# Modern Difference-in-Differences:

## Better understanding some of the recent advances

**Pedro H. C. Sant'Anna**
Emory University

NABE TEC, October 27 2024

# Causal inference with observational data: What can we do?

- In many applications/situations, we do not have experimental data.
- Without an experiment, we need to rely on **observational data**.
- With **observational data**, we have no choice but rely on **additional assumptions** to talk about causal inference.
- Different methods rely on different identification assumptions.
- Examples: DML, DiD, SC, IV, RDD.

# Recent boom of DiD methods

In the last years, we have seen many methodological advances in DiD: (by no means an exhaustive list)

- Athey and Imbens (2022)
- Borusyak, Jaravel and Spiess (2024)
- de Chaisemartin and D'Haultfœuille (2020)
- Goodman-Bacon (2021)
- Sun and Abraham (2021)

→ "Reverse Engineering" causal interpretations for TWFE regressions and propose some alternative DiD estimators

- Callaway and Sant'Anna (2021)
- Sant'Anna and Zhao (2020)
- Lee and Wooldridge (2023)
- Wooldridge (2021)

→ "Forward Engineering" DiD estimators conditional on covariates

- Rambachan and Roth (2023)
- Roth (2022)

→ Issues with pre-tests and how to handle PT as approximation

- Roth and Sant'Anna (2023a,b)

→ Sensitivity to functional form and random treatment timing

- Callaway, Goodman-Bacon and Sant'Anna (2024a)
- de Chaisemartin, D'Haultfoeuille, Pasquier, Sow and Vazquez-Bare (2024)

→ DiD with continuous and multi-valued treatments

- Ghanem, Sant'Anna and Wüthrich (2022)
- Marx, Tamer and Tang (2023)

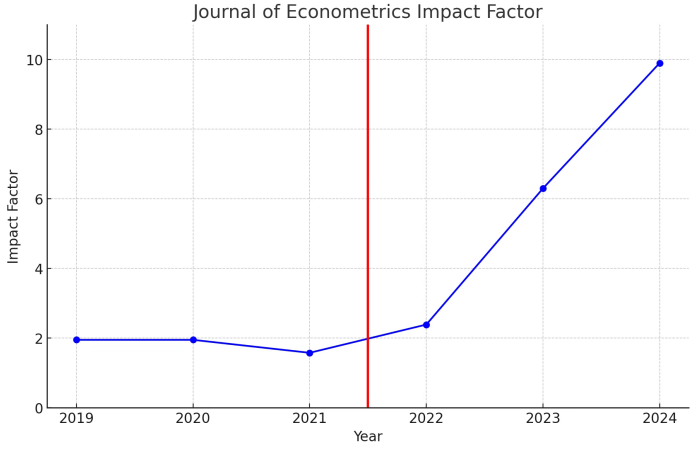
→ Better understanding PT and selection

- Callaway (2021)
- Callaway and Li (2019)
- Tchetgen Tchetgen, Park and Richardson (2024)
- Wooldridge (2023)

→ Nonlinear DiD Models

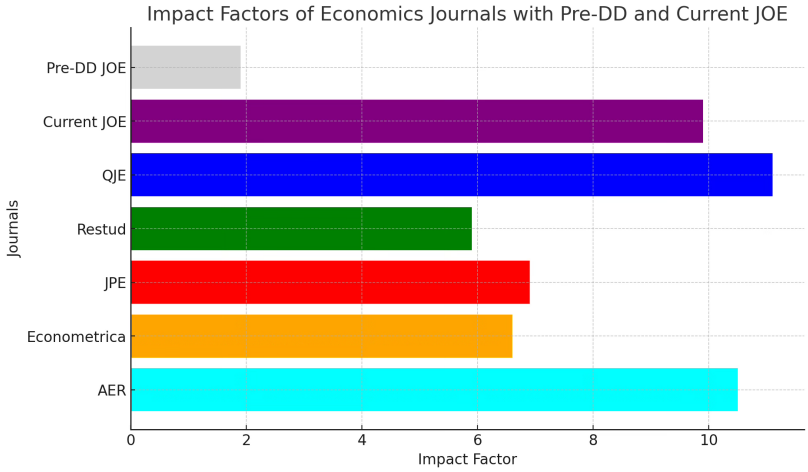
# Academic impact of Difference-in-Differences papers

Scott Cunningham documented the evolution of the Journal of Econometrics Impact Factor.



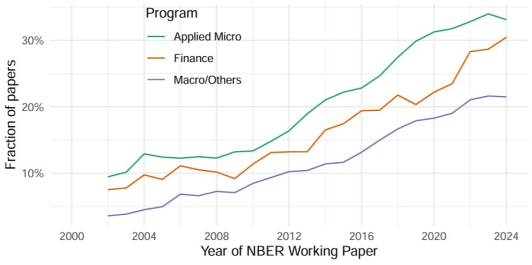
# Academic impact of Difference-in-Differences papers

Scott Cunningham compared impact factors across journals, too

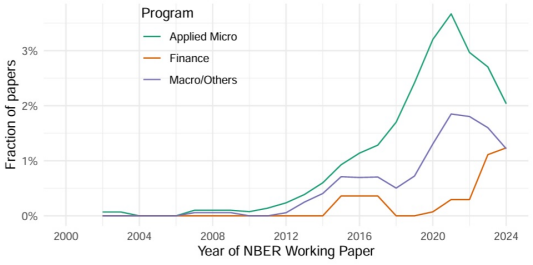


# Popularity of Difference-in-Differences methods

Goldsmith-Pinkham (2024) built on Currie, Kleven and Zwiers (2020) and document the popularity of DiD within economics.




(a) Difference-in-differences

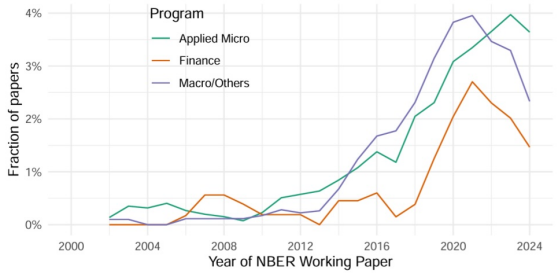



(b) Synthetic controls

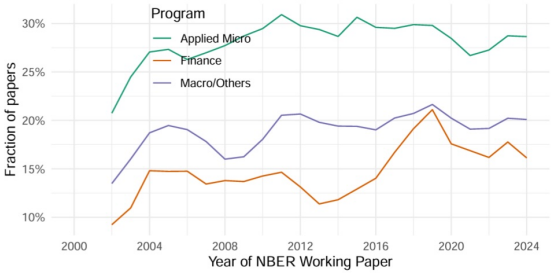
Figure 5: Panel (a) reports the share of papers that mention difference-in-differences or event studies. Figure (b) reports the share of papers that mention synthetic controls (this includes both synthetic difference-in-differences and synthetic control methods). See Table 2 for the breakdown of fields, and the Appendix for definitions on keywords.

# Popularity of Difference-in-Differences methods

Goldsmith-Pinkham (2024): Compare previous plot with IV




(a) Bartik and shift-share instruments




(b) Instrumental variables

Figure 6: Panel (a) reports the share of papers that mention Bartik or shift-share instruments. Figure (b) reports the share of papers that mention instrumental variables. See Table 2 for the breakdown of fields, and the Appendix for definitions on keywords.

# What can explain the empirical popularity of DiD methods?

- Mild data requirements: You need data from before and after the intervention and groups of units that are and are not exposed to the intervention in a given post-treatment period.
- Allows for selection on unobservables: PT is compatible with some types of selection on unobservables.
- Easy-to-use software: Traditional and modern DiD tools are available in Stata, R, and Python, including the very recent ones.
- Possibility to assess the plausibility of PT: We often observe severe pre-treatment periods that we can use to assess the plausibility of PT.
- We can learn about treatment effect dynamics: We can leverage event-study aggregations to understand better how treatment effects evolve with elapsed treatment timing.

# Starting our DiD journey

# 2 x 2 DiD procedures

- Most of us studied DiD using the canonical $2 \times 2$ setup:
  - 2 periods: $t = 1$ (pre-treatment), and $t = 2$ (post-treatment).
  - 2 groups: one group that is first treated in $t = 2$ ($G = 2$), and one group that remains untreated by $t = 2$ ($G = \infty$).
  - Parameter of interest: The average treatment effect among treated units in period $t = 2$,

$$ATT = \mathbb{E}\left[Y_{t=2}(2) - Y_{t=2}(\infty) | G = 2\right].$$

  - Parallel Trends Assumption: In the absence of treatment, the average evolution of untreated outcomes would be the same across the two groups,

$$\mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | G = 2\right] = \mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | G = \infty\right].$$

  - No-Anticipation Assumption: Before treatment starts, units, on average, act as if they are untreated,

$$\mathbb{E}\left[Y_{t=1}(2) | G = 2\right] = \mathbb{E}\left[Y_{t=1}(\infty) | G = 2\right].$$

# Identification

- Under parallel trends and no anticipation, it follows that

$$ATT = \underbrace{(\mathbb{E}\left[Y_{t=2}|G=2\right] - \mathbb{E}\left[Y_{t=1}|G=2\right])}_{\text{Average change for treated group}} - \underbrace{(\mathbb{E}\left[Y_{t=2}|G_i=\infty\right] - \mathbb{E}\left[Y_{t=1}|G=\infty\right])}_{\text{Average change for comparison group}},$$

  a "difference-in-differences" of population means.

# Estimation and Inference

- The most conceptually simple estimator replaces population means with sample analogs:

$$\widehat{ATT}_{DiD} = (\overline{Y}_{g=2,t=2} - \overline{Y}_{g=2,t=1}) - (\overline{Y}_{g=\infty,t=2} - \overline{Y}_{g=\infty,t=1})$$

  where $\overline{Y}_{g,t}$ is sample mean for group $g$ in period $t$.

- Conveniently, $\widehat{ATT}_{DiD}$ is algebraically equal to OLS coefficient $\hat{\beta}$ from either of the TWFE regression specifications

$$\begin{aligned} Y_{i,t} &= \alpha_i + \phi_t + D_{i,t}\beta + \epsilon_{i,t}, \\ Y_{i,t} &= \alpha_g + \phi_t + D_{i,t}\beta + \epsilon_{i,t}, \end{aligned}$$

  where $D_{i,t} = 1[G_i = 2] \times 1[t = 2]$.

# Does TWFE “work” in setups with variation in treatment timing?

# Traditional methods: TWFE regressions

- We know that, in the 2x2 case,

$$Y_{i,t} = \alpha_0 + \gamma_0 1\{G_i = 2\} + \lambda_0 1\{T_i = 2\} + \underbrace{\beta}_{\equiv ATT}(1\{G_i = 2\} \cdot 1\{T_i = 2\}) + \varepsilon_{i,t},.$$

- It is tempting to "extrapolate" from this setup and use variations of the following TWFE specification to estimate causal effects:

$$Y_{i,t} = \alpha_i + \alpha_t + \beta \cdot D_{i,t} + \varepsilon_{i,t},$$

  where dummies $D_{i,t} = 1\{t - G_i \geq 0\}$, where $G_i$ indicates the period unit $i$ is first treated (Group).

- $D_{i,t}$ is an indicator for unit $i$ being ever-treated by period $t$.

- Consider the case where treatment does not turn off: **staggered treatment adoption**.

# Does TWFE “work” in setups with variation in treatment timing?

## Reverse Engineering a Causal Interpretation

# TWFE with staggered adoption: what causal parameter does it recover?

- Does $\beta$ recover any interesting causal parameter of interest?
  - Athey and Imbens (2022), Borusyak et al. (2024), de Chaisemartin and D'Haultfœuille (2024), and Goodman-Bacon (2021) tackle this "reverse-engineering" question (Mogstad and Torgovitsky, 2024).
- Under PT, when treatment effects are heterogeneous/dynamic, $\beta$ does not recover an easy-to-interpret parameter.
- $\beta$ equals a weighted average of ATTs across groups and times, but some weights can be negative.
- Even when weights are non-negative, they are not really intuitive/policy-oriented.
- In my opinion, Goodman-Bacon (2021) explains this most clearly.
  - The TWFE regression specification does not respect the identification assumptions and uses "already-treated" units as a comparison group for "later-treated" units.
  - The weights also change if we were gifted extra pre-treatment periods, for example.

# Event-Study via TWFE specifications

- Treatment effect dynamics is a main source of problems for the "static" TWFE regression.
- But, if that is the case, maybe we could bypass those problems by modeling TE dynamics using variants of the TWFE event-study regression

$$Y_{i,t} = \alpha_i + \alpha_t + \gamma_k^{-K} D_{i,t}^{<-K} + \sum_{k=-K}^{-2} \gamma_k^{lead} D_{i,t}^k + \sum_{k=0}^{L} \gamma_k^{lags} D_{i,t}^k + \gamma_k^{L+} D_{i,t}^{>L} + \varepsilon_{i,t}$$

with the event study dummies $D_{i,t}^k = 1\{t - G_i = k\}$, where $G_i$ indicates the period unit $i$ is first treated (Group).

- $D_{i,t}^k$ is an indicator for unit $i$ being $k$ periods away from initial treatment at time $t$.

# Does this strategy “work”?

# Problems with Event-Study via TWFE specifications: Sun and Abraham (2021)

- Sun and Abraham (2021) bring "bad" news, once again.
  - Even when we impose a parallel trends across all periods and groups and the no-anticipation assumption, the OLS coefficients of the TWFE ES specification are, in general, very hard to interpret.
  - Coefficient on a given lead or lag can be contaminated by effects from other periods.
  - Pre-trends can arise solely from treatment effects heterogeneity.
  - Even under treatment effect homogeneity across cohorts (all share the same dynamics in event-time), the OLS coefficients can still be contaminated by treatment effects from the excluded periods.
- Goldsmith-Pinkham, Hull and Kolesár (2024) show that this contamination problem hold more generally.

# Forward Engineering DiD estimators

# What if we did DiD by hand?

- In the $2 \times 2$ model, we had two paths: TWFE and DiD-by-hand.
- We have already seen that the traditional TWFE specifications would not work well for us.
- A viable alternative is to return to the basics and do DiD-by-hand.
  - Fix the family of target parameters, for example, all post-treatment $ATT(g,t)$'s:

$$ATT(g,t) = \mathbb{E}\left[Y_t(g) - Y_t(\infty) \,|G = g\right], \; for \; t \geq g.$$

    $ATT(g,t)$: Average Treatment Effect at time $t$ of starting treatment at time $g$, among the units that started treatment at time $g$.
  - State your identification assumptions: a type of parallel trends and no-anticipation assumptions.
  - Based on these, get the identification result for the $ATT(g,t)$'s, which should guide the choice of estimation method.
  - Aggregate $ATT(g,t)$'s to form informative and easy-to-interpret causal summary parameters.

# Forward Engineering DiD estimators

## Identification Assumptions

# Identifying Assumptions: No-Anticipation

- Given that we never observe $Y_t(\infty)$ in post-treatment periods among units that have been treated, we need to make assumptions to identify $ATT(g,t)$'s
- **No-Anticipation Assumption**: For all $i,t$ and $t < g, g'$, $Y_{i,t}(g) = Y_{i,t}(g')$.
- Unit treatment effects are zero before treatment takes place.
- This plays a role in the definition of treatment date:
  date of announcement or date of implementation?

# Parallel Trends for all groups and periods

- Contrary to the $2 \times 2$ DiD setup, with variation in treatment timing, one can assume different types of Parallel Trends to hold.
- The choice of PT restricts which comparison group and which pre-treatment periods you can leverage for identifying $ATT(g, t)$
- This is an important difference between different modern DiD estimators.
- Below we impose parallel trends hold across all periods and across all treatment groups:

**Assumption (Parallel Trends for all groups and periods)**

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

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

# ATT(g,t) Estimand: “never-treated” as effective comparison group

- Under no-anticipation and PT as defined above, we can identify the $ATT(g,t)$

$$ATT^{nev}_{unc}(g,t) = \mathbb{E}[Y_{t=t} - Y_{t=g-1}|G=g] - \mathbb{E}[Y_{t=t} - Y_{t=g-1}|G=\infty],$$

  which uses “never-treated” units $G=\infty$ as the effective comparison group for all units and also uses the latest pre-treatment period as the effective baseline period.

- This looks very similar to the $2 \times 2$ DiD formula.

- The difference is now we take a “long difference”.

- This result appears in Callaway and Sant’Anna (2021) and Sun and Abraham (2021).

- Under the PT assumption for all periods and groups, one can also use any other pre-treatment period $t_{pre} < g$ as the baseline period and other cohorts $g' > t$ as the comparison groups, which is implicitly what Borusyak et al. (2024) and Wooldridge (2021) estimands do. See also Lee and Wooldridge (2023) and Marcus and Sant’Anna (2021).

# ATT(g,t) Estimand: not-yet treated as effective comparison group

- If one wants to use all units that have not yet been exposed to treatment by time $t$ as the effective comparison group, we have a different estimand for the same target parameter:

$$ATT^{ny}_{unc}(g,t) = \mathbb{E}[Y_{t=t} - Y_{t=g-1}|G=g] - \mathbb{E}[Y_{t=t} - Y_{t=g-1}|D_t = 0, G \neq g].$$

- The difference from $2 \times 2$ is that now we take a "long difference", and that the comparison group changes over time.

- Same intuition carries, though!

- This result appears in Callaway and Sant'Anna (2021) and de Chaisemartin and D'Haultfœuille (2020), though de Chaisemartin and D'Haultfœuille (2020) focus exclusively on instantaneous treatment effects—they rule out treatment effect dynamics.

- One can also use multiple pre-treatment periods as baseline periods; see, e.g., Borusyak et al. (2024), Lee and Wooldridge (2023), Marcus and Sant'Anna (2021), Wooldridge (2021).

# Forward Engineering DiD estimators

How to handle covariates in staggered DiD applications?

# How to handle covariates in DiD?

- Callaway and Sant'Anna (2021) discussed three different strategies to account for covariates:
  - Regression adjustment (Heckman, Ichimura and Todd, 1997);
  - Inverse-probability weighting (IPW) (Abadie, 2005);
  - Doubly Robust (Sant'Anna and Zhao, 2020).
- Adding covariates linearly into the TWFE will not give you the $ATT(g,t)$'s.

$$Y_{i,t} = \alpha_i + \alpha_t + \gamma_k^{-K} D_{i,t}^{<-K} + \sum_{k=-K}^{-2} \gamma_k^{lead} D_{i,t}^k + \sum_{k=0}^{L} \gamma_k^{lags} D_{i,t}^k + \gamma_k^{L+} D_{i,t}^{>L} + X_{i,t}'\beta + \varepsilon_{i,t}$$

with the event study dummies $D_{i,t}^k = 1\{t - G_i = k\}$.

# Being inspired by the recent developments in Causal ML

- In the last 10 years or so, we have been seeing a lot of advances in Causal ML.
  - Belloni, Chernozhukov and Hansen (2014)
  - Farrell (2015)
  - Belloni, Chernozhukov, Fernández-Val and Hansen (2017),
  - Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey and Robins (2017)
  - Athey and Wager (2018)
  - Athey, Tibshirani and Wager (2019)
  - Chernozhukov, Demirer, Duflo and Fernández-Val (2022).
- All these papers propose estimators that are Doubly Robust/Neyman Orthogonal.
- These ideas have been explored in DiD setups only recently;
  see, e.g., Sant'Anna and Zhao (2020); Chang (2020); Callaway and Sant'Anna (2021).

# Doubly Robust DiD procedure with Panel

Callaway and Sant'Anna (2021) proposed the following DR DiD estimand for staggered setups:

$$ATT^{dr,p}(g,t) = \mathbb{E}\left[\left(\frac{1\{G=g\}}{\mathbb{E}\left[1\{G=g\}\right]} - \frac{\frac{p_{g,t}(X)1\{G>t\}}{1-p_{g,t}(X)}}{\mathbb{E}\left[\frac{p_{g,t}(X)1\{G>t\}}{1-p_{g,t}(X)}\right]}\right)\left(Y_t - Y_{g-1} - m^{G>t}_{Y_t-Y_{g-1}}(X)\right)\right],$$

where

$$p_{g,t}(X) = P(G=g|X, 1\{G=g\}+1\{G>t\}=1), \qquad m^{G>t}_{Y_t-Y_{g-1}}(X) = \mathbb{E}[Y_t - Y_{g-1}|G>t, X].$$

- This is similar to cross-sectional DR formulation but with outcomes measured as "post - pre" instead of "post" and the focus is on ATT, not ATE-type of parameters
- Estimation follows from the plug-in principle, and you may need cross-fitting.

# Forward Engineering DiD estimators

## Aggregation

# Aggregation

# Summarizing ATT(g,t): Event-study / dynamic treatment effects

- The effect of a policy intervention may depend on the length of exposure to it.
- Average effect of participating in the treatment for the group of units that have been exposed to the treatment for exactly $e$ time periods

$$\theta_D(e) = \sum_{g=2}^{T} \mathbf{1}\{g + e \leq T\} ATT(g, g+e) P(G = g | G + e \leq T)$$

- This is perhaps the most popular summary measure currently adopted by empirical researchers.

# Some new DiD tools that I am currently working on

# Heterogeneity analysis

- In many applications, we would like to "deep dive" and better understand how treatment effects vary across covariate strata.
- In cross-sectional analysis, this is usually done by analyzing conditional average treatment effects, $CATE(X_{\text{sub}}) = \mathbb{E}[Y(1) - Y(0)|X_{\text{sub}}]$.
- In staggered DiD setups, the parameter of interest would be

$$CATT(g, t, X_{\text{sub}}) = \mathbb{E}[Y_t(g) - Y_t(\infty)|G = g, X_{\text{sub}}].$$

- We can extend the IPW procedure of Abadie (2005) to the staggered setup.
- More challenging (and interesting) is to derive a DR estimator for the $CATT(g, t, X_{\text{sub}})$ and its best linear approximation. We are doing this in Callaway, Chen and Sant'Anna (2024b).

# Exploring all the content of PT

- The estimators discussed today leverage one pre-treatment period, $t = g - 1$, and uses all the not-yet-treated as the comparison group.

- However, recall that we have assumed that PT holds for all periods and all groups,

**Assumption (Parallel Trends for all groups and periods)**

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

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

- Under this type of PT, we can do better than the currently available DiD estimators.

# Exploring all the content of PT

In Chen, Sant'Anna and Xie (2024), we show the following result:

## Lemma

*Under PT as above, overlap, and no-anticipation, 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,*

$$CATT(g,t,X) = \underbrace{\mathbb{E}[Y_t - Y_{t'}|G = g, X]}_{\equiv m_{g,t,t'}(X)} - \Big( \underbrace{\mathbb{E}[Y_t - Y_{t''}|G = \infty, X]}_{\equiv m_{\infty,t,t''}(X)} + \underbrace{\mathbb{E}[Y_{t''} - Y_{t'}|G = g', X]}_{\equiv m_{g',t'',t'}(X)} \Big), \tag{1}$$

*and, as a consequence,*

$$ATT(g,t) = \mathbb{E}\big[Y_t - Y_{t'}\big|G = g\big] - \mathbb{E}\big[\left(m_{\infty,t,t''}(X) + m_{g',t'',t'}(X)\right)\big|G = g\big]. \tag{2}$$

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

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

# Exploring all the content of PT

- In Chen et al. (2024), we also derive how one should choose the weights such that the DiD estimator is asymptotically efficient (under some regularity conditions).
- This implies that the resulting estimator would have (asymptotically) shorted confidence intervals, leading to more informative inference procedures.
- We also derive efficient estimators for event-study aggregations.
- Stay tuned for the WP!

# Take-way messages

# DiD procedures multiple time periods

- With multiple time periods and variations in treatment timing, TWFE does not respect our assumptions:
- The solution to the TWFE problem is simple:
  - Separate the identification, aggregation and estimation/inference parts of the problem
- Use $ATT(g,t)$ as building blocks so we can transparently see how target parameters are constructed.
- Many different aggregation schemes are possible: they deliver different parameters!
- Can allow for covariates via regressions adjustments, IPW, and DR.

# Difference-in-Differences Checklist

1. Start plotting the treatment rollout (e.g., use panelView R package)
2. Document how many units are treated in each cohort.
3. Plot the evolution of average outcomes across cohorts.
4. Choose the comparison groups and the PT assumption carefully:
   Who decides treatment? What do they know? What type of selection is allowed?
5. Do event-study analysis without any covariates and assess if PT is plausible.
6. If unconditional PT is not plausible, incorporate covariates into the analysis.
7. When using covariates, check for overlap: If control groups are small, problems with overlap will probably arise. If you are OK with extrapolation, use regression adjustment DiD procedures.
8. Do event-study analysis after adjusting for covariates and assess if conditional PT is plausible.
9. Conduct some sensitivity analysis for violations of PT (e.g., use honestDiD R package).
10. If conditional PT is not plausible, look for other methods.

# References

**Abadie, Alberto**, "Semiparametric Difference-in-Differences Estimators," *The Review of Economic Studies*, 2005, *72* (1), 1–19.

**Athey, Susan and Guido Imbens**, "Design-Based Analysis in Difference-In-Differences Settings with Staggered Adoption," *Journal of Econometrics*, 2022, *226* (1), 62–79.

_ **and Stefan Wager**, "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, 2018, *113* (523), 1228 – 1242.

_ **, Julie Tibshirani, and Stefan Wager**, "Generalized random forests," *The Annals of Statistics*, 2019, *47* (2), 1148 – 1178.

**Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen**, "Inference on Treatment Effects after Selection among High-Dimensional Controls," *The Review of Economic Studies*, apr 2014, *81* (2), 608–650.

_ **,** _ **, Iván Fernández-Val, and Christian Hansen**, "Program Evaluation and Causal Inference With High-Dimensional Data," *Econometrica*, 2017, *85* (1), 233–298.

**Borusyak, Kirill, Xavier Jaravel, and Jann Spiess**, "Revisiting Event Study Designs: Robust and Efficient Estimation," *Review of Economic Studies*, 2024, *Forthcoming*.

**Callaway, Brantly**, "Bounds on distributional treatment effect parameters using panel data with an application on job displacement," *Journal of Econometrics*, 2021, *222* (2), 861–881.

**_ and Pedro H. C. Sant'Anna**, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

**_ and Tong Li**, "Quantile treatment effects in difference in differences models with panel data," *Quantitative Economics*, 2019, *10* (4), 1579–1618.

**_ , Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna**, "Difference-in-Differences with a Continuous Treatment," *arXiv:2107.02637 [econ]*, 2024.

**_ , Xiaohong Chen, and Pedro H. C. Sant'Anna**, "Heterogeneous Treatment Effects in Difference in Differences Designs," *Work-in-Progress*, 2024.

**Chang, Neng-Chieh**, "Double/debiased machine learning for difference-in-differences models," *The Econometrics Journal*, 2020, *23* (2), 177––191.

**Chen, Xiaohong, Pedro H. C. Sant'Anna, and Haitian Xie**, "Efficient Difference-in-Differences and Event Study Estimators," *Working Paper*, 2024.

**Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins**, "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, jun 2017, pp. 1–71.

**_ , Mert Demirer, Esther Duflo, and Iván Fernández-Val**, "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments ," *arXiv:1712.04802*, 2022.

**Currie, Janet, Henrik Kleven, and Esmée Zwiers**, "Technology and Big Data Are Changing Economics: Mining Text to Track Methods," *AEA Papers and Proceedings*, May 2020, *110*, 42–48.

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

**_ and _** , "Difference-in-Differences Estimators of Intertemporal Treatment Effects," *The Review of Economics and Statistics*, 2024, *Forthcoming*, 1–45.

**_ , Xavier D'Haultfoeuille, Félix Pasquier, Doulo Sow, and Gonzalo Vazquez-Bare**, "Difference-in-Differences for Continuous Treatments and Instruments with Stayers," *arXiv:2201.06898v4*, 2024.

**Farrell, Max H.**, "Robust inference on average treatment effects with possibly more covariates than observations," *Journal of Econometrics*, 2015, *189* (1), 1–23.

**Ghanem, Dalia, Pedro H. C. Sant'Anna, and Kaspar Wüthrich**, "Selection and parallel trends," *arXiv:2203.09001[econ]*, 2022.

**Goldsmith-Pinkham, Paul**, "Tracking the Credibility Revolution across Fields," *arXiv:2405.20604*, 2024.

**_ , Peter Hull, and Michal Kolesár**, "Contamination Bias in Linear Regressions," *American Economic Review*, 2024, *Forthcoming*.

**Goodman-Bacon, Andrew**, "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 2021, *225* (2).

**Heckman, James J., Hidehiko Ichimura, and Petra E. Todd**, "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, October 1997, *64* (4), 605–654.

**Lee, Soo Jeong and Jeffrey M. Wooldridge**, "A Simple Transformation Approach to Difference-in-Differences Estimation for Panel Data," *Working Paper*, 2023. Available at SSRN: http://dx.doi.org/10.2139/ssrn.4516518.

**Marcus, Michelle and Pedro H. C. Sant'Anna**, "The role of parallel trends in event study settings : An application to environmental economics," *Journal of the Association of Environmental and Resource Economists*, 2021, *8* (2), 235–275.

**Marx, Philip, Elie Tamer, and Xun Tang**, "Parallel Trends and Dynamic Choices," *Journal of Political Economy: Microeconomics*, 2023, *Forthcoming*.

**Mogstad, Magne and Alexander Torgovitsky**, "Instrumental variables with unobserved heterogeneity in treatment effects," Working Paper 32927, National Bureau of Economic Research, Cambridge, MA September 2024.

**Rambachan, Ashesh and Jonathan Roth**, "A More Credible Approach to Parallel Trends," *Review of Economic Studies*, 2023, *90* (5), 2555–2591.

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

**_ and Pedro H. C. Sant'Anna**, "When Is Parallel Trends Sensitive to Functional Form?," *Econometrica*, 2023, *91* (2), 737–747.

**_ and Pedro H.C. Sant'Anna**, "Efficient Estimation for Staggered Rollout Designs," *Journal of Political Economy: Microeconomics*, 2023, *1* (4), 669––709.

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

**Sun, Liyan and Sarah Abraham**, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2).

**Tchetgen, Eric J. Tchetgen, Chana Park, and David B. Richardson**, "Universal Difference-in-Differences for Causal Inference in Epidemiology," *Epidemiology*, 2024, *35* (1), 16–22.

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

**Wooldridge, Jeffrey M.**, "Simple approaches to nonlinear difference-in-differences with panel data," *Econometrics Journal*, 2023, *Forthcoming*.