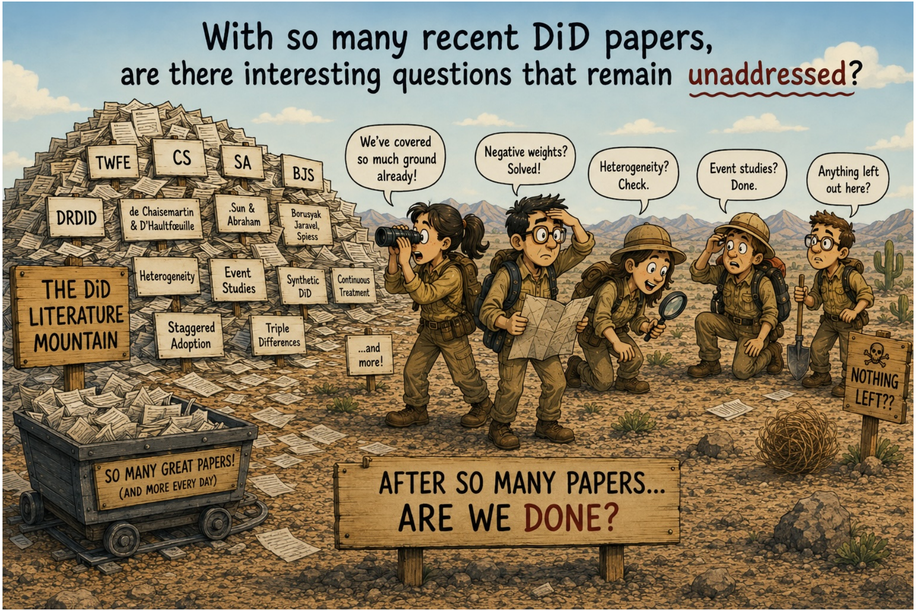
# Efficient Difference-in-Differences and Event-Study Estimators

Xiaohong Chen
Yale University

Pedro H. C. Sant'Anna
Emory University

Haitian Xie
Peking University

2026 North American Meeting of the Econometric Society
June 2026


With so many recent DiD papers,
are there interesting questions that remain unaddressed?
TWFE
CS
SA
BJS
DRDID
de Chaisemartin & D'Haultfœuille
Sun & Abraham
Borusyak Jaravel, Spiess
THE DiD LITERATURE MOUNTAIN
Heterogeneity
Event Studies
Synthetic DiD
Continuous Treatment
Staggered Adoption
Triple Differences
...and more!
We've covered so much ground already!
Negative weights? Solved!
Heterogeneity? Check.
Event studies? Done.
Anything left out here?
NOTHING LEFT??
SO MANY GREAT PAPERS! (AND MORE EVERY DAY)
AFTER SO MANY PAPERS...
ARE WE DONE?

# Some open questions that remain and why they matter

- Take an applied researcher with **several pre-treatment periods** and **multiple untreated / not-yet-treated cohorts**. Concrete design choices arise:
  - Which **pre-treatment periods** to use as baseline?
  - Which **not-yet-treated** cohorts to use as comparison groups?
  - Should we report and compareresults from **multiple estimators**?
- These are not innocuous — they raise open questions with **first-order** stakes:
  - **Precision:** how much do existing estimators sacrifice?
  - **Discipline:** which baselines & comparison groups should we include or report?
  - **Sensitivity:** how much do conclusions depend on these choices?
- Empirical DiD paper makes these choices, usually implicitly.

# Same data, same estimand — very different precision

- Under the same parallel-trends assumptions, prominent DiD estimators on the same data deliver **near-identical point estimates**

- Yet their asymptotic variances can differ by a **factor of two or more**.
  - They are not targeting different parameters.
  - They simply use **different subsets** of the available identifying information.

- In our application, widely used alternatives would need up to **104% more observations** to match the precision of the efficient estimator.

- With a small comparison group, that gap decides **which effects you can detect at all**.


Yes! There is more to discover.
Wait... what's this?
Look!
We found something!
OPEN QUESTIONS
(We haven't fully used the information yet!)
PRECISION
How much information are we leaving on the table?
DISCIPLINE
Which baselines & comparison groups should we use?
SENSITIVITY
How much do conclusions depend on these choices?
EFFICIENT ESTIMATION
How to use all valid information?
OVER-IDENTIFICATION
(the key!)
PLENTY MORE TO DISCOVER!

# DiD is (often) over-identified — and that is the engine

- With **multiple pre-treatment periods** or **staggered timing**, the DiD framework is typically **nonparametrically over-identified**.(Chen and Santos, 2018)
  - Many distinct estimators follow from the same assumptions; modern methods use only some of the moments.

- Over-identification is not only a source of precision. It is what makes the other design questions **tractable** — it governs:
  - how **precise** a DiD estimator can be;
  - **which comparisons** drive each estimate;
  - **diagnostics** for restrictions in tension with the data;
  - how to trade off **precision against bias** when assumptions are uncertain.

# What we do in this paper

- **1. Efficiency.** We derive the **semiparametric efficiency bound** for cohort-specific $ATT(g,t)$ and event-study $ES(e)$ parameters, give **closed-form efficient influence functions**, and propose easy-to-compute estimators that **attain the bound**.

- **2. Design diagnostics.** Different weighting choices can target different estimands under misspecification, so the same over-identification also yields:
  - **incremental over-identification tests** for additional baseline / comparison-group restrictions;
  - **robustness frontiers**: how far a headline estimate moves as restrictions are relaxed, at a stated precision cost (Andrews, Chen and Tecchio, 2025)
  - **weight decompositions**: what accounts for differences across estimators.

- **3. Adaptive estimation.** An **adaptive shrinkage** estimator that trades precision gains against potential bias.(Armstrong, Kline and Sun, 2024)

- Single & staggered timing; with or without covariates; extension to **instrumented DiD**.

# A taste of the econometrics

# DiD framework

- "Short" panel $\{W_i\}_{i=1}^n = \{(Y_{i,1}, \ldots, Y_{i,T}, X_i', G_i)'\}_{i=1}^n$: $n$ large, $T$ finite and fixed.
- Treatment is binary, an **absorbing state**, with possibly different starting dates.
- $Y_{i,t}(g)$: potential outcome for unit $i$ at time $t$ if first treated in period $g$.
- $G_i$: period unit $i$ is first treated, with $G_i = \infty$ if never treated by $T$; $G_i \in \mathcal{G} \subseteq \{2, \ldots, T, \infty\}$.
  - Single date: $G_i = g$ (treated) or $G_i = \infty$ (untreated). Let $G_g = \mathbf{1}\{G = g\}$.

# Target parameters

- Group-time average treatment effects on the treated:

$$ATT(g,t) := \mathbb{E}[Y_t(g) - Y_t(\infty) \mid G = g].$$

- Aggregate over cohorts into event-study summaries ($e = t - g$):

$$\begin{aligned} ES(e) &:= \mathbb{E}\left[ATT(G, G+e) \,\middle|\, G+e \in [1,T]\right], \\ ES_{\text{avg}} &:= \tfrac{1}{|\mathcal{E}|} \textstyle\sum_{e \in \mathcal{E}} ES(e). \end{aligned}$$

# Maintained Assumption: Sampling, Overlap, and No-anticipation

**Assumption (Maintained Assumption (M))**

(i) *(S)* $\{(Y_{i,t=1}, \ldots, Y_{i,t=T}, X_i', G_i)'\}_{i=1}^n$ *is a random sample from* $(Y_{t=1}, \ldots, Y_{t=T}, X', G)'$.

(ii) *(O) For each* $g \in \mathcal{G}$, $\mathbb{E}[G_g|X] \in (0,1)$ *almost surely (a.s.).*

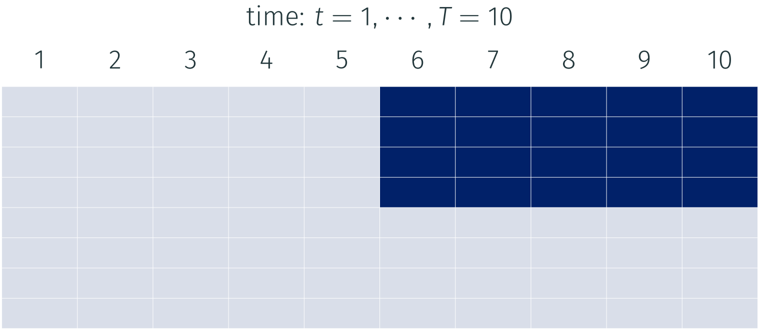
(iii) *(NA) For every* $g \in \mathcal{G}_{trt}$, *and every pre-treatment periods* $t < g$, $\mathbb{E}[Y_{i,t}(g)|G = g, X] = \mathbb{E}[Y_{i,t}(\infty)|G = g, X]$ *almost surely.*

- The maintained Assumption M is not enough to identify ATT or ES type parameters.
- DiD methods impose parallel trends assumptions to identify these parameters: we will discuss them in a bit.

# DiD with Single Treatment Time

# No variation in treatment timing

- Single treatment period at time $g$: $G_i = g$ (treated) or $G_i = \infty$ (untreated).
- $ES(e) = ATT(g, g+e)$.



# Parallel Trends Assumption: Post-treatment periods

**Assumption (PT in the post-treatment periods)**

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

$$\mathbb{E}[Y_t(\infty) - Y_{t-1}(\infty)|G = g, X] = \mathbb{E}[Y_t(\infty) - Y_{t-1}(\infty)|G = \infty, X] \ a.s.$$

- Impose parallel trends only for post-treatment periods.
- Uses only period $g - 1$ as the baseline.
- Without covariates, easy to show that, for any $t \geq g$,

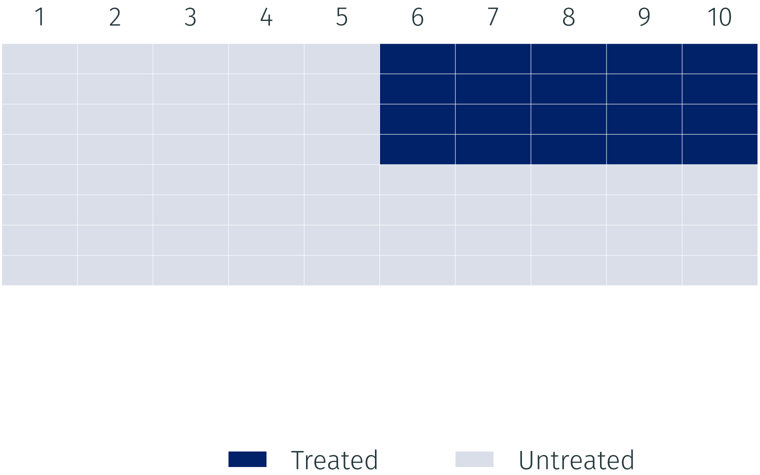
$$ATT(g, t) = \mathbb{E}[Y_t - Y_{g-1}|G = g] - \mathbb{E}[Y_t - Y_{g-1}|G = \infty].$$

- Limitation: If we were gifted 10 more pre-treatment data periods, the estimand for $ATT(g, t)$ would not be allowed to use any of that.

# PT-Post: Implications

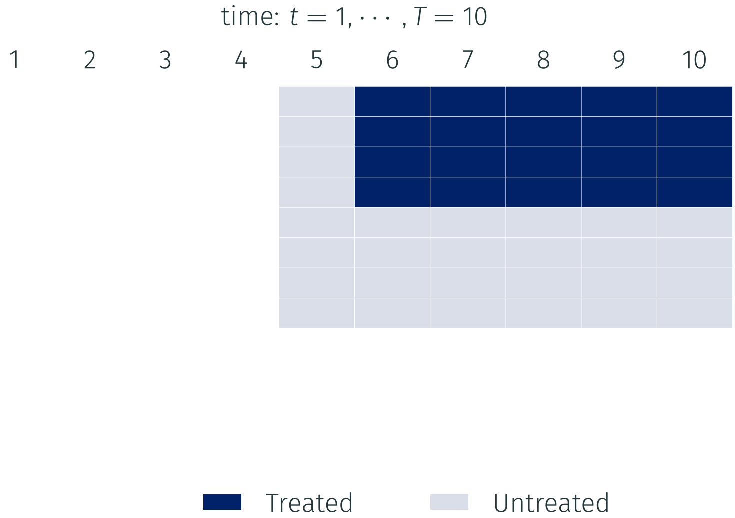
- $ES(e) = ATT(g, g+e)$ for any $e \geq 0$

time: $t = 1, \cdots, T = 10$



# PT-Post: Implications

- $ES(e) = ATT(g, g+e)$ for any $e \geq 0$



# Parallel Trends Assumption: All periods

## Assumption (PT in all periods)

*For each* $t \in \{2, \ldots, T\}$,

$$\mathbb{E}[Y_t(\infty) - Y_{t-1}(\infty)|G = g, X] = \mathbb{E}[Y_t(\infty) - Y_{t-1}(\infty)|G = \infty, X] \ a.s.$$

- Impose parallel trends in all periods.
- Allows us to use any pre-treatment period as a baseline.
- Without covariates, easy to show that for any $t \geq g$ and any $t' < g$,

$$ATT(g, t) = \mathbb{E}[Y_t - Y_{t'}|G = g] - \mathbb{E}[Y_t - Y_{t'}|G = \infty].$$

- If we were gifted 10 more pre-treatment periods of data, we could easily use all of them to compute $ATT(g, t)$.
- **How to do that efficiently, such that we maximize precision?**

# Characterization of the DiD model based on seq. conditional moment restrictions

**Lemma (Moment-restrictions for over-identified DiD with single treatment time)**

*The family of prob. dist. of $(Y_{t=1} \cdots, Y_{t=T}, X', G)$ satisfying Assumptions M and PT-All-g are observationally equivalent to the family of prob. dist. of $(Y_{t=1} \cdots, Y_{t=T}, X', G)$ satisfying Assumption M(i)(ii), and the set of moment restrictions: for all $t \in \{g, \ldots, T\}$, with prob. one,*

$$\mathbb{E}[G_g(ATT(g,t) - CATT(g,t,X))] = 0,$$

$$\mathbb{E}\left[CATT(g,t,X) - \frac{G_g(Y_t - Y_{g-1})}{p_g(X)} + \frac{G_\infty(Y_t - Y_{g-1})}{p_\infty(X)} \middle| X\right] = 0,$$

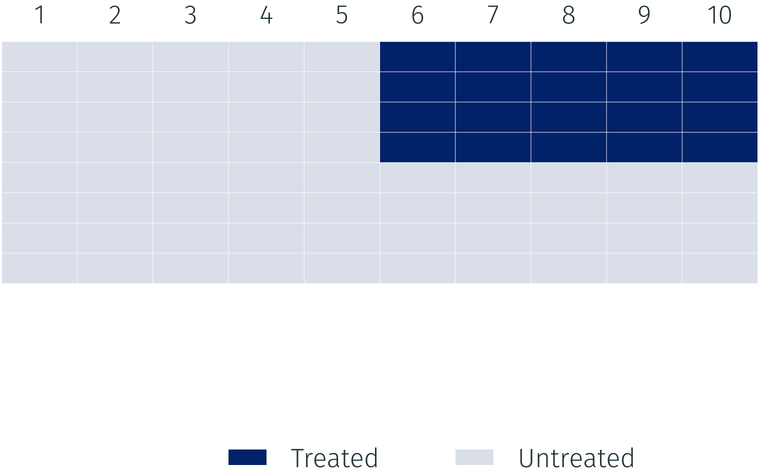
$$\mathbb{E}\left[\frac{G_g(Y_{t'} - Y_1)}{p_g(X)} - \frac{G_\infty(Y_{t'} - Y_1)}{p_\infty(X)} \middle| X\right] = 0, \text{ for all } 2 \leq t' \leq g-1,$$

$$\mathbb{E}[G_g - p_g(X)|X] = 0.$$

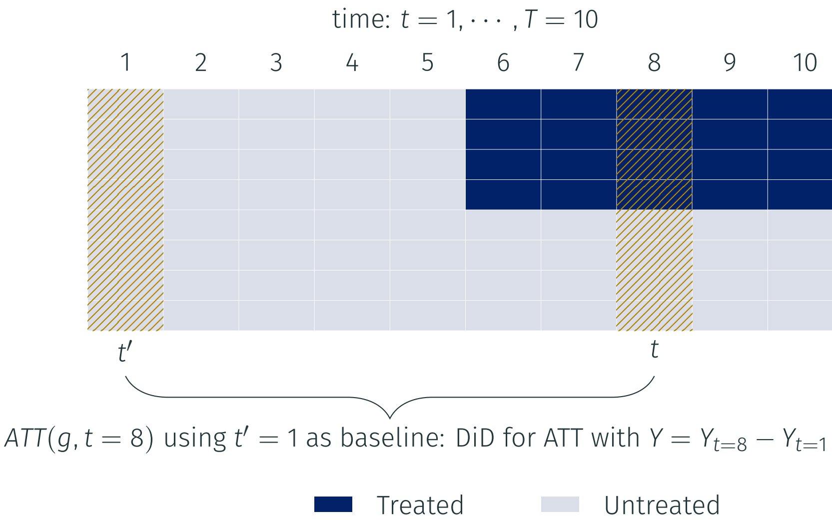
- Semiparametric efficient bound for $ATT(g,t)$: apply Ai and Chen (2012) for seq. moments.

# Understanding the sources of over-identification

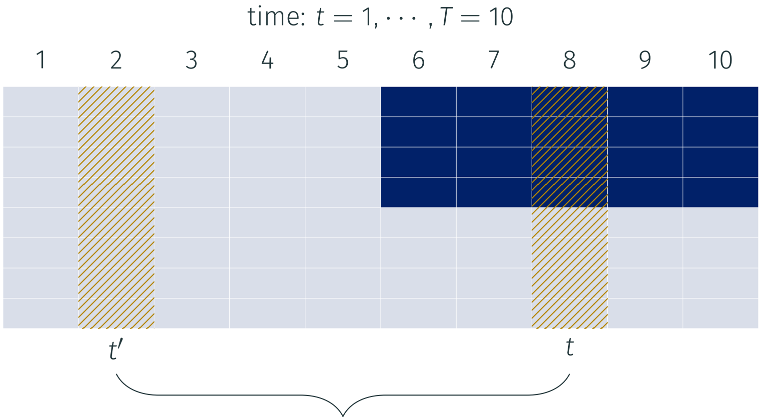
time: $t = 1, \cdots, T = 10$



# Understanding the sources of over-identification



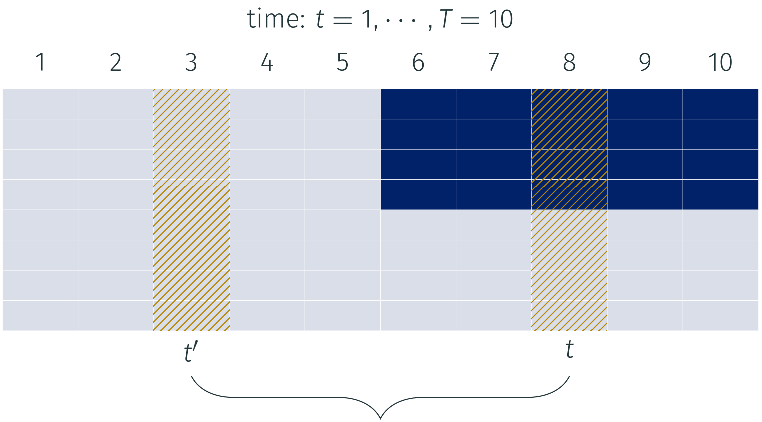
# Understanding the sources of over-identification




$ATT(g, t = 8)$ using $t' = 2$ as baseline: DiD for ATT with $Y = Y_{t=8} - Y_{t=2}$

Treated    Untreated

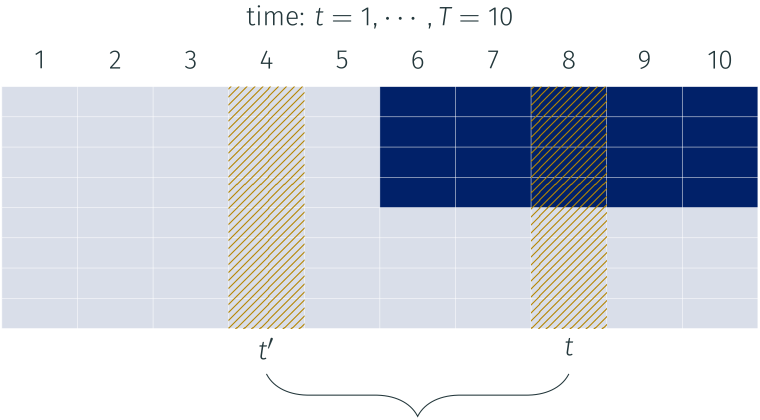
# Understanding the sources of over-identification




$ATT(g, t = 8)$ using $t' = 3$ as baseline: DiD for ATT with $Y = Y_{t=8} - Y_{t=3}$

Treated Untreated

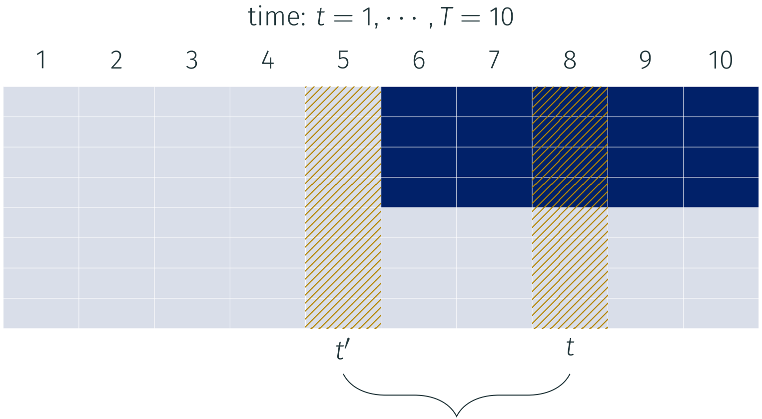
# Understanding the sources of over-identification




$ATT(g, t = 8)$ using $t' = 4$ as baseline: DiD for ATT with $Y = Y_{t=8} - Y_{t=4}$

Treated    Untreated

# Understanding the sources of over-identification




$ATT(g, t = 8)$ using $t' = 5$ as baseline: DiD for ATT with $Y = Y_{t=8} - Y_{t=5}$

Treated    Untreated

# Intuition on how to get semiparametric efficiency

- For each $1 \leq t' \leq g-1$, we fix the baseline period at $t'$, and compute the "efficient influence function" for $ATT(g,t)$ as-if there were only 2 groups, $G = g$ and $G = \infty$, and two periods, $t$ (post-treatment) and $t'$ (pre-treatment)
  - Akin to compute the "DR scores" in DML language.
- Stack all the non-collinear influence functions into a vector, $\mathbb{IF}^{att(g,t)}$.
- Compute the covariance of $\mathbb{IF}^{att(g,t)}$ given covariates, $V_{gt}(X) = \text{Cov}\left(\mathbb{IF}^{att(g,t)} | X\right)$.
- Efficient Influence Function for $ATT(g,t)$ is given by

$$EIF^{att(g,t)} = \frac{\mathbf{1}' V_{gt}(X)^{-1}}{\mathbf{1}' V_{gt}(X)^{-1} \mathbf{1}} \mathbb{IF}^{att(g,t)}.$$

- Next, we explore these results to obtain EIF-based estimands for $ATT(g,t)$, which serve as a blueprint for efficient estimation.

# Using EIF as a blueprint for estimating ATT(g,t)

- The key is to explore that $\mathbb{E}[EIF^{att(g,t)}] = 0$ to get IF-based estimand:

$$ATT(g,t) = \mathbb{E}\left[\frac{\mathbf{1}'V_{gt}(X)^{-1}}{\mathbf{1}'V_{gt}(X)^{-1}\mathbf{1}}\theta_{g,t}(W)\right], \tag{1}$$

where $p_g(X) = \mathbb{E}[G_g|X]$, $\theta_{g,t}(W) = (\theta_{g,t,1}(W), \dots, \theta_{g,t,g-1}(W))'$ is a $(g-1) \times 1$ column vector with

$$\theta_{g,t,t'}(W) = \frac{1}{\mathbb{P}(G=g)}\left(G_g - \frac{(1-G_g)p_g(X)}{1-p_g(X)}\right)(Y_t - Y_{t'} - \mathbb{E}[Y_t - Y_{t'}|G=\infty, X]) \,.$$

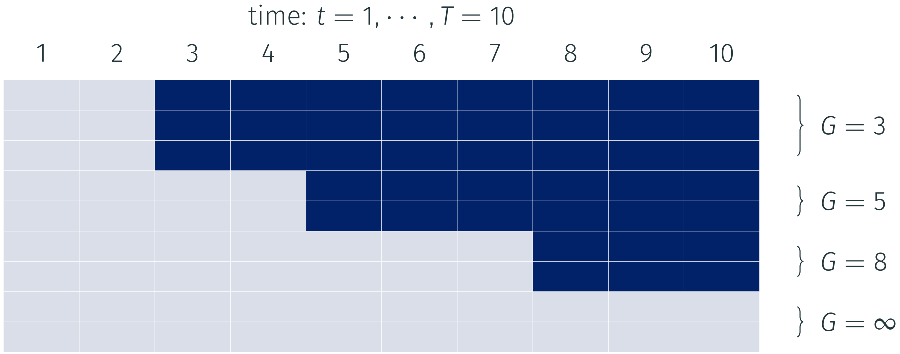
- Here, $\theta_{g,t}(W)$ is a vector of DR DiD "integrands", each being computed pretending we were in the $2 \times 2$ DiD setup of Sant'Anna and Zhao (2020).

- Efficient DiD estimators: apply plug-in principle, or do DML.

# DiD with Staggered Treatment Adoption

# Staggered Adoption

- Multiple treatment starting periods, leading to several treatment groups defined by treatment starting date.
- Each group has their $ATT(g, t)$.





EMORY UNIVERSITY

# PT for all periods

**Assumption (Parallel Trends for all groups and periods (PT-All))**

*For each* $t \in \{2, \ldots, T\}$ *and* $(g, g') \in \mathcal{G}_{trt} \times \mathcal{G}$,

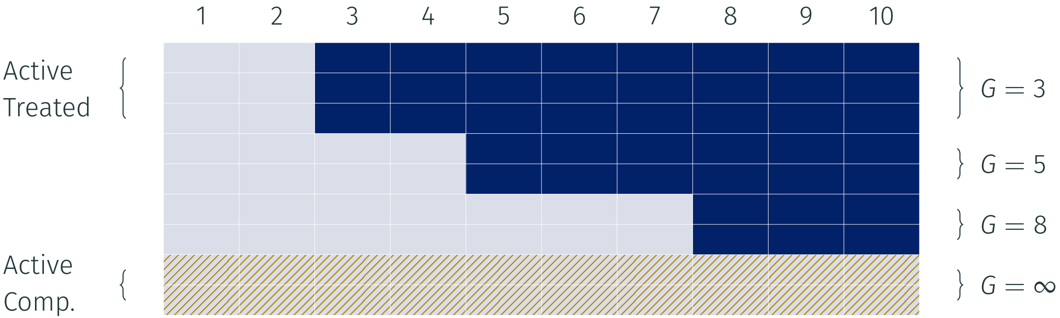
$$\mathbb{E}[Y_t(\infty) - Y_{t-1}(\infty)|G = g, X] = \mathbb{E}[Y_t(\infty) - Y_{t-1}(\infty)|G = g', X] \ a.s.$$

*i.e., conditional on covariates, the average evolution of untreated potential outcomes is the same across treatment groups, in all available periods.*

- Two sources of nonparametric over-identification (in the sense of Chen and Santos (2018)): multiple baseline periods and multiple comparison groups.

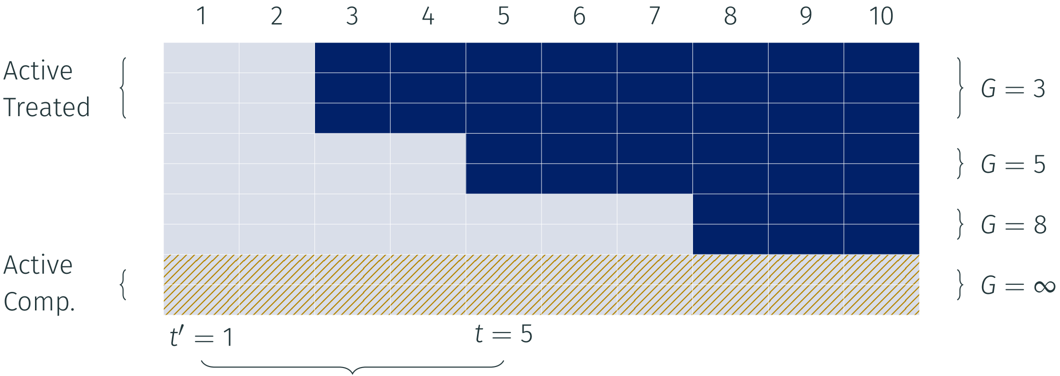
# Understanding the sources of over-identification

$ATT(g = 3, t = 5)$, using one active comparison group $G = \infty$:



# Understanding the sources of over-identification

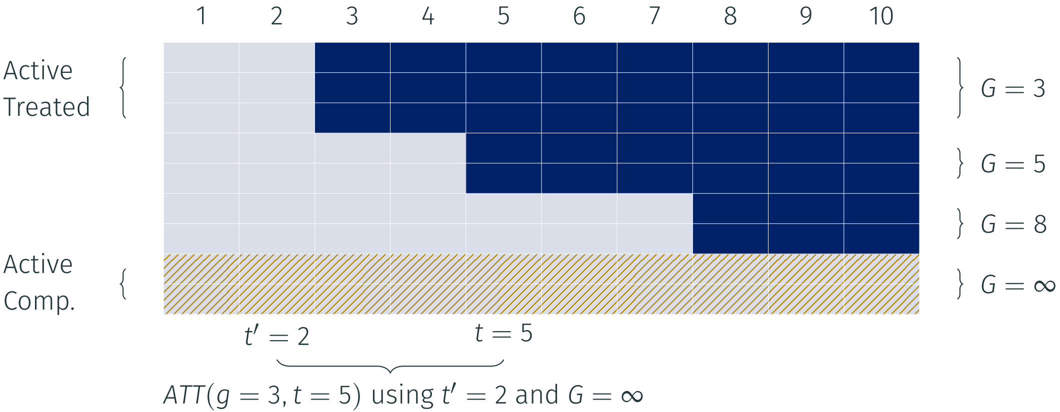
$ATT(g = 3, t = 5)$, using one active comparison group $G = \infty$:




$ATT(g = 3, t = 5)$ using $t' = 1$ and $G = \infty$

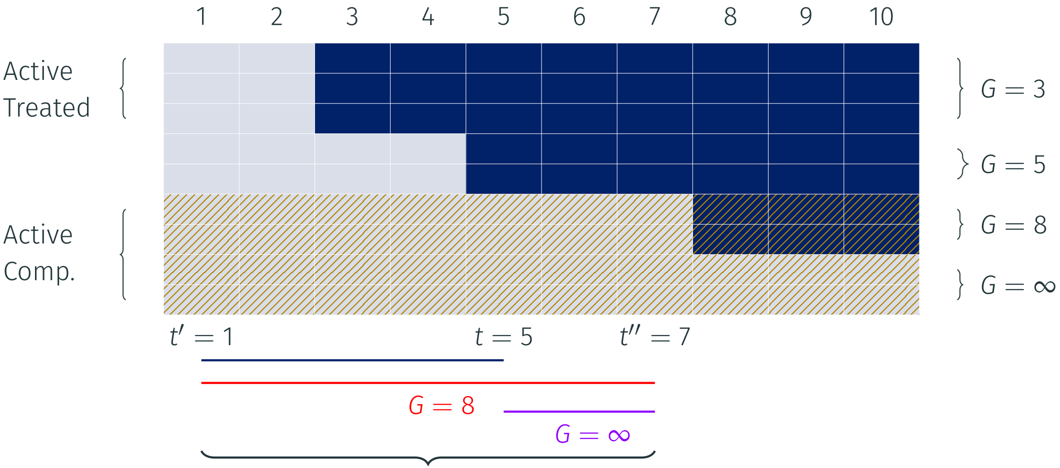
# Understanding the sources of over-identification

$ATT(g = 3, t = 5)$, using one active comparison group $G = \infty$:

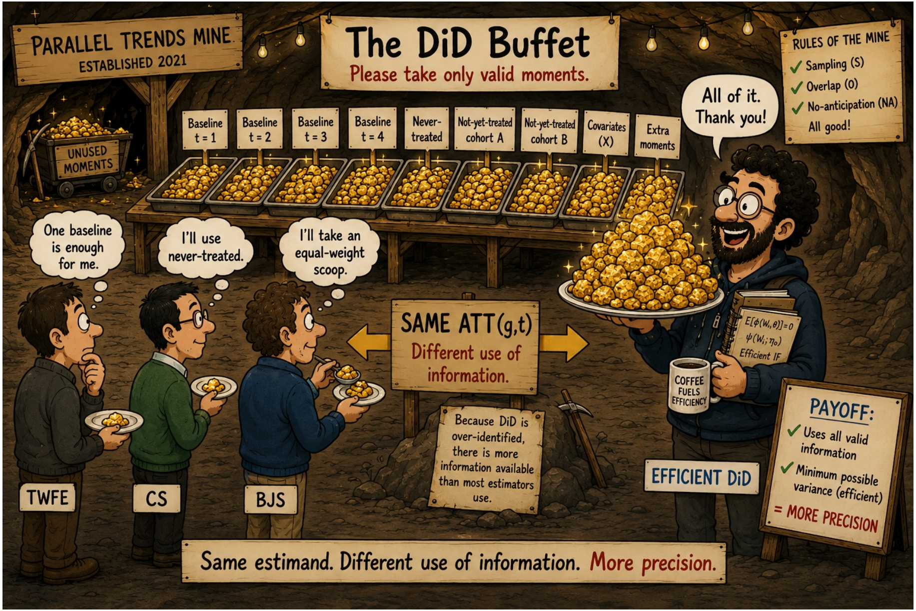


# Understanding the sources of over-identification

$ATT(g = 3, t = 5)$, using two active comparison groups $G = 8$, and $G = \infty$:




$ATT(g = 3, t = 5)$ using $t' = 1$, $t'' = 7$, and $G = \infty, 8$



PARALLEL TRENDS MINE
ESTABLISHED 2021
The DiD Buffet
Please take only valid moments.
RULES OF THE MINE
Sampling (S)
Overlap (O)
No-anticipation (NA)
All good!
UNUSED MOMENTS
Baseline t = 1
Baseline t = 2
Baseline t = 3
Baseline t = 4
Never-treated
Not-yet-treated cohort A
Not-yet-treated cohort B
Covariates (X)
Extra moments
All of it. Thank you!
One baseline is enough for me.
I'll use never-treated.
I'll take an equal-weight scoop.
SAME ATT(g,t)
Different use of information.
Because DiD is over-identified, there is more information available than most estimators use.
E[φ(W,θ)]=0
ψ(W;η₀)
Efficient IF
COFFEE FUELS EFFICIENCY
PAYOFF:
Uses all valid information
Minimum possible variance (efficient)
= MORE PRECISION
TWFE
CS
BJS
EFFICIENT DiD
Same estimand. Different use of information. More precision.

# Exploring all the content of PT

## Lemma (Over-identification in staggered designs)

*Under Assumptions M and PT-All, for every group* $(g, g') \in \mathcal{G}_{trt} \times \mathcal{G}_{trt}$ *and time periods* $(t, t', t'') \in \mathcal{T} \times \mathcal{T} \times \mathcal{T}$ *such that* $t \geq g$, $g > t'$, *and* $g' > \max\{t', t''\}$, *with probability one,*

$$CATT(g,t,X) = \underbrace{\mathbb{E}[Y_t - Y_{t'}|G = g, X]}_{\equiv m_{g,t,t'}(X)} - \Big(\underbrace{\mathbb{E}[Y_t - Y_{t''}|G = \infty, X]}_{\equiv m_{\infty,t,t''}(X)} + \underbrace{\mathbb{E}[Y_{t''} - Y_{t'}|G = g', X]}_{\equiv m_{g',t'',t'}(X)}\Big), \tag{2}$$

*and, as a consequence,*

$$ATT(g,t) = \mathbb{E}\left[Y_t - Y_{t'}\middle|G = g\right] - \mathbb{E}\left[\left(m_{\infty,t,t''}(X) + m_{g',t'',t'}(X)\right)\middle|G = g\right]. \tag{3}$$

*More generally, for any covariate-specific weights* $w^{g,t}_{g',t',t''}(X)$ *that sum up to one, we have that*

$$ATT(g,t) = \mathbb{E}\left[\sum_{(g',t',t'')\in\mathcal{H}^{g,t}} w^{g,t}_{g',t',t''}(X)\left[m_{g,t,t'}(X) - \left(m_{\infty,t,t''}(X) + m_{g',t'',t'}(X)\right)\right]\middle|G = g\right].$$

# Semiparametric Efficient Staggered DiD

- We will now fix $t' = 1$ for simplicity (no loss of generality).
- For $g' \in \mathcal{G}_{\text{trt}}$ and $1 \leq t'' \leq g' - 1$, let $\pi_g = \mathbb{E}[G_g]$ and

$$\mathbb{IF}^{att(g,t)}_{g',t''} = \frac{1}{\pi_g}\Big(G_g((m_{g,t,1}(X) - m_{\infty,t,t''}(X) - m_{g',t'',1}(X)) - ATT(g,t))\Big) + \frac{G_g}{\pi_g}(Y_t - Y_1 - m_{g,t,1}(X)) - \frac{G_\infty}{\pi_g}\frac{p_g(X)}{p_\infty(X)}(Y_t - Y_{t''} - m_{\infty,t,t''}(X)) - \frac{G_{g'}}{\pi_g}\frac{p_g(X)}{p_{g'}(X)}(Y_{t''} - Y_1 - m_{g',t'',1}(X)). \quad (4)$$

- We then stack all the non-collinear IF terms to form $\mathbb{IF}^{att(g,t)}_{\text{stg}}$.

# Efficient Staggered DiD

- Here follows the efficient staggered DiD estimand

$$ATT(g,t) = \mathbb{E}\left[\frac{1'\Omega_{gt}(X)^{-1}}{1'\Omega_{gt}(X)^{-1}1}\theta_{\text{stg}}^{\text{att(g,t)}}(W)\right], \tag{5}$$

where $\Omega_{gt}(X) = \text{Cov}(\mathbb{IF}_{\text{stg}}^{att(g,t)}|X)$, $\theta_{\text{stg}}^{\text{att(g,t)}}(W)$ is the column vector

$$\theta_{\text{stg}}^{\text{att(g,t)}}(W) = (\theta_{g'}^{\text{att(g,t)}}(W)', g' \in \mathcal{G}_{\text{trt}})',, \tag{6}$$

such that, for $g' = g$,

$$\theta_{g'}^{\text{att(g,t)}}(W) = (\theta_{g,1}^{\text{att(g,t)}}(W), \ldots, \theta_{g,g-1}^{\text{att(g,t)}}(W))',$$

and, for $g' \neq g$,

$$\theta_{g'}^{\text{att(g,t)}}(W) = (\theta_{g',2}^{\text{att(g,t)}}(W), \ldots, \theta_{g',g'-1}^{\text{att(g,t)}}(W))',$$

with

$$\begin{aligned}\theta_{g',t''}^{\text{att(g,t)}}(W) = {} & \frac{1}{\pi_g}G_g(Y_t - Y_1 - m_{\infty,t,t''}(X) - m_{g',t'',1}(X)) \\ & - \left(\frac{G_\infty}{\pi_g}\frac{p_g(X)}{p_\infty(X)}(Y_t - Y_{t''} - m_{\infty,t,t''}(X)) + \frac{G_{g'}}{\pi_g}\frac{p_g(X)}{p_{g'}(X)}(Y_{t''} - Y_1 - m_{g',t'',1}(X))\right).\end{aligned} \tag{7}$$

# Semiparametric efficiency result

**Theorem (Efficient DiD with staggered treatment adoption)**

*Under Assumptions M and PT-All, the efficient influence function for $ATT(g,t)$, $t \geq g$, is given by*

$$\mathbb{EIF}_{stg}^{att(g,t)} = \frac{\mathbf{1}'\Omega_{gt}(X)^{-1}}{\mathbf{1}'\Omega_{gt}(X)^{-1}\mathbf{1}}\mathbb{IF}_{stg}^{att(g,t)}.$$

*The semiparametric efficient variance bounds are obtained as the second moments of the efficient influence functions, provided they are finite.*

*See our paper for the efficient influence function of a $ES(e)$, $e \geq 0$.*

# The efficient weights — and why existing estimators fall short

- Optimal aggregation across pre-periods and comparison groups is governed by $V_{gt}(X)^{-1}$. The weights:
  - depend on the **covariance of outcome changes** within each group;
  - **vary** with the event time $t$;
  - are **covariate-dependent** (unlike GMM) — e.g. optimal weights for men may differ from women;
  - are **generally not constant** across pre-treatment periods.

- Most DiD / ES estimators are **not** efficient — they either:
  - use a **single** pre-period $t'$: Dynamic TWFE (DTWFE), de Chaisemartin and D'Haultfœuille (2020), Callaway and Sant'Anna (2021), Sun and Abraham (2021); or
  - take a **simple average** over $t' < g$: TWFE, Wooldridge (2021), Gardner (2021), Borusyak, Jaravel and Spiess (2024).

# Does it matter? An empirical application

# The causal effect of hospitalization on out-of-pocket spending

- Dobkin, Finkelstein, Kluender and Notowidigdo (2018) study the effect of hospitalization on out-of-pocket medical spending using a DiD / Event-Study strategy on the timing of hospitalization.

- We follow the sample construction of Sun and Abraham (2021), using the publicly available Health and Retirement Study (HRS) data from the Dobkin et al. (2018) replication package:
  - Adults hospitalized at ages 50–59, excluding pregnancy-related admissions.
  - Balanced panel spanning waves 7–11 (2004–2012).
  - Final sample: 652 individuals in 4 treatment groups — $G_i = 8$ (252), $G_i = 9$ (176), $G_i = 10$ (163), $G_i = \infty$ (only **65** never-treated).

- Small comparison group $\Rightarrow$ **power is binding**, and precision matters.

# Point estimates are similar across estimators

| Estimator | $ATT(8,8)$ | $ATT(8,9)$ | $ATT(8,10)$ | $ATT(9,9)$ | $ATT(9,10)$ | $ATT(10,10)$ | $ES(0)$ | $ES(1)$ | $ES(2)$ | $ES_{avg}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| EDiD | 3072 | 1112 | 1038 | 3063 | 90 | 2908 | 3024 | 692 | 1038 | 1585 |
| | (806) | (637) | (817) | (690) | (641) | (894) | (486) | (471) | (816) | (521) |
| CS-SA | 2826 | 825 | 800 | 3031 | 107 | 3092 | 2960 | 530 | 800 | 1430 |
| | (1035) | (909) | (1008) | (702) | (651) | (995) | (539) | (585) | (1008) | (647) |
| CS-dCDH | 3029 | 1248 | 800 | 3324 | 107 | 3092 | 3134 | 779 | 800 | 1571 |
| | (913) | (861) | (1008) | (959) | (651) | (995) | (536) | (570) | (1008) | (566) |
| BJS-G-W | 3029 | 1285 | 1021 | 3239 | 77 | 2758 | 3017 | 788 | 1021 | 1609 |
| | (916) | (767) | (851) | (862) | (729) | (957) | (555) | (587) | (851) | (582) |

- As expected when PT is plausible — the estimators agree on the point estimates.

SEs in parentheses. **EDiD**: efficient DiD. **CS-SA** / **CS-dCDH**: Callaway and Sant'Anna (2021) & Sun and Abraham (2021) / de Chaisemartin and D'Haultfœuille (2020) (never- / not-yet-treated). **BJS-G-W**: imputation (Borusyak et al., 2024; Gardner, 2021; Wooldridge, 2021).

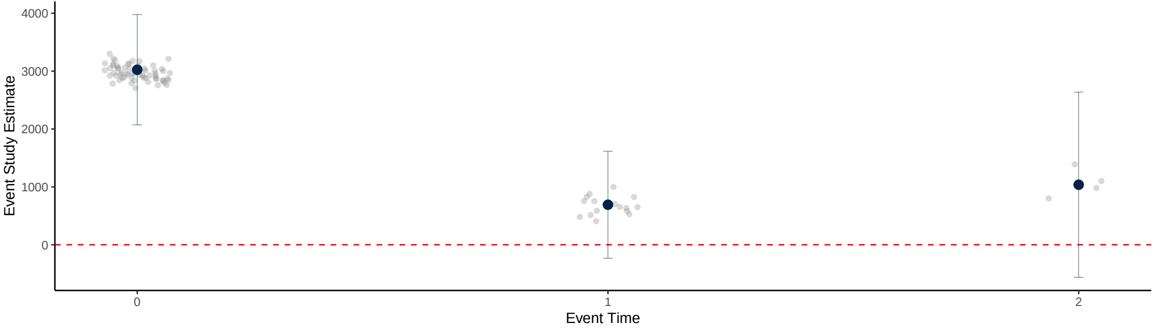
# Our efficient DiD delivers substantial gains in efficiency

- Asymptotic relative efficiency (ARE) of EDiD vs. the other estimators — heuristically, the **relative sample size** they would need to match EDiD's precision.

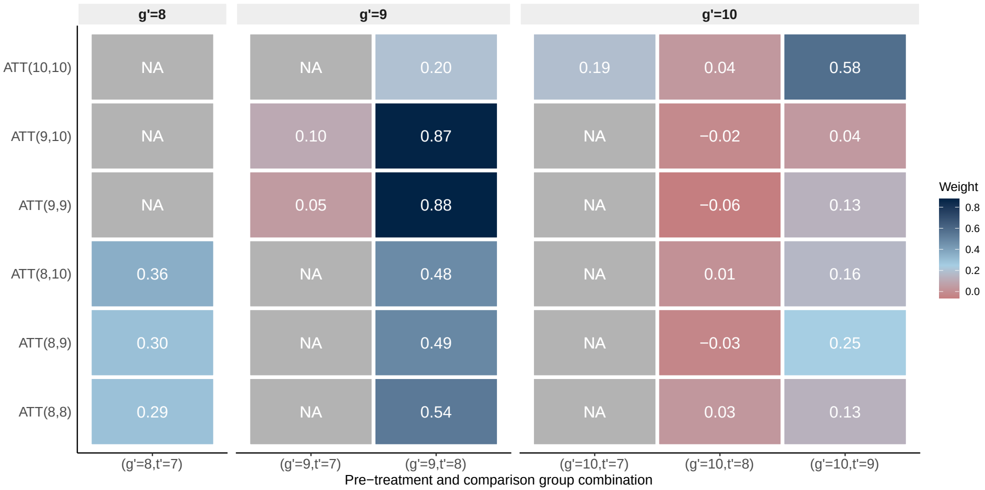
| Estimator | $ATT(8,8)$ | $ATT(8,9)$ | $ATT(8,10)$ | $ATT(9,9)$ | $ATT(9,10)$ | $ATT(10,10)$ | $ES(0)$ | $ES(1)$ | $ES(2)$ | $ES_{avg}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| EDiD | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| CS-SA | 1.65 | 2.04 | 1.52 | 1.04 | 1.03 | 1.24 | 1.23 | 1.54 | 1.52 | 1.54 |
| CS-dCDH | 1.28 | 1.83 | 1.52 | 1.93 | 1.03 | 1.24 | 1.21 | 1.46 | 1.52 | 1.18 |
| BJS-G-W | 1.29 | 1.45 | 1.09 | 1.56 | 1.29 | 1.15 | 1.30 | 1.55 | 1.09 | 1.25 |

- Several alternatives need **up to twice the sample size** (e.g. $ATT(8,9)$: ARE $= 2.04$) for the same precision; the ranking is not fixed across parameters.

# Visualizing event-study stability



# Understanding the efficient weights for the $ATT(g, t)$'s



# Takeaways

# Takeaways

- DiD is usually **over-identified** — and that over-identification is the **engine** behind precision, diagnostics, and sensitivity.
- We deliver **closed-form efficient estimators** that attain the semiparametric bound — single & staggered designs, with or without covariates.
- The same structure gives **incremental over-id tests**, **robustness frontiers**, **weight decompositions**, and an **adaptive** estimator.
- Gains are **first-order**: in Dobkin et al. (2018), alternatives need up to **104% more data** to match our precision.
- Researchers can report not just an estimate, but **which comparisons drive it, what precision they buy, and how much the conclusion depends on including them**.

# Thanks!

pedro.santanna@emory.edu

psantanna.com

@pedrohcgs

@pedrosantanna.bsky.social

# Backup

# How do we estimate the efficient DiD?

- The EIF is a **Neyman-orthogonal** moment, so estimation follows a standard plug-in / DML recipe:
    1. Estimate the nuisances — propensity scores $p_g(X)$ and outcome-change regressions $m_{\cdot,t,t'}(X)$ — by ML and cross-fitting.
    2. Estimate the covariance weights $V_{gt}(X)^{-1}$ (single) or $\Omega_{gt}(X)^{-1}$ (staggered) from the influence functions.
    3. Plug into the EIF-based estimand and average.
- Orthogonality $\Rightarrow$ first-stage error is **second-order**: inference uses the EIF's own variance, with no extra correction.
- Closed-form EIFs $\Rightarrow$ no bespoke variance derivation per estimand.

# Some remarks

- We focus on "fixed-$T$, large-$n$" panels.
- We expect our efficient estimators to work best when $\sqrt{n} \gg T$.
  - PT may be less plausible when $T$ is large.
  - When $T$ is large, there are probably better tools available.
- The degree of efficiency gains depends on the degree of over-identification (e.g. the number of parallel trends across periods).
  - We develop a simple Hausman-style over-identification test to assess the plausibility of PT; we also have visual tools.
  - You can incorporate linear, quadratic, or other known unit-specific trends, if you trust those parametric assumptions.

# Corollary: PT holding only in post-treatment periods

- Under PT-Post, we can only use the period immediately before treatment, $t' = g - 1$.
- The model is just-identified; the efficient influence function is simply $\mathbb{IF}_{g-1}^{att(g,t)}$.
- This generalizes the EIF in Sant'Anna and Zhao (2020), who focused on the much simpler two-period model.

# Calibrated simulations: CPS (single treatment date)

- Empirically-calibrated DGP building on Arkhangelsky, Athey, Hirshberg, Imbens and Wager (2021), single treatment date.
- Short panel $T = 7$ (four pre, three post); heterogeneous effects with $ATT = 0$; outcomes in logs.
- Target: $ES_{\text{avg}} = \big(ATT(5,5) + ATT(5,6) + ATT(5,7)\big)/3$.
- Compared with TWFE, Dynamic TWFE (DTWFE), and Synthetic DiD (SDiD).

# CPS Monte Carlo: relative RMSE and bias

| | Sample size | Relative RMSE | | | | Bias (×10) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | EDiD | TWFE | DTWFE | SDiD | EDiD | TWFE | DTWFE | SDiD |
| 1. Baseline | 50 | 1 | 3.57 | 12.46 | 1.53 | 0.01 | 0.60 | 2.58 | 0.00 |
| | 200 | 1 | 2.32 | 3.37 | 1.95 | 0.00 | -0.01 | 0.00 | 0.00 |
| *Outcome Model* | | | | | | | | | |
| 2. No corr | 50 | 1 | 3.52 | 12.33 | 1.45 | 0.02 | 0.59 | 2.46 | 0.00 |
| | 200 | 1 | 2.27 | 3.33 | 1.95 | -0.01 | 0.00 | -0.01 | 0.00 |
| 3. No M | 50 | 1 | 3.67 | 13.04 | 1.49 | -0.02 | 0.61 | 2.45 | 0.03 |
| | 200 | 1 | 2.17 | 3.00 | 1.68 | -0.01 | 0.02 | 0.00 | 0.01 |
| 4. No F | 50 | 1 | 1.47 | 1.88 | 1.42 | 0.00 | -0.02 | -0.04 | -0.02 |
| | 200 | 1 | 1.64 | 2.18 | 1.63 | 0.00 | 0.00 | -0.01 | 0.00 |
| *Treatment Assignment* | | | | | | | | | |
| 6. Gun law | 50 | 1 | 7.27 | 18.23 | 4.67 | 0.00 | -0.08 | -0.23 | -0.11 |
| | 200 | 1 | 8.70 | 12.94 | 6.05 | 0.00 | 0.03 | 0.02 | 0.02 |
| 7. Abortion | 50 | 1 | 6.99 | 17.19 | 4.77 | -0.01 | 0.55 | 1.75 | 0.33 |
| | 200 | 1 | 8.04 | 12.64 | 5.23 | 0.00 | -0.01 | -0.05 | 0.00 |
| *Outcome Variable* | | | | | | | | | |
| 9. Ln Hours | 50 | 1 | 1.01 | 1.92 | 0.95 | -0.34 | 0.12 | 1.53 | 0.02 |
| | 200 | 1 | 1.24 | 1.95 | 1.21 | 0.01 | 0.05 | 0.02 | 0.07 |
| 10. Ln U-rate | 50 | 1 | 0.82 | 1.44 | 0.82 | 0.73 | -0.24 | -0.36 | -0.26 |
| | 200 | 1 | 1.03 | 1.53 | 1.01 | 0.03 | 0.00 | 0.01 | 0.00 |

# Compustat Monte Carlo: relative RMSE and bias (staggered)

| | Relative RMSE | | | | Bias ($\times$10) | | | |
|---|---|---|---|---|---|---|---|---|
| $\rho$ | EDiD | BJS-G-W | CS-SA | CS-dCDH | EDiD | BJS-G-W | CS-SA | CS-dCDH |
| $\rho = 0$ | 1 | 1.02 | 1.62 | 1.56 | 0.00 | -0.01 | 0.00 | 0.00 |
| $\rho = 0.5$ | 1 | 1.06 | 1.28 | 1.24 | 0.01 | 0.01 | 0.00 | 0.00 |
| $\rho = 1$ | 1 | 1.19 | 1.09 | 1.04 | 0.04 | -0.02 | -0.01 | -0.01 |
| $\rho = -0.5$ | 1 | 1.05 | 2.39 | 2.30 | 0.00 | 0.00 | -0.01 | 0.00 |
| $\rho = -1$ | 1 | 1.62 | 3.24 | 3.43 | -0.01 | -0.01 | 0.00 | 0.00 |

- Efficiency gains persist across serial-correlation regimes $\rho$.

# When parallel trends is uncertain: a toolkit

- The efficiency bound assumes parallel trends across all groups and all periods (PT-All). When those extra restrictions—imposed beyond post-treatment parallel trends—are **uncertain**, the **same over-identification** delivers three tools:
  - **Test** — a Hausman / incremental-Sargan test of PT-All;
  - **Robustness frontier** — how far a headline contrast moves as the PT scope is relaxed, at a stated precision cost;
  - **Adaptive estimator** — smooth shrinkage trading precision against bias, with no pre-test discontinuity.
- Building on Chen and Santos (2018), Andrews et al. (2025), and Armstrong et al. (2024).

# Testing parallel trends across all groups

Hausman-type test of Assumption PT-All (PT-All): compare the efficient $\widehat{ES}$ (consistent and efficient under PT-All) against the just-identified $\widecheck{ES}$ (consistent under PT-Post alone).

$$\widehat{H} = n\left(\widehat{ES} - \widecheck{ES}\right)'\left(\widehat{\text{aCov}}(\widecheck{ES}) - \widehat{\text{aCov}}(\widehat{ES})\right)^{-1}\left(\widehat{ES} - \widecheck{ES}\right) \xrightarrow{d} \chi^2(|\mathcal{E}|).$$

- Reject PT-All when $\widehat{H}$ exceeds the $\chi^2(|\mathcal{E}|)$ critical value; nontrivial local power against local alternatives whose score has nonzero projection onto the span of the estimator-difference influence functions $\widehat{ES}(e) - \widecheck{ES}(e)$, $e \in \mathcal{E}$.
- Model is nonparametrically overidentified (Chen and Santos, 2018): the same objects that deliver the efficiency bound also deliver the test.
- **Incremental Sargan:** for each candidate $(g', t')$ with $g' > t'$ and $t'' = 1$ — each extending the PT-Post moment set by one restriction, $L$ extensions in all — compute a Hausman-type $p$-value $\mathrm{p}_{g',t'}$; order $\mathrm{p}_{(1)} \leq \cdots \leq \mathrm{p}_{(L)}$.
- Holm–Bonferroni step-down (Holm, 1979) at familywise rate $\alpha$ rejects $\mathrm{p}_{(\ell)}$ if $\mathrm{p}_{(\ell)} < \alpha/(L+1-\ell)$ and stops at the first non-rejection — admits extra restrictions only when the data do not reject them.

Building on Chen and Santos (2018); step-down via Holm (1979).

# Robustness frontiers for reported event-study contrasts

For a headline scalar $\theta = a'ES$ (e.g. $ES(e)$ or $ES_{\text{avg}}$), let $\widehat{\theta}_R = a'\widehat{ES}$ (efficient) and $\widehat{\theta}_U = a'\widecheck{ES}$ (conservative). With $\xi = \psi_U - \psi_R$, $D = \mathbb{E}[\xi^2]$, $V_R = \mathbb{E}[\psi_R^2]$, and $\widehat{\text{se}}(\widehat{\theta}_R) = \sqrt{\widehat{V}_R / n}$:

$$H_{\theta,n} = \frac{n(\widehat{\theta}_U - \widehat{\theta}_R)^2}{\widehat{D}} \xrightarrow{d} \chi_1^2 \quad \text{under PT-All (PT-All),} \qquad \boxed{\widehat{\theta}_R \pm \tau \sqrt{H_{\theta,n}}\, \widehat{\text{se}}(\widehat{\theta}_R)}$$

The boxed frontier holds up to $o_p(n^{-1/2})$. Under a PT-Post local alternative with score $s$, $H_{\theta,n} \xrightarrow{d} \chi_1^2(\{\mathbb{E}[\xi(W)s(W)]\}^2 / D)$: nontrivial local power iff $\mathbb{E}[\xi(W)s(W)] \neq 0$.

- **Frontier** = affine estimates $\widehat{\theta}_\lambda = \widehat{\theta}_R + \lambda(\widehat{\theta}_U - \widehat{\theta}_R)$ whose first-order variance stays $\leq (1 + \tau^2)V_R / n$ for a tolerance $\tau > 0$: how far the headline can move as PT scope is relaxed for a precision cost $\tau$.
- $\psi_R$ efficient under PT-All gives $\mathbb{E}[\psi_R \xi] = 0$, so the first-order variance is exactly $(V_R + \lambda^2 D) / n$.
- Inspired by the Andrews et al. (2025) reading of Hansen's $J$ statistic in finite-dimensional GMM; with continuous covariates the conditional-PT violations are generally infinite-dimensional, so $H_{\theta,n}$ is the finite-dimensional reported-parameter Hausman diagnostic (Hausman, 1978) — the overidentification-failure component that moves the reported object.
- **Report:** efficient $\widehat{\theta}_R$, conservative $\widehat{\theta}_U$, $\sqrt{H_{\theta,n}}$ (p-value), and frontier radii for $\tau \in \{0.25, 0.5, 1\}$.

# Robustness frontiers: scope and finite-menu extension

- **Scope.** The frontier does not bound movement under arbitrary parallel-trends violations; it quantifies sensitivity to relaxing the stronger restrictions at a transparent precision cost. For honest CIs under arbitrary violations, the Rambachan and Roth (2022) bounds are complementary to this diagnostic.
- **Finite-menu extension.** To compare $K$ pre-specified baseline/comparison configurations $\widehat{\theta}_1, \ldots, \widehat{\theta}_K$, set $\widehat{d} = (\widehat{\theta}_1 - \widehat{\theta}_R, \ldots, \widehat{\theta}_K - \widehat{\theta}_R)'$ with $D_K = \mathrm{Var}(\sqrt{n}\,\widehat{d})$.

$$J_{\theta,K} = n\,\widehat{d}'\,\widehat{D}_K^{+}\,\widehat{d} \xrightarrow{d} \chi^2_r, \quad r = \mathrm{rank}(D_K) \leq K \ \text{ under PT-All}, \qquad \widehat{\theta}_R \pm \tau\sqrt{J_{\theta,K}}\,\widehat{\mathrm{se}}(\widehat{\theta}_R).$$

- $\widehat{D}_K^{+}$ is the Moore–Penrose pseudoinverse; rank deficiency ($r < K$) arises naturally when alternatives share overlapping baseline/comparison choices.
- When $K = 1$ this reduces to the scalar frontier $\widehat{\theta}_R \pm \tau\sqrt{H_{\theta,n}}\,\widehat{\mathrm{se}}(\widehat{\theta}_R)$.
- Building on the overidentified-model logic of Chen and Santos (2018); plotting the frontier alongside the efficient CI, with $\widehat{\theta}_U$ marked, gives a compact sensitivity summary.

# Adaptive estimation under uncertain parallel trends

Pre-testing $\widehat{ES}_{\text{avg}}$ vs. $\widecheck{ES}_{\text{avg}}$ is suboptimal: hard thresholding pays a variance discontinuity at the hypothesis boundary (Roth, 2022). Take the restricted $\widehat{ES}_{\text{avg}}$ (efficient under PT-All; biased if PT-All fails) and the unrestricted $\widecheck{ES}_{\text{avg}}$ (consistent under PT-Post), with relative-efficiency ratio $\widehat{\rho}^{\,2}$ and overidentification statistic $\widehat{T}_O$:

$$\widehat{ES}^{\text{AKS}}_{\text{avg}} = \widehat{\rho}\,\widehat{\sigma}_U\,\delta^*\left(\widehat{T}_O;\, \widehat{\rho}^{\,2}\right) + \widecheck{ES}_{\text{avg}} - \widehat{\rho}\,\widehat{\sigma}_U\,\widehat{T}_O,$$

$$\widehat{T}_O = \frac{\widecheck{ES}_{\text{avg}} - \widehat{ES}_{\text{avg}}}{\widehat{\sigma}_O}, \quad \widehat{\rho}^{\,2} = \frac{\widehat{\sigma}_R^2}{\widehat{\sigma}_U^2} \in (0,1), \quad \widehat{\sigma}_O^2 = \widehat{\sigma}_U^2 - \widehat{\sigma}_R^2 > 0.$$

- Smooth shrinkage: small $\widehat{T}_O \Rightarrow$ weight on the restricted (efficient) estimator; large $\widehat{T}_O \Rightarrow$ weight on the unrestricted one.
- $\delta^*(\cdot;\rho^2)$ is the smooth shrinkage of Armstrong et al. (2024, Thm 4.1(ii)): it minimizes the worst-case actual-to-oracle MSE ratio over all bias bounds $|B| \in [0,\infty]$ simultaneously, so MSE stays controlled relative to an oracle that knows the violation magnitude—avoiding the pre-test discontinuity.
- Requires the scalar pair $(\widecheck{ES}_{\text{avg}}, \widehat{ES}_{\text{avg}})$ to be jointly asymptotically bivariate normal with consistently estimable covariance, where $\widehat{\sigma}_R^2 = \widehat{\text{aCov}}(\widehat{ES}_{\text{avg}})$, $\widehat{\sigma}_U^2 = \widehat{\text{aCov}}(\widecheck{ES}_{\text{avg}})$. The construction is bivariate by design; a per-$e$ analog applies to each $ES(e)$ with the usual multiple-testing caveats.

# References

**Ai, Chunrong and Xiaohong Chen**, "The semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions," *Journal of Econometrics*, 2012, *170* (2), 442–457.

**Andrews, Isaiah, Jiafeng Chen, and Otavio Tecchio**, "The purpose of an estimator is what it does: Misspecification, estimands, and over-identification," *arXiv:2508.13076*, 2025. Forthcoming, Econometric Society World Congress monograph.

**Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager**, "Synthetic Difference-in-Differences," *American Economic Review*, 2021, *111* (12), 4088–4118.

**Armstrong, Timothy B., Patrick Kline, and Liyang Sun**, "Adapting to Misspecification," *arXiv:2305.14265v4*, 2024.

**Borusyak, Kirill, Xavier Jaravel, and Jann Spiess**, "Revisiting Event Study Designs: Robust and Efficient Estimation," *Review of Economic Studies*, 2024, *91* (6), 3253–3285.

**Callaway, Brantly and Pedro H. C. Sant'Anna**, "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

**Chen, Xiaohong and Andres Santos**, "Overidentification in Regular Models," *Econometrica*, 2018, *86* (5).

**de Chaisemartin, Clément and Xavier D'Haultfœuille**, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

**Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew J. Notowidigdo**, "The economic consequences of hospital admissions," *American Economic Review*, 2018, *108* (2), 308–352.

**Gardner, John**, "Two-stage differences in differences," *Working Paper*, 2021.

**Hausman, Jerry A.**, "Specification Tests in Econometrics," *Econometrica*, 1978, *46* (6), 1251–1271.

**Holm, Sture**, "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 1979, *6* (2), 65–70.

**Rambachan, Ashesh and Jonathan Roth**, "Design-Based Uncertainty for Quasi-Experiments," 2022. arXiv:2008.00602 [econ, stat].

**Roth, Jonathan**, "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends," *American Economic Review: Insights*, 2022, *4* (3), 305–322.

**Sant'Anna, Pedro H. C. and Jun Zhao**, "Doubly robust difference-in-differences estimators," *Journal of Econometrics*, 2020, *219* (1), 101–122.

**Sun, Liyang and Sarah Abraham**, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

Wooldridge, Jeffrey M, "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators," *Working Paper*, 2021, pp. 1–89.