Advances in the Difference-in-Difference Literature

Matt Bombyk @ APEC Skills Workshop

December 6, 2023

Introduction

- The Difference-in-Difference method is a simple, intuitive, and powerful method for causal inference with observational panel data, used especially in policy evaluation.
- DID and its extensions, Regression DD (using two-way fixed effects or TWFE) and regression-based event studies, are extremely popular methods in applied microeconomics.
  - Used in 26 of the 100 most cited papers published by the AER from 2015-2019 (de Chaisemartin & D'Haultfoeuille, 2022)
- Since around 2017, a literature has been building that points out serious flaws with these methods in many cases
- Numerous solutions have been proposed

Introduction

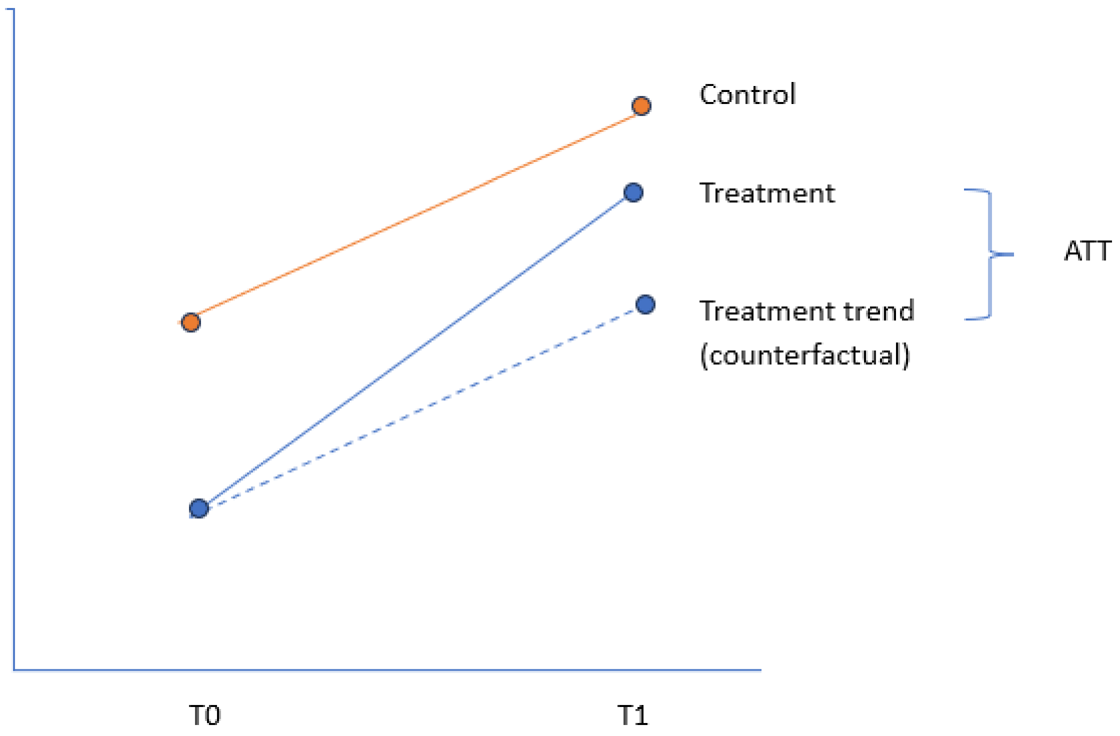
- In this presentation, I will walk through the basic issues, and proposed solutions.
- I will focus on three key papers in the literature, all published in the *Journal of Econometrics*
    1. Goodman-Bacon (2021) – Difference-in-differences with variation in treatment timing
    2. Sun and Abraham (2020) – Estimating dynamic treatment effects in event studies with heterogeneous treatment effects
    3. Callaway and Sant'Anna (2021) – Difference-in-Differences with multiple time periods
- The end of the presentation has an extensive list of resources for further study, including papers, software packages and online resources.

Basic Diff-in-Diff model

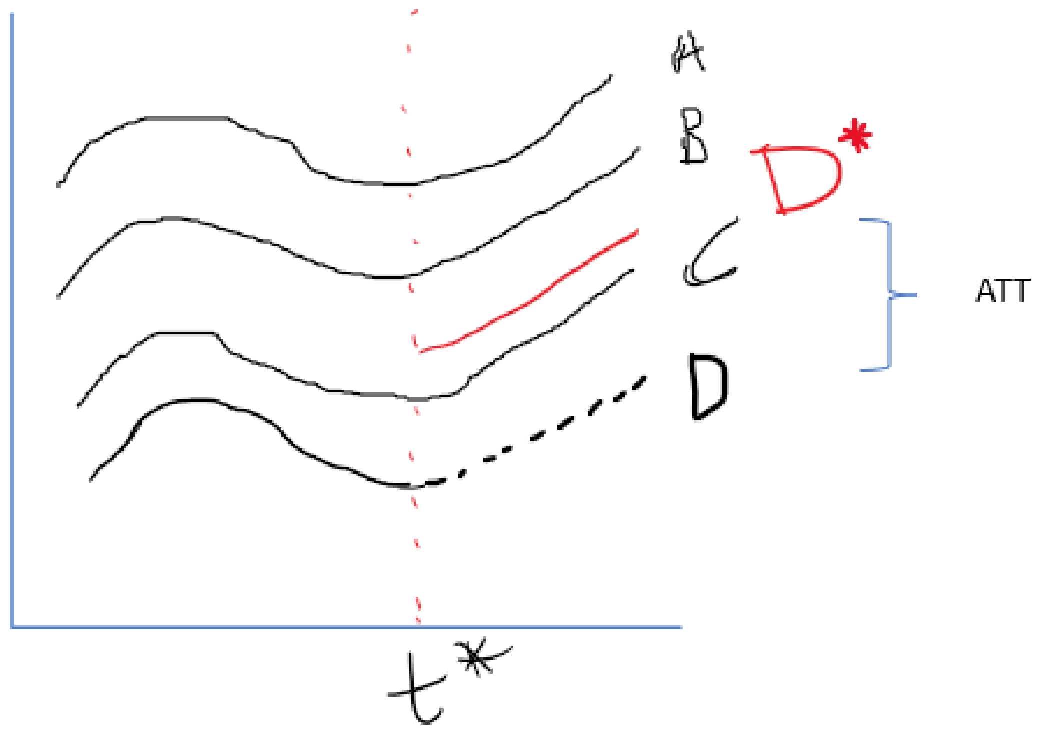
- Two periods (pre, post)
- Two groups (treatment, control)
- Estimand: ATT
- Assumption needed for identification: parallel trends
- Equation:

$$y_{it} = \alpha + \delta_{TRT} + \lambda_{POST} + \beta D_{it} + e_{it}$$

  - δ_{TRT} — treatment group dummy
  - λ_{POST} — post-period dummy
  - D_{it} — an interaction between these



Parallel trends with more than two periods



Regression DD – Two-Way Fixed Effects

- Can extend this idea to multiple time periods, and variable treatment timing
- *Note: covariates can be added. Most of the newly developed models can also incorporate covariates. But I will not cover that issue today.*
- Basic TWFE equation:

$$y_{it} = \delta_i + \lambda_t + \beta D_{it} + e_{it}$$

- δ_i and λ_t are unit and time fixed effects, D_{it} is a treatment indicator
- Complication: heterogeneous treatment effects: $\beta_i \neq \beta_j$ for units $i \neq j$

DID with Variation in Treatment Timing

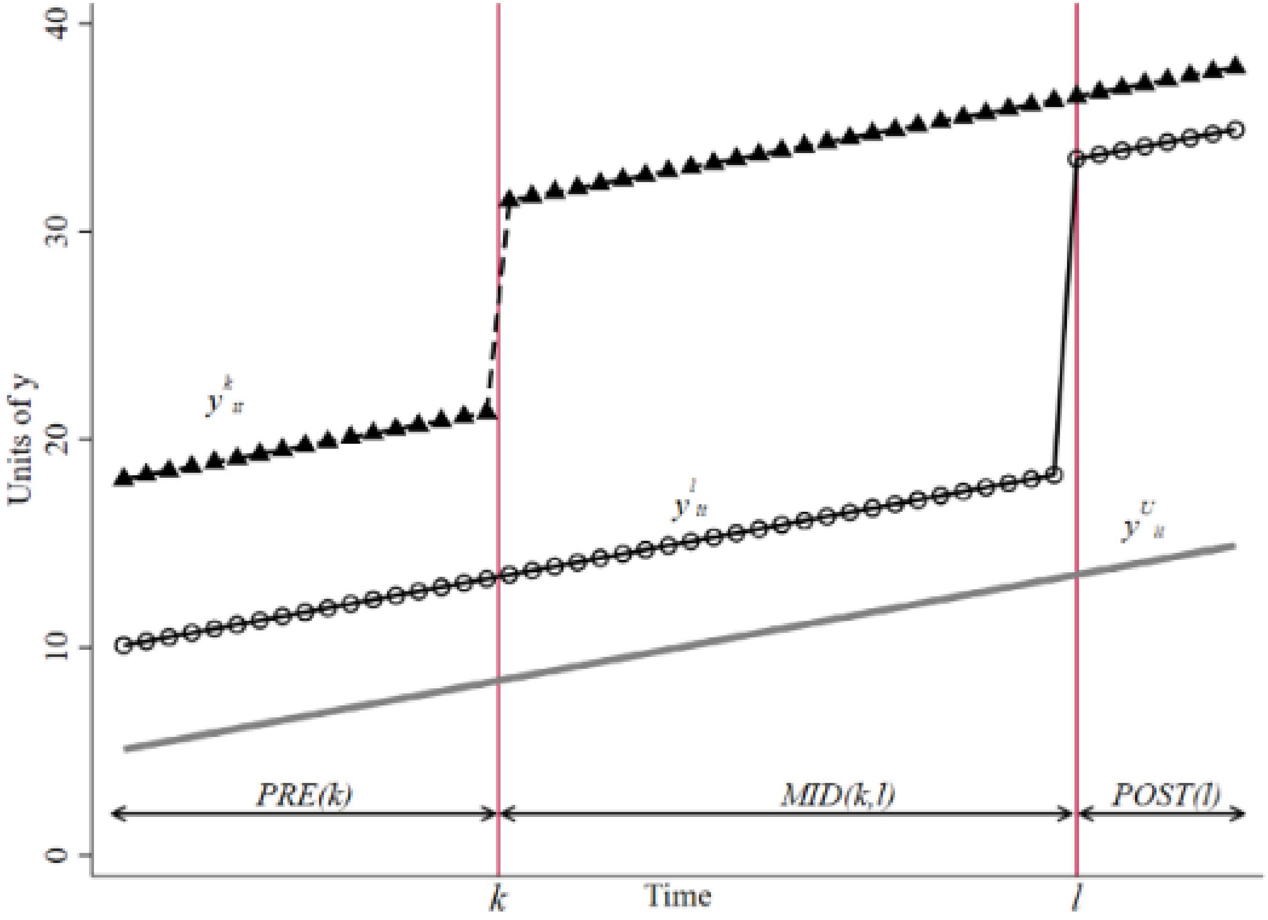
- Goodman-Bacon (2021)
- Underlying issues: dynamic treatment effects, timing variation
- Assumptions needed for TWFE:
  - When there is no variation in treatment timing, TWFE is an unbiased estimator for the ATT.
  - When there is variation in treatment timing, but <u>treatment effects do not change over time</u>, TWFE is an unbiased estimator for a *variance-weighted* average of treatment effects (VWATT), where variance is the variance of the treatment dummy—highest for units treated in the middle of the panel.

Goodman-Bacon (2021)

- Diff-in-Diff Decomposition Theorem
  - Decomposes the TWFE estimator into a weighted sum of 2x2 DD estimators
  - All weights are positive <u>if</u> treatment effects constant over time. Otherwise there can be negative weights.

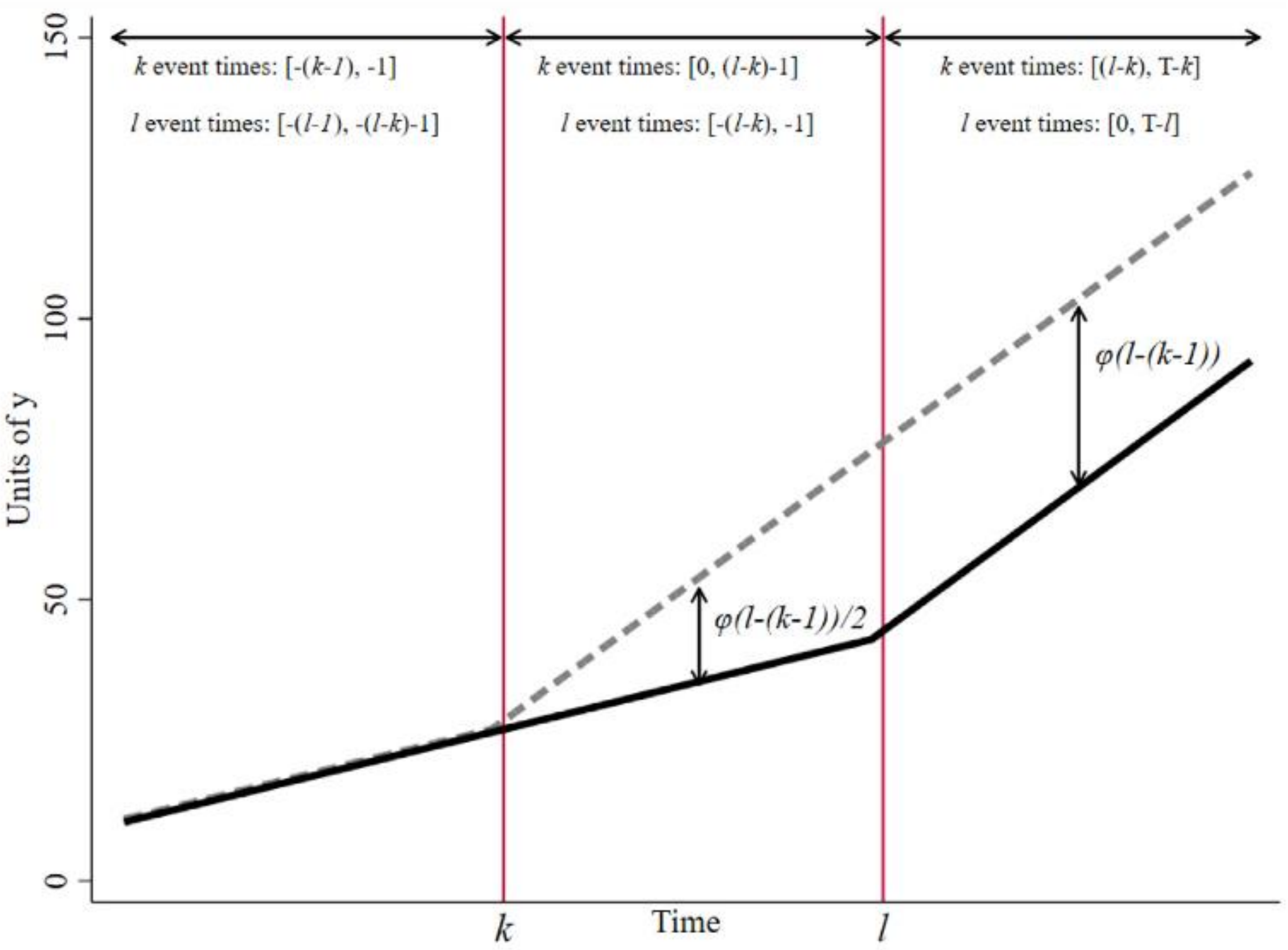
- Key issue: earlier-treated units' post-treatment periods are used as "controls" for later-treated units
- Negative weights are problematic because they can produce estimates outside the convex hull of the actual 2x2 DD estimators.

Early versus late-treated units



Source: Goodman-Bacon (2021)

Non-constant treatment effects



Source: Goodman-Bacon (2021)

DID with Variation in Treatment Timing

- Other helpful decompositions of the TWFE estimand and intuitive explanations can be found in:
  - Imai and Kim (2021) – decomposition in a matching framework
  - de Chaisemartin and D'HaultfŒuille (2020) – TWFE as a weighted average of individual treatment effects in DID
  - Borusyak et al. (2023) – TWFE as a weighted average of individual treatment effects in event studies
  - Gardner et al. (2023) – discussion of bias in TWFE coming from contamination of estimated fixed effects by treated units

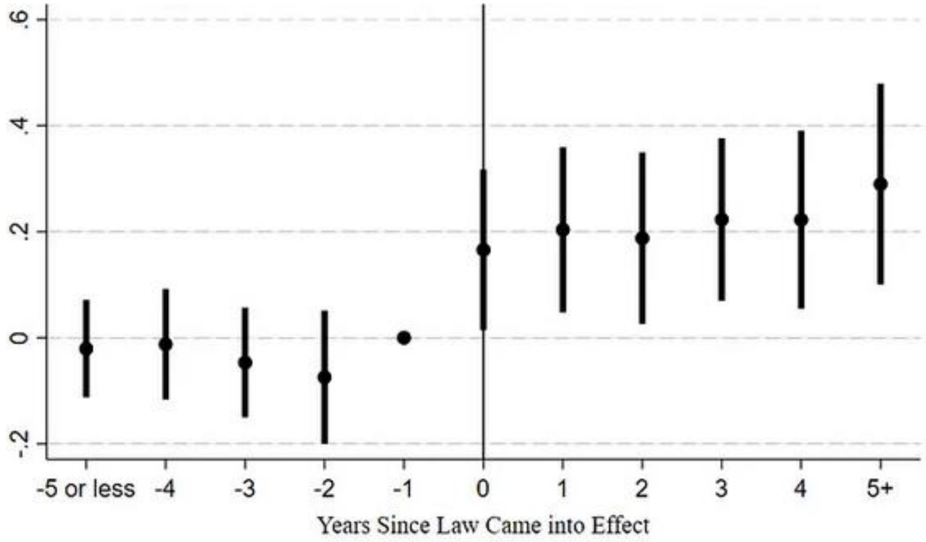
Event Studies

- "Event studies" are a generalization of DID that allow the treatment effects to vary over (event) time
- Very similar to synthetic control
- Equation:

$$y_{it} = \delta_i + \lambda_t + \sum_{k=1}^{N_{pre}} \beta_k^{PRE} D_{it}^{PRE,k} + \sum_{j=0}^{N_{post}} \beta_k^{POST} D_{it}^{POST,j} + e_{it}$$

- $D_{it}^{PRE,k}$ is an indicator for an observation of unit being k periods prior to the treatment time, and $D_{it}^{POST,j}$ is similar for post-treatment periods.
- The researcher can decide how many pre- and post- period indicators to include, which are captured in N_{pre} and N_{post}.
- The β coefficients measure event-time specific treatment effects.

Typical Event Study Plot

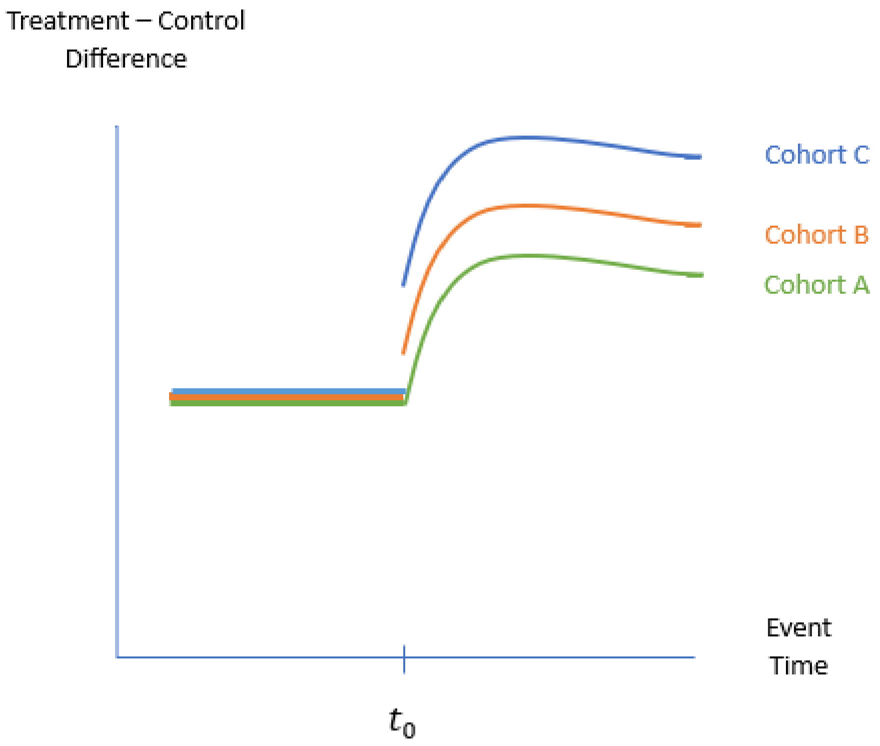


Sun and Abraham (2020)

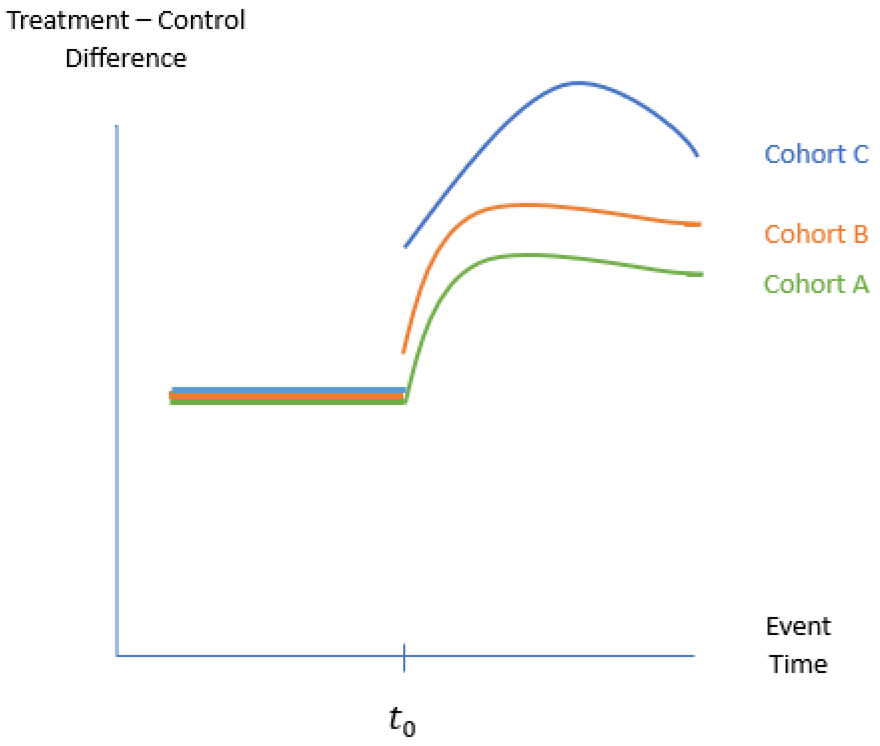
- Event-time-specific dynamic treatment effects contaminated by other periods
- Major issue: interpretation of pre-treatment coefficients as evidence for (or against) parallel pre-trends. This is **problematic**!
- Conditions needed to avoid this:
  - Parallel trends
  - No anticipation
  - Homogenous treatment effect dynamics across units (stronger than the static case)

Homogeneous dynamic treatment effects

Homogeneous – Good



Heterogenous – Bad



Sun and Abraham’s Solution

1. Specify a regression with separate event study coefficients for each cohort

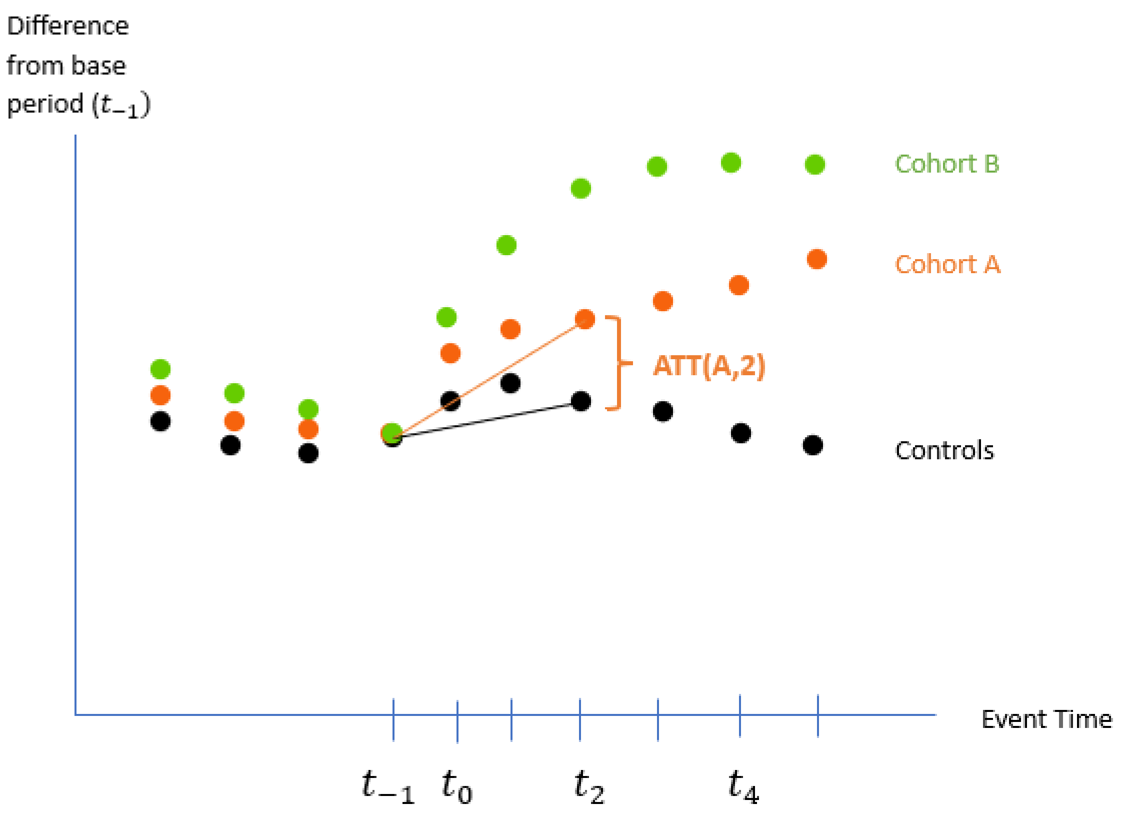
2. Reweight coefficients using the cohort shares, so units are given uniform weight instead of the weight depending on treatment timing

Callaway and Sant'Anna (2021)

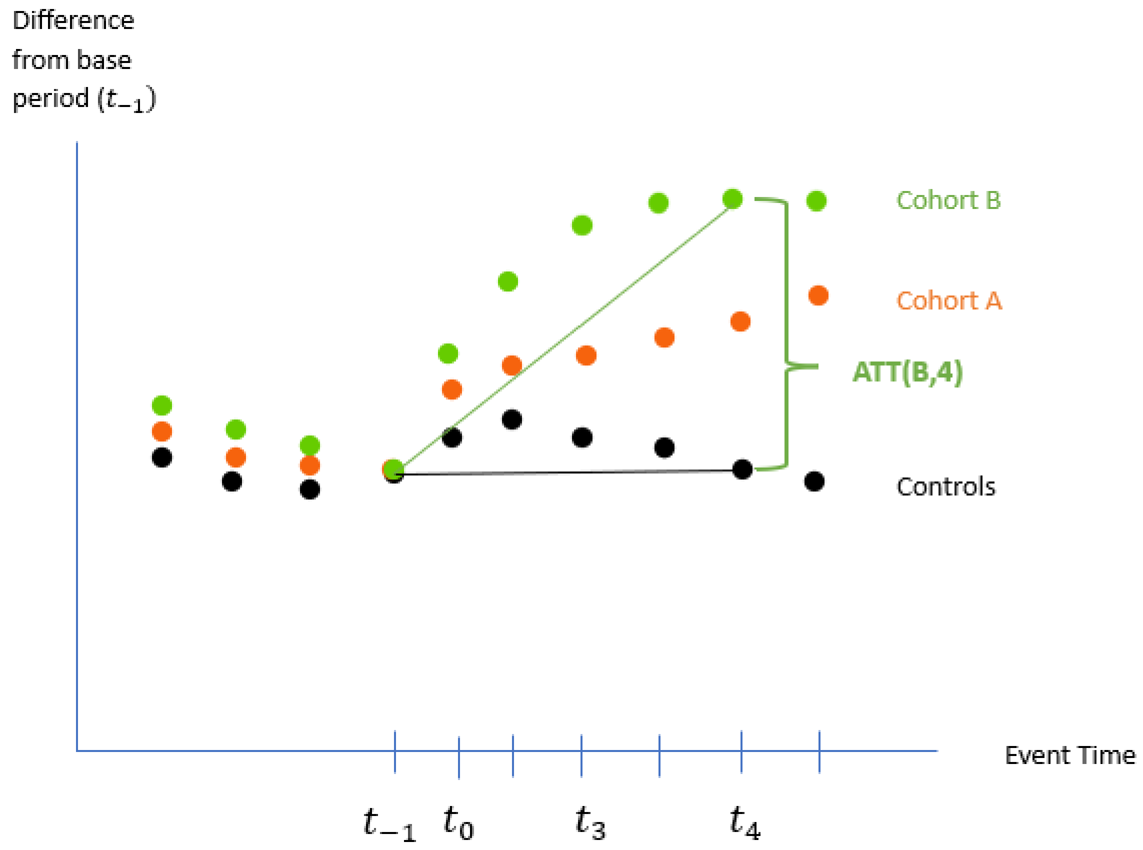
- New Perspective – build up estimates nonparametrically from basic building blocks
- Building blocks – ATT(g,t)
  - Average treatment effect on treated for cohort g in period t
- Can combine ATT(g,t)'s in various ways to estimate DID, event studies, or other parameters of interest

Every cohort and period is compared to the control, to estimate a separate ATT(g,t) parameter

Cohort A, time 2 = ATT(A,2)



Cohort B, time 4 = ATT(B,4)



Aside: covariates and conditional PTA

- An important aspect of C/SA is the doubly robust estimation for achieving parallel trends *conditional on covariates* (conditional PTA or CPTA)

- Three methods proposed to better satisfy parallel trends
  - Outcome regression (OR) – model the evolution of the counterfactual using covariates in a regression model
  - Inverse propensity weighting (IPW) – re-weight control units to better represent treatment units
  - Doubly robust estimation – requires only one of these models to be correctly specified

Other Proposed Solutions

- Callaway & Sant'Anna (2021) – semiparametric DID/ event study with doubly-robust covariate adjustment
    - Sant'Anna and Zhou (2020) – doubly robust estimator for DID
    - Abadie (2005) – IPW for DID
    - Heckman et al. (1997,1998) – outcome regression
- Sun and Abraham (2020) – event study using regression and interacted fixed effects
- Borusyuk et al. (2023) – impute counterfactual using untreated observations
- de Chaisemartin and D'HaultfŒuille (2020) – compute ATT as an average of 2x2 ATT's across all treatment cohorts, using only clean controls
- Gardner et al. (2023) – Two-stage DID/ event study – estimate fixed effects on untreated observations, use them to residualize outcome, then regress residualized outcome on treatment
- Wooldridge (2021) – two-way Mundlak regression – use average of pre-treatment controls in regression
- Stacked DID (e.g. Cengiz et al. 2019) – make a new dataset for each treatment cohort (including only one cohort plus all control units), run separate TWFE regressions for each, and combine the results
- Synthetic DID (Arkhangelsky et al. 2021) – weight units and time periods based on representativeness of the sample

How do I choose which model?

- Most of these use the same assumptions for identification
- Methods of calculating standard errors are different
- Covariates enter in different ways
- Choose the one that makes the most sense for your application, and balances realistic parallel trend assumptions with statistical power
- Don’t include covariates unless you need to, and you know exactly what adding them does to your model and assumptions

Additional Topics and important papers

- Covariates (most models can incorporate)
  - Caetano & Callaway (2023)
- Parallel trends in depth
  - Freyaldenhoven et al. (2019), Marcus & Sant’Anna (2021), Roth (2022), Rambachan & Roth (2023)
- TWFE Diagnostics
  - Goodman-Bacon (2021), Jakiela (2021)
- Continuous treatment and fuzzy DID
  - Callaway et al. (2021), de Chaisemartin & D’HaultfŒuille (2018)
- Inference
  - Each paper uses different assumptions and methods to estimate standard errors
- Synthetic control
  - Very closely related to DID
  - Abadie & Gardeazabal (2003); Abadie, Diamond, Hainmueller (2010)
  - Abadie (2021) – review paper; Arkhangelsky et al. (2021) – synthetic DID

Software for DID

- Asjav Naqvi's DID repository lists most of the packages (and has other helpful resources) https://asjadnaqvi.github.io/DiD/
- Table 2 in Roth et al. (2023) lists packages in R and/or Stata for most of these methods https://www.jonathandroth.com/assets/files/DiD_Review_Paper.pdf
- Stata 18 has native implementation of Callaway & Sant'Anna estimator and other DID and causal analysis features https://www.stata.com/new-in-stata/heterogeneous-difference-in-differences/

Software considerations

- Most of the packages have both R and Stata versions, usually one is better than the other. Best to learn both programs!
  - -did_imputation- is very fully-featured in Stata
  - -csdid- (Stata) or -did- (R) is good in both, but more fully-featured in R
  - -did2s- is much better in R

- Many packages are updated frequently with bug fixes and performance improvements

Additional Resources (1)

- Scott Cunningham has in-depth articles about many of these studies on his substack https://causalinf.substack.com/
- The Econometrics Frontier Group (E-FroG) at APEC has studied many of these papers and has materials on a Google Drive
- Andrew Goodman-Bacon and Janna Johnson gave a workshop on the new DID methods in early 2023, materials are available here (need to request access)
- Two useful seminar series available on the web:
  - https://taylorjwright.github.io/did-reading-group/
  - https://www.chloeneast.com/metrics-discussions.html
- List of applied micro methods papers from Christine Cai: https://christinecai.github.io/PublicGoods/applied_micro_methods.pdf

Additional Resources (2)

- Blog post from Callaway on covariates in DID: https://bcallaway11.github.io/posts/fms-did-time-varying-covariates
- Great site with in-depth DID tutorials: https://diff.healthpolicydatascience.org/
- Jonathan Roth's DID resources: https://www.jonathandroth.com/did-resources/
- Andrew Goodman-Bacon's FAQ on his paper: https://drive.google.com/file/d/1D9t-nQt__tw-1-k6BEAtok1ul1hoQJGm/view
- Review papers: Baker et al. (2021), Roth et al. (2023), de Chaisemartin & D'Haultfoeuille (2022)

References (1)

- Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, *72*(1), 1–19. https://doi.org/10.1111/0034-6527.00321
- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, *105*(490), 493–505. https://doi.org/10.1198/jasa.2009.ap08746
- Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *The American Economic Review*, *93*(1), 113–132. https://doi.org/10.1257/000282803321455188
- Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference in differences. *American Economic Review*, *111*(12), 4088–4118. https://doi.org/10.1257/aer.20190159
- Baker, A., Larcker, D.F., Wang, C.C.Y., 2021. How Much Should We Trust Staggered Difference-In-Differences Estimates? (SSRN Scholarly Paper No. ID 3794018). Social Science Research Network, Rochester, NY. https://doi.org/10.2139/ssrn.3794018
- Borusyak, K., Jaravel, X., Spiess, J., 2023. Revisiting Event Study Designs: Robust and Efficient Estimation. Working Paper. https://doi.org/10.48550/arXiv.2108.12419
- Caetano, C., Callaway, B., 2023. Difference-in-Differences with Time-Varying Covariates in the Parallel Trends Assumption. https://doi.org/10.48550/arXiv.2202.02903
- Callaway, B., Goodman-Bacon, A., Sant'Anna, P.H.C., 2021. Difference-in-Differences with a Continuous Treatment. https://doi.org/10.48550/arXiv.2107.02637
- Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230.

References (2)

- Cengiz, D., Dube, A., Lindner, A., Zipperer, B., 2019. The Effect of Minimum Wages on Low-Wage Jobs*. The Quarterly Journal of Economics 134, 1405–1454. https://doi.org/10.1093/qje/qjz014
- de Chaisemartin, C., D'HaultfŒuille, X., 2018. Fuzzy Differences-in-Differences. The Review of Economic Studies 85, 999–1028. https://doi.org/10.1093/restud/rdx049
- de Chaisemartin, C., D'Haultfœuille, X., 2020. Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. American Economic Review 110, 2964–2996. https://doi.org/10.1257/aer.20181169
- de Chaisemartin, C., D'Haultfoeuille, X., 2022. Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey. Working Paper Series. https://doi.org/10.3386/w29734
- Freyaldenhoven, S., Hansen, C., & Shapiro, J. M. (2019). Pre-event trends in the panel Event-Study design. *American Economic Review*, *109*(9), 3307–3338. https://doi.org/10.1257/aer.20180609
- Gardner, J., Thakral, N., To, L., Yap, L., 2023. Two-Stage Differences in Differences. Working Paper.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014
- Heckman, J.J., Ichimura, H., Todd, P.E., 1997. Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. The Review of Economic Studies 64, 605–654. https://doi.org/10.2307/2971733
- Heckman, J.J., Ichimura, H., Todd, P., 1998. Matching As An Econometric Evaluation Estimator. The Review of Economic Studies 65, 261–294. https://doi.org/10.1111/1467-937X.00044
- Imai, K., Kim, I.S., 2021. On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data. Political Analysis 29, 405–415. https://doi.org/10.1017/pan.2020.33

References (3)

- Jakiela, P., 2021. Simple Diagnostics for Two-Way Fixed Effects. https://doi.org/10.48550/arXiv.2103.13229
- Marcus, M., Sant'Anna, P.H.C., 2021. The Role of Parallel Trends in Event Study Settings: An Application to Environmental Economics. Journal of the Association of Environmental and Resource Economists 8, 235–275. https://doi.org/10.1086/711509
- Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, 90(5) 2555-2591. https://doi.org/10.1093/restud/rdad018
- Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, *4*(3), 305–322. https://doi.org/10.1257/aeri.20210236
- Roth, J., Sant'Anna, P.H.C., Bilinski, A., Poe, J., 2023. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. Journal of Econometrics 235, 2218–2244. https://doi.org/10.1016/j.jeconom.2023.03.008
- Sant'Anna, P. H. C., & Zhao, J. B. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, *219*(1), 101–122. https://doi.org/10.1016/j.jeconom.2020.06.003
- Sun, L., Abraham, S., 2020. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics. https://doi.org/10.1016/j.jeconom.2020.09.006
- Wooldridge, J., 2021. Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators.