# Differences-in-Differences in Multiple time periods

**Benjamin O. Harrison**


**20 April 2025**

Econometrics/Reading Session

# Motivation

# Introduction

- Differences in Differences is one of the popular research designs in empirical work.
- As with any causal research design, the goal is to uncover some causal parameter of interest (ATT).
- Unlike other research designs, such as Pseudo-randomized methods, leverage on randomization to uncover causal effects.
- DiD methods exploit variation in time (before vs after) and across groups (treated vs untreated) to recover causal effects of interest.
- Key Advantage of DiD is that it allows for selection on unobservables and time trends. However, it requires the unobserved characteristics to be time-invariant.

# Running Example

- How did the Expansion of Public Health insurance (Medicaid) under the Affordable Care Act (ACA) affect Mortality?
- The ACA originally mandated that in 2014, all states expand Medicaid eligibility to adults with incomes up to 138% of the federal poverty threshold.
- Many states expanded Medicaid after 2014, but several have not done so as of 2024.

Table 1: Medicaid Expansion Under the Affordable Care Act

| Expansion Year | States | Share of States | Share of Counties | Share of Adults (2013) |
|---|---|---|---|---|
| Pre-2014 | DE, MA, NY, VT | 0.08 | 0.03 | 0.09 |
| 2014 | AR, AZ, CA, CO, CT, HI, IA, IL, KY, MD, MI, MN, ND, NH, NJ, NM, NV, OH, OR, RI, WA, WV | 0.44 | 0.36 | 0.45 |
| 2015 | AK, IN, PA | 0.06 | 0.06 | 0.06 |
| 2016 | LA, MT | 0.04 | 0.04 | 0.02 |
| 2019 | ME, VA | 0.04 | 0.05 | 0.03 |
| 2020 | ID, NE, UT | 0.06 | 0.04 | 0.02 |
| 2021 | MO, OK | 0.04 | 0.06 | 0.03 |
| 2023 | NC, SD | 0.04 | 0.05 | 0.03 |
| Non-Expansion | AL, FL, GA, KS, MS, SC, TN, TX, WI, WY | 0.20 | 0.31 | 0.26 |

**Figure 1:** Taken from Baker et al., 2025

# Setup

# Canonical $2 \times 2$ Setup

- N units: $i = 1, 2, ..., n$
- Two time periods $t = 1, 2$
- Let $Y_{it}(0, 0)$ represent the potential outcome of unit i at time t if it is untreated in both periods.
- Let $Y_{it}(0, 1)$ represent the potential outcome if treated in period 2.
- To simplify notation, let $Y_{it}(g)$ denote the potential outcome when unit i receives treatment at time t = g.
- The observed outcome $Y_{it} = (1 - D_i)Y_{it}(0) + D_i Y_{it}(1)$
- The average treatment effect of the treated at time t is:

$$ATT(t) = \mathbb{E}[Y_{it}(1) - Y_{it}(0)|D_{it} = 1]$$

- In the Medicaid example, the treated group is counties (978) that expanded in 2014, and the control group represents 1222 counties that did not expand in 2014 and 2013.

# Identification of $2 \times 2$ setup

- Recall the $ATT(t) = \mathbb{E}[Y_{it}|D_i = 1] - \mathbb{E}[Y_{it}(0)|D_i = 1)]$
- No Anticipation Assumption: For all treated units $i$ and all pre-treatment periods $t$, $Y_{i,t}(1) = Y_{i,t}(0)$.
- Parallel Trends Assumption: states that if the treatment hadn't occurred, Average outcomes for the treatment and control groups would have evolved in parallel

$$\mathbb{E}\left[Y_{i,t=2}(0) \mid D_i = 1\right] - \mathbb{E}\left[Y_{i,t=1}(0) \mid D_i = 1\right] = \mathbb{E}\left[Y_{i,t=2}(0) \mid D_i = 0\right] - \mathbb{E}\left[Y_{i,t=1}(0) \mid D_i = 0\right].$$

- Using the PT and no anticipation assumption:

$$\mathbb{E}\left[Y_{i,t=2}(0) \mid D_i = 1\right] = \mathbb{E}\left[Y_{i,t=1} \mid D_i = 1\right] + \left(\mathbb{E}\left[Y_{i,t=2} \mid D_i = 0\right] - \mathbb{E}\left[Y_{i,t=1} \mid D_i = 0\right]\right)$$

- Hence the ATT is identified as:

$$ATT(t = 2) = \left(\mathbb{E}[Y_{i,t=2}|D_i = 1] - \mathbb{E}\left[Y_{i,t=1} \mid D_i = 1\right]\right) - \left(\mathbb{E}\left[Y_{i,t=2} \mid D_i = 0\right] - \mathbb{E}\left[Y_{i,t=1} \mid D_i = 0\right]\right)$$

# More on PT Assumption

- The PT assumption can be thought of as a Selection bias Stability assumption:

$$\underbrace{E\left[Y_{i2}(0) \mid D_i = 1\right] - E\left[Y_{i2}(0) \mid D_i = 0\right]}_{\text{Selection bias in period 2}} = \underbrace{E\left[Y_{i1}(0) \mid D_i = 1\right] - E\left[Y_{i1}(0) \mid D_i = 0\right]}_{\text{Selection bias in period 1}}$$

- Recall that when we naively compare differences in means, we run into the problem of selection Bias;

$$\begin{aligned}\mathbb{E}\left(Y_i \mid T = 1\right) - \mathbb{E}(Y_i \mid T = 0) = \mathbb{E}\left(Y_i(1) \mid T = 1\right) - \mathbb{E}\left(Y_i(0) \mid T = 1\right) \\ + \mathbb{E}\left(Y_i(0) \mid T = 1\right) - \mathbb{E}\left(Y_i(0) \mid T = 0\right).\end{aligned}$$

- There is a need to make some assumptions about the treatment's relationship and potential outcomes to deal with selection bias.
- In the Randomized case, we assume the bias away through the randomness of treatment assignment.
- PT allows selection Bias but assumes that they are stable over time.

- As economists, we love regression.
- The beauty of this canonical DiD setup is we can uncover the ATT with a simple linear regression:

$$Y_{i,t} = \beta_0 + \beta_1 1\{D_i = 1\} + \beta_2 1\{t = 2014\} + \beta^{2\times 2}\left(1\{D_i = 1\} \times 1\{t = 2014\}\right) + \varepsilon_{i,t}$$

- $\widehat{ATT}(2)$ yields:

$$\widehat{ATT}(2014) = \left[\left(\widehat{\beta_0} + \widehat{\beta_1} + \widehat{\beta_2} + \widehat{\beta}^{2\times 2}\right) - \left(\widehat{\beta_0} + \widehat{\beta_1}\right)\right] - \left[\left(\widehat{\beta_0} + \widehat{\beta_2}\right) - \widehat{\beta_0}\right] = \widehat{\beta}^{2\times 2}$$

# In empirical settings

- Although the canonical setup is very easy to understand, the empirical applications in the real world deviate from canonical setup
  - How can we include covariates $X$ in our setting?
  - More than 2 time periods.
  - Variation in treatment time.
- Recent literature has shown how these variations present problems in traditional methods.
- As such one must proceed with Caution

# Multiple Time Periods

# Multiple Time periods

- In most empirical settings, like Medicaid, there are multiple periods.
- Additionally, it could be that units adopt treatment at different periods.
- Classify into two frameworks:
  - Case 1: Event Study Framework: One treatment timing, but allows for studying the dynamics of the treatment effect.
  - Case 2: Staggered Adoption Design: Allow units to adopt treatment at different periods. This is more general as it uncovers both treatment effect dynamics for different adoption periods.

# Setup

- Consider $\mathcal{T}$ periods and binary treatment $D_{it}$.
- Assumption 1: Irreversibility of Treatment
  $D_1 = 0$ almost surely. For $t = 2, .., \mathcal{T}, D_{t-1} = 1 \implies D_t = 1$
- Assumption 1 ensures that no one is treated in period 1, and once a unit adopts the treatment, it forever remains treated.
- Define G to represent the time period, a unit i first adopts the treatment.
- let $G_g$ or $G_{ig} = 1\{G_i = g\}$ be a binary variable which gives 1 if a unit receives the treatment in time g.
- units are grouped into cohorts based on their first adoption dates g.
- Let$C\infty = 1\{G_i = \infty\} = 1 - D_{i,\mathcal{T}}$ represent units that never receives the treatment.
- Define $Y_{it}(g)$ as the potential outcome of unit i at time t if he first adopted the treatment at t = g
- The potential outcome is defined over the full treatment path.

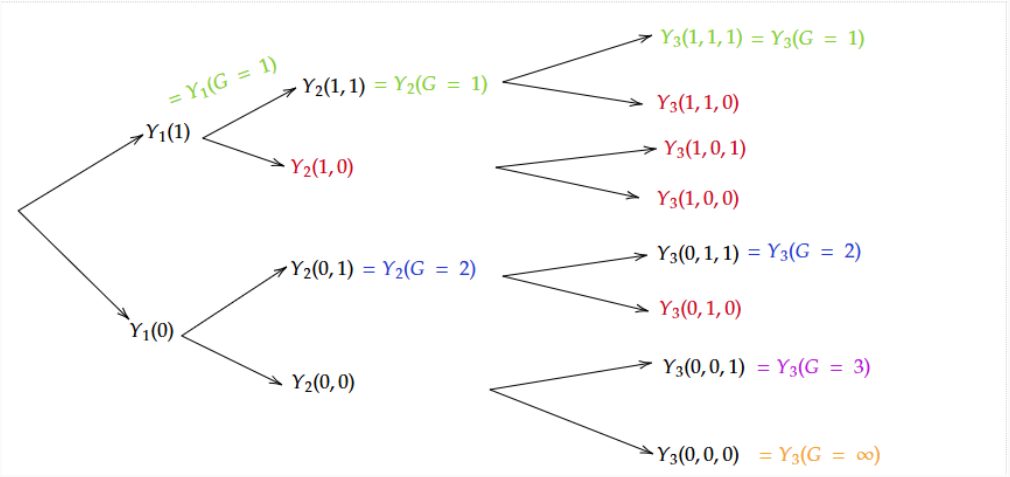



**Figure 2:** Caption

# Comment on Treatment timing

- The event study framework contains only one treatment timing group (Medicaid expansion in 2014).
- In this framework, the question is identifying the causal effect of switching the treatment in 2014 on the mortality rate in 2014, 2015, 2016, ...
- This framework identifies the one-time treatment effect dynamics.
- However, the subtlety in this setup is that if a unit adopts treatment, it is forever treated, and also if a unit doesn't get treated in 2014, it never receives the treatment.
- Essentially, the ES causal estimand is $Y_{it}(1, 1, ..., 1) - Y(0, 0, ..., 0)$, which is one of the many causal estimands of the dynamic causal effect literature.

# Comment 2: Treatment timing

- The Staggered Adoption design generalizes the event study design by allowing units who fail to adopt treatment in earlier periods to adopt it later.
- The SAD therefore computes for each cohort the treatment effect dynamics of that cohort.
- so SAD setup is just a generalization of the event study setup in which the causal estimand Group-time Average Treatment effect parameter

$$ATT(g,t) = \mathbb{E}(Y_{it}(g) - Y_{it}(0)|G_g = 1)$$

- This estimand strips away certain causal estimands like the causal effect of receiving a one-time treatment in which the treatment is turned off once it is administered or perhaps the causal effect of an early adopted treatment as to a late adopted treatment.

# Justification of Assumption 1

- The Justification of this assumption hinges critically on how the researcher defines the ”treatment” variable.
- Case 1: Treatment definition on lasting experience. In this case, the treatment is defined as the status of having been exposed or undergone the intervention.
- Example: Consider the active job training of 2 years. The treatment can be defined as having received the training or as a status of being trained (in this case, irreversible).
- When the treatment is defined as the latter, as in many cases, the causal question encompasses both the effects during active training and the long-term effects of the aftermath of the training experience.

# Justification of Assumption 1

- Example 2: Suppose we are interested in the effect of a crisis or a disaster like an earthquake.
- By defining the treatment as a state or experience, one combines both the effect of the one-time disaster (earthquake event) and the effect of the aftermath of the earthquake.
- However, Policymakers may be interested in isolating the effect of the acute treatment event itself from the longer-term aftermath of that event.
- Example: Evaluating the effectiveness of warning systems or alarms in mitigating harm during the event.

# Can TWFE recover *ATT* ?

# Traditional Methods: TWFE Event study regression

- Can the simple two fixed effects estimator identify the causal effect under this treatment variation framework??

$$Y_{it} = \alpha_i + \phi_t + D_{it}\beta + \epsilon_{it}$$

- Athey and Imbens, 2022; Goodman-Bacon, 2021 proved that naively running a simple TWFE estimation obtains a weighted average of different treatment effect parameters, i.e.

$$\mathbb{E}\left[\hat{\tau}_{\text{did}}\right] = \sum_{t\in\mathbb{T}}\sum_{a\leq t} \gamma_{t,a}\tau_{t,\infty a} = \sum_{t\in\mathbb{T}} \gamma_{t,+}\tau_{t,\infty 1} - \sum_{t\in\mathbb{T}}\sum_{a\leq t} \gamma_{t,a}\tau_{t,a1}.$$

- If the treatment effect is constant over time, then: $\mathbb{E}\left[\hat{\tau}_{\text{did}}\right] = \tau_{\infty 1}$.
- The weights can be negative $\implies$ the effect of the treatment can be positive for all groups, but the TWFE estimation could lead to a negative overall effect.

# Interpretation

- Goodman-Bacon, 2021 proved that the negative weights is due to treatment effect dynamics.
- TWFE OLS specification combines two sources of comparisons:
  - Clean comparisons: DiD's between treated and not-yet-treated units
  - Forbidden comparisons: DiD's between newly-treated and already-treated units
- These forbidden comparisons can lead to negative weights: the "control group" is already treated, so we run into problems if their treatment effects change over time.

# Some Intuition for Forbidden Comparison

- Consider the two-period model, except suppose now that our two groups are always-treated units (treated in both periods) and switchers (treated only in period 2)
- With two periods, the coefficient $\beta$ from $Y_{it} = \alpha_i + \phi_t + D_{it}\beta + \epsilon_{it}$ is the same as from the first-differenced regression $\Delta Y_i = \alpha + \Delta D_i \beta + u_i$

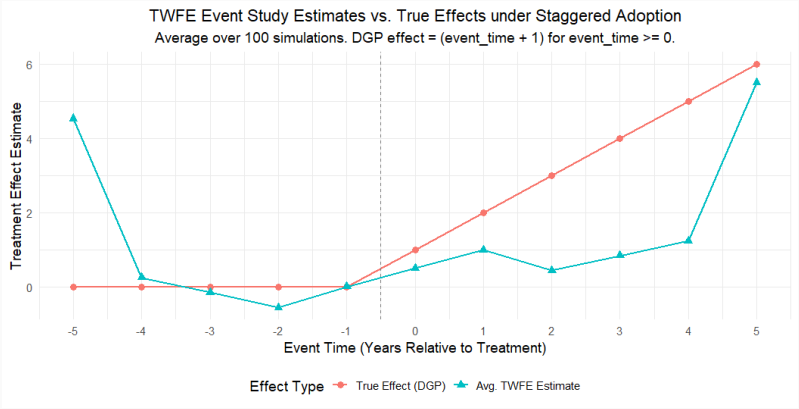
$$\hat{\beta} = \underbrace{\left(\overline{Y}_{\text{Switchers},2} - \overline{Y}_{\text{Switchers},1}\right)}_{\text{Change for switchers}} - \underbrace{\left(\overline{Y}_{AT,2} - \overline{Y}_{AT,1}\right)}_{\text{Change for always treated}}$$

- Problem: if the effect for the always-treated grows over time, that will enter $\hat{\beta}$ negatively!
- Using the Frisch-Waugh-Lovell theorem, we obtain $\beta$ using two-step:
  - regress the treatment indicator $D_{it}$ on $\alpha_i, \phi_t$ (Linear Probability model)
  - regress $Y_{it}$ on $D_{it} - \hat{D}_{it}$ to obtain $\beta = \frac{E\left(Y_{it}\left(D_{it} - \hat{D}_{it}\right)\right)}{\text{Var}\left(D_{it} - \hat{D}_{it}\right)}$
- However, it's well known that the linear probability model for $D_{it}$ may have predictions outside the unit interval

# Stylized Example using simulated data

- So what if one runs a dynamic Event study treatment effect, such as:

$$Y_{i,t} = \alpha_t + \alpha_g + \sum_{e=-K}^{-2} \delta_e^{anticip} \cdot D_{i,t}^e + \sum_{e=0}^{L} \beta_e \cdot D_{i,t}^e + v_{i,t}$$



# Differences-in-Differences with Multiple time periods

# Callaway and Sant'Anna 2021

- TWEF tries to aggregate the effect across time and units, but runs into the issue of negative weights.
- To address these problems, we must
  - Be precise about the **target parameter** (estimand) – how do we want aggregate treatment effects across time/units.
  - Estimate the target parameter using only clean comparison
- The main idea behind Callaway and Sant'Anna, 2021 is to break the problem of identification into a sequence of $2 \times 2$ Canonical DiD estimators.
- Then use the identification methodology to identify each sequence and then provide an aggregation scheme to summarise the sequence of target parameters.

# Identification

- Define $ATT(g,t)$ to be the ATT in period t for units who first adopted treatment at period g;

$$ATT(g,t) = \mathbb{E}(Y_{it}(g) - Y_{it}(0) | G_g = 1)$$

- Under the PT and No Anticipation assumption, the $ATT(g,t)$ is identified as :

$$ATT(g,t) = \underbrace{E\left[Y_{it} - Y_{i,g-1} \mid G_g = 1\right]}_{\text{Change for cohort g}} - \underbrace{E\left[Y_{it} - Y_{i,g-1} \mid C = 1\right]}_{\text{Change for never-treated units}} \quad \forall t \geq g$$

- Hence $ATT(g,t)$ is a $2 \times 2$ canonical DiD estimator which compares the cohort treated group g to the never-treated group.
- The proof follows directly from $2 \times 2$ case under:
  - PT : $E\left[Y_{it}(0) - Y_{i,g-1}(0) \mid G_g = 1\right] - E\left[Y_{it}(0) - Y_{i,g-1}(0) \mid C = 1\right] \quad \forall t, g \quad t \geq g$
  - No anticipation : $\mathbb{E}\left(Y_{it}(g) \mid G_g = 1\right) = \mathbb{E}\left(Y_{it}(0) \mid G_g = 1\right) \quad \forall t, g, \quad t < g$

**Proof.**

$$\begin{aligned}
\mathrm{ATT}(g,t) &= \mathbb{E}\left(Y_{it}(g) - Y_{it}(0) \mid G_g = 1\right) \\
&= \mathbb{E}\left(Y_{it} \mid G_g = 1\right) - \mathbb{E}\left(Y_{ig-1}(0) \mid G_g = 1\right) + E\left(Y_{g-1}(0) \mid G_g = 1\right) - E\left(Y_{it}(0) \mid G_g\right. \\
&\text{By no anticipation: } \mathbb{E}\left(Y_{g-1}(0) \mid G_g = 1\right) = \mathbb{E}\left(Y_{g-1}(g) \mid G_g = 1\right) \\
&= \mathbb{E}\left(Y_t - Y_{g-1} \mid G_g = 1\right) - \left(\mathbb{E}\left(Y_t(0) \mid G_g = 1\right) - \mathbb{E}\left(Y_{g-1}(0) \mid G_g = 1\right)\right) \\
&= \mathbb{E}\left(Y_t - Y_{g-1} \mid G_g = 1\right) - \left(\mathbb{E}\left(Y_t(0) - Y_{g-1}(0) \mid C = 1\right)\right) \\
ATT(g,t) &= E\left[Y_{it} - Y_{i,g-1} \mid G_g = 1\right] - E\left[Y_{it} - Y_{i,g-1} \mid C = 1\right] \quad \forall t \geq g
\end{aligned}$$

□

# Identification with Covariate

# General Framework

- One of the contributions of Callaway and Sant'Anna, 2021 is how covariates play a role in this staggered adoption design.
- Consider a random sample

$$\left\{\left(Y_{i,1}, Y_{i,2}, \ldots, Y_{i,\mathcal{T}}, D_{i,1}, D_{i,2}, \ldots, D_{i,\mathcal{T}}, X_i\right)\right\}_{i=1}^{n}$$

  where $D_{i,t} = 1$ if unit $i$ is treated in period $t$, and 0 otherwise.
- Assumption 3: **Limited Treatment Anticipation**: There is a known $\delta \geq 0$ s.t.

$$\mathbb{E}\left[Y_t(g) \mid X, G_g = 1\right] = \mathbb{E}\left[Y_t(0) \mid X, G_g = 1\right] \text{ a.s.}$$

  for all $g \in \mathcal{G}, t \in 1, \ldots, \mathcal{T}$ such that $\underbrace{t < g - \delta}_{\text{"before effective starting date"}}$.
- Generalized propensity score uniformly bounded away from 1 :

$$p_{g,t}(X) = P\left(G_g = 1 \mid X, G_g + (1 - D_t)\left(1 - G_g\right) = 1\right) \leq 1 - \epsilon \text{ a.s.}$$

# PT assumption based on ”never treated” group

**Assumption (Conditional Parallel Trend based on a ”never treated”)**

*For each $t \in \{2, \ldots, \mathcal{T}\}, g \in \mathcal{G}$ such that $t \geq g - \delta$,*

$$\mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid X, G_g = 1\right] = \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid X, C = 1\right] \text{ a.s.}$$

# PT assumption based on "not-yet-treated" group

**Assumption (Conditional Parallel Trend based on a "not-yet-treated")**

*For each* $(s,t) \in \{2,\dots,\mathcal{T}\} \times \{2,\dots,\mathcal{T}\}, g \in \mathcal{G}$ *such that* $t \geq g - \delta, s \geq t + \delta$

$$\mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid X, G_g = 1\right] = \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid X, D_s = 0, G_g = 0\right] \text{ a.s.}$$

# Choice of the comparison Group

- The choice of the comparison group is based on the application.
- If there exists a group that never receives the treatment in all t, then the ”never treated” group PT is probably the way.
- However, in some applications, the ”never-treated” group may be too different from the treated group; in that case, the ”not yet treated” group may serve as a good comparison.
- The choice is based on which comparison group is more informative.
- Additionally, the not yet treated comparison group may be a larger set in the short run, but in the long run, especially if all units eventually get treated, then we can’t identify ATT after that.
- Lastly, the validity of the DiD estimator in the long run reduces.

# Identification: Never treated as comparison group

- To simplify notation let $\delta = 0$
- With the PT and No anticipation, for $t \geq g$ $\mathrm{ATT}(g,t)$ is non-parametrically identified by the DR estimand

$$\mathrm{ATT}_{dr}^{\mathrm{nev}}(g,t) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}\left[G_g\right]} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_g(X)C}{1-p_g(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_{g,t}^{nev}(X)\right)\right]$$

where $m_{g,t}^{nev}(X) = \mathbb{E}\left[Y_t - Y_{g-1} \mid X, C = 1\right]$

- We can write the

$$\Rightarrow ATT(g,t) = E\left(\mathbb{E}\left(Y_t(g) - Y_t(0) \mid G_g = 1, X\right) \mid G_g = 1\right)$$

$$= E\left[E\left(Y_t - Y_{g-1} \mid G_g = 1, X\right) - \underbrace{E\left(Y_t - Y_{g-1} \mid C = 1, X\right)}_{m_{g,t}(X)} \mid G_g = 1\right]$$

$$= E\left(Y_t - Y_{g-1} \mid G_g = 1\right) - E(m_{g,t}(X)|G_g = 1)$$

- For any binary random variable $G_g$, $E(Z|G_g = 1) = \frac{E(ZG_g)}{P(G_g=1)} = \frac{E(ZG_g)}{E(G_g)}$
- Hence the outcome regression estimator for ATT :

$$ATT(g,t)_{or} = \mathbb{E}\left[\frac{G_g}{\mathbb{E}(G_g)}(Y_t - Y_{g-1} - m_{g,t}(X)\right]$$

# Sketch of Proof ctd

- Similarly;

$$ATT^{nev}_{ipw}(g,t) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}\left[G_g\right]} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_g(X)C}{1-p_g(X)}\right]}\right)\left(Y_t - Y_{g-1}\right)\right],$$

- Finally, the $ATT^{\text{nev}}_{dr}(g,t)$ combines both methods.

# Summarizing ATT(g,t) ’s

# Aggregation Schemes

- ATT(g, t) are very useful parameters that allow us to better understand treatment effect heterogeneity.
- Callaway and Sant'Anna, 2021 proposed an aggregation of $ATT(g,t)$ in an informative way:

$$\sum_{g=2}^{\mathcal{T}} \sum_{t=2}^{\mathcal{T}} \mathbf{1}\{g \leq t\} w_{gt} ATT(g,t)$$

- The weights are chosen by the researcher in an informative way.
- Two simple aggregation methods are :

$$\theta_M^O := \frac{2}{\mathcal{T}(\mathcal{T}-1)} \sum_{g=2}^{\mathcal{T}} \sum_{t=2}^{\mathcal{T}} \mathbf{1}\{g \leq t\} \, \mathrm{ATT}(g,t)$$

$$\theta_W^O := \frac{1}{\kappa} \sum_{g=2}^{\mathcal{T}} \sum_{t=2}^{\mathcal{T}} 1\{g \leq t\} \, \mathrm{ATT}(g,t) P(G = g \mid C \neq 1)$$

# Aggregation con't

- Policymakers are interested in the event study dynamics of the treatment.
- Average effect of participating in the treatment for the group of units that have been exposed to the treatment for exactly $e$ periods

$$\theta_D(e) = \sum_{g=2}^{\mathcal{T}} 1\{g + e \leq \mathcal{T}\} ATT(g, g+e) P(G = g \mid G + e \leq \mathcal{T}, C \neq 1)$$

- This aggregation avoids the pitfall of overweighting earlier adopters, unlike $\theta_W^0$

# Estimation and Inference

# Estimation

- Identification results suggest a simple two-step estimation procedure.
- Estimate the generalized propensity score $p_g(X)$ by $\widehat{p}_g(X)$.
- Estimate outcome regression models for the comparison group, $m^C_{g-1}(X)$ and $m^C_t(X)$, by $\widehat{m}^C_{g-1}(X)$, and $\widehat{m}^C_t(X)$, respectively.
- With these estimators on hands, estimate the $\mathrm{ATT}(g,t)$ using the plug-in principle (you can use IPW, OR or DR estimands!).

- - Under relatively weak regularity conditions,

$$\sqrt{n}(\widehat{\mathrm{ATT}}(g,t) - ATT(g,t)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{gt}(\mathcal{W}_i) + o_p(1)$$

- From the above asymptotic linear representation and a CLT, we have

$$\sqrt{n}(\widehat{ATT}(g,t) - ATT(g,t)) \xrightarrow{d} N\left(0, \Sigma_{g,t}\right)$$

  where $\Sigma_{gt} = \mathbb{E}\left[\psi_{gt}(\mathcal{W})\psi_{gt}(\mathcal{W})'\right]$.
- Above result ignores the dependence across $g$ and $t$, and "multiple-testing" problems.

# Multiplier Bootstrappping Method

- How to construct simultaneous confidence intervals?
- We propose the use of a simple multiplier bootstrap procedure.
- Let $\widehat{\Psi}_{g\leq t}(\mathcal{W})$ denote the sample-analogue of $\Psi_{g\leq t}(\mathcal{W})$.
- Let $\{V_i\}_{i=1}^n$ be a sequence of iid random variables with zero mean, unit variance and bounded third moment, independent of the original sample $\{\mathcal{W}_i\}_{i=1}^n$
- $\widehat{ATT}^*_{g\leq t}$, a bootstrap draw of $\widehat{ATT}_{g\leq t}$, via

$$\widehat{ATT}^*_{g\leq t} = \widehat{ATT}_{g\leq t} + \mathbb{E}_n\left[V \cdot \widehat{\Psi}_{g\leq t}(\mathcal{W})\right]$$

# Bootstrapping Procedure

1. Draw a realization of $\{V_i\}_{i=1}^n$.
2. Compute $\widehat{ATT}^*_{g\leq t}$ as in (3), denote its $(g,t)$-element as $\widehat{ATT}^*(g,t)$, and form a bootstrap draw of its limiting distribution as
$$\hat{R}^*(g,t) = \sqrt{n}\left(\widehat{ATT}^*(g,t) - \widehat{ATT}(g,t)\right)$$
3. Repeat steps 1-2 $B$ times.
4. Estimate $\Sigma^{1/2}(g,t)$ by
$$\widehat{\Sigma}^{1/2}(g,t) = \left(q_{0.75}(g,t) - q_{0.25}(g,t)\right) / \left(z_{0.75} - z_{0.25}\right)$$
5. For each bootstrap draw, compute $t-$ test $^*_{g\leq t} = \max_{(g,t)} \left|\hat{R}^*(g,t)\right| \widehat{\Sigma}(g,t)^{-1/2}$.
6. Construct $\widehat{c}_{1-\alpha}$ as the empirical $(1-\alpha)$-quantile of the $B$ bootstrap draws of $t-test^*_{g\leq t}$.
7. Construct the bootstrapped simultaneous confidence intervals for ATT ( $g,t$ ), $g \leq t$, as
$$\widehat{C}(g,t) = \left[\widehat{ATT}(g,t) \pm \widehat{c}_{1-\alpha} \cdot \widehat{\Sigma}(g,t)^{-1/2}/\sqrt{n}\right].$$

# Conclusion

# Thanks!

benjamin.harrison@emory.edu
benhars.com
@benhars

# References


- Athey, S., & Imbens, G. W. (2022). **Design-based analysis in difference-in-differences settings with staggered adoption.** *Journal of Econometrics*, *226*(1), 62–79.
- Baker, A., Callaway, B., Cunningham, S., Goodman-Bacon, A., & Sant'Anna, P. H. (2025). **Difference-in-differences designs: A practitioner's guide.** *arXiv preprint arXiv:2503.13323*.
- Callaway, B., & Sant'Anna, P. H. (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.