# Staggered-Adoption Difference-in-Differences - Part I

## ECON 2400 - Applied Econometrics II

Juan C. Yamin

Brown University

Spring 2026

# Outline

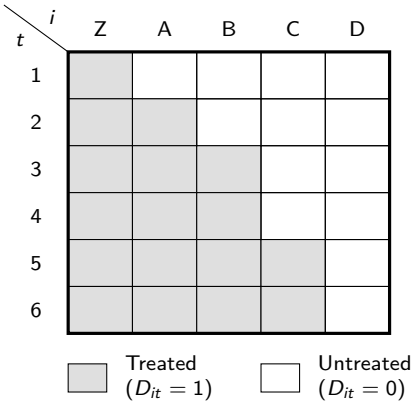
# Outline

# Setting: staggered adoption panel

- Balanced panel:
  - units $i = 1, \ldots, N$
  - periods $t = 1, \ldots, T$.
- Absorbing treatment: $D_{it} = \mathbf{1}\{t \geq E_i\} \Rightarrow$ once $i$ is treated, they remain treated for the remainder of the panel
- Adoption time:

$$E_i = \min\{t : D_{it} = 1\} \in \{1, \ldots, T, \infty\}$$

- "Cohort" = units with the same $E_i$.
- Groups:
  - always treated ($E_i = 1$)
  - switchers ($2 \leq E_i \leq T$)
  - never treated ($E_i = \infty$).



# Potential outcomes and treatment timing

- In general, $Y_{it}(\dots, d_{t-1}, d_t, d_{t+1}, \dots)$ so outcomes at time $t$ may depend on the entire path of assignments.
- If treatment turns on once and stays on, any path is summarized by $e \in \{1, \dots, T\}$:

$$Y_{it}(e) := Y_{it}(0, \dots, 0, \underbrace{1, \dots, 1}_{t \geq e}), \qquad Y_{it}(\infty) := Y_{it}(0, \dots, 0).$$

- To simplify notation, define $Y_{it}(1) := Y_{it}(E_i)$, $Y_{it}(0) := Y_{it}(\infty)$.
- Define the cell-level average effect

$$\tau_{it} := \mathbb{E}[\, Y_{it}(1) - Y_{it}(0) \,].$$

  - $\tau_{it}$ can vary across units $i$ and over time $t$.
  - Because $Y_{it}(1) = Y_{it}(E_i)$, $\tau_{it}$ can implicitly capture dynamic effects.
  - Our estimands will be weighted averages of $\{\tau_{it}\}$ over treated unit-time cells.

# Assumptions for identification

**No anticipation.** The treatment has no causal effect *prior* to its implementation. For all $t < E_i$,

$$Y_{it}(1) = Y_{it}(0).$$

**Parallel trends.** Absent treatment, average outcomes evolve as the sum of a unit effect and a time effect:

$$\mathbb{E}[\, Y_{it}(0)\,] = \alpha_i + \beta_t.$$

$\Rightarrow \alpha_i$ captures permanent level differences across units, while $\beta_t$ captures shocks common to all units. A useful equivalent form is that for any two periods $s < t$,

$$\mathbb{E}[\, Y_{it}(0) - Y_{is}(0)\,] = \beta_t - \beta_s,$$

so untreated changes are the same across units.

# Estimands: weighted averages of cell-level effects

**Treated and untreated cells.** Define the sets of unit-time pairs

$$\Omega_1 := \{(i,t) : D_{it} = 1\}, \qquad \Omega_0 := \{(i,t) : D_{it} = 0\}.$$

$\Omega_0$ includes both *never-treated* observations ($E_i = \infty$) and *not-yet-treated* observations ($t < E_i < \infty$).

**A general target parameter.** For researcher-chosen weights $\{w_{it}\}_{(i,t)\in\Omega_1}$ with $w_{it} \geq 0$ and $\sum_{(i,t)\in\Omega_1} w_{it} = 1$, define

$$\tau_w := \sum_{(i,t)\in\Omega_1} w_{it}\, \tau_{it}.$$

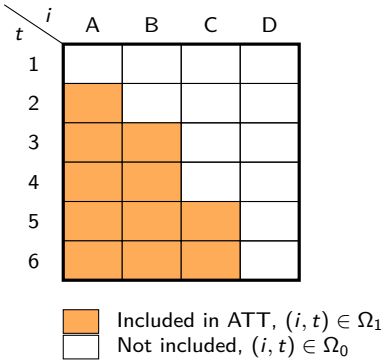
*Interpretation:* different choices of $w_{it}$ correspond to different questions (overall vs dynamic effects, different cohort weighting, etc.).

# Estimand: the (overall) ATT

Recall $\Omega_1 = \{(i,t) : D_{it} = 1\}$. The overall ATT averages the cell effects over treated unit-time pairs:

$$\tau^{ATT} := \frac{1}{|\Omega_1|} \sum_{(i,t)\in\Omega_1} \tau_{it} \qquad \Big(\text{i.e., } w_{it} = 1/|\Omega_1|\Big).$$

- We average *only* over periods in which a unit is treated.
- Never-treated units contribute no treated observations, so they do not enter $\Omega_1$.
- With dynamic effects, this estimand weights cohorts by how many post-treatment periods they contribute (early adopters get more weight if we observe them longer after adoption)



# Estimand: ATT at event time $h \geq 0$ (dynamic effects)

Instead of averaging over all treated observations, align units by *time since treatment started*:

$$h := t - E_i \quad \text{(event time / exposure length)}.$$

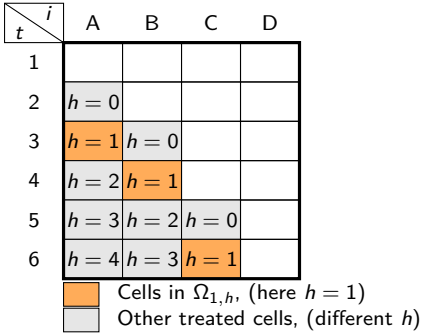
For a fixed $h \geq 0$, define the treated cells at event time $h$:

$$\Omega_{1,h} := \{(i,t) : D_{it} = 1,\ t - E_i = h\}.$$

The event-time ATT is the average effect among those cells:

$$\tau_h^{ATT} := \frac{1}{|\Omega_{1,h}|} \sum_{(i,t)\in\Omega_{1,h}} \tau_{it}.$$

Compare units at the same exposure length $h$. This isolates dynamics (effects after 0, 1, 2, ... periods since adoption).



# Outline

# Two-way fixed effects (TWFE)

Static TWFE specification — to get a single summary statistic of treatment effects:

$$Y_{it} = \alpha_i + \beta_t + \tau^{TWFE} D_{it} + \varepsilon_{it}, \qquad D_{it} = \mathbf{1}\{t \geq E_i\}.$$

A natural hope is that $\tau^{TWFE}$ equals the overall ATT estimand

$$\tau^{ATT} = \frac{1}{|\Omega_1|} \sum_{(i,t)\in\Omega_1} \tau_{it}, \qquad \tau_{it} = \mathbb{E}[Y_{it}(1) - Y_{it}(0)].$$

Under *no-anticipation*, *parallel trends* and a *homogeneous effect* across units and periods:

$$\tau_{it} = \tau \quad \text{for all } (i,t) \in \Omega_1,$$

then $\tau^{TWFE} = \tau = \tau^{ATT}$. If effects vary by cohort or event time (dynamic effects), then $\tau^{TWFE}$ generally does not equal $\tau^{ATT}$.

# What does TWFE estimate with heterogeneous effects?

With staggered adoption and heterogeneous $\tau_{it}$, the TWFE coefficient can be written as an *implicit weighted average* of cell-level effects:

$$\tau^{TWFE} = \sum_{(i,t)\in\Omega_1} w_{it}^{TWFE}\,\tau_{it}, \qquad \sum_{(i,t)\in\Omega_1} w_{it}^{TWFE} = 1.$$

The weights $\{w_{it}^{TWFE}\}$ need not match the researcher-chosen weights defining $\tau^{ATT}$, and they can be **negative**.

**Why? Forbidden Comparisons!** In staggered adoption, TWFE uses *already-treated* units as controls for *later-treated* units. When treatment effects evolve with exposure length, these treated-as-controls comparisons can produce negative weights.

# TWFE mechanics: residualization and a weighted sum

Static TWFE Specification:

$$Y_{it} = \alpha_i + \beta_t + \tau^{TWFE} D_{it} + \varepsilon_{it}.$$

**Step 1. Residualize treatment on fixed effects.** Let $\hat{D}_{it}$ be the fitted value from projecting $D_{it}$ on unit and time fixed effects:

$$D_{it} = \gamma + \alpha_i + \beta_t + \varepsilon_{it}, \qquad \hat{D}_{it} = \hat{\gamma} + \hat{\alpha}_i + \hat{\beta}_t = \bar{D}_i + \bar{D}_t - \bar{D}, \qquad \tilde{D}_{it} := D_{it} - \hat{D}_{it}.$$

(Here $\tilde{D}_{it}$ is the residual from a linear probability model with fixed effects.)

**Step 2. Frisch–Waugh–Lovell.** The TWFE coefficient satisfies

$$\hat{\tau}^{TWFE} = \frac{\sum_{it} \tilde{D}_{it} Y_{it}}{\sum_{it} \tilde{D}_{it}^2},$$

Therefore, a weighted sum:

$$\hat{\tau}^{TWFE} = \sum_{it} \hat{w}_{it}^{fe} Y_{it}, \qquad \hat{w}_{it}^{fe} := \frac{\tilde{D}_{it}}{\sum_{js} \tilde{D}_{js}^2}.$$

# TWFE mechanics: what the FE weights do

Recall $\hat{w}_{it}^{fe} := \dfrac{\tilde{D}_{it}}{\sum_{js} \tilde{D}_{js}^2}$ and $\hat{\tau}^{TWFE} = \sum_{it} \hat{w}_{it}^{fe} Y_{it}$.

Weight identities:

$$\sum_{it} \hat{w}_{it}^{fe} = 0.$$

Moreover,

$$\sum_{it} \hat{w}_{it}^{fe} D_{it} = 1 \qquad \text{and} \qquad \sum_{it} \hat{w}_{it}^{fe} (1 - D_{it}) = -1.$$

The FE weights are proportional to the residualized treatment: $\hat{w}_{it}^{fe} \propto \tilde{D}_{it}$. Because $\hat{D}_{it}$ comes from a linear probability model with fixed effects, it can fall outside $[0, 1]$.

For a treated observation ($D_{it} = 1$), $\tilde{D}_{it} < 0 \iff \hat{D}_{it} > 1$.

Intuition: if unit $i$ is treated in many periods (large $\bar{D}_i$) and period $t$ has many treated units (large $\bar{D}_t$), the additive fitted value $\bar{D}_i + \bar{D}_t - \bar{D}$ can exceed 1. Then the treated cell has a negative residual and therefore a negative FE weight. This is the algebraic manifestation of forbidden comparisons: already-treated observations end up acting like controls for later-treated observations.

# A 2-unit, 3-period example: computing $\tilde{D}_{it}$

Two units $A, B$ and three periods $t = 1, 2, 3$ with $(E_A, E_B) = (2, 3)$. So

$$D = \begin{array}{c|cc} & A & B \\ \hline t=1 & 0 & 0 \\ t=2 & 1 & 0 \\ t=3 & 1 & 1 \end{array}$$

Project $D$ on unit and time fixed effects:

$$\hat{D}_{it} = \bar{D}_i + \bar{D}_t - \bar{D}, \qquad \tilde{D}_{it} = D_{it} - \hat{D}_{it}.$$

Resulting residualized treatment:

$$\tilde{D} = \begin{array}{c|cc} & A & B \\ \hline t=1 & -\frac{1}{6} & \frac{1}{6} \\ t=2 & \frac{1}{3} & -\frac{1}{3} \\ t=3 & -\frac{1}{6} & \frac{1}{6} \end{array}$$

Even though $D_{A3} = 1$, we have $\tilde{D}_{A3} = -\frac{1}{6} < 0$. So the treated cell $(A, 3)$ will receive a negative FE weight.

# The same example: closed form and negative weight on long-run effects

Here $\sum_{it} \tilde{D}_{it}^2 = \frac{1}{3}$, so $\hat{w}_{it}^{fe} = 3\tilde{D}_{it}$ and

$$\hat{w}^{fe} = \begin{array}{c|cc} & A & B \\ \hline t=1 & -\frac{1}{2} & \frac{1}{2} \\ t=2 & 1 & -1 \\ t=3 & -\frac{1}{2} & \frac{1}{2} \end{array}$$

Closed form for the TWFE coefficient:

$$\hat{\tau}^{TWFE} = \sum_{it} \hat{w}_{it}^{fe} Y_{it} = \big(Y_{A2} - Y_{B2}\big) - \frac{1}{2}\big(Y_{A1} - Y_{B1}\big) - \frac{1}{2}\big(Y_{A3} - Y_{B3}\big).$$

Expectation under $\mathbb{E}[Y_{it}] = \alpha_i + \beta_t + \tau_{it} D_{it}$. All $\alpha_i, \beta_t$ cancel, yielding

$$\mathbb{E}[\hat{\tau}^{TWFE}] = \tau_{A2} + \frac{1}{2}\tau_{B3} - \frac{1}{2}\tau_{A3}.$$

The term $-\frac{1}{2}(Y_{A3} - Y_{B3})$ compares treated $A$ to treated $B$ at $t = 3$, giving the early adopter's longer-run effect $\tau_{A3}$ a negative weight.

# Under parallel trends: TWFE is a weighted average of treatment effects

Potential outcome model implies that $Y_{it} = Y_{it}(0) + D_{it}\tau_{it}$. So the TWFE estimator can be written as a weighted sum:

$$\hat{\tau}^{TWFE} = \sum_{it} \hat{w}_{it}^{fe} Y_{it} = \sum_{it} \hat{w}_{it}^{fe} Y_{it}(0) + \sum_{it} \hat{w}_{it}^{fe} D_{it}\tau_{it}.$$

Parallel trends gives $\mathbb{E}[Y_{it}(0)] = \alpha_i + \beta_t$. Since $\sum_t \hat{w}_{it}^{fe} = 0$ for each $i$ and $\sum_i \hat{w}_{it}^{fe} = 0$ for each $t$,

$$\mathbb{E}\left[\sum_{it} \hat{w}_{it}^{fe} Y_{it}(0)\right] = \sum_{it} \hat{w}_{it}^{fe}(\alpha_i + \beta_t) = 0.$$

Therefore.

$$\mathbb{E}[\hat{\tau}^{TWFE}] = \mathbb{E}\left[\sum_{it \in \Omega_1} \hat{w}_{it}^{fe} \tau_{it}\right], \qquad \sum_{it \in \Omega_1} \hat{w}_{it}^{fe} = 1.$$

The ATT estimand uses equal weights $1/|\Omega_1|$ on treated cells. If $\tau_{it}$ is heterogeneous and $\hat{w}_{it}^{fe}$ differs from $1/|\Omega_1|$ (and can be negative), then generally $\mathbb{E}[\hat{\tau}^{TWFE}] \neq \tau^{ATT}$.

# Outline

# Event study (fully-dynamic) specification

As before, let $E_i$ denote unit $i$'s first treatment period and define event time $h := t - E_i$.

**Fully-dynamic event study.** For integers $a \geq 1$ (how many leads) and $b \geq 1$ (how many lags),

$$Y_{it} = \alpha_i + \beta_t + \sum_{\substack{h=-a \\ h \neq -1}}^{b-1} \tau_h \mathbf{1}[t = E_i + h] \; + \; \tau_{b+} \mathbf{1}[t \geq E_i + b] \; + \; \varepsilon_{it}.$$

- The omitted category is $h = -1$, so $\tau_h$ is interpreted relative to that baseline.
- The indicators $\mathbf{1}[t = E_i + h]$ include:
  - *leads* ($h < 0$): "pre-trends / anticipation test"
  - *lags* ($h \geq 0$): "dynamic treatment effects" relative to $h = -1$
- The right-tail bin $\mathbf{1}[t \geq E_i + b]$ pools all event times $h \geq b$ into a single coefficient $\tau_{b+}$.

Common practice: some leads/lags are dropped or binned on the left and/or right

# Fully-dynamic event study: under-identification without never-treated units

Fully-dynamic event study (no binning for simplicity):

$$Y_{it} = \alpha_i + \beta_t + \sum_{h \neq -1} \tau_h \mathbf{1}[t = E_i + h] + \varepsilon_{it}.$$

**Key issue (BJS).** If there are *no never-treated units* (all units eventually adopt), then the path $\{\tau_h\}_{h \neq -1}$ is not point identified:

$$\{\tau_h\} \quad \text{and} \quad \{\tau_h + \kappa(h+1)\} \quad \text{fit the data equally well for any } \kappa \in \mathbb{R}.$$

$\Rightarrow$ To pin down the fully-dynamic path, you need never-treated units or an additional normalization (binning, dropping some leads, or imposing restrictions on pre-trends).

# Event studies: negative weights and contamination

As in static TWFE, $\hat{\tau}_h = \sum_{i,t} w_{it}^{(h)} Y_{it}$, $w_{it}^{(h)} \propto \tilde{D}_{it}^{(h)}$, $D_{it}^{(h)} = \mathbf{1}[t = E_i + h]$, where $\tilde{D}_{it}^{(h)}$ is $D_{it}^{(h)}$ residualized on unit FE, time FE, and other event-time dummies.

Taking expectations (Sun & Abraham, 2021),

$$\mathbb{E}[\hat{\tau}_h] \;=\; \sum_{(i,t):\, t \geq E_i} w_{it}^{(h)} \tau_{it} \;=\; \sum_{e} \sum_{\ell \geq 0} \sum_{i:\, E_i = e} w_{i,e+\ell}^{(h)} \mathbb{E}[Y_{i,e+\ell}(1) - Y_{i,e+\ell}(0)].$$

where the RHS is regrouped by cohort $e$ and event time $\ell$.
Problems:

1. $\tilde{D}_{it}^{(h)}$ (and $w_{it}^{(h)}$) can be negative $\Rightarrow$ Some treated cells enter with negative weight.
2. Some weight can fall on treated cells with $t \neq E_i + h \Rightarrow \hat{\tau}_h$ becomes a signed mix of different $\tau_{it}$'s, not a transparent average "effect at $h$"

   Because $\tilde{D}_{it}^{(h)}$ is orthogonal to every other $D_{it}^{(\ell)}$ with $\ell \neq h$, it must take nonzero values on cells where those other dummies are one, so already-treated cells at $\ell \neq h$ enter the comparison for $\hat{\tau}_h$.

# Event studies: consequences of contamination

Even when the event-study regression is identified, the TWFE coefficient on a given lead/lag generally does *not* isolate the causal effect at that event time.

- **Lags:** $\hat{\tau}_h$ need not equal the ATT at event time $h$ (contaminated by heterogeneity of effects at other horizons)
- **Leads:** estimated "pre-trends" can be nonzero only because lead coefficients can inherit post-treatment effects through the regression's partialling-out (even if PTA holds!)

Note: these problems tend to be small in practice!

# Outline

# Diagnosis 1: Bacon decomposition for static TWFE

Goodman-Bacon (2021): the TWFE coefficient is a convex average of 2×2 DiDs,

$$\hat{\tau} = \sum_k \omega_k \widehat{DID}_k, \qquad \omega_k \geq 0, \ \sum_k \omega_k = 1.$$

**Which comparisons?**

- Treated vs. never-treated (good)
- Early vs. late adopters (good)
- Late vs. early adopters (forbidden)
- Treated during sample vs. always-treated (forbidden)

**Diagnostic.** Report the total weight share on treated-as-control comparisons,

$$\Omega^{TC} := \sum_{k \in \text{treated-as-control}} \omega_k.$$

Large $\Omega^{TC}$ means TWFE is largely identified by treated vs. treated contrasts, which is fragile under dynamic or cohort-heterogeneous effects.

# Diagnosis 2: Where do negative weights show up? (Jakiela, 2021)

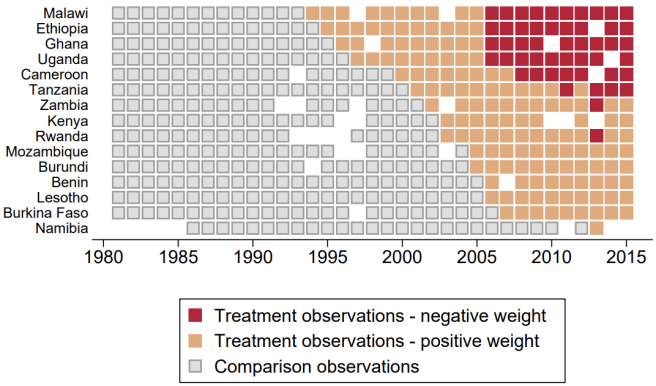



Figure 2A: Treated cells with $\tilde{D}_{it} < 0$ highlighted

**What is computed?**

- Residualize the treatment indicator on FE:

$$\tilde{D}_{it} = D_{it} - \widehat{\alpha}_i - \widehat{\beta}_t.$$

- Treated cells with $\tilde{D}_{it} < 0$ enter with **negative weight**.
- Negative-weight treated cells are effectively serving as controls in the TWFE comparison.

**Quick takeaway:** lots of negative-weight treated cells is a red flag under dynamic or heterogeneous effects.

# Diagnosis 3: What variation identifies TWFE? (Jakiela, 2021)

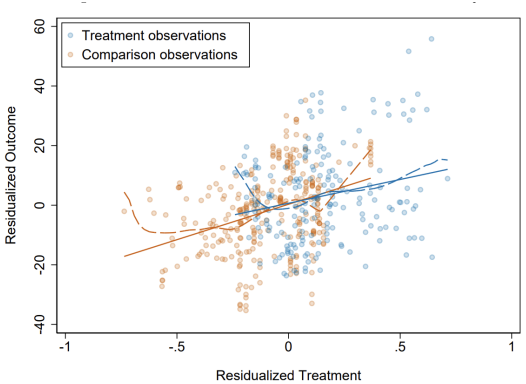



Figure 3A: Residualized outcome $\tilde{Y}_{it}$ versus residualized treatment $\tilde{D}_{it}$

**What is computed?**

- Residualize both outcome and treatment on FE
- TWFE is the slope from regressing $\tilde{Y}_{it}$ on $\tilde{D}_{it}$.
- If treated and comparison observations trace different patterns in the $(\tilde{D}, \tilde{Y})$ plot, a single common-effect summary is strained, and the TWFE slope is mixing different objects.

**Quick takeaway:** visible differences in the residualized relationship suggest heterogeneity plus problematic weighting.

# Diagnosis 4: Sign-robustness to heterogeneity

de Chaisemartin and d'Haultfœuille propose reporting the *minimal amount of treatment-effect heterogeneity* needed for the ATT to have the *opposite sign* of the static TWFE estimate:

$$\sigma_{fe} = \frac{\left|\hat{\tau}_{\text{static}}\right|}{\sigma(\hat{w}_{fe})},$$

where $\sigma(\hat{w}_{fe})$ measures the dispersion of the implied TWFE weights on treated observations.

**How to do it in practice.**

1. Estimate static TWFE and record $\hat{\tau}_{\text{static}}$.
2. Use `twowayfeweights` to compute implied weights $\hat{w}_{fe}$ and the share of negative weights.
3. Report $\sigma_{fe}$:
   - small $\sigma_{fe}$: sign reversal is possible under modest heterogeneity
   - large $\sigma_{fe}$: sign reversal would require very large heterogeneity

# Outline

# Wrap-up and what is next

**Today:**

- Setup: staggered adoption DiD and TWFE/event-study specifications
- Threats: negative weights and contamination (treated-as-controls; cross-horizon mixing)
- Diagnosis: Bacon decomposition, Jakiela diagnostics, and robustness checks (`twowayfeweights`)

**Next TA session: Estimation and Extensions**

- Estimation: group-time ATT, imputation, etc. and practical implementation
- Extensions: continuous treatment, spillovers, and other complications

# Outline

# Notation summary

# References