant'Anna (2023), PDHP workshop slides, based on Goodman-Bacon (2021).

Goodman-Bacon: Which Comparisons Are Clean?

Early vs. Never

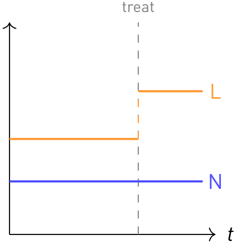
✓ **Clean**



$\hat{\delta}_{EN}$

Late vs. Never

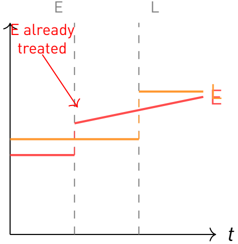
✓ **Clean**



$\hat{\delta}_{LN}$

Late vs. Early

× **Contaminated**



$\hat{\delta}_{LE}$ **biased**

N = never-treated   E = early-treated   L = late-treated $\hat{\delta}^{\text{TWFE}} = w_{EN}\hat{\delta}_{EN} + w_{LN}\hat{\delta}_{LN} + w_{LE}\hat{\delta}_{LE}$

Modern Estimators

Fixing Staggered DiD

The Fix: Clean Comparisons Only

Only compare newly-treated units to never-treated or not-yet-treated units.

- **Key concept:** Group-time average treatment effect $\text{ATT}(g, t)$
  - The average treatment effect for the **cohort** first treated in year g, measured at time t
  - This keeps effects specific to each cohort and time period — no aggregation yet
- **Aggregation:** Once we have $\text{ATT}(g, t)$ estimates, we can aggregate:
  - By cohort: average effect for each adoption-year group
  - By event time: average effect at k years after treatment (event study)
  - Overall: a single summary ATT, using researcher-chosen weights
- This framework (Callaway & Sant'Anna 2021) avoids the bad already-treated comparisons

Comparison of Modern Estimators

Estimator	**Key idea**	**Clean control**	**Event study?**
Callaway & Sant'Anna (2021)	Group-time ATT(g, t), then aggregate	Never- or not-yet-treated	Yes
Abraham & Sun (2021)	Interaction-weighted; saturates cohort$\times$time dummies	Never-treated	Yes
Stacked DiD (Cengiz et al. 2019)	Stack separate clean 2×2 datasets per cohort	Not-yet-treated	Yes
TWFE (standard)	Single regression, all variation pooled	None	Yes (biased w. het. effects)

Modern estimators differ mainly in how they select the clean comparison group and aggregate effects across cohorts.

How to Read a Modern Event Study

- Modern estimators produce event study plots like TWFE — but each $\hat{\beta}_k$ compares a cohort to a *clean* control group at event time k
- **Three things to check:**
  1. **Pre-trends** ($k < 0$): should be near zero — supports parallel trends
  2. **Treatment effect path** ($k \geq 0$): size, direction, persistence
  3. **Confidence intervals**: precision of the estimates
- **Modern vs. TWFE:** The shapes can differ substantially with heterogeneous effects. Modern event studies give a more credible picture.

Application

Inflation Targeting in ASEAN

Setup: Staggered IT Adoption in ASEAN-5

- **Question:** Does adopting flexible inflation targeting improve macroeconomic performance?
- **Outcomes:** Inflation level, inflation volatility (5-yr rolling SD), GDP growth volatility
- **Sample:** 52 developing countries (30 IT adopters), 1980–2022

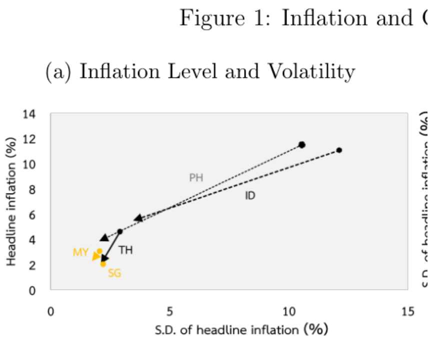
Country	IT Adoption Year	Group
Thailand	2000	Early adopter
Philippines	2002	Early adopter
Indonesia	2005	Late adopter
Malaysia	—	Non-targeter (control)
Singapore	—	Non-targeter (control)

Source: Nookhwun & Waiyawatjakorn (2024), *Asian Economic Policy Review*.

Non-ASEAN developing non-targeters also included as controls.

The Data: Inflation Before and After IT Adoption

- Each arrow traces a country from its **pre-IT** period (tail) to its **post-IT** period (head)
- **IT adopters** (black): Philippines and Indonesia start at high inflation and high volatility — large improvement post-adoption
- **Thailand** (black): starts at low levels; modest further improvement
- **Non-targeters** Malaysia, Singapore (yellow): already low; serve as the control group



Source: Nookhwun & Waiyawatjakorn (2024), Figure 1(a).

Why TWFE Is Problematic Here

- With three different adoption years, TWFE creates **bad comparisons**:
  - When Indonesia adopts IT in 2005, TWFE can use Thailand and Philippines as comparison groups — but they are *already treated* since 2000/2002
  - Their treatment dummy stays on; the problem is that their already-treated outcome trend is used as the benchmark for Indonesia
  - If IT effects are still evolving in Thailand or the Philippines, their inflation trend is not a clean counterfactual
- **Heterogeneity is also real:**
  - Thailand adopted IT at already-low inflation ($\approx$2%); Indonesia and Philippines after high-inflation episodes ($>$10%) — the treatment effect differs across cohorts
- **Conclusion:** Both contaminated comparisons *and* heterogeneous effects are present — exactly the setting where modern estimators are essential

Results: What the Modern Estimators Find

Method choice matters — modern estimators find larger and more significant effects.

Estimator type	Inflation level	Infl. volatility	Growth vol.
Cross-sectional DiD	-1.9^{*}	-0.9	-0.4
Stacked DiD	-4.6^{***}	-2.2^{***}	-0.1
Group-time ATT	-2.7^{**}	-0.7	0.2
IW event study	-3.5^{***}	-0.4	0.5

Estimates in pp. Stars mark significance.
Source: Nookhwun & Waiyawatjakorn (2024), Table 4.

- **Robust:** IT significantly lowers inflation levels (-1.9 to -4.6 pp) across all methods
- **Only stacked DiD** finds inflation volatility reduction — methods disagree on secondary outcomes
- **No method** finds significant effect on growth volatility

What This Tells Us About Causal Identification

A well-designed DiD tells you what the policy did — not whether the policy was a good idea.

- **Why the design is credible:**
  - IT adoption followed the 1997 Asian Financial Crisis and exchange rate regime shifts — an external shock, not an endogenous response to inflation trends
  - Control group (non-IT developing countries) faces similar macro structures and global shocks
  - Pre-trends tests support parallel trends
- **But the design has limits:**
  - IT bundles many simultaneous changes: target announcement, central bank independence, transparency reforms — DiD estimates the package, not any single element
  - Countries may have pursued other reforms alongside IT adoption
- The result is: "IT-adopting countries saw lower inflation" — not necessarily "the inflation target *itself* caused lower inflation"

Conclusion

PIER

DiD Checklist

1. **Is there a clear treatment event and a credible control group?**
   - The control group must be a plausible counterfactual for the treated group
2. **Is the parallel trends assumption plausible?**
   - Look at pre-treatment trends; is there a theoretical reason for trends to diverge absent treatment?
3. **Is treatment adoption staggered across units?**
   - If yes: is there reason to believe effects are heterogeneous across cohorts? Use modern estimators.
4. **Could there be spillovers from treated to control units?**
   - e.g., NJ minimum wage increase attracts workers away from PA restaurants (SUTVA violation)
5. **Is there anticipation of the treatment?**
   - Firms or households may adjust *before* the policy takes effect, contaminating the pre-period

Summary: When to Use Which Estimator

- **Single treatment date, two groups:** Use the canonical 2×2 DiD or TWFE — both give the same estimate; keep it simple.
- **Many units and periods, single treatment date:** TWFE is fine.
- **Staggered adoption, likely homogeneous effects:** TWFE is approximately fine; run an event study plot to check.
- **Staggered adoption, heterogeneous effects likely:** Use Callaway & Sant'Anna (2021) or Abraham & Sun (2021). Report TWFE as a comparison.
- **General advice:**
  - Always plot the event study (pre- and post-treatment)
  - Always test for pre-trends
  - If TWFE and modern estimates diverge, take the divergence seriously — it signals heterogeneity
  - Report the estimates from multiple methods as a robustness check

DiD in Central Banking: Where to Apply It

Any policy affecting some units before others — with a credible control group — is a DiD opportunity.

Setting	Example	Key challenge
Monetary policy framework	IT adoption, staggered (Nookhwun & Waiyawatjakorn 2024)	Staggered timing; regression to the mean
Macroprudential policy	Dynamic provisioning, Spain (Jiménez et al. 2017)	Formula-based bank exposure; regulatory arbitrage
Financial deregulation	Deregulation across 15 countries (Bazot 2024)	Reverse causality; endogenous reform timing
Credit & liquidity programs	COVID loan moratoriums, credit guarantees	Anticipation; all banks treated simultaneously
Exchange rate regimes	FX intervention, managed-float adoption	Spillovers; regime confounded by macro shocks

When DiD fails: all units treated simultaneously (e.g. global shock) $\Rightarrow$ no clean control group exists.

PIER

Common Pitfalls

- **Pre-trend fishing:** Running many specifications until the pre-trends look flat, then reporting only that one — this is specification searching and invalidates inference
- **TWFE with staggered adoption and heterogeneous effects:** The estimate can be biased and difficult to interpret; in extreme cases, the sign can be wrong
- **Ignoring anticipation:** If units know treatment is coming, pre-period outcomes are already affected, and the "pre" period is not truly pre-treatment
- **Spillovers:** If treatment in one unit affects untreated units (e.g., min wage in one state affects workers in neighbouring states), the control group is contaminated
- **Parallel trends in levels vs. logs:** Parallel trends in log-levels does not imply parallel trends in levels, and vice versa. Choose the scale carefully.
- **Confounding concurrent policies:** The treated group may have adopted other policies at the same time as the treatment of interest

Backup

Additional applications and technical notes

Backup: Pre-Trends Example

Dustmann et al. (2022), *Reallocation Effects of the Minimum Wage*, QJE

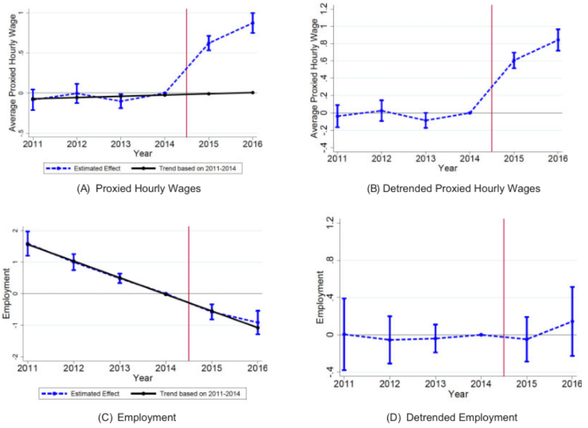


Figure VII. Wage and employment effects of Germany's minimum wage. Pre-period coefficients near zero support parallel trends.

Contiguous counties straddling a state border share local conditions but face different minimum wages.

- **Design:** Pair every county with a minimum wage increase with an adjacent county across the border in a state that does not
- **Why this control group?** Contiguous pairs share commuting zones, industry mix, and local demand shocks
- **Finding:** After accounting for local conditions, employment effects are close to zero and statistically insignificant
- **Lesson:** The choice of control group determines what confounds you remove

Germany's first national minimum wage created sharp variation in treatment "bite".

- **Reform:** Germany introduced a national minimum wage of €8.50/hour in January 2015
- **Treatment variation:** Share of workers below €8.50 before 2015 varied sharply across industries, regions, and firms
- **DiD design:** Compare high-bite vs. low-bite firms/regions before and after January 2015
- **Findings:** Bottom wages rose; small negative employment effect; large reallocation toward higher-productivity firms
- **Lesson:** DiD can exploit variation in treatment *intensity*, not just treated vs. untreated

Backup: Monetary Policy and Bank Risk-Taking

Jiménez et al. (2014, Econometrica)

Same firm, two banks with different capital buffers, same rate change.

- **Question:** Does a lower overnight rate cause banks to take on more credit risk?
- **Data:** Spanish Credit Register 2002–2008
- **DiD design:** Exploit bank capitalisation interacted with overnight rate changes
- Firm$\times$time fixed effects absorb demand-side shocks; within-firm comparison isolates credit supply
- **Lesson:** Within-firm DiD can separate loan supply from borrower demand

Backup: Macroprudential Buffers

Jiménez et al. (2017, JPE)

Formula-based provisioning creates quasi-experimental variation across banks.

- **Policy:** Spain's dynamic provisioning: banks build reserves in booms and draw them down in busts
- **DiD design:** Banks differ in buffer accumulation due to different portfolio mixes
- Same firm borrowing from high-buffer and low-buffer banks isolates supply
- **Finding:** Larger buffers support credit and firm outcomes in bad times
- **Lesson:** Macroprudential tools can provide real insurance, but spillovers/regulatory arbitrage matter

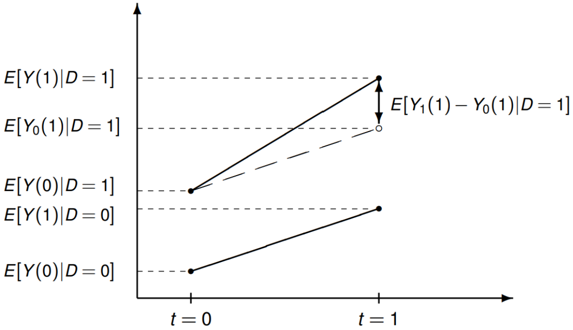
PIER

Technical Notes and Further Reading

The following topics are beyond the scope of this lecture but are important for applied work.

- **Sensitivity analysis under PT violations** (Rambachan & Roth 2023, *ReStud*). Instead of testing whether pre-trends are zero, bound the ATT under the assumption that post-treatment violations are no larger than pre-treatment ones. Implemented in the `HonestDiD` R/Stata package.
- **TWFE fails with covariates, even in 2$\times$2** (Caetano & Callaway 2024). Including time-varying covariates in TWFE produces non-convex weights on unit-level effects even without staggered timing. Use doubly-robust DiD (Sant'Anna & Zhao 2020) instead.
- **Parallel trends variants in staggered settings** (Baker et al. 2025, Ch. 3). PT-GT-Nev (vs. never-treated), PT-GT-NYT (vs. not-yet-treated), and PT-GT-all (vs. all) are distinct assumptions. Each maps to a different clean comparison set and estimator.
- **Balanced vs. unbalanced event studies** (Baker et al. 2025, Ch. 5). The standard aggregated event study uses different cohort compositions at each event time. A balanced event study fixes the cohort set — this matters when early cohorts have shorter pre-periods.
- **Simultaneous confidence bands** (Callaway & Sant'Anna 2021). Pointwise CIs understate uncertainty when you read an entire event study path. Uniform bands correct for multiple testing; available in the `did` package.
- **Forward-engineering checklist** (Baker et al. 2025, Ch. 6). Eight steps: define target parameters $\rightarrow$ state PT assumption $\rightarrow$ choose estimator $\rightarrow$ characterise uncertainty $\rightarrow$ estimate $\rightarrow$ sensitivity analysis $\rightarrow$ heterogeneity $\rightarrow$ iterate. A useful referee and design audit tool.

DiD in Potential Outcomes Notation



- Solid lines: observed outcomes
- Dashed line + open circle: counterfactual $E[Y_0(1)|D=1]$ — what the treated group would have experienced without treatment
- **ATT** = vertical gap at $t=1$:

$$E[Y_1(1) - Y_0(1) \mid D = 1]$$

- Parallel trends assumption: counterfactual trend = control group trend

When Does the Estimator Choice Matter?

- **It matters a lot when:**
  - The adoption is staggered over many periods
  - Treatment effects grow over time (dynamic effects)
  - Different cohorts have very different responses (heterogeneity)
  - There are few never-treated units (much of the variation is in bad-comparison cells)
- **It matters less when:**
  - There is effectively a single treatment date (or adoption is clustered in time)
  - Treatment effects are stable over time and similar across groups
  - There are many never-treated units serving as clean controls
- **Empirical practice:** Run TWFE as the baseline; use C&S or S&A as robustness. If the estimates differ substantially, the heterogeneity matters and you should report the modern estimates as the preferred specification.

PIER

References (1/2)

- Abraham & Sun (2021). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Journal of Econometrics*
- Baker, Callaway, Cunningham, Goodman-Bacon & Sant'Anna (2025). Difference-in-Differences Designs: A Practitioner's Guide. *JEL*
- Callaway & Sant'Anna (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*
- Card & Krueger (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *AER*
- Cengiz, Dube, Lindner & Zipperer (2019). The Effect of Minimum Wages on Low-Wage Jobs. *QJE*
- Cunningham (2021). *Causal Inference: The Mixtape*. Yale UP. Ch. 9
- Dube, Lester & Reich (2010). Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties. *REStat*

References (2/2)

- Dustmann, Lindner, Schönberg, Umkehrer & vom Berge (2022). Reallocation Effects of the Minimum Wage. *QJE*
- Goodman-Bacon (2021). Difference-in-Differences with Variation in Treatment Timing. *Journal of Econometrics*
- Jiménez, Ongena, Peydró & Saurina (2014). Hazardous Times for Monetary Policy: What Do Twenty-Three Million Bank Loans Say About the Effects of Monetary Policy on Credit Risk-Taking? *Econometrica*
- Jiménez, Ongena, Peydró & Saurina (2017). Macroprudential Policy, Countercyclical Bank Capital Buffers, and Credit Supply: Evidence from the Spanish Dynamic Provisioning Experiments. *JPE*
- Manopimoke, P., Nookhwun, N., Sriklay, T. & Tawichsri, T. (2026). Firm Inflation Expectations and the Macroeconomy: Evidence from Thailand. *PIER Working Paper*
- Nookhwun & Waiyawatjakorn (2024). Flexible Inflation Targeting and Macroeconomic Performance: Evidence from ASEAN. *Asian Economic Policy Review*
- Roth, Sant'Anna, Bilinski & Poe (2023). What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. *Journal of Econometrics*